Application of Artificial Intelligence in Business

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
im Fachbereich Humanwissenschaften der

Universität Osnabrück

vorgelegt von

Faisal Khalil

Supervisors and promotion's committee:

Prof. Dr. rer nat. Gordon Pipa, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Osnabrück, Universität Osnabrück, 2022

To those who seek knowledge for the satisfaction of the mind,
and understanding of the world

Acknowledgements

Throughout the writing process of this thesis, I have received a great deal of support and assistance, without which I would not be able to write this thesis.

First, I would like to thank my supervisor Pro.Dr.rer. Gordon Pipa, for giving me continuous support during my Ph.D. He not only helped me scientifically and conceptually but also helped me a lot morally, emotionally, and in the administrative process of foreign office as well. I am also thankful to Gabi Pipa who helped me and my family in settling down in the city that was a great help in difficult times. During my research phase, Prof.Pipa raised a lot of questions and gave me a lot of good suggestions and directions that helped me in formulating the research hypotheses and research paradigm.

Secondly, I am really thankful to the Higher education commission of Pakistan (HEC) which funded the 3 years of the Ph.D. program. I am also thankful to DAAD who has managed and administered the funding process during the funding duration.

I am also thankful to my colleagues at the lab with whom I have been discussing and sharing ideas regularly and they have given me a lot of valuable feedback and suggestions.

Finally, I am thankful to my parents and my family for their continued and unconditional moral, emotional and financial support which enabled me to achieve a lot of success in my life.

Abstract

In this cumulative thesis, I have applied artificial neural network-based (ANNs) methods to solve financial forecasting and business management problems. Purposefully, I have selected three different sectors and identified a practical problem set to apply AI-based methods. First is, the financial forecasting problem where I have collected hourly based specific stocks returns and for the same hour I have parsed, collected, and aggregated textual information from thousands of different news feed, blogs, social media, and print media, converting the information into sentiment index. Subsequently, using the Long Short term Memory model (LSTM) six models I have predicted the stock market returns taking into account the news analytics. The second case study is financial distress prediction with the help of ANNS. In this case, we have parsed a lot of text from SEC- tax filings to calculate sentiment analysis, additionally, we have also computed financial health-related ratios for the selected USA stock exchange companies. As the next step, we have labeled the data with healthy and non-healthy classifications for the selected time window. We have implemented many different, simple ANNS, CNNs, and LSTMs AI models and found that LSTM outperforms all other models and is able to detect financial distress in future time periods very accurately. On the same footprints, we have performed a third case study that is related to business text generation and prediction. In this specific study, we have collected 2.3 billion business text tokens to train transformer bases ANNS. The model is able to generate grammatically and contextually correct business-related information for the given topic. Conclusively study shows that AI is transforming and converging the

business and research in the field to the next level.

Preface

This thesis is written for the completion of the degree requirement at cognitive science, University Osnabrueck. The thesis is written in a cumulative style by combining research output, which consisted of two research papers and one study project (not included in the thesis). The main focus of the thesis is how Artificial intelligence as a tool helps businesses and finance in adopting new methods and technologies. The research article and study project are included as it is in the body of the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the introduction part, we will describe the origin of business, a brief history of the technology, and the current need for technology in business and commerce followed by current cutting-edge technologies and trends in business premises. At the end of the introduction, we will briefly discuss which problems we have tried to focus on and how Artificial intelligence, in general, could be applied in business and finance. We have supplied research papers and a study project as a case studies to highlight the effectiveness of AI to solve the current industrial problems.

1.0.1 Origin of Business

Before studying, analyzing, or researching the technological application of the business, it is relevant and important to study the historical, political, cultural, and economic context of the business. Many historians give focus on the modern history of business but in fact, business history has its roots in ancient Greek, Romans, and even the civilizations of Mesopotamia. The immediate question arises when actually a business for the sake of profit or desire and pursuit of profit is started in human beings. it is hard to find a clear answer but, manufacturing and trade was existing before the period of Mesopotamia urban society but the actual origin of profit-making and, its desire is unknown. The reason is that when we go beyond the urban civilization in the times of Mesopotamia, people were thinking of them as part of a wholistic communal system where they were unable to recognize themselves as separate individuals who, in principle, carry their personal interests. As evidence from the studies of anthropologists, the life of prehistoric peoples in terms of social, economic, and religious aspects was indistinguishable. Their life was strongly influenced by priest or shamans or other religious icons and figures. They more and less struggle for survival, and their thoughts and, responsibilities belonged to those religious figures or gods. So, pursuing any profit-oriented interest for people of this era was difficult. In the Sumerian era, for example, the fertile crescent was warmer as compared to the other parts of the world, which triggered heavy migration of birds, humans, and animals to that region. Some popular cereal-based crops were native to that land. It was easy to cultivate the land with weak tools. So human population settled on the crescent and the Nile on one side, and the Persian Gulf on the other end. So, people adopted cultivation and farming as the focus. Leaders of the community, who were mostly priests and had power due to closeness to God, and give them safety from natural disasters, have started to build, organize maintain good functional irrigation system and protect it from the invaders.

Around 3100 BCE, the art of making bronze is discovered with help of tin. Bronze gives access to more advanced and lethal weapons that increase the defense of priests' states, which led to significant efforts to promote making kings, armies, and arsenals by priests. Kings started developing their military. Bronze armor shields and other battle tools would make them feel more protected from invaders. Thus, the expenses of so-called states of priests increased, and they introduced the system of collection of grains out of quota given to citizens by the state. That was the early implementation of the tax system.

With the passage of time, two main civilizations along with many other neighboring civilizations emerge in the Early Bronze Age, most renowned were the Egyptians and Sumerians. Egypt is ruled by the pharos and the villages were mostly concentrated along the Neil River. Neil has fertile land which was very suitable for cultivation. Moreover, Egypt comprises mostly the desert area and is enriched with gold, copper, and other metals. On the other hand, Sumerians were living in scattered settlements in their region. Due to spending most of the time in the watery irrigation systems, some diseases caught the peasants, and they become fewer, weaker, and slow, which led them to switch their occupation from framing to herding. They start domesticating goats and sheep. So, they started processing the wool and heed of animals to make different items. Sumerians had enough labor, they were good at building wood-related items, ships, temples, and tombs. So, over the period they specialized in manufacturing and clothing but at the same time lacked in metals and gold. So, these necessities give birth to trade.

So, Sumerians started making journeys across all civilizations to exchange their specialties and fill the gap with what they needed. Traveling at that time was very tedious, slow, and full of danger and uncertainties. So, these specialized traders are called "Tamarkus" in Samara. Tamarkus became particularly important for the community, as the import and export of goods were highly dependent on their skill and bargaining power. So, naturally, they got closer to rulers and were awarded generously in the return for trade missions. So, they become influential in society and realize their self-interest or profit by offering their expertise. So, that is how early profit-oriented businesses get started.

1.0.2 A brief history of technology:

To pose the problem set and research question we are trying to study in this thesis, it is especially important to shed light on the evolution of technology into business and finance over the last two centuries.

The modern history of technology can be divided into two clear parts, namely, the pre-industrial revolution and the post-industrial revolution. We will be discussing the post-industrial area, along with a brief overview of the pre-industrial era. According to many anthropologists, primates starts the journey of innovation and technology almost 350,000 hundred years ago. When they came down from trees and adapted their bodies to two feet and spared the other two for grabbing and hunting. Over the passage of time, they start consuming meat and other small animals, fruits leaves, eggs, nuts, beans, and grains. Around 10 to 12 thousand years ago, glaciers started melting due to which the weather got drier, which led to the scarcity of water and fewer options for cultivation. So, they had to choose from two options, moving to suitable places with their belongings like cattle and sheep, etc. or they had to choose some places near water resources and do permanent settlements. So, the Homo sapiens who permanently settled invented agriculture. Settling in permanent places created villages, and villages created the chain of needs that enhanced over the passage of time. So, the quest of utilizing fewer resources and getting maximum gain from them gave birth to the thought of innovation. Let us look at how different areas of science have evolved during the pre-and post-industrial eras.

Meat and other starchy food releases more energy when heated. So, Homo sapiens needed fat, so they started thinking about cooking the food, and for cooking they needed fire. In prehistory, humans learn to make fire. The first invention related to fire was a bow drill. Bow drill was an effortless way to create fire with friction method instead of revolving the wooden stick with hands using a bow and connected shaft

When people started cooking is unknown but in caves of current China, scientists found snail shells and pottery that are said to be 14,500 years ago.

Hunting in groups, and saving the family from the invader and other predators demands lots of fights, fights that cause pain and treatment of pains stated with many religious offerings to make God happy or with trees and plants for example use of willow trees that are enriched with salicylic acid to cure the pain and fiver. The oldest surgery evident is trepidation where a hole in skill is found from a burial in France that is 6500 years ago. Sumerians invented many bronze instruments like scalpels and knives to do surgeries. The code of Hammurabi is the code of medical compensation found in Babylonia.

Similarly, during the excavation archaeologist found many fractured and cured bones evidence of medieval scripture, and medical engraving of surgical tools in Egypt. Animals' bodies work on metabolism, and metabolism depends on water. Metabolic waste is also dependent on water in the shapes of urine and sweating. The process of digestion also needs water, and that is especially important to control the temperature of the human body. Around 10,000 BEC, Homo Sapiens' population increased in North Africa which led to the scarcity of hunting animals. So, H.Sapiens were forced to either move from place to place or do forming and ways to store grains to be used on rainy days. That need gave birth to water management and agriculture. The ancient agriculture system was founded in Ancient Persia in almost 6000 BC. This system, called "Qanat"

Sumerians have used many different techniques to stop the flood and contained the water as the early summer comes with a lot of ice melting and overflow water in the regions. Likewise, Indus valley civilizations had amazingly effective and functional water drainage, reservoir, and storage system in 45BEC . Over the passage of time, the human population grew and trade started in Mesopotamia, the need for mobility emerged as two legs or horses was not able to transport crops, goods, and human. So, the journey started with a floating log that later turned into the barge and the ships. The realistic boat is founded in 4000BC in Egypt. Cattle and horses were domesticated around 2500 and with the invention of the wooden wheel, hope to make wagons that were suitable to daw with castles. Increasing mobility helped the kings to extend their area of control and that led people to settle down even areas other than near rivers and banks and thus villages and towns came into being.

As the population growth of human beings increased, their needs and complexity were magnified. So, the trade and efficient use of storable food and grains needed accounting and management. Thus, in ancient times people have been using several types of shapes, i.e., squares, circles, avoid, and ovals, made of different materials to count the grains and other materials. The evidence of the use of shapes for tokens and numbers is rooted back in 8000 to 3000 BC.

From token and shaping counting the basic arithmetic has been evolved and practically used by the Greece empires. The use of decimal points and positional numbers like those used on the abacus for addition, subtraction, and multiplication were remarkably similar to what is used today. In Greece mathematic zero was not present as a symbol, instead, they were repeating the starting symbol again after the 100th unit place.

Independent to Greece mathematics, at the same time, chines dynasties have additionally discovered and used negative numbers. Hindu-Arabic numerals invented positional notations and decimal system concepts and the new digit zero made mathematics extremely easy and understandable. This Hindu Arabic numeral system is used by Iranian mathematicians and spread to the west afterward.

1.0.3 The transition of human intelligence into artificial intelligence

In this section, we will discuss how human-level cognition transformed into computational methods over the passage of time that led to open avenues for artificial intelligence. How does the brain work? Before we go into the depth of artificial intelligence and deep learning it is important to describe how the brain works.

Neurons and synapses

The brain is made up of billions of cells called neurons. Neurons are the nerve cells in the brain that transmit messages and create thoughts, feelings and movements. A neuron has three parts: dendrites, a cell body, and an axon. Dendrites are branch-like structures that extend out from the cell body. They receive signals from other neurons. The cell body contains the nucleus as well as other structures that support the cell. The axon is the long tail-like structure that sends signals to other neurons. Neurons transmit messages using a combination of electricity and chemicals called neurotransmitters. When the neuron receives a signal from another neuron, the dendrites send the signal to the cell body. If the signal is strong enough, it travels through the axon to the end of the axon where it triggers the release of neurotransmitters. These neurotransmitters travel across a tiny gap called the synaptic cleft and are received by the dendrites of other neurons. This process is called neurotransmission. Neurotransmitters Neurotransmitters are chemicals released by neurons in the brain. They are responsible for most of the communication that happens in the brain. The two most common neurotransmitters are dopamine and serotonin. Dopamine is responsible for the brain's pleasure and reward system. It is released when you do something pleasurable like eating, having sex, or taking drugs. Serotonin is a mood stabilizer. It is responsible for feelings of well-being, satisfaction, and happiness The brain also produces chemicals called endorphins. Endorphins are released by the body in response to pain. They reduce the intensity of pain signals, which is why you feel less pain when you exercise. Temporary Neuronal Connections Neurons in the brain are constantly forming and breaking connections with each other. These connections are temporary and can form in response to any stimulus For example if you want to remember a phone number, your brain will form a temporary connection between the neurons associated with the numbers This is why it is easier to remember something if you repeat it a few times Permanent Neuronal Connections The brain can also form more permanent connections between neurons. These connections form due to experience and are responsible for learning and memory. For example, when you first start driving, you may have a hard time remembering to use your turn signal. However, after a few weeks of driving, you will start to remember to use your turn signal without even thinking about it.

our brain is neuroplastic which help memorizing, achieve the task, and search for an alternate path to achieve the goal that makes the connection of the neuron stronger and deterministic. Let's see what neuroplasticity is and how its works:

Neuroplasticity means that over the passage of time how the brain can adapt itself by creating new connections to the neurons. it can be referred to as "brain plasticity," or "neuroplasticity." Aspects of human brains are pliable, meaning they are adaptive and can be altered in response to environmental or structural changes. Neuroplasticity explains how the human brain is able to adapt, master new skills, and store memories after a traumatic brain injury. The brain communicates with other parts of the body using electrochemical signals. These signals are transmitted through a structure in the brain called a synapse. Stimulating neural pathways through a repetitive memory-forming cognitive function (such as studying or practicing) strengthens synaptic communication between neurons, which results in an increase in neuroplasticity. Additionally, neuroplasticity can occur naturally as we undergo different experiences, changes in the brain may also be activated through neuroplasticity exercises and cognitive training. Research suggests that neuroplasticity may help patients recover from traumatic injuries. Surely, that depends on the extent of the damage, but the process of cognitive rehabilitation is so strong that they rewire and fix the natural communication process and structure that is ultimately very improt for the overall emotional and cognitive health of human beings. Neuroplasticity in stroke patients has been widely researched: after a stroke, certain parts of the brain are impaired and damaged; these areas include those responsible for movement, speech, and communication. When someone suffers a brain injury, neurons in the affected brain regions die and neural pathways become dormant. In order for healing to occur in these areas after an injury, it is vital that individuals undergo rehabilitation as soon as possible so that they can learn how to perform tasks like walking or talking again.

After discussing how primarily neurons, synapsis and neuroplasticity does work let's briefly see how scientists could understand the mechanism of the brain and apply this knowledge to create a new field of Artificial intelligence and deep learning.

Neurons to Artificial Neurons

Artificial neural networks are computational models that work in parallel, unlike computers which have a single processor. Artificial neural networks consist of multiple simple processors that can act alongside one another to model changing systems or to capture the change in the system. Neural networks follow a dynamic computational structure, and do not abide by a simple process to derive the desired output; instead, they combine multiple weighted inputs with reference to a certain threshold and activation function and give out the final value. Neural networks consist of layers of parallel computing processes. Each layer multiplies the number of inputs by an originally established weight, resulting in what is called the internal value of an operation. The internal value further changes or is enhanced and forwarded to the activation function that again maps the input to the output. The output of this layer again becomes the input of the next layer until it reaches the end of the layer called the output layer, where it sees the Truth provided to it. After seeing the true value it tries to map how far or wrong his guess is from the provided truth and comes back and adjusts or returns the weight of each layer and guesses again how near it can reach the truth value. So it repeats itself time and again to minimize this error in guessing the true value. The process is called training and how it is mathematically done is called the backpropagation function. Analogies are useful in understanding neural networks. Learning is similar to how we learn in our normal lives, and it involves performing an action, receiving feedback from a trainer or coach about how to get better at a certain task, and adjusting our behaviors accordingly. Similarly, neural networks require training to describe what should have been produced as a response to the input; this process is called backpropagation through the network. Each layer of the network computes an error value that is fed back through the system and used to adjust the threshold and weights for the next input. In this way, the error keeps becoming marginally lesser each time we run these simulations as we learn how to analyze values correctly. Backpropagation is a procedure that continually adjusts the weights of an artificial neural network in order to minimize errors. After backpropagation, the network is allowed to run without adjustments and may be applied using adjusted weights and thresholds as guidelines. Since the neural network's output statements are not verified during backpropagation, it is necessary to make adjustments during a new training session or leave them as is for the network to run. Many adjustments may need to be made since the network consists of a great number of variables that must be precise for it to function properly. A basic example of an artificial neural network can be examined by teaching the network to convert text to speech. The network should be used on multiple articles and paragraphs, with the training phase consisting of adjusting the parameters and threshold value of the network until it is able to improve the accuracy of its guess. The network may then be tested on new articles in order to determine if it can truly convert text to proper speech. Mathematical and statistical problems such as speech synthesis and recognition, face recognition and prediction, nonlinear system modeling, and pattern classification can be solved using network models.

Types of the artificial neural networks

There are 3 main types of neural networks: Feedforward neural networks, Recurrent neural networks, and Convolutional neural networks.

Feedforward Neural Networks Feedforward neural networks are the simplest type of neural network. They consist of layers of neurons. Each neuron takes a certain number of inputs and produces a single output. The output is calculated by multiplying each of the inputs by a weight. The weights are generated randomly. The outputs from the first layer of neurons are used as inputs for the second layer of neurons. The outputs from the second layer of neurons are used as inputs for the third layer of neurons, and so on. The final layer of neurons is the output layer. The outputs from all of the neurons in the final layer are used to produce the final output.

Recurrent Neural Networks Recurrent neural networks are a more advanced type of neural network. They can be thought of as being like feedforward neural networks, but with a twist. The outputs from the neurons in the final layer are fed back into the network, to be used as inputs for the neurons in the first layer. This is repeated for every layer in the network. The inputs to the neurons in the first layer are a combination of the outputs from the neurons in the previous layer, and the outputs from the neurons in the previous layer of the previous layer. This is repeated for every layer in the network. As well as having an output layer, recurrent neural networks also have an input layer. This layer is used to input data into the network. For example, if the network was trying to predict the stock market, the input layer would be used to input data about the current state of the stock market.

Convolutional Neural Networks Convolutional neural networks are the most advanced type of neural network. They are based on the same principles as recurrent neural networks but use a different technique to calculate the outputs of the neurons in each layer. Instead of calculating the output of each neuron by multiplying the inputs by weight, the output of each neuron is calculated by multiplying the inputs by a 'kernel'. The kernel is a set of numbers, usually between 0 and 1. The kernel is convolved with the inputs to produce the output of the neuron. The kernel is then moved to a new position within the inputs, and the process is repeated. This is repeated until the kernel has been moved to every position within the inputs. The outputs from the neurons are then fed into the next layer, to be used as the inputs for the neurons in that layer. The kernel is then convolved with the inputs to produce the output of the neurons in that layer. This process continues until the kernel has been convolved with the inputs for every layer in the network.

How ML,AI, CV and Deep Learning hlep Industries?

ML and business

Machine learning (ML) is a powerful tool that can help extract meaningful insights from raw data to quickly solve complex business problems. ML algorithms learn from the data iteratively and allow computers to find different types of hidden insights without being explicitly programmed to do so. ML is evolving at such a rapid rate, driven mainly by new computing technologies. Machine learning can help businesses to improve their efficiency and scalability. The tools of artificial intelligence and various ML algorithms have become increasingly popular in business analytics. The factors of growing data volume, easy availability of data, cheaper and faster computational processing, and affordable data storage have led to a significant increase in machine learning. As a result, organizations can now benefit from understanding how bsinesses can use machine learning and apply it to their own processes.

ML can help extract meaningful information from large data sets. When implemented correctly, ML can provide solutions to various business complexities and predict complex customer behaviors. We have seen major technology giants, such as Google, Amazon, Microsoft, etc., develop their own Cloud Machine Learning platforms. Some of the ways in which ML can benefit your business are listed below:

Machine learning can automate tasks that are time-consuming and repetitive, such as data entry and analysis. Machine learning can help you make predictions about future trends and behaviors, based on past data. Machine learning can help you personalize your products and services for each customer, based on their individual needs and preferences. Machine learning can help you detect fraudulent activities, such as fraudulent credit card transactions or insurance claims. Machine learning can help you improve the efficiency of your operations by identifying inefficiencies and bottlenecks. Machine learning can help you make better decisions about pricing, by analyzing customer behavior and trends. Machine learning can help you monitor and optimize your marketing campaigns, by tracking customer engagement and click-through rates. Machine learning can help you understand and analyze customer feedback, to identify areas of improvement for your products and services. Machine learning can help you predict which new products or services your customers might be interested in, based on their past behavior. Customer lifetime value prediction and customer segmentation are some of the major challenges faced by marketers today. Companies have access to a huge amount of data, which can be effectively used to derive meaningful business insights. ML and data mining can help businesses predict customer behaviors, and purchasing patterns, and help in sending the best possible offers to individual customers, based on their browsing and purchase histories.

Deep Learning and Business

There are some possible areas where deep learning is already progressing and increasing the efficiency of the business. For example, in the corporate financial sector with help of blockchain technology, autonomous driving, and health care for diseases diagnose with AI or image segregation. Content filtering and suggestions for the news and media industry. chatbots and virtual assistance systems that mainly fall in customer service with help of AI, Machine translation is another option.

Autonomous Driving

The most promising and fastly growing area where we can see the application of AI is self-driving cars. specifically for business self-driving vehicles can reshape the transportation and logistic industry. In general, there are different use cases of transportation concerning business. For example, public transport like buses and trams can be assisted with fully autonomous or partially autonomous with high assistance control processes. Similarly, Uber is planning to implement autonomous driving in car sharing or taxi services. In general autonomous driving has many benefits, for example, reduction of cost, 24 hours of services without the risk of mental and physical fatigue, with driverless platoons concept less congestion and traffic jams on road. In a nutshell, autonomous vehicles assist businesses in many different ways which lead to an increase in the efficiency of the business. Apart from the passenger transport business, autonomous vehicles can be used for logistic purposes that shrink the delivery time to end customers and create more customer satisfaction. Before we go into details about how actually autonomous driving works and what important models and deep learning innovations are crucial for further development in the field, there are some significant challenges to this industry that are necessary to discuss. For example, the most challenging task is accuracy, reliability, and infrastructural compatibility and capability. For example, different countries follow the different sides of steering that change the overall lane system for oncoming traffic and right-left turn. Another challenge is that every country has its own speed limits, laws, road of signs and different conditions is. Based on such issues the overall required autonomous control accuracy rate must be extremely high as it involves human life. Since it involves human life, it's a very hard decision for the regulatory authorities and governments to allow such operations. A minor incidence can trigger public anger and consequently cause the breakdown of this business model. AVs (autonomous vehicles) cannot make advancements alone as many different technologies are directly linked to the success of self-driving vehicles. For instance, camera technologies, GMS, telecommunication technologies, cloud computational platforms, availability of real-time IoT-based sensor data, data latency, data transfer, and detection and integration into AVs. There are some other administrative issues there for example ownership of indemnity and Demange caused, determining the fault of the party and all liked matters to the insurance and claim is also a sort of big challenge for AVs.

The next challenge is related to the implementation of the concept. It is important that the governments and other regulatory authorities agree to maintain and build the bare minimum and needed infrastructure aligned to the necessities of AVs operations i.e roads, lane markers,s, and signs visibility need a lot of additional financial budget. For wealthy nations might be it is not a big problem but for developing countries as well as rural areas self-driving vehicles might not be on the priority list. After talking about the potential benefit which we can reap and some tough challenges to the sector which the researcher community is closely intact with, let's look at how deep learning evolved up to AVs, how its works, and what path-breaking models the magic of the autonomous vehicles.

The journey of self-driving cars started back in 1986 when Dean Pomerleau buit is test vehicle and Carnegie Mellon University started to develop PANS: A Portable Navigation Platform they later named it NAVLAB ., a mathematical model that could assist in steering any passenger or personal vehicle. If you go into the depth of the paper that was published later on in IEEE following the summary of how

the system works.
The approach of Dean to personal test vehicles works quite differently compared to that of NAVLAB ALV (autonomous land vehicles) vision algorithms but is fundamentally derived from the same concept. The ALV vision algorithm works on input images that comes directly from the video stream of the good definition camera installed on the vehicle. Subsequently, the image is split into still frames and the algorithm then tries to segregate the pixel of different color groups based on certain color thresholding. For example, the road is normally looked grayish and lanes are marked with white dashed or solid lines. Now mathematically, you can separate the significantly different types of color from each other and thus you can draw a boundary around the road and non-road elements in the frame. So NAVLAB (they name the experimental vehicle this name) has done the same. After the image is separated based on road area and non-road area the next step is to convert the 2D image into 3D space and calculate the angle which should center around the visible boundary. So, the problem can be seen as a big mathematical function that takes the image bytes 3D converted image data as an input and maps the angle that is centered in the visible boundaries of the road. In the next step, the calculations were given to steering control to automatically steer the vehicle. The same year at the same university another team of researchers under the leadership of prof. Geoffery Hinton published a paper titled Learning representation from backpropagation errors. That paper was really a resurrection of artificial intelligence that was buried due to recession and some critiques given on the approach via book published that claim that there is in fact not any solid way to give intelligence to the machines. As in result of that book, during the period of 1950 to 1980, not much work has been done on Machine learning and the research community wasn't convinced that machine learning has the potential to solve the problem in a self-intelligent way. So that paper reignite the flare and researchers started implementing the backpropagation process that ultimately led to Artificial neural network development. So coming back to the topic independent from

the NAVLAB team Dean Pomerleau has replaced the mathematical function that takes input and gives back the theta with that backpropagation algorithm. Dean has collected real human driver steering data for certain roads and provided that data to the backpropagation algorithm to solve the relationship and the mapping between the input value, the image in this case, and the given output data that was steering theta in this case. He came to the conclusion that after some training the system reaches very close to real human-level control. This means that the backpropagation algorithm figures out for each epoch how to adjust the weight attached to each node to minimize the error function and in the end, it achieved very good accuracy.
That is how the first ever deep learning algorithm for self-driving cars came into being. Well, that is not the end of the story, rather there were some serious issues that needed to be addressed even after the implementation of the backpropagation algorithm. The Dean's main system could just assist the steering and the computational power needed was way low as compared to what was needed to get the high-quality images frames from the video streaming of the scene and then to predict the theta value based on-road situation. Resultantly, the vehicle could only move very slowly. But after a couple of decades when computation resources get improved, smarter, and faster over the passage of time, that has rejuvenated the concept of the self-driving vehicle. Specifically, after 2016 when Nvidia came up with GPU based processing unit which in fact changed the computational speed in the field forever and gave a new breath to the computational hungry AI models. So computational resources have solved only a tiny bit of the challenge we had. For instance, if the acceleration issue is resolved what about the following questions, when to break and when and how much acceleration should be given to the vehicle in which situation? What to do in case of an emergency braking maneuver is needed and what about the unusual obstacles on the tract? These are some serious questions before industrial-grade autonomous vehicles could be built up.

Let's have look at some pathbreaking concepts in the field of AVs. Well, autonomous driving is not a solo ride of deep learning rather it's a combination of robotics and deep learning. the important bits of autonomous driving could be categorized into these points. It needs perception to view the environment around, it needs to avoid the obstacle it faces with help of a laser or camera, it needs to localize the position, that means where the vehicle is now and it needs the angle from place A to B and how to move from one place to another. So these are three gross levels of self-driving vehicles. The main pillar of a self-driving car is perception, which is how the car sees its surroundings and finds the path and obstacles ahead. AVs uses typically have three sensors i.e. camera for imaging, LIDAR any kind of laser light range sensor to calculate the distance from the obstacle, or any form of RADAR, light ranging based on radioactive waves, and a depth sensor to calculate built the map.

After 2006 when computational resources get ample and sophisticated, the concept of deep neural networks come into the picture, and in the early stages of self-driving car journeys, the LaneNet algorithm gets very popular. The LaneNet was essentially making the segmentation task. The major difference between image segmentation from the traditional deep learning model is that the deep learning model mostly tries to detect the object in the image and at most, where approximately the object boundary is, but the task of segmentation is a dissection of the image at the pixel level and the image is labeled pixel by for the whole image. Thus the labeling task is getting more challenging as well as tedious. Image segmentation did not just come directly but it has to went through the evolution stage of generic level deep learning to highly specialized image segmentation. For example, perception in deep learning first invented convolution neural networks and many different improvements came into the field. For example, YOLO ann SSD were got popular as compared to their counterpart, RCNN, and event faster RCNN because RCNN were fundamentally cutting down the image into multiple pieces and under the hood running separate convolution networks to reach the confidence level to decide what is the object and where does it locate within the image. Thus to avoid this shortcoming researcher has come up with Faster RCNN with some conceptual improvements and different architecture but still, the model was overall very slow and relatively could not get much attention from the research community. Apart from RCNNs being slow and taking more resources but these models are very good at object detection jobs. Chasing the same problem set researchers have come up with the idea of "You look only once " (YOLO) The concept of YOLO was not to cut the image into pieces and then predict but rather just look at the image wholistically at once. The invention of YOLO was a great relief to the task of object detection because it was a much faster algorithm without compromising a lot of accuracy of a task at hand and thus that gives a new horizon to the perception section of autonomous vehicle research.

Following a similar path, self-driving vehicle research is replacing the exing methods of localization and mapping tasks. Deep learning is being used in many different filed and got the attention of many startups and investors. We are hoping that the passenger transport industry, as well as the logistic industry, could be greatly benefited from the improvement on deep learning in the coming days.

AI in corporate finance:

AI is helping the corporate sector to great extent. For example, the corporate sector means mainly the banking sector, any financial institution that offers any financial advisory, or other financial services, insurance companies, credit ratings, and credit counseling services, stock market tradings, and customers' services and overall digital transforming of these industry comes under the umbrella of the financial or corporate sector.
So, deep learning is in general playing a pivotal role in reshaping this sector. The backbone of the financial industry is prediction, the pre-

diction could be in any form, for example, prediction of the market, risk prediction for insurance, bankruptcy prediction for banks, and cash flow predictions for firms are some example of prediction. Now let's see how AI is helping here. For example, banking sector issues credit to its customers not only depending on the income stream of the borrower but also on many other factors and demographics of institutional and non-institutional borrowers. So, with help of millions of previous similar cases AI model can predict the creditworthiness of borrower, the possibility to get the money back, or even the prediction of riskiness involved in the contract. In this way, banks could take extra interest for bearing extra risk. According to a Business Insider report, 75% of respondents at banks with over $100 billion in assets are implementing AI technologies. McKinsey shares that banking and other financial service companies could generate more than $250 billion in value by applying AI technologies in their financial processes. Apart from the need to AI at the technical level, administratively corporate industry has involved a lot of document processing in the shape of contracts, payslips, application Forms, credit rating checks, and many different types of documents. There is a lot of data sitting on these papers which is crucial for the analysis purpose. With help of computer vision technology needed information can be now easily extracted from long thick documents and can process faster without the peril to miss important details. AI algorithms capture the data automatically, process the data even manage the lending operation automatically.

With the advancement of deep learning and natural language processing, it's now possible to create virtual assistants and Robo-financial advisors which round the clock advise a wide range of investors regarding asset and portfolio management and securities buy-sell decisions. For the financial sector, significant effort and budget is needed to deal with security breaches and security maintenance. AI is now lifting the heavy burden and detects the try to the breach in advance very efficiently and in a very cost-effective manner.

Lastly, important stuff that comes under the umbrella of the financial sector is trading. Trading involves a lot of factors for example technical factors and fundamental factors. For a human being, it is tough to analyze such a high amount of data very efficiently but for an AI model, it is a matter of minutes if not seconds. Thus, AI has proven highly effective for AI-based trading. For example, AI and NLP leverage technologies can gauge market sentiments using the information prevailing in the market and can suggest buying and selling very easily which is beyond human capacity.

AI and Health care sector

AI is getting increasingly useful for almost every single field and similarly health eco-system. The potential benefits of AI in the medical industry include helping peoples not to visit doctors or visit less often. The important bit to achieve this goal is to continuously monitor the health of the people and send that data for analysis back to the data center, which is done by the Internet of medical things (IoMT). For example, young people extensively use healthcare-related applications which help them in controlling a lot of factors related to their health. A lot of data means the ability for doctors to understand the health pattern and need for care in a better manner. So it is more based on counseling, guidance, a deep understanding of health conditions, and information about how someone feels.

AI has been long been used in cancer diagnostics with help of scanned image segmentation and neural network. According to the Americal cancer society, AI is helping a great deal in the detection of early-stage cancer and heart diseases with help of imaging and scanning occurred during medical sessions. Adaptation and use of AI triggered a lot of open source research and industrial investments. That essentially creates a plausible environment to collect a lot of medical and health-related big datasets. These datasets are very crucial for further development and predictive medical analysis. So overall it has transformed the decision-making process in the medical and healthcare industry. This data is helpful in making the overall decision about which lifestyle and habits lead to deterioration of health and which economic and environmental factors are important for well being of the people and better health. Due to medical staff shortfall across the world research and practitioner has long been trying to employ robots in the medical field. It ranges from simple machinal robot arms in the manufacturing process up to highly intelligent robots equipped with intelligence and decision making and able to assist in surgical procedures.

For example, IBM Watson is helping the health-related firm to implement a lot of its cognitive science and neuroscience-related technology to strengthen their diagnostic ability, which gives them not only access to the database of the customers but also a lot of data related to similar diseases across the world including the prescription, reports medical histories, treatment procedures and a lot more. So the diagnostic system based on a lot of the data can give a prediction of diagnosis in a very accurate way, which is a great help in the medical industry. In a similar fashion, Google Deep Mind is trying to connect academia, research, practitioners, and patients to solve real-world health problems. The overall focus of Google Deep Mind health is to combine ML, and AI to build a powerful yet filed specific algorithm that works on a similar principle as the human brain does.

Overall lot of research and experiments are taking place in the AI direction right now. Reaching a tested drug from research preclinical trial to general mass takes several years. With the help of AI and health care applications advancement, we are hoping that this time span could be shortened.

NLP, Machine translations and text generation

Before the advancement of natural language processing, statistical models have been used for text classifications, sentiment analysis, summarization, dialogue systems, and similarly for machine translations. Over the passage of time, much has changed in terms of machine translation. The quality of machine translation is magical with help of recurrent neural networks and then subsequently encoder and decoder-based ANNs architecture. In this section, we will shortly introduce machine translation (MT), brief history of MT, ongoing current development in the field, and how overall AI is helping the area of MT as well.

Due to the internet and technology, the world has become a global village, and business and multinational enterprises need to expand their business into different regions, thus they need to offer services locally and need a lot of grip on the local language. Moreover, many students, researchers, NGOs, and other institutions encounter cross-cultural endeavors. The need for translation is deep-rooted and multifaceted. So let's look into the problem of translation at hand.

Translation problem is normally more complex than it seems. For instance, languages are normally not straightforward and there are many stories and metaphorical contexts attached to the words, so the translation of the literal meaning leads to a completely different translation of the second language. So, keeping this challenge in mind, the main task of MT is to carry contextual information within the text. These ambiguities come from syntactic properties, the meaning of the word, the relation of the word, and a lot of background context. Most syntactical problem comes from idiomatic phrases and morphological orientation, whereas semantic problems in the translation are related to the pronominal structure of the sentence. Like what pronoun is referring to. Additional gender of nouns is differently structured in different languages. "I saw the video and it is awesome". Now if we translate this sentence into German it may change its gender and thus need different properties of translation. Sometimes it is very important to know the word knowledge to build the problem causal or temporal relationship between the two clauses. For human beings, this difference is easy to identify but for natural language processing and computers, it is much harder than a human being.

In the early days of machine translation, researchers developed models that took into account syntax, semantics, and language-dependent features. Today, neural machine translation has taken this approach a step further by analyzing human-generated sentences for content and meaning. The first version of machine translation was proposed by Vauquois (1986). The idea was to first start with lexical transfer and gradually step-wise syntactical and semantic transfer of the translation from one language to another language pair. That was a kind of pyramid approach to transfer the context with the translation. The most important step in the development of the NLP was to build large-scale dictionaries, rules, and regulations. In this regard, there are mainly two types of corpus monolingual and parallel. So monolingual corpus helps in understanding the relationship of phrases and many other features of written language and parallel corpus is a bitext collection of text for two language pairs. Before we go into what tools and knowledge base are available in this field, let's take a brief look at the history and then the evolution of NLP.

Both neural networks and machine translation have a long history, with several peaks and dips over the passage of time. Each period is characterized by breakthroughs that ultimately brought spring to the winter time of Machine translation research.

As statistical machine learning has been replaced by neural machine translation in the last two decades. So, many new young researchers in deep learning came up with lots of innovations and path-breaking new models that have crafted MT to a different level.

First, a fully functional translation system was developed in 1976

by Montreal University. It was used to translate the weather forecasts. But non-commercially UA Air Force was using a Russian-English language translation system and the same system was, later on, brought to Europe, and new languages Paris have been developed namely, logos and METAL in 1980. Based on a similar principle DKFI has developed Verbmobill for different language pairs. Later on, after the widespread use of desktop computers, translation was commercialized on a large scale and people, different search engines and software industries start offering translation services to home users.
During the 1980s and 1990s, researchers in artificial intelligence and computational linguistics focused on developing systems that use interlingua to represent meaning independent of a specific language. They developed reversible grammars that could be used for analysis and generation. The success of CATALYST made it a target of considerable interest to commercial translation companies, and it was eventually bought by Systran [1]. The software was used by the US Army and many large companies. Based on Systran, Korean, Japanese, and Chinese versions were developed later on. Although the system was quite successful in translating technical documents, it had many limitations. It failed to cope with idiomatic expressions and it was unable to translate text that was not in the same domain as the training data. The development of statistical machine translation systems in the late 1990s led to the end-to-end development of CATALYST.

Since language translation is characterized by too many features and formalism, it is hard to capture all directions to reach a high-quality translation. So, researchers thought to use existing language translations done by humans and statistical models to try to find the best match of the translation of the sentence within the database of translation, and modify it based on some rules. We can call this data-driven translation. In the late 1980 IBM initiate the idea of statistic machine translation, different from statistical translation, due to the success of the statistical models in the field of speech processing and recognition. This idea was a foundational block of creating solid mathematical bases for machine translation. In subsequent years many notable projects started for example Verbmobil project is notable. In 1998 baseline research of IMB was reimplemented to create tools for statistical machine transition later DARPA(Defense Advanced Research Projects Agency) has shown a lot of interest and also funded many projects that play a significant role in advancement in the field. i.e.BOLT TIDE. USA defense authorities have also shown a lot of interest in the field of automatic translation after the event of 9/11 2001, especially in the Arabic language. A lot of other factors have played an important role here. For example, a computer's processing power and capacity are also very important to analyze a lot of language data. Finally, advancement in deep learning, general artificial intelligence, machine learning models, computational efficiency as well as graphics processing unit gives a completely new transformation to the field of machine translation. Let's have a look at some of the important breakthroughs after the researchers get convinced that a new approach of the neural network could give extremely high-quality translation and cope with the limitations of statical machine translation. In the Late 90s, some researchers have already given clues to shift the machine transition toward recurrent-based neural networks for example see. Forcada and Ñeco (1997) [4]and Castaño et al. (1997 however the data size could be blamed for not having enough success and attention to these models. Hence the resource scarcity surpassed the computational complexity and it took a lot of time for the model to get practically implementable.

Over the passage of time, many projects went open source, and data is made publicly reusable. Later in the year, Schwenk (2007) [6] gave an idea of integration of neural networks and neural language which could be considered as a solid cornerstone for the MT, but this idea rolls out or is optimized very slowly too, mainly due to computational efficiency factor.

The neural turn of the statistical machine translation has started reshaping the formation of MT. the pioneering work done by Schwenk (2007) [6] was on of early construction of basic step stone for mathematical grounds for the neural language model which use continuous speech recognition techniques, that means the estimation of probabilities in continuous space. The same idea was slowly adopted later on not due to conceptual robustness but due to computational power deficiency.

The next important neural language model was proposed by Devlin et al. (2014) [3] with a neural network join model that was purely lexical and could be integrated into other decoders. Some other notable researches include the adaption of convolutional networks into neural language or text-related problems. Generally, Convolutional networks are originally targeted at image classification. They have given input sentence to the network in the shape of word embedding and then these embeddings are used by the kernel that produces a sequence of the sentence and repeats the step multiple times to reach the end of the layer, it reverses the process to decode the sentence back into second language pair at the end to get the translated sentence back. This model is not recurrent but works like sequence-to-sequence models.

Similarly (Cho et al., 2014; Sutskever et al.,2014 [2] [8] were able to create a sequence-to-sequence network that was very important for the long chain of words that carry context, which was also a very significant improvement in previous techniques. The idea was to map two pairs of language with word vectors using the LSTM model that keeps the states or memory of the previous word within the sentence and that plays role in giving weights to transfer learning to the next word or state. That was end to end learning example that produce a BLEU score of 34 units. So, after the initial neural language model concepts enrichment, slowly the research community shifted from statistical machine translation to neural machine translation. After 2016 the whole idea of machine translation completely shifted to neural machine translation.

After the idea of the sequence-to-sequence learning method, an attention mechanism was invented by the research communities that means how to give importance to a specific world in relation to other words. Currently, after the successful implementation of the seq-2seq neural method, the transformer architecture was the most powerful method proposed by the research community. With the success of such a method, NLP and machine translation are at a whole new level.

So, now machine translation not only solves a lot of business-related problems but also helps the government and other official authorities to cope with a lot of language translation-related problems.

Evolution of AI

There is no doubt that artificial intelligence (AI) is rapidly evolving and growing more sophisticated every day. With the rapid expansion of AI capabilities, inevitably, machines will eventually surpass human intelligence. While this may seem like a scary prospect at first, it is important to remember that AI is designed to augment human capabilities, not replace them. In the future, humans and machines will work together to create a smarter, more efficient world. So artificial intelligence evolved on the principle of how the human brain essentially works.

The human brain is the most complex organ in the body, and scientists are still working to understand all of its functions. However, they have identified some of the main areas that are responsible for cognition, or the ability to think, learn, and remember. These include the cerebral cortex, the hippocampus, and the amygdala. The cerebral cortex is the outermost layer of the brain and is responsible for higher-level functions such as planning, decision-making, and language. The hippocampus is located in the middle of the brain and is important for memory and learning. The amygdala is located in the front of the brain and is responsible for emotions and fear. Cognition is a complex process that is not fully understood, but scientists believe that it involves the interaction of these and other brain regions. For a long time, scientists have been trying to understand how part of the brain communicates information and how it passes signals from one part to another is an interesting mechanism for neurons, dendrites, and synapses. mathematicians and practitioners were inspired to form the same mechanism and translate it into a numerical and mathematical formulation that leads to the birth of machine learning and subsequently deep and artificial intelligence

Let us look into a major breakthrough with some pioneering researchers in the area of AI. before we go into the evolution of AI, it is important to mention briefly, the classification or types of AI. There are mainly three types of AI: artificial narrow intelligence, artificial general intelligence, and artificial super intelligence [7]. Let us briefly describe these three types. Artificial narrow intelligence is the kind of AI that we see nowadays around us. The reasoning, problem-solving ability, and decision-making of AI systems are heavily dependent on data or environmental data the system sees and try to predict the next step. Providing some deterministic characteristics of the environment or set of problems narrows the level of artificial intelligence. i.e., SIRI and Alexa assistance from Apple Inc. Machine translation, facial recognition, and basic machine perceptions all fall under the category of narrow artificial intelligence. Then came Artificial General Intelligence, a general intelligence that is equivalent to human-level intelligence, and learned representation that could be applied across the platform and are capable of generalizability. So practically speaking there is no general AI that exists yet, it's just science and fiction, but mankind is moving toward that. The third type of artificial intelligence is "Artificial superintelligence" [7] super intelligence can be defined as an intelligence level that is beyond the human level or which humans cannot achieve. This is a level that surpassed humans in reason, wisdom, problem-solving ability, decision making, and efficiency or seeped to perform certain tasks. Having AI categorized, let's talk about a short glimpse into the recent past of AI and how AI evolved to narrow artificial intelligence. After vetting a lot of literature and surveys regarding AI, we can segregate AI development into three main phases, namely, the Alan Turing era 1950 -1970, followed by the knowledge-based and expert system era from 1970 - 1990, and the modern AI era that is the 1990s to present. Let's briefly look into the main issue milestone achieved in each era one by one.

The Foundation of the AI 1950 - 1970

A famous paper by Alan Turing "Computing Machinery and Intelligence (Turing 1950)" , gave the concept that if the machine can transmit a message that is not different from a human message then we can call it an intelligent machine. following are some of the breakthroughs

In 1950, Alan Turing wrote a paper called "Computing Machinery and Intelligence", in which he proposed a test that would come to be known as the Turing test. The test is simple. A human judge engages in a natural language conversation with one human and one machine, each of which tries to appear human. All participants are placed in isolated locations. If the judge cannot reliably tell the machine from the human, the machine is said to have passed the test.

Alan Turing starts to think about the philosophical phenomenon that there is no reasonable argument exit that machines cannot think intelligently making this prospect clear that machines can think and thus can lean and improve.

Turing has foreseen that more than 70 percent time human interrogators will not be able to distinguish the machine take decision from the human taken decision. That is amazingly true today. Alan further wrote that if a machine can think that is based on conditional reflex and programming instruction machine must be able to learn and improve the reason and decision-making process. That was in fact ground-breaking though and a vision by Alan Turing that gave birth to artificial and machine intelligence. By the time Alan has published his research the journey of AI has started and we can classify AI in different eras during which AI has made exponential growth and development. Let us briefly explain some significant events that took palace in different AI development phases.

In 1956, John McCarthy held a conference at Dartmouth College about artificial intelligence. The conference was attended by several prominent researchers in the field, including Allen Newell and Herbert Simon. In the conference, John McCarthy stated that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." This statement was the beginning of the field of artificial intelligence. The first artificial intelligence program was written in the year 1957. It was called the "Short code". Short code was designed to simulate mathematical reasoning. It used the logic of predicate calculus. This program was created by John McCarthy and Marvin Minsky. In 1959, John McCarthy started the Stanford Artificial Intelligence Laboratory. The laboratory was started with a grant from the National Science Foundation. John McCarthy was the director of the laboratory. The laboratory was one of the first to attract top computer scientists and, in the years that followed, produced some of the most important and influential research in the field.

Computer program learned to play checks got media attention to

Discovery of the perceptron birth of a neural network: paper published by titled "the embryo of an electronic computer that will be able to walk, talk, see, write, reproduce itself and be conscious of its existence." In 1957, psychologist Frank Rosenblatt developed a neural network that attempted to mimic the brain's neurons. He called his first network a perceptron, which is essentially a digital representation of a few brain neurons. The perceptron was designed to classify images into two categories: male or female. It learned these categories over time by scanning in images of men and women as data and hypothesizing that over time the network would learn how to distinguish between the two (or at least see patterns that made men look like men and women like women). This is essentially how we learn new things over time.

In 1958, the New York Times published an article describing Rosenblatt's perceptron as "the embryo of an electronic computer that will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

Although neural networks were first proposed in the 1950s and implemented in the late 1960s, their effectiveness was limited due to their small size. In 1969, Marvin Minsky, who had worked with Rosenblatt as a student, published a book that assessed the work done on neural nets. He determined that these networks were "generally disappointing" and "hard to justify." The book was said to have so impacted the computer science community that Artificial Neural Networks were abandoned for over a decade.

Book published by "Perceptrons" introduced feed-forward layers and two layers structure of perceptrons. The book entails knowledge and literature of almost a decade and lays the basic foundation for neural networks.

Expert systems era 1970 - 1990

Nothing much happened in this era because of financial recessions, less attention, and investment in technologies due to financial constraints. Most of the focus has been given to expert systems and knowledge base systems. The following are the salient achievements of the 3 decades.

A computer program called of work form user and engines in conversation

Hopfiled net first form of backpropagation

ID 3 algo that creates decision trees from giving datasets to classify the set of problem

The modern era of AI from 1990 - Present

We can call this the era of machine learning, deep learning, and reinforcement learning. much happened in that time, both at the invention level, industry level, and academic level. Here are the milestones we have achieved in this era.

Reading hand digits with the help of a convolution neural network

introduces the concept of Q learning with help of the Markov chain decision process that was started.

German science Schmidhuber solves the task with deep learning with almost 1000 layers of network in 1993 at the start of a recurrent neural network. Again, German scientist Scmidhauber and his student came up with the idea of long short-term memory, a breakthrough in the field of natural language.

A robot challenge DARPA Grand Challenge, autonomous diving by the robot

3.2 million label images called ImageNet dataset completely open-source dataset. set a level field for big experience in the future.

The ImageNet competition was won by AlexNet main triggering point for deep learning research [5]

Generative adversarial network that generates the data like image

Alpha Go defeated the world number champion in the game. Alpha Zero mastered chess in less than 4 hours and defeated the best players without losing a single game.

Rransform-based encoder and decoder-based neural network architecture with a special multi-head attention-based mechanism, that we have used in one of our research papers.

Many advances in computer vision. i.e Yolo, edition of yolov5, autonomous card drive with RL and image segmentation.

Open AI advancement in the field of natural language generation, with the discovery of GPT3 is now GPT2 that can write grammatically correct paragraphs and blogs outline with extra ordering context aligned, text generation.

After shedding some light on the history of AI we must be careful about the ethics of AI, this is the area that needs to be researched because there is a lack of legal paradigm that encompasses that ethical issue that might arise due to the advancement of AI, including in near the something if AI reaches the superintelligence level mankind should be prepared for that .

Technology and Business:

In the mid of the 20th century, the age of information started. The technological focus has shifted from mechanical and analog technologies to electronic technologies. Record keeping and information and record access were digitized first, which has been used still in the same way. In this era, transistors, integrated circuits, chips, and the logic gate have been invented and further developed, which give birth to computers, cellular, internet in information technologies. With the commonality of the internet, people started having personal computers which created an interest in the usage of video games, networking, and electronic music. After 1980 computers made their way to businesses like automatic teller machines, industrial robots, and educational institutes. From 1989 to 2005 world web and streaming of the internet, so-called web1.0 got popular.

During the same era industries, a business starts leveraging their operations and procedures with the help of the computer.

Over these years, every sector of business compelled to adopt the digital revolution. The businesses that keep aligning with the momentum of technology survived and those who were unable to upgrade vanished from the map of the business world.

Let us take a short look at how different sectors of business have changed since the happening of the digital revolution at the end of the 20th and the start of the 21st century.

For example, health care is a very important business section for any country. As it involves, the health of people and general policy-making and money circulation for this sector is particularly important. Just for sake of comparison per capita spending of the US is, 7910 $ in 2008, which is, in comparison, almost two times of OECD countries. For this alarming figure, lots of inefficient procedures, unnecessary extra equipment, and malpractices are responsible. Using innovative technologies in the medical industry, for example, the Internet of things, where one device can communicate data to another remote device in real-time this cost could be decreased significantly. The rate of change per capita in health care spending for the period of 1980 to 2008 has dropped significantly as compared to the rate of change for the period of 2008 to 2020. Apart from other things, credit for this decline somehow goes to technological advancements because industrial practitioners can better insight, predict and manage.

Similarly, Education and learning took a giant leap. Earlier, access to information regarding any subject or technology was dependent on access to the top institutions, and libraries. Access to such resources was not available mostly to the inhabitants of the Third World countries. But with help of internet and communication technology, cloud computing, and an indirect revenue-based marketing model, now high-quality courses and knowledge can be learned and accessed anywhere sitting from home.

Soft-computing and modern technologies played a vital role in all walks of life and completely transformed the fundamental structure of the business, but due to brevity, we do not want to go into greater details. Our focal point here is how technology in general and especially Artificial intelligence helped business and corporate finance to grow and transform.

Let us briefly look into how technology, in general, is involved in reshaping the future of financial services.

Over the last few decades technologies have changed the game of doing business, like, as communication, managing customer services, chatbots, data sharing, data storing and security, B2B third-party services integrations and financial prediction has completely reshaped. For example, people can manage their financial assets including cryptocurrency wallets on their mobile phones, even while traveling. Furthermore, banking institutions have shifted their data layers to clouds and hybrid cloud infrastructure that gives more speed, less latent accessibility, and enhanced security. Intelligent and real-time data sharing gives more personalization and insight into spending and data consumption.

With the introduction of REST API, banking analytics and financial data can be consumed by any third-party budgeting and expenditure asset management applications on mobile or web Apps. Data sharing across the platform could be possible due to cloud infrastructures and API protocols.

Blockchain technology has changed the transaction verification and security process. After hype of cryptocurrencies and a lot of investments in this type of currency, many practitioners, researchers, and investors get attracted. Many big names like Jp Morgen and other renowned financial institutions finding ways to implement blockchain technology in the banking sector for payment processing, loan processing, and fraud detection mechanism.

More specifically, currently, Artificial intelligence completely reshaping the financial sector. For example, Asset price predictions with AI Algorithms, market returns chart pattern recognition, and financial advice with artificially intelligent chatbots, that are trained using AI and NLP methods, on tens of thousands of real in-person pieces of advice, events/discussions in the past.

For example, considering other important traditional indicators of financial markets, neural networks are proven to be amazingly effective and more accurate stock market returns forecasting methods as compared to traditional time series-based models. Long-short term Memory so-called LSTM keeps the very long-term states of training weights and thus a very efficient way of predicting stock market returns.

The financial performance of the firm is a key factor for the investors and people who lend money to the business. Traditionally, there are many fundamental indicators i.e., financial ratios that are considered performance indicators. But for example, due to natural Laurange processing, now it is possible to parse the financial reports and SEC fillings/ from 10K FORM (document gives discussion regarding the financial position of the company and estimate the positive or negative sentiments regarding the financial state of the company. A combination of text analysis, financial indicators, and non-linearity-based estimation methods I.e., Natural networks, is able to solve the prediction problem with way more efficiency and accuracy.

When it comes to business there are many different sectors where we can leverage existing business processes and operations with the help of artificial intelligence. Following the same footprint, the thesis stresses the application of Artificial intelligence in business and corporate finance.

So, for this purpose, we have chosen three different business areas as case studies to provide scientific evidence that how artificial intelligence is helping businesses and corporations? These three areas are namely; financial forecasting and new analytics process, financial distress prediction, and business report writing or artificially assisted customers' support area.

To support our thesis, we have published two research papers and one study project that will appear as it is in the incoming sections. The following are the titles of the research publications and a study project.

Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytics Process

Transforming the Generative Pretrained Transformer into Augmented Business Text Writer.

Artificial Neural Networks for Bankruptcy Prediction.

Incoming section, we will place the above three scientific studies as it is. And then we will discuss the overall results, limitations, and some future suggestions at the end.

Chapter 2

Research Papers

2.1 Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytic Process

Author: Faisal Khalil:
co-author/ supervisor: Prof. Dr. rer. Gordon Pipa:
Authors Contribution:
Faisal Khalil:
The research paper is written by Faisal Kahlil during Ph.D. at the department of cognitive science University Osnabrueck.
Prof Gordon Pipa
Prof. Pipa supervised the overall research concept and has given a lot of valuable suggestions, directions, and continuous thought to the research paradigm, that reshape the research paper.
Peer Review:
Published: Computational Economics 60 (1), 147-171
https://link.springer.com/article/10.1007/s10614-021-10145-2

Computational Economics
https://doi.org/10.1007/s10614-021-10145-2

Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytic Process

Faisal Khalil[1] · **Gordon Pipa**[1]

Accepted: 22 June 2021

Abstract

This study tries to unravel the stock market prediction puzzle using the textual analytic with the help of natural language processing (NLP) techniques and Deep-learning recurrent model called long short term memory (LSTM). Instead of using count-based traditional sentiment index methods, the study uses its own sum and relevance based sentiment index mechanism. Hourly price data has been used in this research as daily data is too late and minutes data is too early for getting the exclusive effect of sentiments. Normally, hourly data is extremely costly and difficult to manage and analyze. Hourly data has been rarely used in similar kinds of researches. To built sentiment index, text analytic information has been parsed and analyzed, textual information that is relevant to selected stocks has been collected, aggregated, categorized, and refined with NLP and eventually converted scientifically into hourly sentiment index. News analytic sources include mainstream media, print media, social media, news feeds, blogs, investors' advisory portals, experts' opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. The results of the study indicate that sentiments significantly influence the direction of stocks, on average after 3–4 h. Top ten companies from High-tech, financial, medical, automobile sectors are selected, and six LSTM models, three for using text-analytic and other without analytic are used. Every model includes 1, 3, and 6 h steps back. For all sectors, a 6-hour steps based model

✉ Faisal Khalil
fkhalil@uni-osnabrueck.de

Extended author information available on the last page of the article

outperforms the other models due to LSTM specialty of keeping long term memory. Collective accuracy of textual analytic models is way higher relative to non-textual analytic models.

Keywords LSTM · Natural language processing · News analytic · Sentiment analysis · Stock prediction

1 Introduction

Accurate forecasting of returns is crucial for individual investors, investment banks and corporate investment managers. It is also equally important for investors to foresee the returns accurately and design the investment or trading strategies keeping in view all relevant aspects of forecasting. For many years stock market forecasting studies have been emphasizing the volatility models. Few studies inculcate the role of technological forecasting i.e. artificial intelligence. The efficient market hypothesis (EMH) is proposed by Fama (1998) is under criticism, because the proposed model is in contrast to the behavioral finance concept (Kahneman & Tversky, 1979; Kahneman, 2003; Shefrin, 2008). It has been much debated and considered as a limitation of EMH that this model is not considering the role of investor's sentiments and their behavioral aspect. Technological advancements and inventions of the new artificial intelligence-based model are reshaping the method of forecasting (Wang et al., 2018; Kuo & Huang, 2018; Makridakis et al., 2018). Normally, the artificial intelligence-based model takes previous stock prices and other variables into account but news analytic consideration is less researched. In this specific context news analytic is in the early stages and needs advancements for better forecasting and efficiency of intelligent trading systems. In the last decade, soft computing methods and techniques have grown rapidly that entice researchers to explore more sophisticated techniques for the stock market and time-series predictions. Time series financial modeling has a long history and time-series data is characterized by hidden relationships, high uncertainty, and unstructured in nature. To estimate the behavior of financial time series there are two types of models are available; linear model and non-linear models. Whereas, linear models are affected by techniques like Box Jenkins and Kalman filters, piece-wise regression, and Brown's exponential smoothing. All these theories are turning data into the linear functions. However, recent evidence shows that financial markets behave in a non-linear fashion. In addition to these problems, there are other factors that intact with financial markets like, general economic conditions, political events, news, and investor's psychology that makes a stock market prediction so difficult (Cheng & Chan, 2018; Huang et al., 2007). To address these issues, artificial intelligence has been evolved as a very good technique due to its learning, generalization, and non-linear behavior to overcome these problems and to give better forecasting (Makridakis et al., 2018; Li & Ma, 2010). In this connection, the most relevant techniques are; Recurrent Neural networks, Neural

Networks, fuzzy logic, and genetic algorithm (Hiransha et al., 2018; Ergen et al., 2017; Nelson et al., 2017; AlFalahi et al., 2014). Artificial Neural Networks model are pretty good with flexibility and adaptability to learn from changes and previous trend in a given set of input and predicts the trends based on network training. There is a fair deal of evidence that exists in the literature that models that based on artificial neural networks outperform the traditional time series model, for example, see Adebiyi (2012), AlFalahi et al. (2014), Trippi and DeSieno (1992), Correa et al. (2009) and Hansson (2017). There are many soft-computing techniques available under the umbrella of artificial intelligence but finding appropriate techniques is very important to get accurate forecasting results. Study of Li et al. (2018) and Atsalakis and Valavanis (2009) can be referred here each has surveyed more than 100 articles by researchers who have used fuzzy logic, genetic algorithms, and neural networks and recurrent neural network as modeling techniques in their studies. It is evident from these articles that mostly researcher have used feed-forward neural networks (FFNN), currently, some studies use Recurrent Neural Networks(RNN) multilayer perceptron (MLP) to forecast the stock markets (Arora et al., 2019; Pawar et al., 2019). This survey study also testifies the magnitude of the importance of non-conventional tools for stock market prediction. For the stock market prediction process we cannot rely upon past stock prices and some other variable but we need to embed the impact of market news to achieve maximum accuracy. In the prediction process, it can be very tedious for managers to focus on every news that just pops up and align their investment strategies. A human being can miss much information and even information can be out of his reach as well. So, here natural language processing (NLP) techniques come into play. So, there is an urgent need to automate the news analysis process based on NLP technique so that the investment manager and the corporations can be benefited as well as AI-based predictive models can be supplied with more relevant information instead of just past prices. Natural language processing is a subfield of AI where Algos and deep learning model tries to make computers understand language intuitively near to the human level (Nadkarni et al., 2011). A human being has evolved from thousands of year training to understanding the emotion and feeling of language elicits but computers are struggling with the help of deep learning and AI-based models. In this study, we have used the NLP model (see Fig. 1) with the help naive Bayes classifier to process the raw information that is parsed out of many sources. These sources include mainstream media, print media, social media news feeds, blogs, investors' advisory portals, expert's opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. Detail of the news analytic and sentiment analysis can be seen in Sect. 3.1.2. Many studies propose soft computing techniques for better and most of the researches have focused on the comparison of traditional time series stock prediction models and artificial neural embedded network models. This study contributes to the existing body of knowledge in the following ways: Normally, studies use news information and stock price data for indices. Apart from other motivations to choose indices for the prediction process, one benefit is that data collection and aggregation is relatively easier

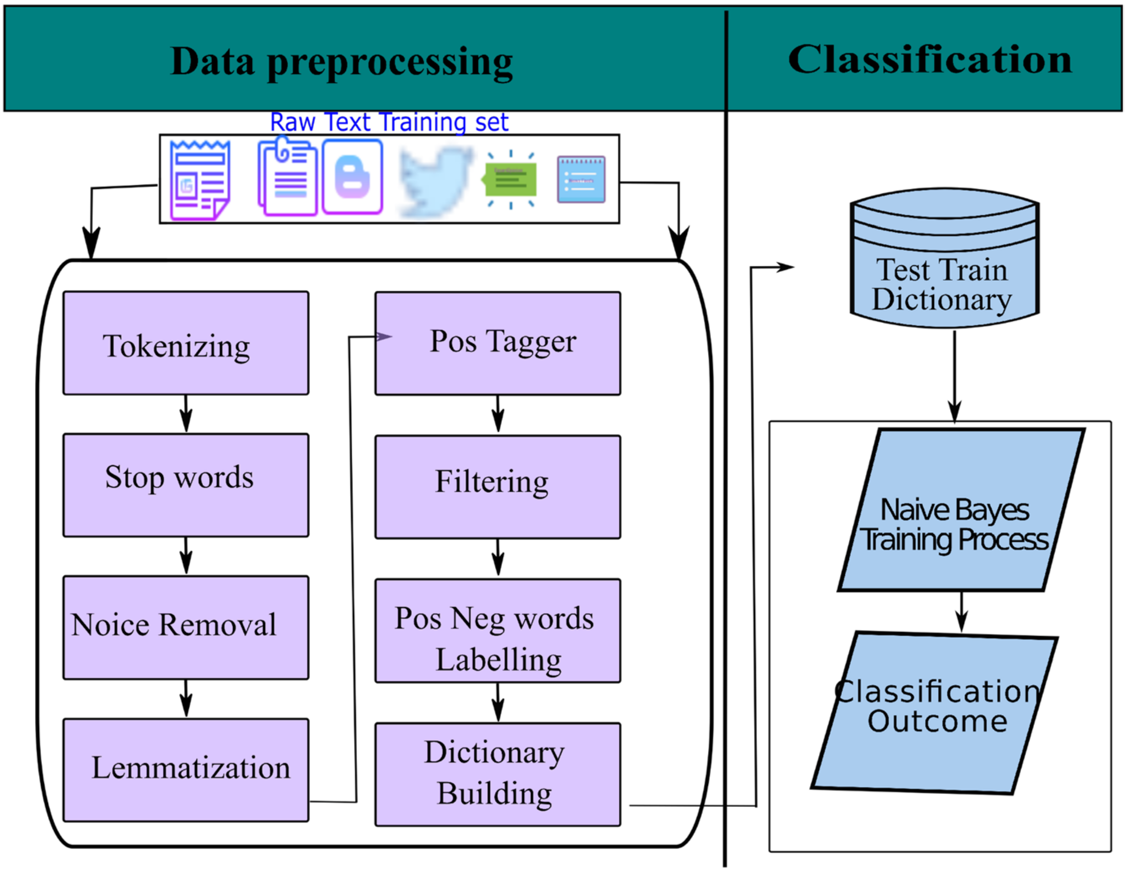

Fig. 1 Natural language processing model

because of its ready availability. However, collecting news information for each company individually and make meaningful sentiments for that stock is challenging. However, this study focuses on individual-level stock and news information that makes this study bit challenging because not only news from all possible sources need to accumulate but also company internal news is also taken care. For example, the company changes the top echelon due to any reason or decides to change the level of dividends, any commentary on 'hashtags' is not covered by prominent media sources but still, they impact upon the prediction. Secondly, this study is emphasizing NLP techniques and the way how to raw news text can be used for sentiments building processes. So, NLP based models are simply efficient in extracting emotion, feelings, and sentiments out of a raw text. Thirdly, this study not using simple neural networks for predictions process but Long Short Term Memory (LSTM) model based upon the newly developed and highly proven performance in different fields. LSTM models are specifically designed to remember the long-term dependencies. A point that makes it different is mostly, LSTM model is supplied with past stock prices as an input to predict the future price of the stock, however, this study has used sentiments, extracted with help of NLP techniques, to predict the stock price and it is evident from results that model with sentiments has significantly increased the accuracy of the model. This study will be generically beneficial to all institutional and individual investors, all kinds of traders, portfolio managers, and specifically for short-term and long-term investors who invest in the equity market, future marks, derivative and foreign exchange market.

The rest of the paper is divided into the following sections; Literature review, Methodology section that discusses data collection processes, sentiment index development process, NLP techniques, and implementation of the study model. Then comes results and their interpretations and finally the conclusion of the study.

2 Literature Review

By exploring existing literature on the application of neural networks and machine learning in the area of business and finance we found that neural network literature is rooted back to 1988. Research articles have started publishing in 1990 exclusively in the area of finance (Zhang et al., 1998). The neural network are gaining popularity in organizations that are investing money in neural network and data mining solutions for the resolution of their problems (Smith & Gupta, 2000). In the year 1988 researchers were more focused on the application of artificial intelligence in the area of production and operational management. In the area of finance more articles are published in the field of financial firm's bankruptcy prediction. Whereas the focus of stock market prediction was restrained to comparison of traditional time series models with ANNS. Despite the fact that a substantial effort has been made for time series prediction via kernel methods (Chang & Liu, 2008), ensemble methods (Qin et al., 2017), and Gaussian processes (Frigola & Rasmussen, 2013), the drawback is that most of these approaches employ a predefined linear form and may not be able to capture the true underlying non-linear relationship appropriately. Recurrent neural networks (RNNs) (Rumelhart et al., 1986; Werbos, 1990; Elman, 1991), a type of deep neural network specially designed for sequence modeling, have received a great amount of attention due to their flexibility in capturing non-linear relationships. In particular, RNNs have shown their success in NARX time series forecasting in recent years (Diaconescu, 2008; Gao & Er, 2005). Traditional RNNs, however, suffer from the problem of vanishing gradients (Bahdanau et al., 2017) and thus have difficulty capturing long-term dependencies. Recently, long short term memory units (LSTM) (Hochreiter & Schmidhuber, 1997) and the gated recurrent unit (GRU) (Cho et al., 2014) have overcome this limitation and achieved great success in various applications, e.g., The reason suggested by researchers is that the neural network has the capability to outperform the time series models because these models can efficiently predict without the requirement of data being following any distribution and linearity. In addition to the comparison of ANNs and traditional model, in literature evidence exists where models based upon artificial intelligence are compared with each other e.g. Tan et al. (2011) have compared three models; ANN, decision tree and hybrid model with the conclusion that ANN has the highest accuracy in stock price prediction.

In early stages of development for financial forecasting using ANNS, researcher and professionals emphasizes on a comparison of traditional time series models and ANNs to measure the better accuracy in forecasting process for an instant see

Swanson and White (1997), Yoon et al. (1993), Kaastra and Boyd (1996), Lawrence (1997) and Kryzanowski et al. (1993). ANNs gives 72 percent accuracy of predicting stock market returns and also able to accurately predict the positive and negative returns by training and validating the neural networks (Kryzanowski et al., 1993).

As for as methods of artificial neural networks are concerned researches have used different ways to mimic the neural networks of the human brain.

Many efforts have been made to solve the issue of linearity, for example, Kernam method has been used by Chang and Liu (2008), Bouchachia and Bouchachia (2008) and Frigola and Rasmussen (2013), with the help of traditional non-machine learning-based model that are unable to capture underlying non-linear relationships. Stock market data is always stochastic and noisy in nature so, LSTM is more suitable. Normally, statistical and metamathematical models are used for financial prediction and these model are handcrafted and aligned with respect to observation and thus compromise accuracy (Tsantekidis et al., 2017). Fischer and Krauss (2018) suggested that LSTM performs well as compared to Random forecast, Deep neural network, and logistic classifier. Recurrent Neural Network(RNN) model gained popularity due to the flexibility of use and coping up the problem of linearity in time series (Rumelhart et al., 1986; Werbos, 1990; Elman, 1991).

Artificial intelligence based expert system is also catering to the needs of auditing, banking sector, credit risk management but along with it matchless benefits there is the dark side of these expert systems of being costly. Omoteso (2012) have studied the cost and benefit analysis of an intelligent system that can predict the future direction and softwares development in this area. It is concluded that in small and medium organization it may be not suitable to apply such system to achieve the marginal benefit by incurring heavy cost. Oreski et al. (2012) apply neural networks to reduce the data dimensionality by coping redundant data and removing irrelevant factors to enhance the predictive ability of genetic algorithm. Similarly, López Iturriaga and Sanz (2015) designed the artificial neural network-based model that have predicted the financial distress of US bank 3 years before the bankruptcy occurs.

2.1 NLP

Recently, natural language processing (NLP) has grown up as powerful techniques for many fields due to its capability to capture sentiments and feeling into the text in more nuanced way. Many applications have started adopting the NLP techniques to give their users better experience (Xing et al., 2018). Though it relatively easy to get the external news with help of many sources but it difficult to access and parse the data through financial statement of company. So, developing information content from companies financial statements is tedious and difficult. Here information means voluntary information disclosed by firm that is not obligatory by law to disclose to stakeholders (Xing et al., 2018). With help of databases this paper includes all sort of internal information whether it reaches to external media or not as well as external news and information.

Textual information extraction and news articles processing rooted back to 1934 (Bühler, 1934; Chomsky, 1956). Previous two decade people have been giving much focus upon bag of word approach to seek sentiments of text with help of stop words and frequencies. Serous drawback of these model is that they are unable to capture the context of sentence. For example, company A is gaining advantage over company B or company B is gaining advantage over A are two completed opposite sentiments but belongs to same bag- of-word. Recent advances like, word to vector representation , word embedding and LSTM have addressed these problems very well. Sentiment analysis is very important phenomena for stock market and financial forecasting (Poria et al., 2016). With increasing use of web.2.0 Standards (Cooke & Buckley, 2008) users have easy access and ways to sharing the information across platform like Facebook, twitter, etc thus market sentiments become importance for financial market. Businesses dealing in financial products and services reshaping their approach to make their application more informed and sophisticated to gain competitive edge over rivals. New NLP techniques are promising them for their required edge. Existing sentiment technique can be broadly categorized into three domains; namely, hybrid , knowledge-based and statistical approach (Poria et al., 2017). Knowledge-based sentiment analysis is based upon list of words and its frequencies- a relatively old approach that categorizes text into different categories and then further compares the frequencies with the lexicon. Second is statistical method, this approach is not only focusing on list of word but also use statistical model to classify the text with help of probabilities. Third category is mixture of these two Lenat (Lenat et al., 1990; Liu & Singh, 2004; Fellbaum, 1998). This study has used hybrid approach with help of modern available NLP techniques that supports programming languages environments as well.

3 Methodology

3.1 Data Pre-processing

This section describes a summary of approaches and methods that have been used to process the data from raw text to machine-readable data. Data preprocessing has been divided into four major sections, namely; hourly stock returns, News Analytics preprocessing, Naive Bayes Classifier and sentiment index development. All three sections give the snapshot of preprocessing of data. Let's briefly describe one by one.

3.1.1 Hourly Stock Data

Hourly stock returns are calculated with help of opening and closing price of all 10 companies. Hourly stock data is obtained from Thoumson Retures data portal. Simple formula for calculating the stock return is as follows:

$$R_{ij} = \frac{closing_{ij} - opening_{ij}}{openingij} \tag{1}$$

whereas R_{ij} is j_{th} stock at i_{th} hours, $closing_{ij}$ is closing price of j_{th} stock at i_{th} hours, *opening ij* is opening price of j_{th} stock at i_{th} hours.

3.1.2 News Analytic Processing

There are many sources through which information flows into the stock exchanges related to a specific stock. News and information sources that have been used in this paper are: mainstream media, print media, social media news feeds, blogs, investors' advisory portals, experts opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. We have collected the news stories from a very well known and reliable database, named; Thomson Returns. Using Thomson Reuters's API we were able to collect new stories if these stories would be related to any of ten stocks which, we have chosen for analysis. The reason for choosing individual stock instead of the stock exchange is; stock exchanges absorb and react to collective level information and thus, specifically event-level information is hard to be separated. Every news story has its timestamp according to GMT and precise at the millisecond level. The time frame for news collection is 10 years, so, collecting every news resulted to have very large text corpus. The timestamp for news is strictly matched with the stock exchange's opening and closing time. Although, we have thrown a lot of use full collected news information that lies outside of the stock exchange opening and closing time window. However, it was necessary to gauge the impact of news analytics on the stock price movement.

3.1.3 Naive Bayes Classifier

After the raw text regarding news is extracted from sources, the text is refined in the way that it can be used in the Navie Bayes Classification model. Originally text was in 'HTML' form with a lot of unnecessary information, but with help of parser and some lines of coding, 'HTML' based- text is refined and filtered into 'lxml' form. 'XML' form of text is accurately and quickly readable by machines. Naive Bayes Classification model has been used to calculate the sentiments out of news text. The Fig. 1 shows how information filters though raw sources to sentiment score. The left column of the diagram shows that Raw text, which includes 'HTML' meta-information in it. The first step is to split the complete sentences into a list of unique words, the process is called tokenizing. Next comes, creating a filter of stop words, these stop words are mostly related to pronouns. At the next stage, the text is filtered from hyperlinks and unnecessary information. In the next step, lemmatization is applied to address spelling mistakes. The list of all words is labeled with a part of speech. Then, data is a little bit more refined to see any redundancies. As a next step, with the help of the already available NLTK database, each word has been assigned with negative or positive labels. In the next two steps data is prepared for test and train dataset - ready to feed to the 'Naive Bayes' Model for training. After the training process is completed each sentence is tested to get sentiments scores out of it. The outcome of the NLP model is utilized in building the sentiment index and LSTM data at the later stages.

3.1.4 Sentiment Index

Following variables are taken into account while building the sentiments index: 'sentiment time window', 'score value', 'class of sentiment score ', 'relevance' of score towards the underpinning stock; time window means how many times news/information, related to the selected stock appeared during 1 h time period. The logic behind keeping the time window to 1 h is that stock exchanges need a bit of time to absorb the information related to an individual stock. secondly, Minute level analysis is too early and day level analysis is too late. Next factor is 'score value'. Sore value is the outcome of a trained NLP model, the process is given in the Fig. 1. Sentiment score values are classified into three categories based on their scores; positive, negative, and neutral. All the negative score are carrying the negative signs and neural sentiment are equal to zero. The scores for all three type of classes are ranges from 0 to 1. Thus 'score value' are summed up during the 1 h time window, if the sum of the score is negative and greater than − 0.10, it is labeled as negative score, if the sum is between − 0.10 and 0.10 it is considered as neural score and, from 0.10 to 0.90, the 'score value' is positive. In the Next step, the sentiment score outcome is finally multiplied by variable 'relevance' to weight the sentiment with respect to its relevance score. 'Relevance score' is percentage number, calculated; the number of times news story mentioned the name of a stock divided by the total count of words in the news story. The mathematical expression of the sentiment index is as under:

$$S_{ij} = \sum_{i \in I} \left(e_i \times R_i \times C_i \right) \tag{2}$$

whereas I = time windows for everyi_{th} and j_{th} stock. $e = \max\left(pos_i, neg_i, neut_i\right) \exists$,

$$C_i = \begin{cases} +1 & argmax\left(pos_i, neg_i, neut_i\right) = 1 \\ 0 & argmax\left(pos_i, neg_i, neut_i\right) = 3.\ R_i = Relevance. \\ -1 & argmax\left(pos_i, neg_i, neut_i\right) = 2 \end{cases}$$

$$P(w_j|x_i) = \frac{P(x_i|w_j).P(w_j)}{P(x_i)} \tag{3}$$

whereas w_i is a particular class (e.g. Negative or positive) and x_i is an given features, $P(x_i|w_j)$ is called the posterior or in other word probability of feature x_i belongs to class w_j, $P(w_j)$ probability of class itself with respect to total sample also called the prior and finally, $P(x_i)$ is called the marginal probability or evidence. based upon above-stated Bayes theorem, conditional class probabilities of the equation can be calculated as follows:

$$P(\mathbf{x_i}|\omega_j) = P\left(x_1|\omega_j\right) \cdot P\left(x_2|\omega_j\right) \cdot \dots \cdot P\left(x_d|\omega_j\right)$$

Posterior probabilities can be calculated with following expression:

$$P(\mathbf{x_i}|\omega_j) = \prod_{k=1}^{d} P(\mathbf{x_k}|\omega_j)$$

So,probability of class can be calculated with this expression:

$$P(\omega_j) = \frac{N_{\omega_j}}{N_c}$$

3.2 Model Equation

Artificial intelligence-based models have proved their importance and efficiency in almost all spheres of life and the field of economics and finance can not be excluded. Our model can be used practically in a variety of ways. For example, online trading expert systems are forced to integrate advanced ways for the prediction process. The current model could be specifically very relevant for the trading system to reshape the prediction process and reduces the effort of organizing and search the relevant market info through millions of text records with either human-based effort or the traditional text filtering approaches. The model already uses sophisticated NLP techniques to include the sentimental-based market information into the model. For example, building the information-related index is very crucial. Keep this point in view we have built a customized sentiment index that collects the market information at one minute level and sums it up for a 1-h window. On one hand, it enables LSMT model to capture high-level precision and on other hand, its overcome the limitation to rely upon daily-based market information. There are many traditional models which try to achieve precise forecasting using economic data i.e. simple regression, Moving Averages, and autoregressive-based models (see.ARMA, ARIMA, ARCH GARCH), simple regression, and a bunch of other time series forecasting models. The universal problem for all these models is the limitation to handle the assumption of linear distribution, handling long past lags, and very strict criteria of data structure. These limitations come with a lot of compromises in terms of efficiency and accuracy. Artificial neural network-based model and most specifically, LSTM is very good at handling long-term dependencies i.e. you can keep tracing the past data without losing the information it carries. Moreover, With help of different activation functions and specific approaches model works flexibly without setting many assumptions Let's elaborate how this model works.

The current model is based upon original scientific publications made by Hochreiter and Schmidhuber (1997). The research is regarded highly by the research community because of its ability to work on long-term dependencies and the ability to remember important information in previous steps. The cases where the dependency of information does not matter much, simple neural network models work fine, but this is not an ideal situation in the practical business world. Stock market prediction, natural language processing, sentimental analysis, and language translation are the example where information of model is highly dependent and context is very important thus recurrent neural network model are good alternatives of simple neural networks. Here is a short description of how the model of this study is fitted.

Hidden state function can be written in the following way:

$$h_t = f(h_{t-n}, X_t) \tag{4}$$

So, the hidden state of LSTM model has been written with the help of the following equation.

$$h_t = tanh(W_h h_{t-n} + W_x X_t) \tag{5}$$

Weight matrix is first multiplied with current input.Previous time steps hidden states are one by one multiplied with weight matrix for hidden state. Finally, tanh has been applied on result after adding both, current input and previous time steps hidden states. Now output layer of LSTM model is as under:

$$O_t = \sigma(W_o.[h_{t-1}, X]_t + b_o) \tag{6}$$

whereas W is weight matrix for output layer and h_t we have calculated in Eq. 5.

Equations 5 and 6 simply shows how hidden and output layers of the LSTM model are formulated but this formulation is not much different from simple neural network models. The true secret of LSTM model lies in its unique way of developing cell and memory state with help of gating mechanism.

3.2.1 Signalling and Gates

Gates are basically fully connected feed-forward networks that receive information, applies functions, usually sigmoid activation functions, and do point-wise operations and then return outputs. Thus, we have applied here sigmoid activation function that spits outputs between the range of 0 and 1. So, all the outputs values closer to 0 are considered unimportant and cell deletes them, on the other hand, all information that is close to 1 is important for the prediction process and therefore updated in cell state. In this section, we will describe how signals and gates for LSTM work. Not all information in cell state is important to know for the prediction process and overflow of unnecessary information means disinformation. Primarily there are three gates of LSTM, namely: forget gate, input gates, and the output gate.

Forget Gates Forget gate receives information from current input and earlier hidden layer input, it applies the sigmoid function on this number and multiplies it with previous cell state. This decides that whether we want information in previous cell state with respect to new information and $t-1$ information in state C_{t-1} . The mathematical equation of forget gate is as under:

$$f_t = \sigma(W_f.[h_{t-1}, X]_t + b_f) \tag{7}$$

Input Gate This is the second part of the signalling process. In the first part, we have decided that the previous cell state is importation to keep or not. Now it is time to store new essential information on cell state, that will be later judged again by forgetting gate with respect to its importance for the model learning process. Input gate is a multiplication of t − 1 hidden state and t input by input weight matrix, that will be later merged into the new candidate. The activation function of the input gate is sigmoid. Mathematical equation of i_t is as under:

$$i_t = \sigma(W_i.[h_{t-1}, X]_t + b_i) \tag{8}$$

New Candidate Similar to input gate new candidate is multiplication of hidden state's current input with weighted matrix of new candidate denoted with symbol $\tilde{C}_t$ with combination of i_t new candidate will decide with how much information model wants to write on new cell state. Mathematical equation of $\tilde{C}_t$ is as under:

$$\tilde{C}_t = \tanh(W_c.[h_{t-1}, X]_t + b_c) \tag{9}$$

Now cell sate is updated with help of input gate and new candidate the equation is as follows:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{10}$$

Output layer is the multiplication of the weight matrix of the output layer by previously hidden state and current input.

$$O_t = \sigma(W_o.[h_{t-1}, X]_t + b_o) \tag{11}$$

Finally output h_t is product of output layer and hidden state and mathematical expression of h_t is as under:

$$h_t = o_t \times tanh(C_t) \tag{12}$$

3.2.2 Model Optimization

As a model optimization function Stochastic Gradient Descent (SGD) has been used in this study. As our model is not supposed to be linear so slop of non-liner error between two point can be calculated with help of derivative as under:

$$\frac{f(x) = \Delta f(x)}{\Delta x} = \lim_{\Delta \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{13}$$

Cost of the model is always an outcome of the specific function. In our model cost is the difference between the actual price of the entity—predicted price of the entity and based on Mean Square Errors. There are two major parameters that need to be tuned to reach the global minimum level of error.

$$\frac{\delta f}{\delta_\beta} \tag{14}$$

As there in our function of cost two parameters are involved namely, α and β. Because there are two parameters we need partial derivation δ.

$$\frac{\delta f}{\delta_{\alpha}} \tag{15}$$

In the direction of the slop we can calculate all possible partial derivatives and map them on a vector and can be called gradient vector. Mathematical expression is as under:

$$f : R^n \rightarrow R : \nabla f = \begin{bmatrix} \frac{\delta f}{\delta \theta_1} \\ \frac{\delta f}{\delta \theta_2} \\ \vdots \\ \frac{\delta f}{\delta \theta_n} \end{bmatrix} \tag{16}$$

θ is the point toward slop to achieve the global minima and δf changes in function due to change in slop. So, in this way, we can make a vector of all possible partial derivatives to go down to hill.

So, gradient descent update rule is as under:

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla_{\theta} f \tag{17}$$

whereas θ_{new} is updated parameter θ_{old} is old parameter '-' sign means we want to go downhill η is step size that model should take on slop line to go down hill,∇_{θ} is gradient with respect to parameters.

3.2.3 RMS Prop

To really speed up the model learning and error reduction, RSMprop algorithm has been used in the model. The idea behind this algo is to divide the gradient decent into two parts, a gradient that moves in vertical and gradients that moves in a horizontal direction. Vertical movement is called oscillation that is not much beneficial of error reduction. Thus, this algorithm focus on horizontal movement to achieve the global minima.

$$s_{dW} = \beta s_{dW} + (1 - \beta)(s_{dw})^2 \tag{18}$$

$$W = W - \alpha \frac{s_{dw}}{\sqrt{s_{dW}} + \varepsilon}$$
$$s_{db} = \beta s_{db} + (1 - \beta)(s_{db})^2$$
$$b = b - \alpha \frac{s_{db}}{\sqrt{s_{db}} + \varepsilon}$$

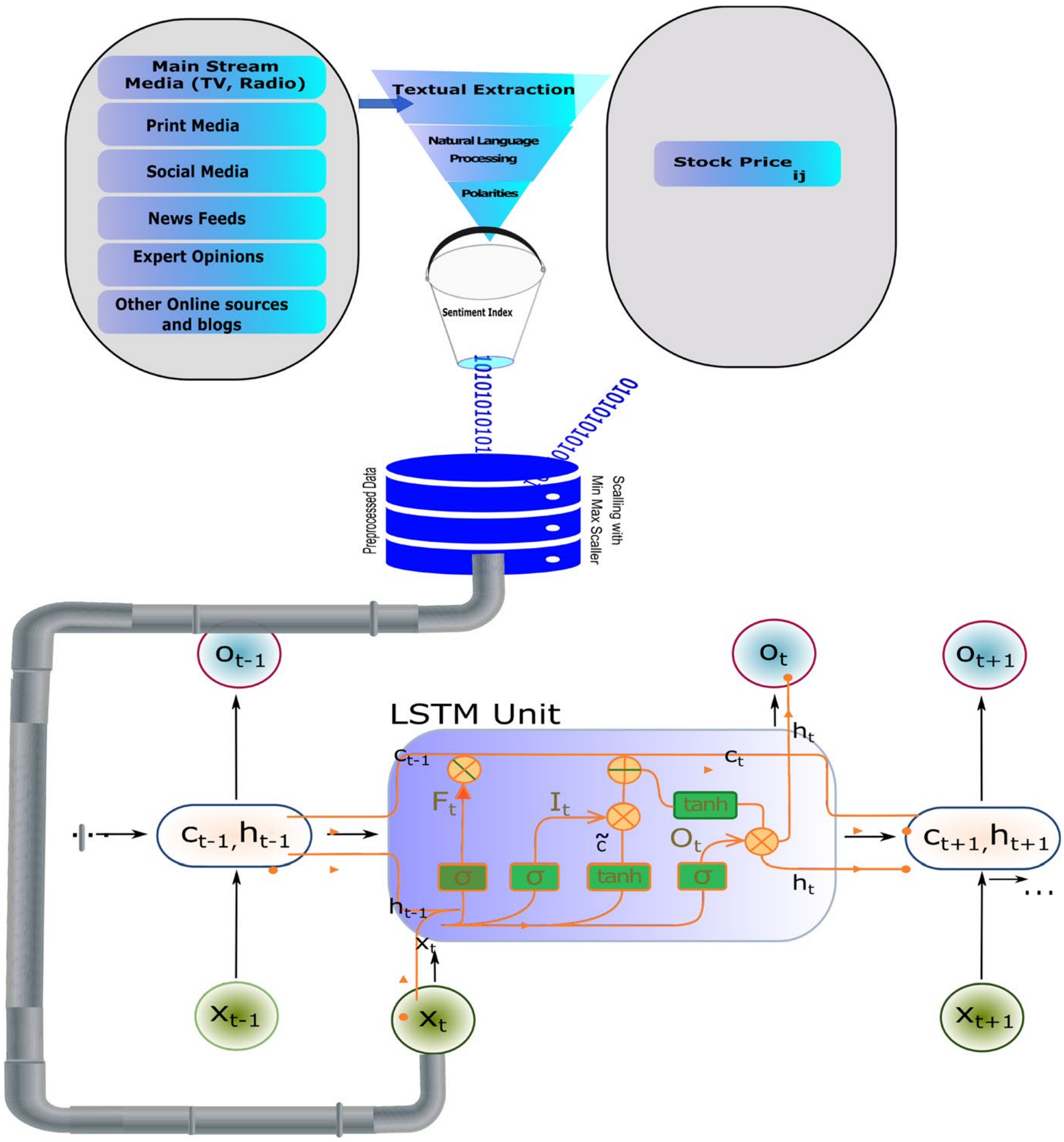

Fig. 2 Study model

whereas s_{dW} is gradient in horizontal direction and s_{db} is gradient in vertical direction. α is learning rate and β is simply parameter for moving average that separate for s_{dW} and s_{db}. Whereas, $(s_{dw})^2$ square of past gradient. ϵ is very small value to avoid dividing by zero.Moving average is effective in this algo because it gives higher weight to current value of gradient and less weight to square of past gradient.

Overall schematic of the study model is as follows.

Now, we will start next section where we have described our results of model (Fig. 2).

Table 1 Training loss for all models

Name/error	With sentiments			Without sentiments		
	MSE	MAE	RMS	MSE	MAE	RMS
Apple shift1	2389.420	48.087	48.882	2590.420	49.087	50.899
Apple shift3	0.546	0.602	0.739	0.606	0.662	0.778
Apple shift6	0.449	0.485	0.670	0.700	0.660	0.880
Bank of America shift1	3.066	1.230	1.751	2048.201	45.225	45.257
Bank of America shift3	0.004	0.043	0.060	8344.511	90.952	91.348
Bank of America shift6	0.010	0.071	0.101	8391.930	91.208	91.607
Cisco shift1	57.922	7.370	7.611	121.695	10.872	11.032
Cisco shfit3	6.259	2.416	2.502	112.721	10.374	10.617
Cisco shfit6	0.027	0.122	0.163	112.424	10.360	10.603
Ford shift1	6.314	2.385	2.513	45.308	6.676	6.731
Ford shift3	0.002	0.029	0.041	137.890	11.625	11.743
Ford shfit6	0.005	0.048	0.070	200.572	13.990	14.162
General Electric shift1	54.135	7.189	7.358	197.547	13.958	14.055
General Electric shift3	0.356	0.573	0.597	140.152	11.659	11.839
General Electric shift6	0.010	0.068	0.100	140.001	11.653	11.832
Intel shift1	127.881	11.022	11.308	202.233	13.988	14.221
Intel shift3	47.959	6.733	6.925	57.391	7.208	7.576
Intel shift6	0.037	0.141	0.193	53.460	6.927	7.312
JP Mogen shift1	511.131	21.853	22.608	1637.782	40.050	40.470
JP Morgen shift3	421.752	19.834	20.537	1359.767	36.464	36.875
JP Morgen shift6	79.236	8.529	8.901	874.040	29.119	29.564
Microsoft shift1	508.827	21.728	22.557	144.701	10.353	12.029
Microsoft shift3	387.778	18.945	19.692	99.047	8.231	9.952
Microsoft shfit6	28.659	5.090	5.353	29.534	4.361	5.435
Pfizer shift1	138.511	11.664	11.769	17.021	3.797	4.126
Pfizer shift3	32.070	5.604	5.663	3.812	1.680	1.952
Pfizer shift6	0.032	0.143	0.179	170.684	12.307	13.065
Well Frago shift1	361.245	18.773	19.006	874.213	29.413	29.567
Well Frago shift3	214.780	14.464	14.655	568.722	23.650	23.848
Well Frago shift6	0.852	0.867	0.923	326.240	17.755	18.062
Sum of errors	5379.275	236.108	243.427	29,003.325	624.264	637.437

4 Results and Interpretations

This section shows the result of model and gives short detail and analysis of the results. Volume wise top ten companies from four major sectors has been selected for analysis purpose. Prediction accuracy results are given in the Table 1. Some figures and tables are omitted from the result section on account of brevity.

The Fig. 3 shows the top ten companies with the highest trade volume during the period of 2008–2016. These top ten companies are the sample that is under the

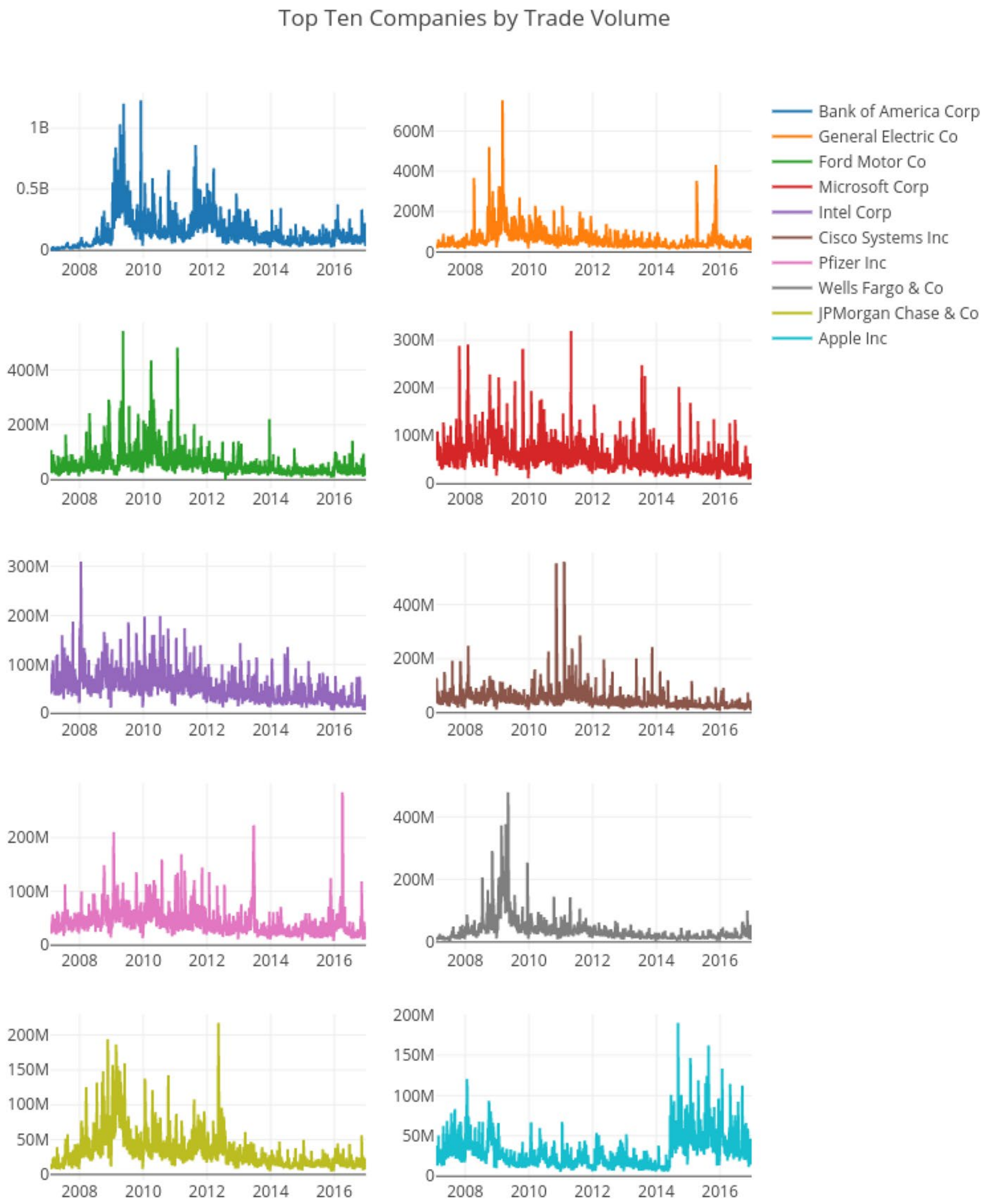

Fig. 3 Top ten companies from different sectors

study in this paper. These ten companies roughly are big names in the financial, IT, medical, electronics and auto-mobile sectors. The reason for selecting diversified companies is to show the reflection of big sectors onto study model. Due to data collection issue, the latest year of study is 2016 but year of the study does not matter in study because purpose of the study to investigate prediction accuracies with machine learning based models and importance of textual analytic (Figs. 4, 5, 6, 7, 8, 9, 10).

There are some interesting results to show regarding the above-shown figures. Six types of different models have been applied to each of the ten companies. The first

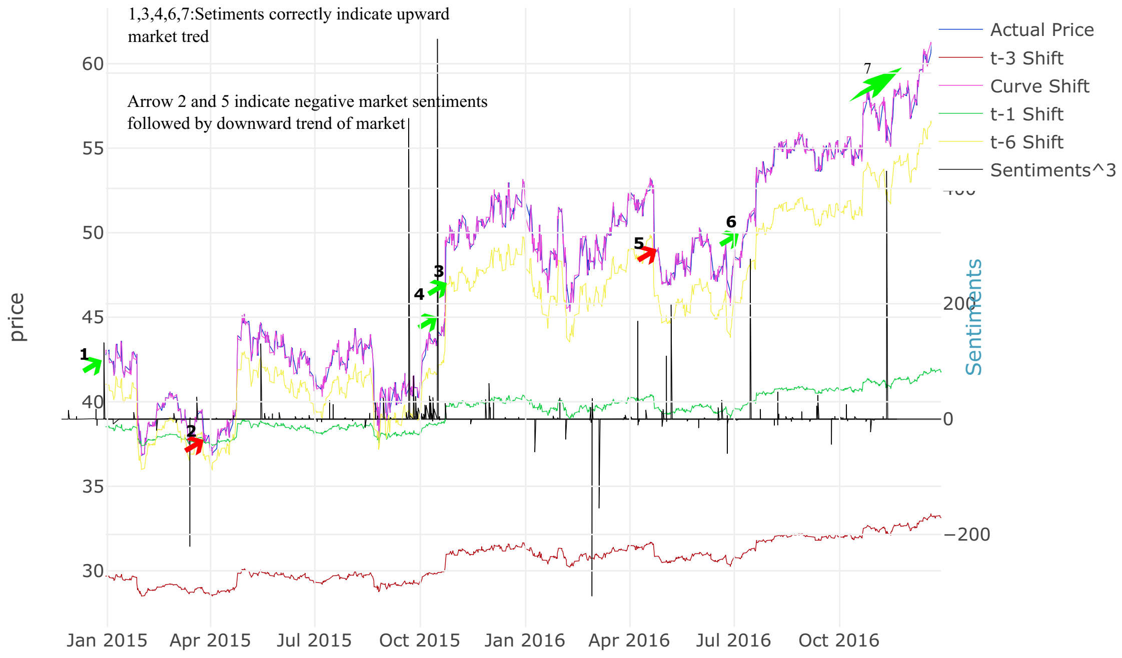

Fig. 4 Microsoft price predictions with sentiments

Fig. 5 Intel price predictions with sentiments

three models are related to the stock market prediction with embedding the company related sentiments and the other three are related to the forecasting without sentiments. There are six types of models for each group i.e with sentiment and without sentiments. The legends in the figures for both types of groups i.e with sentiments and without sentiments is the same. Let's describe the six types of models. The actual price is a simple plot of Actual price for a given period. '$t-3$' is a model

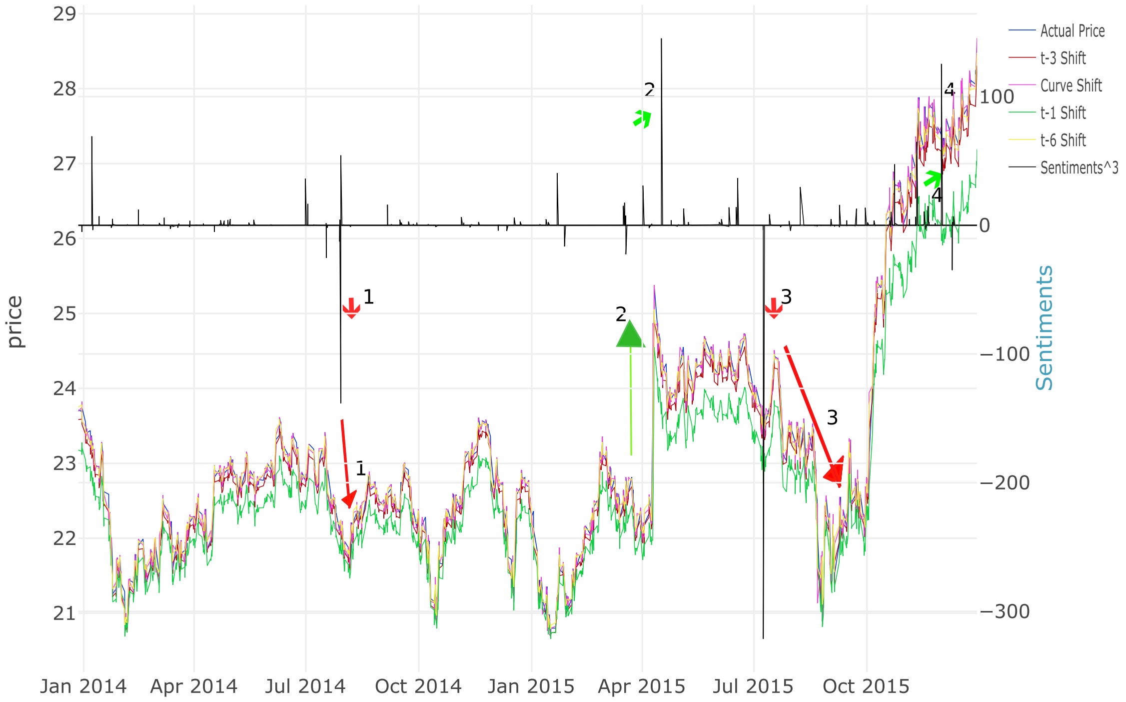

Fig. 6 GE price predictions with sentiments

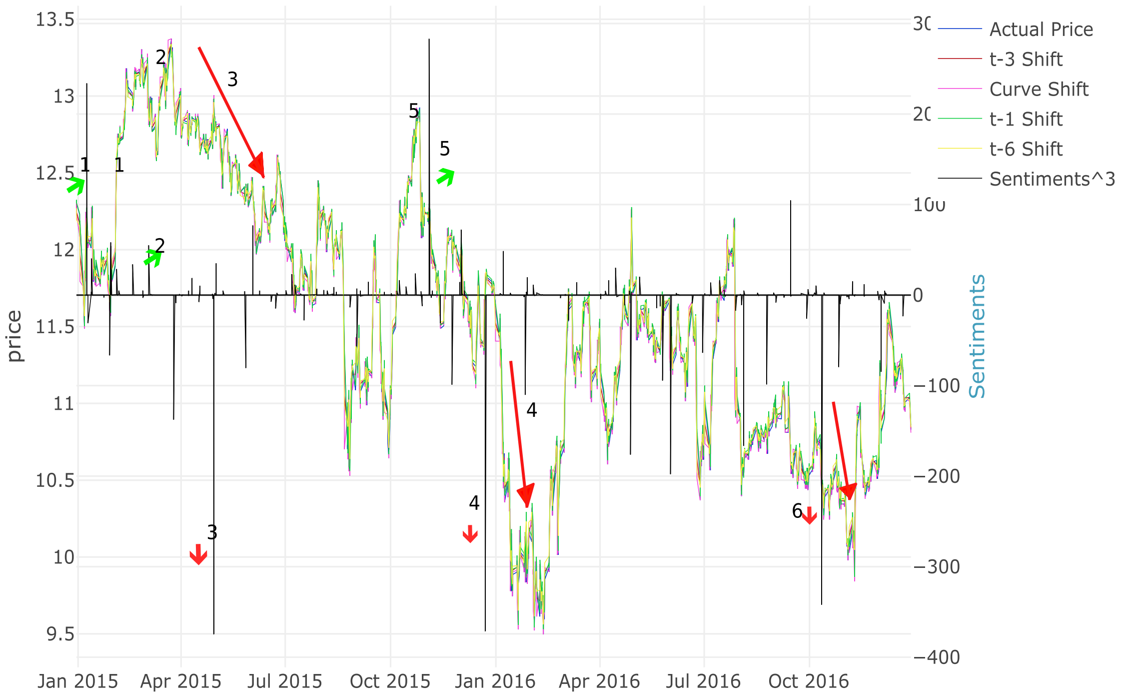

Fig. 7 Ford stock predictions with sentiments

that is based upon 3 h time windows. That essentially means that the model knows the actual price of 1 h in the future as a label for t − 3 time price. 't − 3'. The curve shift is a simple 1-h curve shift without sliding windows. For the sliding window, we mean that model gets a price and certain t and gets the output as a window of 1, 3, and 6 h future prices as the label. So in the curve shift is just the next day as and as

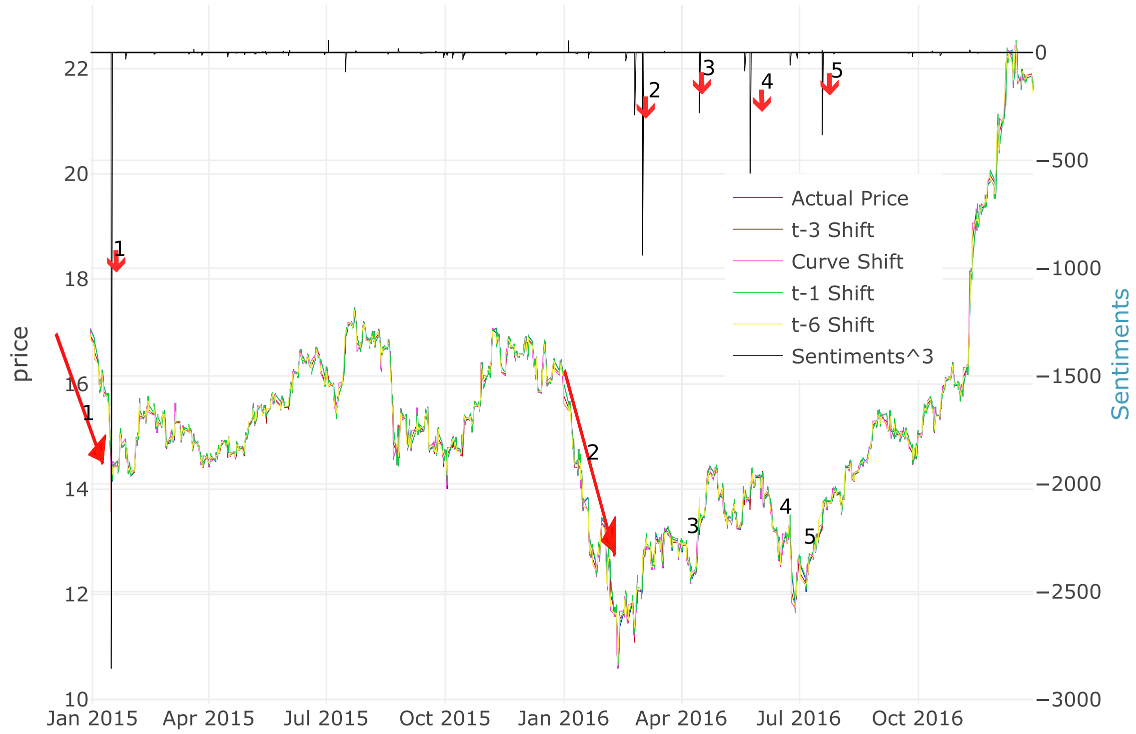

Fig. 8 Bank of America price prediction without sentiments

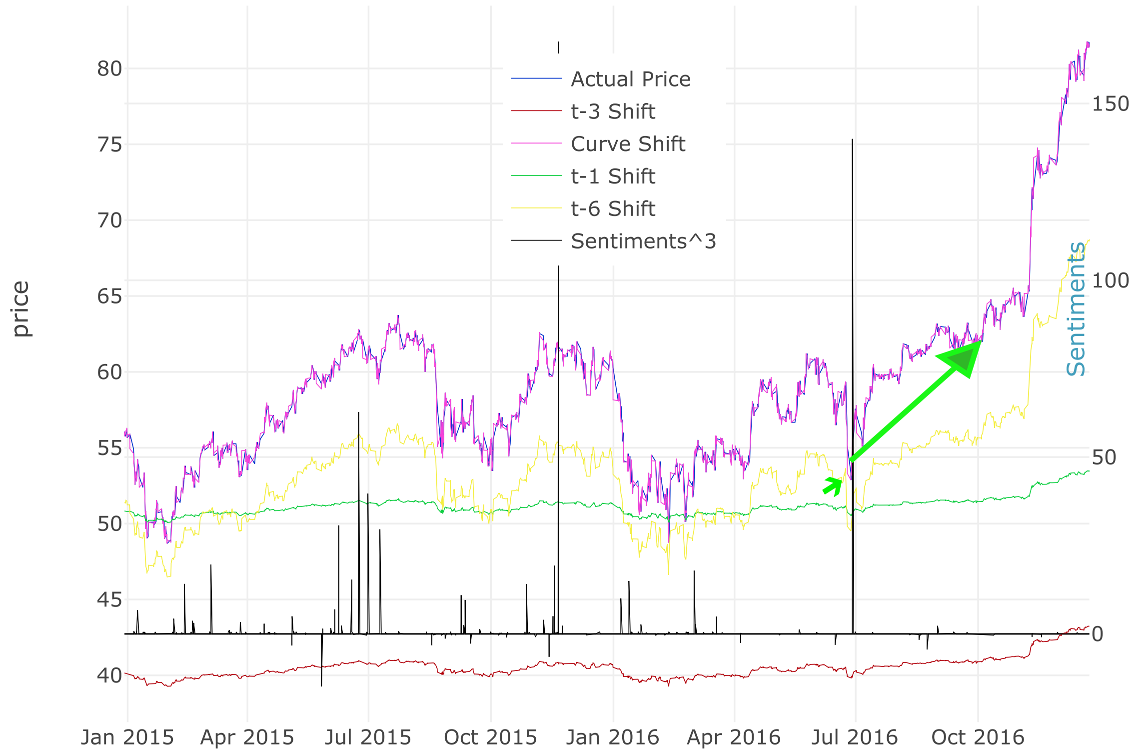

Fig. 9 JP Morgan price prediction with sentiments

the label for current t price. Similarly, t − 1 and t − 6 are sliding windows of 1 and six-time steps respectively. sentiment curve legend only appears in the graphs where the sentiment index has been used as an input in all input models.

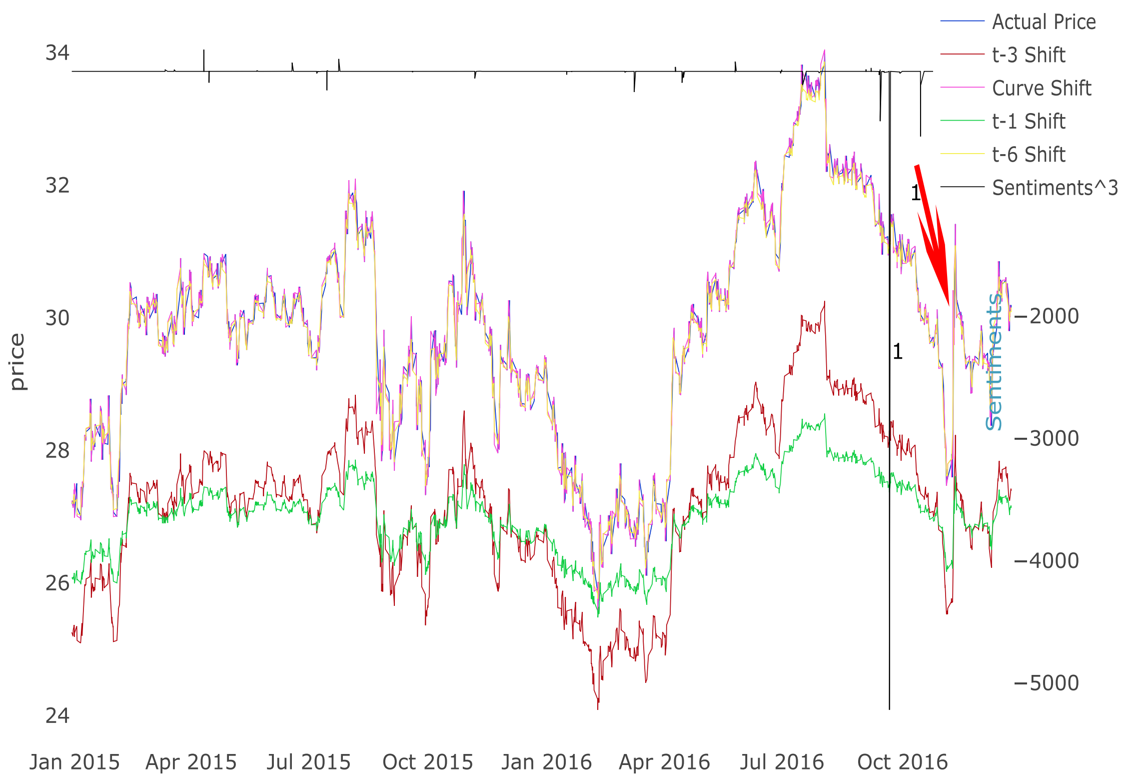

Fig. 10 pfe price prediction with sentiments

The model (t − 1) is looking back to 1 h past data along with sentiments and try to predict the price of the next hour and so on. The second model is looking 3 h back and the third model is looking 6 h back. The purpose behind selecting these different models is to get the idea to what extent the model needs past information to be able to render better results. It can be observed from figures that model is consistently giving very good results when we have given it 6 h of information of the company, as compared to three and one time steps respectively. In most of the cases, the one-time step is relatively the least accurate model and the reason is obvious that the model is getting less information. The sentiment line is plotted on the secondary axis of the figures, as scales of both axes are very different, so to avoid convolution and getting a better overview, we have used the secondary axis. All sentiment scores are exponential with x_i^3. Referred to method section for detailed formulation and algorithm for sentiment scores building. Selected companies are quite large and famous around the world thus the frequency of the company's information is high. So there are many cases of small sentiments that don't influence the market much. So, in the exponentiation process most strong sentiments get prominent and gives better visualization for analysis purpose. The sentiment line is giving a very insightful and meaningful indication for the next market direction. Sentiments on average are getting 2–3 h advance the company-specific information and that information reflects stock direction very effectively.

The results of all six models of study are given in the Table 1. For comparison purposes, three different criteria of model accuracies are given. This is a very comprehensive table that shows a complete training process and achieved the prediction accuracy of the study model. To get a generic overview, the sum for two panels, namely, with sentiments and without sentiment is given at the end of the table. It is obvious from the results that the error sum using all criteria is greater in a case where models don't use company-specific textual analytic.

5 Conclusion

In the recent past, the basic way of operating businesses and corporations, penetrating the new market and reaching to the customers and providing the financial services is exponentially influenced by the new wave of data sciences and artificial intelligence. The research study is motivated by the same phenomenon and empirically investigates the forecasting the stock prices with out-of-box cutting edge soft-computing techniques. The forecasting process is inherited with three unique parts: text analytic, hourly Sentiment index building process, and LSTM AI-based model. First, company-specific text information has been collected, aggregated, classified, and cleansed from thousands of different Thomson Reuters's based information channels that include, mainstream media, print media, social media, blogs, investors advisory services, discussion forums, brokers commentaries. Useful information was lurking in a pile of unwanted information, using natural language processing techniques information is cleaned and useful features have been extracted to be fed to Naive Bayes Classifier to get its sentiments. The second part is building an hourly sentiment index. Though much information from raw text has been collected at the end only three important features related to the sentiment have been preserved for the index building process, namely class of sentiment, the direction of sentiment, and relevance of the sentiments to a specific company. Sentiment information can come anytime round the clock but stock exchange only works for a specific time range. So first, the time of sentiment is matched with the operational time of stock exchange, then based on the self-discovered equation sentiment index has been built. Most of the research studies in similar direction use daily bases stock market values along with other variables but uniquely this study uses an hourly based model for the forecasting process. The reason for the hourly-based model is that getting the accurate influence of the information because 1 day is too late and minute-interval is too early, thus, the direction of stock may not be aligned with sentiments. Third, and final part of the study is the usage of the LSTM neural network model that works in a very special way when it comes to time series or long term dependency of the information.

The results of the study show that sentiments are playing a very important role in the prediction process. Exponentiated sentiments are concisely followed by the big companies traded at US major stock exchanges. That makes our new way of measuring the sentiments robust. Top ten companies from High-tech, financial, medical, automobile sectors are selected, and six LSTM models are applied, three for using text-analytic and other three without analytic being used. Every model includes 1, 3, and 6 h steps back. For all sectors, a 6-h steps based model outperforms the others due to LSTM specialty of keeping long term memory. collective accuracy of having textual analytic models is way higher relative to non-textual analytic models.

Limitation of study Limitation of the study includes a waste of a lot of useful information due to matching the time of news and information strictly with an opening and closing time of stock exchange. However, top companies are operating worldwide and universal time varies across the globe. Almost every hour round the clock information regarding these companies is coming in. but during time matching processing very useful information of almost 18 h has been thrown out. The next challenge in this connection is to come up with a sophisticated mechanism to cope with the issue.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Adebiyi, A. A. (2012). A model for stock price prediction using the soft computing approach.

AlFalahi, K., Atif, Y., & Abraham, A. (2014). Trading and fuzzy logic. *Int J Intell Syst,29*(2), 1–23.

Arora, N., et al. (2019). Financial analysis: Stock market prediction using deep learning algorithms. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)*, Amity University Rajasthan, Jaipur-India.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Systems with Applications, 36*(3 Part 2), 5932–5941. https://doi.org/10.1016/j.eswa.2008.07.006.

Bahdanau, D., Cho, K., & Bengio, Y. (2017). Learning to compute word embeddings on the fly. *Iclr*. https://doi.org/10.2507/26th.daaam.proceedings.070. arXiv: 1409.0473v7

Bouchachia, A., & Bouchachia, S. (2008). Ensemble learning for time series prediction. In *First Int. Work. Nonlinear Dyn. Synchronization*.

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache* [*Linguistics theory: Representation function of language*]. Jena Fischer.

Chang, P. C., & Liu, C. H. (2008). A TSK type fuzzy rule based system for stock price prediction. *Expert Systems with Applications, 34*(1), 135–144.

Cheng, C. H., Chan, C. P., & Yang, J. H. (2018). A seasonal time-series model based on gene expression programming for predicting financial distress. *Computational Intelligence and Neuroscience, 2018,* 1067350.

Cho, K. R., Huang, C. H., & Padmanabhan, P. (2014). Foreign ownership mode, executive compensation structure, and corporate governance: Has the literature missed an important link? Evidence from

Taiwanese firms. *International Business Review, 23*(2), 371–380. https://doi.org/10.1016/j.ibusrev.2013.06.005.

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory, 2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813.

Cooke, M., & Buckley, N. (2008). Web 2.0, social networks and the future of market research. *International Journal of Market Research, 50*(2), 267–292. https://doi.org/10.1177/147078530805000208.

Correa, M., Bielza, C., & Pamies-Teixeira, J. (2009). Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. *Expert Systems with Applications, 36*(3), 7270–7279.

Diaconescu, E. (2008). The use of NARX neural networks to predict chaotic time series. *Computer* (Long Beach Calif).

Elman, J. L. (1991). Distributed representations. Simple recurrent networks and grammatical structure. *Machine Learning*. https://doi.org/10.1023/A:1022699029236. arXiv:1206.2944

Ergen, T., Kozat, S. S., & Member, S. (2017) Based on LSTM neural networks. *IEEE Transactions on Neural Networks and Learning Systems Efficiency*, 1–12.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49*(3), 283–306.

Fellbaum, C. (1998). A semantic network of English: The mother of all WordNets. In *EuroWordNet A Multiling. database with Lex. Semant. networks*, Dordrecht, pp. 137–148. https://doi.org/10.1007/978-94-017-1491-4_6

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054.

Frigola, R., & Rasmussen, C. E. (2013) Integrated pre-processing for Bayesian nonlinear system identification with Gaussian processes. In *Proceedings of IEEE Conference on Decision and Control*. https://doi.org/10.1109/CDC.2013.6760734. arXiv:1303.2912

Gao, Y., & Er, M. J. (2005). NARMAX time series model prediction: Feedforward and recurrent fuzzy neural network approaches. *Fuzzy Sets Systems*. https://doi.org/10.1016/j.fss.2004.09.015.

Hansson, M. (2017) *On stock return prediction with LSTM networks*. https://lup.lub.lu.se/student-papers/search/publication/8911069

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science, 132*(Iccids), 1351–1362. https://doi.org/10.1016/j.procs.2018.05.050.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, W., Lai, K. K., Nakamori, Y., Wang, S., & Yu, L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology & Decision Making, 6*(1), 113–140. https://doi.org/10.1142/S021962200700237X.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing, 10*(3), 215–236. https://doi.org/10.1016/0925-2312(95)00039-9.

Kahneman, D. (2003). Maps of bounded rationality: Economist psychology for behavioral. *American Economic Review, 93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392.

Kahneman, D., & Tversky, A. (1979). Prospect theory—An analysis of decision under risk.pdf. *Econometrica, 47,* 263–292. https://doi.org/10.2307/1914185.

Kryzanowski, L., Galler, M., & Wright, D. W. (1993). Using artificial neural networks to pick stocks. *Financial Analysts Journal, 49,* 21–27. https://doi.org/10.2469/faj.v49.n4.21.

Kuo, P. H., & Huang, C. J. (2018). A green energy application in energy management systems by an artificial intelligence-based solar radiation forecasting model. *Energies, 11*(4), 819.

Lawrence, R. (1997). Using neural networks to forecast stock market prices. *Methods*, pp. 1–21. http://people.ok.ubc.ca/rlawrenc/research/Papers/nn.pdf

Lenat, D., Guha, R., & Pittman, K. (1990). Cyc: Toward programs with common sense. *dlacmorg*. https://dl.acm.org/citation.cfm?id=79176

Li, Y., & Ma, W. (2010) Applications of artificial neural networks in financial economics: A survey. In *2010 International symposium on computational intelligence and design*, pp. 211–214. https://doi.org/10.1109/ISCID.2010.70

Li, Y., Jiang, W., Yang, L., & Wu, T. (2018). On neural networks and learning systems for business computing. *Neurocomputing, 275,* 1150–1159. https://doi.org/10.1016/J.NEUCOM.2017.09.054.

Liu, H., & Singh, P. (2004). ConceptNet—A practical commonsense reasoning tool-kit. *BT technology Journal, 22*(4), 211–226. https://doi.org/10.1023/B:BTTJ.0000047600.45421.6d.
López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications, 42*(6), 2857–2869. https://doi.org/10.1016/j.eswa.2014.11.025.
Makridakis, S., et al. (2018). Forecasting the impact of artificial intelligence, part 3 of 4: The potential effects of AI on businesses, manufacturing, and commerce. *Foresight: The International Journal of Applied Forecasting, 49,* 18–27.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE, 13*(3), e0194889.
Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association, 18*(5), 544–551.
Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. In *2017 International joint conference on neural networks (IJCNN)* (pp. 1419–1426). IEEE.
Omoteso, K. (2012). The application of artificial intelligence in auditing: Looking back to the future. *Expert Systems with Applications, 39*(9), 8490–8495. https://doi.org/10.1016/j.eswa.2012.01.098.
Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications, 39*(16), 12605–12617. https://doi.org/10.1016/j.eswa.2012.05.023.
Pawar, K., Jalem, R. S., & Tiwari, V. (2019). Stock market price prediction using LSTM RNN. In *Emerging trends in expert applications and security* (pp. 493–503). Springer.
Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2016.06.009.
Poria, S., Chaturvedi, I., Cambria, E., & Hussain, A. (2017). Convolutional MKL based multimodal emotion recognition and sentiment analysis. In *Proceedings of IEEE international conference on data mining*, ICDM. https://doi.org/10.1109/ICDM.2016.178
Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A dual-stage attention-based recurrent neural network for time series prediction, pp 2627–2633. https://doi.org/10.24963/ijcai.2017/366. arXiv:1704.02971
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*. https://doi.org/10.1038/323533a0. arXiv: 1011.1669v3
Shefrin, H. (2008). A behavioral approach to asset pricing. *America (NY), 71,* 046123.
Smith, K. A., & Gupta, J. N. D. (2000). Neural networks in business: Techniques and applications for the operations researcher. *Computers & Operations Research, 27*(11–12), 1023–1044.
Swanson, N. R., & White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics, 79*(4), 540–550. https://doi.org/10.1162/003465397557123.
Tan, S., Wang, Y., & Wu, G. (2011). Adapting centroid classifier for document categorization. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2011.02.114.
Trippi, R., & DeSieno, D. (1992). Trading equity index futures with a neural network. *Journal of Portfolio Management, 19,* 27–27.
Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2017) Using deep learning to detect price change indications in financial markets. In *25th Eur Signal Process Conf EUSIPCO 2017*, 2017-January, pp. 2511–2515. https://doi.org/10.23919/EUSIPCO.2017.8081663
Wang, M., Zhao, L., Du, R., Wang, C., Chen, L., Tian, L., & Stanley, H. E. (2018). A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Applied Energy, 220,* 480–495.
Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE, 10*(1109/5), 58337.
Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review, 50*(1), 49–73. https://doi.org/10.1007/s10462-017-9588-9.
Yoon, Y., Swales, G. S., & Margavio, T. M. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society, 44*(1), 51–60. https://doi.org/10.2307/2584434.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14,* 35–62. https://doi.org/10.1016/s0169-2070(97)00044-7.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Faisal Khalil[1] · **Gordon Pipa**[1]

Gordon Pipa
gpipa@uni-osnabrueck.de

[1] Institute of Cognitive Science, 49090 Osnabrück, Germany

2.2 Transforming the Generative pretrained Transformer into Augmented Business Text writer

Author: Faisal Khalil:

co-author/ supervisor: Prof. Dr. rer. Gordon Pipa:

Authors Contribution:

Faisal Khalil:

The research paper is written by Faisal Kahlil during Ph.D. at the department of cognitive science University Osnabrueck.

Prof Gordon Pipa:

Prof. Pipa supervised the overall research concept and has given a lot of valuable suggestions, directions, and continuous thought to the research paradigm, that reshape the research paper.

Peer Review:

Accepted at: Journal of Big Data Springer

https://www.researchsquare.com/article/rs-1170589/latest.pdf

Noname manuscript No.
(will be inserted by the editor)

Transforming the Generative pretrained Transformer into Augmented Business Text writer

Faisal Khalil · Prof.Dr.rer.Gordon Pipa

Received: date / Accepted: date

Abstract This study uses transformers architecture of Artificial neural networks to generate artificial business text for a given topic or theme. The implication of the study is to augment the business report writing, and general business writings process with help of Generative pretrained transformers (Generative Pretrained Transformer (GPT)) networks. Main focus of study is to provide practical use case for GPTs models with help of big data. Our study model has 355 million model parameters and trained for three months on GPU enable devices using 2.3 billion text tokens(is available as open-source data now). Text tokens are collected with help of rigorous preprocessing, which includes; shortlisting of Subreddits of Fortune 500 companies and industries, listed on US-based social news aggregation online portal called "Reddit". After shortlisting, millions of submission of users during the five years, are parsed to collect the URLs out of it. 1.8 million working URLs are scrutinized. Business text is parsed, cleaned, and converted into word embeddings out of Uniform resoruce Locator (URLs). The result shows that both models; conditional interactive and random sampling, generate text paragraphs that are grammatically accurate and stick to the given topic.

Keywords Natural Language Generation · Transformers · Business text generator

PhD candidate at Institute of Cognitive Science,Universität Osnabrück,Germany
Wachsbleiche 27,D-49090 Osnabrück
Tel.: 49-(0)541-969-3380
E-mail: fkhalil@uni-osnabrueck.de

Prof.Dr.rer.Gordon Pipa Professor and Chair of Neuroinformatic Group at Institute of Cognitive Science,niversität Osnabrück,Germany
Wachsbleiche 27,D-49090 Osnabrück
E-mail: gpipa@uni-osnabrueck.de
Tel.: 49-(0)541-969-3380

1 Introduction

With the passage of time, the field of artificial intelligence, and machine learning have been made progress by leaps and bounds. Nearly all fields are getting benefits from the cutting-edge technologies to leverage their processes, and Deep learning is one of them. Big tech giants are reformulating their strategies to align with AI and ML. Deep learning is a branch of Machine learning that enhances the model learning process with its deep layered architecture. Like many other walks of life, Deep learning has won its spurs as a very effective and efficient technique for natural language processing related tasks. Since, computers are unable to understand the natural language, enabling them to understand the natural language and to process the information in a useful fashion has long been under the researchers' and practitioners' focus.

This study is inspired by the new method implement by the Google Brain team Vaswani et al. (2017) and the work of OpenAI Radford et al. (2019). Before introducing the transformers implement by the above-cited research work, it is important to shed the light on the recent past of Natural language processing (NLP). Although Natural language Processing (NLP) has deep roots in the past and the first breakthrough was the well-known paper of Alan Turing 'Computing Machinery and Intelligence' Turing (2009), real progress in the field has been made in the late 1980s - when machine learning algorithms came into the picture. The machine learning revolution has permanently changed the approaches to address NLP related problems. At the start, mostly much stress has been given to rich text features embedding - to enables Artificial Neural Networks (ANNS) to understand the rich text in numerical form. Later these embeddings are given to an end-to-end neural network that essentially maps the input and output, i.e McClelland and Rumelhart (1989). Later one, seminal work published related recurrent neural network Rumelhart et al. (1985). Recurrent models are very important for natural language processing because natural language caries lexical, syntactical, and semantic context in it- thus previous words or characters are very important to solve machine translation and text prediction tasks. In the year 2002 Jürgen Schmidhuber and his students Gers et al. (2002) came up with a better idea for neural network application that involves long-term dependencies, named, Long Short Term Memory (LSTM). Long Short Term Memory (LSTM) devises some gating and sates mechanism that keeps import information from the previous sequence and also memories the previous state that finally accumulates to the current state to predict the next sequence. Many enhancements have been made by the research community in the recurrent neural network model. The most highlighted models are seq2seq (sequence to sequence) Sutskever et al. (2014a), Jacovi et al. (2018). Seq2seq models essentially work with encoders and decoders recurrently to encode the output of the previous sequence and combine it with the current input. The next enhancement in recurrent model is attention mechanism, see Xu et al. (2015a),Yang et al. (2016). Attention mechanism has been proven very well in machine translations, where two pairs of sentences of two languages are mapped together with encoders and decoders.

So, looking back to the short history of the evolution of the natural language processing techniques, we understood one common limitation of all these models concerning solving the NLP task is the models are computational resources hungry and very slow. NLP corpus normally involves an enormous amount of training data, long-term dependencies, and recurrent nature. These factors make the training process very slow to achieve the desired result. Addressing this problem, the research community has come up with multilayered attention head and encoder decoders - formally called TransformersVaswani et al. (2017). The current study uses a similar approach to generate the domain specific text, and detailed methodology is discussed in 3.2. We have used a recently developed transformer neural network architecture. This architecture is primarily used for Google translation works in two different blocks, namely, encoders and decoders. We have only used the decoder part. We have provided the model with a 2.3 billion text token during the training. The model has 355 model parameters and has been trained for 3 months to reach a 2.6 training loss value. Above-mentioned 2.3 billion text tokens are collected after rigorous data preprocessing steps. US-based social news aggregation and discussion forum has been selected for data collection purpose. Almost 700 Subreddits are shortlisted for the purpose of getting URLs out of it. Millions of submissions for five years have been considered. Submission means any post, comment, or reply by the user. Users often redirect towards URLs for clarification. So, 1.8 million URLs are collected from the submissions, and validation and functionality of all URLs have been confirmed. With the help of a parser, these URLs are parsed and cleaned to get the text. Finally, 2.3 billion ready to feed to the model word embedding has been generated. In rest of the paper; literature review, Methodology of the study and model, results of the study and limitation and future suggestion have been given respectively.

1.0.1 Research gap

After getting the flashback of the evolution of the NLP and recent developments of NLP, we can see one common problem for all Natural language understanding problems is creating a relationship matrix between the words or characters and giving importance to the specific word at a specific place. Solving this problem is very important for all NLP-related niches, for example, Natural language understanding, Natural language generation and, machine translation. In this connection, we have mainly two problems to be solved. Problem no 1 is again giving importance to the words and specific place in the sentence and creating correlation or context to each word embedding based on their usage. The second very problem is supplying a lot of data or in other words a lot of instances to the model to learn the placement and relational pattern of the characters or words. Giving a lot of data needs a lot of words' embeddings matrix that leads to extremely slow model training and a lot of computation resources. So, the computational and efficiency problem is more lethal as it seems to get a breakthrough of problem No. 1. The research community either could wait for the computation resources to get more efficient and faster enough to solve the problem at hand, or they must have to come up with an optimal solution. So, the solution to this problem was attention mechanism Vaswani et al. (2017) and most specifically transformer architecture of neural networks, formally called encoders decoders Radford et al. (2019). well, fair enough transformer can, theoretically, overcome the above-mentioned problems and give a new horizon to the landscape of NLP and NLG, but we need to provide a lot of real-life use cases and proof of concept to supplement this new ANNS architecture. After this conceptual breakthrough, the next challenge is to come up with a lot of data and preprocess that much big data to supply it to these new models to proof the concept of the conceptional invention. Our paper is exactly filling this gap here by coping with the challenge of developing the proof of concept and practicality of this new advancement of NLP and deep learning. So, in this journey the most important step is to find a use case; so, we have chosen business-related reports and text writing. In the next subsection, we will give precise details where and how this concept can be used in a commercial setting and what benefit it can promise. Coming back to the current point, getting a lot of business-related data is very important as well very hard because of a lot of irrelevant text and without the authenticity of being business text. So, involving humanized efforts to tag data is very costly and not plausible. So, we decided to use "reddit" a platform, widely used, and each post is voted by the community. In this way, we could get human checked data in huge volume, related to the business problems. it is also relevant to mention here that we did not parse data from "reddit" directly, rather we have only collected URL links from the posts, and then we parse complete URLs text. So, our main contribution here is rather less on the theoretical side and more on the practical side. As we have retuned and adopted the existing theoretical concept in a more practical setting to provide its proof of concept. after having this discussion, it's very relevant to provide one hypothetical application instance and possible commercial usage of this study. So, next subsection talks about the hypothetical ideal use case and overall generic use cases of the study.

1.0.2 Hypothetical use case:

Let's here create a practical scenario. In the office and business management, there are a lot of reports and text writing, for example, Manager X has to give a job placement ad for a consultancy firm, or, he has to write an advertisement. He has to write a small report about his product and its competitor in the industry he is operating to get external funding. In such cases the grammar is not only an important factor but are pinning words other people are using in the industry to influence more or clarity of text is maybe more important. Let say a software application helps Manager X in two ways; first, gives a context or appropriate usage of words replacement based on millions of other use cases already people used in similar instances. Second, if he writes "Apple Inc.". the application suggests him, i.e., "Apple has launched iPhone pro max. in 2020 that gave them xxx hundred thousand $ annual revenue". So, now Manager X can save a lot of time and energy in surfing google in searching facts and figures. if some assistance is provided on how he can paraphrase any keywords, could improve business writing greatly. I know that requires a lot of work on front-end development too, but the Black box part would be NLG here.

1.0.3 Practical Implication:

The study has great potential for real-world practical uses: for example, next-word prediction, topic modeling to extract text out of scanned images, contextual soundness of the business writing, and suitability of word usage even if it's grammatically correct in the first place. Any subject-specific knowledge, language usage, and vocabulary are always different compares to generic languages. Many companies and start-ups have software applications that are using a similar approach but use general language text. Here is a list of some: Gmail salutation and common words autofill used during the email GoogleEMail (2021), Grammarly Grammarly (2021) gives words context suggestion and content clarity based on the text they have trained upon. At the start of registration, they asked for purpose of use. Maybe something like Grammarly business writer or something similar could be the very practical use of this study. Reverso Translator gives translation based on the frequency of usage of the word in literature along with text, except where the looked-up words have been used. There is the potential of usage of such tool is there where one can give the accurate context of the only business-related text. Lastly, we did know at the time of conducting this research, but one online platform emerges now which is using augmented writing approach with greater success having a top-level firm in their customers' portfolio, i.e. see Writing (2021). This would be a very true practical usage of such a study. There is not only business related application of language generation model but also applied to many filed. i.e van Deursen et al. (2020) introduced Generative Examination Networks (GEN) to generated chemical space.Similarly, Bagal et al. (2021)

2 How deep learning integrate into corporate sector?

The literature on the Natural Language Processing is root back in the 1940s. After parsing the literature, the evolution of NLP can be segregated into different phases; for example, the journey started from machine translation problems, followed by the computers and information technology revolution - that triggered the AI applications into this area. After AI and machine learning came into the picture — complex task solving ability has been improved with less time — thus grammatical structure has been focus more. After advancements like deep learning and reinforcement learning, NLP has now entered into artificial text generation and generated text is hardly differentiates from human written.

Though the research community of that time had been working on NLP, the first scientific paper was published by the MIT language department head, William. N Locke and A.Donald Booth, head of the Brick-Beck collage Locke and Booth (1956). Machine Translation (Machine Translation (MT)) started with three dominant languages of that time, English, Russian, and a bit of Chinese. Computational resources were too scarce and much effort had to be exerted on converting data in bits ALPAC (1966). Early birds in this area have given focus to syntactical computational processing of language, and it was important to first draw the basic structure for the language Plath (1967). Work of Ceccato (1967) some researchers have tried to shift the focus from the syntactical to semantic oriented language processing. Ceccato tried to co-relational analysis between the same pattern of a pair of languages and tried to achieve the semantic driven language processing. Winograd (1972) and Woods (1978) have seen the 1960s transformational grammar theory is a misfit of computational grammar and analysis and not offering much in terms of semantics. The computational confidence approach is given by Woods' and Winograd's enriched the previous work in a semantic path.

Later on, in the 80s, AI came into the picture and the community has shifted their focus toward a machine leaning based approach for solving the existing dilemmas of NLP in a pure semantics way Schank (1980). In this decade, researchers have realized that the NLP task such as building the word representation to use in AI-related networks and pining the context is very hard. Some note able work of the 1980s is as follows: Briscoe et al Briscoe et al. (1987) have built a general-purpose grammatical formalism including syntactical analyzer for the English language with help of suboptimal software, named Grammar Development environment (Grammar Development environment (GED)). They also program software to build and manage a large grammar base. Towards the direction of speech recognition, Young et al Young and Chase (1998) have led to major US speech recognition projects, called, Continuous speech recognition (Continuous speech recognition (CSR)) and (Long vocabulary speech recognition (LVCSR)). The paper includes tools and methods for news transcription, text dictation, and transcriptions.

The next phase of the NLP development is the 1990s, that mostly focuses on a combination of lexical and syntactical approach for natural language processing. After lot of twists and struggle of almost two decades, the statistical and probabilistic approach has been adopted for classification tasks in NLP Sparck Jones (1992). Later on, these models became raw sources of machine learning related techniques to solve the NLP complexities. for example, Manning and Schuetze Manning and Schütze (1999) have worked on information retrieval, feature extraction out of it, and analyzing the textual information with statistical models. Mani and Maybury Maybury (1999) have used terminological logic to built a knowledge base for automatic information extraction and text summarizing. By the end of the 1990s, dialogue speech system and language processing had expanded the horizon with multilingual text machine translations, speaker-independent speech to speech dialogue system. Wahlster (2000) has worked on project Foundation of Speech-to-Speech Translation- so-called, 'Verbmobil'. This multilingual (German, English, and Japanese) takes input in a speaker-independent manner and translates them into other desired languages. it also handles domain-specific business spoken dialogues and translates into other languages with approximately 80 percent accuracy. The struggle of many years make the NLP researchers, practitioner, and industry realize that linguistic resources are inevitable for the further development in this filed, thus, two institutions, "British National Corpus" BNC (2020) and "WordNet" Fellbaum (1998) are come into being. The next era of natural language processing started after 2001. Though many models have been proposed by the researchers which were other than neural networks, we are only discussing the neural network-oriented important models in this paper.

Bengio et al Bengio et al. (2003) proposed tri-gram state-of-the-art neural probabilistic model. They have used a neural network for the probability function. The idea is based on the conjecture that unseen words get a higher probability to be predicted based on the similarity of the words - on which the network is trained. The next word prediction approach has many practical uses commercially, for example, see the work of Kannan et al. (2016) that can generate a small short semantic reply of the email.

The next advancement in the field of NLP is multitask learning, off-course this method is not only confined to the NLP but a general enhancement in the neural network world. Collobert and Weston Collobert and Weston (2008) have tried to implement this technique for transfer learning. Vector representations of the words have been fed as an input to the model to do word prediction and then learning of the current model was transferred to the other independent model to achieve a similar but not the same task. The multi-task learning approach was first introduced by the Caruana Caruana (1998). Once, so-called, word vector representations are fed to the neural network, they start learning the context and association of each work with the other. Transfer learning makes it possible to share the learned weight across the models for generalization and incremented learning approach. During the optimization process, it is very important which parameter to transfer. Ruder Ruder et al. (2019) proposed that the sharing parameter can also be learned during the learning process. See also similar research McCann et al. (2018). In this connection, the next milestone was "vectors representation" of the text, so-called word embeddings. This basic word embedding idea was first floated by mikolov Mikolov et al. (2013). They have proposed that removing the hidden layer while training the word embedding is giving more promising outcomes. Later on, this idea paved the way for the concept 'word2vec' and originally adapted to two popular approaches, namely, bags-of-words and skip grams. This phenomenon has triggered the research interest in this direction and many researchers have enrich this concept see; Arora et al. (2016), Mimno and Thompson (2017), Antoniak and Mimno (2018), Wendlandt et al. (2018). The current direction of the word embedding is to train a very large corpus and use used pre-trained embeddings for multilingual models in an independent and unsupervised fashion. for example, see. Artetxe et al. (2018), Conneau et al. (2017), Søgaard et al. (2018).

In the year 2013 and 2014 neural network architectures are being applied to NLP, the most obvious choice was recurrent, recursive, and convolutional neural networks. simper Elman Elman (1990) RNNs were replaced with LSTM by Hochreiter and Schmidhuber (1997) because of long-term context dependencies in input text. secondly, convolutional networks are originally dealt with computer vision areas but also implemented in NLP for example see the work of Blunsom and Grefenstette (????) and Kim (2014). The obvious plus of the using convolutional network is they are more parallel and local context based on layers rather than past state contrary to the LSTMs.

Concerning recurrent neural networks, the next enhancement was a sequence to sequence modeling (seq2seq). Seq2seq model is using the same recurrent architecture of the neural networks, but the important bit is disguise in encoding and decoding procedures. The input sentence is first encoded into a vector representation. The decoder then tries to decode the predicted symbols based on the encoder state sequentially. The sequence to sequence model was proposed by Sutskever et al Sutskever et al. (2014b). Later on, in the year 2016 Google Google (2020) has decided to change its monolithic sentence based machine translation to complete neural network-based. Now, seq2seq models are the foundation of language generation models and further developments, i.e transformer-based neural network architectures. Similarly, image captioning Vinyals et al. (2015) is using the same technique to generate the image captions automatically. The seq2seq model leads toward attention mechanism and transformers based approaches. The basic limitation of the seq2seq network is that it tries to compress the whole sequence of the sentence and then convert it into a fixed-length vector. Thus, the model cannot look into the hidden state. Attention mechanism, by contrast, looks into the hidden state of the model combine them to realize how much stress should be given to a specific word. Attention Bahdanau et al. (2014) was the core innovation in the field of neural machine translation that permanently replace the traditional methods of machine translation. Have a look on different flavors of attention based networks and their application; reading comprehension Hermann et al. (2015) , entity parsing Vinyals et al. (2016) , image captioning Xu et al. (2015b).

The pretrained model has gain popularity among the NLP research community. The main advantage of the pretrained model is that it is context agnostic and unsupervised model. Labeling for the NLP task can be very costlier and challenging. So, the pretrained model captures the meaning and context of one language and the leanings can be transformed into the other language to get the meaning and context generation or translation. The pretrained model was first proposed by Dia and Le Dai and Le (2015). The current study is also based on pretrained multi head attention based model.

3 Methodology

In this section we have described how data is prepossessed and then processed data is fed to the model is discussed in detail. The completely prepossessed data will be available as an open-source data for further research and development.

3.1 Data Preprocessing

In this section, we have described the process of data preparation for model training. Everything else with respect to the neural network model is similar to many other applications of ANNS, but the main concept here is to leverage the training process with an enormous amount of training data. Websites could be the potential source of a lot of textual data as well as a great deal of diversity in it, but the bottleneck with websites' data is the validity of data and too much unnecessary information in it. Following the research by Vaswani et al. (2017) we have adopted a similar approach and choose 'Reddit' reddit (2021a) - a USA based social news aggregation and discussion platform with 330 million users reddit (2021a) to collection the website URLs to parse the data form. To ensure the validity and usefulness of the web URLs , only those links have been taken that contained more than 3 'karma'. 'Karma' is so-called assurance given by the other user about the validity of comments and discussion. In this way, we have got a human level quality check on the data. Once we have devised the mechanism of data quality, the next filer was to get the URLs that are only related to the business and Fortune 500 companies . Most of the top 500 companies have their discussion and news profile on 'Reddit' called 'Subreddit'. 'Reddit' has a very large community and thus, thousands of submissions are committed on a daily basis. The raw data, ranging from 2005 to 2017, is first programmatically collected with help of the 'Reddit' programming interface reddit (2021b) and stored in the 'BigQuery' database. In the next step, we have extracted all the URLs having 'karma' ranking more than 3 from the daily submission of the users. These URLs are verified, whether they are working or not and at the end 1,852,482 working URLs list was prepared to parse the textual data from 'Hyper Text Mark Langauge (HTML)' tags. With the help of parallel computing and a computer grid, 20 GBs of text files have been collected from all working URLs. These 20 GB text files are gain filtered for some unnecessary characters and symbols. Finally, the 2,302,554,291 text token were collected to be converted into word embeddings. The process is shown in figure

1a that depicts a flow of data preprocessing with help of a schematic diagram. preprocessing involves:

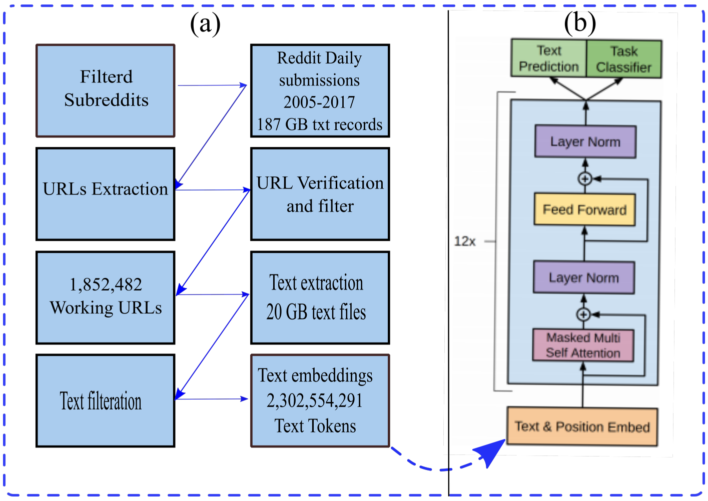

Fig. 1 Data preprocessing and network architecture

3.2 Methods

Next comes the transformer neural network model applied to preprocessed data. The Transformer model takes all words tokens are encoded into words embeddings, that is nothing but the numbers that represent each word. Normally, transformers have two parts, encoders and decoders, but we have only used the decoders part of the Transformer because both encoder and decoder are feasible for machine translations- that is not the case in this study. See figure 2 how general transformer works, originally designed for machine translation problems. This architecture was later adopted and modified by many researcher and lab to improve NLP and translation related problems. If you pay closer attention to the paper Vaswani et al. (2017), you will realize transformers are also basically a from of transfer learning where sentence of the language one are pass through many layers of self-attention and feedforward neural network layers and update the training weights keeping the relationship of each word within the sentence and position of each words into mind, whereas, learned weighted of language one are transferred to feedforward layer of decoder part to learn the nature of relationship and position or grammatical aspect into mind when model tries to predict the words in the second language. That is how essence and context of sentence are translated correctly. So our case is rather different from machine translation, thus second language inputs' weight are not possible here.So, we stick to the decoder part of the model as a main model architecture. coming back to the point of data processing, Words embedding are stored and converted into NumPy zip format for simplicity purposes. first, we

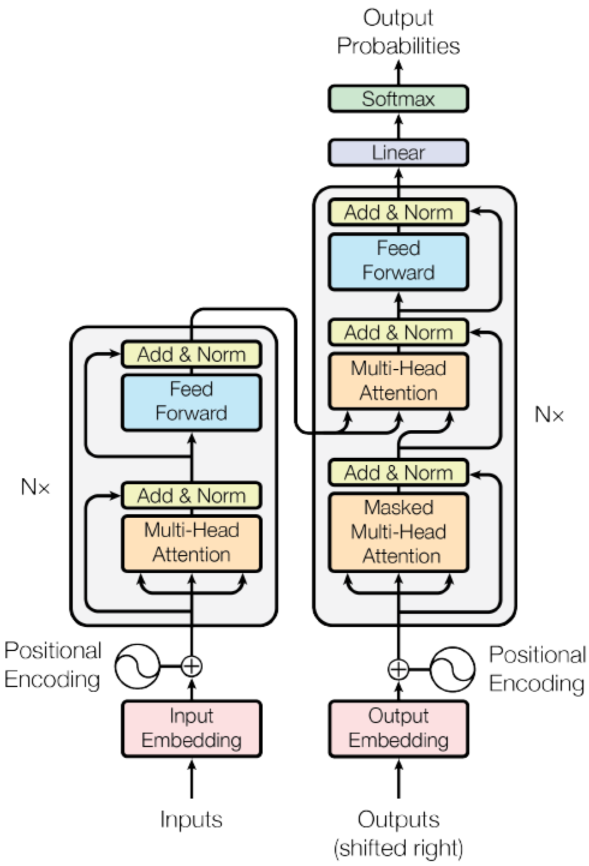

Fig. 2 Transformers general network architecture Vaswani et al. (2017)

will see the high-level representation of the model, and then we will look into how the self-attention layer is working. The model gets the words embedding as input, it assigns positional encoding to each word. The positional encoding keeps the position of the word into a sentence to capture the context efficiently, contrary to random order. Word embedding along with its positional information passes through the self-attention layer. The self-attention layer is twelvefold layers.

For analogy purpose, we can say this layer create many copies of the sentence and map the relationship and importance of each word in the sentence to figure out how much attention to the specific words is to be given. That is why it is called a multi-head self-attention layer. We can plunge into the self-attention layer to see how it is working. Input vector $\mathbf{X}_1..\mathbf{X}_N$ is multiplied by three different vectors, namely, Query vector($\mathbf{q}_1$), Keys vector ($\mathbf{K}_1$) and value vector($\mathbf{V}_1$). The vector is random weights of dimension 64 and the output of these matrices' multiplication is $\mathbf{W}^Q, \mathbf{W}^k, \mathbf{W}^v$. In the next step, we get the dot product of ($\mathbf{q}_1 \cdot \mathbf{K}_1....\mathbf{K}_N$) for sentence (1....$n$.). To stabilize the gradient process, each output is then divided to the ($\sqrt{d_k}$), whereas, d is dimension of the vector k. This operation gives us scores for each word. higher the sores means that more attention should be given to that word. In the next step all the scores for on word related to all other words should be summed up into a variable $\mathbf{Z}$:

$$\mathbf{Z} = softmax\left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_k}}\right) \times \mathbf{V} \tag{1}$$

This is the final calculation of one out of many self-attention layers, that is to be fed - in a matrix shape, to the feed-forward neural network. To focus on different positions of the words in the sentence we need, multiple representational subspaces, subspace is achieved with the help of multiple head or copies of the attention layer. so ;

$$\begin{aligned} \mathbf{Q}_i....\mathbf{Q}_n &= \mathbf{W}_i\mathbf{X} \\ \mathbf{K}_i....\mathbf{K}_n &= \mathbf{W}_i\mathbf{X} \\ \mathbf{V}_i....\mathbf{V}_n &= \mathbf{W}_i\mathbf{X} \end{aligned} \tag{2}$$

whereas, $i...n$ is the number of attention layers. $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ is the query, key, and value vector and $\mathbf{X}$ is the word embedding input matrix. So, every attention layer produces a $\mathbf{Z}$ matrix and depending on how much attention layers being chosen, in our case 12. The attention output matrices $\mathbf{Z}_1....\mathbf{Z}_{12}$ are multiplied with the weights' matrix jointly for all layers, called $\mathbf{W}_O$. The resulting matrix is input for a fully connected feed-forward network. The final output of the feed-forward network is then decoded back to the words to generate the sequence of the sentence. For the clarity of the dimensions of the different matrices, please refer to the table 1

Matrix	Dimension
$X_1 \ldots X_n$	upto 512 depends onlength of sentence
Every W	64
X $\times W$	DX×64
Z	DX×64
W_o	DX×64

Table 1 The table gives the dimension of the different matrices

4 Results

In this section, we have described the results of our study. In this section, we have presented text samples that are generated by our trained model. The results include a sample from both conditional and unconditional samples. Conditional sampling means that we have provided a certain keyword to the model as an input and the model has returned a text paragraph related to that given keyword, however, unconditional means random samples generated by the trained model. Training loss summary of the 'Tensorboard' model is given in the Appendix in section8. To support out the accuracy of model and the sample are not appear out of chance, we have given 100 randomly generated sample by the model in the appendix under section 8

We have trained the model up to 460,000 steps. Since the model has almost a 355Million model parameter and more than 2.3 billion text token, the model requires extremely excellent computation power and time. The model has been trained for 3 months on a single GPU and settles on a loss value of 2.6. This value of loss for the text-based model is quite reasonable because the language model always involves complex grammatical chains like dependencies and structures that are not easy to capture. The next two subsections provided real-time model generated text, both based on conditional and unconditional random outputs.

4.1 Interactive Conditional outputs of the Model

This subsection provides 5 different output samples of the interactive conditional sampling method of the study model. This is so-called interactive model outputs, in which the model communicates with the user. The user gives input/keywords to the model and the model generates a text paragraph that mostly talks about the given keyword/topic. Given below are the table boxes that show output against five different user given inputs.

Table 2 Results of Interactive conditional samples

Key:	**conditional output:**
Apple iPhone	7 Plus: 3-month free trial 15% off $150+ purchase iPhone 7 Plus: 3-month free trial 15% off $150+ purchase 4G LTE: iPhone 6s: 2 year data plan 3 GB data, $49.99/mo 8GB data, $59.99/mo 4G LTE: iPhone 6s: 2 year data plan 3 GB data, $49.99/mo 8GB data, $59.99/mo iPhone 5s: 2 year data plan 5 GB data, $45.99/mo 8GB data,$69.99/mo

output generated by the trained model against user given topic/key.

4.2 Unconditional outputs of the model

In the section below we have given some table boxes that contain the random sample output generated by the model. this is an artificial text written by the model. If we observe the generated paragraphs, it is very clear that the text is following the grammatical rule mostly and topics of the sample pointing towards the business-related text. An enormous amount of sample can be produced on demand, due to the brevity of this article we have only given some sample.

Table 3 Results of Interactive conditional samples

Key:	**conditional output:**
oil&price	“which make up the global oil industry. The used state-of-the-art computer models to show that the price of crude from North Dakota has increased by a factor of 40 over the past year. In the past year, it has risen by 28 percent, a period of unprecedented energy pricing, according to the report.That is an almost unheard-of percentage increase, said Tim Leissner, an assistant professor at the University of Michigan who specializes in energy and economic modeling, which is used in energy forecasts. ”Usually, the percentage increases are pretty much the same year after year, so that’s a pretty consistent pattern,” said Leissner, who was not involved in the report. ”Normally what we would expect is that the price goes up because of oil production, and it generally goes up at the end of the year.”But that is no longer the case, said Leissner.”The oil industry has a new president in Trump, and his agenda is very aggressive on a range of issues. And the industry is not only seeing this increase, but it is actually seeing it drop,” Leissner said. The increase in North Dakota oil production has gone from 6.5 million barrels per day in the first Automatic half of 2017 to 7.3 million in the second half of 2017 and 7.3 million in the first quarter of 2018, according to the report. And in the last two months alone, North Dakota’s production has fallen from 4.6 million barrels per day in the first half of 2018 to 3.3 million in the second half of 2018. The increase in oil production in North Dakota has also affected a number of other energy prices, with natural gas prices rising about 12 percent in the first half of 2017 and natural-gas prices rising about 13 percent in the first half of 2018.The data comes as the Trump administration prepares to review the Obama rule that requires oilfields on federal lands to submit to federally supported research and exploration programs. The rule will affect oil and gas development in the state. Leissner said the oil market will be monitored for ”the next few years” for the first time in decades. ”This kind of increases is a big deal, given that oil is the only major product in the world that people are going to be producing for years to come,” he said.The research is published in the Journal of Energy Technology. The report, which included”

output generated by the trained model against user given topic/key.

Table 4 Results of Interactive conditional samples

Key:	**conditional output:**
Microsoft Windows	“users, if you don’t already have them or you’re thinking about installing them on your Windows Vista/Windows 7 PC, you’ve got to pay the price. There’s also the issue that you’re going to have to run Microsoft on your current PC, which isn’t ideal if you’re a Mac user. If you only do this once, and you’ve got Mac OS X installed, Windows 7 will work. The problem with this is, it’s a pain to use on those platforms. Even if you’ve got Mac OS X on your computer, it’s going to be a pain to use in Windows, and if you want to use Windows Vista on a Mac, it’s going to be a pain on Windows. So we’re going to need to figure out a way to get Windows 7 running on existing Macs. We’ve been trying to do so, so far. And at some point, we’ve found something that’s going to let us do it. We’ve been working with the folks over at Xiph.ai and are able to install Automatic and upgrade and upgrade, and I don’t want to go into too much detail, but we actually managed to install Windows 7 on Mac OS X through the Xfce application. And if you do this just once and you’re running Windows 7 on a Mac, it will run. Xfce and Window 7 To get Xfce running for Windows, right-click in the Xfce launcher, and then choose “Xfce Installer”. This will open a pop-up window in which you can choose what to install. The first option is “Mac OS X 10.4.5”. You can choose the installer from the list below. If you’re using the installer option, you’ll notice that Xfce has been created for you. And you’ll also see an “Xfce.app” folder for Windows. In order to get Xfce running on Windows, we’ll need to run the installer from that folder. Open Xfce by going to its application menu. Select “Install Xfce” from the search box. You’ll notice it’s not”

output generated by the trained model against user given topic/key.

Table 5 Results of Interactive conditional samples

Key:	**conditional output:**
S&P 500	“index fell 1% and the S&P 500 and Dow Jones Industrial Average gained 0.7% and 0.4%.The Dow closed up 8.2%.Meanwhile, the S&P 400 surged 3.8%, with the Nasdaq Composite index up 8.3%. The Nasdaq rose 0.4%, while the S&P 500 3.6%.While Wall Street was off the back foot, on the financial markets, the Dow rose 1.8% and the Nasdaq rose 0.8%.In the energy market, the Standard & Poor’s 500 index dipped 1.2%, the Nasdaq index rose 3.6% and the S&P lost 4.1%. In real estate, the S&P 500 rose 0.8% and the Nasdaq climbed 1.2%.Meanwhile, the Nasdaq’s gains were smaller in the oil energy group, while it gained 0.4% and the S&P 500 rose 0.3% in residential and business.The Nasdaq rose 2.0% and the S&P 500 gained 0.8% in the energy sector.At the top of the industry, the Automatic S&P 500 closed 6.9%, with the Nasdaq up 2.6% and the S&P 500 up 2.1%.In the energy group, the Dow rose 8.7%, with the Nasdaq rising 2.4% and theS&P500 adding 4.5% in the energy segment.And in residential, the S&P 500 rose 1.7% and the Nasdaq rose 1.3%.In the business sector, the S&P 500 rose 2.5% and the Nasdaq climbed 1.1%.The Dow was up about 7.8% as the S&P 500 fell 1.4%, while the Nasdaq declined 1.2%.The Dow hit a record high of 11,868. The S&P 500 was up 11.4% and the Nasdaq rose 0.4%.The financial sector was trading at a record high as investors worried about the economy and a potential Federal Reserve interest rate hike, which has taken hold following a sharp decline in the central bank’s benchmark interest rate.The Dow closed up 8.7%, the S&P 500 rose 3.8%, and the Nasdaq rose 3.9% ”

output generated by the trained model against user given topic/key.

Table 6 Results of Interactive conditional samples

Key:	**conditional output:**
heath care industry	“in a manner that protects the privacy of patients. We will continue to take this necessary step to ensure that our industry is transparent and accountable,” the department said in its statement. In addition, the department plans to ”deregulate the sale of any health care services without a patient consent,” and ”work towards an industry-wide ban on involuntary sex trafficking. The Department of Justice has been involved throughout the issue of sex workers and the use of technology, and is actively working with regulators and private-sector partners to better address the issue of human trafficking, the company stated. ”With recent advancements in technology for technology’s sake, sexual exploitation is no longer confined to the virtual realm, but now has become a reality,” it added. The department said it ”will continue to work with state, federal and private sector partners to improve their policies and laws regarding human trafficking, including protections and the use of technology to protect victims of sex trafficking.” The government’s goal is to help ensure that ”we do not, in any way, further endanger any of our citizens, particularly those whom we have entrusted with our law enforcement and public Automatic safety responsibilities.” The department added, however, that it ”will not tolerate law enforcement personnel, employers, contractors and contractors to be complicit or assist in unlawful human trafficking, as that might undermine our mission.”¡—endoftext—¿”This domain is no longer available, possibly due to a change in the registrar or the domain administrator.”¡—endoftext—¿”What started as a series of tests to see if it could survive the cold, the hot and the humidity is turning into a real life experiment that will test the idea that this tiny space heater might have a practical purpose. For this project, the team set out to conduct the tests using just a standard fridge - just a standard fridge, for short. But they wanted to see what it could do if the temperature inside it dropped by 10C - a scenario that will be a reality in 2017. The first test will bring in an air temperature of -30C - something which will kill any kind of bacteria as well as fungi in about two days. As you can imagine, that wasn’t the worst of the worst in the test. A typical fridge - even the kind you’ll find in the kitchen - was just as bad as the prototype at catching bugs. The team has now developed a small freezer for the freezer, which can withstand even lower”

output generated by the trained model against user given topic/key.

Table 7 Results of Non-conditional samples

Sample	**Machine Generated Text**
Sample1:Topic Microsoft	SA-4, but some of the other models that it supports can be set, as can the version of Windows. However, most users will need to install the free and limited OS X Lion operating system. To get it the free version, you can purchase it from the Apple site. In the past few years the Windows phone market has grown significantly. Microsoft has been aggressively supporting Microsoft Office applications in Windows 8. The company is still selling a wide selection of Office applications. The reason Microsoft has decided to target the Android market, especially is that, as a company that is heavily invested in smartphone users, Microsoft will be less apt to change direction. Still, it is unlikely that you're going to be surprised to find that Windows Phone has managed to outcompete Android in terms of support level and functionality. That has been the case for years now, and not just because of the various operating systems. However, things have changed for the better. The development time has gotten slower, and the hardware has gotten more modern. There have also been increased efforts to make the operating system itself more user friendly. Microsoft has been steadily working to increase the range of features available on the platform. In the case of Windows Phone 8, this means that it supports the latest version of the operating system, the Universal Windows Platform (UWP), which has evolved in ways that made the system more accessible to the new users. The new version of the Windows Phone OS also integrates with the new Metro UI, which has been available for several years already. This means that the interface is easier to navigate with each new update. There is an obvious difference between the Windows Phone version of the Windows OS and other systems, which makes the difference in the Windows Phone OS much stronger. As a result, the Windows Phone OS is likely to enjoy a much wider appeal. The Windows Phone OS is a much more mature OS, however, and it may prove to be even more attractive once it is officially supported by Microsoft. In that case, it seems that the Windows Phone OS can only prosper as long as Redmond will provide more hardware devices that can be run this OS. It is, of course, a very hard problem to solve. However, it seems inevitable that this issue will play a greater role in Microsoft's future strategy. Microsoft has been focused on offering a wide range of popular consumer and enterprise computing options in order to take advantage of the growing mobile market. There is also a good chance that the introduction of Windows and Office to the marketplace will bring a greater opportunity for Windows to become a mainstay for mobile devices. Further Reading¡—endoftext—¿A new video showing a drone carrying a baby to her birth, while also showing her running and jumping, is set to debut at the London premiere of David Cronenberg's "Puff Daddy," at the V&A on Wednesday. Puff Daddy follows a teenage girl whose father is killed in an accident and has been left with an orphaned daughter. Her ex is a young woman from a nearby suburb who has a passion for flying and is looking for a way to give back to her community. "You can't have a child without a parents," said Cronenberg, who directed "All the Money in the World" and "O Brother, Where Art Thou?" alongside Brian De Palma and John Frankenheimer, as well as "Shakespeare's Son," "All the Money in the World" and "The Other People's'" alongside Tom McCarthy. The director also showed off new CGI footage of the film's main characters, including the first scene where they're shown playing with the baby and flying. "Puff Daddy" is shot in a sequence that features a close-up of the girl and the baby. "We wanted to create a visual effect in ways that were visually appealing," Cronenberg said. The scene with the daughter flying was filmed in the streets of North London, but Cronenberg said the scene in the film's last shooting, "Rudolph the Red-Nosed Reindeer," will also appear in the visual effects package. The project also features a "flying baby" sequence that was filmed in a nearby suburb.

Automatic output generated by the trained model

Table 8 Results of Non-conditional samples

Sample	**Machine Generated Text**
Sample2:Topic: Health Care	," said the chief executive officer of the British medical charity, Beaumont Hospital. "In recent years we have seen a surge in the number of young people coming into this country seeking to change our society, but the risks of those who do become radicalised remain too high. "Young people like Mika are at risk of radicalisation and may be vulnerable to becoming radicalised themselves through viewing social media as a possible means of radicalising themselves or others. "Our advice is to work closely with the police and other relevant authorities to help these young individuals to understand the risks in the communities they may come into contact with in the future and to talk to parents about their responsibilities." It has long been feared that social media is inextricably linked to radicalisation. Earlier this year it emerged that the police were monitoring 4 million posts on Twitter, Facebook, Kik and Line, all forms of instant communication, for signs of terrorism. But Dr John Ralston of the University of Oxford has claimed that although "social media has been used for a long time in the UK, and indeed throughout Europe, some parts of society have never noticed it." He said that while people in certain sections of the community have been concerned at the recent rise in extremism, many young people in other parts of the population have not. "The vast majority of young people at one time or another have encountered such people through social media."¡—endoftext—¿The biggest financial institutions have the greatest exposure to the market, yet they are the most transparent, according to Transparency Market Research (TMR). The research group analyzed more than 700 leading financial companies, looking for those who reported some form of transparency, or disclosed more information than allowed. The findings, based on the organization's annual survey of 1,250 U.S. companies, show that firms with the largest exposure to the financial market have disclosed the highest amounts of transparency – even though they are less transparent than the average firm. Those firms with the most transparency are also the companies that the study found to the highest risk of market abuse, including: • A number of the top 50 firms made disclosures in excess of 30 percent of their company size. • The majority of firms did not disclose their disclosure forms during the year.

Automatic output generated by the trained model.

Table 9 Results of Non-conditional samples

Sample	**Machine Generated Text**
Sample3:Topic Energy Market	to have more time with people," he says. "I am sure I will enjoy sitting on my porch with the trees behind me. It will be a really relaxing time."For all the progress the film has made this year, however, as the box office has surged, so has that of the franchise. The latest installment, starring Tom Hiddleston, will bow on Monday, while the two-part drama "Hiddleston: Longmire" will arrive in U.S. cinemas on February 19, followed by its worldwide debut in April.¡—endoftext—¿Hodl has been selling the technology for years. On the shelf of any Walmart, Walmart.com or Amazon, it's not uncommon to come across a shelf full. But now a new technology to turn your clothes into a new energy-generating asset comes to San Francisco and Silicon Valley. As the world warms and urban temperatures rise, the amount of energy stored in the fabric of the clothes and other products increases. This process has the potential to revolutionize a whole range of industries and technologies—as well as the way people buy and use clothes. Firms that design, produce, market, sell or install this technology face a number of problems, including the technology's limitations—which includes the inability to make use of the wind. And in a major market, technology can be confusing to customers. If an installer sees a new technology on a product, its not clear where to go. Is it a product or an energy-generating product? It's an industry with a number of big names, all of which are focused on the same thing, but each company has a different brand. Sprinting out of the closet Fifty years ago, it was a relatively new idea that was made possible at the beginning of the Cold War by advances in the nuclear power and hydrogen bomb. This meant that we could get more power with less fuel than any previous technology had before. But more power, as it turns out, is less efficient than it used to be, and that has led to more problems than it is solving. The technology was called the "energy storage." The technology used energy (or energy storage) to cool a part of fabric. It was the first technology for storing energy. Then this technology fell out of fashion. In the mid-1970s, it was popularized by Motorola, a firm that produced smart refrigerators that could store energy and heat the refrigerator. This technology became the fuel cell and energy storage industry before it fell out of favor.Sprinting out of the closet The reason we still have to solve problems with energy storage is partly due to the sheer size of the industry. It takes lots of expertise to be able to make use of the technology, but we still find ourselves getting it wrong. The biggest reason are twofold. First, the technology was extremely effective early on, but as technology has made more efficient, then it has gotten difficult to get it right. Second, the technology is often confusing and difficult to integrate into existing products. If you're in an industry where the problem with storing energy is a lack of innovation, you may find the technology confusing and cumbersome and you may have to learn how to use multiple energy storage products to get the right kind of energy. The companies behind the two companies—Gigasolar, in Palo Alto, and GigaSolar, in Sunnyvale—are trying out new energy storage technologies that make an energy store. Sprinting out of the closet is a common solution to those issues. However, there are companies that are able to solve the first problem using their previous technology without ever looking at the second problem. One company that has done this is Gigasolar, in Sunnyvale and Stanford. Gigasolar's company mission is to get rid of the technology gap between what the customers need and what the companies are built for. It's not a simple business. They have to design a product that solves the problem before it can be fully put to use. The company had to do this. Gigasolar, which is run by the founders Peter and Adam Zweig, are very involved in figuring out how to work within the energy delivery space. They have a number of patents that cover multiple products. In these patents, they have some of the best energy energy storage patents that you can find. For instance: They use carbon dioxide to store energy in the atmosphere and then they release carbon dioxide back into the atmosphere to convert it into new energy.

Automatic output generated by the trained mode.

Table 10 Results of Non-conditional samples

Sample	Machine Generated Text
Sample4:Topic Retailing	production are not the same as production at a retail store like you are likely to see at the Superstore. I recently did a bit of testing for this.You Can't Buy This There are two types of products that can make it through the Superstore checkout process. The first are the items that are sold using the same pricing structure as the retail store. For example, the price for the T-Shirt above only goes up \$0.75 with the same color. All I had to do was go to the Superstore and double check my shirt price with the same information as I use to determine if my T-Shirt cost me half what it cost on the same item in the Retail Store. The T-Shirt above, when bought online, has been \$0.75 up charge, but on sale is \$0.85 for the same size of shirt (you can see why the \$0.75 difference is so important!). The second type of product that can help you make it through the checkout process are items with the same price that you'll find on the retail store. This is usually with items like apparel. For example, the same shirt is \$.25 on the Superstore (a T-shirt cost me \$0.85 on the Retail Store for the same size of Shirt), as long as they look the same they're actually a bargain. If you're a Clothing shopper and shop your clothing online, you'd probably get the price of those same garments for \$0.75. In my opinion, the Superstore prices for the same items are more consistent, so I've decided to include them. To find the Superstore price for Clothing, you need to purchase a Clothing App item online, take the price as your price, then click the purchase button. You should then follow the steps above to determine the price for that item. A good starting point would be the same price that you pay for the Clothing App item in the Retail Store. Here's an example of the Clothing App item that I would purchase on the Retail Store. The Shirt above, I bought for \$27 online.If you purchase this on the Retail Store, \$27 would be the price you'd pay online: \$37.99. If you'd purchased it on the Superstore it would be \$0.80. I would take my clothing price and multiply it by the retail value of the Shirt (i.e. \$0.80 is what I think the T-shirt is on sale at the Superstore). I would assume this will be the same price as my T-Shirt. To do this, I just have to multiply the actual price by my price, then subtract my retail value from my Retail Store price to create a final price of \$0.40 for the T-shirt. As you can see, my total Clothing App price is \$38.50. To keep track When you're shopping online, take in one value from each App item for those two different sizes of the Shirt (the same ones) then calculate that value and convert to a per-item price for the product you're looking at while shopping. The final price should be that. So let's say that we have the Shirt below for \$25 that we think would be the same \$25 price for the same Shirt on the Retail Store. After we calculated the per-item value for the Shirt, we would add \$0.80 to our total clothing price for that T-Shirt to calculate our final price. In this case, that final price would be \$0.80. This is the final value of that T-Shirt – \$0.80.

output generated by the trained model.

4.3 Discussion

In this section, we are going to discuss the results of the study and how these results stratify the problem inference of the study. The main focus of this study is to testify the validity and useability of current theoretical development in the field of natural language generation and generally Natural language processing. As we stressed out previously the long dependencies chain of one word to other words, placement of the word in a given sentence and relational space of the word and characters is the big challenge of language generation-related problems. This problem was very difficult for recurrent neural network models to cope up with. So, the researchers came up with different theoretical concepts. In this connection, we are providing practicality, useability, and proof of concept of the model in our study. For this purpose, we have provided two types of results, interactive conditional and non-interactive random samples. How we have trained the model, iteration, and loss graph can be observed in the appendix section. However, the main objective of the study was to generate the text that sticks to the overall topic of text, formally called topic modeling, secondly grammatically correctness, and thirdly, somehow related to business only. If we closely read the results of section 1, we have provided the model with random business-related diverse keywords from all different business genres. Model is not only able to collect very related text, but also supply some facts and figures. Moreover, the linkage of sentences and story making is very decent. That does the perfect job for the hypothetical use case and highlighted research gap. Additionally, for the robustness of the model, we have also created a random sample of text generations with thousands of instances. Due to brevity, we have provided some samples here, and we have given link to could to access all other thousands of samples. In both types, an interactive and non-interactive, model is achieving the initial goal for context, relatedness, and topic modeling. of course, this is just a founding block to generate any meaningful commercial soft application. we need to assemble other pieces of puzzles, namely front-end development, scripting, and mapping for words PACs of words matching the counts and statistics concerning the whole database, with many more bumps and stager on the journey down this road.

5 Conclusion

The current study is focusing on the application of Natural language processing in the field of business writing. In the recent past, the Deep Learning research community has come up with a new architectural style of deep layered AI models that are aligned with the specific need of natural language and text generation. The transformer is one of those models that are proven to very accurate and effective in context and grammar capturing in the text.

Response to the possible question, what is the purpose of the study very briefly? The study uses a generative pretrained neural network model. The model is fed with a lot of business-related preprocessed text data acquired parsing the 1.8 million URLs collected from Reddit. As a result of the trained model, user can give keywords or some topic to the model and model produced paragraph that completely sticks to the given topic, provided that the topic or keyword is in the domain of business or management sciences. These features or results provided by the study can be utilized in automatic paragraph prediction to assist the business report writer or any relevant person involves in the writing process. As there are many applications available for next word prediction generally but paragraph prediction is lacking.

Now, let us give little more details on how data is preprocessed, and the model is trained to get the output? A large amount of quality data is very important for language processing models. To address the quality issue we have chosen 'Reddit'; news aggregation and content sharing platform. Although Reddit covers a lot of different topics, we have shortlisted 'subreddits' - topic-specific Reddits. There is a huge amount of Reddit submission every day and 'KARMA' vote is given to the post that is helpful for the community. So we have collected 1.8 million URLs from those submissions that have 'KARMA' vote greater than three. In the next separate step, we have collected and cleaned all the text available in the URLs.

In the end, 2.3 billion text tokens have been fed to the model. The model has 355 million parameters. After three months of model training, the model can generate grammatically correct and aligned with business topic text as a model output. In the coming subsection, we have discussed what could be the practical application of the model and future suggestions along with some limitations of the study.

5.1 Implications and future work suggestion

There are many possible implications of this study. One possible use is market intelligence report writing. Possibly a piece of software can be developed to auto-complete the paragraphs for business intelligence report writing. Any business-related industry can be benefited with help of paragraphs prediction instead of just word prediction. In this way, the speed of efficiency of the user can be enhanced significantly. As for future suggestions are concerned, we think that text token prefixed by the theme or topic of the text can make this model more useful. For example during the training text, at the start of the text, we can provide what this piece of text is talking about. In this way, we can have greater control over the output of the model we can generate real-time long reports based on specific keywords. The report is just one example we can utilize the model is much more effective ways. We hope that the research community is maybe already doing something in that direction.

5.2 Limitations of the study

We have tried to do the study at our best, but there certain technical limitations of the study. Since the models related to text generation is usually based on an enormous amount of training data; that is a very important factor to capture the grammatical structure and relatedness of the topic, this study only relies upon the text generated from those web URLs that were discussed in business-related Subreddits. The study may be improved significantly with help of having more sources of training data and more computational power.

6 List of abbreviations

NLP Natural language Processing
GPT Generative Pretrained Transformer
URLs Uniform resoruce Locator
ANNS Artificial Neural Networks
LSTM Long Short Term Memory
MT Machine Translation
GED Grammar Development environment
CSR Continuous speech recognition
LVCSR Long vocabulary speech recognition
HTML Hyper Text Mark Langauge

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

7.4 Competing interests

The authors declare that they have no competing interests

7.5 Funding

Not Applicable

7.6 Authors' contributions

The research study is written by Faisal Khalil during the PhD at University Osnabrueck, Germany. Prof.Dr.Gordon Pipa has supervised and improve the study in many aspects.

7.7 Acknowledgements

No Acknowledgements

7.8 Authors' information (optional)

Mr. Faisal khalil is doing PhD at university Osnabrueck, Department of cognitive sciences. His major areas are artificial intelligence and Deep learning. His main focus is application of AI in the areas of business and corporate finance.

References

ALPAC (1966) Language and machines computers in translation and linguistics

Antoniak M, Mimno D (2018) Evaluating the stability of embedding-based word similarities. Transactions of the Association for Computational Linguistics 6:107–119

Arora S, Li Y, Liang Y, Ma T, Risteski A (2016) A latent variable model approach to pmi-based word embeddings. Transactions of the Association for Computational Linguistics 4:385–399

Artetxe M, Labaka G, Agirre E (2018) A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. arXiv preprint arXiv:180506297

Bagal V, Aggarwal R, Vinod P, Priyakumar UD (2021) Molgpt: Molecular generation using a transformer-decoder model. Journal of Chemical Information and Modeling 62(9):2064–2076

Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473

Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of machine learning research 3(Feb):1137–1155

Blunsom P, Grefenstette E (????) Nal kalchbrenner, et al. 2014. a convolutional neural network for modelling sentences. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics

BNC (2020) British national corpus. URL https://www.english-corpora.org/bnc/, accessed on : 2020-07-04

Briscoe T, Grover C, Boguraev B, Carroll JA (1987) A formalism and environment for the development of a large grammar of english. In: IJCAI, Citeseer, vol 87, pp 703–708

Caruana R (1998) Multitask learning. autonomous agents and multi-agent systems

Ceccato S (1967) Correlational analysis and mechanical translation

Collobert R, Weston J (2008) A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning, pp 160–167

Conneau A, Lample G, Ranzato M, Denoyer L, Jégou H (2017) Word translation without parallel data. arXiv preprint arXiv:171004087

Dai AM, Le QV (2015) Semi-supervised sequence learning. In: Advances in neural information processing systems, pp 3079–3087

van Deursen R, Ertl P, Tetko IV, Godin G (2020) Gen: highly efficient smiles explorer using autodidactic generative examination networks. Journal of Cheminformatics 12(1):1–14

Elman JL (1990) Finding structure in time. Cognitive science 14(2):179–211

Fellbaum C (1998) Towards a representation of idioms in wordnet. In: Usage of WordNet in Natural Language Processing Systems

Gers FA, Schraudolph NN, Schmidhuber J (2002) Learning precise timing with lstm recurrent networks. Journal of machine learning research 3(Aug):115–143

Google (2020) Alphabet inc. URL https://www.google.com, accessed on : 2020-07-04

GoogleEMail (2021) Gmail. URL https://www.google.mail.com/, accessed on : 2021-11-15

Grammarly i (2021) Grammarly. URL https://app.grammarly.com//, accessed on : 2021-11-15
Hermann KM, Kocisky T, Grefenstette E, Espeholt L, Kay W, Suleyman M, Blunsom P (2015) Teaching machines to read and comprehend. In: Advances in neural information processing systems, pp 1693–1701
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735–1780
Jacovi A, Shalom OS, Goldberg Y (2018) Understanding convolutional neural networks for text classification. arXiv preprint arXiv:180908037
Kannan A, Kurach K, Ravi S, Kaufmann T, Tomkins A, Miklos B, Corrado G, Lukacs L, Ganea M, Young P, et al. (2016) Smart reply: Automated response suggestion for email. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 955–964
Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:14085882
Locke WN, Booth AD (1956) Machine translation of languages. American Documentation (pre-1986) 7(2):135
Manning CD, Schütze H (1999) Foundations of statistical language processing
Maybury M (1999) Advances in automatic text summarization. MIT press
McCann B, Keskar NS, Xiong C, Socher R (2018) The natural language decathlon: Multitask learning as question answering. arXiv preprint arXiv:180608730
McClelland JL, Rumelhart DE (1989) Explorations in parallel distributed processing: A handbook of models, programs, and exercises. MIT press
Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
Mimno D, Thompson L (2017) The strange geometry of skip-gram with negative sampling. In: Empirical Methods in Natural Language Processing
Plath W (1967) Multiple path analysis and automatic translation. Amsterdam, The Netherlands: North-Holland
Radford A, Wu J, Amodei D, Amodei D, Clark J, Brundage M, Sutskever I (2019) Better language models and their implications. OpenAI Blog https://openai com/blog/better-language-models
reddit (2021a) Reddit. URL https://www.reddit.com/, accessed on : 2020-07-15
reddit (2021b) reddit. URL https://www.reddit.com/dev/api/, accessed on : 2020-07-15
Ruder S, Bingel J, Augenstein I, Søgaard A (2019) Latent multi-task architecture learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 33, pp 4822–4829
Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. Tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science
Schank RC (1980) Language and memory. Cognitive science 4(3):243–284
Søgaard A, Ruder S, Vulić I (2018) On the limitations of unsupervised bilingual dictionary induction. arXiv preprint arXiv:180503620
Sparck Jones K (1992) Thesaurus. Encyclopedia of artificial intelligence 2:1605–13
Sutskever I, Vinyals O, Le QV (2014a) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, pp 3104–3112

Sutskever I, Vinyals O, Le QV (2014b) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, pp 3104–3112
Tensorbaord (2020) Google tensorboard. URL https://www.tensorflow.org/tensorboard, accessed on : 2020-10-15
Turing AM (2009) Computing machinery and intelligence. In: Parsing the turing test, Springer, pp 23–65
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
Vinyals O, Kaiser Ł, Koo T, Petrov S, Sutskever I, Hinton G (2015) Grammar as a foreign language. In: Advances in neural information processing systems, pp 2773–2781
Vinyals O, Blundell C, Lillicrap T, Wierstra D, et al. (2016) Matching networks for one shot learning. In: Advances in neural information processing systems, pp 3630–3638
Wahlster W (2000) Mobile speech-to-speech translation of spontaneous dialogs: An overview of the final verbmobil system. In: Verbmobil: Foundations of speech-to-speech translation, Springer, pp 3–21
Wendlandt L, Kummerfeld JK, Mihalcea R (2018) Factors influencing the surprising instability of word embeddings. arXiv preprint arXiv:180409692
Winograd T (1972) Understanding natural language. Cognitive psychology 3(1):1–191
Woods WA (1978) Semantics and quantification in natural language question answering. In: Advances in computers, vol 17, Elsevier, pp 1–87
Writing TA (2021) Textio augmented writing. URL https://textio.com/, accessed on : 2021-11-15
Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015a) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015b) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E (2016) Hierarchical attention networks for document classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1480–1489
Young SJ, Chase LL (1998) Speech recognition evaluation: a review of the us csr and lvcsr programmes. Computer Speech & Language 12(4):263–279

8 Appendix

The figure 3 shows a model training summary in graphical shape. The graph is produced with help of “Tensorlflow” tool called “Tensorboard”. Tensorboard is a tool for viewing the hidden layers and mechanism of the ANN models - written by Google to increase the efficiency of “Tensorflow” library Tensorbaord (2020).

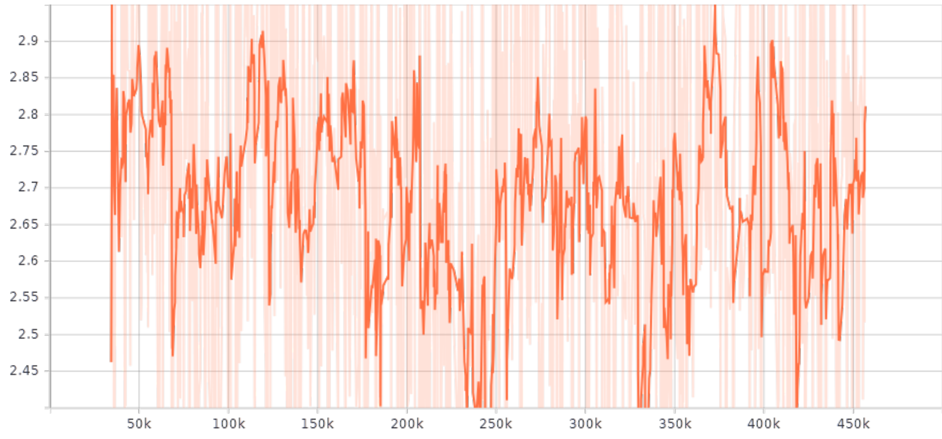

Fig. 3 Tensor-board Training loss summary

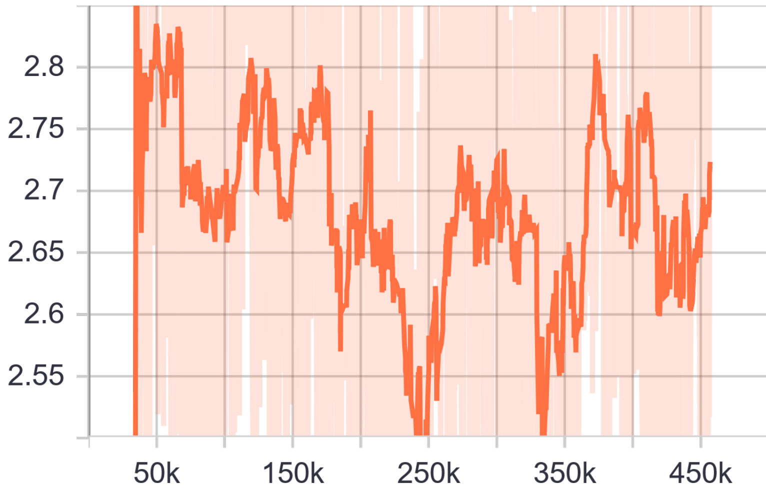

Fig. 4 Tensor-board Test loss summary

8.1 Random Samples

In the section, we have given the 'Microsoft 'OneDrive' shared folder link which contains 2,284 samples that are generated by the model during the training process. The random sample has been generated roughly after every 200 training steps. Samples can be accessed via following link:

Click here to see the all random samples!!

Chapter 3

Conclusion

This thesis is based upon a cumulative style of writing for the fulfillment of the degree requirement of Ph.D. The thesis is comprised of two scientific published papers and one collective study project with master's students done at the cognitive science department, University Osnabrück. The primary focus of the thesis is the Application of Artificial intelligence in business and finance. For this purpose, we have chosen, three different major sectors of the business, namely, financial forecasting, banking sector, and generic business. in the introduction section, we have briefly discussed the origin of the business, and history of technology, the current trends in business, and how AI is leveraging the business operations.

Data pre-processing, data mining, natural Lagrange processing, Sentiment analysis, new analytics, LSTM neural network, and Transformer-based neural network architecture have been used as a method in these research papers. In the coming paragraph, we will conclude each research study and its practical implication in the industry.

In the first research paper, we tried to highlight the key areas of financial forecasting. Financial forecasting has been used for several decades to maximize the return of stock and other tradable financial assets. Second important thing is that within the surge of access to the information and hype of social media, the stock returns are no more only based on technical and fundamental indicators but, get highly influenced by the information regarding that stock. Many traditional methods have been developed and applied for stock market return forecasting, most common are time-series and moving average based stock market return forecasting methods I.e., ARAMA ARIMA, ARCH, GARCH, EGARCH, etc. The problem with these methods is that all models required a basic assumption of data linearity. But data linearity is mostly practically not possible for time series type of data. So, we have tried to address this problem and applied artificial intelligence methods to forecast market returns. In addition to the AI models, to perform information analysis, we have added a news analytics process in this research paper we have parsed, collected, aggregated text, and news information from several diverse sources like blogs, main steam news, print media, expert commentary, company's internal announcements event, etc. This textual information eventually has been used to convert into positive and negative sentiments with help of basic machine learning methods and then we have devised our own sentiment index formulation instead of using the traditional count-based sentiment index method. We decided on hourly-based returns data as well as an hourly-based window of information sentiments. Using LSTM, we have performed six different models including and excluding news analytics to forecast the top 10 USA-based stocks chosen from different sectors each to represent the different sectors of business and to include diversity for the model robustness purpose. The model shows that information i.e., new sentiment analysis plays a vital role in the volatility of stock returns, and stocks need 3 to six hours to show the impact of the news sentiment on returns of the stock.

Our research has many possible practical implications, for example, many famous intraday trading and long-term trading platforms are providing trading, management, and forecasting opportunities to their customers with the help of AI technologies. Similarly, information and news have a greater impact on fluctuations of the returns, so, this research could directly benefit similar kinds of applications area, for example, stock market, option, future, or crypto trading as the fundamental principle or forecasting and trading are the same. So, with the help of the current AI forecasting method and especially with the combination of the natural language processing financial forecasting surely has gone to the next generation and our research gives scientific proof of concept for that.

The second research study is related to the general business and marketing operations that are involved mostly in customer service and market research areas. For example, chatbot support to the customer provides nearly human-level support and answers to the questions. Similarly, for analysis and comparison purposes, businesses need to collect primary information, statistics, and figures related to any company without the human being involved. Another use case of our research study is to help the business report writing process in the way that the system gives suggestions about writing style which is aligned to the industry or commonly used writing norm within the industry. All these implications are very practical and most businesses encounter them during the normal course of business.

Our study encompasses the same premise and gives the solution to above mention problem using current development in the field of AI and transforms neural network architecture that is an advanced form of normal recurrent neural networks.

So, we have parsed, filtered, and analyzed 187 gigabytes of the text data to get 2.3 billion text tokens for the model training purpose. Our transformer model has 355 million parameters and took almost 3month on grid computing power to train. The idea was to form a very big text corpus filter only business-related text and train a model to answer the business problems and get the business-related writing style or formalities. With the vision in mind that later new updated information patch, as well as training, could be supplied to the model weights and it would be able to get current knowledge related to the business. Resultantly, our study model can write small reports regarding different business-related themes and is able to answer the point chatbot style questions. The other implication of this study is auto text completion in the business report writing process. At the end of the training, the model was tested on different keywords related to a diverse range of business-related keywords and model was able to write relevant, grammatically, and contextually correct information.

The last case study is from the banking sector because banking is also a big industry. For investors and moneylenders, it is important to know in a transparent way how well the firm will be financially doing in near future? For this purpose, we have collected financial ratio data as well as form 10 K data of the US banking sector. We have used natural Laurange processing and regex (python library to parse text on rule bases) to extract the relevant section from the form 10K then that is converted into words embedding and used in recurrent and other convolutional network models to predict the financial distress for the banks in near future. We concluded that conforming with existing research', ANN models could predict financial distress of bankruptcy with more accuracy. Furthermore, we have tried to estimate many different ANNs-based models and learned that LSTM is the best predictor of bankruptcy prediction. Surprisingly, even a convolution neural network could not outperform the simple LSTM. The likely reason is that the convolutional network is more efficient for special and temporal data but in our case, we have temporal data only and LSTM can deal with such data very well.

The practical implication of this study is that investors and other stakeholders can easily make their decisions based on the prediction that not only involves fundamental analysis but also textual analytics.Even it is equally important for the policymakers and strategies with the firm to take appropriate actions to improve the financial health of the company. If they could investigate the situation beforehand.

3.0.1 Discussion

Over the passage of time, there are a lot of new technologies, methods, models, applications, and systems have emerged. Knowledge base and different academic and commercial fields of studies have become highly interdependent and interdisciplinary in nature. The success of one technology is highly dependent on the advancement and enrichment of other tools or technology. Keeping the same principle in mind we have tried to focus on the role of artificial intelligence methods in the area of business and especially in corporate finance. Generally speaking, all kinds of businesses are in the phase of the digital transition and come across a lot of adaptation to technology. During the previous three to four decades artificial intelligence has evolved as a strong and useful tool that has its traces in every walk of life. The current thesis is shedding light on the application of artificial intelligence in business.

I have not done the thesis in theoretical format but following the empirical footprint, I have provided research papers and one study project as a proof of concept. The research paper are published during the Ph.D. process and included in the thesis but the study project was supervised by me and written by a group of students, thus not included in the thesis. In the coming paragraphs, I will shortly discuss the reason for selecting the research papers, a short description of what I have achieved, and how that research could be useful for industrial implementations. Subsequently, I will discuss in which direction overall this field is going, what are challenges need to keep in mind, and what are future research directions.

When it comes to business the important sector is the stock market and trading. Stock market predictions have long been linked and dependent on software, to either create an automatic buy/sell signal or automatically complete the transaction process. We have tried to use artificial neural networks and natural language processing developments to get accurate sentiments from the market. Getting the sentiment of investors and stakeholders is quite a tough job because the sentiments are subjectively spread across multiple channels and it's very hard to even consolidate them in one place. For example, just imagine that news, and current information about company X is available everywhere on the internet. Information could be in any form, company internal policy news, blog, Twitter feeds, Instagram, Facebook, mainstream media, print media, social feeds, and discussion forums. The second big challenge is getting positive and negative sentiments out of all consolidated aggregated information. Important to note that in stock markets information is updated on minutes or hourly levels. That means we had to somehow map the updated information with the stock price for every hour to see the impact of sentiments on price change. So, in short, we have developed a sentiment index to consolidate the minutes level information into one-hour windows keeping in mind the importance of the news (not all text is important for market movement) relevance to the company and some other weight mechanisms. We have used recurrent neural network models to predict the next hour's companies' stock movement based on the previous price and sentiment index which we have proposed on our own. This method is based on a very practical approach and is much needed in the industry. Many companies offer intelligence trading already trying to implement artificial neural network approaches to predict and they market their solutions on the buzz of AI.

I haven't seen many companies offer the option of market sentiments in a meaningful way but there are many interesting projects going in this direction. In this connection, our own built sentiment index and methodology could greatly help the industry to build very practical solutions.

A second case study was related to generating business-related text autonomously. The idea is to create a process with help of transformer neural network architecture to make it able to generate business text based on a single keyword. The main process is training the very recently invented state of art ANN model on a very large business text and making it able to learn all business-related context, info, facts, and figure so it could be able to generate paragraphs based on keywords supplied. The research is probably just a foundational framework but there are many great applications that could come through and infact there are interesting applications in progress right now in the commercial arena. For example with the development of the NLP now it is possible for google to just give a summary of very important and relevant information when users search for it. In the backend similar approach of text summarization is working. Similarly, a lot of grammar and spelling suggestion is also based on the same principle. They get trained on the very large corpus or textual dataset and with help of smart ANN architecture, they are able to carry context and produce a very good auto-correct and grammar check. Google has now implemented a text auto-complete option in Gmail writings. The idea of our research paper was to create text generation specifically for business report writing and the general business writing process. For example, the business has a lot of record keeping and market research, and other business reporting writing. AI writing assist cloud help user in rewriting grammatically and factually correct sentences. Especially for the non-native writer of any specific language, it would be a great help if AI writers could help them to write or even generate augment writings that would be factually, grammatically, and syntactically correct. Overall there is a great potential for AI to assist in business processes.

Our next case study was regarding financial bankruptcy prediction with help of NLP and ANNS methods. With help of master's students in the Cognitive science department, we have tried to study the problem of financial distress prediction well before time so that companies could be prepared to manage or avoid financial disasters. It is also equally good for investors and lenders of that company to know how well the company performs where they have invested.

So, the financial performance of the company is normally dependent on many different factors for example fundamental environment in which any certain company is operating. For example economic condition and macroeconomics of that country or even the overall financial environment of the world. These factors are normally not under the control of the firms and are thus called systematic risk factors to the company. However, we have studied idiosyncratic risk factors for example we have collected data of financial ratios regarding the health of the company. The second part of the data was parsed through 10K FORMS of the specific companies to get the sentiment of management regarding the financial position of the company.

So, data collection process was very tough due to the data of healthy and non-healthy firms is not easily available and the matching datasets and collection financial ratio data of those companies or even if the data is not available we needed to calculate the data missing data from the financial statement of the company. We have applied many different recurrent and convolutional neural network models to rigorously check out the approach. LSTM models were the again best performer because they have the availability to capture the long chain of dependencies of the text as well as the previous lags on which the prediction is dependent. Practically there is a lot of potential in the industry for such a solution. Many new firms and startups die due to bad financial situations and not taking timely action or timely preparation to avoid the financial crunch.

Although we have tried to produce some case studies as proof of concept which show how artificial intelligence is helping business operations, there is a lot more that could not be covered in this thesis because it would take ages to cover the complete spheres where AI could help. For example, almost in every commercial activity, ANNS could greatly help in one way or another. The second thing to notice is that ANNS models are data hungry and the accuracy of the target goal highly depends on the amount of data. The field is emerging and effort is being exerted to create open-source financial databases. But right now it's a big limitation for the growth of the industry. The second main issue is the amount of time required to complete the training process. With GPU the task got much faster but still, GPU is too costly secondly standalone GPUs do not help rather need a fleet of GPUs called grid computing or parallel computing. Thus, this limit is getting resolved rapidly with new techniques, efficiency, and new architecture which consume much less computing power. We faced a similar issue during the research phase.

In the end, we would like to highlight some areas which are right now on the higher end in adopting the AI related methods. For example, autonomous driving is getting adapted at a faster pace, and in fact, AV depends on AI mostly. The second prominent area is logistics. For example, managing the warehouses and supply chain with the help of drones and intelligent robots. On similar conventions, the health care and medical industry is greatly benefiting from the AI method, for example, DeepMind at google was able to propose protein folding structures with very high accuracy, which would otherwise need years of research and resources to accomplish the tasks b humans. Last but not least in the field of natural language processing, there is some breadth-taking mesmerizing innovation going on. For example, OpenAI neural language models can do many practical tasks with very high accuracy. Following similar footprints, DALL-E is able to create very relevant images with help of written captions or descriptions. We are hoping the incoming year's businesses need to be completely aligned with new AI innovations to complete and survive in the highly innovative and competitive environment.

3.0.2 Limitations

Artificial intelligence models are really replacing the traditional methods when it comes to financial management and business, but this is not without challenges and a lot of limitations. We have faced some limitations too during conducting this research. Some of the prominent are as follows:

The field of artificial intelligence is very dynamic and rapidly changing but the research publication process is very sluggish and time taking. The proposed method, inference, and conclusion could get outdated very easily which means the researcher is unable to use current and more advanced methods for the research paradigm. Another thing is that Artificial intelligence models are data and resource-hungry. In a practically setting acquiring the big data to train the models to get meaningful results is one of the big challenges. Even if data is available the structure of the data is meant for mostly old-time series analysis and is not ready for machine learning approaches. Similarly, recurrent models are keeping the previous step state of the model up to a long chain which needs a lot of GPUS-based computing resources. Most of the time single computer regardless of how good computing it has is simply not sufficient and you must combine the virtual computing resources for analysis with help of grid computing. So, you must modify, rewrite, or structure the exiting stater code to be able to run on the grid that is normally programmatically overly complex and gives you results in batches, which need to be combined to get a holistic picture. This task itself is time-consuming and tough.

3.0.3 Future work suggestions

There are many future areas to work on in the context of the application of AI in business. We want to highlight some future projects that can be a very good industrial example of the AI application of AI in business. First, the Computer vision approach could solve many texts related problems that are very complex to handle otherwise. For example, document reading and categorization with the help of computer vision. Computer vision is normally considered for normal object recognition tasks but if we treat text as an object and map the relation of each word and its location we can convert the text from scanned images into digital text. Secondly, there is enormous potential for AI in the field of logistics and supply chain management. For example, there are two types of logistics and warehousing, one is warehousing for big corporations that store and manage their inventories on their own, and the other is specialized logistic warehousing provider to small and medium companies which want to outsource them, but they don't want to invest a huge amount in warehousing by their own. Often inventory counting and tally processes happen on interval bases and real-time inventory status is impossible to update due to mismatching and human error. As a solution, we can implement the ComputerVision inventory counting method embedded in flying drones which is hovering in front of inventory stands and count and send the information of inventory in real-time giving enhanced accuracy and more UpToDate inventory status. So, in general, computer vision and natural language processing have enormous potential in the area of generic business. Recurrent neural networks and now their more advanced versions called transformers and encoder decoder-based state-of-the-art AI networks can solve more complex forecasting problems with greater accuracy which was never possible before. The third future work suggestion is related to trading applications, the trading application could try to predict the next pattern not based on any fundamental or technical analysis but purely based on computer vision or it can guide layperson in stock purchasing with the help of NLP based chatbots. So, this is also an incredibly good future application of AI. Of course, there are countless commercial areas where AI can help, we have realized these three ideas where one can work, namely, Scanned document digitalization with the help of computer vision, Inventory counting, and management system with the help of rove and computer vision, and trading applications that guide a common person in financial asset based on the artificial chatbot and market pattern recognition.

Bibliography

[1] Systran. 25

[2] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014. 27

[3] L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–563, 2014. 27

[4] M. L. Forcada and R. P. Ñeco. Recursive hetero-associative memories for translation. In *International Work-Conference on Artificial Neural Networks*, pages 453–462. Springer, 1997. 26

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 33

[6] H. Schwenk. Continuous space language models. *Computer Speech & Language*, 21(3):492–518, 2007. 26, 27

[7] O. Strelkova. Three types of artificial intelligence. 2017. 29

[8] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014. 27

1