



Affective Analytics and Visualization for Ensemble Event-Driven Stock Market Forecasting

Praveen Joshi^(✉) and Haithem Afli^(✉)

ADAPT Centre, Cork Institute of Technology, Cork, Ireland
{praveen.joshi,haithem.afli}@adaptcentre.ie

Abstract. This study provides a comprehensive and objective valuation of factors included in the prediction of the stock market price of an organization. It examines the earlier financial-based model, machine learning-based model, and state of the art achieved deep learning model in contrast to the proposed method. The proposed model also evaluates the methodologies adapted while building the aforementioned models. For evaluation purposes, the dataset has been curated from one of the social media platforms- ‘twitter’, business news website- ‘Financial Times’, and from ‘Quandl’ to obtain financial indicators respectively to an organization. Big Data technologies have been deployed to accumulate and pre-process the data. An essential goal of this study is to support the visualization of such multidimensional data for exploration and analysis. Different tools and methodologies were evaluated to provide concise information at a particular instant by processing cumulative data of the organization under study. Affective analysis of news and twitter has been done with early defined models, and custom formulas derived in the research and the same have been plotted in real-time over the dashboard. Once specified pattern and transformation have been made, the transformed data curated is evaluated overall existing models and the proposed hybrid model. The model holds the ability to drive decisions after incorporating quantitative data such as stock price indicators, as well as qualitative data curated from social media platforms and business news website.

Keywords: Computer systems organization heterogeneous (hybrid) systems · Computer systems organization data flow architectures · Software and its engineering abstraction · Modeling and modularity · Theory of computation streaming models · Theory of computation data modeling · Human-centered computing scientific visualization · Computing methodologies natural language processing · Computing methodologies philosophical/theoretical foundations of artificial intelligence · Computing methodologies machine learning · Computing methodologies modeling and simulation

1 Introduction

In the artificial intelligence (AI) community, stock price prediction has always been one of the biggest challenges [51]. Starting from the initial time, professional

traders developed lots of analytical methods to cater to this problem, in which few methods which caught the eye were fundamental analysis, quantitative analysis, and many more [56]. However, none of the earlier financial methods showed any noticeable results to determine the stock market prices. Prediction of stock prices has been seen beyond the capability of professional traders who, in general, are driven by greed and fear were not able to make rational decisions about buying and selling in the stock market [40] and traditional AI that is supposed focus on imitating human [56]. Recent days have seen the exponential growth of the artificial neural network in terms of underhood capability to approximate any complex continuous. These developments enable an artificial intelligence system to find out the more complicated relationship between the features and the target class and ability to cater lots of data with the advancement in algorithmic architecture to parallelize for a huge amount of data [23].

This research project is aimed to capture the events happening in real life, which can be a merger of a company, changes made in the hierarchy of an organization in respective company and subsidiary companies to the change in own nation's growth where that particular organization is expanded and change in the relationship amongst different countries. The study tries to captures these events in the form of sentiment scores and in-depth emotion score to measure events impact on the rise and fall of the stock price for the concerned organization.

Social media platform provides a space to every individual around the world to put up their concerns and thoughts regarding the different aspects of an organization, which can depict their likeliness towards the new launch of a product, service released by the organization or any event in association to the respective organization. These opinions and flowing thoughts have been captured concerning the organization and processed to analyze the sentiments and emotions of the general people, which is then further passed upon to the predictive models to relate such events to the rise and fall of the organization's stock price.

Business News is another platform that captures the sentiments and emotions of the people who bring in the money into the stock market. Business NEWS helps in capturing critical events as compared to the social media platform as it targets explicitly to bring insight from the event. As experts in the financial area derive the notion of the event, analysis done over Business News holds vital importance. All engineered features processed from the Business NEWS are again passed onto the predictive models to learn the correlation amongst those features to predict the stock price movement.

The financial indicators and its derived features with the existing formulated financial models have been deriving the investment in the stock market until recent years. They have been powerful to capture enough trends and movement of the stock market. With the exponential growth of other platforms now these models can be empowered with more advanced features and hybrid AI models can be introduced to capture feature relation, which is still not known to financial experts.

This study aims to assess the hybrid predictive models and their capability to make advancement in the financial model by incorporating many powerful features directly related to the response of the events from the general expert or non-expert people in finance around the world. Amid finding the best model for prediction, the study also tries to visualize the continuous stream of tweets and periodic feeds of business news to visually infer and get a notion of attachment of stock prices to the affective response of the peoples.

2 Background

This section provides a brief about the stock price and how artificial intelligence can help in inferring the prices of stocks. As Big Data streaming is captured from different sources and analyzed, thereby, this section also aids in creating a notion of Visual Analytics (VA) tools.

2.1 Stock Marketing

A stock market is a place in real or virtual, which provides “trading” facilities on corporations’ stock and derivatives to the investors to trade securities and stocks of a corporation or mutual organization. The stock exchange is a regulatory body that governs the issue and redemption of securities. It also facilitates investments, income, capital events, and dividends [49]. The stock market is also known as the secondary market as it involves trading between two individuals where individuals can be organization, corporation, broker, or investor [51]. Stock prices are highly volatile, but the notion of price remains the same where if the stock is highly in demand, its price will rise, whereas on the contrary, if the market notion is against the notion of the company, there will be a dip in the prices. All companies whose stocks can be purchased over the stock exchange are known as “listed companies.”

2.2 Introduction to Artificial Intelligence

Artificial Intelligence or AI as we call it is still a field of ongoing research and experiment. AI is a field in which human intelligence is replicated into machines which enhances them from simple mechanical devices to enhanced intelligent and self-sufficient machines. AI has a different sense of understanding of different people. Some think that AI should closely replicate the human behaviour and thought process while others think that it should be free from the notion of emotion and should be able to figure out the best way rationally. AI is like an umbrella which takes into consideration various fields and their respective perspectives and techniques be it from philosophy, mathematics or computer science. Many think that the notion of AI itself is a modern idea, but the vision has been there since last 50 or so years. It was Alan Turning who brought AI into trend after the introduction of Turing Machines (1937) [52] which was a model of ideal self-sufficient, intelligent computer based on which he developed the theory

of AUTOMATA. After this, the first developed artificial network the MP neuron was work of Walter Pitts and McCulloch in 1943 [35]. Ever since then researchers all across the globe have been trying to imitate the process of the human brain. A simple machine qualifies as an AI machine if it can perform all the work that a human can with the help of his brain. If a machine can impersonate human behaviour is qualified to be called as an artificially intelligent machine.

1. Machine Learning

Machine Learning is an Artificial Intelligence approach to enable the systems to automatically learn and refine itself from experience without providing any external code for all the features [7]. The heart of Machine Learning lies in the development of programs that retrieve the data and utilise it to learn and improve. The learning procedure involves observing the data for the patterns present and make future decisions based on the patterns and examples provided to the system. The goal is to make the system to learn and adjust on its own without any interference of human.

In Machine Learning there two major categories of division as described underneath:

- (a) **Supervised Machine Learning Algorithms:** It utilises past learning into the new data to make predictions of future events, only when data is pre-labeled.
 - (b) **Unsupervised Machine Learning Algorithms:** It is used in scenarios where the data are not labelled or classified. The data is explored to draw inferences to determine the hidden structure from unlabelled data.
- Semi-Supervised Machine Learning Algorithms:** It utilises both labelled and unlabeled data for learning. It is used in cases where labelled data need resources for further training.
- Reinforcement Machine Learning Algorithms:** This method produces action to interact with the environment. It involves the trial and error search and delayed reward.

2. Deep Learning

Deep learning is one of the three subsets of the significant broad classification of machine learning where algorithms are inspired by the structure and the functioning of the human brain. Deep learning allows the computational models to learn multiple levels of abstraction within the data in the multiple computational layers of the composed model [27]. These methods worked astonishing well in multiple domains by improving the existing state-of-the-art.

3. Transfer Learning

In today's era, even with the abundance flow of data, there are some domains where a lesser amount of work have been done. Less research and low reachability to such topics have led to significantly fewer data in those domains. Even if the problem is similar to one of the existing problem which has been tackled with the help AI, it is sporadic that new problem also follows the same distribution as to the problem already been solved. In such cases, if knowledge transfer is done correctly, it can lead to significant performance improvement

in the model, taking away the much more painful task of addressing and labelling more data. In the recent decade, transfer learning has emerged as one of the new learning frameworks to address the problem of labelled data scarcity [43]. There are two main types of Transfer Learning techniques:

- (a) **Networks as feature extractors:** In this approach, features are extracted from some interim processing layer of a computational deep learning model, and the values coming out of those network at this stage are used as feature vectors. This feature vectors down the processing pipeline are used with different models for specific tasks in another domain different than on what the deep learning model was initially trained upon [21].
- (b) **Fine tuning pre-trained networks:** In this approach, a pre-trained network is used as a starting point; then continuous efforts are made in order to fine-train the pre-existing weights such that they can generalise well over the new task [6].

2.3 Artificial Intelligence in Stock Market Prediction

Recent days have witnessed numerous research that was carried with the help of artificial intelligence to predict stock prices. In AI-enabled stock market prediction, it is noticeable that not only the combination of indicators but new features also came into existence and incorporated into the traditional and latest AI model. In early 2008, in order to simulate the market, a genetic algorithm (GA) in combination with a support vector machine (SVM) was introduced [13]. Where GA was able to simulate the indicator variables, but feature selection was not introduced hence made the model computationally very expensive. The subsequent year 2009, to cut down the computational cost attached to GA, proposed another 3 stage approach that bought down computational need multi-fold [28]. The three-stage approach proposed initial technical analysis over the indicator variable based on historical data, secondly selection of features amongst them, and then applying SVM. This machine learning and statistical model outperformed the earlier model. In the year 2010, with intense research over the market, new indicators were introduced, which extended the horizon for the number of factors that were included in earlier days to many more independent variables that were proven to impact the stock prices. Also, because of new predictors variables and the availability of data, new ways were exploited to build new models. One of the models was based on the artificial neural network with the given exploded indicator variables [24]. This model outperformed the existing machine learning (ML) model and financial mathematical model, giving rise to a new wave of ANN-based models for stock price prediction.

2.4 AI Stock Market Prediction with Financial Indicators

In stock market financial indicators such as open, high, low and close (OHLC) holds great importance. These trading indicators can show the captivity of the organisation in the real world. OHLC is considered to be complete in order to show the

behaviour of an organisations' stock prices. In general, for an extended period, it can provide useful insights about not only the trading strength but also the price gaps. For example, if we plot low and high points for the day, it can provide insights regarding the day level volatility of the stock in stock exchange [17].

Significant data processing capabilities not only extended our horizon of finding the influences of financial factors over a long period which were not known earlier but also helped in advancing our decision making capability. With progress in the capability of harnessing the Big Data and to nurture it for a particular requirement also opened a whole new world for building more data-driven models and the ability to incorporate more number features.

In 2015, the deep learning based event-driven stock price model showed significant improvement as compared to all previously ML-based model [16]. DL based model was able to show a remarkable 6% improvement when compared to earlier model SP 500 stock historical data. With extensive data and new implementation of models also made stock predictions of stocks to reach near to the actual future price of the stocks. One of the recent research paper published in 2017, was also able to provide essential baselines when the different architecture of the ANNs was deployed to predict stock market price [11]. It was able to show that ANN as compared to the existing model was able to identify more hidden context from the data. It was also able to increase covariance estimation when it was subjected to the covariance-based market analysis.

2.5 AI Stock Market Prediction with Textual Data

Widespread adoption of technologies not only bought the world together but also enabled individuals to share their thought, ideas and experiences over the worldwide forum. These thoughts and experiences started building the sentiments amongst the people of same interest leading to favouritism and boycott of a product or organisation with a higher impact and over a global level.

In very recent years, 2009, new research was carried which showed the dependency of the financial news over the stock price prediction [47]. This paper introduced a new world of natural language processing (NLP) to qualitative financial data. Now qualitative data also came into consideration as one of the major contributing factors towards the stock price prediction. Financial news was processed using NLP to create a bag of words including only noun phrases and the named entities for the financial domain. The closeness of model prediction on the real-time financial news was very close to the receding stock price of the impacted organisation. Not only data from different forums but with a general inclination towards Twitter as a social platform also narrowed down the horizon to collect and compute the sentiments of the people regarding any organisation. With hashtag functionality and constrained length of tweets made Twitter favourable amongst the researchers to get motivated towards finding sentiments from the twitter data. In 2013, Twitter tweets and time series data of stocks formed the baseline of another research [50]. Topic-based analysis of the Twitter data and its incorporation with past historical data of stock price variation of a particular organisation showed the worth of the twitters' tweet and power that

it was able to pass on to the quantitative financial data. With advancement in NLP, another paper published in 2015, was able to find a relation between the specific topics impacts on a specific organisation [41]. This method explained how data collected from different forms could be processed and cleaned to the data which is of real concern in the prediction of stocks for that particular company. Paper was able to bring the notion of defined topics for specific industries and change in sentiments for that particular topic. It was also able to bring down the cost of pre-processing needed for the data to be of any importance as suggested by the earlier research. Data used to build the model incorporated existing topic modelling approaches with newly proposed methods and historical financial data. Research not only showed the improvement which was gained over the existing approaches by 2.07% for 18 stocks over a year time but was also able to capture the sentiment analysis contribution in stock price prediction in the real world effectively.

1. Sentiment Analysis Tasks

Exponential growth in an individual's power to access the internet and so for social media has led to a flood of thoughts and ideas that are shared per second across the world. Ease of curation of such thoughts and experiences of individuals for an organisation has given rise to sentiment analysis. As Zhang, Lei explained, Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organisations, individuals, issues, events, topics, and their attributes [2,32]. Over last decade numerous research tried to capture the opinions which are individuals as customers, people of the state and far most as the human to find the influence over the organisations, countries and towards global topics. It was not only able to find the rationale of individuals belonging to a particular geographical area but all over the global level and tried to measure its impact on different higher level organisations. Sentiment analysis is broadly categorised and studied under three categories which are document level, sentence level and aspect level [55].

- (a) **Document-level** sentiment analysis task tries to classify the sentiments as neutral, positive or negative based on the overall sentiment captured in a document. The document-level analysis assumes the notion that each document will be talking about one context. One tweet can also be considered a document which can represent the opinion of an individual regarding some product or organisation.
- (b) **Sentence level** sentiment analysis task provides the capability to capture the sentiment of the document on the sentence level. In sentence level, before going for sentiment analysis, the sentence is mostly checked for the subjectivity. Subjectivity classification helps in avoiding objective opinion which is none other than facts [31]. Only sentences with high subjectivity are taken into consideration to undergo sentiment analysis task. Sentence level sentiments are also captured in three classes neutral, positive and negative respectively.

- (c) **Aspect level** sentiment analysis task focuses on summarising and bringing out overall sentiments from people's opinion for any particular entity also known as targets. Aspect level sentiments express sentiments for each aspect of an entity. For example, if an organisation is considered as an entity, then the salary package can be considered as one aspect and employee perks as different aspect and so on. When aspect level sentiments are captured, one can easily find out whether salary or employee perks are useful in the organisation.
2. **Emotions Analysis Task** are closely related to the sentiment with more analysis of the inferred polarity. For example, negative sentiment can be caused by sadness or anger, while a positive sentiment can be caused by happiness or anticipation. Thus, following the way in sentiment analysis, many deep learning models are applied to detect emotions [55]. Zhou proposed an Emotional Chatting Machine (ECM) that can generate appropriate responses grammatically relevant and emotionally consistent based on GRU [55]. Their system is modelling the emotion factor, using emotion category embedding, internal emotion memory, and external memory. A bilingual attention network model was proposed by Wang [54] for code-switched emotion prediction. Abdul-Mageed and Ungar, built a large, automatically curated dataset for emotion detection using distant supervision and then used GRNNs to model fine-grained emotion [1]. They extended the classification to model by Plutchik [44], in which he proposed 8 primary emotion dimensions as shown in Fig. 1.

2.6 AI Stock Market Prediction with Twitter Data Analysis

The rising number of blogs and social media platform in the last decade provided a mean for people to put forward their opinion regarding the entities which can be an organisation or individuals'. This massive amount of opinionated data mining has provided a mean by which we can quickly capture the sentiments of targeted individuals regarding any product, organisation or government institution. In social media, Twitter gained favouritism from the world-wide community in terms of its use and became one of the dominant platforms to convey opinions. With the limit in tweet length and worldwide acceptance also grabbed attention from the AI community. Many research papers were published to establish the tweets sentiments relations to the stock market prediction in the last decade. In 2010, A.Pak proposed twitter corpus for sentiment analysis. In that particular corpus tweets were tagged to specific emotions manually, such that happy emoticons signified positive sentiment and sad emoticons signified negative sentiments [42]. In 2013, Twitter tweets with the time series data were able to provide significant results in stock market prediction [50].

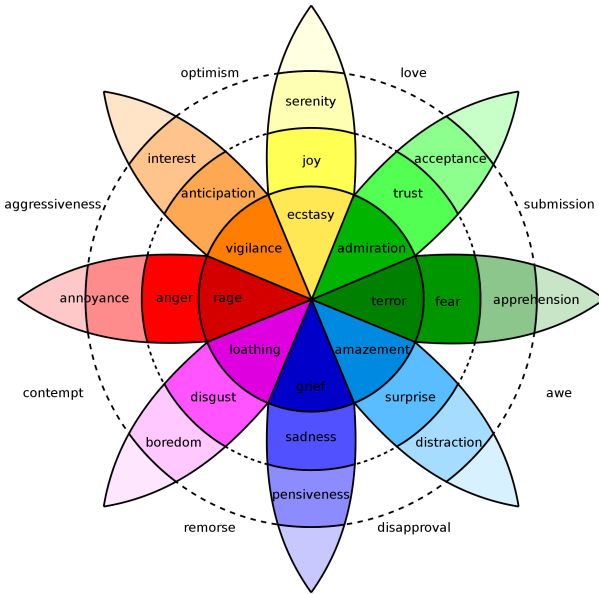


Fig. 1. Plutchik’s wheel of emotion

2.7 AI Stock Market Prediction with News Data Analysis

While in last decade there were plenty of articles and research paper were published in data mining and time series to predict stock market prices, but there is very handful of papers which covered text mining in stock market prediction. Some of the earliest research paper who started using business news for financial forecasting [10, 26], did a remarkable job but still due to the absence of news and public opinion about the particular organisation and nature of high volatility of stock price there was still scope to enhance the model. Most of the sentiment classification involves the training of the system based on the labelled documents by the experts or generated by the system. In 2004, Mittermayere proposed NewsCATS engine which was able to classify the news into three classes namely Good, Bad and No-movers. In this based on the category of the News movement of the stocks was predicted [38]. The AZFinText system proposed in 2010 is a regression system which also tried to predict the stock market prediction based on the news [48].

2.8 Big Data Visualisation of Streaming Data

Big Data poses a computing challenge because of its rapid velocity, immense volume, and a wide variety of [25]. With ever-increasing, human-centered systems are creating enormous amounts of data. As this enormous high volume of a wide variety of information can easily be generated and collected at a very high

speed; Have created a necessity of Big Data visualization and visual analytics in a diverse real-world application.

In the last few years, there has been the development of various tools and techniques to visualize patterns in the textual data. Which most popular ones try to find the co-occurrences of the entities [9,18]. Also, there has been a multi-fold increment in the software to visualize such data [4]. Tableau is one of the most commonly used software for the analysis; It's the ability to hook with different ingestion system make its favorable choice amongst the developers and higher managers to crunch data quickly into multiple axis [14]. Alongside different systems were built to not only visualize the Big Data but to provide end to end solution to the Big Data visualization problem. ELK stack is one of the commonly used stacks which is used in the industry [20].

2.9 Contribution

The main contribution of this work is to analyze and develop the architecture to provide visualization aid to the prediction of the stock market prices. In prior research, most of the time, groups were mainly focused upon either the qualitative or the quantitative data, which have concentrated upon the more modest algorithms to solve that task without any practical validation. As the state of the art algorithms comprises the machine learning models which made this research novel in the financial domain but at the same time deployment of deep learning (DL), model made it to lose the explainability aspect of it. The visualization aid was aggregated to the framework to make it more explainable to mitigate the drawback of using deep learning algorithms.

Our approach extends the existing framework described in Sect. 2.8. In addition to developing together with the components that worked best in abstract design, we also created a pipeline that can ingest from multiple platforms in parallel without any concerns of qualitative or quantitative data. Framework in integrated with state of the art machine learning models of the finance domain and at the same time, the ingested data was represented in real-time over the tableau dashboard to comprehend the model's prediction.

Our findings are that model prediction does not need to come at the expense of the explainability. For the approaches to building the framework, we have developed components extension to integrate seamlessly, which provided the necessary aid of visualization to the financial model prediction. Although the data ingestion requires unique formulation and filtering, the process itself is straight forward and easily accessible.

3 Big Data Pre-processing and Visualization of Tweets, Business NEWS and Financial Indicators

This section defines the end-to-end pipeline that has been utilized in this study for analysis of qualitative as well as quantitative data flowing in from the different platforms. Although data is scrapped based on the scrapping policies and API

utilized to connect to the different platform, but once we have a data stream, the ingestion, processing tunnel and storage space remains the same for the project.

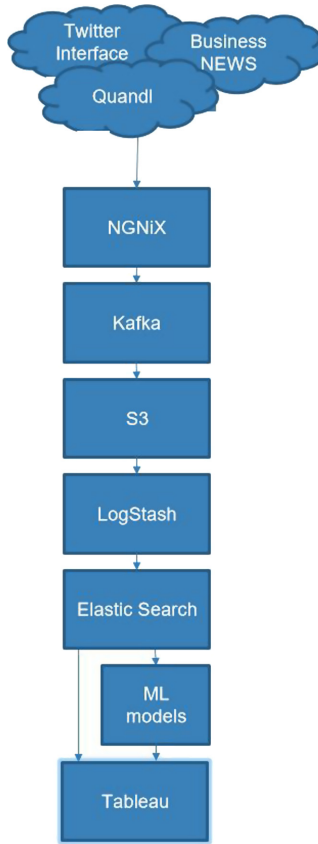


Fig. 2. Proposed architecture

Figure 2 describes the proposed architecture that has been used in the project for creating a visualization pipeline and also used for the data modeling part. In an initial study, all the platforms for the study were evaluated based on scrapping policies and availability of API, and three different platforms were found best for the study. Twitter interface API has been used to collect the twitter data from the social media perspective, Financial times feed was consumed to gather the business NEWS, and Quandl API was consumed to collect the financial indicators of the Microsoft stocks. Based on API and feed, Kafka streaming tunnel with NGNiX was created for continuous monitoring and streaming of data from the different platforms. S3 has been consumed to create a data lake for the data getting ingested from the Kafka pipeline. Over S3, the logstash component was built whose primary aim was to provide server-side processing pipeline, and

output was served to elasticsearch. In elastic search, data can be found in a much more meaningful way as it forms a definite structure. This processed and ready to use data is then consumed by the machine learning models to predict the stock price as well as same data was consumed by tableau for intermediary visualization and analysis to get a notion of how the model should behave and then the model was tweaked if any discrepancies in the visual and inference from machine learning model are found.

Data set curated from different platforms for the Microsoft organization is from the 4th of April 2015 till the 28th of March 2019. Whereas due to less number of data points, no development set have been taken out from the data set, and training and the testing split is based on the dates. Training data points are taken from the 4th of April 2015 till the 1st of January 2019, and testing data is the point is from the 2nd of January 2019 till the 28th of March 2019. Data set Overview is provided in the Table 1

Table 1. Dataset overview - train and test split

Company	Start date	End date	No. of days	Train	Test	f	'f (VIF)'
Microsoft News	4/4/2015	3/28/2019	1027	961	66	25	23
Microsoft Twitter	4/4/2015	3/28/2019	1419	1332	87	26	16
Microsoft Finance	4/4/2015	3/28/2019	983	923	60	32	9
Microsoft News BERT	4/4/2015	3/28/2019	1027	961	66	768	–
Microsoft Twitter BERT	4/4/2015	3/28/2019	1419	1332	87	768	–
Combine	4/4/2015	3/28/2019	1423	1336	87	79	16

In the Table 1, 'f'- refers to the number of features in the data set and 'f(VIF)''-represents the number of features after removing the correlated features from the data frame. Also, the table 'f(VIF)' is null for Microsoft finance and NEWS BERT because features arrangement holds a semantic representation of each textual document, and hence the relation can break if the correlation feature removal is implemented in the data set.

3.1 Data Pre-processing

The need for data pre-processing is only required by the raw textual data curated from the Twitter platform and financial times website.

Pre-processing of Scrapped Textual Data

The textual dialogues are processed using *ekphrasis*¹ tool [5] in which series of operation are performed. A brief visual description of this tool is described in Fig. 3 as well as components are explained underneath:

¹ <https://github.com/cbaziotis/ekphrasis>.

PREPROCESSING PIPELINE FOR TEXTUAL DATA

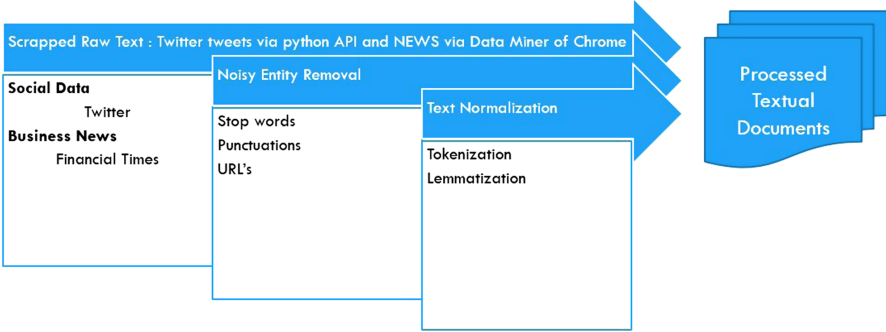


Fig. 3. Pre-proessing pipeline for textual data

1. Noisy Entity Removal: Twitter being a social networking platform on the global level, which makes tokenisation of twitter data most complicated task. It is essential to keep the words intact with the corresponding emotions attached to it. Also, creative writing use for new emotion generations and hashtag should be considered. Textual data curated from the Business NEWS platform are much more formal hence requires fewer efforts for cleaning. The goal here is to remove any stop words, punctuation’s, URLs with censored words, and not to remove complex emoticons.
2. Text Normalization: This step involves tokenising the processed data coming after the above-stated stage. Tokenised words are then lemmatised so that each word can be visualised as root words.

3.2 Feature Extraction Techniques

Feature Extraction from Textual Data

1. **Sentiment Analysis.** For capturing sentiment analysis different libraries and ontology have been used:

(a) **TextBlob**

TextBlob is a library supporting Python 2 and 3 for the original processing data. It provides a simple application interface which helps in efficiently leveraging everyday natural language tasks such as part-of-speech (POS) tagging, extraction of entities based on the POS tags, sentiment analysis and more [33]. TextBlob under the hood utilises NLTK and pattern library which are widely used and accepted in natural language processing NLP community. In recent years, TextBlob gained wide acceptance in AI community which can be readily determined by the number of a research paper using it as a tool for sentiment analysis [3, 34, 53].

(b) **PySentiment**

PySentiment is the library for sentiment analysis which is built on top of the dictionaries. Two dictionaries which are used by this library is namely Harvard iv-4 by Harvard University and Loughran and McDonald Financial Sentiment Dictionary.

i. Harvard Institute provides **HIV4** dictionary. This dictionary provides 185 features for each of the 11789 words. One hundred eighty-five features in this dictionary represent the different aspects of the word ranging from sentiment, affiliation, psychological, emotions and many more.

ii. In 2012, **Loughran and McDonald Financial Sentiment Dictionary (LM)** consisted of 84330 financial words with their sentiments was published [36]. This dictionary after its release in public domain assisted much research to captures sentiments from financial articles, business news and much more

2. **Emotion Analysis:** “Words are associated with emotions,” as quoted in the research paper NRC emotion Lexicon [39]. In order to capture the emotions from the tweets and business news, there are different deep learning models available [8], but to make the architecture lightweight, tokenized processed documents are mapped against the NRC emotion lexicon. Eight emotions are captured in the process. As tweets posted over the twitter platform for a day is more than ten thousand, hence more advanced emotion normalization score system is used to compute the emotions.

Score Formation for Emotions:

$$\frac{\sum_{i=1}^l \frac{\sum_{i=1}^d \text{emotion}_i \text{ appearing in a tweet}}{\text{length of the tweet}}}{\text{number of tweets per day}} \quad (1)$$

In Eq. 1, ‘i’ refers to the single tweet of the day ‘d’ refers to total number of tweets for a day. ‘i’ refers to the total number of days.

3. **Bidirectional Encoder Representations from Transformers (BERT): State-Of-The-Art Textual Representation of Textual Documents.** BERT provides the pre-trained vectors representation of the words, which can be used further with the various AI models. BERT architecture is a frame to provide representations by joint conditional probabilities both from the left and right context for all the processing layers [15]. BERT vectors are used in the experiment to utilise the shallow transfer learning models to enhance the capabilities of the current predictive models. BERT is used as a service, to convert processed text both for twitter and business NEWS to its corresponding vector. As there are multiple models in BERT, current experiment utilises **BERT-Base-Uncased** which holds the capability to represent the word in the 768 dimensions.

Feature Extraction from Financial Indicators. As platform ‘Quandl’ provides the financial indicators such as OPEN, CLOSE, Adj CLOSE, VOLUME

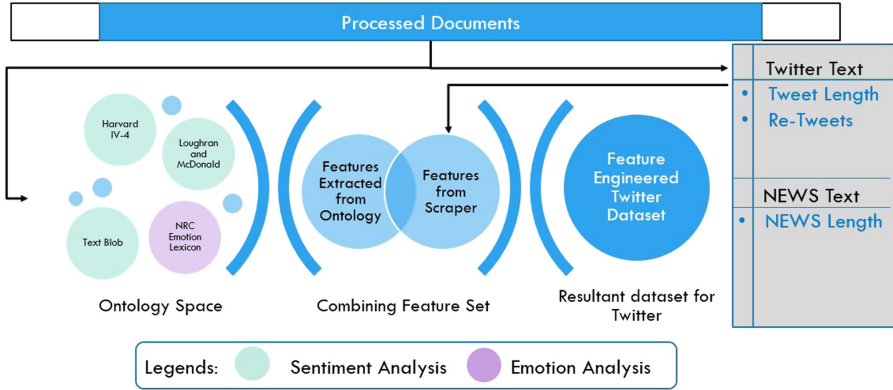


Fig. 4. Textual: feature engineered data set preparation

and DATE for a range of specific duration. Existing research in the area of stock market prediction can help the system to derive out a significant number of derivative features from the information provided by the Quandl platform. As for generating a label for the dataset, as explained in Sect. 3.3, the system is using the open price of the stock of the current and successive day. Hence all derived features are built upon the OPEN financial indicator. From OPEN financial indicator corresponding Fourier transformation have been derived based upon the wavelet research [30], the moving average is computed as a feature with a lag of 2, 7 and 21 days [22], Moving Average Convergence Divergence MACD [12,46], Upper and lower bounds [29], exponential moving average [37] with lag of 12 and 21 days, momentum and log momentum [19].

3.3 Formulation of the Feature Engineered Data Set and Label

Feature extraction techniques Sect. 3.2 are used to build overall feature engineered dataset for the current experiment. Qualitative textual data is converted into quantitative data with the help of the feature extraction techniques, and the custom score mechanism explained in Sect. 3.2. In parallel derivative financial indicators have been developed based on the prior research in the field of finance.

Formulation of Target Labels. As the system tries to predict the rise or fall for particular days under test for the current organisation. Hence, opening stock price is taken as a measure to compute the label for the particular day. The formula for computing the target label is provided underneath:

$$TargetLabel_t := if (OpenIndex_t \leq OpenIndex_{t+1}) then 1 else 0 \quad (2)$$

In this equation, t depicts the day under evaluation or for which label is to be assigned and $t + 1$ represents the next day. According to the equation, if the market is going up, one is assigned as label whereas on the contrary 0 will be assigned as a label for the fall of the stock price.

Formulation of the Textual Feature Engineered Data Set. Feature extraction techniques provided a way to extract the features and to give quantitative meaning to them. All the features from the analysers are combined to formulate the overall data-set. In addition to features coming from the analyser twitter also have one more feature provided by the Twitter API, which is also taken into consideration. Architecture for textual feature engineered data set is visualised in Fig. 4.

Formulation of the Financial Indicators Based Feature Engineered Data Set and Label. As described in the previous section, all the derivatives of the OPEN indicator formulated the new engineered feature. Visualisation of data set formulation of the financial indicator is provided in the Fig. 5.

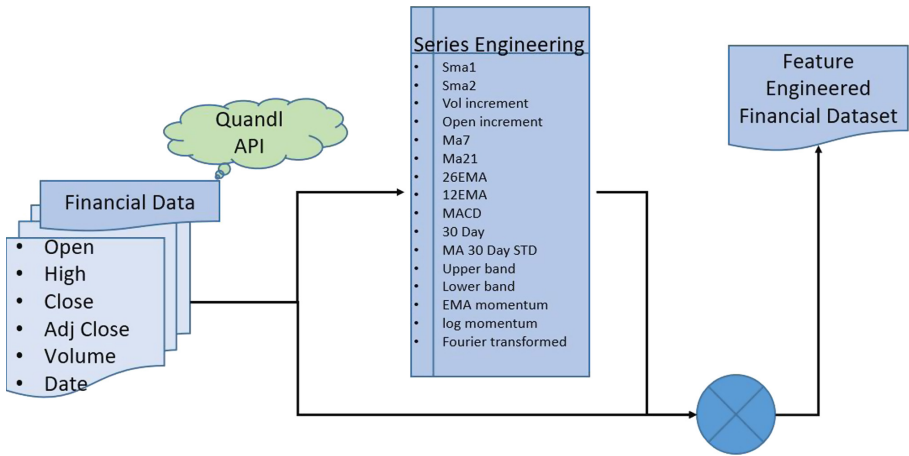


Fig. 5. Financial indicator: feature engineered data set preparation

3.4 Twitter Data Accumulation and Visualization

Once we formulate the definite dataset as described in Table 1. Emotion analysis over the tweets is done with the help NRC [2] lexicons, and individual emotion score is further amplified with the custom score Eq. 1. Two positive emotions, namely- 'trust' and 'joy' are evaluated against the OPEN index of the stock market and visualised in Fig. 6. Two negative emotions namely- 'anger and 'sadness' are evaluated against the OPEN index of the stock market and visualised in Fig. 7.

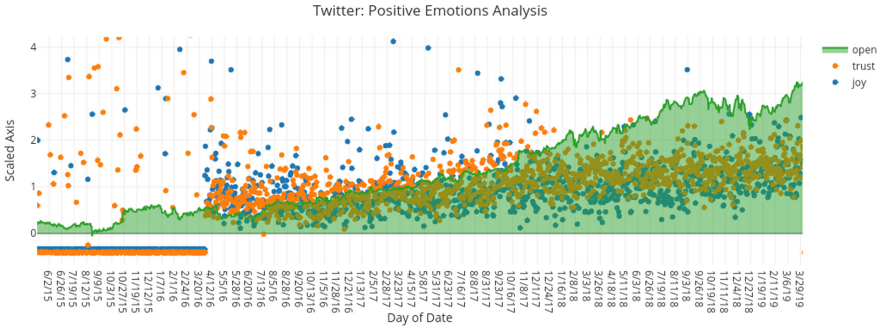


Fig. 6. Positive emotions in Tweets

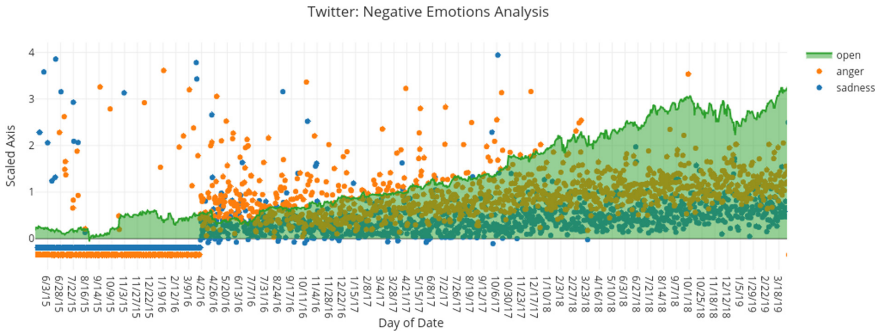


Fig. 7. Negative emotions in Tweets

Sentiment Analysis of tweets is done with the help of TextBlob 1a library and pysentiment 1b library. In the pysentiment library, two libraries are used, Harvard Institute dictionary and Loughran and McDonald Financial Sentiment Dictionary to capture the sentiments flowing in the tweets concerning the organisation. All the sentiments are averaged out and visualised in Fig. 8.

3.5 Business NEWS Data Accumulation and Visualization

Emotion analysis over the Business NEWS is done with the help NRC 2 lexicons, and individual emotion score is further amplified with the custom score Eq. 1. Two positive emotions, namely- ‘trust’ and ‘joy’ are evaluated against the OPEN index of the stock market and visualised in Fig. 9. Two negative emotions namely- ‘anger and ‘sadness’ are evaluated against the OPEN index of the stock market and visualised in Fig. 10.

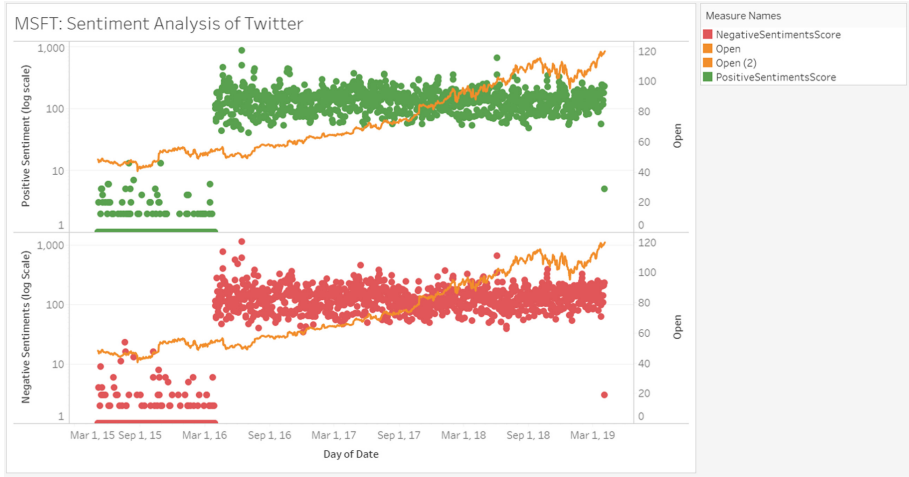


Fig. 8. Sentiments vs opening price for Microsoft stocks based on Tweets

Sentiment Analysis of Business NEWS is done with the help of TextBlob 1a library and pysentiment 1b library. In the pysentiment library, two libraries are used, Harvard Institute dictionary and Loughran and McDonald Financial Sentiment Dictionary to capture the sentiments flowing in the tweets concerning the organisation. All the sentiments are averaged out and visualised in Fig. 11.

4 Architecture for Stock Market Prediction

As experiment evaluates data gathered from the social platform, business NEWS, and financial indicators with the state-of-the-art models. Hence two strategy has carried forward to build the hybrid architecture to improve the performance of the earlier existing systems.

4.1 Hybrid Architecture Based on Best Model Selection Strategy

The first strategy followed is to build an Architecture that can incorporate quantitative as well as qualitative data. Hence prediction from the best performance models for each of the platforms is taken together and given to the voting classifier. The voting classifier then uses the soft voting technique to assign a weight to different models based on the classifier, which are housed in the voting classifier, as visualized in the Fig. 12.

4.2 Hybrid Architecture Based on Shallow Transfer Learning Model

The second strategy is the formulation of the effectiveness of the state-of-the-art shallow networking based transfer learning technique in the form of BERT

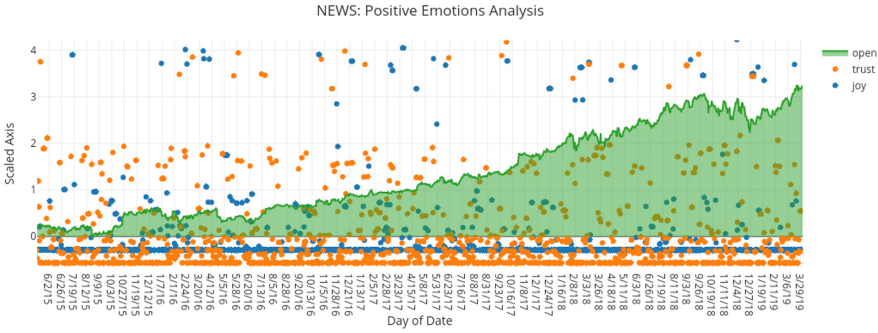


Fig. 9. Positive emotions in business NEWS

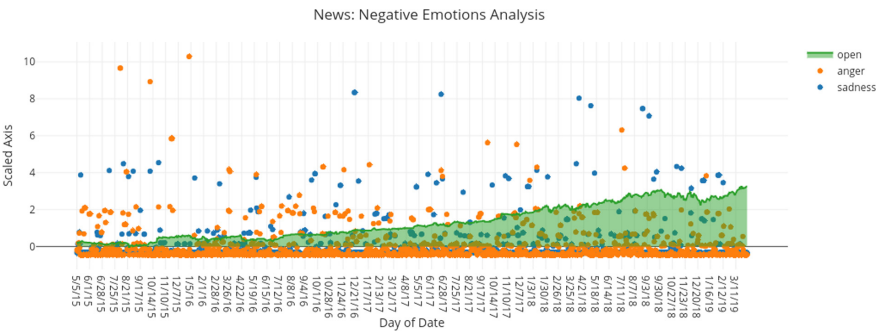


Fig. 10. Negative emotions in business NEWS

vectorization. A tweet and an abstract of NEWS formed an independent document of variable length. Each of the documents then goes through the BERT vectorization service, where it gets converted into the fixed-length vector. The fixed-length vector of tweets and news are independent of each other. Once all the fixed-length vector for the whole day is identified, then the average fixed-length vector is formed for that particular day, which then merges (twitter and news) to create a day data point. A high context level diagram is provided in Fig. 13.

4.3 Hybrid Architecture Based on Engineered Feature Dataset

The third strategy is the formulation of the dataset. The dataset preparation is done by combining all the features extracted from twitter, business news, and Quandl collected financial data. Once data is formulated, it is then subjected to the machine learning models for the training and inference purpose. A high context level diagram is provided in Fig. 14.

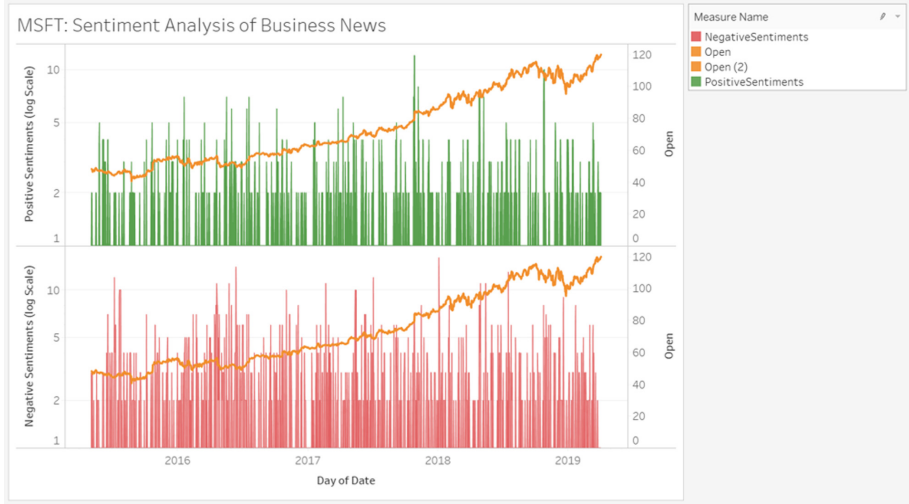


Fig. 11. Sentiments vs opening price for Microsoft stocks based on business NEWS

5 Evaluation Metrics

As current experiment carried out is a supervised problem, hence matrices evaluated for the comparison of the results from the different machine learning model and the deep earning models has been done based on accuracy, precision, recall and F1-Score.

1. Accuracy:

Accuracy is the ratio of total correct predictions that have been made in all the classes in the classification problem. Mathematically it can be visualised as the ratio of true positive and true negative with all the data points present in the data-set. Mathematical formula of accuracy is given underneath:

$$\frac{\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l} \quad (3)$$

In Eq. 3, 'tp' represents the true positive from the model. 'tn' represent the true negative from the model. 'fn' represents the false negative from the model. 'fp' represents the false positive from the model.

2. Precision:

Precision defines the exactness of the system. It is defined as the ratio of true positives identified by the model over actual number of positive marked by the model. Mathematical formula of the precision is given underneath:

$$\frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i}}{l} \quad (4)$$

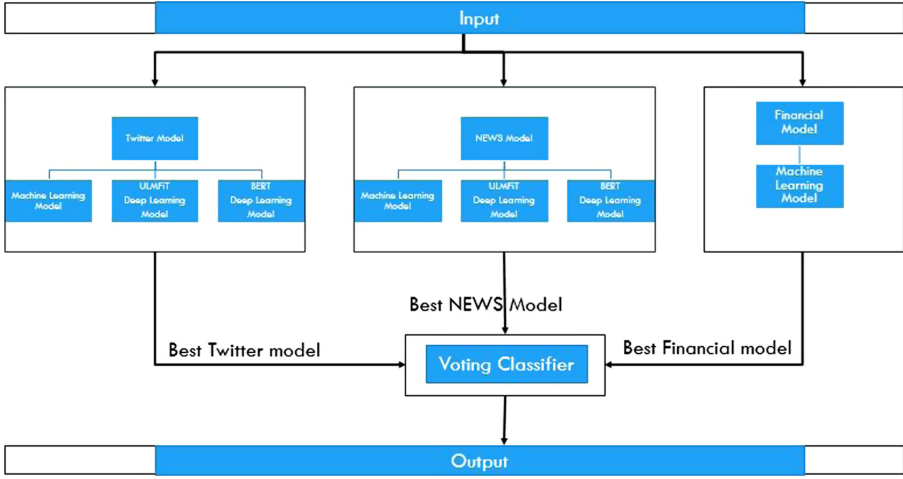


Fig. 12. Hybrid architecture based on best model selection strategy

In Eq. 4, ‘tp’ represents the true positive from the model. ‘tn’ represent the true negative from the model. ‘fp’ represents the false positive from the model.

3. Recall:

Recall helps in evaluating the completeness of the model. It is the ratio of predicted positive over the ground truth positive classes. Mathematical formula of recall is given underneath:

$$\frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i}}{l} \tag{5}$$

In Eq. 5, ‘tp’ represents the true positive from the model. ‘tn’ represent the true negative from the model. ‘fn’ represents the false negative from the model.

4. F1-Score:

F1-Score is computed by evaluating the harmonic mean of the precision and recall. Mathematical formula of F1-score is given underneath:

$$\frac{(\beta^2 + 1) Precision_M Recall_M}{\beta^2(Precision + Recall)} \tag{6}$$

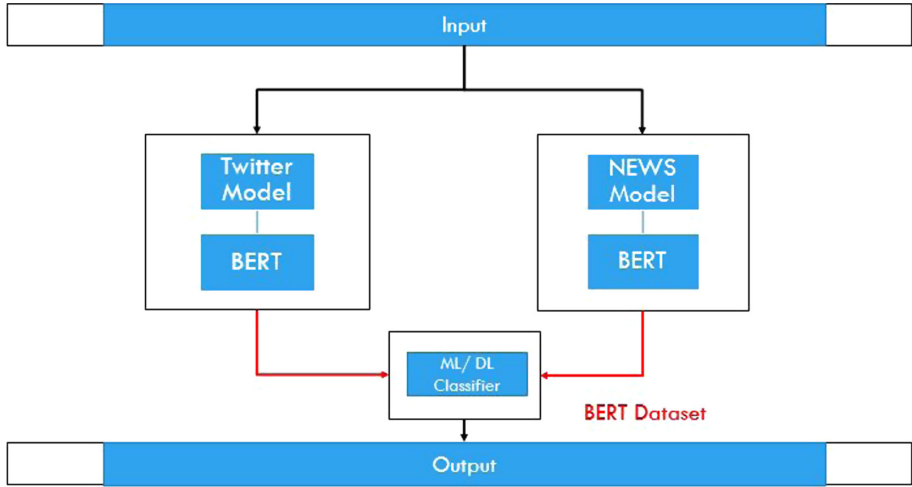


Fig. 13. Hybrid architecture based on shallow transfer learning model

6 Results

As current experiment carried out is a supervised problem, hence metrics evaluated for the comparison of the results from the different machine learning model and the deep learning models has been done based on accuracy, precision, recall, and F1-Score.

6.1 Hybrid Architecture Based on Best Model Selection Strategy

Best model from the different platforms has been selected. From Twitter models the dense, deep neural network has been taken, from Business NEWS models Naive Bayes model has been taken and from financial indicator models Random Forest has been taken; Individual output coming from each of the best models will be given to Voting Classifier to make the prediction. Results obtained from the architecture is described in Table 2.

Table 2. Evaluation metric for hybrid architecture based on best model selection strategy

Model	Accuracy	Precision	Recall	F1-score
VotingClassifier	0.5714	0.63	0.57	0.59

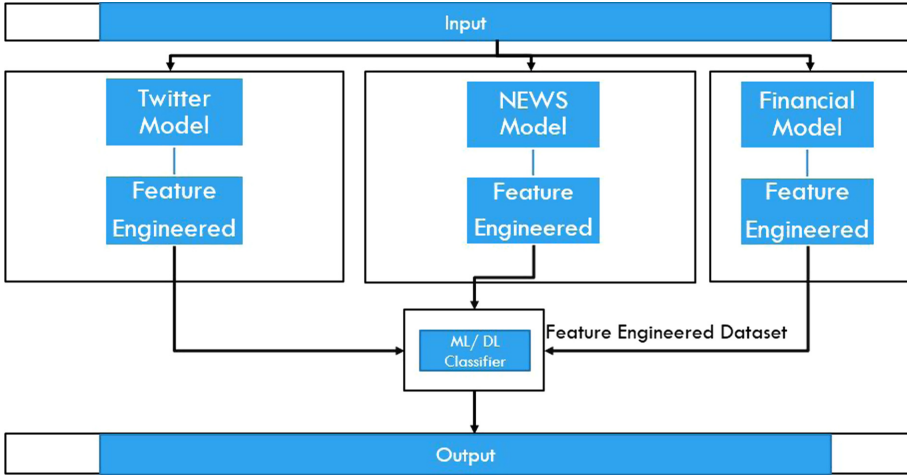


Fig. 14. Hybrid architecture based on engineered feature dataset

6.2 Hybrid Architecture Based on Shallow Transfer Learning Model

Dataset evaluated in this section is a resultant dataset obtained after the merger of the BERT vector of Twitter documents and Business NEWS articles on a daily basis. The evaluation result over machine learning model is provided in Table 3 and evaluation on deep learning model is provided in Table 4.

Table 3. Evaluation metric for hybrid architecture based on shallow transfer learning ML model

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.4827	0.53	0.48	0.49
RandomForest	0.6091	0.98	0.61	0.74
xGBoosting	0.592	0.6	0.6	0.49

6.3 Hybrid Architecture Based on Engineered Feature Dataset

Accumulated feature engineered datasets, from multiple platforms, are taken and evaluated with the machine learning and deep learning models. The evaluation result of Machine Learning models on the framed dataset is provided in Table 5 and deep learning-based models evaluation is provided in Table 6

7 Validation of the AI2VIS4BigData Reference Model

Section confirms and maps the proposed architecture of the study to the AI2VIS4BigData reference model [45], as shown in Fig. 15. This mapping is

Table 4. Evaluation metric for hybrid architecture based on shallow transfer learning DL model

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.4827	0.53	0.48	0.49
DenseNetwork	0.4367	0.43	0.44	0.44
CLSTMNetwork	0.5402	0.6	0.54	0.56
CBiLSTMNetwork	0.5172	0.69	0.52	0.58
CGRUNetwork	0.5517	0.66	0.55	0.59
CRNNNetwork	0.5747	0.82	0.57	0.66

Table 5. Evaluation metric for hybrid architecture based on engineered features dataset ML model

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.4367	0.84	0.44	0.54
RandomForest	0.7241	0.72	0.72	0.72
xGBoosting	0.69	0.69	0.69	0.69

necessary to validate the proposed system but also to provide a useful gateway to extend this research and possible collaboration in the future. As in the AI2VIS4BigData reference model for processing step ‘Data Management & Curation,’ our data ingestion pipeline, as proposed in Sect. 3, can directly be used. In other processing steps, as mentioned in the AI2VIS4BigData reference model, ‘Interaction & Perception’ tableau can facilitate the meaningful visualization needed for explanation of the inference made by the AI model.

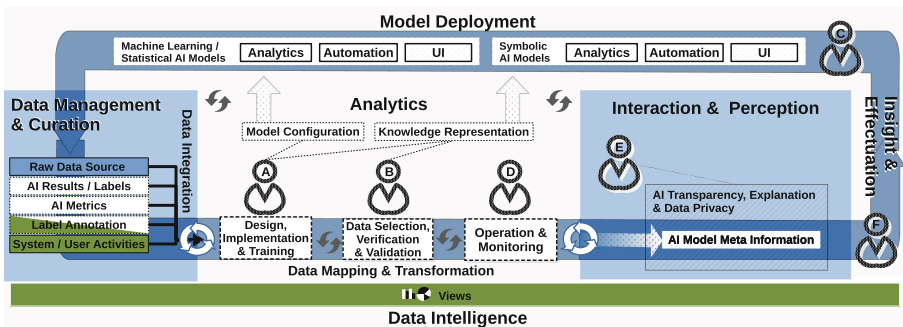


Fig. 15. AI2VIS4BigData reference model

Table 6. Evaluation metric for hybrid architecture based on engineered features dataset DL model

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.4367	0.84	0.44	0.54
DenseNetwork	0.6896	0.7	0.69	0.69
CNN_LSTMNetwork	0.5632	0.57	0.56	0.56
CNN_Bi-LSTMNetwork	0.5057	0.52	0.51	0.5
CNN_GRUNetwork	0.6206	0.62	0.62	0.62
CNN_RNNNetwork	0.5862	0.59	0.59	0.59

8 Conclusion and Future Work

Amongst all the Hybrid architecture, the Random Forest model was able to outperform all the other machine learning and deep learning model by the significant margin. Accuracy of **72.41%** and weighted average precision of **72.00%** shows the balanced inclination of the model towards the respective two classes, which are rise and fall of the stock price for the subsequent day.

The present research has been carried out to provide the feasibility study of the social media platform and Business NEWS over the stock market prediction. The findings in terms of affective analysis visualization and model building showed a significant correlation amongst the social media platform and Business NEWS for stock price prediction. It also captured the results obtained from the state-of-the-art methodologies over the research problem. As a remark, even though there is high volatility in the stock market but with the amount of data flowing in different social media platforms and righteous Business NEWS, in coming future, it will be very much possible to capture the stock price movement with multiple such platforms efficiently.

In future, underneath mentioned directions can be explored to build better visualization platform which can provide explainability to black-box machine learning model:

- More complex emotions can be captured with a correct mathematical formula, which can improve the efficiency of the system.
- More hybrid models strategy can be evolved and evaluated, and deep neural network on most the cases underperformed in the current experimental setup.
- Complex features can be developed in the financial indicator as they showed prominent results as individual models.
- Parallel research on multiple different platforms demands scalability. Scalable modules can be developed to capture the events in real-time.

Acknowledgments. This research was conducted with the financial support of ADVANCE CRT PHD Cohort under Grant Agreement No. 18/CRT/6222 and at the ADAPT SFI Research Centre at Cork Institute Of Technology. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through

the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant # 13/RC/2106.

We want to thank Johnson Controls for contributing the idea of moves, which helped in accomplishing the research with significant results.

References

1. Abdul-Mageed, M., Ungar, L.: EmoNet: fine-grained emotion detection with gated recurrent neural networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 718–728 (2017)
2. Afli, H., Maguire, S., Way, A.: Sentiment translation for low resourced languages: experiments on Irish general election Tweets (2017)
3. Arai, K., Tolle, H.: Method for real time text extraction of digital manga comic. *Int. J. Image Process. (IJIP)* **4**(6), 669–676 (2011)
4. Bassil, S., Keller, R.K.: Software visualization tools: survey and analysis. In: Proceedings of the 9th International Workshop on Program Comprehension, IWPC 2001, pp. 7–17. IEEE (2001)
5. Baziotis, C., Pelekis, N., Doukeridis, C.: DataStories at SemEval-2017 task 6: siamese LSTM with attention for humorous text comparison. In: Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada (2017)
6. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: Advances in Neural Information Processing Systems, pp. 153–160 (2007)
7. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
8. Boucekif, A., Joshi, P., Boucekif, L., Afli, H.: EPITA-ADAPT at SemEval-2019 task 3: detecting emotions in textual conversations using deep learning models combination. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 215–219 (2019)
9. Braun, P., Cuzzocrea, A., Keding, T.D., Leung, C.K., Padzor, A.G., Sayson, D.: Game data mining: clustering and visualization of online game data in cyber-physical worlds. *Procedia Comput. Sci.* **112**, 2259–2268 (2017)
10. Bushee, B.J.: The influence of institutional investors on myopic R&D investment behavior. *Account. Rev.* **73**(3), 305–333 (1998)
11. Chong, E., Han, C., Park, F.C.: Deep learning networks for stock market analysis and prediction: methodology, data representations, and case studies. *Expert Syst. Appl.* **83**, 187–205 (2017)
12. Chong, T.T.L., Ng, W.K.: Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. *Appl. Econ. Lett.* **15**(14), 1111–1114 (2008)
13. Choudhry, R., Garg, K.: A hybrid machine learning system for stock market forecasting. *World Acad. Sci. Eng. Technol.* **39**(3), 315–318 (2008)
14. D’Agostino, M., Gabbay, D.M., Hähnle, R., Posegga, J.: *Handbook of Tableau Methods*. Springer, Heidelberg (2013)
15. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)

16. Ding, X., Zhang, Y., Liu, T., Duan, J.: Deep learning for event-driven stock prediction. In: IJCAI, pp. 2327–2333 (2015)
17. Duarte, J.: Trading Options for Dummies. Wiley, Hoboken (2017)
18. Dubois, P.M., Han, Z., Jiang, F., Leung, C.K.: An interactive circular visual analytic tool for visualization of web data. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 709–712. IEEE (2016)
19. Fama, E.F., French, K.R.: Size, value, and momentum in international stock returns. *J. Financ. Econ.* **105**(3), 457–472 (2012)
20. Fuente, A.D.D., Andreassen, O., Charrondière, C.: Monitoring mixed-language applications with elastic search logstash and kibana (ELK). In: Proceedings of ICALEPCS, pp. 9–12 (2015)
21. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint [arXiv:1408.5882](https://arxiv.org/abs/1408.5882) (2014)
22. Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M.: Stock market prediction system with modular neural networks. In: 1990 IJCNN International Joint Conference on Neural Networks, pp. 1–6. IEEE (1990)
23. Kordonis, J., Symeonidis, S., Arampatzis, A.: Stock price forecasting via sentiment analysis on Twitter. In: Proceedings of the 20th Pan-Hellenic Conference on Informatics, p. 36. ACM (2016)
24. Krollner, B., Vanstone, B., Finnie, G.: Financial time series forecasting with machine learning techniques: a survey (2010)
25. Laney, D.: 3D data management: controlling data volume, velocity and variety. *META Group Res. Note* **6**(70), 1 (2001)
26. Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., Allan, J.: Language models for financial news recommendation. In: Proceedings of the Ninth International Conference on Information and Knowledge Management, pp. 389–396. ACM (2000)
27. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436 (2015)
28. Lee, M.C.: Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Syst. Appl.* **36**(8), 10896–10904 (2009)
29. Levy, H.: Upper and lower bounds of put and call option value: stochastic dominance approach. *J. Finance* **40**(4), 1197–1217 (1985)
30. Li, J., Shi, Z., Li, X.: Genetic programming with wavelet-based indicators for financial forecasting. *Trans. Inst. Measur. Control* **28**(3), 285–297 (2006)
31. Liu, B., et al.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, vol. 2, pp. 627–666 (2010)
32. Lohar, P., Dutta Chowdhury, K., Afli, H., Mohammad, H., Way, A.: ADAPT at IJCNLP-2017 task 4: a multinomial Naive Bayes classification approach for customer feedback analysis task (2017)
33. Loria, S.: TextBlob documentation (2018)
34. Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., et al.: TextBlob: simplified text processing. *Secondary TextBlob: simplified text processing* (2014)
35. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**(4), 115–133 (1943). <https://doi.org/10.1007/BF02478259>
36. McDonald, B.: Loughran and McDonald financial sentiment dictionary (2012)
37. Melvin, M., Taylor, M.P.: The crisis in the foreign exchange market. *J. Int. Money Finance* **28**(8), 1317–1330 (2009)

38. Mittermayer, M.A.: Forecasting intraday stock price trends with text mining techniques. In: 2004 Proceedings of the 37th Annual Hawaii International Conference on System Sciences, p. 10. IEEE (2004)
39. Mohammad, S.M., Turney, P.D.: NRC emotion lexicon. National Research Council, Canada (2013)
40. Naresh, G., Thiagarajan, S., Mahalakshmi, S.: Is there any real market indicator to predict stock index returns? A sem approach (2016)
41. Nguyen, T.H., Shirai, K., Velcin, J.: Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **42**(24), 9603–9611 (2015)
42. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: *LREc*, vol. 10, pp. 1320–1326 (2010)
43. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2009)
44. Plutchik, R.: The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **89**(4), 344–350 (2001)
45. Reis, T., Bornschlegl, M.X., Hemmje, M.L.: Towards a reference model for artificial intelligence supporting big data analysis. In: Proceedings of the 2020 International Conference on Data Science (ICDATA 2020) (2020, to appear)
46. Rosillo, R., De la Fuente, D., Brugos, J.A.L.: Technical analysis and the Spanish stock exchange: testing the RSI, MACD, momentum and stochastic rules using Spanish market companies. *Appl. Econ.* **45**(12), 1541–1550 (2013)
47. Schumaker, R.P., Chen, H.: Textual analysis of stock market prediction using breaking financial news: the AZFin text system. *ACM Trans. Inf. Syst. (TOIS)* **27**(2), 12 (2009)
48. Schumaker, R.P., Chen, H.: A discrete stock price prediction engine based on financial news. *Computer* **43**(1), 51–56 (2010)
49. Setty, D.V., Rangaswamy, T., Subramanya, K.: A review on data mining applications to the performance of stock marketing. *Int. J. Comput. Appl.* **1**(3), 33–43 (2010)
50. Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X.: Exploiting topic based Twitter sentiment for stock prediction. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 24–29 (2013)
51. Soni, S.: Applications of ANNs in stock market prediction: a survey. *Int. J. Comput. Sci. Eng. Technol.* **2**(3), 71–83 (2011)
52. Turing, A.M.: On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **2**(1), 230–265 (1937)
53. Upadhyaya, B., Khomh, F., Zou, Y.: Extracting restful services from web applications. In: 2012 Fifth IEEE International Conference on Service-Oriented Computing and Applications (SOCA), pp. 1–4. IEEE (2012)
54. Wang, Z., Zhang, Y., Lee, S.Y.M., Li, S., Zhou, G.: A bilingual attention network for code-switched emotion. In: Proceedings of the International Conference on Computational Linguistics (COLING 2016) (2016)
55. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: a survey. *Wiley Interdisc. Rev. Data Min. Knowl. Discov.* **8**(4), e1253 (2018)
56. Zheng, A., Jin, J.: Using AI to make predictions on stock market