Augmented Textual Features-Based Stock Market Prediction

SALAH BOUKTIF[ID], **(Member, IEEE), ALI FIAZ, AND MAMOUN AWAD**

Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding author: Salah Bouktif (salahb@uaeu.ac.ae)

ABSTRACT Due to its dynamics, non-linearity and complexity nature, stock market is inherently difficult to predict. One of the attractive objectives is to predict stock market movement direction by using public sentiments analysis. However, there is an active debate about the usefulness of this approach and the strength of causality between stock market trends and sentiments. The opinions of researchers range from rejecting the relationship to confirming a clear causality between sentiments and trading in stock markets. Nevertheless, many advanced computational methods have adopted sentiment-based features, yet did not attain maturity and performance. In this paper, we are contributing constructively in this debate by empirically investigating the predictability of stock market movement direction using an enhanced method of sentiments analysis. Precisely, we experiment on stock prices history, sentiments polarity, subjectivity, N-grams, customized text-based features in addition to features lags that are used for a finer-grained analysis. Five research questions have been investigated towards answering issues associated with stock market movement prediction using sentiment analysis. We have collected and studied the stocks of ten influential companies belonging to different stock domains in NASDAQ. Our analysis approach is complemented by a sophisticated causality analysis, an algorithmic feature selection and a variety of machine learning techniques including regularized models stacking. A comparison of our approach with other sentiment-based stock market prediction approaches including Deep learning, establishes that our proposed model is performing adequately and predicting stock movements with a higher accuracy of 60%.

INDEX TERMS Machine learning, model stacking, sentiment analysis, stock movement direction prediction, textual features extraction, tweets mining.

I. INTRODUCTION

Stock market movement prediction has massive benefits in academia and industry. In particular, accurate prediction helps investors make decisions and gain profit in the stock exchange. However, this prediction task is challenging due to the financial data nature that comprises noise, non-stationary, high degree of uncertainty, and chaotic characteristics. Moreover, the complex interaction of political and economic factors makes market prediction more difficult [1]. To develop an effective market trading strategy, it is essential to collect the appropriate data to learn stock movement patterns and trading behaviors. Many researchers have shown that social media data can be a valuable resource to recognize investors' patterns and decisions. In particular, sentiment

The associate editor coordinating the review of this manuscript and approving it for publication was Haruna Chiroma[ID].

analysis (SA), opinion mining, natural language processing (NLP), information retrieval, and structured/unstructured data mining [2] have been utilized to analyze and discover sentiment from texts and other communication mediums. Over the past few years, there has been an exponential growth in the use of social network platforms in sharing views, ideas, and reviews [3]. In particular, information pertaining to public moods in real time can be obtained and subsequently processed using these social media platforms.

Recently, there has been a debate on the usefulness of the public emotions expressed through social media in predicting the stock market movement. Indeed, some researchers have shown that sentiments and news could affect stock market movement and serve as potential predictors for trading decisions [4]–[6]. Some other researchers have questioned such techniques and affirmed minimal effectiveness of sentiment analysis in predicting the stock market movement [7]–[9].

Another halfway opinion, which we embrace, has emerged advocating that stock prices of certain companies are more susceptible to public sentiments than others; hence, mining sentiment analysis can lead to more predictable stock market if we use the appropriate analysis for these companies [10].

Besides, capturing and translating sentiment into numbers can generate different abstractions and conflicting interpretations of sentiment features. Moreover, several techniques of sentiment extraction can perform differently in various stock contexts. In other terms, the features and aspects of a company, a product, or a stock have significant roles in interpreting sentiments. For example, it has been shown that the rhythm of a company's stock price variation is affected by the volume of tweets followers interested in that stock/company [11]. However, the study did not suggest affirmative dependency between volume of tweets per unit of time.

In this work, we argue that using sentiment analysis to predict stock market movement is not mature enough and more exploration is needed to answer the following research questions:

- RQ1: What is the combination of textual analysis techniques that is most appropriate for stock market movement prediction?
- RQ2: Should we focus on leveraging data mining techniques to better discover embedded relationships between stock market and sentiment representations?
- RQ3: Does the predictability of sentiment analysis models depend on specific stock/company characteristics such as domain, stockholder backgrounds, volumes and origins of the available posts, etc.
- RQ4: Are we extracting and representing sentiments in the most efficient way in the context of stock market prediction? Or we need to explore various extraction and representation methods.
- RQ5: Should we focus on specific aspect of sentiment analysis rather than tackling generic NLP problems? Can aspect driven sentiment analysis be more effective?

In this paper, we attempt to give some answer elements to these research questions. In particular, we are advocating that finer-grained textual and sentiment analysis would predict the stock market movement direction more accurately. Therefore, we conduct a rigorous empirical study using various machine learning models (RQ2) for predicting stock direction. We also investigate the use of various text features, sentiment features, N-grams, and lags of historical stock prices in prediction (RQ1). We collected and explored stock data and tweets from ten NASDAQ-100 companies, namely, Google, Yahoo, Amazon, Apple, Alibaba, Tesla, Microsoft, IBM, Facebook, and Bitcoin. However, for space limitation, we evaluate ML models against six companies covering five different stock domains (RQ3), namely, Information technology (Microsoft and Apple), Electronic trading (Amazon), Services and Computing (IBM), Social Network (Facebook), and Car industry (Tesla). With regards to (RQ4), we will further evaluate

different machine learning algorithms when presented with various extracted text features and feature selection methods. Additionally, we will explore two different approaches for sentiment features extraction, namely, supervised-based approach using TextBlob, and Lexicon/Rule-based approach using VADER. Again, for the sake of a finer-grained sentiment analysis, we propose to focus on aspect-based sentiments (RQ5) that could be derived using the sentiment scores of certain individual words representing some stock market aspects.

The uniqueness of this research lies in addressing and testing the following research hypotheses;

1) Sentiments and public opinions are related to the actual changes in stock prices and an effective prediction model can be build based on that.
2) The combination of various historical stock prices, text-based and lagged features complemented by feature selection and regularized model stacking, could make considerable impact on the predictive performance of stock prediction model.
3) Individual characteristics of a company can influence the stock predictability.

The remainder of this paper is organized as follows. In Section II, we present the related works. We introduce the research backgrounds in machine learning models and feature selection in Section III. We describe our research methodology in Section IV. In Section V, we analyze and discuss our experimental results. Finally, we conclude our research and highlight future research directions in Section VI.

II. LITERATURE REVIEW

The literature is rich with sentiment analysis papers discussing the usage of tweets, financial news along with other relevant information to predict the stock movement [2], [4]. In this section, we focus on the recent research efforts on the prediction of stock market trend using social media data. However, there is a debate in causality between sentiments and trading in stock market. Researchers are divided between affirming such relationship and rejecting it. The two poles of opinions are inherited from two old theories. The first one derives from the Efficient Market Hypothesis (EMH) [12] where the stock market is believed to react instantaneously to any given news and that it is impossible to consistently outperform the market. The second pole of opinions comes from the Random Walk Theory (RWT) [13] where the stock market prediction is seen to be impossible and prices are determined randomly and outperforming the market is infeasible. Therefore, in our review, we distinguish between works advocating the prediction of stock market movements using sentiment analysis and works that do not believe in that.

A. STOCK PREDICTION ADVOCATES RELATED WORK
Literature has numerous research work that supports using sentiment analysis to predict stock market trends. One of the early works is by Bollen et al. where they found that

social, political, cultural and economic events are significantly correlated to the public mood levels (POMS) and accurate prediction results can be obtained using machine learning models when sufficiently large and representative data is available [5]. In another research, Bollen et al. have investigated on the public mood-driven assets. In particular, they have studied the correlation between public mood state and the closing value of the Dow Jones Industrial Average (DJIA) over time [3]. The Granger's causality analysis was used to determine correlations between features, and a self-organizing fuzzy neural network was trained to predict DJIA values on the basis of various combinations of past DJIA values and public mood data. Results demonstrated that DJIA predictions can be significantly enhanced by the presence of specific public mood features.

Ichinose and Shimada [14] proposed one-day stock price model to predict the Nikkei Stock Average using the content of articles in Yahoo-Japan-finance news. They used SVM and linear perceptron machine learning models. Simple bag-of-words (BOW) as well as BOW weighted by the absolute and actual values of the volatility score were used as model inputs. Various combination of models and features were tested and an accuracy of approximately 60% was obtained. Pagolu et al. investigated the correlation among changes in stock prices and the public mood extracted from tweets [15]. Logistic regression, random forest and SVM models were used for stocks movement prediction using N-grams and Word2vec features extracted from Microsoft stock related tweets. They concluded that a strong correlation exists between ''rise'' and ''fall'' in stock prices and the public opinions expressed through tweets.

Ly et al. in a very recent work [16], have also advocated the predictability of stock market. They have evaluated various machine learning (ML) algorithms and observed the daily stocks movement considering transaction cost and no transaction cost. They have shown that traditional machine learning algorithms have a better performance in most of the directional evaluation indicators without considering the transaction cost, however, DNN models showed better performance when considering transaction cost. Ly et al. work has focused on the techniques of modeling rather than on the feature selection and sentiment analysis problems.

In their analysis of the Chinese stock market, Sun et al. extracted data from microblogs, chat rooms and web forums. They identified a strong correlation and Granger causality between chat room posts sentiments and stock movement [6]. The performance evaluation of their trading strategy revealed reasonable and promising portfolio returns. Kollintza-Kyriakoulia *et al.* predicted the closing price of the next day for stocks of five companies, based on technical analysis, news articles and public opinions [17]. They used symbolic aggregate approximation and dynamic time warping to study the existence of a relation between stock closing price, tweets, and news articles. Both linear and SVM models showed improved results on those time periods where

patterns between stock price and the textual information were identified.

Bouktif and Awad [18] proposed an approach based on Ant Colony Optimization for combining Bayesian classifiers that used 19 public mood states collected from social network as input attributes to predict stocks movement directions. The results exhibited significant performance of the stock market prediction model as well as preserved results interpretability. Shah *et al.* analyzed the effects of news sentiments on the stock market [19]. They developed a dictionary-based sentiment analysis model and reported that news articles are powerful indicators for predicting short-term stock price movement. Picasso *et al.* used machine learning techniques to predict stock movement using indicators of technical analysis and the sentiment of news articles as inputs, to forecast the trend of a portfolio of twenty companies listed in NASDAQ100 index [20]. The predictive model showed robustness and effectively classified both positive and negative trends in the portfolio of stocks.

A closer work to ours is that of Hu *et al.* [21] where the authors address the challenge of using online content to predict stock market trend. They propose to imitate the three principles of the human beings learning process, namely, sequential content dependency, diverse influence, and effective and efficient learning. While Ziniu et al. were considering these principles, in our current work, we are covering five research questions that advocate not only the above three principles but also how to apply them along with other principles on the sentiment analysis for stock market movement direction. In particular, we propose using lagged features to consider sequential dependency (first principle), we take advantage from the big data obtained via text mining of a large corpus of tweets and from the historical stock prices (OHLCV) to consider diverse influence (second principle), and we use rigorous feature selection methods to optimize the learning process (third principle). Beyond the coverage of these principles, we are advocating and exploring a finer-grained sentiment analysis that rather focus on aspect-based sentiments (RQ5) that could be driven by the polarity and subjectivity scores of certain individual words or groups of words (i.e., N-grams) representing some stock market aspects. In addition, we are leveraging data mining techniques to better discover the embedded relationships between stock market and sentiments representations. This is achieved by proposing a regularized model stacking (an ensemble-learning method), as suggested by Shah *et al.* [22], on the top of a number of individual models using SVM, Naïve Bayes, ANN and XGBoost.

B. STOCK PREDICTION CRITICS RELATED WORK

In their research, Mudinas et. al. used various sentiment signal sources and different time periods to investigate the relationship between sentiment signals and stock price movements. Experimental results indicated that some stocks in some time periods exhibited strong cross-correlation

however, it was absent in other cases [23]. Porshnev *et al.* examined the prediction accuracy improvement of stock market by using data on psychological states of Twitter users [7]. They explored the use of two different lexicon-based approaches, namely, frequencies of words and the eight basic emotions in Twitter data. The results indicated that the addition of information from Twitter did not significantly augment predictive accuracy.

Li *et al.* [8] conducted experiments using five years historical Hong Kong Stock Exchange prices and news articles. Although the proposed models with sentiment analysis outperformed the bag-of-words model at the individual stock, sector and index levels, they did not perform well because they merely focused on sentiment polarity. Oliveira *et al.* [9] used several sentiment analysis methods to compare five popular lexical resources and two novel lexicons. Their sentiment indicators were based on daily words and individual tweet classifications using data from nine major technological companies. They found scarce evidence that sentiment indicators can explain the stock returns.

Lachanski *et al.* [24] thoroughly scrutinized the work of Bollen *et al.* [3] with an attempt to replicate the findings. They could neither predict the stock market out-of-sample accuracy nor reproduce the p-value pattern of Bollen's work. Serious concerns about the validity of Bollen's results were also raised in their review.

In this work, we are looking for a consensus between the two opinions. We are empirically investigating the predictability of stock market movement direction by using not only sentiments polarity or Bag-of-words, but also multiple finer-grained textual features. A vigorous and robust assessment of classification performance is conducted and presented. In addition, Latent Dirichlet Allocation (LDA) for tweets corpus authentication and Granger causality test are used to check whether twitter sentiments lags exhibit explanatory power for the stock movement prediction.

III. BACKGROUND
This section presents a brief overview on the categories of text-based features, machine learning models, classifiers stacking, feature selection approaches and performance metrics for evaluation used in this research.

A. SENTIMENT ANALYSIS
Sentiment analysis can be used to understand the public opinion towards a company or a product. Thus, we can automatically classify sentiments of millions of posts or tweets without requiring manual annotation. Sentiment classification is traditionally performed using both supervised and unsupervised methods, namely, machine learning and lexicon-based approach [22], [25]. In general, a machine learning algorithm attempts to minimize a cost function.

1) MACHINE LEARNING BASED SENTIMENT EXTRACTION
Supervised machine learning models are built from large labelled instances of text or sentences. These are modeled as

TABLE 1. Feature categories used and collected.

Feature Category	Definition
Sentiment Polarity	Sentiment orientation (positive, negative, or neutral). Polarity scores are within the range [-1,1] where 1 means positive statement and -1 means a negative statement.
Sentiment Subjectivity	Sentences with subjective expressions comprise opinions, beliefs and views while sentences with objective nature normally represent facts. Subjectivity scores are within the range [0,1], where 0 is very objective and 1 is very subjective
N-Gram	A contiguous sequence of words from a sentence emphasizing ordering. In this context and due to the limited tweet length, we extract unigram and bigram features. Bigrams can capture relationships between the words that are frequently used in sequence or co-occurrence.
Customer Tweet Features	Punctuation and special symbols that emphasize emotion or attitude [26] such as length of tweets, count of hashtags, emoticons, URLs, capital words, punctuation and question marks ('?' and '!') etc.
Lagged Features	Lagged features are the values at prior time steps that allow varying amounts of recent history to be brought into the forecast. Lagged historical prices and sentiments can have a significant effect on company returns [27]. We must be careful of the number of lags to be considered, as the higher the lag the fewer days that are available for training.
Calendar Related	It comprises calendar day, month, quarter, weekend/weekday etc.

a classification problem and use machine learning algorithms such as Naive Bayes, Support Vector Machine, and Maximum Entropy. However, due to possible dissimilarities in jargon of the source domain used to train the sentiment model and target domain to which it is applied, actual sentiment orientation may be affected.

2) LEXICON/RULE BASED SENTIMENT EXTRACTION
Lexicon based or unsupervised approaches classify data using dictionaries of words annotated with their positive or negative semantic orientation. These algorithms look up the text or sentence to find all known words and then combine their individual semantic orientations by averaging or summing their associated scores or values. A major drawback in this approach is that there is no a mechanism to deal with context dependent words.

B. CATEGORIES OF TEXT FEATURES
In stock market prediction, intrinsic informative features of text data can be extracted from tweets corpus. Various feature categories comprising sentiment polarities and subjectivities, N-grams, special character counts and lag values of features are discussed here. These feature categories will help in building better prediction models [26]. Table 1 present these categories definitions.

C. MACHINE LEARNING MODELS
In stock direction prediction, a variety of methods have been adapted amongst which supervised machine learning have

remained quite popular. In this research several supervised machine learning algorithms are trained to determine how the model performance varies when presented with historical price and text related features.

One of the widely used model is Naïve Bayes. The model is easy to build, interpret, and particularly useful for textual data. This model provides a family of probabilistic classifiers that are based on the Bayes theorem that assumes a strong independence characteristic within its feature vectors. However, such independence assumption is often violated. Though, these classifiers still tend to perform very well. Logistic regression is another widely used statistical modelling technique. The probability of a target/outcome is modelled as a logistic function of a linear combination of features. The model is fast to train with outputs having intuitive probabilistic interpretation. The algorithm can be regularized to avoid overfitting but it is usually limited to linear separable data.

SVM is also widely used machine learning technique for stocks classification. It searches for the optimal margin hyperplane to divide the up and down movement classes, such that the margin between these classes is maximized. This is performed by mapping the data into higher dimensional space using the kernel trick. SVM has shown resistance to overfitting problem, eventually achieving good generalization performance. As an ensemble method, Random Forest is used classification based on decision trees. The algorithm grows 'n' decision trees considered as weak classifiers, where each tree provides different kind of classification. All the trees are then merged into a forest. Trees are grown using sampling with replacement from a dataset and prediction conflict is resolved based on majority voting. Combining the results of these multiple trees helps to correct individual decision tree's tendency to overfit training set [28].

XGBoost or extreme gradient boosting is an improved version of gradient boosted machine learning algorithm that results in a high computational speed with an improved model performance by building more stable base models, thus reducing chances of overfitting. The boosted trees are fitted sequentially so that each new tree gives more weight to the mistakes of previous trees and therefore minimizing the loss [29]. In addition, XGBoost algorithm is able to rank the various features based on their importance during model construction. Artificial Neural Network (ANN) inspired by biological neurons has become very important method for stock market prediction because of its ability to deal with uncertainty, noise, and incomplete data, as well as subtle functional relationships in the data [30]. Two or three-layer feed-forward neural network is commonly used for stock prediction problems where the output layer has a single neuron with sigmoid transfer function. This results in a continuous value output between 0 and 1. A threshold of 0.5 is used to determine the up or down movement prediction.

Stacking is a meta learning approach that works by using a meta-classifier that learns from other algorithms prediction with the goal to generate an overall system that performs better than the individual classifiers. Different models can be used as the base and meta-learners in the stacking framework. In order to prevent overfitting, some form of regularization is used for the first and second level learners. Commonly used stacking framework is advocated to have several base and one meta-model, where the meta-model learns from the predictions of base models. The stacking approach has shown good results in stock prediction tasks [31].

D. FEATURE SELECTION AND EXTRACTION

In general, high-dimensional problems require an extensive amount of data to accurately train a machine learning model. Feature selection is the process of selecting a subset of features to circumvent the problem of curse of dimensionality, reducing training times and improving model performance [32]. Bagging and boosting tree-based models can be used to assess feature importance. Bagging uses multiple base learners that are generated in parallel and having equal weights in the ensemble committee. Boosting incrementally builds the ensemble model where the base learners are generated sequentially by training each new model to emphasize samples previously misclassified. While training a tree, the boosting algorithm computes how much each feature decreases the weighted impurity in the tree. Then the features can be ranked according to this measure.

SVM recursive feature elimination (SVM-RFE) is another technique that uses a linear kernel to select features by recursively considering smaller sets of features. Features importance is ranked by the model coefficients. Linear and kernel principal component analysis (PCA) are also used to reduce the number of features which can improve the training performance [33]. Whereas Kernel PCA is simply the nonlinear form of PCA, that can better make use of the complicated spatial structure of high-dimensional features.

IV. METHODOLOGY

To predict stock movements, we propose a machine learning approach that consists of five main steps, as depicted in Fig. 1. In step I, we scrape two sets of stocks data from online resources. These datasets include: (a) Open, High, Low, Close and Volume (OHLCV) price data on Amazon, Apple, Microsoft, IBM, Facebook, and Tesla stocks from Yahoo finance (b) Public tweets about these companies from Twitter. Both of these datasets span from 2008 to 2018. In step II we perform data pre-processing followed by extraction of various informative features from tweets using NLP techniques. In Step III we fit machine learning models, as designated in Section III, and evaluate model accuracies. If the accuracy is less than a certain threshold T, Step IV is triggered. In step IV, we perform feature selection and transformation and then refit the machine learning models with improved set of features. Models accuracies are then compared with those in previous step. Finally, in step V we further try to improve classification performance using regularized model stacking. Latent Dirichlet Allocation (LDA) is used for topic modelling to validate the legitimacy of tweets corpus while
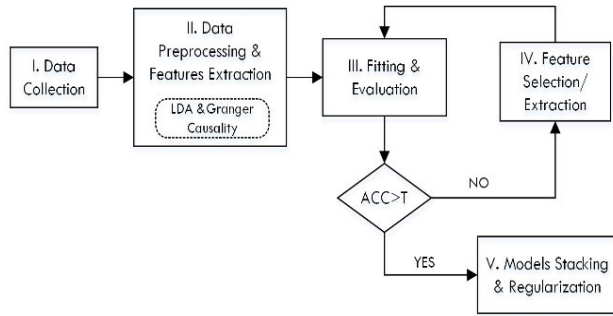
FIGURE 1. Research methodology.

TABLE 2. Distribution of data.

	Amazon			Apple		
	Training	Testing	%age	Training	Testing	%age
↑	1049	232	47.85	1184	305	52.93
↓	1092	304	52.14	1066	258	47.06
	Microsoft			Tesla		
↑	1071	244	49.51	884	222	50.89
↓	1053	288	50.48	854	213	49.10

Granger Causality analysis is used to check whether there is a statistically significant causality between sentiments driven from tweets and the stock returns.

We aim to forecast stock market movement using sentiment and texture features. We assume that the mood and the content of tweets reflect general feelings of society toward selected company's stock. Formally, we will be given a training set that will have an N points of the form $(x_i \ldots y_i), \ldots, (x_n \ldots y_n)$., where x_i is the set of features that includes sentiment features, text features, n-grams, OHLCV prices, lags of features, and calendar/tweets counts. y_t is the class to be predicted and it is defined as in equation 1:

$$y_t = \begin{cases} 1, & if \ p_t > p_{t-1} \\ 0, & otherwise \end{cases} \tag{1}$$

where p_t is the price of the stock price at time t. In other words, y_i indicates whether the stock price is '1 = up' or '0 = down'. We will apply different machine learning algorithms to predict y_i.

A. DATA SETS AND PREPARATION

We have collected historical stock price data for ten well-known NASDAQ-100 companies namely Amazon, Apple, Microsoft, IBM, Facebook, Google, Yahoo, Alibaba, Bitcoin, and Tesla for the period of 2008 to 2018 from Yahoo finance API representing more than 2800 trading days. The data consists of open, high, low and close bid prices as well as the time and the volume of the bids.

Additionally, we have collected tweets for the same period for those companies (See Github.com.Available online: https://github.com/mxawad2000/tweets_data_code). However, we report the results on six famous companies, due to the space limit, namely, Amazon, Apple, Microsoft, IBM, Facebook, and Tesla. Tweets are extracted using keywords, hashtag, cashtag and stock ticker. These tweets encompass both stock related tweets as well as people views on different company products and services. Careful choice of keywords was made to boost the number of relevant extracted tweets to get closer to the overall real market pattern. As a summary description of the collected datasets, they comprise approximately one hundred seventy thousand tweets for each company. For example, the stock movements of training and

test data for Amazon, Apple, Microsoft and Tesla stocks are shown in Table 2. The green/red (Up/Down) arrows indicate the direction of the price movement either it is up or down. Since the target variable, i.e., the stock movement direction, is nearly balanced in our datasets, the accuracy is an appropriate metric to evaluate the prediction performance.

Data preparation of the collected data is ensured by two operations. The first one is data preprocessing. It aims at cleaning and formatting the data and the second one is a validation step. It tends to validate the data against its semantic consistency with the stock topics.

1) TWEETS PRE-PROCESSING

Cleaning and standardization of tweets play an essential role for analysis of social media posts to make it noise-free. We have applied the following pre-processing steps on the tweets.

a) Filtering out spams, non-English, and context irrelevant tweets to the company or stock.

b) Removing retweets, tweets which contain the string "RT", and removing very short tweets with length less than 20 characters.

c) Uppercase characters are changed to lower case, thus preventing repetition of the same words in feature vector.

d) Discarding tweets with more than 90% of content matching with some other tweets.

e) Stemming and lemmatization: we performed morphological analysis to get root form of the words. Also, we used NLTK dictionary to remove stop words.

f) Removing punctuations, URLs, hashtags, and user IDs from the tweets.

2) LATENT DIRICHLET ALLOCATION (LDA) FOR CORPUS VALIDATION

In order to validate the downloaded tweets corpus against their relation with the stock topics, we use LDA technique to discover topics within the corpus [34]. LDA is a generative statistical topic model that can be used to classify text in a document to a particular topic. It treats each document as a mixture of topics, and each topic as a mixture of words. Within each topic we can determine the words with certain probabilities. Table 3 shows the example of the top two topics for Amazon and Apple stocks discovered using LDA.

TABLE 3. Topic modelling on tweets corpus.

Company	Topic Number	Word Probabilities
Amazon	1	0.142*"stock" + 0.072*"amzn" + 0.069*"company" + 0.066*"inc" + 0.019*"wallstreet" + 0.017*"google" + 0.016*"lower" + 0.015*"wall" + 0.014*"big" + 0.013*"street"
	2	0.170*"stock" + 0.149*"amazon" + 0.100*"inc" + 0.065*"nyse" + 0.064*"invest" + 0.062*"market" + 0.039*"investing" + 0.032*"stockmarket" + 0.028*"facebook" + 0.016*"news"
Apple	1	0.283*"stock" + 0.240*"apple" + 0.030*"inc" + 0.028*"market" + 0.019*"near" ' 0.014*"low" + 0.013*"investor" + 0.011*"buy" + 0.011*"service" + 0.011*"price"
	2	0.053*"deal" + 0.047*"new" + 0.046*"plus" + 0.037*"earnings" + 0.034*"lte" "+ 0.031*"unlocked" + 0.030*"aapl" + 0.027*"cellular" + 0.027*"factory" + 0.023*"drop"

Within each topic, the most probable words to appear in that topic are shown as word probabilities. As seen in Table 3, the topics and the associated keywords within each topic are related to the technology company's stocks and their products, thus it seems like a better fit for our prediction task.

B. FEATURE EXTRACTION FOR STOCKS
After necessary preparation of dataset, we extract useful features that can be used for stock movement prediction. These feature vectors would then be aligned with the ''up'' and ''down'' movement labels. Features are stored in a matrix format of N×M, where N is the number of samples and M is the number of features in the training set. The output matrix 'Y' is a N×1 matrix that stores the labels. Fig. 2 depicts different categories of the extracted features from tweets. We note that OHLCV (Open High, Low, Close, Volume,) stock prices are extracted from yahoo finance and the calendar related features characterize each day attached to the tweets.

1) SENTIMENT FEATURES EXTRACTION
We have explored two approaches for sentiment extraction from tweets, namely, machine learning based (SE1), and lexicon/rule-based (SE2) approaches. The sentiment extraction approach that is performing better in predicting stock movement would be used in our experimentation.

We have used TextBlob, a Python API for NLP, for sentiment analysis. The API gives the sentiment polarity and subjectivity scores for each tweet. Since there are multiple tweets per day, these sentiment features are aggregated on a daily basis to match the stock return data and represent total market sentiment indicator for that day. Fig. 3 shows examples of sentiments scores, for respectively positive, negative
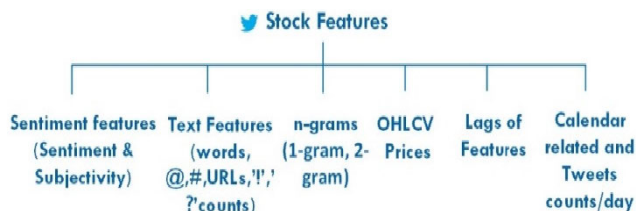

FIGURE 2. Extracted features for stocks prediction.
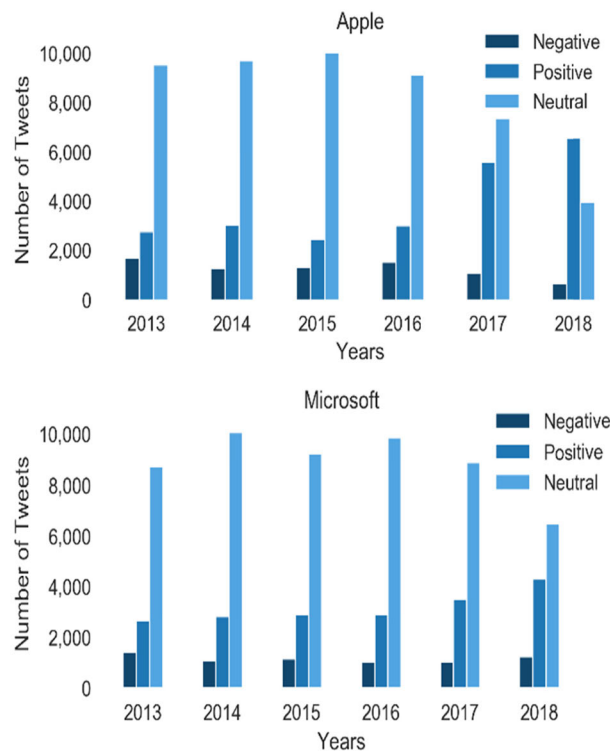

FIGURE 3. Sentiment counts for positive, negative and neutral tweets.

and neutral tweets for Apple and Microsoft spanning years 2013 to 2018.

For Lexicon/Rule Based sentiments extraction, we have used VADER (Valence Aware Dictionary and sEntiment Reasoner) which is a python library API for social media text analysis [35]. This unsupervised approach does not need any labeled data since the model is constructed from a generalizable, valence-based gold standard list of lexical features along with their associated sentiment intensities. The sentiment polarities are obtained by using unsupervised approach.

2) N-GRAM FEATURES EXTRACTION
For finer-grained analysis, we have extracted N-grams features from the tweet's corpus. Specifically, every appearing word sequence of length '1' and '2' is extracted from the tweets to form a dictionary of words and phrases. When building unigram and bi-gram features, we ignored the terms that have a frequency strictly higher or lower than a selected threshold, e.g., tenth and ninetieth percentile. This filtering helps to deal with the terms that appear too frequently
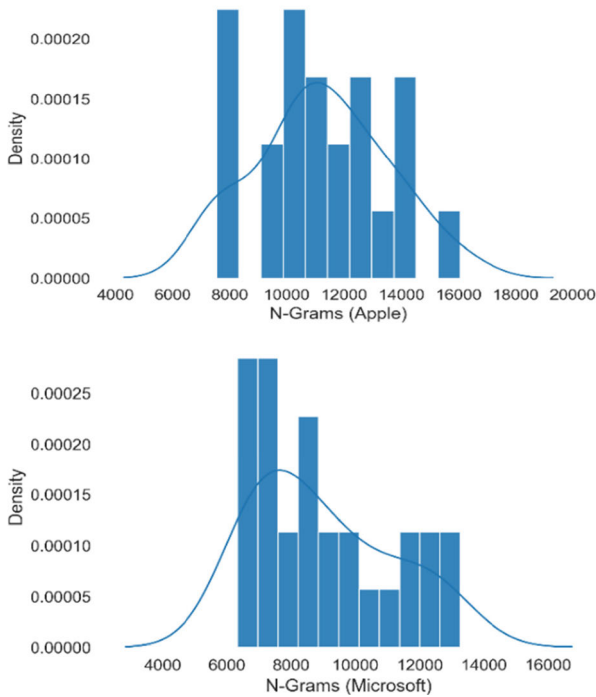
FIGURE 4. Distribution of N-Grams.

FIGURE 5. Count of tweets specific features for Amazon and Apple tweets.

(very common used English terms) or infrequently (rare used terms). Moreover, we considered the top features ordered by term frequency across the tweets corpus, thus making the extracted feature matrix denser. Distribution plots of N-grams for Apple and Microsoft tweets are shown in Fig. 4.

Some simple statistical analyses show that each company's tweets corpus exhibits a number of terms that are too frequent and indicate some aspects related to the activities of the company. Examples of these top-4 terms for Tesla company include the following bi-grams: ''tesla model'', ''tesla motor'', ''tesla roadster'' and electric car''. For Amazon, top-4 terms include the following 1-grams: ''book'', ''order'', ''buy'' and ''ship''. For Apple the terms ''iphone'', ''ipad'', ''buy'' and ''share'' are the top 4 1-grams. In Apple's tweet corpus, a statement saying 'I like the iphone but not the ipad' contains two opposing sentiments for two different Apple products. Behind these opposing sentiments, two aspects of Apple products that can better reflect the public opinion regarding Apple company. The same analysis can be carried for the other companies. For instance, we can consider different aspects of the Amazon company, like ordering, shipping and buying-service aspects. Therefore, the sentiment scores can be used to identify the most positive and negative tweets with respect to particular company aspects.

3) OTHER-TEXTUAL-FEATURES EXTRACTION
These are counts of some tweets features such as hashtags, mentions, capital words, URLs and punctuations. Fig. 5 depicts these feature counts from the Amazon and Apple tweets corpus.
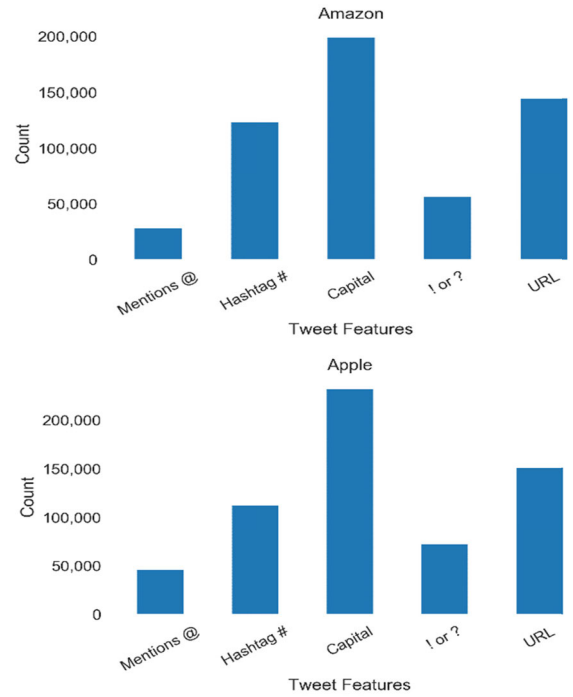
4) LAGGED FEATURES EXTRACTION
Lags of sentiment polarities, sentiments scores and other textual features are also created. An important consideration addressed here was to create an appropriate number of lags, as the higher the lag, the fewer the days that are available for training.

C. GRANGER CAUSALITY TEST FOR HYPOTHESIS CHECK
Since we are using lags of sentiments as features, we need to ensure that lagged terms are not simply redundant and exhibit explanatory power for the dependent variable which is the stock's closing price. Granger causality analysis is a linear regression-based technique that helps identify the correlation between two time series [36]. For our research, we want to examine whether daily changes of lagged Twitter sentiment polarities are useful to predict the movement of stock prices. First, we make the stock's closing prices time series stationary by taking log difference and performing Dickey-Fuller test to check stationarity. The null hypothesis states that the time series is non-stationary. Results for Dickey-Fuller test for the three stocks are shown in Table 4.

The results in Table 4 show that the test statistics value is less than 1% and 5% critical values and a p-value much smaller than 5% significance level, therefore we can reject the null-hypothesis and conclude that the time series is stationary. Next, we use Granger causality analysis between Twitter sentiment polarity and actual stock prices of the four companies. P-values for the highest four lags are shown in Table 5.

Results of Granger causality analysis shows that for the stock return of the companies, there exists a statistically

TABLE 4. Dickey-fuller test to check stationarity.

	Amazon	Apple
Test Statistic	-9.81583	-9.19
p-value	5.50e-17	1.16e-15
Critical Value (1%)	-3.53272	-3.4326
Critical Value (5%)	-2.85756	-2.8625
	Microsoft	**Tesla**
Test Statistic	-9.7406	-8.1111
p-value	8.5e-17	1.2e-12
Critical Value (1%)	-3.4328	-3.4335
Critical Value (5%)	-2.8626	-2.8629

TABLE 5. Granger causality test for the four stocks.

Lag No.	p-value	Lag No.	p-value
	Amazon		**Apple**
5	0.0244	**9**	0.0379
6	0.0219	**10**	0.0221
7	0.0439	**11**	0.0207
8	0.0412	**12**	0.0308
Lag No.	**p-value**	**Lag No.**	**p-value**
	Microsoft		**Tesla**
1	0.0437	**1**	0.022
2	0.0390	**2**	0.048
27	0.0366	**3**	0.168
28	0.0415	**4**	0.1423

significant causality between stock and the sentiments driven from the tweets for different lags at the 5% level of significance. Therefore, sentiment features can be used in our model as they encompass predictive information for the stock closing prices. For Tesla stocks for example, we found only two immediate lags to be statistically significant for the prediction.

D. MODEL TRAINING AND FEATURE SELECTION

The data set is divided into two parts. The training set comprises 80% of the dataset while the rest of the data is reserved for out-of-sample evaluation. Train set ranges from 2008 to 2017, while test set from 2016 to 2018 for the sliding window time series split validation. Several different learning algorithms were fitted on training data comprising Naïve Bayes, Logistic Regression, SVM, ANN, random forest and XGBoost. Extracted features were normalized before being subjected to each of these machine learning algorithms. Additionally, these exhaustive features were undergone into feature selection and dimensionality reduction. Specifically, we have applied Random Forest, XGBoost, and Recursive Feature Elimination techniques for feature selection. For dimensionality reduction, we have used Linear PCA and Kernel PCA transformation. Several experiments were then performed with the six machine learning models and five features subsets to discover how accurately the stock price can be predicted using the compact features dataset for each particular company.

TABLE 6. Machine learning techniques.

Symbol	Learning Technique
SVM	Support Vector Machines
LR	Logistic Regression
ANN	Artificial Neural Networks
RF	Random Forest
XGB	Extreme Gradient Boosting
SM	Stacked Model

TABLE 7. Feature selection techniques.

Technique	Symbol
SVM-RFE	SVM based Recursive Feature Elimination
PCA	Principal Component Analysis
K-PCA	Kernel PCA
RF	Random Forest
XGB	Extreme Gradient Boosting

TABLE 8. Scenario setup for out experiments.

ID	Feature Categories and Pre-Processing used in the Scenario
S1	Combination of historical stock prices, sentiment features generated by two different sentiment extraction approaches and lags of both prices and sentiments are used in model training.
S2	Combination of historical prices, sentiments, N-grams, text-characteristics, tweets count and lagged values are used in training.
S3	Integrated Features: combination of features processed with PCA, K-PCA, RF, XGB, and RFE techniques for dimensionality reduction. Top features that explain 90% of the variance in the data are used in training.
S4	Model stacking is applied on base classifiers with highest reported accuracies after performing the feature selection step in Scenario 3. A logistic regression model is used as the high level or meta-classifier to help improve classifier's generalisation [37]. The predicted probabilities for the stock movement from base classifiers will become meta-features for the logistic regression model, see Fig. 9.

V. EMPIRICAL RESULTS AND ANALYSIS

In this section, we present and analyze the empirical results that aim at answering the research questions RQ1, RQ2 and RQ3 formulated in Section 1. Furthermore, a comparison of the final augmented feature models is also made to current deep learning approaches in the stock domain.

A. EXPERIMENTAL SCENARIOS

We designed four experimental scenario-settings to predict the movement of stocks of selected six companies, namely, Amazon, Apple, Microsoft, Facebook, IBM, and Tesla. In each scenario, we use different set of features and applied feature selection techniques. Table 6 and Table 7 list the machine learning methods and feature selection techniques used in these scenarios. Table 8 explains the different scenario settings.

During initial experimentation, we have found that neither past historical prices nor sentiment analysis in their simple representation (i.e., polarities and subjectivities) is not useful
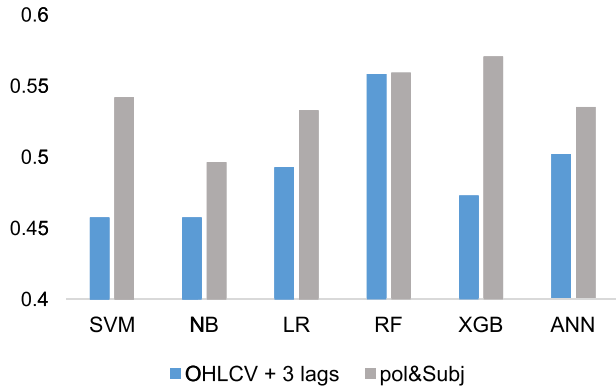
FIGURE 6. Comparison between OHLCV vs. Sentiments for Amazon.
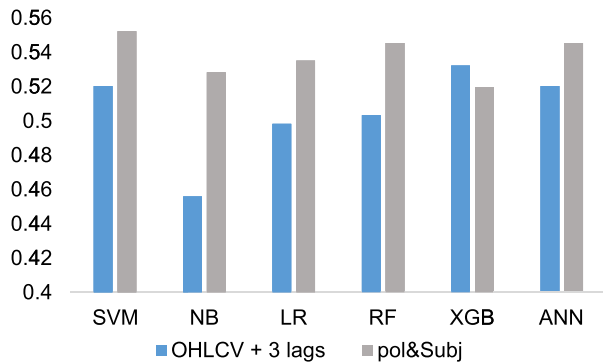

FIGURE 7. Comparison between OHLCV vs. Sentiments for Apple.

FIGURE 8. Results of Scenario 1.

FIGURE 9. Accuracy in all datasets as in Scenario 3 for all six stocks Amazon, Apple, Microsoft, Test, Facebook and IBM.

for prediction of stock direction. Fig. 6 and Fig. 7 view respectively, for Amazon and Apple stocks, the comparison in accuracy of two set of features, namely, OHLCV prices and sentiment polarity-based features. The results for Amazon and Apple cases show that machine learning techniques trained on sentiment features perform better, yet the overall accuracy is still low. Theoretically, these past trends are intuitive sources to track the history of stock movement and to reflect the trader/investor psychology, however, this allows only to improve slightly the accuracy above the majority class baseline. Additionally, it seems that polarity and subjectivity features are not enough representative of the expressed sentiments. They rather over simplify the public moods.

In Scenario 1, a combination of historical prices and sentiments extracted from two different sentiment analysis approaches along with their lags are used as inputs for model training. Fig. 8 shows the model accuracies for Amazon and Microsoft stocks achieved with scenario 1 using three set of features and the machine learning models described in Table 6.

The testing accuracies varied between the techniques and the feature categories. As seen in the experimental results, sentiment features extracted using the lexicon-based approach (SE2) gave better model performance than sentiments extracted with machine learning based-approach (SE1). Hence, SE2 would be utilized in subsequent scenarios. The lexicon-based sentiment extraction approach SE2 is more successful in dealing with social media texts as it considers
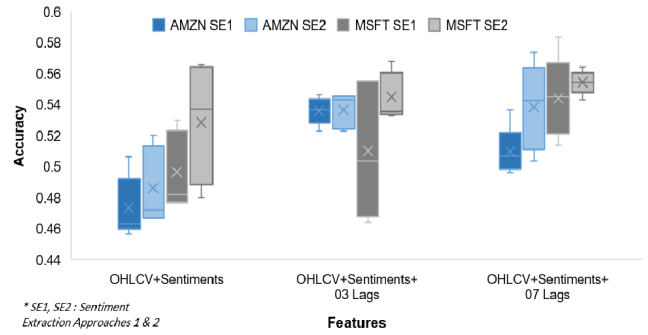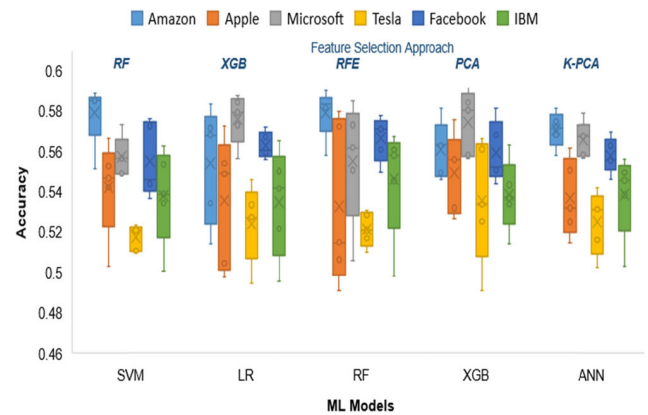
social media slangs, acronyms, emoticons and independent of model training on a text corpus belonging to a particular domain.

In Scenario 2, we integrate more fine textual features merging historical prices, sentiments, N-grams, text-characteristics, tweets volume and lags. As a result of integration, the training set becomes highly dimensional and sparse. As a consequence, the trained models suffered from the problem of overfitting and accuracy degraded by 2-3% for all learning techniques. Henceforth, it becomes important to use feature selection techniques to remove redundant and irrelevant features, thereby presenting most significant subset of features.

In Scenario 3, we apply feature selection techniques presented in Table 7 to empower the classification models and avoid the overfitting problem. The augmented and integrated features have contributed positively in improving the accuracy of the models. For example, the highest accuracy achieved with Scenario 1 and depicted in Fig. 8 is 57% while the highest accuracy achieved with Scenario 3 and depicted in Fig. 9 is 59%. Although the accuracy has improved, a preliminary finding with respect to the pairings of feature selection and machine learning for stock market prediction, is that there is no universal combination of techniques (i.e., machine learning and feature selection) that will perform the best in all the circumstances of the widely diverse and dynamic stocks markets. Notice that Scenario 3 is depicting
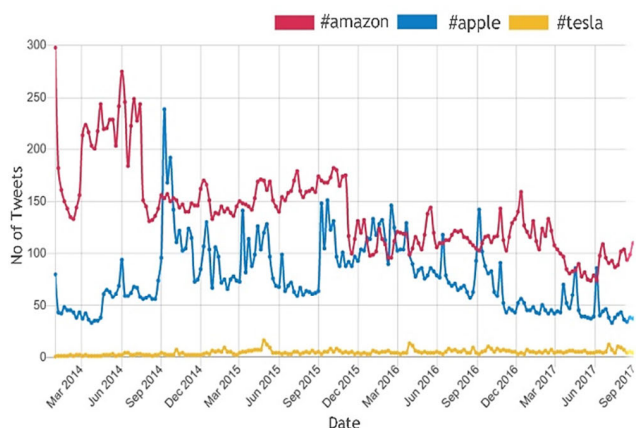
FIGURE 10. Hashtag trend for twitter for stocks.

FIGURE 11. Stacked model architecture.

TABLE 9. Performance metrics of stacked model for all six stocks Amazon, Apple, Microsoft, Test, Facebook and IBM.

Stock	Accuracy
Amazon	0.602
Apple	0.59
Microsoft	0.603
Tesla	0.578
Facebook	0.589
IBM	0.587

the accuracies for the six testbed stocks, namely, Amazon, Apple, Microsoft, Test, Facebook and IBM.

For each testbed stock, we report different accuracies depending on used pairing of feature selection technique and machine learning model.

The testing accuracy has improved for all learning techniques when features are augmented and carefully selected as shown in Fig. 9. For Tesla data set, improvement is low because it is relatively newer company with less products and services in the market compared to Amazon, Apple, Microsoft, IBM, and Facebook, therefore, fewer people tweets about the company products/services as evident by hashtag trend analysis (2014-2017) depicted in Fig. 10.

To support research question RQ2, we perform Scenario 4, where we advocate using stacked architecture encompassing best performing multiple classifiers obtained in Scenario 3. Our preference of stacking classifiers is motivated, on the one hand, by the importance of leveraging the already used machine learning to improve the predictability of stock market direction. As suggested by Shah *et al.* [22], ensemble classifier is very promoting when using sentiment analysis and a safer way to increase the predictive accuracy while avoiding the overfitting problem. On the other hand, Lv *et al.* in a very recent work [16], have evaluated various ML algorithms and observed the daily stocks movement considering transaction cost and no transaction cost have shown that traditional machine learning algorithms have a better performance in most of the directional evaluation indicators without considering the transaction cost, however, DNN models have shown better performance when considering transaction cost. A third reason of our preference of stacked classifiers is that these latter are not constrained by large data sets and the very large cost of model training. This reason is more legitimate in this work because we are not only trying to leverage ML techniques for stock market prediction but we are also trying to answer several other research questions.

We boost classifiers prediction capabilities by combining multiple classifiers and applied multiple methods of feature selection in model stacking ensemble, Fig. 11. Models are
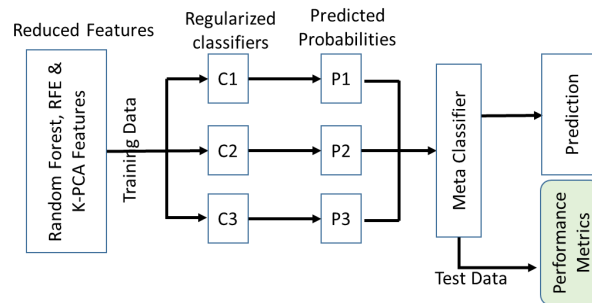
carefully regularized and scrutinized to determine whether their hyper-parameters setting could lead to overfitting. The final accuracy and performance metric results for the four stocks for short-term forecast up to 14 days ahead are shown in Table 9. We can observe that accuracy has improved for all datasets reaching 60% in the case of Microsoft and Amazon.

In summary, our experiments indicate that augmenting text features contributes positively in predicting stock market movement which supports RQ1. Then we attempted to capture both the past trends of the stocks market and public moods. Intuitively, these sources reflect the trader/investor psychology, however, these features alone did not improve the predictive ability of our classifiers. This is mainly because of the polarity and subjectivity features that are not enough representative and omitting some aspect details of the expressed sentiments.

Therefore, they rather over simplify the public moods. Additionally, the results show that machine learning techniques perform differently as expected, however, applying feature selection and dimensionality reduction (i.e., Scenario 3) and stacking the models in an ensemble fashion (i.e., Scenario 4) can improve prediction supporting RQ2. Other possibilities of using sophisticated learning approaches such as deep learning might need to be further investigated.

Regarding the hypothesis assuming that prediction depends on certain circumstances of the stock (RQ3), the results obtained for Tesla stock are relatively lower that obtained on the other stocks despite the augmented and integrated feature set in Scenarios 1,2 and 3. This is most likely due to the fact that Tesla is a new company with less products in the market and subsequently less followers and smaller volume of tweets.

TABLE 10. Two-sample t-test for Amazon at level 0.05.

Model	DNN	SVM	RF	Final Model
Mean	0.52483	0.51687	0.51457	0.59878
Std. Dev	0.01288	0.01387	0.01784	0.02186
p-value	0.000259	0.00001	0.000031	-
t-value	4.99234	7.75289	6.4134	-

TABLE 11. Two-sample t-test for Microsoft at level 0.05.

Model	DNN	SVM	RF	Final Model
Mean	0.52927	0.50012	0.5453	0.59877
Std. Dev	0.03620	0.04257	0.00712	0.02124
p-value	0.00203	0.00027	0.00014	-
t-value	3.77489	4.98513	5.53816	-

B. STACKED MODEL RESULTS SIGNIFICANCE

In this experiment setup, we show whether the average accuracy attained in the stacked model is really significant or due to random change. A comparison of the accuracy of our final stacked architecture model is performed with the baseline model including Deep Neural Network (DNN), Support Vector Machines and Random Forest. The training is conducted using the features as described in scenario 1 (i.e., Table 8). A 10-fold cross validation [38], [39] is used to develop a better estimation of model prediction accuracy. We create many train/test splits for short-term forecast (7-14 days ahead) using sliding window approach and average the errors over all these splits as shown in Fig. 12.

A one-tailed two-sample t-test for difference in means is then performed with the following hypothesis:

$$\text{Null Hypothesis H}_0 : \mu_1 - \mu_2 = 0$$
$$\text{Alternate Hypothesis H}_1 : \mu_1 > \mu_2$$

where μ_1 is the mean accuracy score of the final model and μ_2 is the mean accuracy for the other models

As shown in these results in Table 10, Table 11, and Table 12 for Amazon, Microsoft, and Facebook, respectively, a low p-value <0.05 and high t-value for the various models ascertains that the accuracy of the final proposed model is greater than DNN, SVM and random forest models at 0.05 level of significance for Amazon and Microsoft and at 0.10 level for Facebook.

Next, we compare our stacked model with various studies employing deep learning approaches reported in literature. The reported accuracies of various models including DNN, long short-term memory (LSTM), convolutional neural network (CNN) and CNN-LSTM models vary in the range 52-65% as described in Table 11.
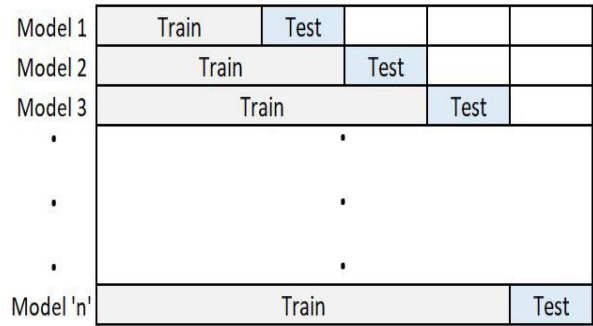

FIGURE 12. Sliding window time series split for model validation.

TABLE 12. Two-sample t-test for Facebook at 0.10 level.

Model	DNN	SVM	RF	Final Model
Mean	0.51283	0.51687	0.51457	0.5795
Std. Dev	0.02375	0.01387	0.01784	0.03486
p-value	0.047	0.0003	0.00001	-
t-value	4.81029	6.02053	7.57356	-

TABLE 13. Comparison with other deep learning models.

Paper Ref.	Deep Learning Models	Directional Accuracy (%)
[41]	DNN, LSTM	52.6, 60.6
[42]	Deep CNN	57.88
[43]	Deep CNN	61, 57.8, 57.4
[44]	LSTM	55.9
[45]	LSTM	60-65
[46]	CNN-LSTM	56.84
[21]	HAN-SPL	47.8

Comparing the accuracy of our model with these results in Table 13, manifests that our proposed model has a good prediction accuracy. Our empirical findings are in line with the opinion supporting the usefulness of sentiments for stock prediction as advocated in many reported works [5], [6], [14]–[22], [39]–[46].

However, there are still many questions that remain unanswered because of the complexity and dynamicity of the relationship between sentiments and stock movement direction.

VI. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have participated in the debate on the usefulness of sentiment analysis for stock market movement prediction. Being motivated by the large amount of valuable knowledge available in social media, we have utilized Twitter data as our information corpus to predict ups and downs of six well known NASDAQ companies. In particular, we advocated that using finer-grained textual and sentiment analysis would provide better predictive ability to discover the stock market movement direction. In the current work, we have

proposed a methodology that starts by extracting multiple text-based features to enrich the representation of sentiments. Then it applies diverse feature selection methods to contextually choose the appropriate feature sets for different circumstances, and ends by stacking individual models to get the best of base stock direction classifiers. As demonstrated in our empirical investigation, different machine learning algorithms and feature selection methods performed differently for various stocks, which would not be the case if the stock market had followed a random pattern. We conclude that rise-and-fall in stock prices of a company is affected by the public opinions or emotions expressed on twitter. Only we need more sophisticated ways of sentiment analysis to predict the stock market direction.

For the sake of maturing sentiment analysis for stock market predictions, more researches are needed towards (1) improving the representation of sentiment as a set textual features and (2) leveraging the abilities of machine learning algorithms. Accordingly, given the dynamic and complex nature of stock time series data, our future research involves mainly the investigation of variations of deep learning techniques for both sentiment features engineering and prediction modelling of the stock movement. This will involve for instance, extracting embedded words and stock aspects using approaches like Word2Vec, GloVe and FastText.

REFERENCES

[1] P. D. Yoo, M. H. Kim, and T. Jan, "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation," in *Proc. IEEE Int. Conf. Comput. Intell. Modelling, Control Automat., Int. Conf. Intell. Agents, Web Technol. Internet Commerce*, vol. 2, Nov. 2005, pp. 835–841.

[2] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[3] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.

[4] M. Zembik, "Social media as a source of knowledge for customers and enterprises," *Online J. Appl. Knowl. Manage.*, vol. 2, no. 2, pp. 132–148, Aug. 2014.

[5] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 450–453.

[6] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, "Predicting stock price returns using microblog sentiment for chinese stock market," in *Proc. 3rd Int. Conf. Big Data Comput. Commun. (BIGCOM)*, Aug. 2017, pp. 87–96.

[7] A. Porshnev, I. Redkin, and A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis," in *Proc. IEEE 13th Int. Conf. Data Mining Workshops*, Dec. 2013, pp. 440–444.

[8] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14–23, Oct. 2014.

[9] N. Oliveira, P. Cortez, and N. Areal, "Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from Twitter," in *Proc. 3rd Int. Conf. Web Intell., Mining Semantics (WIMS)*, 2013, pp. 1–8.

[10] L. Bing, K. C. C. Chan, and C. Ou, "Public sentiment analysis in Twitter data for prediction of a Company's stock price movements," in *Proc. IEEE 11th Int. Conf. E-Bus. Eng.*, Nov. 2014, pp. 232–239.

[11] R. L. Peterson, *Trading on Sentiment: The Power of Minds Over Markets*. Hoboken, NJ, USA: Wiley, 2016.

[12] E. F. Fama, "The behavior of stock-market prices," *J. Bus.*, vol. 38, no. 1, p. 34, Jan. 1965.

[13] B. G. Malkiel, *A random walk down Wall Street*. New York, NY, USA: W.W. Norton & Company, 1973.

[14] K. Ichinose and K. Shimada, "Stock market prediction from news on the Web and a new evaluation approach in trading," in *Proc. 5th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2016, pp. 78–81.

[15] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPES)*, Oct. 2016, pp. 1345–1350.

[16] D. Lv, S. Yuan, M. Li, and Y. Xiang, "An empirical study of machine learning algorithms for stock daily trading strategy," *Math. Problems Eng.*, vol. 2019, Apr. 2019, Art. no. 7816154.

[17] F. Kollintza-Kyriakoulia, M. Maragoudakis, and A. Krithara, "Measuring the impact of financial news and social media on stock market modeling using time series mining techniques," *Algorithms*, vol. 11, no. 11, p. 181, Nov. 2018.

[18] S. Bouktif and M. Adel Awad, "Predicting stock market movement: An evolutionary approach," in *Proc. 7th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage.*, 2015, pp. 159–167.

[19] D. Shah, H. Isah, and F. Zulkernine, "Predicting the effects of news sentiments on the stock market," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2018, pp. 4705–4708.

[20] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Syst. Appl.*, vol. 135, pp. 60–70, Nov. 2019.

[21] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2018, pp. 261–269.

[22] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *Int. J. Financial Stud.*, vol. 7, no. 2, p. 26, May 2019.

[23] A. Mudinas, D. Zhang, and M. Levene, "Market trend prediction using sentiment analysis: Lessons learned and paths forward," 2019, *arXiv:1903.05440*. [Online]. Available: http://arxiv.org/abs/1903.05440

[24] M. Lachanski and S. Pav, "Shy of the character limit:' Twitter mood predicts the stock market' revisited," *Econ J. Watch*, vol. 14, no. 3, pp. 302–345, 2017.

[25] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[26] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "NRC-canada: Building the State-of-the-Art in sentiment analysis of tweets," 2013, *arXiv:1308.6242*. [Online]. Available: http://arxiv.org/abs/1308.6242

[27] D. E. Allen, M. McAleer, and A. K. Singh, "Daily market news sentiment and stock prices," *Appl. Econ.*, vol. 51, no. 30, pp. 3212–3235, Feb. 2019.

[28] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1. Aug. 1995, pp. 278–282.

[29] Machine Learning Mastery. (Aug. 21, 2019). *A Gentle Introduction to XGBoost for Applied Machine Learning*. Accessed: Aug. 1, 2019. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/

[30] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, Nov. 1996.

[31] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers," *Int. J. Speech Technol.*, vol. 26, no. 1, pp. 25–33, Nov. 2006.

[32] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Aug. 2007.

[33] H. Yu, R. Chen, and G. Zhang, "A SVM stock selection model within PCA," *Procedia Comput. Sci.*, vol. 31, pp. 406–412, 2014.

[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[35] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. Conf. Weblogs Social Media*, May 2014, pp. 216–225.

[36] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[37] K. M. Ting and I. H. Witten, "Issues in stacked generalization," *J. Artif. Intell. Res.*, vol. 10, pp. 271–289, May 1999.

[38] L. Oneto, *Model Selection and Error Estimation in a Nutshell*. Cham, Switzerland: Springer, 2020.

[39] S. Bouktif and M. A. Awad, "Ant colony based approach to predict stock market movement from mood collected on Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2013, pp. 837–845.

[40] S. Bouktif, A. Fiaz, and M. Awad, "Stock market movement prediction using disparate text features with machine learning," in *Proc. 3rd Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Oct. 2019, pp. 1–6.

[41] D. Shah, W. Campbell, and F. H. Zulkernine, "A comparative study of LSTM and DNN for stock market forecasting," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 4148–4155.

[42] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," in *Proc. 7th Int. Conf. Cloud Comput. Big Data (CCBD)*, Nov. 2016, pp. 87–92.

[43] H. Gunduz, Z. Cataltepe, and Y. Yaslan, "Stock market direction prediction using deep neural networks," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[44] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1419–1426.

[45] Y. Wang, Y. Liu, M. Wang, and R. Liu, "LSTM model optimization on stock price forecasting," in *Proc. 17th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci. (DCABES)*, Oct. 2018, pp. 173–177.

[46] M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara, and A. G. Evsukoff, "Deep learning for stock market prediction using technical indicators and financial news articles," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

ALI FIAZ received the B.S. degree in mechatronics and control engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2008, the M.S. degree in industrial engineering from the University of Engineering and Technology, Peshawar, Pakistan, and the master's degree in computer science from COMSATS, Islamabad, Pakistan, in 2016. He is currently a Research Assistant with the Software Engineering Department, United Arab Emirates University. His research interests include the areas of machine learning, deep learning, optimization, and data mining.

SALAH BOUKTIF (Member, IEEE) received the degree in engineering and the master's degree in industrial computing from the School of Computer Science, University of Tunis, Tunisia, and the Ph.D. degree (Hons.) in computer science from the University of Montreal, in 2015. He is currently an Associate Professor with the College of Information Technology (CIT), United Arab Emirates University. Before joining CIT, in 2007, he was a Postdoctoral Fellow with the Department of Computer Engineering, Polytechnic School of Engineering, Montreal. He has been working as a Senior Software Engineer with Tunisian Railway Company (Known as SNCF) for three years. He has published many papers in highly ranked journals and conferences such as the IEEE ICSM, the IEEE ASE, ACM GECCO, the ACM/IEEE ASONAM, the IEEE ICWS, Elsevier IST, *ACM Transactions on IMS*, and *Plos One*. His research interests include energy prediction, data mining, big data analytics, search-based software engineering, and software quality assurance.

MAMOUN AWAD received the B.S. degree in computer science from the University of Baghdad, in 1994, the M.Sc. degree in computer science from Wichita State University, in 1999, and the Ph.D. degree in software engineering from The University of Texas at Dallas, in 2005. Since August 2006, he has been an Associate Professor with the College of Information Technology, United Arab Emirates University, where he also teaches and conducts research. His research interests include on the applications of data mining, big data, and AI.

● ● ●