ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse97942020.pdf
https://doi.org/10.30534/ijatcse/2020/97942020

A GCC Stock Market Classification Model using Sentiment Analysis based on HNBCs

Ghaith Abdulsattar A.Jabbar Alkubaisi[1], Siti Sakira Kamaruddin[2], Husniza Husni[3], Nura Said Al-Saifi[4]
[1]Muscat College, Sultanate of Oman, ghaith.alkubaisi@outlook.com
[2]Universiti Utara Malaysia, Malaysia, sakira@uum.edu.my
[3]Universiti Utara Malaysia, Malaysia, husniza@uum.edu.my
[4]Muscat College, Sultanate of Oman, nura@muscatcollege.edu.om

ABSTRACT
Sentiment analysis has become one of the most common method to classify stock market behaviour. Moreover, sentiment analysis has gained a lot of importance in the last decade especially due to the availability of data from social media such as Twitter. However, the accuracy of stock market classification models is still low, and this has negatively affected the stock market indicators. In this research, a model for GCC stock market classification based on sentiment analysis is constructed. It is designed to enhance the classification accuracy by the incorporation of tweet timestamp and location features, stock market domain expert labelling technique and the construction of a hybrid Naïve Bayes classifiers to classify the stock market sentiments. The methodology for this research consists of six phases. Data collection, labelling technique, data pre-processing, classification, performance and evaluation, and the final phase is recognition for the stock market behaviour. The model produced a significant result in classifying stock market behaviour with accuracy more than 89%. The model is beneficial for investors and researchers. For investors, it enables them to formulate their plans based on accurate indicators whereby it reduces the risk in decision making. For researchers, it draws their attention to the importance of feature engineering, labelling technique, and the classifiers hybridization in enhancing the classification accuracy.

Key words: Expert labelling, Hybrid Naive Bayes classifiers, Sentiment analysis, Stock market classification, Timestamp-Location features.

1. INTRODUCTION

Investors and businesspeople need to decide on an effective approach to improve the outputs of their investments and to avoid massive financial losses, mainly on investment in the stock market [1], [2]. The stock market is important because a company's stock prices play a pertinent role in all economic sectors [3], [4]. The global increment of the stock exchanges has raised the need for an in-depth decision-making tool using a stock market classification model [5], [6].

Accurate classification of the data sources in the stock market domain is necessary for investors to make suitable decisions, such as selling or buying stocks [7], [8], [9]. These kinds of investments need a pattern [10], [11] to assist decision makers in the stock market reach the right decision with minimal risk [11], [12]. To determine a suitable pattern, trends must be followed by observing the reactions of consumers to everything related to the product, such as the quality and price reported in company financial reports.

This reaction pattern can be easily captured by following social media to identify consumer reactions, such as posts and tweets [13], [14], [15] as well as companies' financial reports [16], [17]. For example, when a service company shares a tweet as an announcement about new services, if the customers tweet positively this will lead to an increase in the demand of that company's stocks. On the other hand, negative tweets will lead to a decrease in demand for the stocks. The last possibility is a neutral tweet, which does not affect the demand for stocks [13], [18]. These reactions represent valuable data: tweets can be mined and analyzed in order to construct useful indicators of these data to support and help the decision makers in selling or buying stocks on the stock exchange [19], [1].

Accessing a pattern with high degree of accuracy is still a major challenge for researchers in the domain of stock market classification using sentiment analysis on [20], [21], [22]. Stock market classification models still suffer from low accuracy in classification [23], [24], [25]; this affects the reliability of the stock market indicators which are extracted by following and analyzing social media data and stock prices [13], [26], [27].

Various factors can affect the results of a stock market classification model, including features such as company name, location and volume [28], [29], [30], [31], [32]; labelling technique [33], [34], [35]; and classification method [29], [30], [36].

From the dataset, important features such as the number of keywords and the interval of data collection can affect the performance of the model [29], [32], [31]. An instance of labelling technique affecting a stock market classification model is training using prior lexical knowledge; auto-labelling uses a lexical base to define the keywords' polarity as positive, negative or neutral [37], [34].Machine learning (ML) classifiers depend on the research requirements [29], [36]; for example, dataset size (large or small), dataset type (numbers or text), features, and parameter sets. This research therefore focuses on these factors because they have a significant direct effect on the reliability and accuracy of the stock market classification model [38], [39], [40], [41]. To predict stock market behaviour, different ML classifiers are used to perform sentiment analysis on Twitter, including the Support Vector Machine (SVM), Decision Trees and Naïve Bayes Classifiers (NBCs) [42], [43]. These classifiers must be chosen according to the requirements of the research domain and the research data [44], [45].

This research focuses on specific spatial and temporal features to distinguish the relationship of every feature to classification accuracy. It also assumes that expert labelling directly impacts classification accuracy. This assumption requires the parameters to be defined to show the effects of labelling technique individually. As the research domain involves the stock market, the classification findings have to be prompt and precise, with the ability to classify the different labelled datasets that represent real consumer reactions. The features, dataset and research domain that constitute the requirements of the research become the basis for the implementation of sentiment analysis on Twitter. This is done by hybridizing NBCs to produce a stock market classification model with increased accuracy of classification, and hence supporting the stock market decision-making process.

NBCs are selected because Naïve Bayes (NB) models with appropriate pre-processing are competitive with more advanced methods such as the SVM classifier [46], [47]. NB models are also essential here for the following reasons. First, the role of each of the proposed features of the classification theory can be proven through naïve assumptions, that is by assuming the independence of features; NB models simplify the calculation of probabilities by supposing the probability of each attribute belonging to a given class value is independent of all other attributes [42], [48]. Even though this assumption is strong, it provides fast and effective results [47], [49].

Secondly, the role of expert labelling in classification accuracy can be proven by defining parameters that subsequently tune the ratio of the labelled and unlabelled tweets. By hybridizing Semi-Supervised Naïve Bayes (SSNB) [50], [51] with Multinomial Naïve Bayes (MNB) [52], [53] and Bernoulli Naïve Bayes (BNB) [54], [55], the performance of the classification model using an expert labelled dataset and a general auto-labelled dataset [56], [57] can be judged.

Thirdly, NBCs represent four model estimators [42], [43]; the different models work to support the proposed model by endorsing the evaluation of the final results and the validity of each test using Bayes theorem and Naïve assumption [58], [59], [57].

Eventually, this research aims to classify consumer reactions for the purpose of assisting investors in stock market exchange to realize the situation of the company in the market based on the classification model's results. All NB models are simple to build; users are easily trained, and the method provides quick results [60], [61], [62], [63]. These models' attributes support the research domain requirements as users require rapid training and results, as well as providing high accuracy and excellent performance [13], [27].

2. RELATED WORK

This section reviews the latest stock market classification models using sentiment analysis on Twitter as shown in section 2.1 and 2.2.

2.1 Stock Market Classification Model using Sentiment Analysis on English Tweets

[64] has developed a stock market classification model using sentiment analysis of English tweets. He first tested the effectiveness of different ML methods on identifying a positive or negative sentiment from the tweets, applying his findings to two tasks: first, examining the relationship between Twitter sentiment and the stock prices; and secondly, locating which words in the tweets correlate to the impact in stock prices through a post-analysis of price shift and the tweets. Figure 1 shows the stock market classification model developed by [64].

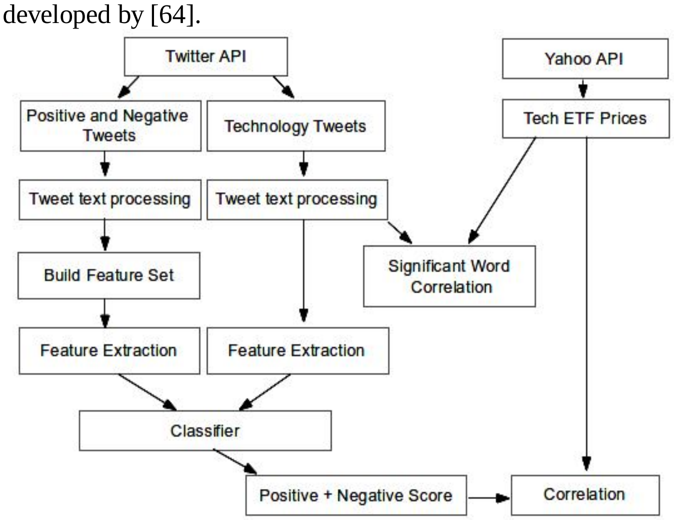

Figure 1: The Proposed Model for Stock Price and Significant Keyword Correlation

[65] also proposed a stock market classification model using sentiment analysis on English tweets. The model has four steps, Figure 2 shows [65] model.

Figure 2: The Proposed Model by [65]

[65] supported their model with Azure ML, as shown in Figure 3.

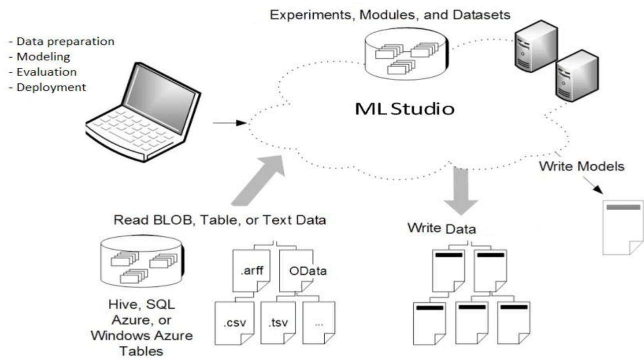

Figure 3: MS. Azure ML [66]

[67] proposed a model to classify the Indonesian stock market using simple sentiment analysis based on ML algorithms including SVM, NB, Decision Tree, RF, and Neural Network algorithms to classify tweets. Figure 4 shows the model proposed by [67].

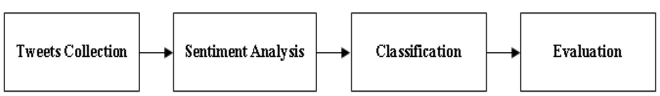

Figure 4: The Proposed Model by [67]

Finally, [68] developed model a to collect and process past tweets then test the effectiveness of different ML methods such as BNB and SVM for providing a positive or negative sentiment on the tweet. They employed the same ML methods to analyze how the tweets are correlated with the stock market price behaviour. Figure 5 shows the model proposed by [68].

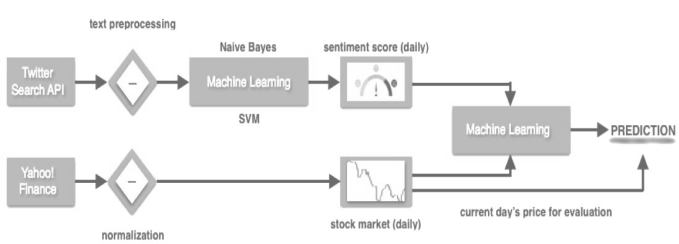

Figure 5: The Proposed Model by [68]

2.2 Stock Market Classification Model using Sentiment Analysis on Arabic Tweets

[69] study represents one of the few studies that has undertaken in Arabic sentiment analysis. It explains the significance of the pre-processing step as a key factor in achieving a high level of accuracy of sentiment analysis. The research was introduced for Saudi stock market tweets. It endeavored to clarify the relationship between Saudi tweets and the Saudi market index, using different implementations of KNNs, SVM, and NB algorithms. Figure 6 shows the model proposed by [69].

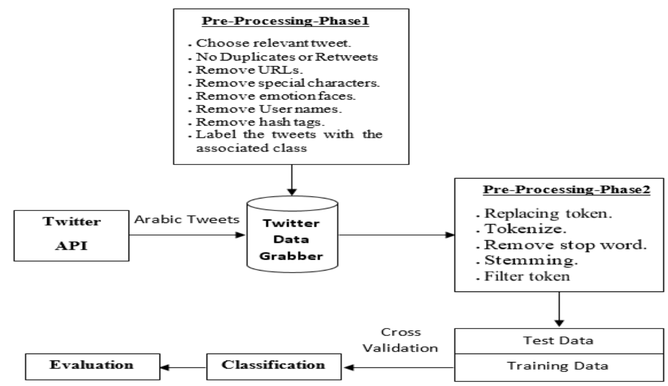

Figure 6: The Proposed Model by [69]

[70] also have proposed a sentiment analysis model as shown in Figure 7 for the SA stock market using sentiment analysis on Arabic tweets. It classifies the tweets into positive, negative or neutral, focusing on the role of the neutral class. The model was built based on the hybridization of ML classifiers such as SVM and NB, and the data pre-processing.

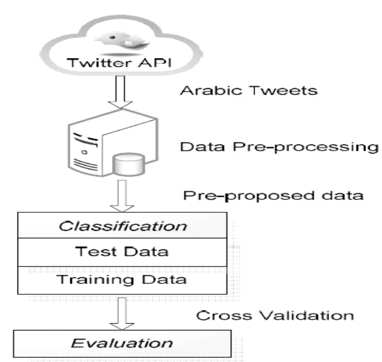

Figure 7: The Proposed Model by [70]

The last model in the domain of Arabic tweets was proposed by [71]; they aimed to enhance the labelling process, which has a direct impact on the reliability of the classification. Figure 8 shows their proposed model.

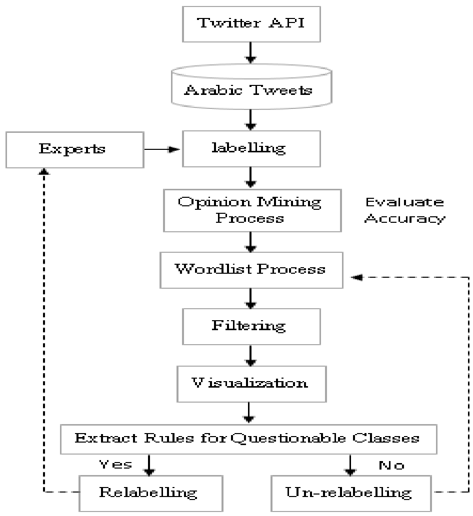

Figure 8: The Proposed Model by [71]

In summary, the common gaps observed in the reviewed stock market classification models are as follows. First, the collected tweets are not related to a specific company's services or products (To track the analysis of supply and demand based on consumer reactions about these services or products). Secondly, the labelling technique applied needs to be closer to the research domain (stock market). Thirdly, features are selected based on the generated corpus of tweets, rather than on the research domain fundamentals and requirements; this increases the size of the important features and reduces the size of the unnecessary features. Fourthly, the researchers randomly selected a supervised ML without relying on the research domain requirements, such as the size of the data set and the required features. Fifthly, all but one [70] classified the collected tweets as positive or negative polarity, ignoring the neutral position. Sixthly, the reviewed classification models suffered from low levels of accuracy, even that by [69]. Finally, greater reliability is required, for example by classifying more than one data set [72], [73], [74], [75].

3. THE GCC STOCK MARKET CLASSIFICATION MODEL

This research has constructed and implemented a stock market classification model to serve foreigner investors looking to invest in stock markets outside their own countries, especially in GCC, by providing them with accurate information about the behaviour a company's stock over a specific period. The model analyses the tweets using sentiment analysis based on HNBCs. Figure 9 shows the model using sentiment analysis of English tweets based on HNBCs as ML classifier.

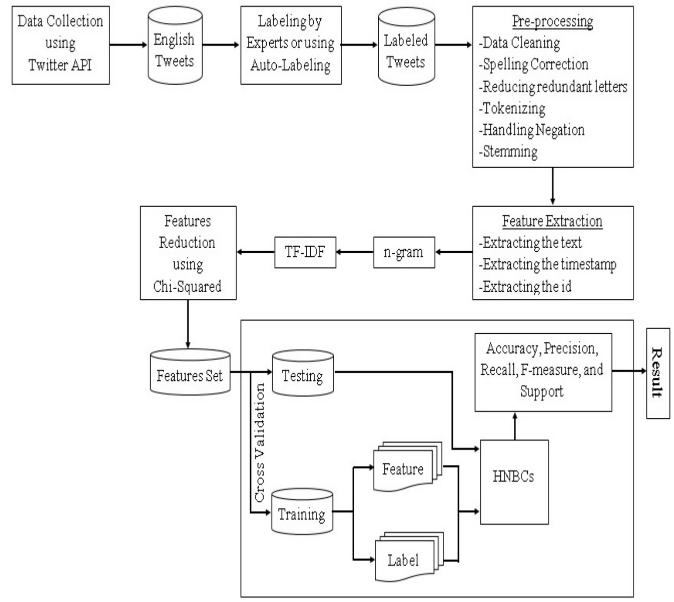

Figure 9: The Implemented Stock Market Classification Model using Sentiment Analysis on English Tweets based on HNBCs.

Figure 10 shows the model for Arabic tweets, based on HNBCs1 as the ML classifier.

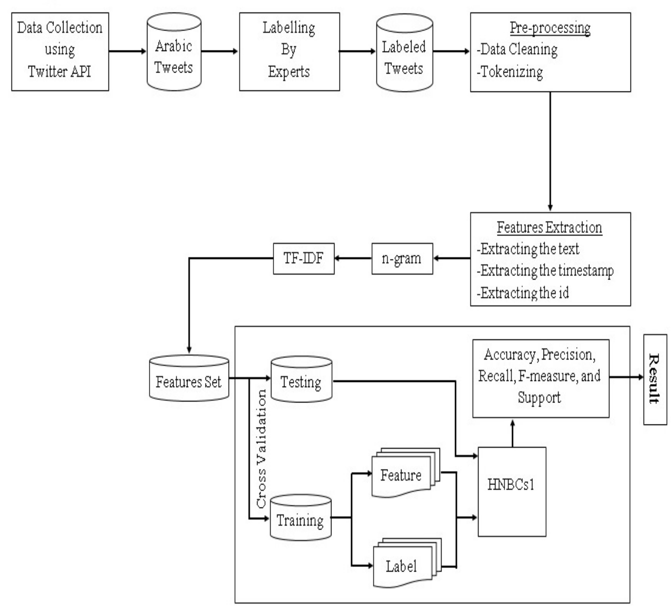

Figure 10: The Implemented Stock Market Classification Model using Sentiment Analysis on Arabic Tweets based on HNBCs

3.1 Data Collection using Tweets Collector

This research implements Tweepy method to collect targeted tweets from particular pages on Twitter like Almarai Company-Saudi Arabia, the Dubai Mall (DM), Etisalat United Arab Emirates (UAE), Dubai Marina Mall (DMM), and AlSafi Arabia (ASA). The implemented model needs to import spreadsheets (database) for use in the Python interpreter, it must rely on the Comma-Separated Values

(CSV) module. Algorithm 1 shows the implemented method.

Algorithm 1: Tweets collector using Tweepy library

1. Set Python interpreter
2. Import Tweepy library from (https://github.com/tweepy/tweepy)
3. Import csv module
4. Set Twitter API credentials
5. Define access key, access secret, consumer key, and consumer secret
6. Begin
   a. Get all tweets method
   b. Input the name to be searched and written on the screen (@ Twitter Page)
   c. Make an initial request for most recent tweets
   d. Save most recent tweets (order by timestamp) and the identification of the oldest tweet
   e. Repeat steps 3 and 4 until new data is exhausted (using while loop)
   f. Save all as ".csv" form
7. End
8. Return dataset.csv

The main objective for this research is to design and implementation for a GCC stock market classification model using sentiment analysis on Twitter based on HNBCs, so it focuses on consumer reactions and stock market behaviour in GCC market beside studying these five companies' tweets and stock separately. Almarai and ASA are represented on the Saudi Arabia (SA) stock market, DM and DMM on the Dubai Financial Market (DFM) under the title EMAAR Malls, and Etisalat UAE on the Abu Dhabi Securities Exchange (ADX). Tweets from all five companies have been collected from their official sites on Twitter. The research also needs a suitable data set to achieve its objectives, so the number of Twitter followers for each company is another important reason for selecting these five companies. Almarai (@almarai) has more than 496,000 followers on Twitter, DM (@TheDubaiMall) more than 814,000, Etisalat UAE (@etislalat) more than 2,000,000, DMM (@DXBMarinaMall) more than 10,000, and ASA (@alsafiarabia) more than 46,000.

3.2 Labelling Technique

In this research, this phase has done by experts in the domain of stock market analysis based on the stock market concepts and consumer reactions analysis to assign the appropriate sentiment weight for each tweet such as positive, negative, and neutral. The real sentiment weight is related to the effect of a tweet on stock behaviour such as positive or negative affection. Eventually, each tweet has labelled as positive: 1, negative: 2 and neutral: 0. All data sets have been sent to the experts in the domain of stock market and consumer reactions analysis to specifying the appropriate polarity for each tweet.

3.3 Tweets Pre-processing and Feature Representation

This research has implemented the pre-processing steps which are starting with transformations. Then, constructed two functions for a specific feature extraction there are temporal and spatial function. The last process of the pre-processing phase includes n-gram, Term Frequency-Inverse Document Frequency (TF-IDF) feature representation, and features reduction only for the English tweets. Algorithms 2 and 3 show the implemented method for the English and Arabic tweets transformation respectively.

Algorithm 2: English Tweets Transformations

1. Import [Dataset.csv]
2. Define cleaning data as a function
3. Begin
   a. Remove the urls
   b. Remove the special characters, Arabic tweets, and all non-English letters
   c. Remove numbers
   d. Remove hashtags
   e. Remove emojis
   f. Remove smileys
   g. Remove the reserved-words
   h. Remove the punctuation
   i. Handling Negation: removing the stop words except 'NOT' from the token because in sentiment analysis 'NOT' can play a major role in differentiate with other text. English stop-words such as is, are, etc.
   j. Spelling correction: this step is to correct the spelling for each word.
   k. Reduce redundant letters such as "cooool" to "cool"
4. End
5. Return cleaned dataset
6. Tokenize dataset
7. Begin
   a. Segment the text into sentences
   b. Break sentences into words
   c. Return tokens-dataset (string)
8. End
9. Begin
10. Stemming: is the process of reducing a word into its stem such as wait, awaited, and waiting all three have the same importance
11. End
12. Return the last [cleaned dataset]

Algorithm 3: Arabic Tweets Transformations

1. Import [dataset.csv]
2. Define cleaning data as a function
3. Begin
   a. Remove the urls
   b. Remove the special characters, unnecessary characters, English tweets, and all non-Arabic letters
   c. Remove numbers
   d. Remove hashtags
   e. Remove emojis
   f. Remove smileys
   g. Remove all reserved-words
   h. Remove the Arabic punctuation
   i. Remove Arabic stop words such as "من" and "غير"
4. End
5. Return cleaned dataset
6. Tokenize dataset
7. Begin
   a. Segment the text into sentences
   b. Break sentences into words
   c. Return tokens-dataset (string)
8. End
9. Return the last [cleaned dataset]

Algorithms 4 and 5 show the implemented method for the English and Arabic tweets filtration respectively.

Algorithm 4: English Tweets Filtration

1. Import the dataset
2. Begin
   a. Extract the text from each document
   b. Extract the timestamp from each document
   c. Extracting the location from each document by user id
   d. Represent all features using n-gram
   e. Compute TF-IDF
   f. Create TF-IDF vectorizer
   g. Reduce the generated feature set using chi-squared
3. End
4. Return the features with TF-IDF scores

Algorithm 5: Arabic Tweets Filtration

1. Import the dataset
2. Begin
   a. Extract the Arabic text from each document
   b. Extract the timestamp from each document
   c. Extracting the location from each document by user id
   d. Represent all features using n-gram
   e. Compute TF-IDF
   f. Create TF-IDF vectorizer
3. End
4. Return the Arabic features with TF-IDF scores

3.4 Classification based on HNBCs

HNBCs represent two classifiers, HNBCs1 and HNBCs2. Each classifier has the ability to classify data set in CSV format with the following array design: [id; timestamp; tweet.text; tweet's label]. Algorithm 6 shows the implemented method for the main model.

Algorithm 6: The Main Model

1. Import labelled dataset.csv
2. Begin
   a. Dataset Transformations
   b. Dataset Filtration
   c. Utilize n-gram
   d. TF-IDF Vectorize
   e. Feature Reduction using Chi-Squared
   f. Build the classification Model
   g. Select the classifier: HNBCs1, HNBCs2
3. End
4. Output the result of performance and evaluation

Algorithms 7 and 8 show the implemented method for the HNBCs1 and HNBCs2 respectively.

Algorithm 7: HNBCs1

1. Import the dataset.csv
2. Import Ensemble-Classifier
3. Define (Classifier1: MNB, Classifier2: BNB)
4. Begin
   a. Create a frequency table
   b. Create a binary table
   c. Calculate the probability for each class besides feature
   d. Determine the maximum probability
   e. Ensemble-Classifier = voting-classifier (estimators = [(Classifier1: MNB), (Classifier2: BNB)], voting = 'soft')
   f. Predicted = cross validation predicts (Ensemble-Classifier)
5. End
6. Output the performance and evaluation (print 'Accuracy, precession, recall, F-score, support')

Algorithm 8: HNBCs2

1. Import the dataset.csv
2. Define (Classifier2: BNB, Classifier 3: SSNB)
3. Begin
   a. Create a binary table for all generated features in the training set
   b. Create the binary unknown features matrix (UF-binary)
   c. Determine the initial weight for each single known and unknown document
   d. Calculate the probability of the features versus the classes
   e. Calculate the probability for each class
   f. Calculate the probability for all unknown document class
   g. Calculate the weight (W) for all unknown data
   h. Repeat for step number c while the maximum iteration is completed (depends on the size of the uploaded documents) or while observed no significant for the weight of the unknown data
   i. Create a binary feature vector for all features that related to a certain tested dataset
   j. Calculate the conditional probabilities for all classes
   k. Determine the result that has a maximum probability
   l. Predict the class for the classified document.
4. End
5. Return the predicted class (label)
6. Output the performance and evaluation (print 'Accuracy, precession, recall, F-score, support')

3.5 Performance and Evaluation

The aim of this phase is to evaluate the performance of each implemented HNBC. The evaluation covers the following measurements: precision, recall, F-measure, support, and accuracy. The most significant measurement for the implemented model is accuracy because it reflects the extent of the improvements in the proposed over other models using the same classifiers (NBCs) and same kind of data set. Measurement of accuracy is especially significant because it reflects the ratio of correct pattern recognition and specified polarity.

3.6 Stock's Behaviour

Decision making in stock market investment is based on the ratio of the data analysis accuracy, meaning that accurate analysis will lead to a safe decision; this is the main goal of this research. From another perspective, the goal involves increasing the classification accuracy by applying and constructing valid methods to help the decision makers in investing in the stock. They are provided with accurate information about a specific company, based on sentiment analysis of the company's tweets. Stock behaviour has three conditions: rising, falling and stable. Each behaviour reflects the size of the related polarity. For example, when the majority of tweets are negative, this will lead to a fall in the stock's behaviour as the demand for the company's products or services drops. Figure 11 shows this in a graphical form.

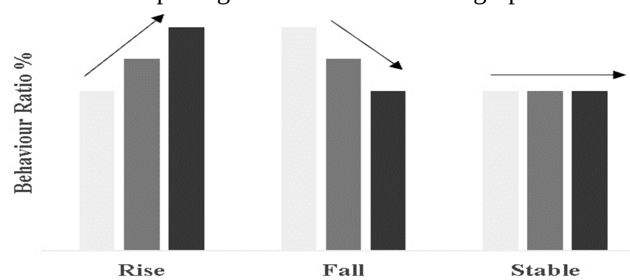

Figure 11: Stock's Behaviour

The behaviour ratio in Figure 11 is related to the percentage of rising, falling and stable stock behaviour on the stock market indicator. For example, when the stock indicator rises to 30%, the behaviour will be positively affected by 30%, and similarly with the other two behaviours.

4. EXPERIMENTAL RESULTS

This section presents the results of the HNBCs implementation using six different data sets: Almarai and ASA Arabic tweets, and Almarai, DM, DMM, and Etisalat UAE English tweets. The performance and evaluation of the model has five measurements as mentioned in Section 3.5. The following sections present the performance and evaluation results for the HNBCs.

4.1 HNBCs1 Performance and Evaluation

Tables 1 and 2 show the implementation results for HNBCs1 using Almarai and ASA Arabic labelled tweets. Almarai tweets were collected from 18 September 2016 to 25 May 2017, and the number of tweets for this test is 3,214. 3,168 ASA tweets were collected from 11 June 2015 to 4 January 2019.

Table1: HNBCs1 using Almarai Arabic Tweets

Class	Precision	Recall	F-Measure	Support
0	0.82	0.93	0.88	1053
1	0.95	0.88	0.91	1942
2	0.82	0.87	0.84	219
Total	0.90	0.89	0.90	3214

Classification Accuracy=**89.45%.**

Table 2: HNBCs1 using ASA Arabic Tweets

Class	Precision	Recall	F-Measure	Support
0	0.81	0.99	0.89	949
1	1.00	0.87	0.93	2120
2	0.62	0.94	0.75	99
Total	0.93	0.91	0.91	3168

- Classification Accuracy=**90.8%.**

The second data set contains 3,236 Almarai tweets after translation into English. Table 3 shows results of this second implementation.

Table 3: HNBCs1 using Almarai English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.83	0.95	0.89	1073
1	0.96	0.88	0.92	1945
2	0.83	0.87	0.85	218
Total	0.91	0.90	0.90	3236

- Classification Accuracy=**90.26%.**

The third test was based on the three data sets originally in English. For DM 3,097 tweets were collected from 26 March 2015 to 24 November 2017; for DMM, 3,159 tweets collected from 30 December 2014 to 29 January 2019; and for Etisalat UAE, 3,217 tweets collected from 8 February 2017 to 25 May 2017. Tables 4, 5, and 6 respectively show HNBCs1 these results.

Table 4: HNBCs1 using DM English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.01	0.02	0.01	60
1	0.97	1.00	0.99	2982
2	1.00	0.69	0.82	55
Total	0.96	0.98	0.97	3097

- Classification Accuracy=**97.51%.**

Table 5: HNBCs1 using DMM English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.90	0.92	0.91	1153
1	0.92	0.96	0.91	1820
2	0.98	0.41	0.58	186
Total	0.91	0.91	0.91	3159

- Classification Accuracy=**91.24%.**

Table 6: HNBCs1 using Etisalat UAE English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.95	0.94	0.95	2186
1	0.86	0.90	0.88	831
2	0.79	0.82	0.82	200
Total	0.92	0.92	0.92	3217

- Classification Accuracy=**92.1%.**

All achieved a high level of classification accuracy. In practice, when the number of neutral and negative tweets is less than positive tweets, this reduces the classification challenge because the model's learning algorithm learns more about the positive label and generates more relationships between the features and labels. The classification was done using six different data sets based on HNBCs1; each test shows that HNBCs1 has high classification accuracy even though size and language are different in the classified data sets.

Eventually, the classification performance measurements such as accuracy, precision, recall, and F-score reflect the classification reliability specifically when executed using high-bias data, so high precision and recall mean the model learning algorithm returned more relevant than irrelevant data. The F-score also relies on precision and recall, so it reflects the classification's reliability because when the F-score is high this means high precision and recall. Tables 1-6 show that the precision, recall, and F-score values are often close.

4.2 HNBCs2 Performance and Evaluation

The HNBCs2 classifier represents hybridization between BNB and SSNB. Tables 7-10 show the results with the same four data sets.

Table 7: HNBCs2 using Almarai English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.89	0.88	0.88	1073
1	0.93	0.92	0.92	1945
2	0.81	0.91	0.86	218
Total	0.91	0.90	0.91	3236

- Classification Accuracy=**90.48%.**

Table 8: HNBCs2 using DM English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.13	0.08	0.10	60
1	0.98	0.94	0.96	2982
2	0.24	0.91	0.38	55
Total	0.95	0.92	0.93	3097

- Classification Accuracy=**91.8%.**

Table 9: HNBCs2 using DMM English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.97	0.91	0.94	1153
1	0.94	0.98	0.96	1820
2	0.90	0.92	0.91	186
Total	0.95	0.95	0.95	3159

- Classification Accuracy = **94.99%.**

Table 10: HNBCs2 using Etisalat UAE English Tweets

Class	Precision	Recall	F-Measure	Support
0	0.98	0.93	0.96	2186
1	0.87	0.96	0.91	831
2	0.88	0.96	0.92	200
Total	0.95	0.94	0.94	3217

- Classification Accuracy = **94.25 %.**

HNBCs2 has a high level of classification accuracy with all data sets. The hybridization improves learning as it is based on the supervised learning of BNB and the semi-supervised learning of SSNB, with the expert labelling technique.

Finally, the performance and evaluation results for all the classifiers show that HNBCs have a high level of classification accuracy using different data sets and but with the same parameter values and labelling by experts. The hybrid model performance measurements such as accuracy, precision, recall and F-score were high, reflecting the reliability of the classification. High precision and recall mean the model learning algorithm generates more relevant relationships between features and labels (positive, negative, or neutral). These relationships lead to high classification accuracy and reliability at the same time. The two classifiers HNBCs1 and HNBCs2 were based on the hybridization which combined individual advantages in a single model. The combination was based on supervised learning using a multinomial classifier (MNB), multivariate classifier (BNB) and the semi-supervised classifier (SSNB). The performance and evaluation results reflect the importance of selecting appropriate classifiers to meet the research requirements.

5. BENCHMARKING

The comparison depends on the following standards [76], [77], [78]:

1. Source of the data set, for example online, internal, and logs.
2. Type of data set, for example text, and tweets.
3. Size of the data set, for example the same or different.
4. The implemented classifier, for example the characteristics and domain.

In this research, a comparison was made between the HNBCs and the baseline NBCs to see how hybridization, selected features and expert labelling improve the stock market classification accuracy. The same data set of tweets is used, and the classifiers are Naive Bayes Classifiers. Table 11 shows the classification accuracy for the baseline NBCs and the HNBCs using Almarai Arabic and English tweets.

Table 11: HNBCs Classification Accuracy vs NBCs

ML Classifier	Accuracy using Almarai English Tweets	Accuracy using Almarai Arabic Tweets
HNBCs1	90.26%	89.45%
HNBCs2	90.48%	-
NB	87.4%	87.9%
MNB	80.37%	75.97%
BNB	79.45%	75.42%
SSNB	77.9%	-

The table shows that the classification accuracy for the HNBCs is higher than for the baseline models, despite using the same data sets, English and Arabic tweets. Also, regarding the models reviewed in Section 2, Table 12 shows the results of the classification accuracy using the baseline NB classifiers against the results of the classification models constructed in this research.

Table 12: HNBCs Classification Accuracy vs the Reviewed Classification Models based on NB

Model	English Dataset				Arabic Dataset	
	Almarai	DM	DMM	Etisalat UAE	Almarai	ASA
HNBCs1	90.26%	97.51%	91.24%	92.1%	89.45%	90.8%
HNBCs2	90.48%	91.8%	94.99%	94.25%	-	-
Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation by Linhao Zhang (2013).	86.1%				-	
Stock Price Prediction using Linear Regression based on Sentiment Analysis. Yahya Eru Cakra and Bayu Distiawan Trisedya, (2015).	56.50%				-	
Stock Price Forecasting via Sentiment Analysis on Twitter. John Kordonis, Symeon Symeonidis, and Avi Arampatzis, (2016).	80.6%				-	
Analysis of the Relationship Between Saudi Twitter Posts and the Saudi Stock Market. Hamed AL-Rubaiee, Renxi Qiu, and Dayou Li, (2015).	-				56.28%	
The Importance of Neutral Class in Sentiment Analysis of Arabic Tweet. Hamed AL-Rubaiee, Renxi Qiu, and Dayou Li, (2016).	-				76.86%	

Table 12 shows that the HNBCs achieved high classification accuracy using data sets of different size and language.

5. CONCLUSION

In conclusion, a GCC stock market classification model for sentiment analysis of tweets based on HNBCs was proposed and implemented. The model aims to provide stock market investors and decision makers with more accurate and reliable stock market indicators based on a real data set, labelled by experts. The research achieved its objectives and makes the following contributions:

a. Inclusion of a spatial and temporal function to extract the tweets' id and timestamp, incorporating them into the processed text before the data representation phase.
b. The expert labelling for the stock market classification model.
c. Construction of HNBCs for GCC stock market classification using sentiment analysis of Arabic tweets.
d. Construction of two different GCC stock market classifiers, HNBCs1 and HNBCs2 using sentiment analysis of English tweets.
e. Improving classification accuracy by constructing a stock market model using sentiment analysis of tweets.

6. RECOMMENDATIONS AND FUTURE WORK

Recommend that spatial and temporal features are necessary in stock market classification models that use sentiment analysis of tweets. Employing expert labelling techniques to construct the learning model reduces risks in decision making by increasing the reliability of the sentiment analysis, a necessary process for any classification model based on supervised ML methods. At the same time, the process of selecting a ML method must be based on the suitability of the selected method to the research domain and the purpose of the research.

For future work, we intend to increase the size of the data set for training and testing, to further improve the classification model outputs. At the same time, we are looking to add more related and specific spatial and temporal features, according to the decision-making concepts and fundamentals of the research domain.

REFERENCES

1. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). **Text mining for market prediction: A systematic review**. *Expert Systems with Applications, 41*(16), 7653-7670.
2. Ren, R., Wu, D. D., & Liu, T. (2018). **Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Mach**ine. *IEEE Systems Journal*.
3. Baker, M., Stein, J. C., & Wurgler, J. (2003). **When does the market matter? Stock prices and the investment of equity-dependent firms**. *The Quarterly Journal of Economics, 118(3)*, 969-1005.
4. Pan, L., & Mishra, V. (2018). **Stock market development and economic growth: Empirical evidence from China**. *Economic Modelling, 68*, 661-673.
5. Bartov, E., Faurel, L., & Mohanram, P. (2017). **Can Twitter help predict firm-level earnings and stock returns?**. *The Accounting Review, 93(3)*, 25-57.
6. Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). **Using Twitter trust network for stock market analysis**. *Knowledge-Based Systems, 145*, 207-218.
7. Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). **Using artificial neural network models in stock market index prediction**. *Expert Systems with Applications, 38*(8), 10389-10397.
8. Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., & Johnson, J. E. (2016). **Bridging the divide in financial market forecasting: machine learners vs. financial economists**. *Expert Systems with Applications, 61*, 215-234.
9. Zhong, X., & Enke, D. (2017). **Forecasting daily stock market return using dimensionality reduction**. *Expert Systems with Applications, 67*, 126-139.
10. Smedt, T. D., & Daelemans, W. (2012). **Pattern for python**. *Journal of Machine Learning Research, 13*(Jun), 2063-2067.
11. De Fortuny, E. J., De Smedt, T., Martens, D., & Daelemans, W. (2014). **Evaluating and understanding text-based stock price prediction models**. *Information Processing & Management, 50*(2), 426-441.
12. Nguyen, T. H., Shirai, K., & Velcin, J. (2015). **Sentiment analysis on social media for stock movement prediction**. *Expert Systems with Applications, 42*(24), 9603-9611.
13. Bollen, J., Mao, H., & Zeng, X. (2011). **Twitter mood predicts the stock market**. *Journal of Computational Science, 2*(1), 1-8.
14. Vu, T.-T., Chang, S., Ha, Q. T., & Collier, N. (2012). **An experiment in integrating sentiment features for tech stock prediction in twitter**. *Proceedings of the workshop on information extraction and entity analytics on social media data*, (pp. 23-38).
15. Kumar, V., Choi, J. B., & Greene, M. (2017). **Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects**. *Journal of the Academy of Marketing Science, 45*(2), 268-288.
16. Chanthinok, K., Ussahawanitichakit, P., & Jhundra-indra, P. (2015). **Social media marketing strategy and marketing outcomes: A conceptual framework**. *Proceedings of the Allied Academies International Conference: Academy of Marketing Studies*, (Vol. 20, No. 2), (p. 35). Jordan Whitney Enterprises, Inc.
17. Cao, G. (2017). **The Impacts of Information on Stock Prices Assessed by Social Media Sentiment**.

Proceedings of the IEEE International Conference on the Internet of Things (iThings), (pp. 1-8).
18. Zhang, X., Fuehres, H., & Gloor, P. A. (2011). **Predicting stock market indicators through twitter** "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, *26*, 55-62.
19. Kim, Y., Jeong, S. R., & Ghani, I. (2014). **Text opinion mining to analyze news for stock market prediction.** *Int. J. Adv. Soft Comput. Its Appl*, *6*(1).
20. Khaidem, L., Saha, S., & Dey, S. R. (2016). **Predicting the direction of stock market prices using random forest**. *arXiv preprint arXiv:1605.00003*.
21. Navale, G., Dudhwala, N., Jadhav, K., Gabda, P., & Vihangam, B. K. (2016**). Prediction of Stock Market Using Data Mining and Artificial Intelligence**. *International Journal of Engineering Science, 6539*.
22. Jeon, S., Hong, B., & Chang, V. (2017). **Pattern graph tracking-based stock price prediction using big data**. *Future Generation Computer Systems*, *80*, 171-187
23. Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). **Mining opinion components from unstructured reviews: A review**. *Journal of King Saud University-Computer and Information Sciences, 26*(3), 258-275.
24. Arvanitis, K., & Bassiliades, N. (2017). **Real-Time Investors' Sentiment Analysis from Newspaper Articles**. In *Advances in Combining Intelligent Methods*, (pp. 1-23): Springer, Cham.
25. Yong, S., TANG, Y.-r., CUI, L.-x., & Wen, L. (2018). **A text mining based study of investor sentiment and its influence on stock returns**. *Economic Computation & Economic Cybernetics Studies & Research, 52*(1).
26. Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). **More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates**. *Journal of Marketing, 77*(1), 87-103.
27. Lin, J., & Ryaboy, D. (2013). **Scaling big data mining infrastructure: the twitter experience**. *ACM SIGKDD Explorations Newsletter*, *14*(2), 6-19.
28. Bonde, G., & Khaled, R. (2012). **Extracting the best features for predicting stock prices using machine learning.** *Proceedings on the International Conference on Artificial Intelligence (ICAI),* (p. 1).
29. Prusa, J., Khoshgoftaar, T. M., & Dittman, D. J. (2015). **Using ensemble learners to improve classifier performance on tweet sentiment data**. In *the IEEE International Conference on Information Reuse and Integration (IRI)*, (pp. 252-257).
30. Prusa, J. D., Khoshgoftaar, T. M., & Dittman, D. J. (2015**). Impact of Feature Selection Techniques for Tweet Sentiment Classification**. In *the Twenty-Eighth International FLAIRS Conference*.
31. Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). **A Bayesian classification approach using class-specific features for text categorization**. *IEEE Transactions on Knowledge and Data Engineering, 28*(6), 1602-1606.
32. Iacomin, R. (2016). **Feature optimization on stock market predictor**. In *the IEEE International Conference on Development and Application Systems (DAS)*, (pp. 243-247).
33. Donmez, P., Carbonell, J. G., & Schneider, J. (2009). **Efficiently learning the accuracy of labeling sources for selective sampling.** *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 259-268).
34. Hasbullah, S. S., Maynard, D., Chik, R. Z. W., Mohd, F., & Noor, M. (2016). **Automated Content Analysis: A Sentiment Analysis on Malaysian Government Social Media.** *Proceedings of the 10th ACM International Conference on Ubiquitous Information Management and Communication*, (p. 30).
35. Canuto, S., Gonçalves, M. A., & Benevenuto, F. (2016). **Exploiting new sentiment-based meta-level features for effective sentiment analysis**. *Proceedings of the ninth ACM international conference on web search and data mining,* (pp. 53-62).
36. Prusa, J. D., Khoshgoftaar, T. M., & Seliya, N. (2016). **Enhancing Ensemble Learners with Data Sampling on High-Dimensional Imbalanced Tweet Sentiment Data**. In *the twenty-ninth international the FLAIRS Conference*.
37. He, Y., & Zhou, D. (2011). **Self-training from labeled features for sentiment analysis**. *Information Processing & Management, 47*(4), 606-616.
38. Jiang, L., Wang, D., Cai, Z., & Yan, X. (2007). **Survey of improving naive Bayes for classification**. In *the International Conference on Advanced Data Mining and Applications,* (pp. 134-145). Springer, Berlin, Heidelberg.
39. Sathyadevan, S., Sarath, P. R., Athira, U., & Anjana, V. (2014**). Improved document classification through enhanced Naive Bayes algorithm**. In *the IEEE International Conference on Data Science & Engineering (ICDSE)*, (pp. 100-104).
40. Yang, A., Zhang, J., Pan, L., & Xiang, Y. (2015). **Enhanced Twitter Sentiment Analysis by Using Feature Selection and Combination**. In *the International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec)*, (pp. 52-57).
41. Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). **Sentiment analysis leveraging emotions and word embeddings**. *Expert Systems with Applications, 69*, 214-224.
42. Marsland, S. (2015). ***Machine learning: An algorithmic perspective*** (Second ed.): CRC press.
43. Raschka, S., & Mirjalili, V. (2017). ***Python Machine Learning***: Packt Publishing Ltd.
44. Tanwani, A. K., Afridi, J., Shafiq, M. Z., & Farooq, M. (2009). **Guidelines to select machine learning scheme for classification of biomedical datasets**. In *the European Conference on Evolutionary Computation,*

Machine Learning and Data Mining in Bioinformatics, (pp. 128-139). Springer, Berlin, Heidelberg.

45. Ali, R., Lee, S., & Chung, T. C. (2017). **Accurate multi-criteria decision making methodology for recommending machine learning algorithm**. *Expert Systems with Applications, 71*, 257-278.
46. Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). **Tackling the poor assumptions of naive bayes text classifiers**. *Proceedings of the 20th international conference on machine learning (ICML-03),* (pp. 616-623).
47. Ting, S., Ip, W., & Tsang, A. H. (2011). **Is Naive Bayes a good classifier for document classification**. *International Journal of Software Engineering and Its Applications, 5*(3), 37-46.
48. Chandrasekar, P., & Qian, K. (2016). **The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier**. *Proceedings of the 40th IEEE Annual Computer Software and Applications Conference (COMPSAC),* (Vol. 2), (pp. 618-619).
49. Kuhn, M., & Johnson, K. (2013). ***Applied predictive modeling***, (Vol. 26). New York. Springer.
50. Han, W., Nan-feng, X., & Zhao, L. (2011). **An enhanced EM method of semi-supervised classification based on Naive Bayesian**. In *the Eighth IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD),* (Vol. 2), (pp. 987-991).
51. Bhattu, N., & Somayajulu, D. (2012). **Semi-supervised Learning of Naive Bayes Classifier with feature constraints**. *Proceedings of the 24th International Conference on Computational Linguistics,* (pp. 65-78).
52. Tan, S., & Zhang, J. (2008). **An empirical study of sentiment analysis for chinese documents**. *Expert Systems with Applications, 34*(4), 2622-2629.
53. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). ***Data Mining: Practical machine learning tools and techniques***. Morgan Kaufmann, *Acm Sigmod Record*, *31*(1), 76-77
54. Di Nunzio, G. M., & Sordoni, A. (2012). **How well do we know Bernoulli**? *IIR, 835*, 38-44.
55. Raschka, S. (2014). **Naive bayes and text classification i-introduction and theory**. *arXiv preprint arXiv:1410.5329*.
56. Jain, A., & Mandowara, J. (2016). **Text classification by combining text classifiers to improve the efficiency of classification**. *International Journal of Computer Application (2250-1797), 6*(2).
57. Catal, C., & Nangir, M. (2017). **A sentiment classification model based on multiple classifiers**. *Applied Soft Computing, 50*, 135-141.
58. Sarkar, B. K., & Sana, S. S. (2009). **A hybrid approach to design efficient learning classifiers**. *Computers & Mathematics with Applications, 58*(1), 65-73.
59. Jain, A., & Mandowara, J. (2016). **Text classification by combining text classifiers to improve the efficiency of classification**. *International Journal of Computer Application (2250-1797), 6*(2).
60. Gong, Z., & Yu, T. (2010). **Chinese Web Text Classification System Model Based on Naive Bayes**. In *the IEEE International Conference on E-Product E-Service and E-Entertainment (ICEEE),* (pp. 1-4).
61. Anjaria, M., & Guddeti, R. M. R. (2014). **Influence factor based opinion mining of Twitter data using supervised learning**. *Proceedings of the 6th IEEE International Conference on Communication Systems and Networks (COMSNETS),* (pp. 1-8).
62. Mahajan Shubhrata, D., Deshmukh Kaveri, V., Thite Pranit, R., Samel Bhavana, Y., & Chate, P. (2016). **Stock Market Prediction and Analysis Using Naïve Bayes**. *International Journal on Recent and Innovation Trends in Computing and Communication, 4*(11), 121-124.
63. Alkubaisi, G. A. A. (2019), **The Role of Ensemble Learning in Stock Market Classification Model Accuracy Enhancement Based on Naive Bayes Classifiers**, *International Journal of Statistics and Probability*, 9(1), 36-47.
64. Zhang, L. (2013). ***Sentiment analysis on Twitter with stock price and significant keyword correlation*** (Doctoral dissertation). Available from The University of Texas at Austin, Department of Computer Science.
65. Qasem, M., Thulasiram, R., & Thulasiram, P. (2015). **Twitter sentiment classification using machine learning techniques for stock markets**. In *the IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI),* (pp. 834-840).
66. Mund, S. (2015). *Microsoft azure machine learning*: Packt Publishing Ltd.
67. Cakra, Y. E., & Trisedya, B. D. (2015). **Stock price prediction using linear regression based on sentiment analysis**. *Proceedings of the IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS),* (pp. 147-154).
68. Kordonis, J., Symeonidis, S., & Arampatzis, A. (2016). **Stock price forecasting via sentiment analysis on Twitter**. *Proceedings of the 20th ACM Pan-Hellenic Conference on Informatics*, (p. 36).
69. Hamed, A.-R., Qiu, R., & Li, D. (2015). **Analysis of the relationship between Saudi twitter posts and the Saudi stock market**. In *the IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS),* (pp. 660-665).
70. Hamed, A.-R., Qiu, R., & Li, D. (2016). **The importance of neutral class in sentiment analysis of Arabic tweets**. *Int. J. Comput. Sci. Inform. Technol, 8*, 17-31.
71. AL-Rubaiee, H., Qiu, R., Alomar, K., & Li, D. (2018). **Techniques for Improving the Labelling Process of Sentiment Analysis in the Saudi Stock Market**. *International Journal of Advanced Computer Science and Applications*, *9*(3), 34-43.
72. Last, M., Tassa, T., Zhmudyak, A., & Shmueli, E. (2014). **Improving accuracy of classification models induced from anonymized datasets**. *Information*

Sciences, *256*, 138-161.

73. Korjus, K., Hebart, M. N., & Vicente, R. (2016). **An efficient data partitioning to improve classification performance while keeping parameters interpretable**. *PloS one, 11*(8), e0161788.
74. Maguluri, L. P., & Ragupathy, L. (2019). **A New sentiment score based improved Bayesian networks for real-time intraday stock trend classification**. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(4).
75. Lim, S. B. (2019). **Classification and Big Data Usages for Industrial Applications**. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(4), 1117-1122.
76. Kuncheva, L. I., Bezdek, J. C., & Duin, R. P. (2001). **Decision templates for multiple classifier fusion: an experimental comparison**. *Pattern Recognition, 34*(2), 299-314.
77. Chaovalit, P., & Zhou, L. (2005). **Movie review mining: A comparison between supervised and unsupervised classification approaches**. *Proceedings of the 38th IEEE Annual Hawaii International Conference on the System Sciences*, (pp. 112c-112c).
78. Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). **Document-level sentiment classification: An empirical comparison between SVM and ANN**. *Expert Systems with Applications, 40*(2), 621-633.