

# Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers

Ghaith Abdulsattar A. Jabbar Alkubaisi<sup>1</sup>, Siti Sakira Kamaruddin<sup>1</sup> & Husniza Husni<sup>1</sup>

<sup>1</sup> School of Computing, Universiti Utara Malaysia, Sintok, Malaysia

Correspondence: Ghaith Abdulsattar A. Jabbar Alkubaisi, School of Computing, Universiti Utara Malaysia, Sintok, 06010, Malaysia. E-mail: ghaith.alkubaisi@outlook.com

Received: December 27, 2017

Accepted: January 10, 2018

Online Published: January 11, 2018

doi:10.5539/cis.v11n1p52

URL: <http://dx.doi.org/10.5539/cis.v11n1p52>

## Abstract

Sentiment analysis has become one of the most popular process to predict stock market behaviour based on consumer reactions. Concurrently, the availability of data from Twitter has also attracted researchers towards this research area. Most of the models related to sentiment analysis are still suffering from inaccuracies. The low accuracy in classification has a direct effect on the reliability of stock market indicators. The study primarily focuses on the analysis of the Twitter dataset. Moreover, an improved model is proposed in this study; it is designed to enhance the classification accuracy. The first phase of this model is data collection, and the second involves the filtration and transformation, which are conducted to get only relevant data. The most crucial phase is labelling, in which polarity of data is determined and negative, positive or neutral values are assigned to people opinion. The fourth phase is the classification phase in which suitable patterns of the stock market are identified by hybridizing Naïve Bayes Classifiers (NBCs), and the final phase is the performance and evaluation. This study proposes Hybrid Naïve Bayes Classifiers (HNBCs) as a machine learning method for stock market classification. The outcome is instrumental for investors, companies, and researchers whereby it will enable them to formulate their plans according to the sentiments of people. The proposed method has produced a significant result; it has achieved accuracy equals 90.38%.

**Keywords:** sentiment analysis, twitter, stock market classification model, classification accuracy, naïve bayes classifiers

## 1. Introduction

In the recent years, one of the most popular domain using sentiment analysis on Twitter is the stock market (Yan, Zhou, Zhao, Tian, & Yang, 2016). Investors need a classification model to reduce risks in decisions making (Xu & Keelj, 2014). Being a public data source, Twitter data are readily available and accessible by everyone. Firms can directly acquire opinions of their customers from their tweets, and they can also predict their sentiments based on the acquired data. Twitter has more than 140 million active users who share almost more than 400 million tweets every day, sharing their thoughts with the public. They do also talk about trending topics or share incidents and experiences to get views and feedbacks from other users. These tweets help users and organizations to collect valuable information in different domains (Atefeh & Khreich, 2015). Firms, therefore are taking a further step by using Twitter as a data source as it is conveniently available and accessible. Data from tweets cover different kinds of events, news, as well as economics data such as stock markets, financial data associated with market indicators, e-commerce and marketing statistics (Geser, 2011; Thompson, 2008).

This study discusses the methods, models, and techniques to help investors in the stock market in making low-risk decisions using the sentiments of people. Many new models have been developed nowadays, but each model has its advantages and disadvantages. Current stock market classification models are still suffering from low classification accuracy (L. Zhang, 2013; Meesad & Li, 2014; Skuza & Romanowski, 2015; Navale, Dudhwala, Jadhav, Gabda, & Vihangam, 2016; Arvanitis & Bassiliades, 2017). Hence, this weakness in the models has a direct effect on the reliability of stock market indicators such as “series of statistical figures” and “financial reports” that explains the stock behaviour ( Bollen, Mao, & Pepe, 2011; J. Lin & Ryaboy, 2013; Ludwig et al., 2013).

Features of data, labelling techniques, and data classification techniques are instances of factors that affect the accuracy of classification model results (Jiang, Wang, Cai, & Yan, 2007; Sathyadevan, Sarath, Athira, & Anjana,

2014; Yang, Zhang, Pan, & Xiang, 2015; Alkubaisi, Kamaruddin, & Husni, 2017). This study focuses on these three factors concerning stock market classification. Features like size of keywords and the period of data collection are examples of the factors (Iacomin, 2016; Shah et al., 2016). One labelling technique, such as self-training uses prior lexical knowledge to justify the keywords polarity to be positive, negative and neutral (Zhou, Zhang, & Zimmerman, 2011; Hasbullah, Maynard, Chik, Mohd, & Noor, 2016). To produce significant results, a suitable machine learning classifier is used in classification (Padmaja & Fatima, 2013). These factors are focused in this study because they are essential in the domain of stock market and may have a direct impact on the accuracy and reliability of the stock market classification model (Jiang et al., 2007; Sathyadevan et al., 2014; Yang et al., 2015; Giatsoglou et al., 2017).

This analysis will help companies to examine their customers' views and feedback on services and products they are providing. Hence, improving their services in future. This will not only help them to attract new customers in future but also assist them in retaining current ones. Investors too can get feedback on the status of different companies in which they can invest in its stock.

This study uses sentiment analysis from tweets to generate a stock market classification model. This model will benefit in improving the accuracy of decision making. Besides this, the research also covers the issues regarding the analytical techniques that affect the accuracy of classification which are reflected by consumer reactions from Twitter. Accurate classification in the domain of stock market needs specific data with particular features and polarities. Thus, the role of analytical approach through the process of data analysis prepares and creates appropriate data with specific features and polarities to achieve high accuracy with the required reliability for classification (Schumaker & Chen, 2009; Larsen, 2010).

Furthermore, several classifiers have been used by researchers to perform sentiment analysis on the stock market data source. This study performs sentiment analysis on Twitter by proposing HNBCs to produce a stock market classification model to increase the classification accuracy and reliability to support decision makers in the field of the stock market. The proposed method represents combined algorithms, including Baseline Naïve Bayes (NB), Multinomial Naïve Bayes (MNB) (Tan & Zhang, 2008), and Semi-Supervised Naïve Bayes (SSNB) (Bhattu & Somayajulu, 2012). This study hybridizes the previous three algorithms for the following five main reasons:

- a. Assuming features independence (Chandrasekar & Qian, 2016).
- b. Deal with labelled data (Su, Shirab, & Matwin, 2011; Bhattu & Somayajulu, 2012).
- c. Adopt probability value (Fang, Hu, Li, & Tsai, 2013).
- d. Suitable for daily microblog tracking like (tweets) (Di Nunzio & Sordoni, 2012).
- e. Finally, it is appropriate for fast training with the quick result (Anjaria & Guddeti, 2014).

This study also proposes essential features namely: temporal and spatial features in relation to the stock market to improve and increase the accuracy of stock market classification to support sound decision making by investors and business people (Adam, Marcet, & Nicolini, 2016). Additionally, expert labelling is employed to improve the output of labelling step because every stock market has a specific feedback. Irrelevant words will be dropped, and concentration will be on the goal of the classification model to increase the exactness of input data in order to develop more related data before classification (Al-Ayyoub, Essa, & Alsmadi, 2015).

The other parts of this study are organized as follows: Section 2 discusses the related works regarding the relationship between stock market classification and sentiment analysis technique by explaining the feature selection, labelling technique, and the classification. Section 3 describes the proposed methodology. Section 4 illustrates the experiment and discussion. Section 5 discusses the recommendations and future work. The final section is the conclusion to the study.

## **2. Related Works**

### *2.1 Feature Selection*

A feature is an individual measurable property of a phenomenon was observed (Bishop, 2006), Feature selection represents an essential factor to improve the classification accuracy of any stock market classification model (Iacomin, 2016; Tang, He, Baggenstoss, & Kay, 2016; Yousef, Saçar Demirci, Khalifa, & Allmer, 2016; Lee & Kim, 2017). (Zhang, 2013) and (Makrehchi, Shah, & Liao, 2013) have primarily focused on the time feature in which data was collected from Twitter in sentiment analysis. The researchers found a significant correlation between stock returns and individual's reactions. In fact, valuable data in the domain of stock market should include several features like time, location, targeted audience, brand, and kind of service, but the most important

for the decision makers who are looking to invest in the stock market are time, brand and location (Janssen, van der Voort, & Wahyudi, 2017; Alkubaisi et al., 2017).

In this study, two features are selected to achieve more accurate results for classification. These two features are temporal and spatial features. The spatial feature represents the tweets based on location and temporal represents tweets based on timestamp (Song & Xia, 2016). It is important to note that data without timestamp and the location cannot support the decision makers in the field of stock market (Ruth & Hannon, 2012; Fernández-Avilés, Montero, & Orlov, 2012; Song & Xia, 2016).

## 2.2 Labelling Technique

An automatic labelling technique is used in the tagging step. This, however, affects the accuracy of classification results. This labelling technique automatically identifies sentiment expressed in a tweet based on the prior general lexicon not specifically relevant to the research area (Zhou et al., 2011). For example, stock market performance (Bollen, Mao, & Zeng, 2011), crime prediction (X. Wang, Gerber, & Brown, 2012) and tourism information (Shimada, Inoue, Maeda, & Endo, 2011). According to (Ahuja, Rastogi, Choudhuri, & Garg, 2015), the automatic labelling is still unable to tell the real public moods since the data is usually collected from logs directly, whereby the resulting labels need to be studied and tabulated to represent the actual users' sentiments. The logs include relevant and irrelevant data about users' Tweets. When research focuses on a specific area, it becomes necessary to focus on data pertinent to the research area. (Attigeri, MM, Pai, & Nayak, 2015) suggested that automatic labelling step could be improved for better sentiment calculations to achieve the required dataset which enhances the accuracy of the classification. Thus, the researchers used cumulative assessment of the sentiments regarding news article or tweets for sentiment analysis. Supplementing the automatic labelling problem, (F. Wang, Zhang, Xiao, Kuang, & Lai, 2015) stated that this sentiment calculation lacks specific dataset reflecting real feedback from the consumers. Real consumers' feedback will lead to improving the accuracy of the final result for the stock market classification model.

The existing stock market classification models which utilize the auto-labelling methods are based on the prior lexical knowledge, and they do not have a specific lexicon which is prepared before the sentiment analysis by the experts. Therefore, if the previous lexical knowledge does not match the domain of classification, the model will not achieve the required classification accuracy because it will not give the accurate polarity weight to the sentence (Giatsoglou et al., 2017; Jha & Mahmoud, 2017).

Anyhow, the labelling step should be compatible precisely with the domain of feedback and the aim of classification (Jha & Mahmoud, 2017). Expert analysts can accurately fit the field of input in the area of the stock market because they know the real polarity for each keyword (Liu, 2012). Besides, the integrated dataset with relevant words require features and actual polarity to generate more suitable learning algorithm which is used to extract patterns appropriate for research (Miranda & Abreu, 2015; Hasbullah et al., 2016).

## 2.3 Classifiers

Presently, the supervised machine learning methods such as NBCs are widely used in the domain of online data classification, concurrent with the rapid growth of internet users (Fiarni, Maharani, & Pratama, 2016). Sentiment analysis is a process that uses computational linguistics, Natural Language Preprocessing (NLP), and text mining to identify text sentiments as positive, negative or neutral. This technique has been known in the text mining field as emotional polarity analysis, opinion mining and review mining (Mostafa, 2013). In addition to calculating sentimental scores, the sentiment acquired from the text is compared to a dictionary definition in order to determine the strength of that opinion. Studies on sentiment analysis focus on text written in English, such as sentiment lexicons, whilst applying this to other languages could generate a domain adaptation problem (Cambria, Schuller, Xia, & Havasi, 2013).

The process of defining polarity of words is not natural because it depends on the domain, such as the stock market, health, and education (Padmaja & Fatima, 2013). Sentiment analysis technique affects the training set before classification, preparation, and detection of the polarity for the dataset helps to improve the classification accuracy (Abdelwahab, Bahgat, Lowrance, & Elmaghraby, 2015).

Machine learning approach for sentiment analysis such as NBC is widely used in the field of online data classification (Islam, Wu, Ahmadi, & Sid-Ahmed, 2007; Nagwani & Verma, 2014). The rapid growth of internet users and the need to analyses consumer's reaction from social networks has led to achieving a suitable pattern that has worked to support the decision of investors in the domain of stock market classification (T. H. Nguyen, Shirai, & Velcin, 2015). Analyzing consumers' reaction also helps to represent valuable data, and information and knowledge from social media such as tweets are extracted and analyzed to build useful indicators of

available datasets (Kim, Jeong, & Ghani, 2014). The baseline NBC needs to define two variables for different classifiers; the first variable for documents and the second one for the tokens. NBC uses these variables separately, which leads to increasing the period of training dataset (Bartov, Faurel, & Mohanram, 2015; W. A. Hussien, Y. M. Tashtoush, M. Al-Ayyoub, & M. N. Al-Kabi, 2016). The field of stock market needs fast train data set to follow up ongoing updates for the stock behaviour.

NBCs are probabilistic supervised machine learning classifiers that use Bayes rule whereby all features included in the data have assumed features independence. This means that there is no relationship between different features values (Vinodhini & Chandrasekaran, 2012). Furthermore, NBCs represent the most used supervised machine learning classifiers in the domain of text mining and data mining applications because it is characterized by simplicity and effectiveness (Islam et al., 2007; T. H. Nguyen et al., 2015). NBCs have four models, GNB, MNB, BNB, and SSNB. Hybridizing algorithms of these classifier models with different numbers of parameters and features lead to achieving the optimization (Nguyen & Armitage, 2008; Aggarwal & Zhai, 2012). This means that there is no relationship between different features values (Vinodhini & Chandrasekaran, 2012). This study adopts and combines three Bayes algorithms to formulate a better algorithm known as HNBC for the stock market classification model. The combined algorithms include NB, MNB, and SSNB.

Finally, NBCs have many advantages, one of them is fast training with fast results, and applying NBCs does not require much training. Simple training data is adequate to build a model and make a classification with low storage requirements (Netti & Radhika, 2015; Kharde & Sonawane, 2016). Assuming feature independence, features represent the attributes or characteristics of data. By using NBCs, each feature will be calculated independently, that will lead to raising the accuracy (Chandrasekar & Qian, 2016).

### 3. The Proposed Model

The proposed model is developed to break down the implementation processes of generating an enhanced classification model for stock market. The operations of model development begin with data collection phase (tweets datasets). It is essential to ensure that the collected tweets from Twitter are well pre-processed. The other phases implemented in this model are expert labelling phase, classification phase using HNBCs, as well as performance evaluation phase.

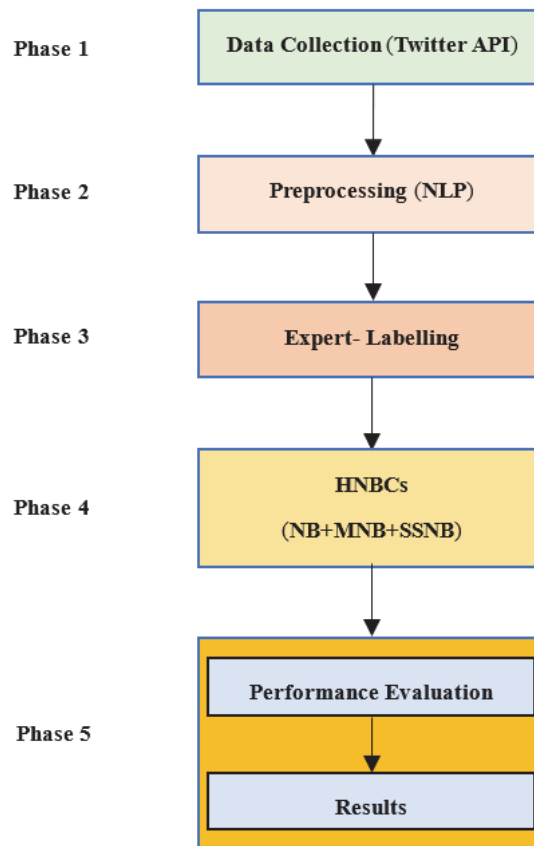


Figure 1. The proposed classification model

Figure 1 consists of five phases: firstly, the collection of dataset from Twitter using Twitter Application Program Interface (API) because the streaming can provide a continuous stream of the information with updates (Go, Bhayani, & Huang, 2009). Secondly, pre-processing the data set using NLP (Abdelwahab et al., 2015). NLP processing starts with the extraction process to extract the required features (Yang et al., 2015) and ends with transformations step (Kouloumpis, Wilson, & Moore, 2011). Thirdly, is the application of the expert labelling technique whereby the datasets are classified into positive, negative and neutral polarity. Fourthly, is the classification level using HNBCs. Finally, the classification model performance evaluation using F-measure, recall, precision, support, and accuracy.

### 3.1 Phase One: Data Collection (Twitter API)

This study focuses on tweets that reflect the consumer reactions about Al-marai Saudi Arabia (@almarai). In Twitter, there are two types of API's that are used to gather tweets, namely Twitter REST API and Twitter Streaming API (Go et al., 2009). This study utilizes an internal library using REST API, and Application Only Authentication (OAuth) required by the Twitter4j library. OAuth is used in Twitter4j library Twitter, and it supports access and provides authorized access to its API (Yamamoto, 2014). APIs streaming can provide a continuous stream of the information with updates. This study searches API and streaming API for Twitter sentiment analysis. Streaming API can access real-time tweets data using queries. This study chooses search API which is a REST API since it enables users to retrieve recently posted tweets by specific queries using HTTP methods (Makice, 2009). Moreover, it can filter results based on time, region, and language (Li, Lei, Khadiwala, & Chang, 2012).

The return of queries is a list of JSON objects containing tweets and metadata. These objects involve username, location, time and re-tweets (Aramaki, Maskawa, & Morita, 2011).

### 3.2 Phase Two: Data Pre-processing

The format of collected tweets needs to undergo pre-processing for data labelling and improve the classification result (W. A. Hussien et al., 2016). This study has adopted the whole steps of the pre-processing method, starting from text cleaning, removing white space, extracting abbreviation, removing stop words, and negation handling. The last process in pre-processing method works on extracting all relevant contents and the required features from tweets and dropping the irrelevant contents. This is named filtering step and all steps before it, are called transformations (Abdelwahab et al., 2015). The following explains the main two categories of the pre-processing method:

- a. Transformations step (Kouloumpis et al., 2011):
  1. Tokenizing text by splitting it using spaces.
  2. Removing stop words like or, also, and so on.
  3. Eliminating URLs, usernames, hashtags, Twitter symbols, punctuations and references.
  4. Reducing redundant letters such as "coool" to "cool".
- b. Filtering step: This step is related to the content extraction process from the collected tweets after the transformations step (Yang et al., 2015). This study focuses on extracting two essential features, which are Spatial and Temporal feature of tweets. There are two ways to obtain spatial information about tweets: the first one is automatically collecting the accurate spatial information available on Twitter and the second is approximately inferring the location of the user from the user profile (Chandra, Khan, & Muhaya, 2011; Jiang, Cai, Zhang, & Wang, 2013). This study also uses shape-let temporal selection. This type of feature selection assume that the time series are independent and identically distributed (H. Wang, Wu, Zhang, & Zhang, 2016). Figure 2 shows the data pre-processing steps

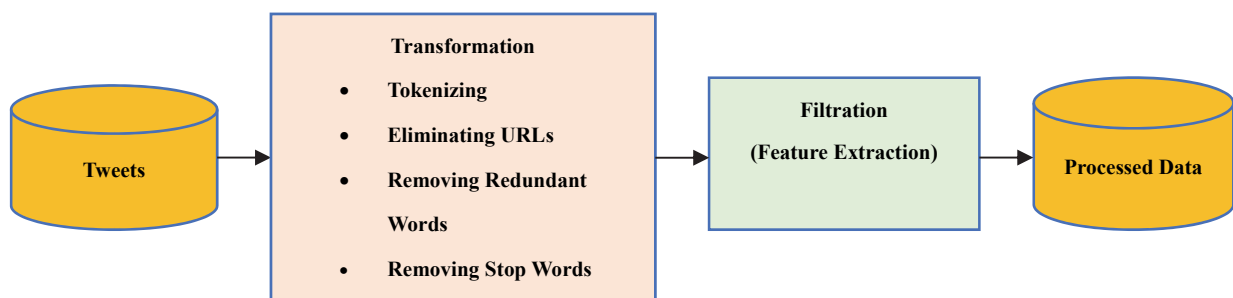


Figure 2. Data pre-processing phase

For the temporal feature, we have selected the timestamp for each tweet. We have analyzed tweet time, and tweet class are not correlated to each other. Therefore, it does not make sense that the tweet time includes the feature set, but in the advanced analysis, we have seen that preceding tweet’s class and current tweet’s class correlate. To implement temporal relation in tweet analysis, we have ordered all tweets by time, and subsequently, we have added the last preceding tweets class name for the current tweet. For the spatial feature, we have selected location id for each tweet instead of location only. However, it is impossible to take the location of the tweets from Twitter for academic purpose, but we can simulate the tweet location as a fixed location id depending on the location of the company and beneficiaries from the company products and services. This perspective assumes that all individual tweet locations have a unique location id, in which the tweet mentioned correlates with the tweet sentiment class.

3.3 Phase Three: Expert Labelling Technique

This study advocates the expert labelling technique to define the polarity (positive, negative and neutral) of data after pre-processing phase. Labelling technique by experts with experiences in the domain of stock market play the primary role in enhancing the polarity result which leads to raising the classification accuracy (Kouloumpis et al., 2011). Table 1 shows an example of the expert labelling technique.

Table 1. Examples of expert labelling and defining the polarity

Tweet	Polarity	Description of Sentiment
Demand declined on dairy products this month.	Negative (2)	This news will affect the related shares negatively.
Low fuel prices.	Positive (1)	This news has a positive effect on different kinds of investment. For example, in the domain of car trading, this will lead to increasing demand of car purchasing.
The USA rap singer felt bored when he was watching the color of the chocolate tray.	Neutral (0)	This news does not have any effect on the shares related to the chocolate companies.

The numbers (0, 1, 2) in Table 2 refer to: 0 as neutral polarity, 1 as positive polarity, and 2 as negative polarity.

3.4 Phase Four: Classification

This study has applied more specifically classification method to identify a suitable pattern in the domain of stock market classification model by hybridizing NBCs. NBCs have four parameter estimators, namely GNB, MNB, BNB, and SSNB. The hybridization of NBCs for this study depends on combining the following three algorithms Baseline Naïve Bayes (NB), MNB, and SSNB. This study has chosen these three classifiers because their characteristics match the requirements. For instance, MNB handling more than one feature at the same time, and this study is already focusing on different types of features like spatial and temporal. Furthermore, SSNB is suitable for small labelled dataset, and this research deals with small dataset includes 3246 tweets. The pseudo code of HNBCs algorithm is presented as follows:-

**HNBCs:**

1: Create a frequency table for all the features in the train set against every single class in every single document. Extra class refers to unknown data.

Supposed known documents feature matrix is shown as  
 $F(i, j, k) \ i = 1.. \#class, j = 1.. \#feature, k = 1.. \#document_k$   
 unknown document feature matrix is shown as  
 $UF(k, j) \ k = 1.. \#document_{uk}, j = 1.. \#feature$

2: Determine the initial weight of every single known and unknown document.

$$W(i, k) = \begin{cases} 1, & \text{if } document_k \in class_i \\ 0, & \text{otherwise} \end{cases}$$

3: Calculate the probability of the features against the classes.

$$\theta(i, j) = \frac{\sum_{k=1}^{\#document_i} W(i, k)F(i, j, k) + \alpha}{\sum_{k=1}^{\#document_i} \sum_{j=1}^{\#feature} W(i, k)F(i, j, k) + \alpha n}$$

$$i = 1.. \#class, j = 1.. \#feature$$

In particular case  $\alpha = 1$  and  $n = \#feature$

4: Calculate each class probability.

$$P_c(i) = \frac{\#document_i}{\#all\ document}, i = 1.. \#class$$

5: Calculate all unknown document class probabilities.

$$P_k(i) = \log(P_c(i)) \cdot \sum_{j=1}^{\#feature} UK(k, j) \cdot \log(\theta(i, j))$$

6: Calculate unknown data weights.

$$W(i, k) = \frac{P_k(i)}{\sum_{j=1}^{\#class} P_k(j)} \text{ where } i = 1 \dots \#class, k = 1.. \#document_{uk}$$

7: Return Step 3 while the maximum iteration is completed or there is no significant difference in unknown data weight.

8: Create a frequency table for all the features for certain test element. Supposed feature matrix is shown as

$$F_t(l) \quad l = 1.. \#feature \text{ of test element}$$

9: Calculate the conditional probabilities for all the classes.

$$P(i) = \log(P_c(i)) \cdot \sum_{l=1}^{\#feature} F_t(l) \cdot \log(\theta(i, l))$$

10: Determine the result which has maximum probability.

$$result = \underset{i}{\operatorname{argmax}}(P(i))$$

### 3.5 Phase Five: Performance and Evaluation

The purpose of empirical evaluation focuses on how to measure the accuracy of HNBCs. Measurements that would be used in this evaluation process to compare the proposed method with baseline NB classifiers includes; Recall, Precision, F-measures, support, and accuracy.

Our equation is given below as:

- Recall = Sensitivity = Total Positive Rate is a proportion of cases that were correctly identified as positive. It is defined as  $[ TP / (TP + FN) ] = [ d / (c + d) ]$ .
- Precision =  $[ TP / (TP + FP) ]$ .
- F1-Measure, is a measure that combines precision and recall =  $2 [ Precision.Recall / Precision + Recall ]$ .
- Support is defined as the number of occurrences of each class.
- Accuracy is defined as the portion or part of the sum total number of classification that is correct. It is given as  $[ (a + d) / (a + b + c + d) ]$  or  $[ (TP + TN) / (TP + FP + FN + TN) ]$ .

## 4. Experiment

This study uses experimental test to generate knowledge from social networks which leads to achieving a stock market classification model with high accuracy and reliability.

### 4.1 Collected Tweets and Pre-Processing

A collection of tweets was compiled for a period of eight months, from 18-September-2016 to 25-May-2017. The data set includes latest 3246 tweets; these tweets related to Al-marai company account on Twitter, which currently produces the best dairy products in the Saudi Arabia and Arab Middle East Countries (Euromonitor, 2016). Table 2 shows a random sample of the collected tweets after pre-processing. All tweets are in the Arabic language.

Table 2. Random Samples of Collected Tweets

Default Location	Timestamp	Tweets (in Arabic)	Tweets After English Translation
8.3173E+17	15/02/2017 05:00	صباح الخير	Good Morning.
7.93317E+17	01/11/2016 04:59	في الوقت الحالي توزيع منتجات المراعي في دول مجلس التعاون الخليجي والأردن ومصر	At present, Almarai products are distributed in GCC, Jordan, and Egypt.
7.78097E+17	20/09/2016 05:02	نأسف لا يمكننا الرد على ملاحظتك دون استلامنا للعينة وتقديمها للفحص الفني	Sorry, we cannot reply now. We should receive the sample first, then only it can be submitted to the technical laboratory.

#### 4.2 Expert-Labeling Technique

After tweets collection and pre-processing, tweets were categorized into three categories: (1: Positive, 2: Negative and 0: Neutral). Classification of Tweets depends on the impact of tweets. If it has a positive significance from the stock market, the tweet is labelled as positive, negative significance as negative, and other tweets are treated as a neutral tweet. This phase was run by two experts; both are specialists in the domain of finance and marketing.

The expert labelling depends on the relationship between tweet text and stock market concepts. Furthermore, after the expert labelling, a word-net is generated. This represents a correlated word with the field of stock market. Table 3 shows the tweets after expert labelling step.

Table 3. Tweets After Expert Labelling

Default Location	Timestamp	Tweets (in Arabic)	Tweets After English Translation	Class
8.3173E+17	15/02/2017 05:00	صباح الخير	Good Morning.	0
7.93317E+17	01/11/2016 04:59	في الوقت الحالي توزيع منتجات المراعي في دول مجلس التعاون الخليجي والأردن ومصر	At present, Almarai products are distributed in GCC, Jordan, and Egypt.	1
7.78097E+17	20/09/2016 05:02	نأسف لا يمكننا الرد على ملاحظتك دون استلامنا للعينة وتقديمها للفحص الفني	Sorry, we cannot reply now. We should receive the sample first, then only it can be submitted to the technical laboratory.	2

#### 4.3 Performance and Evaluation

This section shows the results of Applying the proposed model, benchmark, and discussion.

##### 4.3.1 Results

Table 4 shows the results of the performance and evaluation phase by applying HNBCs with dataset has the following attributes:

- All tweets have labelled by the experts
- Spatial feature = True
- Temporal feature = True

HNBCs Accuracy: 90.38 %.



Table 4. Performance and Evaluation Results of HNBCs

Class	Precision	Recall	F1-Score	Support
0	0.90	0.87	0.88	1081
1	0.92	0.92	0.92	1946
2	0.80	0.90	0.85	219
Average/Total	0.90	0.90	0.90	3246

Table 4 shows the class precision, recall, and F1-score for HNBCs: 90% for the precision, 90% for the recall and 90% for the F1-score.

#### 4.3.2 Benchmark

Table 5 shows the results of the performance and evaluation phase after applying MNB with dataset has the following attributes:

- 10% of tweets have labelled by the experts
- 90% of tweets have labelled by using auto-labelling (lexical based)
- Spatial feature = False
- Temporal feature = False

MNB Accuracy: 82.53 %.

Table 5. Performance and Evaluation Results of MNB

Class	Precision	Recall	F1-Score	Support
0	0.87	0.64	0.73	1081
1	0.81	0.93	0.87	1946
2	0.81	0.85	0.83	219
Average/Total	0.83	0.83	0.82	3246

Table 5 shows the class precision, recall, and F1-score for MNB: 83% for the precision, 83% for the recall and 82% for the F1-score.

#### 4.3.3 Discussion

We can observe that this study has achieved a high accuracy equals 90.38% for HNBCs with all classes (positive, negative and neutral). These results will enable the decision makers and investors in the domain of stock market exchange to make a safe decision with low risk because these results depend on facts regarding the domain of stock market. Facts like the necessary to spatial and temporal features besides the role of the stock market experts in achieving real sentiment analysis. High classification accuracy with real sentiment analysis will lead to generating accurate and reliable reports and indicators on the company's stock. Moreover, the simulation of assumptions in the domain of stock market decision making like the importance of timestamp and location features lead to add more reliability for the classification results.

For example, if the classification model comes with accuracy equal 75%, That means any decision related to this analysis has 25% of inaccuracies. From the simple previous example, we can inference that decision making in any field needs to high accuracy in the classification with simulation matching to the concepts and assumptions of that research domain.

Finally, from the results in table 4 and table 5, we can notice that machine learning methods which using sentiment analysis on Twitter like NB classifiers produce high, real and reliable accuracy by simulating the domain features and preparing the dataset using NLP methods.

### 5. Recommendations and Future Work

We recommend that, spatial and temporal features is necessary for the stock market classification model because it will lead to the increase in the importance and value of generated information besides raising the reliability of the produced reports and indicators. At the same time, expert labelling will reduce the risk of decision making in the domain of stock market exchange because it is working to increase the reality of sentiment analysis which represents a necessary process to the supervised machine learning methods.

For future work, we intend to add more related spatial and temporal features depending on the decision-making concepts and fundamentals of the field of the stock market exchange concurrently improving the automated-expert labelling system by generating an appropriate lexical model which has the ability to update all related keywords and vocabularies.

## 6. Conclusion

The five steps of the model presented in this study starts with data collection, filtration, determination of the polarity according to sentiments of people, classification by enhancing NBCs and ends with the performance and evaluation stage. This model works to increase classification accuracy to help decision-makers in the domain of stock market exchange and the investors who are looking for more investment in the stock market so that they can make more accurate and precise decision. This model also reduces people's apprehension on the reliability of the model. This method has mainly been founded on sentiment analysis approach, by employing expert labelling technique and features, namely, spatial and temporal. This study proposes HNBCs as a machine learning method for stock market classification. A hybrid algorithm has been adapted from NB; these hybrid algorithms incorporate two different NB algorithms based on their specific functionalities.

## References

- Abdelwahab, O., Bahgat, M., Lowrance, C. J., & Elmaghraby, A. (2015, 7-10 Dec. 2015). Effect of training set size on SVM and Naive Bayes for Twitter sentiment analysis. Paper presented at the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).
- Adam, K., Marcet, A., & Nicolini, J. P. (2016). Stock market volatility and learning. *The Journal of Finance*, 71(1), 33-82.
- Aggarwal, C. C., & Zhai, C. (2012). Mining text data: Springer Science & Business Media.
- Ahuja, R., Rastogi, H., Choudhuri, A., & Garg, B. (2015). Stock market forecast using sentiment analysis. Paper presented at the Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on.
- Al-Ayyoub, M., Essa, S. B., & Alsmadi, I. (2015). Lexicon-based sentiment analysis of arabic tweets. *International Journal of Social Network Mining*, 2(2), 101-114.
- Alkubaisi, G. A. A., Kamaruddin, S. S., & Husni, H. (2017). A Systematic Review on The Relationship Between Stock Market Prediction Model Using Sentiment Analysis on Twitter Based on Machine Learning Method and Features Selection. *Journal of Theoretical and Applied Information Technology*, 95(24), 6924-6933.
- Anjaria, M., & Guddeti, R. M. R. (2014, 6-10 Jan. 2014). Influence factor based opinion mining of Twitter data using supervised learning. Paper presented at the 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS).
- Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: detecting influenza epidemics using Twitter. Paper presented at the Proceedings of the conference on empirical methods in natural language processing.
- Arvanitis, K., & Bassiliades, N. (2017). Real-Time Investors' Sentiment Analysis from Newspaper Articles Advances in Combining Intelligent Methods (pp. 1-23): Springer.
- Atefeh, F., & Khreich, W. (2015). A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1), 132-164.
- Attigeri, G. V., MM, M. P., Pai, R. M., & Nayak, A. (2015). Stock market prediction: A big data approach. Paper presented at the TENCON 2015-2015 IEEE Region 10 Conference.
- Bartov, E., Faurel, L., & Mohanram, P. S. (2015). Can Twitter Help Predict Firm-Level Earnings and Stock Returns? Available at SSRN 2631421.
- Bhattu, N., & Somayajulu, D. (2012). Semi-supervised Learning of Naive Bayes Classifier with feature constraints. Paper presented at the 24th International Conference on Computational Linguistics.
- Bishop, C. M. (2006). Pattern recognition. Machine learning, 128, 1-58.
- Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *ICWSM*, 11, 450-453.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.

- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15-21.
- Chandra, S., Khan, L., & Muhaya, F. B. (2011). Estimating twitter user location using social interactions--a content based approach. Paper presented at the Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.
- Chandrasekar, P., & Qian, K. (2016, 10-14 June 2016). The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier. Paper presented at the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC).
- Di Nunzio, G. M., & Sordoni, A. (2012). A visual tool for bayesian data analysis: the impact of smoothing on naive bayes text classifiers. Paper presented at the Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval.
- Euromonitor. (2016). *Dairy in Saudi Arabia*. Retrieved March 22, 2017, from <http://www.euromonitor.com/dairy-in-saudi-arabia/report>
- Fang, X., Hu, P. J. H., Li, Z., & Tsai, W. (2013). Predicting adoption probabilities in social networks. *Information Systems Research*, 24(1), 128-145.
- Fernández-Avilés, G., Montero, J. M., & Orlov, A. G. (2012). Spatial modeling of stock market comovements. *Finance Research Letters*, 9(4), 202-212.
- Fiarni, C., Maharani, H., & Pratama, R. (2016, 25-27 May 2016). Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique. Paper presented at the 2016 4th International Conference on Information and Communication Technology (ICoICT).
- Geser, H. (2011). Has Tweeting become Inevitable? Twitter's strategic role in the World of Digital Communication. *Sociology In Switzerland: Towards Cybersociety and Vireal Social Relations*. Accessed on, 9.
- 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.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1, 12.
- 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. Paper presented at the Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication.
- Hussien, W. A., Tashtoush, Y. M., Al-Ayyoub, M., & Al-Kabi, M. N. (2016). Are emoticons good enough to train emotion classifiers of arabic tweets? Paper presented at the Computer Science and Information Technology (CSIT), 2016 7th International Conference on.
- Iacomin, R. (2016). Feature optimization on stock market predictor. Paper presented at the Development and Application Systems (DAS), 2016 International Conference on.
- Islam, M. J., Wu, Q. J., Ahmadi, M., & Sid-Ahmed, M. A. (2007). Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers. Paper presented at the Convergence Information Technology, 2007. International Conference on.
- Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338-345.
- Jha, N., & Mahmoud, A. (2017). Mining user requirements from application store reviews using frame semantics. Paper presented at the International Working Conference on Requirements Engineering: Foundation for Software Quality.
- Jiang, L., Cai, Z., Zhang, H., & Wang, D. (2013). Naive Bayes text classifiers: a locally weighted learning approach. *Journal of Experimental & Theoretical Artificial Intelligence*, 25(2), 273-286.
- Jiang, L., Wang, D., Cai, Z., & Yan, X. (2007). Survey of improving naive Bayes for classification. Paper presented at the International Conference on Advanced Data Mining and Applications.
- Kharde, V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. arXiv preprint

- arXiv:1601.06971.
- 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).
- Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! *ICWSM*, 11(538-541), 164.
- Larsen, J. I. (2010). Predicting stock prices using technical analysis and machine learning.
- Lee, J., & Kim, D.-W. (2017). SCLS: Multi-label feature selection based on scalable criterion for large label set. *Pattern Recognition*.
- Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. C. (2012). Tedas: A twitter-based event detection and analysis system. Paper presented at the 2012 IEEE 28th International Conference on Data Engineering.
- Lin, J., & Ryaboy, D. (2013). Scaling big data mining infrastructure: the twitter experience. *ACM SIGKDD Explorations Newsletter*, 14(2), 6-19.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, 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.
- Makice, K. (2009). Twitter API: Up and running: Learn how to build applications with the Twitter API: "O'Reilly Media, Inc."
- Makrehchi, M., Shah, S., & Liao, W. (2013). Stock prediction using event-based sentiment analysis. Paper presented at the Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on.
- Meesad, P., & Li, J. (2014, 8-11 Dec. 2014). Stock trend prediction relying on text mining and sentiment analysis with tweets. Paper presented at the Information and Communication Technologies (WICT), 2014 Fourth World Congress on.
- Miranda, F., & Abreu, C. (2015). Handbook of Research on Computational Simulation and Modeling in Engineering.
- Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.
- Nagwani, N. K., & Verma, S. (2014). A Comparative Study of Bug Classification Algorithms. *International Journal of Software Engineering and Knowledge Engineering*, 24(01), 111-138.
- 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.
- Netti, K., & Radhika, Y. (2015, 10-12 Dec. 2015). A novel method for minimizing loss of accuracy in Naive Bayes classifier. Paper presented at the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC).
- 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.
- Nguyen, T. T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys & Tutorials*, 10(4), 56-76. <https://doi.org/10.1109/SURV.2008.080406>
- Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis-An Assessment of Peoples' Belief: A Survey. *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, 4(1), 21.
- Ruth, M., & Hannon, B. (2012). Positive Feedback in the Economy Modeling Dynamic Economic Systems (pp. 71-76): Springer.
- Sathyadevan, S., Sarath, P. R., Athira, U., & Anjana, V. (2014, 26-28 Aug. 2014). Improved document classification through enhanced Naive Bayes algorithm. Paper presented at the Data Science & Engineering (ICDSE), 2014 International Conference on.
- Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news:

- The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 12.
- Shah, R. R., Yu, Y., Verma, A., Tang, S., Shaikh, A. D., & Zimmermann, R. (2016). Leveraging multimodal information for event summarization and concept-level sentiment analysis. *Knowledge-Based Systems*, 108, 102-109.
- Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011). Analyzing tourism information on twitter for a local city. Paper presented at the Software and Network Engineering (SSNE), 2011 First ACIS International Symposium on.
- Skuza, M., & Romanowski, A. (2015). Sentiment analysis of Twitter data within big data distributed environment for stock prediction. Paper presented at the Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on.
- Song, Z., & Xia, J. C. (2016). Spatial and Temporal Sentiment Analysis of Twitter data. *European Handbook of Crowdsourced Geographic Information*, 205.
- Su, J., Shirab, J. S., & Matwin, S. (2011). Large scale text classification using semi-supervised multinomial naive bayes. Paper presented at the Proceedings of the 28th International Conference on Machine Learning (ICML-11).
- Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, 34(4), 2622-2629.
- 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.
- Thompson, C. (2008). Brave new world of digital intimacy. *The New York Times*, 7.
- Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6).
- Wang, F., Zhang, Y., Xiao, H., Kuang, L., & Lai, Y. (2015). Enhancing Stock Price Prediction with a Hybrid Approach Based Extreme Learning Machine. Paper presented at the 2015 IEEE International Conference on Data Mining Workshop (ICDMW).
- Wang, H., Wu, J., Zhang, P., & Zhang, C. (2016). Temporal Feature Selection on Networked Time Series. arXiv preprint arXiv:1612.06856.
- Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from twitter posts. Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction.
- Xu, F., & Keelj, V. (2014, 14-17 July 2014). Collective Sentiment Mining of Microblogs in 24-Hour Stock Price Movement Prediction. Paper presented at the 2014 IEEE 16th Conference on Business Informatics.
- Yamamoto, Y. (2014). Twitter4J-A java library for the twitter API: sep.
- Yan, D., Zhou, G., Zhao, X., Tian, Y., & Yang, F. (2016). Predicting stock using microblog moods. *China Communications*, 13(8), 244-257. <https://doi.org/10.1109/CC.2016.7563727>
- Yang, A., Zhang, J., Pan, L., & Xiang, Y. (2015, 16-18 Nov. 2015). Enhanced Twitter Sentiment Analysis by Using Feature Selection and Combination. Paper presented at the Security and Privacy in Social Networks and Big Data (SocialSec), 2015 International Symposium on.
- Yousef, M., Saçar Demirci, M. D., Khalifa, W., & Allmer, J. (2016). Feature Selection Has a Large Impact on One-Class Classification Accuracy for MicroRNAs in Plants. *Advances in bioinformatics*, 2016.
- Zhang, L. (2013). Sentiment analysis on Twitter with stock price and significant keyword correlation.
- Zhou, L., Zhang, P., & Zimmerman, H. (2011). Call for papers for a series of special issues: Social commerce. *Electronic Commerce Research and Applications*.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).