Comparative Analysis of Bayesian Approaches and Variant Methods in the Financial Field

Yilin Chen

Department of Statistics, University of Michigan-Ann Arbor, Ann Arbor, United States of America

cyilin@umich.edu

Abstract. It is important for researchers to select the most appropriate model for specific financial tasks. This comparative analysis describes the strengths, limitations, and trade-offs of Bayesian approaches and variant methods. Moreover, this study aims to compare Bayesian approaches and their variant methods in the financial field and assess their effectiveness and relevancy for stock price prediction. This study starts with a brief overview of Bayes' Theorem, Naïve Bayes (NB), multinomial Naïve Bayes (NBM), and Gaussian Naïve Bayes (GNB). In the prediction of the Brazilian stock market, NBM demonstrates superior performance compared to other models in terms of accuracy, recall, and F1-score. In the meantime, the GNB and linear discriminant analysis-based combined model (GNB_LDA) exhibits superior performance in accuracy and F1-score, while the model incorporates GNB, standardization, and factor analysis (GNB_Z-Score_FA) excels in mean specificity in the prediction of seven different stocks. This study also suggests further exploration of hybrid models that combine the strengths of Bayesian models with other techniques to enhance financial analysis and decision-making.

Keywords: Component, Bayes' theorem, Multinomial Naïve Bayes, Gaussian Naïve Bayes.

1. Introduction

Bayesian statistics and modeling emerges as useful tool in the field of data analysis and provides a powerful framework for inference and decision-making under uncertainty based on Bayes' Theorem. Unlike traditional frequentist statistics, Bayesian statistics do not rely on hypothesis tests. It combines the quantification of uncertainty as well as prior information into the analysis. Numerous fields of science employ Bayesian inference, such as finance and behavioral science.

The cornerstone of Bayesian statistics was first laid out in an article by Reverend Thomas Bayes on inverse probability and then published by Richard Price. [1]. This essay describes how to predict the likelihood of a future occurrence simply based on previous experiences. The famous theorem currently known as Bayes' theorem was not published until 1825 by Pierre Simon Laplace [2]. In mathematics, the concept of Bayes' theorem has been well-established for a long time; they just became popular in practical statistics in the last 50 years. Nowadays, modern Bayesian practitioners have access to a variety of information and methods that enable the development of customized models and computational methods for specific issues [3]. As a useful tool, Bayesian models have many advantages, such as being suitable for small sample sizes and integrating information from multiple sources with varying levels of accuracy [4, 5]. Secondly, they may be used with other Bayesian analytic techniques like Markov chain Monte Carlo (MCMC) [6]. Thirdly, Bayesian models are also feasible to utilize data in order to build model structures determined by the specific subject and conditional probability distributions. They are solved analytically, providing quick replies to queries and a natural approach to dealing with missing data. Additionally, the compiled version of a Bayesian model offers a mechanism to minimize data overfitting [7].

Although Bayesian statistics and modeling are indispensable tools, some recent artificial intelligence studies show that Bayesian modeling is facing some ongoing challenges. First, the sophistication of financial data frequently necessitates complicated models, which results in computationally intensive Bayesian calculations. Additionally, Bayesian models can also be subjective and biased since they include prior assumptions and experts. Dimensionality is another issue, as financial models often involve high-dimensional data, which can lead to computational inefficiency if Bayesian models are used. This paper aims to equip researchers with a more profound understanding of the principles, applications, and challenges associated with Bayesian Statistics and modeling in finance. The paper begins with an introduction to the fundamental principles of Bayesian statistics. This is followed by explaining how it differs from some of its variants. Next, some applications in finance are discussed and compared. In addition, future directions and emerging trends in Bayesian statistics and modeling are explored. This review is intended to inspire further advancements and encourage the adoption of Bayesian methods across diverse fields.

2. Methodology

2.1. Overview

1) Bayes' Theorem and Naïve Bayes: The basic concept of Bayesian statistics and modeling is related to conditional probabilities, which compute the chance of event X, known as the chance of another event or event Y [8]. This concept is generally termed as Bayes theorem and is applied in many research fields to obtain information on the likelihood of an event, given the information of another one. It is given by:

The basic concept of Bayesian statistics and modeling is related to conditional probabilities, which compute the chance of event X, known as the chance of another event or event Y [8]. This concept is generally termed as Bayes theorem and is applied in many research fields to obtain information on the likelihood of an event, given the information of another one. It is given by:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(X)}. \quad (1)$$

In the given equation, $P(X|Y)$ represents the conditional probability of event X known event Y, $P(Y|X)$ represents the conditional probability of event Y known event X. $P(X)$ represents the probability of event X [8].

Van de Schoot et al. suggested that Bayesian modeling is a systematic process and requires a step-by-step assessment of the problem, event likelihood, and the evaluation of the outcome or the results [8]. In this first step of 'Prior,' the background information regarding the problem of interest is derived, and then a suitable methodological framework, including hypotheses, is formulated. Then likelihood of the events associated with the problem of interest is derived, followed by the posterior stage, whereby the suggested methodologies are applied to obtain the desired outcomes. Finally, posterior inferences are drawn from the analysis, and hypotheses are validated, which in turn adds to the existing knowledge on the topic. Overall, this process allows us to determine the influence of new knowledge on what is already known about a certain phenomenon. This workflow is roughly presented in Fig. 1.

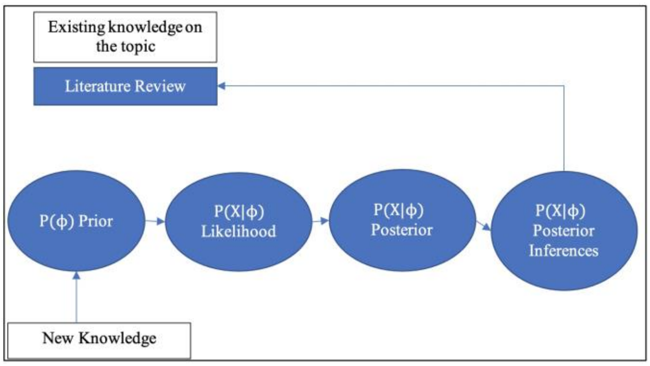

Figure 1. Bayesian Modeling Process

Naïve Bayes (NB) is an algorithm that applies Bayes' Theorem to perform classification tasks in machine learning. NB takes the assumption of the independence of each feature used for classification [8]. This is known as the "naive" assumption, and it also simplifies the calculation. Despite this assumption, NB often performs well in practice, especially when the features are reasonably independent or when there is an enormous amount of data.

2) Multinomial Naïve Bayes Model: One of NB variant models that works well with multinomially distributed data is multinomial Naïve Bayes model (NBM). In this model, the frequencies of specific events are produced by a multinomial distribution. Hence, it is often used in text categorization and frequently used as a starting point in text classification [9]. Additionally, it is also a strong competitor of support vector machine (SVM) if the data are appropriately preprocessed [9].

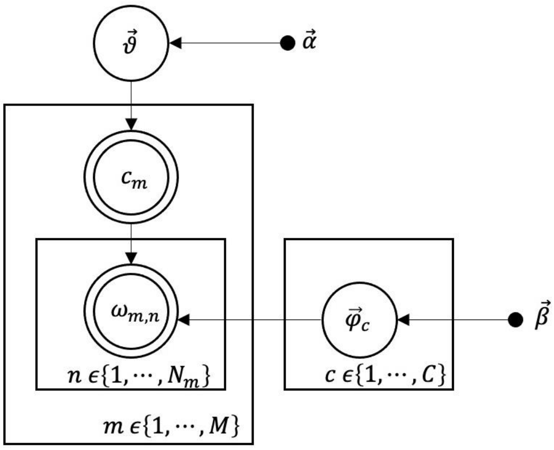

Figure 2. Processes of NBM Model [9]

Fig. 2 displays the Bayesian multinomial graphic model NB. As a generative process, the process of each step is described below:

- Create a multinomial vector $\vec{\vartheta}$ for Dirichlet $\vec{\alpha}$;
- For classes c, c=1… C, create a multinomial vector $\vec{\varphi}_c$ for Dirichlet $\vec{\beta}$;
- For documents m, m=1…M, create a class c_m for multinomial vector $\vec{\vartheta}$;
- For every single word n in document m, n=1…N_m, create a word $w_{m,n}$ for multinomial vector $\vec{\varphi}_{c_m}$.
- The maximum a posteriori (MAP) estimates, also known as the $\max p$ of $\vec{\vartheta}$ and $\vec{\varphi}_c$ after the whole document is trained by the model, would be:

$$\hat{\vec{\vartheta}} = \arg\max \text{Dirichlet}\left(\vec{\vartheta} \mid \vec{n} + \vec{\alpha}\right) \prod_{c=1}^{C} \text{Dirichlet}\left(\vec{\varphi}_c \mid \vec{n}_c + \vec{\beta}\right), \quad (2)$$

Where $\hat{\vec{\vartheta}}$ is the MAP estimates of $\vec{\vartheta}$ and $\vec{\varphi}_c$, D is a set of documents that are trained in this model, and $\vec{n}_c$ is a multinomial vector of every word in class c.

- The MAP parameter estimates of $\vec{\vartheta}$ and $\vec{\varphi}_c$ following a Dirichlet distribution can be further expressed as:

$$\hat{\vartheta}_c = \frac{n^{(c)} + \alpha_c - 1}{M + \sum_{c'=1}^{C} (\alpha_{c'} - 1)}, \quad (3)$$

$$\hat{\varphi}_{c,v} = \frac{n_c^{(v)} + \beta_v - 1}{n_c + \sum_{v'=1}^{V} (\beta_{v'} - 1)}, \quad (4)$$

Where v is the number of unique words in D, v=1,…,V, $\hat{\vartheta}_c$ is the MAP parameter estimate of $\vec{\vartheta}$, $\hat{\varphi}_{c,v}$ is the MAP parameter estimate of $\vec{\varphi}_c$. n_c is the number of words in class c, $n^{(c)}$ is the total document counts associated with each class in D, and $n_c^{(v)}$ represents the number of times of word v occurs within each class in D, α_c=2c, and β_v=2v.

The complete Bayesian inference should be:

$$c = \arg\max \frac{\alpha_c + n^{(c)}}{\sum_{c'=1}^{C} \alpha_{c'} + M} \frac{\Gamma\left(\sum_{v=1}^{V} \beta_v + n_c\right)}{\prod_{v=1}^{V} \Gamma\left(\beta_v + n_c^{(v)}\right)} \frac{\prod_{v=1}^{V} \Gamma\left(\beta_v + n_c^{(v)} + \tilde{n}^{(v)}\right)}{\Gamma\left(\sum_{v=1}^{V} \beta_v + n_c + \tilde{n}\right)}. \tag{5}$$

Note that $\Gamma(\cdot)$ represents the Gamma function, and $I(\cdot)$ is the indicator function, whereas $\tilde{n}^{(v)}$ is the frequency of appearance of any word v in the document given, and $\vec{w}$ is the given document.

3) Gaussian Naïve Bayes Model: Gaussian Naïve Bayes (GNB) is a classification method that relies on a probabilistic approach and the assumption of a Gaussian (normal) distribution. The normal distribution is a statistical model which characterizes continuous random variables in nature. It is bell-shaped, and its parameters are mean μ and standard deviation σ. The mean signifies the average value of distribution, while the standard deviation measures its spread around the mean [10]. GNB assumes that each feature or predictor has its own independent predictive power for the output variable [10]. By combining the predictions from all features, GNB generates the final prediction, which represents the probability of classifying the dependent variable into different groups. The final classification is eventually given to the group with the highest likelihood. Unlike NBM, the GNB model assumes that the information provided is continuous rather than discrete, and the likelihood of an event can be evaluated by a normal distribution [10]. Thus, conditional probability in GNB model is given by the following:

$$P(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}, \tag{6}$$

Where $P(a)$ represents the chance of event x occurring, $a=1 \dots A$, μ is the variance of a, and σ^2 is the variance of a.

2.2. Relevant Improvement Strategies

Some related improvement strategies of Bayesian models were created. They aim to enhance the performance of Bayesian models, make them more applicable and address the challenges associated with complex real-world problems. Semi-supervised learning algorithms and fine-tune model parameters are two typical improvement strategies [11, 12].

1) Semi-Supervised Learning Models: Semi-supervised learning models help increase the accuracy of NBM model. Moreover, it improves the classification quality by training the model on the classified corpus data by a NBM model. In this algorithm, a new sample set is generated, say S, which is then appended to the class of documents or the corpus generated initially D [11]. The model classifies each text into a specific class of documents C using the conditional probability of some class C given the sample set S, which is equivalent to $P(C|S)$ [11]. Then a semi-supervised learning algorithm is applied to the existing corpus of classified documents. Once these documents are appended, this algorithm trains the dataset and then returns the results back to the original model [11].

2) Fine Tuning Model Parameters: Fine tuning of parameters in a model which assumes that all observations are independent can enhance the accuracy and prediction score of the model. Ozkaya suggested that fine tuning can be computed using the following formula to enhance the accuracy of GNB models [12].

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN}, \tag{7}$$

Where TP refers to the situation when a positive condition is accurately recognized as positive, while TN indicates the correct identification of a negative condition as negative. Similarly, FP refers to the scenario where a negative condition is mistakenly identified as positive, and FN represents the misidentification of a positive condition as negative in a classification test.

2.3. Application and Discussion

1) The Study of the Brazilian Stock Market: In this study, lexicon-based and machine learning-based, are evaluated and compared for their performance in predicting Brazilian stock market datasets. After pre-processing the data by removing accents and punctuation marks, filtering stop-words and numbers, different techniques such as lemma extraction and stemming are applied [13]. Specifically, the models are represented by the first string in each row NB, and NBM, and the pre-processing methods are represented by the words that follow [13]. According to TABLE I, NBM achieves the highest F1-score of 0.778, surpassing the second highest F1-score obtained by the NBM Lemma model by 1.83%. Additionally, NBM demonstrates the highest recall and precision values of 0.779 and 0.777, respectively, which are at least 1.69% and 1.83% higher than the other models. These findings indicate the superior performance of NBM in accurately predicting stock market data. However, despite the highly positive outcomes, there remains scope for further improvement.

Table 1. The Precision, Recall, and F1-Score for NBM Models [13]

Model	Precision	Recall	F1-score
NB	0.747	0.74	0.742
NBM	0.779	0.777	0.778
NB Lemma	0.743	0.74	0.741
NBM Lemma	0.764	0.764	0.764
NB Stemming	0.736	0.733	0.734
NBM Stemming	0.751	0.753	0.752
NB Lemma Verb	0.669	0.669	0.669
NBM Lemma Verb	0.69	0.694	0.686
NB Title	0.699	0.702	0.696
NBM Title	0.736	0.724	0.726
NB Summarization	0.711	0.713	0.712
NBM Summarization	0.759	0.759	0.753

2) The Study of 7 Different Stocks on Yahoo Finance:

Table 2. Accuracy Score of Different GNB Models [14]

Datasets	GNB	GNB_Z-Score	GNB_Min-Max	GNB_PCA	GNB_Z-Score_PCA
Apple	0.5361	0.6241	0.6241	0.5342	0.6444
Abbot	0.5713	0.6954	0.6954	0.5361	0.8639
CarMax	0.5583	0.6982	0.6982	0.5111	0.8509
S&P 500	0.5722	0.6704	0.6704	0.5472	0.7444
Tata Steel	0.5461	0.7232	0.7232	0.5012	0.8713
Hindustan Petroleum	0.5197	0.6085	0.6085	0.5126	0.6953
Bank of America	0.5472	0.7111	0.7111	0.5056	0.8333
Mean	0.5501	0.6758	0.6758	0.5211	0.7862
Datasets	**GNB_Min-Max_PCA**	**GNB_FA**	**GNB_Z-Score_FA**	**GNB_Min-Max_FA**	**GNB_LDA**
Apple	0.7241	0.6370	0.7111	0.7314	0.8769
Abbot	0.8398	0.8407	0.8462	0.8528	0.8861
CarMax	0.8648	0.7963	0.7907	0.8176	0.8870
S&P 500	0.7546	0.4537	0.7269	0.7583	0.8259
Tata Steel	0.8809	0.8701	0.8616	0.8637	0.9142
Hindustan Petroleum	0.7548	0.5832	0.6700	0.6700	0.9092
Bank of America	0.8435	0.8509	0.8407	0.8454	0.8713
Mean	0.8089	0.7188	0.7782	0.7913	0.8815

This study examines the effectiveness of the GNB algorithm in forecasting stock prices. The raw data are preprocessed, then divided the data into training and test sets using random sampling from three separate stock markets [14]. The outcomes, shown in TABLE II, indicate that the model combined the GNB with linear discriminant analysis (GNB_LDA), achieves the highest accuracy of 0.9142 on the Tata Steel dataset. On the other hand, the model that combined the GNB with principal component analysis (GNB_PCA) has the lowest accuracy value of 0.5012. The GNB_LDA model again stands out with the highest mean accuracy value of 0.8815. Conversely, the GNB_PCA model yields the lowest mean accuracy value of 0.5211, suggesting lower overall prediction accuracy.

The measured F1-scores results in TABLE III. demonstrate that the GNB_LDA model outperforms all models in terms of F1-score. The highest F1-score recorded among all models is 0.9167, which is achieved by the GNB_LDA model. In contrast, the lowest F1-score recorded is 0.4166, produced by the GNB_PCA model. The best and worst F1-score are both produced on the Tata Steel dataset. For mean F1-scores, the GNB_LDA model has the highest mean F1-score of 0.8843 among all models.

TABLE IV shows the specificity results for the various stock data sets that are examined, shown as the model incorporates GNB, standardization, and factor analysis (GNB_Z-Score_FA) has the greatest mean specificity. Specifically, the model incorporates GNB, and factor analysis (GNB_FA) achieves the highest specificity value of 0.9800 on the S&P 500 dataset. The GNB_PCA model has the lowest specificity value of 0.1030 on the Abbot dataset. Similarly, the GNB_LDA model displays the highest average specificity of 0.8921, while the GNB_PCA model demonstrates the lowest average specificity of 0.2532 within the GNB models considered. Overall, this study concludes that the GNB algorithm may accurately forecast stock values when it combines with the right methods [14].

Table 3. F1-score of Different GNB Models [14]

Datasets	**GNB**	**GNB_Z-Score**	**GNB_Min-Max**	**GNB_PCA**	**GNB_Z-Score_PCA**
Apple	0.6962	0.5365	0.5365	0.6964	0.5362
Abbot	0.6662	0.6803	0.6803	0.6980	0.8740
CarMax	0.6175	0.6766	0.6766	0.5629	0.8540
S&P 500	0.6583	0.6139	0.6139	0.7074	0.7058
Tata Steel	0.4860	0.7461	0.7461	0.4166	0.8747
Hindustan Petroleum	0.6161	0.7074	0.7074	0.6777	0.7659
Bank of America	0.6304	0.7342	0.7342	0.6454	0.8454
Mean	**0.6244**	**0.6707**	**0.6707**	**0.6292**	**0.7794**
Datasets	**GNB_Min-Max_PCA**	**GNB_FA**	**GNB_Z-Score_FA**	**GNB_Min-Max_FA**	**GNB_LDA**
Apple	0.6740	0.5220	0.6494	0.6875	0.8783
Abbot	0.8443	0.8590	0.8534	0.8635	0.8976
CarMax	0.8653	0.7835	0.7839	0.8108	0.8891
S&P 500	0.7237	0.3114	0.7053	0.7398	0.8175
Tata Steel	0.8760	0.8762	0.8626	0.8648	0.9167
Hindustan Petroleum	0.7990	0.7031	0.7532	0.7532	0.9119
Bank of America	0.8516	0.8546	0.8473	0.8542	0.8790
Mean	**0.8048**	**0.7014**	**0.7793**	**0.7963**	**0.8843**

Table 4. Specificity Score of Different GNB Models [14]

Datasets	GNB	GNB_Z-Score	GNB_Min-Max	GNB_PCA	GNB_Z-Score_PCA
Apple	0.1099	0.8728	0.8728	0.1200	0.9423
Abbot	0.3094	0.8004	0.8004	0.1030	0.8443
CarMax	0.4184	0.7927	0.7927	0.4069	0.8599
S&P 500	0.3538	0.9018	0.9018	0.2131	0.9672
Tata Steel	0.6717	0.6413	0.6413	0.6544	0.8544
Hindustan Petroleum	0.2754	0.2774	0.2774	0.1621	0.4037
Bank of America	0.3283	0.6359	0.6359	0.1132	0.7698
Mean	0.3524	0.7032	0.7032	0.2532	0.8059
Datasets	**GNB_Min-Max_PCA**	**GNB_FA**	**GNB_Z-Score_FA**	**GNB_Min-Max_FA**	**GNB_LDA**
Apple	0.9423	0.9424	0.9523	0.9363	0.9284
Abbot	0.8743	0.7665	0.8603	0.8343	0.8343
CarMax	0.8925	0.8868	0.8522	0.8848	0.9002
S&P 500	0.9571	0.9800	0.8834	0.9161	0.9632
Tata Steel	0.9326	0.8326	0.8652	0.8674	0.8957
Hindustan Petroleum	0.5487	0.1843	0.3416	0.3416	0.9006
Bank of America	0.8038	0.8415	0.8132	0.8000	0.8226
Mean	0.8502	0.7763	0.7955	0.7972	0.8921

3) Comparison of Bayesian Models and Neural Networks: Advantages and Disadvantages: Bayesian models are discovered to be helpful for a considerable variety of applications. However, when compared to other machine learning techniques, such as neural networks, Bayesian models offer several distinct advantages and disadvantages. The most significant advantage is that Bayesian models are often easier compared to neural networks, which must make numerous judgments on hidden layers and variants [8]. In comparison to a neural network which is virtually unreadable by humans, Bayes Networks are advantageous to both humans and computers because of their graphical appearance [8]. The most significant disadvantage is Bayesian models presume that all input variables are independent. The accuracy of the Bayes classifier may be affected if that assumption is violated [8]. The correlation and dependency between the input variables may be handled by a neural network with the right network configuration [8].

4) Future Directions of Bayesian Modeling in the Finance Area: Bayesian models have a bright future as a thriving area of research. Portfolio optimization, risk management, and financial forecasting are some of the up-to-the-moment research topics in the finance area. To improve financial decision-making, researchers might concentrate on creating new Bayesian models or hybrid models in the future. For example, in portfolio optimization tactics, Bayesian models may include uncertainty and expert viewpoints. Such operation helps enhance risk management by precisely calculating risk measures and giving probabilistic projections for financial variables. Bayesian models can aid in more reliable financial analysis and decision-making by overcoming these issues.

3. Conclusion

This comparative analysis aims to provide valuable insights into the effectiveness and applicability of Bayesian models and their variants for financial tasks. This analysis provides an overview of Bayes' Theorem, NB, NBM, and GNB, two relevant improvement strategies, and two applications. In the prediction of the Brazilian stock market, NB, NBM, along with four different preprocessing methods, 12 models in total are proposed. Whereas in the prediction of 7 different stocks on Yahoo Finance, GNB model and 7 of its variant models are proposed. The results are recorded and compared, and some relevant improvement strategies are also given. Comparatively, in the prediction of the Brazilian stock market, results indicate that the most optimistic accuracy, recall, and F1-score are provided by NBM. NBM demonstrates the highest F1-score, precision, and recall. In the prediction of 7 different stocks on Yahoo Finance, the GNB_LDA model performs the highest accuracy score, and F1-score. Additionally, the GNB_LDA model also has the greatest mean accuracy score, and mean F1-score. On the other hand, GNB_Z-Score_FA has the highest specificity and mean specificity. Overall, the choice between different Bayesian approaches in the financial field depends on the specific problem, available data, and the degree of interpretability required. In the future, research should continue to explore hybrid approaches. By combining the strengths of Bayesian models with other techniques, researchers can enhance financial analysis and decision-making processes.

References

[1] T. Bayes, "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S," Philosophical transactions of the Royal Society of London, vol. 53, pp. 370-418.

[2] P. S. Laplace, "Pierre-Simon Laplace philosophical essay on probabilities: translated from the fifth french edition of 1825 with notes by the translator," Springer Science & Business Media, vol. 13, 1998.

[3] D. Ashby, "Bayesian statistics in medicine: a 25 year review," Statistics in medicine, vol. 25, 2006, pp. 3589-3631.

[4] P. Myllymäki, "B-course: A web-based tool for Bayesian and causal data analysis," International Journal on Artificial Intelligence Tools, vol. 11, 2002, pp. 369-387.

[5] B. G. Marcot, R. S. Holthausen, M. G. Raphael, "Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement," Forest ecology and management, vol. 153, 2001, pp. 29-42.

[6] S. Kuikka, V. Olli, "Uncertainties of climatic change impacts in Finnish watersheds: a Bayesian network analysis of expert knowledge," Boreal Environment Research, vol. 2, 1997.

[7] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modeling," Ecological modeling, vol. 203, 2007, pp. 312-318.

[8] S. McCloskey, "Probabilistic reasoning and Bayesian networks," Proc. Neural Netw. Mach. Learn. (ICSG), 2000.

[9] S. Xu, L. Yan, W. Zheng, "Bayesian multinomial Naïve Bayes classifier to text classification," Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech, vol. 11, 2017.

[10] A. H. Jahromi, T. Mohammad, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," 2017 Artificial Intelligence and signal processing conference (AISP), 2017.

[11] Y. Reddy, P. Viswanath, B. E. Reddy, "Semi-supervised learning: A brief review," Int. J. Eng. Technol, vol. 7, 2018, p. 81.

[12] U. Ozkaya, S. Levent, "Fine-tuning models comparisons on garbage classification for recyclability," Unpublished.

[13] B. Neuenschwander, "Sentiment analysis for streams of web data: A case study of Brazilian financial markets," Proceedings of the 20th Brazilian Symposium on Multimedia and the Web, 2014.

[14] E. K. Ampomah, G. Nyame, Z. Qin, "Stock market prediction with gaussian naïve bayes machine learning algorithm," Informatica, vol. 45, 2021.