Comparison of Data Mining Classification Techniques Naïve Bayes, J48, IBK And SMO Using Commercial Dataset

D.O.I - 10.51201/12483

https://doi.org/10.51201/12483

Written by

1- Yasser Mahmood Hussain

General directorate of education in Al-Anbar- Iraq

2- Dr. Murtada Khalafallah Elbashir

Jouf University, Sakaka, KSA

3- Dr. Mohammed Babiker Ali Mohammed

Dean of faculty of computer science and information technology – University of Al Butana

Abstract:

Commercial and mall datasets need to using the most known data mining techniques to classify goods and to predict the user requirements , behavior and company strategies. Classification applied to the financial dataset used which is called Wholesale customers dataset which can be collected by any other company or mall be filling the attributes with the information required which is contains 440 instances of 8 attributes, it refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories.

In this paper many classification algorithm is been tested with the dataset used and compared by using many characteristics to explain the best algorithm used for such data classification.

In order to classify or predict commercial data many algorithms used by companies and malls, these algorithms helps the owners and management of these companies for digging the data and analyze it and then extract meaningful data which can be used to take decisions related to it.

Four data mining classification algorithms is been used Naïve Bayes, J48, IBK and SMO and results is discussed in details.

Keywords - Data Mining, Classification,Naïve Bayes, J48, IBK, SMO

1. **Introduction**:

Data mining techniques are able to classify the data of a wide variety of data sets based on their importance. Important data are those that will be accessed in the near future and those most frequently used. This process allows us to migrate the application' future data requests in the hybrid storage system levels based on the importance [1]. This will reduce the application execution elapsed time. I/O intensive application and their users vary in their access patterns. This makes the process of data classification depends on the application type. Application types include, but not limited to, database applications, image processing applications, marketing, finance, voice, and text recognition ... etc. Different classification or data mining approach may provide better accuracy for a type of an application than the others [2].

In this study a comparison of four known classifier is done to choose the better classifier to the commercial data and whole sale dataset.

2. Theoretical background:

Theoretical study of the classification algorithms and how it is applied and calculated is important to understand the effectiveness of these algorithms on the datasets tested

2.1 Naïve Bayes:

The classification of Naïve Bayes is very scalable because it requires a number of parameters that are linear variables in the problem of learning [3]. However, the Maximum-likelihood can be carried out by the evaluation of a closed-form expression, that usually includes linear time instead of expensive iterative approximation that is helpful for the other types of the classifiers [4].

Naïve Bayes model is accessible for the data built up very large datasets. However, the data has the specifications of simplicity, along with the ability to perform the highly refined methods of classification [5].

Bayes theorem is important in providing the way for the calculation of posterior probability P (c|x) from P(c), P(x) and P (x|c) as given by the equation [6]:

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

From the equation as given above:

- $P(c)$ is the prior possibility of *class*.
- $P\ (c|x)$ is the later possibility of *class* given *predictor*.
- $P(x)$ is preceding possibility of *predictor*
- $P\ (x|c)$ is probability which is the likelihood of *predictor* given *class*.

2.2 J48 Decision Tree:

The decision tree incorporates predictive machine learning and by basing on different attribute values from the given available data it is helpful to decide the targeted value (dependent variable) [6].Moreover, different attributes within the decision tree can be denoted through internal nodes, the possible values of these attributes within the samples being observed can be studied in various branches in between the nodes. However, the function of the terminal nodes is to guide about the ending values (classification) of the dependent variables [7].

The predictable attribute is called as the dependent variable as its value is highly dependent upon the values of all the other attributes. Moreover, prediction of dependent variable's value can be carried out by other variables which are basically called as the independent values within the data set [8].

Moreover, a simple algorithm is followed by J48 Decision tree classifier. For the classification to be carried out for the new item; there must be a need for the creation of a decision tree that is based upon the attributed values of the available data for training. However, when there is a need to encounter the data for training, it efficiently identifies the attributes that can clearly discriminate various such instances [9]. Moreover, this feature is very important as it is helpful in telling the most about the data instances for the classification and gaining of the highest information [10]. Although, if there will be any value within its category for which the data instances that have no ambiguity from among the possible values of this feature have same values for the targeted variables then, in this case, there is a need to terminate the branch and then assign it to the obtained targeted value [11].

2.3 SMO

Sequential Minimal Optimization (SMO) is an algorithm which is used for resolving issues such as Quadratic Programming (QP) which ascends whilst the support vector machines are trained. SMO is considered to be an interactive algorithm used to resolve issues related to optimization as mentioned earlier. The algorithm provides a solution to this problem by breaking the issue into several series of sub-problems which can later be solved analytically. Due to the linear equality constraint containing. [12]

The Lagrange multipliers αi , the slightest possible problem includes two of those multipliers. Further on, for any two multipliers $\alpha 1$ and $\alpha 2$, the constraints are then reduced to:

$$0 \leq \alpha_1, \alpha_2 \leq C,$$
$$y_1 \alpha_1 + y_2 \alpha_2 = k,$$

This problem can then be resolved analytically whereby one requires a minimum of a 1D quadratic function. The negative sum is K which is over the other terms in the equality constraint that is fixed in every iteration. [13]

2.4 IBK

For the classification or regression, a non-parametric method for the recognition of the pattern includes the ***k*-nearest neighbor's algorithm** (***k*-NN**) [14]. However, in all two instants, input consisted of the training examples of the *k* closet within feature space [15]. However, yield (output) that was resulted showed either *k*-NN is dependent on the regression or classification or not.

- In *k-NN classification*, output consisted of the class membership. The classification is based on the majority of the votes of its neighbors, so assigning the utmost important class is among its *k* nearest neighbors from which *k* is a positive integer that is mostly small. If $k = 1$ then the single nearest neighbor is just allotted among the individual closest neighbor.
- In *k-NN regression*, the output obtained is the property value of the object. The mean value of all the values of its *k* nearest neighbors.

A type of instance-based learning and lazy learning is a type of *k*-NN through which the functions approximation is carried out locally and the calculation is delayed till the classification [16]. So, *k*-NN algorithm is regarded as the easiest of all learning algorithms.

3. **proposed system:**

The four classification algorithms used in this paper were programmed using java programming language and all four algorithms were displayed in the main windows of the system, ten fields were assigned to each algorithm, the shows all regarding information about the prediction of each algorithm.

Assigned to each algorithm a button with a caption of the same name, when pressed, a prediction and all regarding information is shown in all ten fields assigned to that algorithm.

The system is mainly split up to two main steps which is described in figure (1):

- **Preprocessing step** : where Word Occurrence is done in this step to perform tokenization which is mainly used uniform the keywords used and removing the stop words such as (a, the , …etc) and to perform streaming where the word return to its main root this is the preprocessing to the data before classify those data.
- **cross validation and classification algorithm** : the data will split up to training set and tested set the training set will be assigned to one of the classification algorithm.

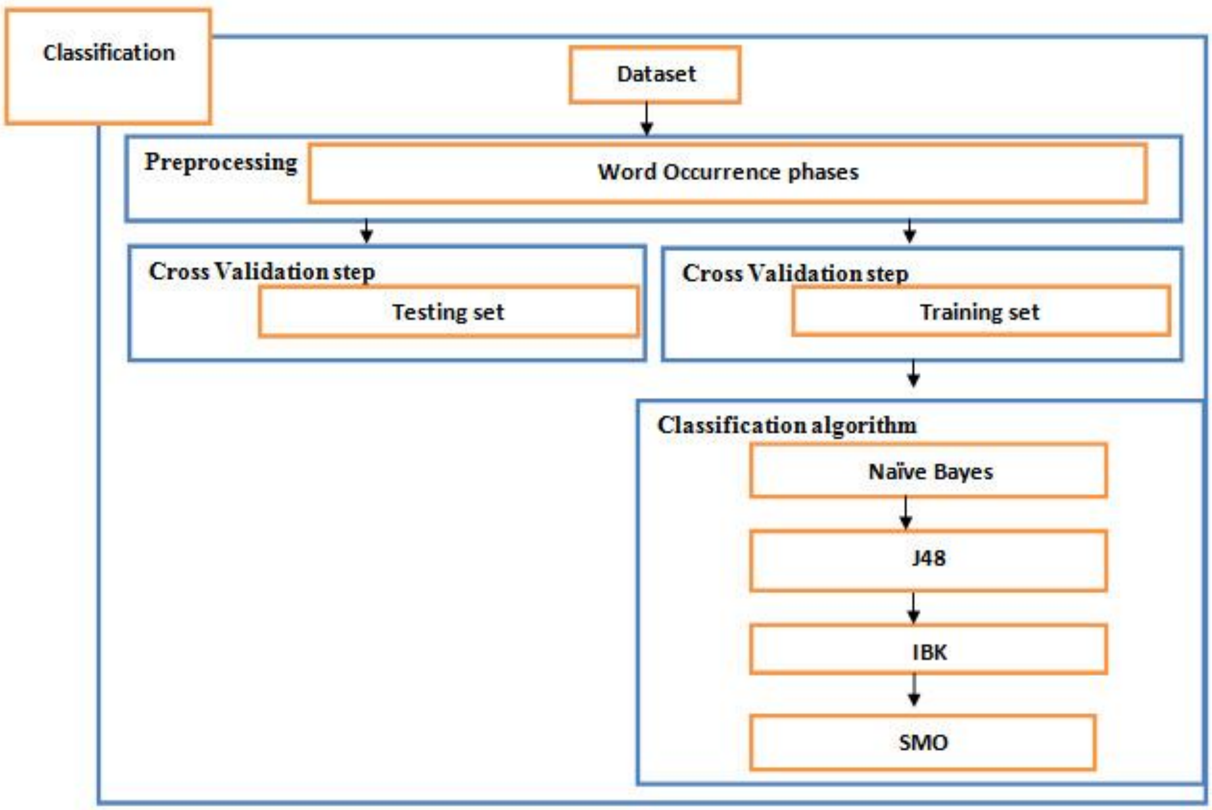

Figure (1) system architecture

The system will explain the related data of each method in the main window to compare with other algorithms, Figure (2) shows the main window of the system:

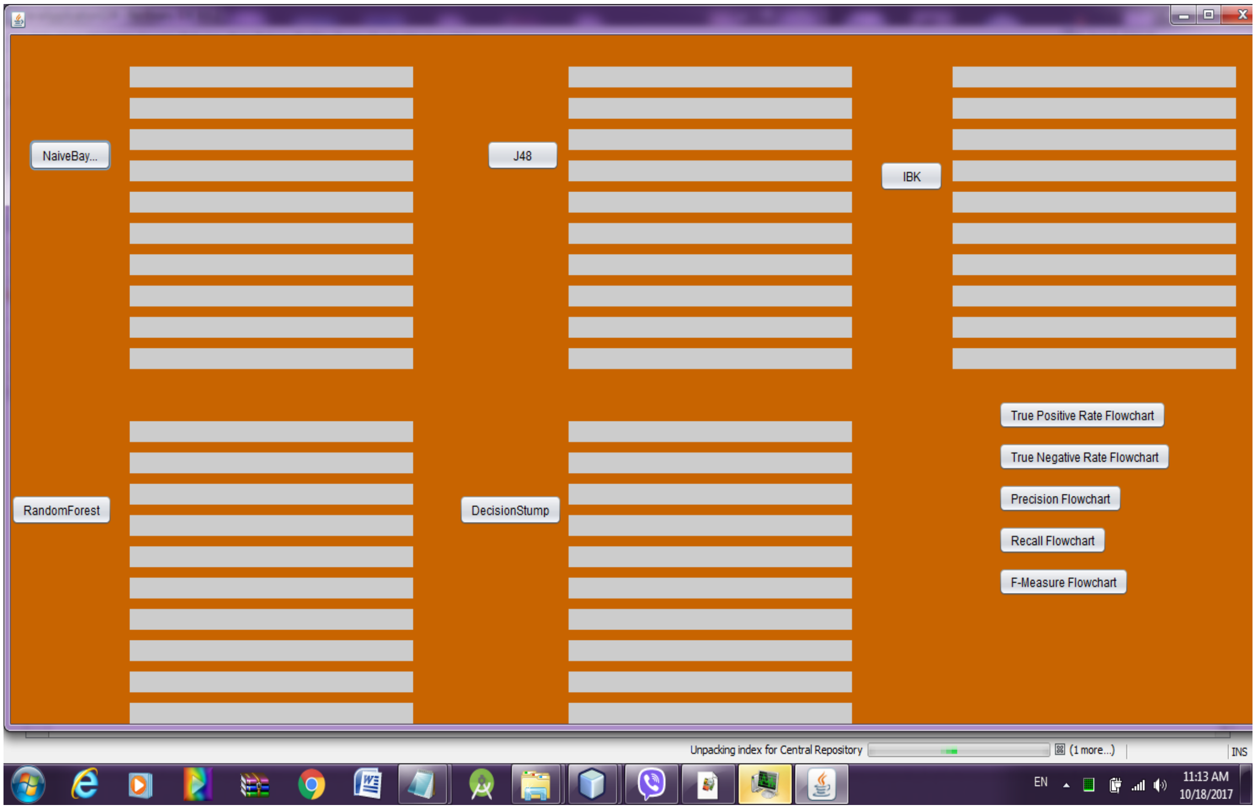

Figure (2) system main window

As shown in figure (2) related information to the classification process of each algorithm is calculated by the system such as (true positive, true negative, precision, recall and F-Measure) and each algorithm when execution had done have its own data as shown in figure (3).

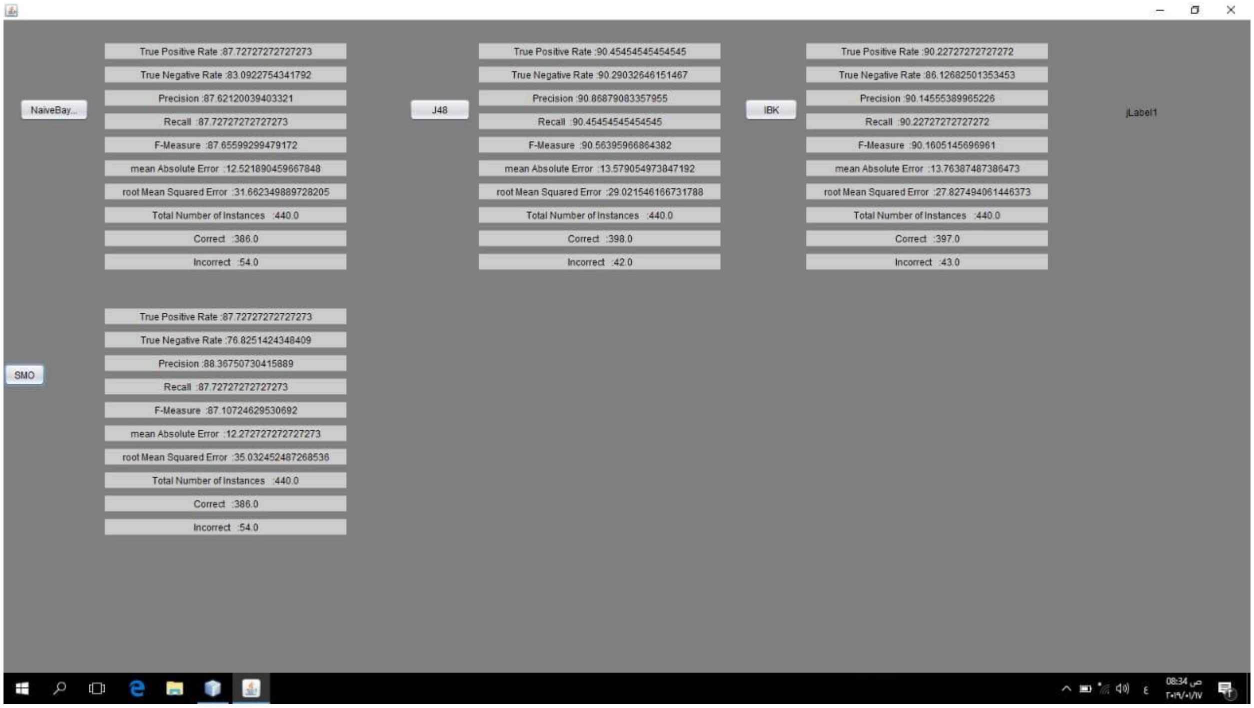

Figure (3) execution of the system and results obtained

4. EXPERIMENTAL RESULTS:

The first part of the system is the preprocessing part to solve the word of occurrences and to prepare data for classification and the training data after the cross validation is started will be classify, and when the dataset classified related information will shown up to user as shown in table (1) and figure (4) these values is founded by calculating true positive which is founded the correct predicts of the positive class ,and the true negative which is founded the correct predicts of the negative class.

Table (1): Classification Results-1

Algorithm	True Positive Rate	True Negative Rate	Precision	Recall	F-Measure
Naïve Bayes	87.72	83.09	87.62	87.72	87.65
J48	90.45	92.92	90.86	90.45	90.56
IBK	90.22	86.12	90.14	90.22	90.16
SMO	87.72	76.82	88.36	87.72	87.10

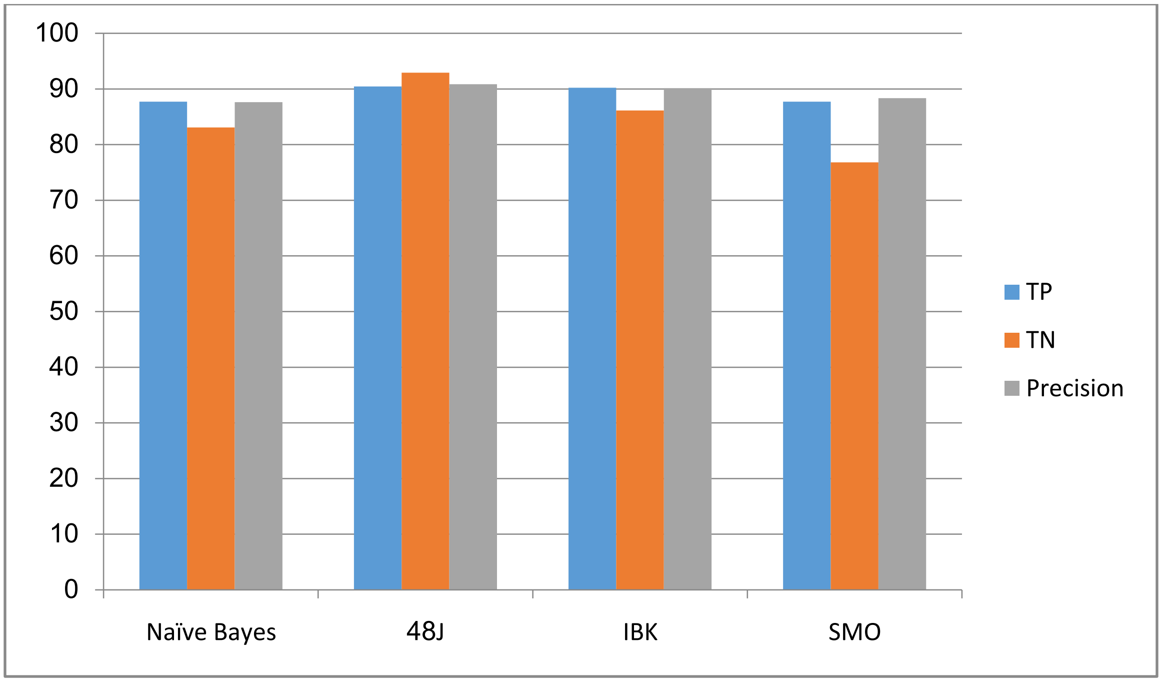

Figure (4) experimental results chart of true positive, true negative and precision

As it`s noticeable, from comparing the results of the four used algorithms in the system, the J48 algorithm seems to show more efficiency than the other algorithms in the all five aspects of (True Positive Rate, True Negative Rate, Precision, Recall, F-Measure).

Table (2): Performance of the classifiers

Algorithm	Mean Absolute Error	Root Mean Squared Error	Total Number of Instances	Correct	Incorrect	accuracy
Naïve Bayes	12.52	31.66	440	386	54	95.6%
J48	13.57	29.02	440	398	42	98%
IBK	13..76	27.82	440	397	43	95.3%
SMO	12.27	35.03	440	368	54	95.8%

Reference to table (2) and figure (5) the result obtained from executing the classification algorithms on the same dataset the J48 algorithm is best classifier comparable to others with accuracy of 98% with highly classified correct instances as well as incorrectly classified instance than other three classifiers.

True positive and negative shows the frequency of correct and incorrect predictions. It compares the actual values in the test dataset with the predicted values in the trained model.

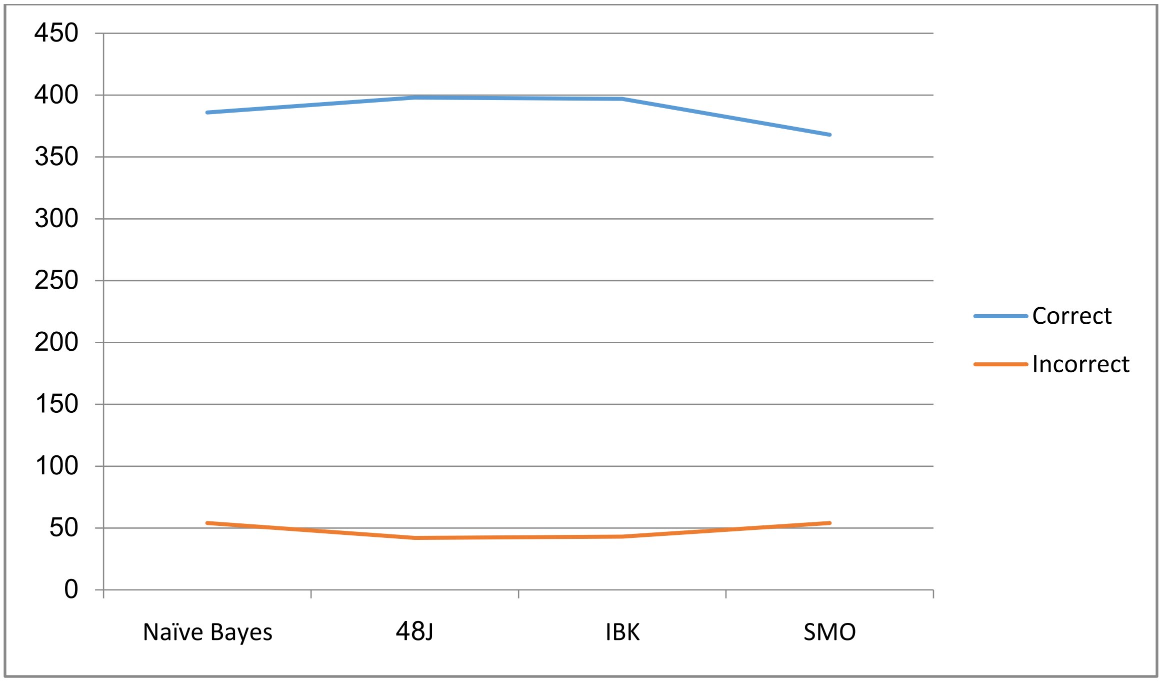

Figure (5) correct and incorrect classification rate

Accuracy were calculate via the global equation

Accuracy = TP/ (Total number of classification)

And precision is calculated via

$$TP / (TP+FP)$$

Where

Recall is founded via

$$TP / (TP+FN)$$

TP=true positive, TN= true negative, FP= false negative and FN= false negative

Conclusion:

In this paper, the accuracy of classification techniques is evaluated based on the selected classifier algorithm. Specifically, we used four popular data mining methods: Sequential Minimal Optimization (SMO), IBK, J48 Tree and Naïve Bayes. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for commercial applications. The performance of J48 shows the high level compare with other classifiers. Hence J48shows the concrete results with Wholesale customer's dataset. Therefore J48classifier is suggested for classification of commercial data to get better results with accuracy, low error rate and good performance.

References:

1. *Ge, Zhiqiang, et al. "Data mining and analytics in the process industry: the role of machine learning." IEEE Access 5 (2017): 20590-20616.*
2. *Board, Financial Stability. "Artificial intelligence and machine learning in financial services." November, available at: http://www. fsb. org/2017/11/artificialintelligence-and-machine-learning-in-financialservice/(accessed 30th January, 2018) (2017).*
3. *Fouladi, Ramin Fadaei, Cemil Eren Kayatas, and Emin Anarim. "Frequency based DDoS attack detection approach using naive Bayes classification." 2016 39th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2016.*
4. *N'Guessan, Assi, Issa Cherif Geraldo, and Bezza Hafidi. "An approximation method for a maximum likelihood equation system and application to the analysis of accidents data." Open Journal of Statistics 7.1 (2017): 132-152.*
5. *Wang, Sida, and Christopher D. Manning. "Baselines and bigrams: Simple, good sentiment and topic classification." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012.*
6. *Chen, Jingnian, et al. "Feature selection for text classification with Naïve Bayes." Expert Systems with Applications 36.3 (2009): 5432-5435.*
7. *Bhargava, Neeraj, et al. "Decision tree analysis on j48 algorithm for data mining." Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering 3.6 (2013).*
8. *Patil, Tina R., and S. S. Sherekar. "Performance analysis of Naive Bayes and J48 classification algorithm for data classification." International journal of computer science and applications 6.2 (2013): 256-261.*
9. *Zhao, Yongheng, and Yanxia Zhang. "Comparison of decision tree methods for finding active objects." Advances in Space Research 41.12 (2008): 1955-1959.*
10. *Kaur, Gaganjot, and Amit Chhabra. "Improved J48 classification algorithm for the prediction of diabetes." International Journal of Computer Applications 98.22 (2014).*
11. *Drazin, Sam, and Matt Montag. "Decision tree analysis using weka." Machine Learning-Project II, University of Miami (2012): 1-3.*
12. *Chen, Chao, Barnan Das, and Diane J. Cook. "A data mining framework for activity recognition in smart environments." 2010 Sixth International Conference on Intelligent Environments. IEEE, 2010.*
13. *Chaurasia, Vikas, and Saurabh Pal. "A novel approach for breast cancer detection using data mining techniques." (2017).*
14. *Frank, Eibe, et al. "Weka-a machine learning workbench for data mining." Data mining and knowledge discovery handbook. Springer, Boston, MA, 2009. 1269-1277.*
15. *Chetty, Girija, Matthew White, and Farnaz Akther. "Smart phone based data mining for human activity recognition." Procedia Computer Science 46 (2015): 1181-1187.*
16. *Ragab, Abdul Hamid M., et al. "A comparative analysis of classification algorithms for students college enrollment approval using data mining." Proceedings of the 2014 Workshop on Interaction Design in Educational Environments. ACM, 2014.*