Volume-02
Issue-01
Feb-2019

ISSN: 2663-0931(Print)
ISSN: 2663-2594(online)
International Journal of
Interdisciplinary Current Advanced Research
www.ijicar.in

Predict and Analysis of Employees Performance in Bank Using Classification Algorisms: A Case Study of Commercial Bank of Ethiopia

[*1] Kedir Eyasu Abdulkadir, [2] Fulea Amena Tolfsa,
[*1] *Lecturer Department of computer Science, Mettu University, Mettu, Ethiopia.*
[2] *Lecturer Department of Accounting And Finance, Mettu University, Mettu, Ethiopia.*

*Note: * Indicates corresponding author*

ARTICLE DETAILS

Article History:

Received Date:11/02/2019

Revised Date: 15/02/2019

Accepted Date:25/02/2019

*Published Online:*10/03/2019

Keywords

Commercial bank improvement, Data mining, J48, PART, Naïve Bayes, Classification algorisms

[*]Corresponding Author

Email: kedir.eyasu@gmail.com

(Kedir Eyasu Abdulkadir)

ABSTRACT

In today's globalization and cut throat competition the banks are struggling to gain a competitive edge over each other. Apart from execution of business processes, the creation of knowledge base and its utilization for the benefit of the bank is becoming a strategy tool to compete. In recent years the ability to generate, capture and store data has increased enormously. The information contained in this data can be very important. The wide availability of huge amounts of data and the need for transforming such data into knowledge encourage IT industry to use data mining. The banking industry around the world has undergone a tremendous change in the way business is conducted. The banking industry has started realizing the need of the techniques like data mining which can help them to compete in the market. Leading banks are using Data Mining (DM) tools for customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. This paper is aimed to predict and analysis what will happen in the banking area based on the past experience of Ethiopian Commercial Bank workers. In the study the researchers used one famous data mining algorithm (sub-task of predictive analysis modeling, which is classification). The target algorithms found in classification, i.e. PART, J48 decision tree and naïve Bayes are applied here to predict and analysis employee's performance using classification model by applied on datasets already collected and preprocessed. Methodology used for the study was Hybrid Data Mining techniques and data sets was collected from commercial bank of Ethiopia using secondary data collection method. CRISP-DM method was applied to understand both the business and on datasets to identify and compare classifiers to predict and analysis. Finally, For J48 decision tree 98.8651 %, For PART classifiers results 99.2372 % and for Naïve Bayes classifier we got 73.3767%. So, since the J48 and Naïve Bayes algorithms are less in having better accuracy, the researcher selected PART classifier for prediction and analysis employee's performance. However, bank business strategy improving for employees in order to enhance customer attractiveness is future work.

1. Introduction

Data Mining (DM) aims to extract useful knowledge from raw data. Interest in this field arose due to the advances of Information Technology and rapid growth of business and scientific databases. These data hold valuable information such as trends and patterns, which can be used to improve decision making. Two important DM tasks are classification and regression. Both tasks use a supervised learning paradigm, where the intention is to build a data-driven model that learns an unknown underlying function that maps several input variables to one output target. Several learning models/algorithms are available for these tasks, each one with its own advantages. In a real-world setting, the value of a supervised DM model may depend on several factors, such as predictive capability, computational requirements and explanatory power. Often, it is important to have DM models with high predictive capabilities on unseen data. Computational effort and memory requirements are particularly relevant when dealing with vast datasets or real-time systems *(Cortez & Embrechts, 2013).*

In the financial services industry throughout the world, the traditional face-to-face customer contacts are being replaced by electronic points of contact to reduce the time and cost of processing an application for various products and ultimately improve the financial performance. The computerization of financial operations, use of internet and

automated software's has completely changed the basic concept of business and the way the business operations are being carried out. The banking sector is not an exception to it. It has also witnessed a tremendous change in the way the banking operations are carried out *(Moin & Ahmed, 2012).*

The intention of this paper is to predict and analysis the workers of Commercial Bank of Ethiopia based on the datasets analysis using classifier model and to recommend the banks the better way of enhancing the effectiveness and efficiency of workers to serve their customers. As a result, the bank business increases when customer's satisfactions improved in well-organized manner.

2. Statement of The Problem

Today the advancement of technology is highly increasing. Due to this reason majority of our day to day activity including banking system are shifting from manual system to automated one. As a result, there are volumes of data produced even within single bank. Those data are source of essential information and knowledge. But, due to lack of individuals who can extract those hidden knowledges and because of lack of attention for data produced daily, customers didn't get sufficient service from the banks and also banks do not have enough information about the behavior of their workers. Lack of knowledge how to processing this huge data that available in bank which generated in ever day transaction. Analyze available data related to employees in order to improve the motivation of worker from time to time need much attention. As a result, their motivations increase the productivity of bank and attract the customer as business root. Therefore, the aim of this study is to apply data mining for constructing a predictive model for determining employee's performance rate. To this end, this study attempts to explore and answer the following research questions

1. What are the relevant attributes for predicting a performance using data mining techniques?
2. Which data mining algorithm is the good predictor to solve the problem of performance rate?

3. Objectives of The Study

- To Predict and Analysis of Employees Performance in Bank using classification algorism Case Study: Mettu Commercial Bank of Ethiopia

A. General Objective

- The general objective of this study is to Predicting and Analysis of Employees Performance in Bank using classification algorism

B. Specific Objectives

- This research was building predict and analysis model for employee's performance in bank, there are specific objectives formulated to achieve the main objective.
- To categorize attributes which could affect bank worker's recitals
- To prepare quality datasets for training and testing that used for predicting and analysis of performance.
- To demonstrate which algorithm of the classification is best to predicting and analysis performance rate.

4. Scope of The Study

The data was collected from Commercial Bank of Ethiopia, Mettu district, main branch. The researchers implemented classification rule in data mining to predict something for the well-functioning of banks and creating satisfied customers. The algorithms for this study were J48 decision tree, PART and Naïve Bayes. The model which only predicts and analysis employees' performance rate without including other confidential datasets of bank. The limitation of this study is the scarcity of encoded data while accessing the saved data from the bank.

5. Significance of The Study

The main aim of this study is to implement the classification algorithm to predict what will happen in the banking area. Based on this the outcome of the study was provided better information that can help both banks, customers and employee to develop better relationship and to have business transactions within short period of time. This study was used the classification algorism to provide or discover information or knowledge that can help both customers and bankers to adopt better culture for future business activities. It was providing the essential steps to be followed when other researchers want to do the study by using classification system. It helps the higher officials of the Commercial Bank of Ethiopia to know the effectiveness and efficiency of their workers. The study provides supports for the stated problems about the life status of employees after analysis and predicting using available datasets. The significance of the predicting and analysis of performance of employees in bank is the major one in business running environment. Generally, the study, model was provided a support on the bank case for bank worker and government to understand the performance rate and it helps the whole banks.

6. Research Methodology

The overall activity of this thesis were used a Hybrid model which is six step knowledge discovery process model. The Hybrid DM process model describes procedures that are performed in each of its steps (*Jinhong et al., 2009).* Due to the nature of the problem and attributes in the dataset, classification mining task were selected to build the classifier models. J48 decision tree, Naïve Bayesian and PART rule induction algorithm is used for this study for developed the model. It is undoubted that the data mining techniques has different types of models which can solve the real world problem which is really integrated with the data base, record, data ware house and files. The Hybrid model contains both the KDD and CRISP-DM at one place. Therefore, the researcher was used the Hybrid DM processes model for this specified study. Because, this model is really fruit full in the prediction of health oriented problem and has many option of feedback mechanism more than CRISP DM (Fayyad et al., 1998). The hybrid DM processes model provides the six step processes to generate the important knowledge and model development. Therefore, the study is completed under the guidance of the Hybrid DM processes model.

7. Research Design

This research is designed to identify the determinant of predict and analysis of employee's rate. To explore the application of data mining on this particular research, hybrid (Ciso.et al) data mining methodology was employed. The explanations of those models are discussed under literature review and chapter three of this study. The hybrid model provides the six step processes to generate the important knowledge and model development i.e. understanding of domain, understanding of data, preparation of data, data mining, and evaluations of discovered knowledge and uses of knowledge.

In this study the researchers used the CRISP-DM (CRoss Industry Standard Process for Data Mining) method to clearly understand the business transaction of the Commercial Bank of Ethiopia. In CRISP-DM there are six steps:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation &
- Deployment (Moro et al., 2011).

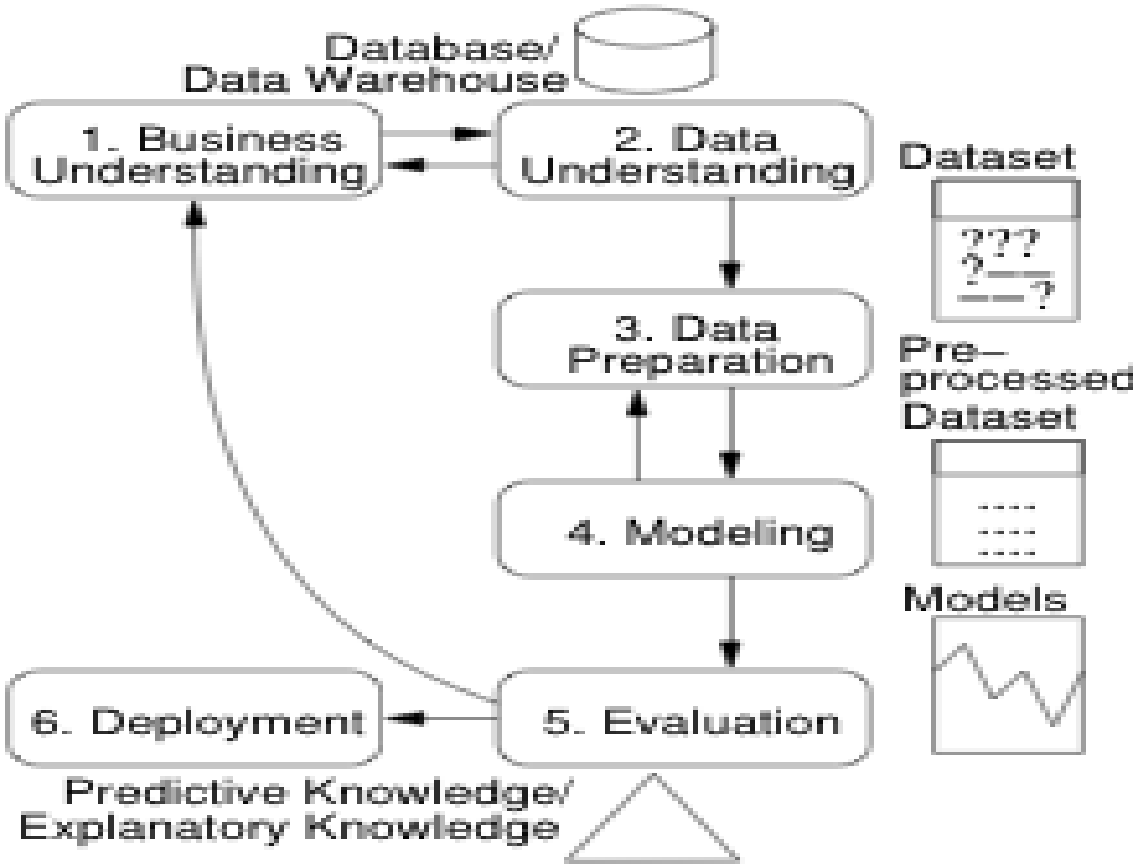

Figure 1: CRISP-DM model (adopted from Moro et al., 2011)

8. Literature Review

Predictive modeling is the process by which a model is created or chosen to try to best predict the probability of an outcome. In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data. Classification is a predictive data mining technique, makes predication about values of data using know results found from different data. Predictive models have the specific aim of allowing us to predict the unknown value of a variable of interest given known values of other variables. Predictive modeling can be thought of as learning a mapping from an input set of vector measurements x to a scalar output y. Classification maps data into predefined groups are classes. It is often referred to as supervised learning because the classes are determined before examining the data. They often describe these classes by looking at the characteristic of data already known to belong to the classes *(Pandey & Pal, 2011).*

Baradwaj and Pal in 2011 did the research on Mining Educational Data to Analyze Students Performance. In the study they use classification task with algorithm of decision tree. By this task they extracted knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling *(Baradwaj & Pal, 2012).*

Bhardwaj and Pal in 2011 they did research on data mining which is a prediction for performance improvement using classification. They collected records of 300 students and tried to identify students as high learner and slow learner. They also studied the records by using Bayes classification in order to identify the difference between high and slow learners. Finally, they decided that Present study shows academic performances of the students are not always depending on their own effort. Their investigation shows that other factors have got significant influence over students' performance *(Bhardwaj & Pal, 2012).*

Moin and Ahmed's work showed that the main tools of data mining and their application. Especially the researchers tried to indicate the application of data mining sub tasks their algorithms in bank area. Activities like risk management, fraud detection, marketing and customer relationship management are main issues addressed in the study *(Moin & Ahmed, 2012).*

Ravisankar and other researchers in 2010, they tried to study on detection of financial statement fraud and feature selection using data mining techniques. Recently, high profile cases of financial statement fraud have been dominating the news. The researchers here used data mining techniques such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) to identify companies that resort to financial statement fraud. Each of these techniques is tested on a dataset involving 202 Chinese companies and compared with and without feature selection. PNN outperformed all the techniques without feature selection, and GP and PNN outperformed others with feature selection and with marginally equal accuracies *(Ravisankar, et al., 2011).*

In order to determine how data mining techniques (DMT) and their applications have developed, during the past decade, Liao and other researchers reviewed data mining techniques and their applications and development, through a survey of literature and the classification of articles, from 2000 to 2011. Keyword indices and article abstracts were used to identify 216 articles concerning DMT applications, from 159 academic journals (retrieved from five online databases), this study surveyed and classified DMT, with respect to the following three areas: knowledge types, analysis types, and architecture types, together with their applications in different research and practical domains. A discussion dealt with the direction of any future developments in DMT methodologies and applications: (1) DMT is finding increasing applications in expertise orientation and the development of applications for DMT is a problem-oriented domain. (2) It is suggested that different social science methodologies, such as psychology, cognitive science and human behavior might implement DMT, as an alternative to the methodologies already on offer. (3) The ability to continually change and acquire new understanding is a driving force for the application of DMT and this will allow many new future applications *(Liao, et al., 2012).*

9. Data Preparation and Business Understanding

9.1 Domain Understanding

This initial step involves working closely with domain experts to define the problem and determine the research goals, identifying key people, and learning about current solutions to the problem. It also involves learning domain-specific terminology. For this purpose, the researcher was selected domain expert by clarifying the problem and aim of the study. Justifying the selected attribute for the stated problem regarding to the prediction of the bank workers predict and analysis rate is very essential. Banks are institutions which work highly in business transactions. They serve thousands of customers daily. Each and every customer has their own behavior. So, banks should understand the need of users and motivate the workers to enhance their effectiveness and efficiency. The more skilled man power the bank has the better service it can provide for its customers sufficient.

9.2 Data Understanding

The data the researchers got from the bank is data about the workers of Commercial Bank of Ethiopia. It gives some information about the workers. For instance, their educational status, salary, their job within the bank and others. The data collected from the bank for this research contains about twenty-two attributes. Most of the attributes have its own characteristics and can determine the well-functioning of the workers this step includes collecting sample data and deciding which data, including format and size, will be needed. For this purpose, Secondary data was used as a source of information. The secondary data included different worker's data which is found in bank that data is

saved in the form of Excel. The data set stored for three years record from the year 2016 to 2019. The total data set obtained from bank was data is around 5375 instances with 19 attributes.

9.3 Data Preparation

Data preparation concerns all activities needed to construct the final dataset for modeling purposes. The tasks are most likely to be carried out multiple times and may not be in any prescribed manner. Different datasets tend to expose new issues and challenges. With the goals in mind, it is important to choose the right data mining algorithms, techniques and tools which are expected to give best results with the provided data. Dependencies among different subsets of attributes are expected to be exhibited by different subsets of data. Most often, not all variables are used in analyzing and modelling process. This phase was conducted repetitively for determining suitable attributes to be used as predictors and target (output) (*Siraj & Abdoulha, 2007).*

For data preparation purpose the researchers tried to remove missing value from the collected data. Only a few numbers of instances are reduced because they have no much effect on the data collected and the number of remaining records is sufficient for data mining purpose. This step concerns deciding which data was used as input for DM methods in the subsequent step. It involves sampling, running correlation and significance tests, and data cleaning, which includes checking the completeness of data records, removing or correcting for noise and missing values. This step covers preprocessing the data and making ready the maternity ward data for model development. For this study WEKA version 3.7 is the major tool to drive the study forward. WEKA is a good tool for the hybrid model since the model has different types of steps to preprocess the data. There is also other tool of data mining like Rapid miner and TANAGRA, but due to the researcher's familiarity with WEKA tool and WEKA tool has so many java dependent environment and preprocessing option more than Rapid miner and Tanagra. Also so many researchers have been used this tool before, for the purpose of mining hidden knowledge in the stored data. Additionally, MS-Excel is the primary tool which the used for the purpose of data cleaning and preprocessing situation.

9.4 Data Mining Techniques

Here the data miner uses various DM methods to derive knowledge from preprocessed data. The selected algorithm for this study is J48 decision, PART rule induction and Naïve Bayes applied on the data to dig out the hidden knowledge by using Weka tool for the prediction maternal mortality rate.

9.5 Data Transformation

This stage is very essential in data mining. It means that preparing or categorizing the collected data according to suitability of data or records to be analyzed in data mining tools. In this study, we have collected data with 15 attributes and5000 instances or horizontal records. We tried to prepare those attributes and assigned value for each of them.

Table 1: data transformation

Attributes	Data type	Value assigned
Age	Numeric	Value >= 20
Sex	Nominal	{male, female}
Region	Nominal	{Oromiya, Dire dawa, finfine, SNNPR, Amhara, Gambela,}
Profession	Nominal	{certificate, degree, diploma, masters}
Field of study	Nominal	{accounting, management, computer science, information science, network administrator}
Income	Numeric	{>=2500}
Work division	Nominal	{front casher, audit manager, security, front casher manager, customer service manager, branch manager, system administrator, database administrator, network admin, fresh

		casher, money control class, accountant, district manager}
Experience	Numeric	>=0 years
On job status	Nominal	{active, semi active}
Marital status	Nominal	{yes, no}
No of children	Numeric	>=0
House hold	Nominal	{yes, no}
Car hold	Nominal	{yes, no}
Contact	Nominal	{yes, no}
Saving account	Nominal	{yes, no}
Current account	Nominal	{yes, no}
Loan	Nominal	{yes, no}
Mortage	Nominal	{yes, no}
Tax-free investment	Nominal	{yes, no}

Experiment-1-J48 Pruned Tree

Decision trees are data-mining methodologies applied in many real-world applications as a powerful solution to classification problems. In decision tree experiment, the performance of J48 classifier in predicting diabetes status of patients was evaluated. The experiment was conducted with the default parameters of WEKA. From the total dataset of **5375** records, **5314**were correctly classified and the remaining **61**instances were incorrectly classified. Decision tree is a graphical representation of the relations that exist between the data in the database. It is used for data classification. The result is displayed as a tree, hence the name of this technique. Decision trees is mainly used in the classification prediction and analysis. It is a simple and a powerful way of representing knowledge.

```
Time taken to build model: 0.56 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        5314               98.8651 %
Incorrectly Classified Instances        61                1.1349 %
Kappa statistic                          0.9772
Mean absolute error                      0.0086
Root mean squared error                  0.0677
Relative absolute error                  2.5843 %
Root relative squared error             16.624  %
Coverage of cases (0.95 level)          99.814  %
Mean rel. region size (0.95 level)      34.6605 %
Total Number of Instances             5375

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.991     0.014      0.984     0.991     0.988      0.999    YES
                 0.986     0.009      0.993     0.986     0.989      0.999    NO
                 1         0          1         1         1          1        TAX FREE INVSTEMENT
Weighted Avg.    0.989     0.011      0.989     0.989     0.989      0.999

=== Confusion Matrix ===

    a    b    c   <-- classified as
 2445   21    0 |    a = YES
   40 2867    0 |    b = NO
    0    0    2 |    c = TAX FREE INVSTEMENT
```

Figure 2: J48 pruned experiment output

Experiment -2-Part Classifier

PART is a rule-based classifier uses a set of IF-THEN rules for classification. An IF-THEN rule is an expression of the form IF condition THEN conclusion. The “IF”-part (or left-hand side) of a rule is known as the rule antecedent or pre condition. The “THEN”-part (or right-hand side) is the rule consequent (Chen, 2009). This experiment conducted under percentage splits technique using 99.2372% of instances for training and the remaining for testing with default parameters of Weka and Correctly Classified Instances are 5334 and Incorrectly Classified Instances are 41 which means 0.7628 %from Total Number of Instances of 5375.

```
Time taken to build model: 2.69 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        5334               99.2372 %
Incorrectly Classified Instances        41                0.7628 %
Kappa statistic                          0.9846
Mean absolute error                      0.0065
Root mean squared error                  0.0611
Relative absolute error                  1.9744 %
Root relative squared error             15.0089 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      34.4558 %
Total Number of Instances             5375

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.987     0.003      0.996     0.987     0.992      1        YES
                 0.997     0.013      0.989     0.997     0.993      1        NO
                 0         0          0         0         0          1        TAX FREE INVSTEMENT
Weighted Avg.    0.992     0.008      0.992     0.992     0.992      1

=== Confusion Matrix ===

    a    b    c   <-- classified as
 2435   31    0 |    a = YES
    8 2899    0 |    b = NO
    1    1    0 |    c = TAX FREE INVSTEMENT
```

Figure 3: J48 pruned experiment output

Experiment -3-Naïve-Bayesian Classifier

It is one of the classifiers that is probabilistic model which incorporate strong independence assumptions. This experiment conducted under percentage splits technique using 90% of instances for training and the remaining for testing with default parameters of Weka and Correctly Classified Instances are 3944 and Incorrectly Classified Instances are 1431.Total Number of Instances of 5375.


```
Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3944               73.3767 %
Incorrectly Classified Instances      1431               26.6233 %
Kappa statistic                          0.4587
Mean absolute error                      0.238
Root mean squared error                  0.3428
Relative absolute error                 71.8123 %
Root relative squared error             84.2095 %
Coverage of cases (0.95 level)          99.8326 %
Mean rel. region size (0.95 level)      64.738  %
Total Number of Instances             5375

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.641     0.188      0.743     0.641     0.688      0.812    YES
                 0.812     0.359      0.727     0.812     0.767      0.812    NO
                 1         0          1         1         1          1        TAX FREE INVSTEMENT
Weighted Avg.    0.734     0.28       0.735     0.734     0.731      0.812

=== Confusion Matrix ===

    a    b    c   <-- classified as
 1581  885    0 |    a = YES
  546 2361    0 |    b = NO
    0    0    2 |    c = TAX FREE INVSTEMENT
```

Figure 4: J48 pruned experiment output

9.6 Classification Model Evaluation

This is the fifth step of the Hybrid DM processes model which evaluates the discovered knowledge and rules are evaluated for the predicting and analysis performance of bank employees. The performance and accuracy of the model on the data is investigated depending on the model responds and discussing with Domain expert weather the knowledge is valid for prediction and analysis performance.

Table 2: Performance prediction and analysis of Classifiers

Model Evaluation	Correctly classified instances		Incorrectly classified instances		Time taken to build model	Precision	Recall	F Measure
Classifiers	Instances	Percentage	Instances	Percentage	Time/seconds			
PART	5334	99.2372 %	41	0.7628 %	3.52	0.992	0.992	1
J48	5314	98.8651 %	61	1.1349 %	0.2	0.989	0.989	0.999
Naive Bayes	3944	73.3767 %	1431	26.6233 %	0.05	0.734	0.731	0.812

10. Result Discussion

For this study we used three classifier algorithms, namely J48 decision tree, PART and Naïve Bayes. We used all these three classifiers to predict the effectiveness and efficiency of workers of Commercial Bank of Ethiopia in Mettu city. The accuracy of the results that we got from each classifier is different. For J48 decision tree 98.8651 %, For PART classifiers results 99.2372 % and for Naïve Bayes classifier we got 73.3767%. So, since the J48 and Naïve Bayes algorithms are bad in having better accuracy, we choose PART classifier for prediction and analysis purpose. It has only 41 instances which are classified incorrectly as false positive and false negative. The performance of PART is better when we compare to its. This means that the classification is better in classification of performance of the employees in the bank based on their experience and motivation of work.

So, based on the obtained result the bank needs to be motivating his employees with different reward and let them educate. But, the in confusion matrix, the number of instances classified as 'No' are classified shows that workers of Commercial Bank of Ethiopia are not effective and efficient since they have challenges that hinder them not to act as expected. As a result, the bank need to improve and solve the challenge in order to enhance the quality of work and to be profitable is basically important. Generally, Study shows to enhance much profitability and competency of bank better to focus employment motivation and create competency level for employees.

11. Conclusion

This study is aimed to predict the efficiency and effectiveness of workers of Commercial Bank of Ethiopia by using the classifiers known as PART, J48 decision tree and Naïve Bayes. The datasets were from Commercial Bank of Ethiopia Mettu city main branch. It was with 19 attributes and 5375 instances. CRISP-DM method was applied to understand both the business and on datasets to identify and compare classifiers to predict and analysis. As a result, based on the accuracy and analysis of obtained from the algorisms. The effectiveness of the workers within the bank will contribute for fulfillment of the customer satisfaction and its better culture for fast and competent business transaction.

12. Recommendation

The researchers investigate applying data mining techniques for predicting and analysis bank employees for better enhancing bank business transaction and customer attractive. The more information is discovered the better approach there will be for the customers. The more workers are treated in proper manner the better the service delivered for the individuals around business area. Finally, we recommend that this study should be including database datasets.

References

1. Ahmadi, M. A., Ebadi, M., Shokrollahi, A., & Majidi, S. M. J. (2013). *Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir.* Applied Soft Computing, 13(2), 1085-1098.
2. Baradwaj, B. K., & Pal, S. (2012). *Mining educational data to analyze students' performance.* arXiv preprint arXiv:1201.3417.
3. Bhardwaj, B. K., & Pal, S. (2012). Data Mining: *A prediction for performance improvement using classification*. arXiv preprint arXiv:1201.3418.
4. Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2010). WEKA manual for version 3-7-3. The University of WAIKATO.

5. Chaudhari, B., & Parikh, M. (2012). *A Comparative Study of clustering algorithms using Weka tools. International Journal of Application or Innovation in Engineering & Management* (IJAIEM), 1(2), 154-158.
6. Cortez, P., & Embrechts, M. J. (2013). *Using sensitivity analysis and visualization techniques to open black box data mining models. Information Sciences*, 225, 1-17.
7. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). *Data mining techniques and applications–A decade review from 2000 to 2011. Expert systems with applications,* 39(12), 11303-11311.
8. Mohamed, W. N. H. W., Salleh, M. N. M., & Omar, A. H. (2012, November). *A comparative study of reduced error pruning method in decision tree algorithms. In Control System, Computing and Engineering (ICCSCE)*, 2012 IEEE International Conference on (pp. 392-397). IEEE.
9. Moin, K. I., & Ahmed, D. Q. B. (2012). *Use of data mining in banking*. International Journal of Engineering Research and Applications, 2(2), 738-742.
10. Moro, S., Laureano, R., & Cortez, P. (2011). *Using data mining for bank direct marketing: An application of the crisp-dm methodology.* In Proceedings of European Simulation and Modelling Conference-ESM'2011 (pp. 117-121). Eurosis.
11. Pandey, U. K., & Pal, S. (2011). *Data Mining: A prediction of performer or underperformer using classification*. arXiv preprint arXiv:1104.4163.
12. Patil, T. R., & Sherekar, S. S. (2013). *Performance analysis of Naive Bayes and J48 classification algorithm for data classification.* International Journal of Computer Science and Applications, 6(2), 256-261.
13. Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). *Detection of financial statement fraud and feature selection using data mining techniques.* Decision Support Systems, 50(2), 491-500.
14. Sharma, T. C., & Jain, M. (2013). *WEKA approach for comparative study of classification algorithm.* International Journal of Advanced Research in Computer and Communication Engineering, 2(4), 1925-1931.
15. Siraj, F., & Abdoulha, M. A. (2007). *Mining enrolment data using predictive and descriptive approaches.* Knowledge-Oriented Applications in Data Mining, 53-72.

