A PROPOSED MODEL TO MEASURE OF THE PERFORMANCE OF ATM MACHINE

Abeer A. Amer, Mohamed Ebied , Nevine Makram Labib

***Abstract**: Data Mining (DM) plays an important role in banking sector to help improve customer experience especially when using automated teller machine (ATM). The role of ATM has been maximized recently due to the national trend of digital transformation and the emergence of COVID-19 pandemic since the ATM provides banks' customers with financial transaction method in public places without any human interaction. which led to the difficulty of financial institutions in measuring the performance of automated teller machines and counting data Whether it is data with a direct or indirect impact on the machines This paper presents a model that helps measuring the efficiency of the ATM by building data warehouse (DW) (Schedule server of ATM, Location, transaction and customer relationship management (CRM)and compares between two DM algorithms (decision tree and naive baye). The decision tree (C5) has been proved to be more convenient method for measuring an ATM's efficiency.*

***Keywords**: computer model, phase transformation, austenite decomposition, CCT-diagram.*

***ITHEA Keywords**: I.6 SIMULATION AND MODELING*

***DOI**: https://doi.org/10.54521/ijita28-01-p02*

Introduction

Automated Teller Machine (ATM) is a digital telecommunication device that offers financial institutions' customers a method of financial transactions in a public space without the need for a human [1]. Most ATMs are connected to international bank networks, enabling people to withdraw and deposit money from machines that are not belonging to the bank or country where they have their account. Automated Teller Machines are one of the most important cash distribution channels for banks [1][3]. Bank clients would be able to conduct their normal transactions such as bill payment, transfer funds and cash request without worrying about banks' working days, business hours, etc[1][2].

As for DW play an important role in financial and banking institutions, as they deal with huge amounts of data, (DW) "is defined process for collecting and managing data from varied sources to provide meaningful insights. Through central repositories of integrated data from one or more different sources. They store current and historical data in one single place. It is a process of transforming data into information and making it available to users in a timely manner to make a difference" [4].

However, ATMs handle many different data on a daily basis, which makes evaluating devices difficult. Here, the importance of DM in banks in data analysis in line with bank requirements is defined as a determination procedure to discover patterns in big data. DM uses different methods from machine learning systems in database analysis. DM adopts a basic process in which intelligent methods of data extraction are related. The overall goal of the information management process is to remove data from the information index and convert it into a support structure for further use [5].

Related Work

DW and DM in banking sector was discussed in many recent studies as follows:

First study presents a bank model developed can help the loan department of the banks to increase their customer potential and capability of applying for loan. They can help the bank to identify more transactions to discover more knowledge Through the bank data warehouse. Future work can focus on generating more test cases to execute on the data warehouse and results can be analyzed to further enhance banking practices and operations [6].

Second study, the researcher presents a bank model that focuses on the loan department of a bank DW and shall explain how business intelligence plays a role in improving the loan analysis for the banks. Through DW simply consolidates historical data at one place and is used to support strategic and business decision making. It helps any enterprise in making business decisions for the welfare and growth of the organization. DW works as a knowledge worker in the role of decision making and data analysis. It helps the managers to resolve problems within a defined timeline and facilitate effective decision making [7].

Third study found out ways on how to analyze ATM in pakistan's largest banks to get rid of cash problem, feeding into a time-series prediction loop, through optimization algorithms such as: linear regression, ridge linear regression, LASSO Model Ridge CV Model, LASSO LAR, Bayesian Ridge Regression, Random Forest Regression, time series prediction, RRN, RRN with time series and ARIMA. In the final analysis selected, the linear algorithm provided an optimal solution and 98% of accuracy with that approach on the new dataset [8].

Fourth study presented a model that provides a score to an ATM location. In order to efficiently capture the spatially dynamic features, two concurrent prediction models have been used, the local model and the global model. The results obtained from the energetic features using the models were very promising and encouraging. Maintaining the volume of local model as 0.65 and 0.35 assigned to the global model from the estimated score of the county regions, can maximize revenue if there is a venture for opening a new ATM network [9].

A final study showed that 353 ATM card users have assessed the ATM performance through regression analyses of the performance of 25 attributes adopted from empirical studies, which in turn provides a perfect model for predicting customer satisfaction. Reliability and responsiveness are the key service quality dimensions of ATM banking. The analyses revealed 12 key features that influence customers’ satisfaction with ATM banking services [10].

According to the previous researches it was found that proposed data mining techniques to measure ATM performance, based on the analysis only but there was no merging between the data warehouse to collect different sources to help in the analysis in more appropriate way to measure the performance of the ATM.

The Proposed Model

DW (ATM) can be described in four phases. Phase one is data source phase which contains relational databases (DB) and Excel sheets. Phase two is the analysis phase in which data is selected, extracted, transformed, integrated and loaded into data marts. Phase three is DW phase which includes several data marts such as: ATM transactions, ATM customer complaint, credit card complaint, ATM location and ATM schedule. Phase four is the analysis phase which is divided into one sub-phase of DM techniques and the evaluation phase of these techniques.

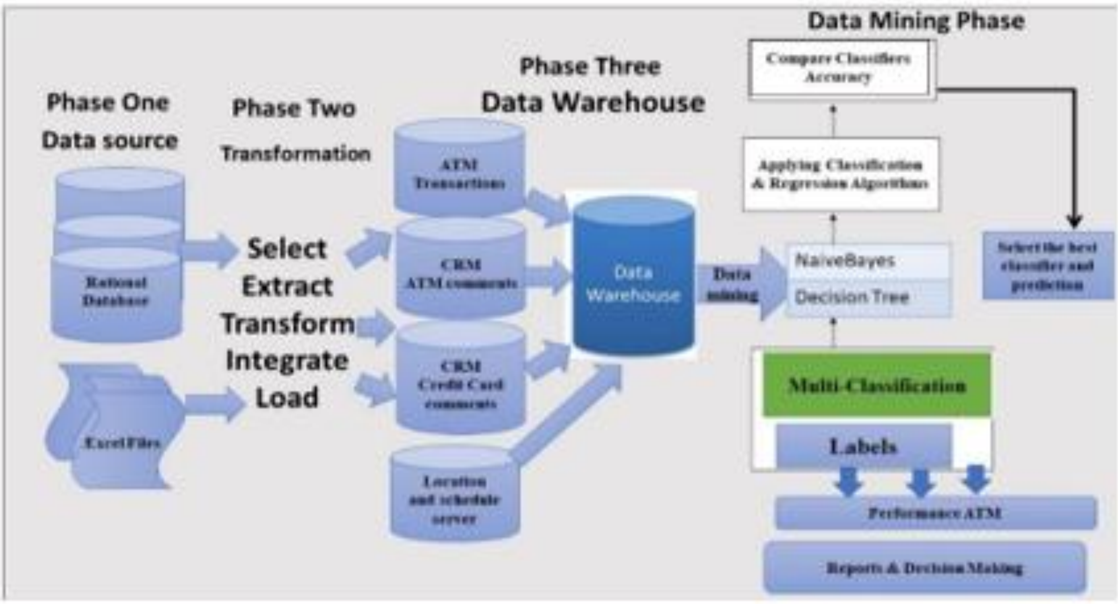

Figure 1. The ATM Model.

1.1 Phase One: (Data Sources)

The paper herein collected data sources through ATM transactions over a fiscal year from 2017 to 2018 for the bank. The Automated teller machines of these transactions are used to build three relational databases and three excel sheets, for example:

1.1.1. Designing Transactions DB:

The Transactions of ATM relational DB can be seen in figure 2, which outline five tables with their relationships as follows:

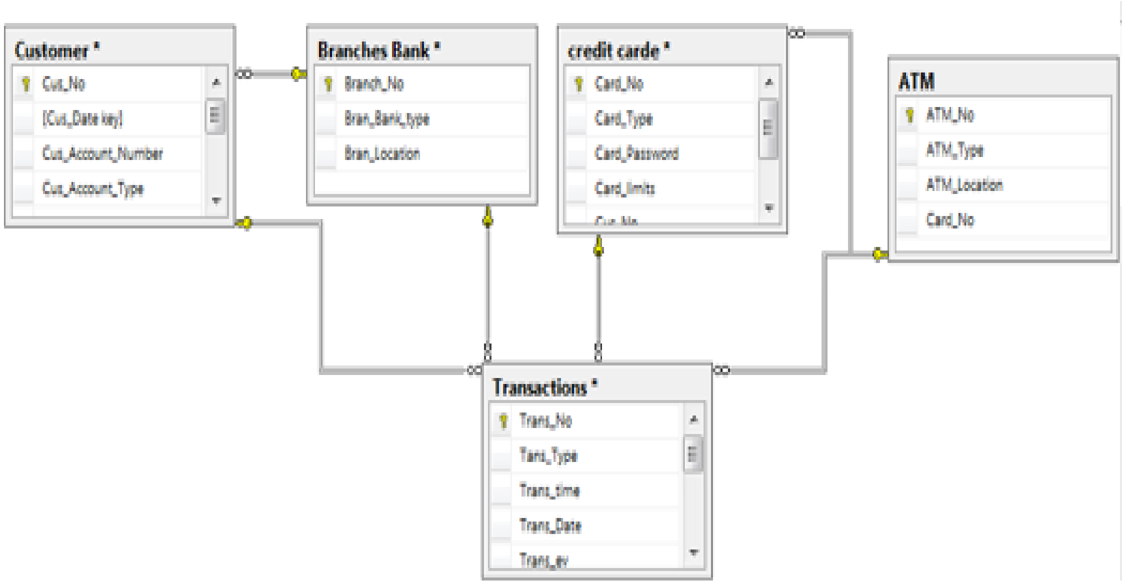

Figure 2. Transactions of ATM DB .

Customer table stores data about transactions with a CUS_NO primary key that has a single to multiple relationships with the transaction table. Credit card table stores data about customer's credit card using a primary key card no. that has a single to multiple relationships with transactions table. The ATM table stores data about ATM using a primary key ATM No. that has an indirect relationship with transactions. The Branches Bank table stores data about branches of the selected bank using a primary key Branch NO. that has a single to multiple relationships with transaction table.

1.2 Phase Two: Transformation Process

The data to be filled in data mart will be selected in order to be extracted from the data sources, SQL server database and text to the destination source MS SQL Server Data Transformation Services (DTS).

Data Transformation Services (DTS)can provide data access and custom transformation specifications. Therefore, the tables will be used for transforming. The data must be designed, cleaned, converted into a correct format, by then the corrected data shall be inserted into the data mart.

1.3 Phase Three and Four: Data Marts and DW Phase

Starting the data mart requires data loading firstly using all the historical data. Fact table is updated every time according to data mart.

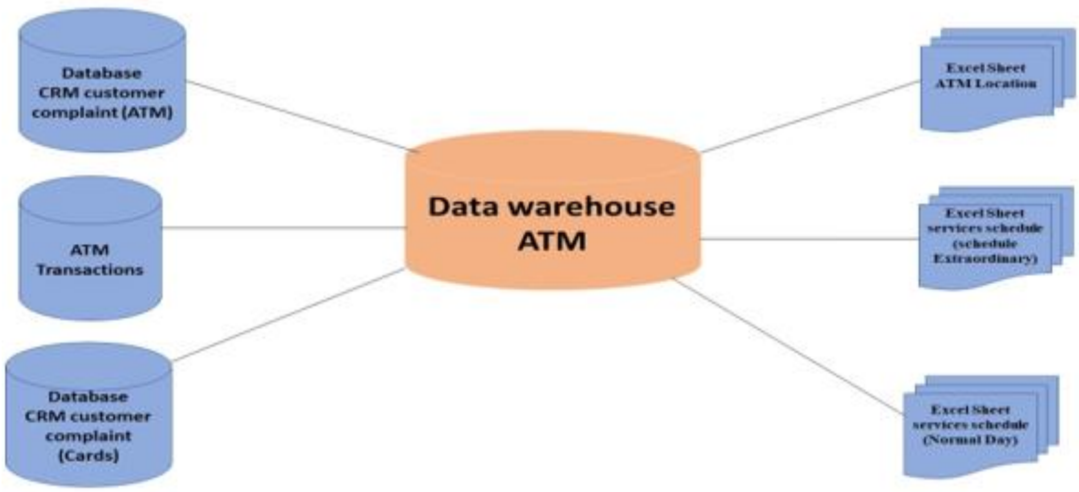

Figure 3. Operational Source Entity Relationships Diagram

1.3.1. Fact Table

Table for snow flow schema should be specified as shown in figure 3. The primary keys in fact table are Customer ID, Product ID and Order Date ID, that are made up of all of its foreign keys.

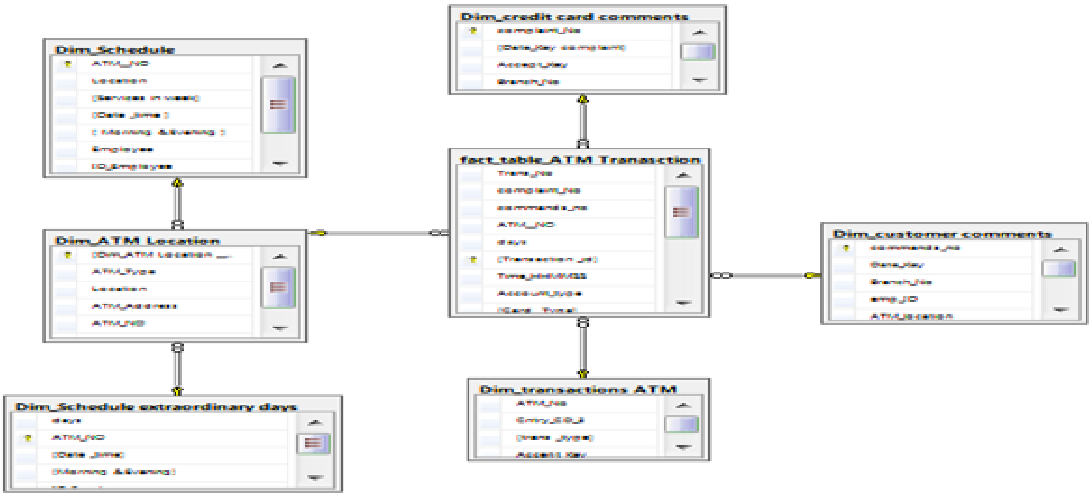

Figure 4. Snow Flake Schema

1.3.2. Dimension Tables:

There are six dimensions tables as shown in figure 3, transactions ATM dimension database, customer comments (ATM) dimension database, Credit card comments, dimension database and location and Schedule (Normal or extraordinary) dimension excel file.

1.4 Selection of Algorithms Phase

The selected algorithms depend on the dataset binary classification or multiclass. Accordingly, a comparison of two different classification algorithms, such as Naïve Bayes and Decision Tree, was made to evaluate the performance of the algorithms through several binary classification evaluation. Additionally, the performance of the algorithms by several multi-classification evaluation metrics highlights a difference in classification algorithms: Naïve Bayes & Decision Tree. Through the dataset was split into two parts, a training set which was 80% used to train the model of the actual dataset, remaining 20% it was used as a cross-validation to train the model. In order to evaluate the performance of the algorithms by several multiclass classification evaluation, in order to measure its accuracy, precision, recall, F-Measure, false positives rate (FPR) and true positives rate (TPR), which are defined by equations by means of a confusion matrix, and finally, Time taken to build model per second.

1.4.1 .Decision Tree by using C5.0:

R implementation of the supervised DM algorithm C5.0 that can generate a decision tree. This algorithm uses an information entropy computation to determine the best rule that splits the data, at that node, into purer classes by minimizing the computed entropy value[11].

Table 1: Comparison of several different ATM Number Classifiers using Decision Tree via C5.0Method

Classifier ATM.NO	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean	Accuracy	Kappa	Time (s)
1188181	1	1	0.9026136	0.9996554	0.9486593	0.5	0.9481805	0.8112	2
1188236	1	1	0.8905537	0.9997562	0.9420007	0.5	0.9392664	0.8116	3
2200001	1	0.9864	0.9732239	1	0.9864303	0.49045	0.9871307	0.9741	3
2200003	1	0.9733	0.9481029	1	0.9733602	0.48725	0.9745326	0.9489	3
202509700	0.7380	0.7654	0.7924528	1	0.8842105	0.27967716	0.872093	0.5571	3
259880800	0.8941	0.8859	0.8332882	1	0.9090641	0.396041595	0.9029921	0.7766	3

Table 1 presents more than ATM Number and the researcher implementation of Decision Tree using C5.0 .The result of it was approximate. Thereupon, the ATM Number of (2200001, 2200003 and 1188181) are more accurate in two ways of techniques.

The result of the technique presents the accuracy percentage between different ATM Number by using decision tree. In the following charts, the resolution percentage for each machine is displayed, where the machine 2200001 achieves the highest accuracy percentage of 98.70 % and the remained percentage in machine 2200001 is 1.30%. Thereafter, there is machine number 2200003 with an accuracy of 97.50% and the rest is 2.50%. Next, appears machine 1188181 for accuracy of 94.70% and the rest is 5.30%, then machine 1188236 for accuracy of 93.80% and the rest is 6.20%. Machine No 259880800 shows accuracy of 90.60% and the rest is 9.40%. At the end, the last machine 202509700 shows accuracy of 87.20% and the rest is 12.80%. The result of the technique is outlined in figure [5].

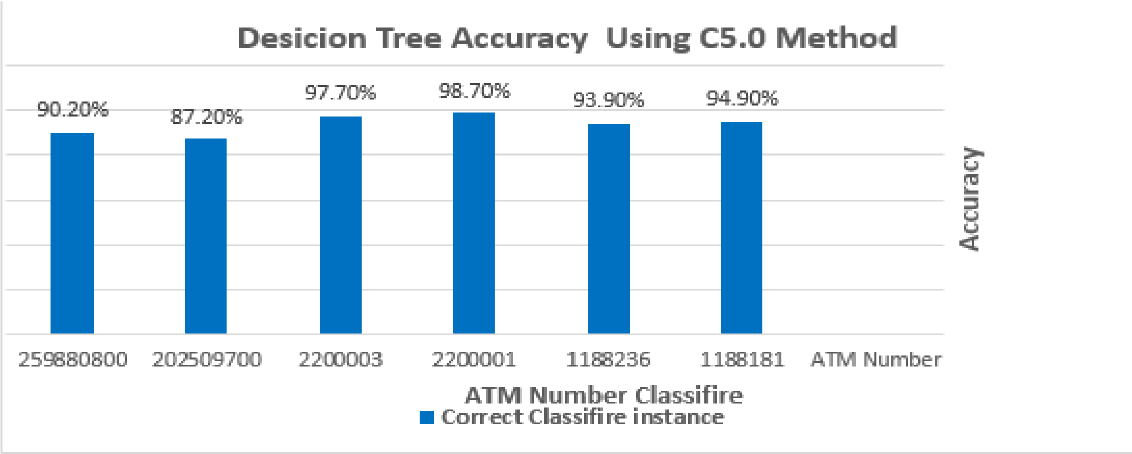

Figure 5. Accuracy percentage between different ATM Numbers by Decision Tree C5.0.

1.4.2. Decision Tree using Rpart:

R function (Rpart) is an implementation of the CART [Classification and Regression Tree] supervised machine learning algorithm, which is used to generate a decision tree. The R implementation is called r-part for Recursive Partitioning. [11] Similar to C50, rpart uses a computational metric to determine the best rule that can split the data, at that node, into purer classes.

Table 2: Comparison of several different ATM Number Classifiers using Decision Tree via Rpart Method

Classifier ATM.NO	Sensitivity	Specificity	Precision	Recall	F-Measure	G-mean	Accuracy	kappa	Time (s)
1188181	1	1	0.900829	1	0.9478275	0.5	0.9476855	0.7912	2
1188236	1	1	0.8881564	0.9995054	0.9405468	0.5	0.9385448	0.8101	3
2200001	1	0.9809	0.9746405	1	0.9871574	0.49045	0.9877413	0.9531	3
2200003	1	0.9745	2200003	1	0.9742937	0.48725	0.97515	0.9529	3
202509700	0.7308	0.7654	0.78	1	0.8764045	0.27967716	0.872093	0.8571	3
259880800	0.8941	0.8859	0.8377049	1	0.911686	0.39604	0.9064714	0.7766	3

Table 2 presents more than ATM Number and the researcher implementation of Decision Tree using R-Part. The result of it was approximate. Thereupon, the ATM Number of (2200001, 2200003 and 1188181) are more accurate in two ways of techniques.

The result of the technique, in the figure, shows the accuracy percentage between different ATM Number by using decision tree. In the following chart, the resolution percentage for each machine is displayed, where the machine2200001 achieves the highest accuracy percentage of 98.70% and the remained percentage in machine 2200001 is 1.30%. Thereafter, there is machine number 2200003 with an accuracy 97.50% and the rest is 2.50%. Next, appears machine no. 1188181 for accuracy of 94.70% and the rest is 5.30%, then machine 1188236 for accuracy of 93.80% and the rest is 6.20%. Machine no. 259880800 shows accuracy of 90.60% and the rest is 9.40%. At the end, the last machine 202509700 shows accuracy of 87.20% and the rest is 12.80%. The result of the technique is outlined in figure[6].

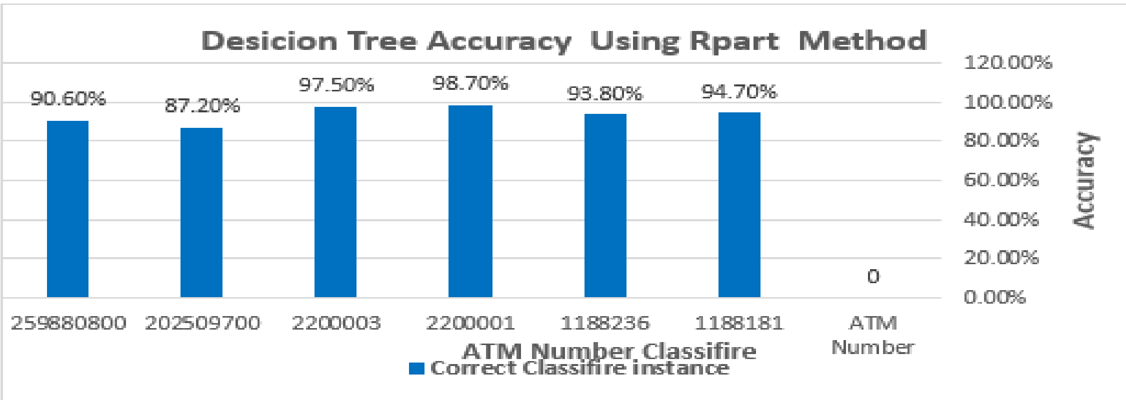

Figure 6 The Accuracy Percentage between different ATM Number using Decision Tree Rpart

1.4.3 Naïve Bayes

Classification algorithm for binary (two-class) and multi-class classification problems. Rather than attempting to calculate the probabilities of each feature value, it is assumed to be conditionally independent by looking at the given class value[11].

Table 4 Comparison of several different ATM Number Classifiers using Naïve Bayes Method.

Classifier ATM.No	Sensitivity	Specificity	precision	Recall	F-Measure	G-mean	Accuracy	Kappa	Time (s)
1188181	0.8599932	0.867342	0.8599932	0.867342	0.863652	0.99576357	0.8688	0.8636	2
1188236	0.8275136	0.7475137	0.8275136	0.7475137	0.7854819	0.94920724	0.7986	0.7854819	3
2200001	0.9781084	0.9717051	0.9781084	0.9717051	0.9748962	0.99671591	0.9766	0.9748962	3
2200003	0.9404319	0.8526027	0.9404319	0.8526027	0.8943662	0.95101644	0.9063	0.8943662	2
202509700	0.7867299	0.7904762	0.7867299	0.7904762	0.7885986	0.99763035	0.7949	0.7885986	3
259880800	0.8391738	0.8521982	0.8391738	0.8521982	0.8456359	0.9923583	0.8491	0.8456359	2

Table 3 presents more than ATM Number and the researcher implementation of the Naïve Bayes Method, which display that ATM Number of (2200001, 2200003 and 1188181) are more accurate. The result of the technique, in figure 8, can show the Accuracy Percentage between different ATM number using naïve bayes.

In the following chart, the resolution percentage for each machine is displayed, where the machine2200001 achieves the highest accuracy percentage of 97.60% and the remained percentage in machine 2200001 is 2.40%. Thereafter, there is machine number 2200003 for accuracy of 90.60% and the rest is 9.40%. Next, appears machine 1188181 for accuracy of 86.80% and the rest is 13.20%. Machine no.1188236 shows accuracy of 80.00% and the rest is 20.00%. The next machine 259880800 shows accuracy of 79.90% and the rest is 20.10% and last machine 202509700 has accuracy of 79.50% and the rest is 20.50%. The result of the technique is outlined in figure 7.

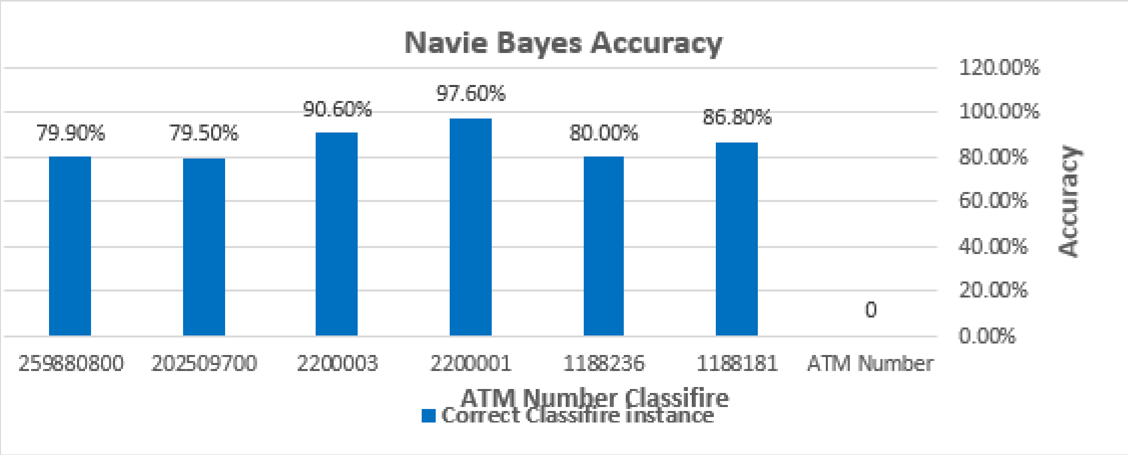

Figure 7 Accuracy Percentage between different ATM Number by Naïve Bayes.

Comparison Between the Two Algorithms

Table 5 Comparison of several different ATM Number Classifiers using Naïve Bayes Method.

	Algorithms		
ATM.No	**Decision Tree C.5**	**Decision Tree Apart**	**Naïve Bayes**
1188181	0.9481805	0.9476855	0.863652
1188236	0.9392664	0.9385448	0.7854819
2200001	0.9871307	0.9877413	0.97448962
2200003	0.9745326	0.97515	0.89436602
202509700	0.872093	0.872093	0.7904762
2598808800	0.9029921	0.9064714	0.8456359

In table4, concerning the results, proves that the best classification algorithm was decision tree algorithm, which had the highest accuracy of ATM numbers: (1188181) is 94.8%, (2200001) is 98.8 % and (2200003) is 97.6%. While the lowest accuracy classification was naïve bayes of ATM numbers (1188181) is 86.8%, (2200001) is 97.6 % and (2200003) is 90.6%, in a 2 second time taken. After the comparison between two data mining classifiers, the researchers chose the decision tree algorithm to show the best result.

Conclusion

Based on the propose model to measure the performance ATM machine , the model collect data from different sources that help data analysis;(directly and indirectly) and show the performance of machines then building a data warehouse can be described in four phases (data sources , Transformation Process, Data Marts and DW) to gather different types of data which were collected in one place and Choosing the appropriate algorithm from the two applied algorithms (Decision Tree or Naive Bayes) so after Compared algorithms to choose the best result(accuracy), decision tree C5 is selected as the high accuracy to use in analyzing the performance of machines.

Future Work

The future work will cover other problems of ATM that affect directly and indirectly in the machine's performance. These problems may include ATM's location and cash problems and machine maintenance as direct problems; and Electronic cards problems as indirect problem affecting machine performance.

Bibliography

[Aldajani,, 2009] Aldajani, Location of banking automatic teller machines based on convolution, Computers & Industrial Engineering, 2009. 57(4): p. 1194-1201., vol. 4, no. 57, pp. 1194-1201, 2009.

[Aboolian, 2007] R. O. B. a. D. K. Aboolian, Competitive facility location and design problem, European Journal of Operational Research, vol. 1, no. 182, pp. 40-62, 2007.

[Alhaffa, 2011] A. a. W. A. Alhaffa, A Market-Based Study of Optimal ATM'S Deployment Strategy, International Journal of Machine Learning and Computing, vol. 1, no. 1, p. 104, 2011.

[Armenise,2010] Armenise, A generative solution for ATM cash management,International Conference of Soft Computing and Pattern Recognition, no. IEEE, 2010.

[Habib, 2009] Q. a. S. Habib, Allocation of heterogeneous banks' automated teller machines, International Conference on Intensive Applications and Services, no. IEEE, p. 2009.

[Mathur S, 2021] L. G. S. P. P. Mathur S, Optimizing OLAP Cube for Supporting Business Intelligence and Forecasting in Banking Sector., Journal of Information Technology Management., vol. 1, no. 13, pp. 81-99, 2021.

[Souza ML, 2020] R. Souza ML, A survey on decision-making based on system reliability in the context of Industry, Journal of Manufacturing Systems, vol. 1, no. 56, pp. 133-56, 2020.

[Rajwani, 2017] A. e. a. Rajwani, Regression analysis for ATM cash flow prediction, International Conference on Frontiers of Information Technology (FIT), no. IEEE, 2017.

[Hasheminejad, 2018] S. a. Z. R. Hasheminejad, ATM management prediction using Artificial Intelligence techniques: A survey. Intelligent Decision Technologies, vol. 3, no. 11, pp. 375-398., 2018.

[Mwatsika, 2016] C. Mwatsika, Impact of ATM banking performance on customer satisfaction with the bank in Malawi.,Int J Bus Econ Res, vol. 1, no. 5, pp. 1-9, 2016.

[Wu, 2008] X. e. a. Wu, Top 10 algorithms in data mining., Knowledge and information systems, vol. 1, no. 14, pp. 1-37, 2008.

[Serengil, 2016] Ş. a. A. Ö. Serengil, Workforce optimization for bank operation centers: A machine learning approach., 2016.

[Minnen, 2006] D. e. a. Minnen, Performance metrics and evaluation issues for continuous activity recognition, Performance Metrics for Intelligent System, no. 4, 2006.

[Grabmeier, 2007] J. a. L. L. Grabmeier, Decision trees for binary classification variables grow equally with the Gini impurity measure and Pearson's chi-square test, International Journal of Business Intelligence and Data Mining, vol. 2, no. 2, pp. 213-226, 2007.

[Mwatsika, 2016] C. Mwatsika, "Impact of ATM banking performance on customer satisfaction with the bank in Malawi," Int J Bus Econ Res, vol. 1, no. 5, pp. 1-9, 2016.

[Bilir, 2018] C. a. A. D. Bilir, Optimization Of ATM And Branch Cash Operations Using An Integrated Cash Requirement Forecasting And Cash Optimization Model., Business & Management Studies: An International Journal, vol. 1, no. 6, pp. 237-255., 2018.

Authors' Information

Abeer A. Amer *– Associate professor Department of Computer sciences and Information Systems Department Sadat Academy for Management Sciences; e-mail:* abamer_2000@yahoo.com

Mohamed Ebied*– Researcher in Computer sciences and Information Systems Department Sadat Academy for Management e-mail:* ebied8111@gmail.com

Prof. Dr. / Nevine Makram Labib *– Professor of Computer Science and Information Systems Chair of Computer sciences and Information Systems Department Sadat Academy for Management Sciences ; e-mail:* nevmakram@gmail.com