Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7923058

ADAPTIVE HGB REGRESSION MODEL FOR REVENUE BASED ATM CASH DEMAND FORECASTING

T.V. Balakrishnan[1], R Kalaiarasi[2],
[1]Research Scholar, School of Computer Science, TNOU, Chennai, India
[2]Asst.Professor, School of Computer Science, TNOU, Chennai, India

ABSTRACT
Forecasting cash demand at Automated Teller Machines (ATMs) is one of the most difficult tasks in today's financial system. If an ATM runs out of cash, the bank's reputation will suffer, and the business will incur some expenditure as a result of the decreased usage of these devices by customers. Considering the usage of 'Any Bank ATM by Any Bank Customer', the volume of transaction increased multifold and cash availability at all the time is necessary. Banks earn revenue, whenever cross usage of bank's ATM take place. Looking into the customer service as well as revenue earning perspectives, banks are keen to have right cash availability. It is aimed to have the continued ATM service with neither a customer's transaction is refused due to a lack of cash, nor does idle cash jeopardize the bank's ability to make money from it. The revenue earning is considered one of the yardsticks of the ATM services and thereby A suitable regression machine learning approach to solve the ATM cash prediction model, with the revenue aspect, is taken up as part of emerging trends in computing. This article introduced the Adaptive HindGradiantBoosting Model Regressor (HGBRegressor). The Adaptive HGBRegressor may be utilized to deliver the best and most precise forecast on ATM transactions. That produces high prediction accuracy.
Keywords: *ATM, Adaptive HGB Regressor, Bayesian Optimization, Off-us transactions, hyper tuning, hyper parameters*

I. INTRODUCTION

Automatic Teller Machines (ATMs) are devices used by financial organizations to automate the provision of a range of services in a public place or region. The dispensing of cash from a user's corresponding account is one of the key services offered by these ATMs. Maintaining a seamless ATM network while dealing with a steady stream of transactions is very difficult for banks. Refilling an ATM with cash is known as ATM replenishment [1]. Customers must have a debit or credit card on file with the bank to access the account via ATMs [2]. The present financial institutions are confronted with addressing inventory and replenishment ideal policies when managing a large number of ATMs as the network of ATMs becomes denser and consumers visit them at a higher rate. A big holding cost will arise from an excessive ATM refill, but an insufficient cash inventory will increase the frequency of replenishments and the likelihood of stockouts, as well as customer displeasure [3]. If the entire quantity of money in the ATM exceeds the customer's demand, there will be unused cash and security concerns, and if the amount of money in the ATM is much less than the required, there will be client unhappiness [4]. If a bank relies on customer deposits for cash, it may recirculate deposits made at branches or use deposits held at its cash processing center. The transaction is originated by the customer at the ATM (Acquirer), using the bank issued debit / credit card (Issuer). If the acquirer and issuer banks are same then the transaction is called “On Us” and if different, it is “Off Us” transaction. In case of Off Us transaction, money belong to different bank than the bank issued the card, as per the Finance Regulatory (RBI), a fee is levied to the issuer to settle to acquirer, in addition to transacted amount. This additional revenue is also one of the factors in considering the ATM installation location, for ex. all Off Site ATMs (ATMs installed in malls or common places other than Branch Premises). Predicting ATM cash demand utilizes various parameters like historical trends, geographic location, seasonal considerations and Revenue generation. One of the most crucial jobs, however, is selecting the most efficient model for accurately forecasting an ATM's cash demand using the revenue model. Because each model has its own merits and drawbacks, no one model can always be deemed the best. Various issues, in particular, have different components and qualities that must be addressed independently determined [5]. Hence this paper proposed the ATM cash demand prediction based on the revenue yield, using the Adaptive HGB regression model.

II. LITERATURE REVIEW

In [6] the author suggested that the Service Oriented Architecture (SOA) may be used to design and deploy automated teller machine (ATM) services, which are useful for financial transactions. The use of Business Process Execution Language (BPEL) is used to integrate diverse services and orchestrate them. The development and validation of a suitable BPEL process are critical. In [7] the author provided mathematical assistance to the present system utilizing Data Analysis and Machine Learning (ML). The results of this study, as well as modeling methodologies, were used to a dataset obtained from a bank, which included data from 5 ATMs from 2011 to 2017. In [8] the author provided that the global or near-global optimum solutions, evolutionary and swarm optimization (ESO) based approaches play a critical role in resolving the aforementioned operational challenges. In [9] the authors suggested that the regression model, Autoregressive Integrated Moving Average (ARIMA) model, and Master Limited Partnership (MLP) neural network model are all used to forecast profitability volumes. In [10] the results imply that financial institutions may minimize loan losses, enhance security level, simplify conformance tasks, and enhance customer targeting by using Artificial Intelligence (AI). In [11] the author introduced a novel functionality centered on time-frequency economic status event 2-D reconstructions. Decision tree model is utilized as a technique, whereas genetic algorithms is employed for optimization. The created algorithm is put to the test on actual banking data, demonstrating the usefulness of the findings in a real-world setting. In [12] the author suggested that the fintech technology is used to power automated banking services. These technologies, on the other hand, provide the potential to improve low-income clients' financial inclusion indices. In [13] using the E-Service Quality model, the author investigated the impact of self-service technology quality characteristics, particularly ATMs, on user satisfaction. Technology optimism's effects on the mentioned previously connection are also studied. In [14] the authors analyzed that the scenario under consideration involves the protection of ATM, which is a sensitive issue in both banking and public security. Because of the essential concerns in this setting, even seemingly little improvements in surveillance system accuracy or responsiveness may make a significant difference. In [15] individuals try to falsify payments in this method. In 2019, card fraud is expected to cost the globe Rs 74.26 billion. As a result, individuals all over the globe are seeking for techniques to recognize suspicious transactions to find a solution to this issue. They are trying to solve this issue by exploiting machine learning faults that are dynamic. The study's main purpose is to investigate the relationship among services and the motivation to use self-service technology, with fundamental and staff assistance serving as regulators and disposition functioning as an intermediary [16]. The goal of this research is to precisely estimate the elimination of the posed double costs [17]. This paper presents a unique mathematical model that combines the two well-studied location and vehicle routing challenges for ATMs in the banking sector [18]. The goal of this research is to develop a model that can predict whether or not credit card users would pay off their obligations. Potential unpaid risks may be forecasted with high accuracy using the suggested methodology, and appropriate measures can be done promptly. They employ a support vector machine, a conventional artificial intelligence technique, to anticipate clients' payment status for the next months [19]. This research investigates the elements that influence people's preference for cash versus cards. A self-administered questionnaire was used to obtain both data in a cross-sectional survey. A convenience sample strategy was used to recruit 521 bank clients from one of India's leading banks for the research [20]. In [21] the author suggested that the aforementioned are implemented in the investigation by constructing the operator of the methods while taking into account the nature of the issue. The purpose of this project is to reduce the risk of PIN theft. They demonstrated a sixth sense technology that may be used to do traditional cash transaction activities utilizing a gestural interface instead of tapping the password registration number, which will help reduce theft connected with PIN monitoring. The testing sample model was constructed, and the results are verified and made available [22].

III. RESEARCH METHODOLOGY

(a) Research Approach: The study has the objective of prediction of cash demand of ATM based on the revenue generation out of off-us transactions. IDE.Jupiter or google colab is a web application used as an open source for sharing and creating documents that consists of equations, livecode, text and visualization narrated. This tool involved for data cleansing, data preprocessing like data transformation, value simulations modeling by statistics ML tools and data visualizations. Iterating with various regression algorithms with the base dataset to fetch out the best model and application of optimization algorithm suitable to the best model to tune the best model to produce the best results, has been the approach of the research work. This work extends to obtain the improved accuracy in predicting the cash demand in ATMs exactly based revenue generation along with the several labels acquired from collected datasets. In addressing the cash demand predict through regression analysis, essential flows are followed to arrive the model outcomes. Comparing the outcome of base model and the tuned model (called Adaptive HGB Regression) is done with the evaluating metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R- Squared value and adjusted R- Squared. By matching to test data set existing and predicted values, arrived to the error function value. The flow of the analysis is stated as below

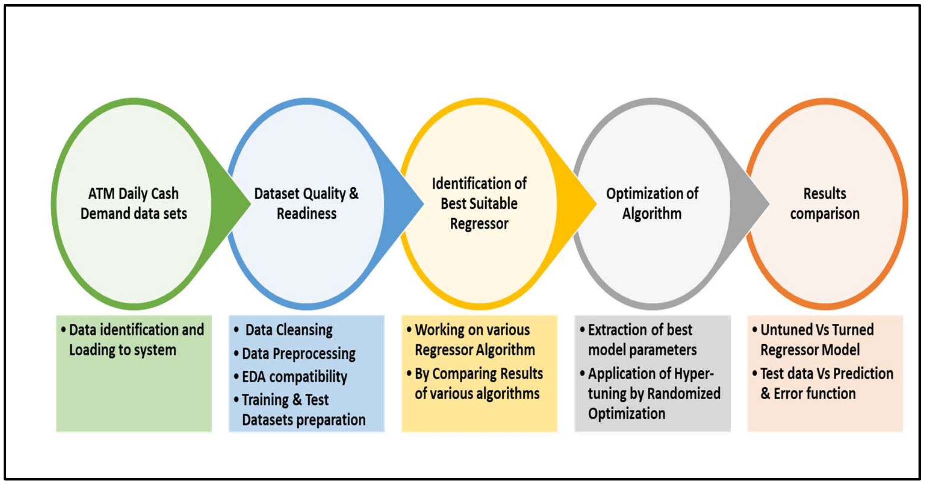


Fig. 1: Proposed architecture of ATM Revenue based Cash demand prediction

(b) Research Dataset Preparation: The **Data collection** or acquisition, is the daily total cash withdrawal from certain ATMs for six years[26], transactions involve more than 10500 records under 12 attributes, as database, considered for research analysis. The dataset contents to have various feature of the transactions, includes, own bank transactions, other bank transactions, their volume, total amount withdrawal, sequence of work nature, festivity indicator, holiday sequence. ATM name and Transaction Date drives the insights of the data. As part of **Data Cleaning,** necessary computational procedures are to be adopted to address the data with inaccuracy, ill structured & Missing of records, with duly best practiced statistical applications. The qualified mean, depicting patterns from days within the month and dates of months in the yea have been applied, as $x(m) = MED(y(m - 2r) \ldots\ldots y(m)\}(1)$, selective to day or dates. As part of **Data preprocessing**, transformations to effective datatypes and formats done using LabelEncoder and data points connectivity using EDA (Exploratory Data Analysis) are anlyzed on the datasets. The revenue or fee, additional income to the bank, is derived out of the translation of the off-us transactions, taken place in the ATM, in to fee amount. As per RBI, cross fund-based transaction will attract Rs.17/- & non-fund based Rs.7 per transaction, levied to issuer to settle to acquirer. Hence, the additional attributes, for the computed revenue ie minimum profit and maximum profit, are created and processed using the fee cost with the 'off us' transactions of the dataset. The maximum profit attribute is considered pivotal and predictable target. Some of dataset EDA reports on the transactional data vs years and ATMs vs Off-us transactions count, is as under

Fig. 2 Dataset share by Year

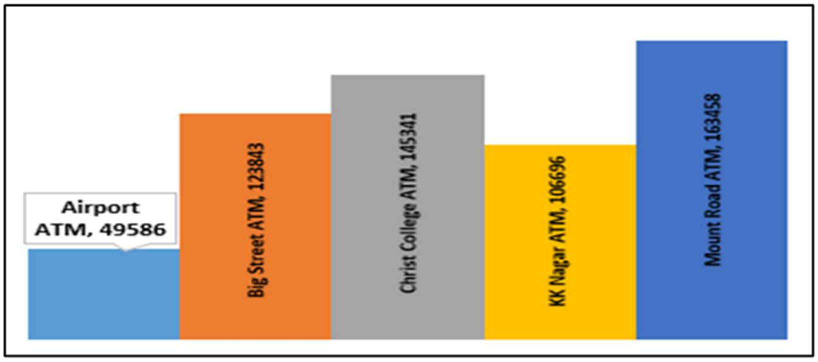


Fig. 3 ATMs to Off-us transactions count

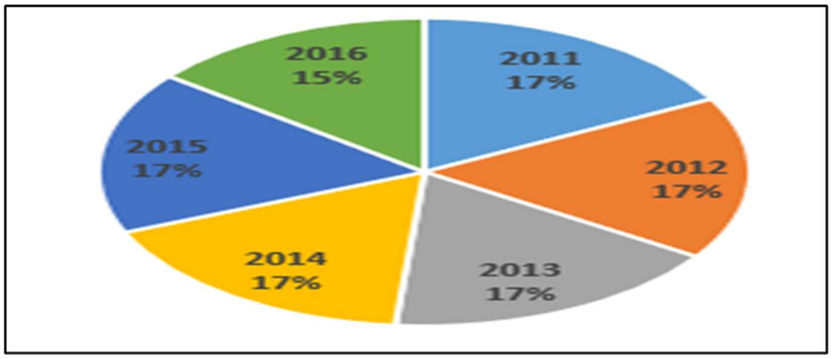
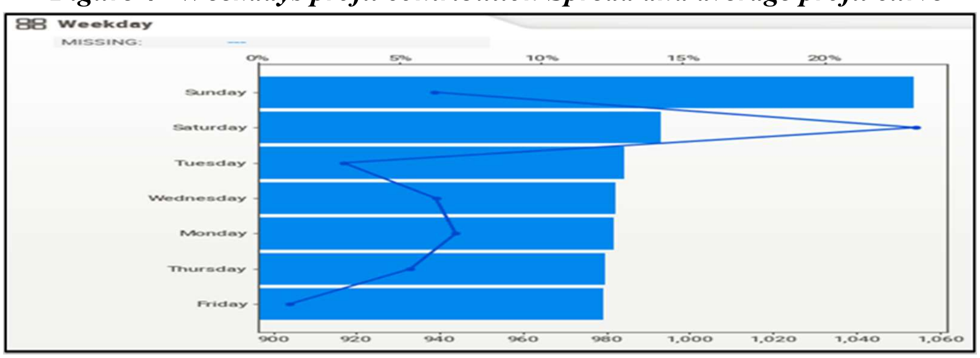


Figure 4   Weekdays profit contribution Spread and average profit curve



Different features of the datasets are valuated to find the relationship among using Correlation Matrix, which understand and reflect the relationship of data points.

(c) Research base model Identification: In order to identify to the suitable algorithm, the preprocessed data is set to have dataset segregation to training and testing to 80, 20 ratio. There are various regression algorithms exist in Machine Learning software, on which training data set is applied to get result effective model. This process is accomplished by a very standard and result promising algorithm process, 'lazypredict' in the supervised mode and produces the efficiency using metric yardsticks such as Root Mean Squared Error (RMSE), R- Squared value and adjusted R- Squared. It is found that Histgradiant Boosting Regressor (HGBRegressor) is the most suitable and effective algorithm model for the given dataset and its effective metrics are MAE: 111.8138, MSE: 23865.8667, RMSE: 154.4858, RMSLE: 5.0401, R-Squared: 0.9341, Adjusted R-Squared: 0.9338. The output of top value shown regressors metrics are shown below.

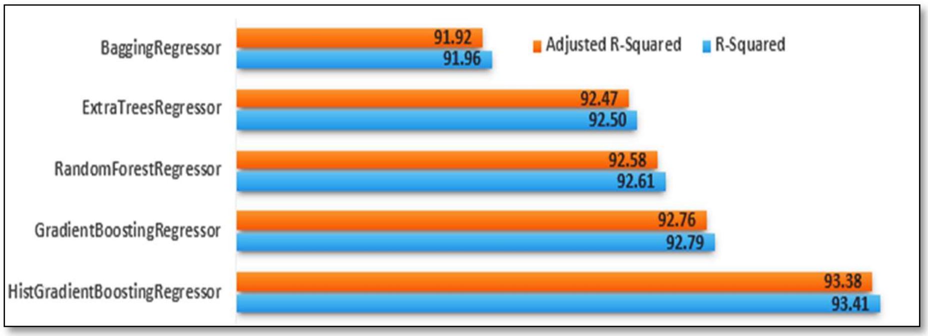


Fig. 5 Comparison of Top five regressors model performance

(d) Adaptive HGBRegressor by Hyper tuning with Bayesian Optimization (BO)

The base model HGB Regressor exhibits parameters which include but not limited to tree max depth, child weight, learning rate, estimators etc helps to further improvise the new nodes upon the tree architecture for optimization using Bayesian search approach. These model parameters can be given as base hyperparameter to BO shown in figure 7. This helps the gain function to accelerate the accuracy of model by appropriate feature to be used in splitting a node or not.

```
mod_1.get_params()

{'early_stopping': 'auto',
 'l2_regularization': 0.0,
 'learning_rate': 0.1,
 'loss': 'least_squares',
 'max_bins': 255,
 'max_depth': None,
 'max_iter': 100,
 'max_leaf_nodes': 31,
 'min_samples_leaf': 20,
 'monotonic_cst': None,
 'n_iter_no_change': 10,
 'random_state': None,
 'scoring': 'loss',
 'tol': 1e-07,
 'validation_fraction': 0.1,
 'verbose': 0,
 'warm_start': False}
```

Figure 6 Effective Model Parameters

Bayesian inference technique is typically applied to complex black-box functions in which all new assessment demands a significant investment in processing power. This is an effective in terms of the function evaluations necessary to accomplish the optimization. Equation 1 expresses the BO algorithm, which has been focused on problem solving. The objective function with only recognized values from same are used to have the hyper rectangular in BO which has no loss of generality and enhances the prediction process by Gaussian process GP prediction. Given a sample set is represented as $y = (y_1, y_2, \dots, y_n)^T \in \Omega^n$ and evaluation set is represented as $x = \left(f(y_1), f(y_2), \dots f(y_n)\right)^T$. GP context to focus on objective function actual value and interprets the evaluation vector y as a realization of X in the potential value of $x^* = f(y^*)$. The realization of a Gaussian random variable is an entire distribution is a multivariate Gaussian. When $\mathbb{C}_X$ signifies the covariance matrix and E_X represent the expectation vector, then the resultant Likelihood has the subsequent analytical formulation is expressed in equation 3.

$$L(\alpha, \theta_1, \theta_2, \dots, \theta_m \,|x_1, x_2, \dots x_n) = \frac{1}{\sqrt{(2\pi)^n |\mathbb{C}_X|}} . \exp\left((x - E_X)^T \mathbb{C}_X^{-1} (x - E_X)\right)$$

Hence, the hyperparameters evaluations have been acquired by exploiting the likelihood function L. Typical shaping of hyperparameter using BO is discussed in Algorithm 1. The k sub-index represent the variable state at iteration k and the instance GPk represent the model of Gaussian process at the kth iteration.

Algorithm 1 Hyperparameter using BO

Input = Iteration as k, loss function as f and loop count as n

Step 1: Fit the Gaussian process with the respective kernel as Gp_k to the data Y_k *and* X_k

Step 2: For k=1 to n do and select yk by random sampling.

Step 3: Compute the exact loss function $x_k \longleftarrow f(y_k)$

Step 4: If $x_k \leq x_{best}$ then

$y_{best} \leftarrow y_k, \quad x_{best} \leftarrow x_k$

End if

End for

Step 5: Maximize the infill criteria ICk over Ω to find the new iteration

$$y_{k+1} = {}^{argmax}_{y \in \Omega} IC_k(y)$$

Step 6: Evaluate f and set $x_{k+1} = f(y_{k+1})$

Step 7: Update the data of hyperparameter as

$$Y_{k+1} = Y_k \cup \{y_{k+1}\} \text{ and } X_{k+1} = X_k \cup \{x_{k+1}\}$$

Step 8: Repeat 5 to 7 till the y_{best} and x_{best} is attained

Step 9: Return

Fig. 7 Bayesian Optimization Algorithm

This Hyper tuned algorithm is named as ***Adaptive HistGradient Booster Model***, increases the learning performance of the model and helped in enhancing the accuracy of model by BO with HGBMRegressor as the tuned HGBMRegressor.

IV. RESULT AND DISCUSSION

This paper creates a mechanism for predicting the ATM cash demands of particular ATMs based on the revenue earning on the off us transaction in financial bank sector. We built a regression technique utilizing transaction history of revenue from each ATM. Adaptive HGBoost Regression, a machine learning approach, has a high prediction accuracy. Linear regression prone to a lot of noise and overfitting. Hence proposed method has efficient when compared to the all-regression algorithms and each ATM must be stocked with the appropriate quantity of cash based on the revenue contribution approach. The promising regressor HGBRegressor is identified by comparing various models, which is the best untuned ML model and the finer parameters form HGBRegressor is input to BO approach to have iterative improvement achieved as hyperparameter tuning. Statistical metrics are used to assess the performance of the predicted value using evaluation ERR metrics like Mean Absolute Error (MAE), RMSE, Mean Square Error (MSE), R2 and Adjusted R2.

Metric	HGBRegressor	Adaptive HGBRegressor
MAE	111.8138	112.5313
MSE	23865.8667	24132.8611
RMSE	154.4858	156.1224
RMSLE	5.0401	5.0911
R-Squared	93.4126	94.3389
Adjusted R-Squared	93.3846	94.3106

Fig. 8. Tuned Vs Untuned ERR Performance

The tuned HGBMReggresor, ie the adaptive HGBMRegressor is improving the best accuracy level of prediction in ML model thereby the cash demand prediction for a bank's ATMs, based on the Off-us transactions and thereon the revenue. The adaptive HGB regression model is compared on the test dataset towards the prediction and the same is shown below

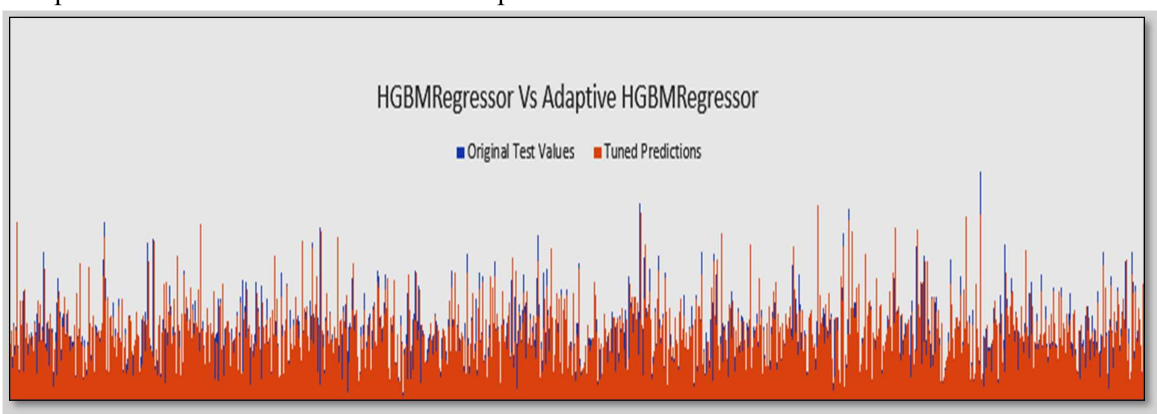


Fig. 9. Original Vs Predicted values

The efficiency of the prediction is attained by 94.34% model effectivity has a lower error percentage ie around 10% as the volume of cash transactions in ATMs are discrete in nature and prediction of demand to a reasonable level accuracy will make the Financial Institutions, in deploying money effective and efficient manner.

V. CONCLUSION

The accurate prediction of cash in ATMs is a necessary yet difficult operation. Considering present market driven agile approaches, financial institutions look more on Profits and in case of ATM channel, the additional revenue generation at acquirer is strength and facilitate to well establish the ROTI (Return On Technological Investments). To avoid consumer displeasure and money waste, money in ATMs should always be appropriately balanced. On ATM transactions, the Adaptive HGBoost Regressor may be used to provide the best and most exact prediction. The reduction in RMSE, MSE, and MAE in this article demonstrates the improvement in accuracy. The suggested approach (Adaptive HGBR) has a 94 percent accuracy level. The ATM cash prediction system simulation demonstrated effectiveness, suggesting that this AHGB Regression (proposed) is capable of generating trustworthy rules that may enhance existing cash management. The scope of expansion of the project is there to further analyze the non fund based revenue generation vs fund based revenue generation as ATM services include various transactions other that cash dealings.

Reference

1. Rafi, M., Wahab, M.T., Khan, M.B. and Raza, H., 2020, January. Atm cash prediction using the time series approach. In 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.
2. SingaporeDutta, M., Psyche, K.K., Khatun, T., Islam, M.A. and Islam, M.A., 2018, August. ATM card security using bio-metric and message authentication technology. In 2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET) (pp. 280-285). IEEE.
3. Ekinci, Y., Serban, N. and Duman, E., 2021. Optimal ATM replenishment policies under demand uncertainty. Operational Research, 21(2), pp.999-1029.
4. Jadwal, P.K., Jain, S., Gupta, U. and Khanna, P., 2019. Clustered support vector machine for atm cash repository prediction. In Progress in Advanced Computing and Intelligent Engineering (pp. 189-201). Springer,
5. Fallahtafti, A., Aghaaminiha, M., Akbarghanadian, S. and Weckman, G.R., 2022. Forecasting ATM Cash Demand Before and During the COVID-19 Pandemic Using an Extensive Evaluation of Statistical and Machine Learning Models. SN computer science, 3(2), pp.1-19.
6. Nath, A. and Rath, S.K., 2021. Service oriented architecture-based design of bank-ATM and its verification with Petri-net. International Journal of Simulation and Process Modelling, 16(4), pp.343-352.
7. Chavan, V.V., Manjaly, J.J. and Ali, M.A., 2021, September. Automated Teller Machine Cash Demand Prediction. In 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) (pp. 142-147). IEEE.
8. Krishna, G.J. and Ravi, V., 2019. Evolutionary computing applied to solve some operational issues in banks. In Optimization in Industry (pp. 31-53). Springer, Cham.
9. Dadteev, K., Shchukin, B. and Nemeshaev, S., 2020. Using artificial intelligence technologies to predict cash flow. Procedia Computer Science, 169, pp.264-268.
10. Königstorfer, F. and Thalmann, S., 2020. Applications of Artificial Intelligence in commercial banks–A research agenda for behavioral finance. Journal of behavioral and experimental finance, 27, p.100352.
11. Ketenci, U.G., Kurt, T., Önal, S., Erbil, C., Aktürkoğlu, S. and İlhan, H.Ş., 2021. A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering. IEEE Access, 9, pp.59957-59967.
12. Correa, J.C., Dakduk, S., van der Woude, D., Sandoval-Escobar, M. and Lopez-Llamasa, R., Low-Income Consumers' Disposition to Use Automated Banking Services.
13. Othman, A.K., Hamzah, M.I. and Hassan, L.F.A., 2020. Modeling the contingent role of technological optimism on customer satisfaction with self-service technologies: A case of cash-recycling ATMs. Journal of Enterprise Information Management.

14. Rossi, A., Rizzo, A. and Montefoschi, F., 2018, September. ATM Protection Using Embedded Deep Learning Solutions. In IAPR Workshop on Artificial Neural Networks in Pattern Recognition (pp. 371-382). Springer, Cham.
15. Thakral, M., Singh, R.R., Jain, A. and Chhabra, G., 2021, October. Rigid Wrap ATM Debit Card Fraud Detection Using Multistage Detection. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 774-778). IEEE.
16. Guan, X., Xie, L., Shen, W.G. and Huan, T.C., 2021. Are you a tech-savvy person? Exploring factors influencing customers using self-service technology. Technology in Society, 65, p.101564.
17. Poorzaker Arabani, S. and Ebrahimpour Komleh, H., 2019. The improvement of forecasting ATMS cash demand of Iran banking network using convolutional neural network. Arabian Journal for Science and Engineering, 44(4), pp.3733-3743.
18. Nazari-Ganje, N. and Mirzapour Al-E Hashem, S.M.J., 2020. An integrated location-inventory routing problem for ATMs in banking industry: a green approach. In Modeling and Optimization in Green Logistics (pp. 27-52). Springer, Cham.
19. Yontar, M., Dağ, Ö.H.N. and Yanık, S., 2019, July. Using Support Vector Machine for the Prediction of Unpaid Credit Card Debts. In International Conference on Intelligent and Fuzzy Systems (pp. 377-385). Springer, Cham.
20. Nambiar, B.K. and Bolar, K., 2022. Factors influencing customer preference of cardless technology over the card for cash withdrawals: an extended technology acceptance model. Journal of Financial Services Marketing, pp.1-16.
21. Danaci, E., Alkaya, A.F. and Gültekin, O.G., 2019, July. An Empirical Analysis of Swarm Intelligence Techniques on ATM Cash Withdrawal Forecasting. In International Conference on Intelligent and Fuzzy Systems (pp. 1235-1242). Springer, Cham.
22. Kumar, S.P. and Shanmugasundaram, N., 2018. Pin number theft recognition and cash transaction using sixth sense technology in ATM/CDM. International Journal of Engineering and Technology, 7(2), pp.178-180.
23. Vangala, S. and Vadlamani, R., 2020. ATM Cash demand forecasting in an Indian Bank with chaos and deep learning. arXiv preprint arXiv:2008.10365.
24. Ghasemi, Z., Afshar Kermani, M. and Allahviranloo, T., 2021. Exploring the Main Effect of e-Banking on the Banking Industry Concentration Degree on Predicting the Future of the Banking Industry: A Case Study. Advances in Fuzzy Systems, 2021.
25. Ekinci, Y., Serban, N. and Duman, E., 2021. Optimal ATM replenishment policies under demand uncertainty. *Operational Research*, *21*(2), pp.999-1029.
26. Dataset availability from Kaggle site URL: https://www.kaggle.com