A Long-Short-Term-Memory Based Model for Predicting ATM Replenishment Amount

Muhammad Asad
Department of Computer Science FAST National University of Computer and Emerging Sciences
Karachi, Pakistan
asad.shamshad@gmail.com

Muhammad Shahzaib
Department of Computer Science FAST National University of Computer and Emerging Sciences
Karachi, Pakistan
Muhammad.Shahzaib@outlook.com

Yusra Abbasi
Department of Computer Science FAST National University of Computer and Emerging Sciences
Karachi, Pakistan
abbasi.yusra14@gmail.com

Muhammad Rafi
Department of Computer Science FAST National University of Computer and Emerging Sciences
Karachi, Pakistan
Muhammad.Rafi@nu.edu.pk

***Abstract*— The advent of Automatic Teller Machines (ATMs) enable self-service, time-independent, easy to use, mechanism through which a financial institution supports large number of services to its users. Cash withdrawal from the ATM is still one of the major transactional loads for these networks. ATM cash replenishment is the process by which ATM machines are filled with the cash so that the users can withdraw it. The rapid adaptation and standardization of these network give rises to many challenging problems that requires intelligent management of these resources. ATM cash replenishment amount prediction is one such problem, predicting the right amount for everyday use such that the minimum amount of cash always available before the next replenishment. In this way there will be no customer dissatisfaction through empty ATM. The paper proposes a machine learning approach to ATM replenishment amount prediction by using a data driven approach for the estimation of right amount for each ATM or some group of ATMs. The data comprises of replenishment of 2241 ATMs for last 22 months from 6 different Banks of Pakistan. The Long Short-Term Memory (LSTM) based model produce Root Mean Squared Error (RMSE) of 132.53, which is quite encouraging for this problem.**

***Keywords*— LSTM, Neural Networks, Regression, Time Series, Cash Prediction.**

I. INTRODUCTION

A. Problem Background

Automatic Teller Machines (ATMs) are the devices financed and managed by financial institutions to provide facility to humans to make transaction in public places. Banks usually refill ATMs per day and this cash replenishment is manual till now, cash in ATMs is kept up to 40 percent more than its real need, This excess amount leads to business loss as that amount can be used by Banks for investments and various purposes that can generate revenue for them, therefore we have developed automatic forecasting mechanism to make cash replenishment prediction more accurate to avoid business loss and customer dissatisfaction. To develop this model, we used real data of 2241 ATMs from 22 months and to make prediction much accurate we have taken into account seasonality effects as well as location effect, we have used historical data of ATM to predict future demand through Machine learning approach.

B. Problem Statement

Given the data generated at an ATM for a period of *'T'* days we want to predict the amount required at some specific *'d'* day for upcoming transactions.

$$f(A_i) = \sum_{t=i}^{m} x \qquad (a \leq x \leq x_i)$$

where

A_i = i th ATM, for which cash will be predicted

x = Predicted cash needed for i th ATM

x_i= Upper limit for predicted amount

a = Actual cash needed

t = Start of cycle

m = Last day of cycle

C. Multivariate Model

Fig. 1 The Multivariate Model demonstrates the overall process of how data is analyzed to make predictions and the flow of steps model follows in real-time. The model works in a way that it obtains the data generated by ATM. After fetching, pre-processing is done on data which include data cleaning, data analysis, interpolation of cash, scaling, feature engineering and clustering i.e. creating exogenous inputs and calculating cash withdrawal. The data has been prepared to perform training and development and after tuning parameters and getting best weights, the model is evaluated and ready to forecast cash replenishment demand.

II. LITERATURE REVIEW

Numerous authors had done work on the problem of cash forecasting. The existing literature and the work done about

978-1-7281-8855-3/20/$31.00 ©2020 IEEE

ATM cash prediction is divided into the sections as I-0, II-B, II-C, II-D, II-E, II-F, II-G, II-H, and II-I based on the different techniques and approaches.

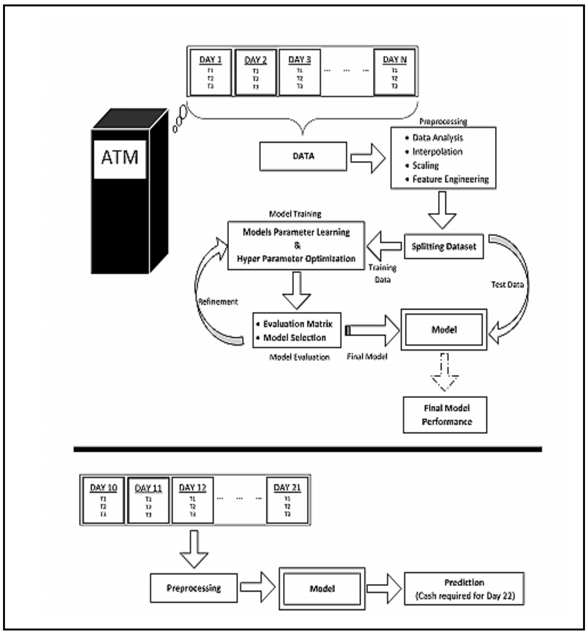


Fig. 1: Multivariate Model

A. *ATM Cash Management Prediction using Artificial Intelligence Techniques:*

The less amount of work done and surveys conducted in the past related to ATM cash management and prediction motivated [1] to conduct a survey on ATM management prediction and its related issues i.e. forecasting cash demand, fraud detection, ATM failure, user interface, replenishment strategy, ATM location, customer behavior, etc. All the aspects tried to be covered and mainly divided into three sections/studies. First, research on data sets and according to different data sets the ATMs were classified into 7 branches of different behaviors as per data sets. Second, all Artificial Intelligence techniques used to achieve ATM cash prediction in past were summarized. Third, the issue counter which has been countered by many researchers that there is no point of comparison as each and every research is different from other in terms of data set, features, techniques, and evaluation metrics. There is no benchmark data set available for this problem.

B. *Cash Demand Forecasting for ATM's Using Different Financial Modelling Techniques:*

The nature of cash withdrawal from ATMs is unpredictable due to seasonality effects. It is hard to make optimized model considering unexpected cash withdrawal throughout the year. Around 40% excess of cash placed in ATMs to avoid cash-out-of-flow situations. [2] conducted a survey on Cash Demand Forecasting related techniques which encompasses two modelling techniques i.e. Support Vector Regressor and Artificial Neural Network and which one give the optimized results. It found out that ANN is complex and hard to understand as compared to SVR which is more feasible in practical situations which is weekly withdrawal with 97.72% accuracy, but ANN has more space in future work which can lead to better cash management due to its nature of learning the behavior of cash withdrawal.

C. *Cash prediction Using Machine Learning:*

To overcome the problem of cash-related cost in ATM's [3] used machine learning approach i.e. Support Vector Regression and for training it, the parameters use mapping function, cost of error C and the width of the insensitive tube, data records of 15 ATMs from past 2 years were used and then they were used to forecast the daily cash demand for every ATM. [4] also used machine learning approach, he used Linear Regression, Ridge Linear Regression, Experiment with Lasso, Time series Regression and LSTM and when compared all of them he found 98 percent accuracy in Linear Regression.

D. *Cash Prediction Using Neural Network:*

Cash prediction is affected by some nonlinear factors like seasonality and location. Taking into account these factors flexible ANN (Artificial Neural Network) was used by [3] For every ATM he used three-layer feed forward neural network which was trained using Levenberg Marquardt optimization method and RMS (root mean square) error between predicted and 3 real value. In addition to Neural Network approach [4] used LSTM which is recurrent neural network which doesn't produced the desired results as compared to the approach discussed in section II-C where Linear Regression produced a better result compared to LSTM.

E. *Cash Prediction Considering Seasonality Effect:*

[5] used a hybrid approach using time series data and rules derived from the trends observed. Considering the data and region i.e. Middle East, they focused on the festive seasons' withdrawals, weekends withdrawals, and 10th of every month which is believed to be the pay day of employees. After training models on 2 years of data it was found that the cash predicted by the Neural Network managed to save 26 percent and SVR (Support Vector Regression) save 25 percent as compared to the cash used in past replenishment.

F. *Cash Prediction Through Clustering:*

Cash demand in ATMs is affected by location i.e. ATMs situated in crowded area or near markets have high cash demand, therefore [6]. demonstrated an approach based on clustering that cluster nearby ATMs and forecast group demand. Similarly, seasonality effects were considered by Clustering ATMs based on similar day of week pattern where each ATM cash withdrawal time series is translated into day-of-the-week seasonality sequence and then this sequence is discretized. The results after clustering turned out better when compared to individual forecast.

G. *Forecasting of Cash Flow in ATM Based on Cubic Exponential Smoothing Method:*

As cash in ATM belongs to interest free funds, the efficient amount needs to be replenished in an ATM to minimize the loss faced by banks, as extra cash in ATM will be a waste of resource. [7] tried to forecast optimize cash requirement and made the surplus of cash required in ATM less. To achieve this, smoothing parameter of cubic exponential smoothing forecast model. Three exponential smoothing forecast model were used to predict the cash which minimized the prediction error.

H. Improvement of Demand Forecasting Models with Special Days:

Forecasting less cash results in decrease of customer satisfaction, but if it is forecasted too high then the bank will face huge loss resulted in lost interest. [8] targeted the data set such as special days in UK to enhance the performance of forecasting models used earlier. Their studied found out that 19 special days in UK has significance effect on cash withdrawals. By enhancing data through this study, they achieved 21.57% SMAPE for forecasting.

I. A Flexible Neural Network for ATM Cash Demand Fore-casting:

Neural Networks has great applicability when it turns out to learn on large data sets which make it generalize for future predictions. [9] trained a neural network consist of three-layer feed-forward with Levenberg-Marquardt algorithm on the past data of ATM. The past data was updated after every week with ANN predictions. Also, regularization and generalization in data was done to make ANN efficient and saved it from overfitting which gives 5-10% error on stimulated data and for real data it gives 25-30% error.

III. DATASET

The data we have is replenishment level data of 2241 ATMs from last 22 months from 6 different Banks of Pakistan. The data consist of 609,932 rows. The data is consisting of the amount replenished in ATMs or in other words the count of notes added in respective cassettes at different timestamps. The modern ATMs are consisting of four cassettes for different value of notes. So, the data consist of how many notes added and left in these cassettes after one complete replenishment, status of ATM machine, start and end of replenishment period. The extracted data is composed of 29 columns such as Notes added/left, ATMID (ID of ATM), Start Time and End Time of replenishment, Event ID, Status, etc. Each column gives important information for prediction and data handling. Event ID and Status column used to find the insights in data where rows in connected table were missing.

A. Preprocessing

In this section we have covered the exploratory data analysis and preprocessing of data. After exploring the data, we have found out outliers, missing values and other problems caused by the replenishment process convention, which includes the replenishment period. The Replenishment period vary from one day to several weeks for ATM to ATM and also with in individual ATM. The problem varying replenishment period, outliers in notes were focused first to proceed further.

In order to find the outliers in the data we used Boxplot to identify outliers as the number of notes and maximum capacity of a cassette is not given. After identifying the outliers, we first convert them to 'NaN' then interpolate them. Converting the outliers to 'NaN' lead to lose of information if typical interpolation techniques were used. Instead of linear or quadratic interpolation we interpolate them by the average of cash withdrawal of every month at that day as it gives better results compared to quadratic interpolation.

There were some other important factors which affect the replenishment such as which day of week it is? is it a salary week? is it end/start of month? is today holiday? is it weekend? In order to counter these, we synthesized new features which were Weekdays (7 Weekdays-Bit), Salary Week (0-1) Holiday (0-1), Weekday (0-1), Weekend (0-1). In the data of ATM event there was information about events occurred in an ATM which helps in differentiate from ATM out-of-cash and ATM out-of-service.

IV. PROPOSED APPROACH

In this section, we have covered the problem of ATMs having varying replenishment period and cash withdrawal which are important to address before proceeding towards prediction of ATM replenishment amount.

A. Clustering of ATMs

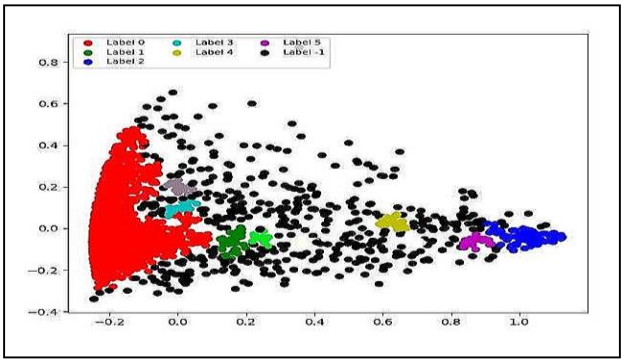
As discussed, that the data set is consist of 2241 ATMs from every region of Pakistan. So, it was an important task to cluster the ATMs on basis of similarities to extract relevant data.

In order to achieve this, we engineered three features using central tendencies of the data to cluster the ATMs. The important characteristics of ATMs were based on how much cash withdrawal from the ATM? How many times it got replenished in a month? How many days it takes to complete one cycle of replenishment. Considering these similarity measures we created the following features for each ATM:

- Average monthly cash withdrawal
- Average number of times an ATM got replenished in one month
- Average number of days an ATM took to be replenish again

We calculated mean for monthly cash withdrawal and number of times ATM replenished in one month, but for number of days taken by ATM to get replenished we took the median of each month and then calculated the mean of them. As due to ATM out of service error, some ATMs average disturbed so median represent the best value for this feature. On the basis of these three attributes we clustered our data set using Density Based Spatial Clustering (DBSCAN). Considering our data set, the DBSCAN produced 10 clusters as per their properties. The cluster with highest number of ATMs and which was needed to perform operations was selected. The Fig.2 below shows ATMs in 2D and different color shows the clusters to which the ATMs belong.

Fig. 2: ATMs in 2D and result in clustering



The cluster consisting of red points were identified as 'high demand in less time'. This cluster was based on 1293 ATMs with average monthly replenishment of 28 days and average time taken by ATM for one replenishment cycle of 1.5 days. The black points are those points which failed to be a part of any cluster.

C. *Neural Networks*

Neural networks are one of the most widely used technique for time series forecasting. It's capability to learn weights based on the historical data fed to the network produced more accurate results for the future. We used NN for predicting ATM replenishment amount using the cash withdrawal as a target variable. In order to predict we used Dense Neural Network and it maps the input parameters on the desired Cash Withdrawal variable. The Dense Neural Networks works in a way that output of each neuron is fed to all the neurons in the next layer making it a dense network passing values which make it perform well in learning the weights and predicting the values.

The weights assigned to each feature is based on the importance and effect of that variable on data as data is more sensitive to some features such as salary week, weekend, and holidays.

Dataset of starting fifteen and a half month which is 70 percent were used to train the neural networks and the remaining 30 percent which is six and a half months were used for testing purpose.

The 'high demand in less time' cluster used to apply neural network. The simple dense neural network produced 199 RMSE on training and 196.5 RMSE on testing data.

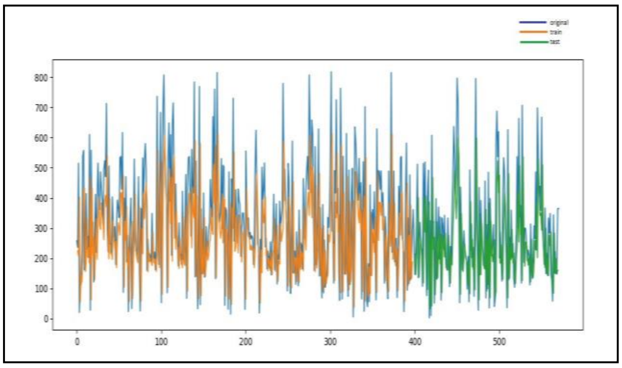
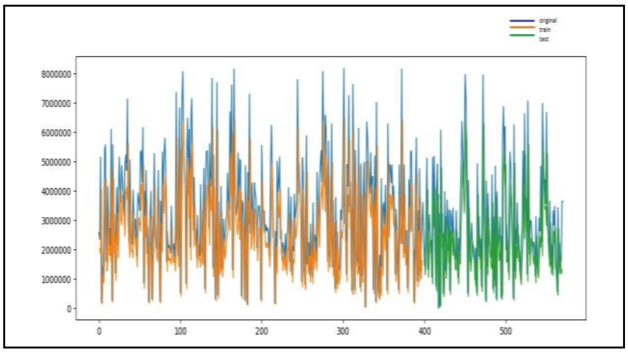


Fig. 3: Results with Dense Neural Network

In Fig. 3 the blue line indicates the original, orange line shows the training and green line shows the testing data. It can be seen the NN is struggling in predicting the extreme values both maximum and minimum of the series.

The sliding window method is used to enhance the performance of Neural Network. We used the look-back value of 7 which means to look back the past 7 values to predict the next day i.e. $t_i = f\,(t_i{-}1,\ t_i{-}2,\ t_i{-}3,\ t_i{-}4,\ t_i{-}5,\ t_i{-}6,\ t_i{-}7)$.

Fig. 4: Results with Dense NN and Look-back-7

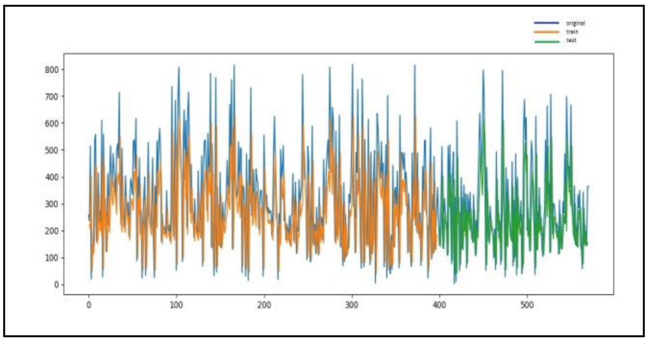


In the Fig .4 graph blue denotes actual series, orange denotes prediction on training set, green denotes prediction on testing set. Where as t_i is the current day. Root Mean Square Error (RMSE) were used as a loss function as it penalized more and make learning faster to reach the optimal solution as compared to Mean Absolute Error. The NN with sliding window method reduced the RMSE to 179.1 on training and 177 RMSE on testing data. In Fig. 4 it can be seen that NN started to predict the minimum and maximum values.

C. Long-Short-Term-Memory - LSTM

We have used LSTM for solving the ATM replenishment amount prediction problem. LSTM is a Recurrent Neural Network which used back propagation for learning. LSTM is based on memory blocks which contains different gates using sigmoid and tanh function to assigned weights instead of using neurons. Sigmoid function used as gating. Function that decides which information to retain or lost by assigning a value (0-1). A masking layer is used to counter missing dates in data. The masking layer assigned 0 to missing dates and LSTM does not take those masked values in consideration while learning. Function that decides which information to retain or lost by assigning a value (0-1). A masking layer is used to counter missing dates in data. The masking layer assigned 0 to missing dates and LSTM does not take those masked values in consideration while learning.

Fig. 5: Results with LSTM



In Fig. 5 graph blue denotes actual series, orange denotes prediction on training set, green denotes prediction on testing set.

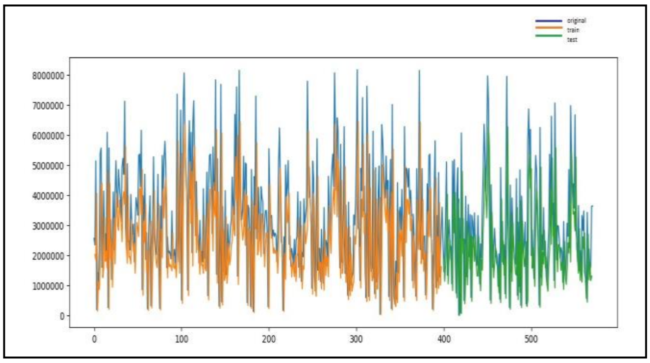


Fig. 6: Results with LSTM and Look-back-7

The RMSE loss function is used with the LSTM as mentioned in the above window. After several iterations with different combinations we got 181 RMSE on training data and 179 RMSE on testing data. In Fig. 5 the LSTM struggle to capture the min. and max. same as Dense Neural Network. Sliding window with a look-back value of 7 days is used with the LSTM for enhancement. The LSTM with 7 look-back values reduced the RMSE of training data to 133.5 and of training data to 132.53. In Fig. 6 the LSTM able to learn predict more accurately.

V. EXPERIMENTAL STUDY

There are multiple architecture of Neural Network (NN). Neural network is a powerful tool in the world of artificial intelligence, but the traditional neural network does not track the previous observation while predicting future value of series. We have used both traditional neural networks also known as Vanilla neural network or Dense neural network with sliding window approach. In time series prediction considering previous values give great advantage to the learning model. Here comes the Recurrent Neural Network also known as RNN to help.

Recurrent Neural Network retains the information of the previous values of the series as we move to the next one. It is represented as the loop over the Neural Network in graphs. Recurrent Neural Network is nothing but consecutively connected neural network in which each neural network passes the result to the next neural network.

The architecture of Recurrent Neural Network we used is called Long-short Term Memory (LSTM). LSTM is widely used with time series data. It also solves the vanishing gradient problem. Unlike neural networks, the neuron in LSTM is called memory block. It is composed of two main components. First is memory, which is also called internal state or cell state. It keeps the memory of the previous value of the sequence. Second are gates also called regulators which decides whether or not to activate the memory block.

There are three gates in the common LSTM Memory block:

- Input Gate: decides whether to updating memory or not
- Output Gate: decides what should be the output of the memory block
- Forget Gate: decides whether to erase the memory or not

At input gate there is a sigmoid layer and a 'tanh' layer whereas the sigmoid layer chooses what to update in the cell memory and tanh performs the update.

At output gate is the sigmoid layer followed by a 'tanh' layer, it decides what will be the output of the LSTM memory block. the output is based on the cell state filtered by the sigmoid layer

At forget gate, there is also a sigmoid gate. It decides which part of the cell state is to be erased and what to be retain till the next iteration. In Window Method previous values of the sequence is provided as a feature. For example, if we predict the y_t value of the sequence. we will provide y_{t-1}, y_{t-2} .... y_{t-n} where the value of n is called Window size. Window size is also called Look back, which denotes the timestep the model will look back for predicting the next value. For dense neural network we provide the previous values as the input feature. In case of sliding window approach combined with neural network, it is required to pass specific numbers of previous values to the model that is window size, which is not possible in every case. Whereas the use of LSTM solves that problem as the architecture of LSTM is robust to such dependencies.

VI. EVALUATION MEASURE

We have used root mean squared error (RMSE) for evaluation of the trained models. It measures the squared sum of variation of prediction value from expected result. RMSE is usually used for evaluation of the regression problems like climatology and forecasting [10]. It measures the spread of samples from the line of best fit. RMSE is the square root of sum of errors as expressed in the formula.

$$RMSE = \sqrt{\frac{\sum_{t=-1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

where,

$\hat{y}_t$= predicted value for the t^{th}sample.

y_t= Actual value for the t^{th} sample.

T = Total Samples.

Best RMSE score is zero in ideal case where as in real prediction of regression problem can never be absolute correct. The worst score can reach infinity based on how far is the samples from the actual value. RMSE is also preferred over MAE because it does not use the absolute function [11]. It tends to give high weights to the larger errors than the other evaluation metrics. As we are predicting the cash need of the ATM, we wanted to avoid high prediction then the actual need. RMSE is best evaluation metric for such use cases. RMSE is considered a good evaluation measure for the evaluation of model that is trained using large data set. RMSE is the most used evaluation metrics both in Statistics and machine learning studies.

VII. RESULTS AND DISCUSSION

After running experiment with the proposed approach, we have found the following RMSE.

Algorithm	Train RMSE	Test RMSE
LSTM	181	179
LSTM with Look Back 7	133.5	132.53
Dense NN	199	196.5
Dense NN with Look Back 7	179.1	177

Based on the results mentioned above it is feasible to use LSTM with look-back-7 over Dense Neural Network. LSTM outperformed Dense neural network in accurately predicting the ATM replenishment amount, though it is hard to understand in business scenarios but it is easy to implement for time series data. Better results were found if we perform experiment on cluster of ATMs having similarities rather than individual ATMs. The results showed the performance on the 'high demand in less time' cluster. Same approach can be used for other 9 identified clusters. In these clusters it is find that many clusters consist of ATMs which do not required to be replenished frequently or required a large amount to be replenished. Since, the Root Mean Squared (RMSE) we got by LSTM with look-back-7 is 132.53 which is much lower than Dense neural network and quite encouraging for problem of predicting ATM replenishment amount while using the replenishment data.

VIII. CONCLUSION FUTURE WORK

The paper proposed an LSTM based model for predicting ATM replenishment amount. It is a challenging problem to identify the right amount of cash always available for the customers for transactions. The back-end database schema that is the replenishment log from ATM software is used to formulate the model. Overall, the model performed quite well on training and test set data. The Long-Short-Term Memory (LSTM) based model produce Root Mean Squared Error (RMSE) of 132.5, which is quite low in comparison to different models suggested in the literature. The practical implementations of the proposed model can increase customer satisfaction and decreased cost for maintaining ATMs. In the future work, we foresee to clusters ATMs in term of transactions patterns and cash demands similarity to forecast a cash demand. In this way a simple model can be used to serve a group of ATMs.

IX. REFERENCES

[1] S. M. H. a. R. Z. A. m. p. u. A. I. t. A. s. I. D. T. 1. Hasheminejad, "ATM Management Prediction Using Artificial Intelligence Techniques," vol. 11, 2017.

[2] T. D. and K. A. a. D. Srinivasakumar, "A Survey On Cash Demand Forecasting For ATMs Using Different Financial Modelling Techniques," *The International Journal of Science and Technologies,* vol. 4, 2016.

[3] R. a. D. D. a. B. L. Simutis, "Cash Demand Forecasting for ATM Using Neural Network and Support Vector Regression Algorithms," *Continous Optimization and Knowledge Based Technologies,* vol. 1, 2008.

[4] A. Rajwani, T. Syed, B. Khan and S. Behlim, "Regression Analysis For ATM Cash Flow Prediction," *International Conference of Frontiers of Information Technology,* 2017.

[5] P. V. a. K. M. Ranade, "ATM Cash Flow Management," *International Journal of Innovation Management and Technology,* vol. 6, p. 5, 2015.

[6] Y. Ekinci, J.-C. Lu and D. a. Ekrem, "Optimization of ATM cash replenishment with group-demand forecasts," *Expert Systems with Application An International Journal,* 2014.

[7] Y. a. G. E. Zhang, "The Study of Forecasting of Cash Flow in ATM based on Cubic Exponential Smoothing Method," *An Indian Journal,* vol. 10, 2014.

[8] Catal, C. a. Fenerci, A. a. Ozdemir, B. a. Gulmez and Onur, "Improvement of Demand Forecasting Models with Special Days," *Procedia Computer Science,* vol. 59, 2015.

[9] Simutis, R. a. Dilijonas and L. a. F. Darius and Bastina, "A Flexible Neural Network for ATM Cash Demand Forecasting," 2007.

[10] B. a. Gary, " Mean absolute Error and Root Mean Square Errorr: which is the better metric for assessing model performance?," 2017.

[11] Chai and T. a. R. R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?– Arguments against avoiding RMSE in the literature. Geoscientific model development," vol. 3, 2014.

[12] K. Venkatesh, V. Ravi, A. Prinzie and D. V. d. P. Cash, "Cash Demand Forecast in ATM by Clustering and Neural Networks," *European Journal of Operational Research,* vol. 232, 2014.

[13] C. Ramırez and G. Acu˜na, "Forecasting Cash Demand in ATM Using Neural Networks and Least Square Support Vector Machine," vol. 7042, 2011.

[14] H. a. Jeffery, "Introduction to Neural Networks with Java," 2008.

[15] Serengil, S. Ilkin and O. a. Alper, "ATM Cash Flow Prediction and Optimization with ANN," *International Journal of Engineering Research and Development,* vol. 11, no. 1, 2019.

[16] Kamini, Venkatesh, Vadlamani, Ravi, Prinzie, Anita and V. d. P. a. Dirk, "Cash Demand Forecasting In ATMs By Clustering and Neural Networks," *European Journal of Operational Research,* vol. 232, 2014.

[17] M. Rafi, M. T. Wahab, M. B. Khan and H. Raza, "ATM Cash Prediction Using Time Series Approach," *"ATM Cash Prediction Using Time Series Approach," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET),* no. 10.1109/iCoMET48670, p. 6, 2020