



**Arab American University**  
**Faculty of Graduate Studies**

**Automated Teller Machines Availability Forecasting Model Using Machine  
Learning Algorithms**

By

**Lana Rateb Omar Al-zeir**

Supervisor

**Dr. Mohammad Jubran**

**This thesis was submitted in partial fulfillment of the requirements for the  
Master's degree in**

**Data Science and Business Analytics**

**2/2024**

**© Arab American University -2024. All rights reserved.**

## Thesis Approval

### **Automated Teller Machines Availability Forecasting Model using Machine Learning Algorithms.**

By

**Lana Rateb Omar Al zeir**

This thesis was defended successfully on 28-Feb-2024 and approved by:

Committee members

Signature

1. Dr. Mohammad Jubran: Supervisor



2. Dr. Majdi Owda: Internal Examiner



3. Dr. Radi A. Jarrar: External Examiner

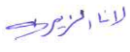


## Declaration

I declare that the thesis titled "Automated Teller Machines Availability Forecasting Model using Machine Learning Algorithms " is my work, has been composed solely by myself does not contain work from other researchers, and has not been submitted for any other degree or scientific work except the reference is made.

Name of The Student: Lana Rateb Omar Al-zeir

ID: 202012335

Signature: 

Date: 23/5/2025

## **Dedication**

To the memory of our martyrs, who sacrificed their lives  
on the way to freedom and independence.

To Dr. Marwan Barghouthi and all pure men and women,  
who faithfully serve our country.

To my parents who taught me patience, determination, and  
courage.

To my husband who always stands next to me and offers  
love, help, and support.

To my children, the air I breathe and hope I live for.

To my beloved brothers and sisters, the rock that I always  
lean on.

To all my friends and colleagues.

## **Acknowledgment**

I want to express my deep thanks and gratitude to the gentlemen at the Bank of Palestine, for their cooperation in providing me with the data used in this thesis.

I would also like to thank Dr. Mohammad Jubran for supervising me and providing me with valuable comments and constructive suggestions, without which this work would not have been brought to light.

## **Abstract**

Automated Teller Machine is the most widespread bank service, it is also the most widely used among electronic banking services, and it represents an important tool for the bank to spread, promote, and attract clients. There are several factors to take into consideration in ATM cash management, forecasting the withdrawal amount is the first in priority for any bank, replenish with the optimal amount saves expenses and maximizes the profit.

Based on raw ATM transactions data from the Bank of Palestine for 2017-2018, these data were preprocessed, and new features were engineered using different approaches, 3 models were utilized (Random Forest, XGBoost, and LSTM), and LSTM gained the best result in tow tasks, first to predict the next-hour withdrawals for 172 ATMs grouped into hourly time series and considered as one ATM, with R2 of 91% and MAPE of 0.37; the other task was to predict the rolling mean transactions amount of the next even hours (2-20), with the best performance of with R2 96%-99% and MAPE of (0.28-0.08).

This project find out that the cash withdrawal amounts are highly affected by the days of salary transfer of the Palestine National Authority and the hours of the transactions. And demonstrated that utilizing the designed features and suitable machine-learning algorithms allows for accurate predictions of withdrawals from these machines.

## Table of Contents

Thesis Approval.....	I
Declaration.....	II
Dedication.....	III
Acknowledgment.....	IV
Abstract.....	V
Table of Contents.....	VII
List of Tables.....	X
List of Figures.....	XI
List of Appendices.....	XII
List of Abbreviations.....	XIII
List of Equations.....	XV
1 Chapter One: Introduction.....	1
1.1 ATM cash management.....	1
1.2 Problem Statement.....	2
1.3 Objectives.....	2
1.4 Contribution.....	3
1.5 Thesis Organization.....	3
2 Chapter Two: Literature Review.....	4
2.1 Group the ATMs into clusters according to specific characteristics.....	4
2.2 Each ATM's transaction represents a one-time series.....	9
2.3 Conclusion.....	19
3 Chapter Three: Exploratory Data Analysis.....	20
3.1 Introduction.....	20
3.2 Data Collection.....	20
3.3 Correlation.....	20
3.4 ATM Transaction Analysis.....	21
3.5 Time Series Analysis for ATM Transactions.....	26
3.6 Conclusion.....	26

4. Chapter Four: Proposed Methodology .....	27
4.1. Introduction.....	27
4.2. Dataset .....	27
4.3. Data Cleaning .....	29
4.3.1. Drop nonsuccessful transactions .....	29
4.3.2. Missing values .....	29
4.3.3. Outliers Detection.....	29
4.4. Data Preprocessing .....	29
4.4.1. Currency Exchange .....	29
4.4.2. Grouping transactions on an hourly basis .....	30
4.5. Feature engineering.....	30
4.6. Data Normalization.....	31
4.7. Time series decomposing.....	32
4.7.1. Deseasonalization.....	34
4.7.2. Detrending.....	34
4.8. Machine Learning Algorithms.....	34
4.8.1. Random Forest RF.....	34
4.8.2. eXtreme Gradient Boosting XGBoost.....	36
4.8.3. Long Short-Term Memory LSTM.....	37
4.9. Metrics Selection .....	40
4.10. Conclusion.....	41
5. Chapter Five: Experiments and Result Analysis.....	42
5.1. Introduction.....	42
5.2. Experimental Environment .....	42
5.3. Dataset preparing for models .....	42
5.3.1. Train & Test Set Split .....	42
5.3.2. Time-based cross-validation. ....	42
5.4. Generating Models.....	43
5.5. Parameter Fine-Tuning .....	43
5.6. Results.....	46

5.6.1	Random Forest Results.....	46
5.6.2	XGBoost results.....	48
5.6.3	LSTM results.....	51
5.7	Discussion.....	53
5.7.1	Error term-based comparison.....	53
5.7.2	Graph-based comparison.....	54
5.7.3	Time-based comparison.....	56
5.7.4	Features-based comparison.....	56
5.8	Conclusion.....	56
6.	Chapter Six: Thesis Conclusion and Future Work.....	57
6.1	Thesis Conclusion.....	57
6.2	Future Work.....	58
	References.....	59
	Appendices.....	67
	المُلخَص.....	71

## List of Tables

Table 2-1 Summary of ATM Datasets in the Literature for Clustered ATMs .....	6
Table 2-2 Summary of Models Used for ATMs' transactions represent clustered ATMS .....	8
Table 2-3 Input Features [3] .....	9
Table 2-4 Features in [14] .....	12
Table 2-5 Summary of ATM Datasets in the Literature for Each ATM's transactions represent one-time series .....	14
Table 2-6 Summary of Models Used for ATMs' transactions represent one-time series .....	17
Table 4-1 BOP statistics 2017-2018 [32] [33] .....	28
Table 4-2 Statistical Summary .....	28
Table 4-3 New Features Description .....	30
Table 5-1 Parameter Fine-Tuning .....	44
Table 5-2 LSTM trial and error combinations results .....	45
Table 5-3 Epoch number tuning results.....	45
Table 5-4 RF error terms value the best results for each model are shown in bold .....	46
Table 5-5 XGBoost error terms values the best results for each model are shown in red. ....	48
Table 5-6 LSTM error terms value the best results for each model are shown in red. ....	51
Table 5-7 Models error terms comparison .....	53
Table 8-1 Raw data fields description .....	<b>Error! Bookmark not defined.</b>
Table 0-2 Raw Data Features .....	<b>Error! Bookmark not defined.</b>

## List of Figures

Figure 2-1 Schematic of the proposed methodology [4] .....	11
Figure 3-1 Heat map for correlation between the dataset's features .....	21
Figure 3-2 Currency percentages.....	22
Figure 3-3 box plot for the day of the week demands.....	22
Figure 3-4 transactions number per week of the month .....	23
Figure 3-5 box plot for the day of the month demands .....	23
Figure 3-6 Number of transactions per day of the month .....	24
Figure 3-7 Total withdrawals Percentages vs Hour .....	24
Figure 3-8 Hourly data for each month .....	25
Figure 3-9 Issuing Bank vs ATM transactions count.....	25
Figure 3-10 Total Daily Transactions Amount .....	26
Figure 4-1 Proposed Methodology.....	27
Figure 4-2 dys_bet_trans Vs same_dys_avg .....	31
Figure 4-3 Time Series Components [41] .....	32
Figure 4-4 Transaction's amount series components .....	33
Figure 4-5 Random Forest regressor [46] .....	35
Figure 4-6 General Architecture of XGBoost [51].....	36
Figure 4-7 LSTM cell.....	37
Figure 5-1 Expanding window [45] .....	43
Figure 5-2 sliding window [45].....	43
Figure 5-3 'Actual vs predicted values of Transactions amount for RF.....	47
Figure 5-4 Rolling Hours vs R2 for RF.....	47
Figure 5-5 Highest 5 features importance .....	48
Figure 5-6 Actual vs predicted values of Transactions amount for XGBoost .....	49
Figure 5-7 Rolling Hours vs R2 for XGBoost.....	50
Figure 5-8 Highest 5 features importance .....	50
Figure 5-9 Actual vs predicted values of Transactions amount for LSTM.....	52
Figure 5-10 Rolling Hours vs R2 for LSTM .....	52
Figure 5-11 Loss vs validation for LSTM.....	53
Figure 5-12 'Actual vs predicted values of Transactions amount for RF.....	55
Figure 5-13 Actual vs predicted values of Transactions amount for LSTM.....	55

## List of Appendices

Appendix (1)	Correlation heat map
Appendix (2)	Hourly seasonality
Appendix (3)	Raw data fields description
Appendix (4)	date- based features that were tested and excluded

## List of Abbreviations

Abbreviation	Description
ID CNN	one-dimensional Convolutional Neural Network
ATM	Automated Teller Machines
ARIMA	Autoregressive Integrated Moving Average
BOP	Bank of Palestine
CNN	Convolutional Neural Network
DL	Deep learning
DTW	Discrete time wrapping
ERPB	Euro Retail Payments Board ERPB
FCN	Fully Connected Network
GMDH	Group Method of Data Handling
GNN	Graph Neural Networks
GRNN	general regression neural network
ILS	Israeli shackles
JOD	Jordanian Dinar
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long short-term memory
ML	Machine Learning
MLFF	Multilayer Feed Forward Neural Network
NN	Neural Network

PNA	Palestine National Authority
RPROP+	Resilient backpropagation algorithm
USD	United State Dollar
WNN	Wavelet Neural Network
WNN	wavelet neural network
XGBoost	eXtreme Gradient Boosting

## List of Equations

Equation 3.1 Correlation.....	19
Equation 4.1 Min – Max scalar.....	33
Equation 4.2 Sliding model.....	34
Equation 4.3 Multiplicative model.....	35
Equation 4.4 Detrending by differencing.....	36
Equation 4.5 Forget Gate .....	41
Equation 4.6 Input Gate .....	41
Equation 4.7 Input Gate .....	41
Equation 4.8 Cell state.....	41
Equation 4.9 Input Gate .....	42
Equation 4.10 Input Gate .....	42
Equation 4.11 The coefficient of Determination (R <sup>2</sup> ) .....	43
Equation 4.12 Mean Absolute Percentage Error (MAPE): .....	44
Equation 4.13 Root Mean Square Error (RMSE) .....	44
Equation 4.14 Mean Absolute Error (MAE) .....	44

## **1. Chapter One: Introduction**

### **1.1. ATM cash management**

Financial inclusion has recently received increasing attention from decision makers in various countries of the world, this term aims to enable marginalized and financially inactive groups and to improve their access to diversified banking services, electronic financial services can be considered as one of the approaches to achieve this goal, which enable bank's clients to execute their financial transactions wherever and whenever they need, Automated Teller Machine ATM is the most service used of all other electronic services [1], according to Palestine Monetary Authority PMA reports of the fourth quarter of 2022 [2], there are 724 working ATM in Palestine, 159 ATMS of them are owned by Bank of Palestine, and more than 219,735 ATM active cards, with an increase of about 32% comparing to 2021, these card holders made more than 18.7 million transactions during 2022, with a total value of about 5.855 billion dollars, most of them are in peak time related to salary withdrawals, these statistics stress the importance of ATM cash management optimization.

Euro Retail Payments Board ERPB Working Group [3] declared that despite the increase in digital payment ways, the main access to cash in the Eurozone is still through traditional means like ATMs and bank branches, also, Author [1] conducted a study to define the quality of the electronic services in the banking sector in Jordan, and using a questionnaire tool; the author found that 60.4% answered that they use ATMs the most over other banking electronic services. Almost 60% agreed they would be happy to fulfill their banking needs via the Internet.

This vast number and amount of transactions in peak times represent a challenge for both bank clients and bank administrations; bank clients must wait a long time in queues and be exposed to ATM transaction failure because of a lack of cash or ATM Technical error, on the other hand, bank administrations are suffering from repeated ATM replenishment in addition to low client satisfaction, as they obliged to maintain a minimum amount of cash in their ATMs [2]. The solution to this situation is to offer a model for both sides that suggests the most available ATM for bank clients and forecasts the time needed for the next ATM replenishment.

ATM cash management optimization aims to avoid manual work, at the beginning, banks managed ATM cash based on trial and error until they gained enough experience and written

operating policies were valid. Banks tried to avoid out of service by depositing too much cash in ATMs, losing the profit of idle cash in ATMs, many ATMs were found to hold up to 40% more cash in their ATMs than the optimal requirement, although financial consultants' advice is not to exceed 20% of this excess cash [4].

Banks always aim to maximize their profitability, so they pay great attention to their customers, customer attitudes, satisfaction, and behavioral intentions are always under the spot and are affected by the quality of the electronic service affects in [1], ATM transactions are being under focus for the last few decades, researchers' approaches in this field analysis differed in its purpose, data size, and methodology used, as well as their results.

Machine learning is all about gaining knowledge and useful insights out of data that are derived by computers [31], in addition to traditional statistical models, machine learning is one of the most recent approaches in ATM analysis, these studies utilized different algorithms in terms of its working principle, and collect and preprocess a diversity of features and datasets.

## **1.2. Problem Statement**

This research aims to find out how to efficiently employ historical ATM transactions and suitable machine learning to predict anticipated withdrawal amounts to optimize ATM cash management.

## **1.3. Objectives**

The answer to the above-mentioned problem statement is the main objective of this research, which will guarantee these subsidiary goals:

- enhance financial inclusion by improving the availability of ATMs' cash.
- Save Clients time and effort in using ATMs as the cash will be available all the time.
- Help BOP to organize ATM replenishment more effectively and efficiently by using adequate amounts with better timing.
- Help banks achieve higher customer satisfaction, cut operational costs, and raise their service level.
- This model can propose a new financial inclusion indicator by increasing successful ATM transactions.

#### **1.4. Contribution**

To the researcher's knowledge, this is the first research for ATM cash management optimization, not on the level of Palestine, but on the level of the Arab world. A novel dataset was used, which revealed the reality and nature of the Palestinian society's behavior in using ATMs, this study identified the factors affecting withdrawal amounts; mainly the days of salary transfer of the Palestine National Authority and the hours of the transactions, by applying an appropriate machine learning algorithm, which can be utilized to enhance cash availability in ATMs.

#### **1.5. Thesis Organization**

Chapter 2 provides a literature review of the studies focused on the forecasting of withdrawal amount via machine learning regression models. An in-depth data exploratory analysis was carried out and the proposed methodology was clarified in chapters 3 and 4 respectively, while the executed experiments results illustration was included in Chapter 5, and then shows and compares the result of the experiment, at last, Chapters 6 and 7 will represent the conclusion and future work.

## 2. Chapter Two: Literature Review

Machine learning techniques aim to fit real-world facts and give better results for new datasets [5], these techniques are significantly used in ATM cash forecasting. The researchers dealt with the ATM data in two ways:

### 2.1. Group the ATMs into clusters according to specific characteristics

Consider each ATM as the unit of analysis, while others used clustering ATM data; clustering criteria differ from one paper to another, one is ATM location like in [10], or by the pattern in the withdrawal of day of the week feature in [8], the withdrawal is totaled for each cluster and is treated as one ATM for the analysis process.

Venkatesh et al. [6] aimed to make the replenishment plan more efficient and effective for banks, so they tried to forecast for 56 days period on a weekly period, the approach was Cluster-then-predict, so they advised to cluster ATM data as the first step for data preparation; they transformed the transactions of each ATM into a sequence of 7 values length, each value represents the average of all daily withdrawals on the same day of the week, the similarity between these sequences was criteria for clustering ATMs into four clusters, then developed 4: general regression neural network (GRNN), Group Method of Data Handling (GMDH), multi-layer feed-forward neural network (MLFF), and wavelet neural network (WNN) to forecasting, these model were trained and tested with NN5 dataset, Out of their four models, GRNN gave the best results with SAMPE of 18.44, and they conclude that ATM replenishment plans will save a lot of operational costs if they were prepared on clustering basis.

Jadwal et al. [7] used a sample of 11 ATMs from the NN5 data set, each ATM transaction was considered a single time series, then they clustered these time series using K-means with 3 clusters according to the distance measured between time series., the distance was measured by Discrete time wrapping (DTW), the NN model used was Resilient backpropagation algorithm (RPROP+) model was used for both clustered and non-clustered data, they approved that this approach resembled [6], average RMSE of the three clusters (6.58) is better than those of Non-clustered (8.5).

Ekinci et al. [8] aimed to determine the optimal interval between every two replenishments to establish trucking logistics schedules, they grouped them according to location, and collect features about their locations and residents, the arithmetic mean is found for the daily withdrawal of each group, then they aggregated the daily withdrawal for each cluster, they added categorical features like the day of the week, week of the year, and before special events, numerical features were added also; like Counting recreational, economic and social buildings, and counting the age groups of the population in the area

of each ATM, a regression model is built and tuned for each group, then the predicted amount is divided among ATMs in the same group, based on the percentage of each of them in the withdrawals, the last step was to use the for each group in predicting the optimal replenishment amount for each ATM within the group individually, the dataset in this experiment was the transactions of 152 ATMS in Istanbul, which varies in a period of data of each ATM, the author compared the actual amount of five ATM group to the of each ATM predicted amount using its proportion, they found the average MAPE of 0.202577, while the average MAPE of the comparison to the predicted amount by individual ATM was 0.2269 MAPE, in the contrast to these result, some ATMs had better MAPE in the individual ATM approach.

Finding the optimal cash demand strongly affects the amount and schedule for replenishment, realizing this cause-effect relation represents a serious need and critical challenge for banks, Asad. et al. [9] made use of this relation and of central tendency measures, and engineered three new features, the mean of monthly amount, the median number of times ATM replenishments, and the mean of days for the next replenishment, they used a dataset consisting of ATM feed amounts with it for 2241 ATMs for 22 months from 6 banks in Pakistan, then they grouped ATMs depending on Density-Based Spatial Clustering, and 10 clusters were yielded, Dense NN and LSTM were built and trained, authors tested the two models with and without and a look-back value of 7 days, the least RMSE was produced by LSTM with a look-back, its value was 132.53.

The Summary of ATM Datasets in the Literature for Clustered ATMs is included in Table 2-1 below.

Table 2-1 Summary of ATM Datasets in the Literature for Clustered ATMs

Reference	Name of the Dataset	Number of ATMs	Country	Number of Records (rows)	Period of the collected	Number of features	List of features
[6]	NN5	111	UK	735 observations of each ATM	2 years, 18/3/1996 to 22/3/1998	---	111 time series of ATM transaction
[7]	NN5	11	UK	735 observations of each ATM	2 years, 18/3/1996 to 22/3/1998		3 time series of ATM transactions, each denotes centroid of each cluster
[8]	--	152	Turkey		Up to 2.5 years		<p>Week of the year, Week before special days like Father's, Mother's, and Valentine's Day</p> <p>Interest points count like Cafe, colleges, and malls in 1000 m around the ATM</p> <p>number of social buildings, business-related buildings in the street, commercial buildings</p>

							population in the street, number of people who are at the age of 15-19,20-44,45-54, older than 55
[9]	---	2241	Pakistan	609,932	22 months	32	(ID of ATM), Start Time and End Time of replenishment, Event ID, Status, etc.  They add the mean of the monthly amount, the median number of times ATM replenishments, and the meant of days for the next replenishment
[16]							Scotland special day, Valentine's Day, St. David's Day, Mothering Day, St. Patrick's Day, Easter Sunday, Easter Monday, Good Friday, St George's Day, Early May Bank Holiday, Spring Bank Holiday, Father's Day, Orangemen's Day, Summer Bank Holiday, Halloween, Thanksgiving Day, Boxing Day, Christmas Day, and New Year, the day before the weekend flag, quarter factor

[10]	Middle Eastern Bank	---	Iran	26	3 years	8	Last Day Amount 2. Last Week Average Amount 3. Last Week Amount 4. Salary Day (10th of every month) 5. Month 6. Day of Week 7. Weekend (is it a weekend) 8. Peak Period (19-24th of every month)
------	---------------------	-----	------	----	---------	---	--

Table 2-2 Summary of Models Used for ATMs' transactions represent clustered ATMS

Reference	Forecasted value	Models used	Metric used to evaluate the model	The value of each metric
[6]	ATM center's daily cash demand	GRNN, MLFF, GMDH, WNN	SMAPE	The best result is <b>18.44%</b> for GRNN
[7]	56 days of cash demand	ANN	RMSE	8.5
		ANN with Clusters		<b>6.58</b>
[8]	Weekly cash withdrawal	LR	MAPE	22.69
[9]	Optimal ATM cash replenishment	Dense NN	RMSE	177.00
		LSTM		<b>132.53</b>
[16]	56 days horizon,	Neural Networks, ARIMA, and Exponential Smoothing	SMAPE	The best is for Exponential Smoothing of <b>21.57</b>
[10]	The next day cash demand	MLP and Polynomial Regression PR	Surplus cash amount	MLP is better and predicted less cash amount than PR

Features included have a high impact on its performance, (Catal et al., 2015) [11] emphasize that including 19 special days in the UK when analyzing the NN5 dataset, in addition to flagging Saturday

and Sunday, and taking under consideration month and quarter features, and data were exposed to exponential smoothing, they found that enhancing features improve error term even they did not use complex computational intelligence techniques, as they achieved average SMAPE of 21.57 % in forecasting 56 days cash withdrawals with Exponential Smoothing and outperforming the other two models: Neural Networks and ARIMA.

Also, many other papers extracted a lot of features from the transaction date feature, some of these features are: Last Day Amount, Last Week's Average Amount, Last Week's Amount, Salary Day, Month, Day of Week, is it Weekend, is it Peak Period, like in [10], and is a holiday like in [12], authors of [13] extracted 29 features showed in table 2-1, feature extractions is significant, it enables researchers to depend on the most informative and essential features by feature selection, [14] added day of the month, is last day name within the month and which working day of the week, Experiments of [13] showed that cash withdrawals have a seasonal trend based on date time features and they are predictable, also indicated that unused balance can be decreased dramatically depending on accurate predictions.

## 2.2. Each ATM's transaction represents a one-time series

In this section, several studies were presented in which the authors treated ATM as a one-time series.

Serengil et al. [13] proposed the use of a fully connected neural network (NN) model for cash demand predictions at ATMs to optimize ATM cash flow management. They used a layers NN model with 6500 individual ATMs all over Turkey. The authors selected the 29 features listed in Table 2-3 to be digested by the model. The model is designed to select potential replenishment ATM candidates using daily projections and assessing the associated expenses. Transportation costs and prospective negative interest costs are both considered in the cost function. The model is fitted using data retrieved from 41 ATMs where a 30% reduction in cash expenditures was achieved through the proposed method.

Table 2-3 Input Features [3]

1. Is it a religious holiday?	2. Week of Month
3. Is it before a religious holiday?	4. Week of year
5. Day of the month	6. Is Work Day?
7. Yearly deviation	8. Year

9. Is it Father's Day?	10. Is Monday?
11. Is the First day of the month	12. Is Tuesday?
13. Is the First work day of the month	14. Is Wednesday?
15. Is it half day?	16. Is Thursday?
17. Month of the year	18. Is Friday?
19. Is it Mother's Day?	20. Is it Saturday?
21. Season	22. Is Sunday?
23. Trx Amount of 1 day earlier	24. Is Middle of the Month?
25. Trx Amount of 2 days earlier	26. Is Middle Workday of the Month
27. Trx Amount of 3 days earlier	28. Is it an exceptional salary day?

Gorodetskaya et al. [15] proposed a global methodology for forecasting various non-stationary temporal data in automatic mode to find the optimal use of cash flow in ATMs. The proposed method consists of five steps: designing feature space, outlier detection, features selection (greedy way), building the model, and selection of hyperparameters. The design of the feature space included an extensive set of features such as one-hot encoding of calendar features (day of the week, month, weekend, holiday), time series values for previous days, rolling statistics grouped by calendar features (average, variance, minimum, maximum), and events of massive payments (advance payment, salary). Based on data collected from ATM transactions over 3 years and using the Random Forest and gradient boosting models with sliding window cross-validation, they showed that the forecasting demand in ATMs can be achieved with about 18% mean absolute percentage error.

Sarveswararao et al. [16] propose a two-stage ATM cash demand forecasting. In stage one, the chaos present in the time series is modeled by reconstructing the state space of each series using the lag and embedding dimension. In addition, eight exogenous features are designed to represent the day of the week and whether it is a weekend or not. In stage two, nine different ML and one DL models are fitted and evaluated to predict cash demand forecast. This methodology is presented graphically in Figure 2-1. Based on transactions collected from 40 ATMs for two consecutive years, the least Symmetric mean absolute percentage error (SMAPE) with the Random Forest (RF) model achieved the best forecasting results, followed by one-dimensional Convolutional Neural Network 1D CNN and Long short-term memory LSTM models.

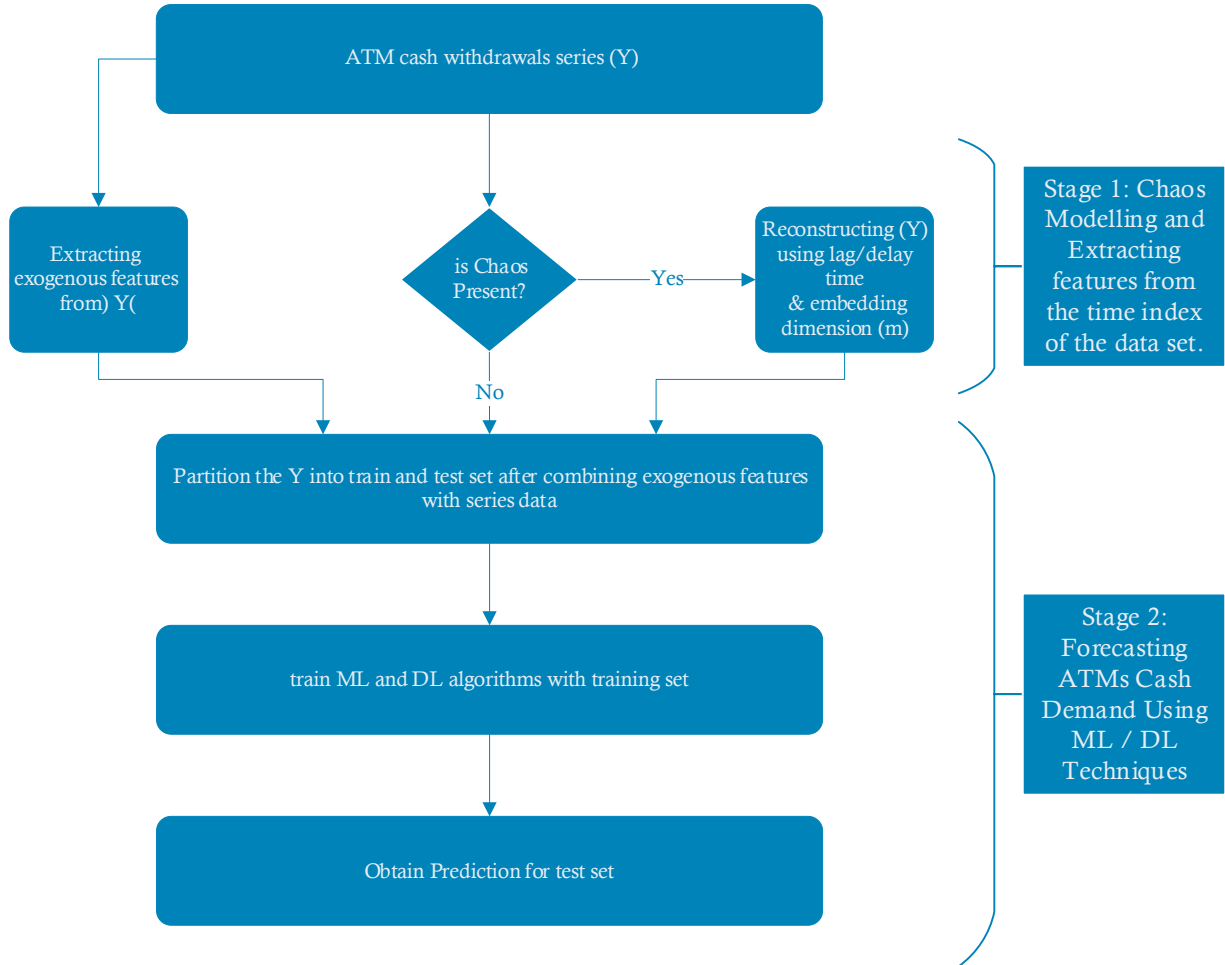


Figure 2-1 Schematic of the proposed methodology [4]

Kamini et al, [17] used the Lancaster University NN5 data set, which consists of the transaction of 111 ATMs in the UK, and covered 735 days. They proposed two methods to predict the next 30 days' cash withdrawals, these two methods are a non-chaotic approach by deseasonalization and imputing, over four models: Autoregressive Integrated Moving Average (ARIMA), and neural network models (Wavelet Neural Network WNN, Multilayer Feed Forward Neural Network MLFF, and Graph Neural Networks GNN). The other was the chaotic approach; where the authors reconstructed the phase space for these models by using the chaotic parameters, including embedding dimension and delay time to analyze the chaos of each ATM time series. For a sample of 30 ATMs, the best forecasting results were achieved using the GRNN model with SMAPE of 14.71%, Kamini et al,

which was better than [6], ignored extracting any date-related feature like the day of the week and if weekend features, lose their effect in raising forecasting accuracy.

Gökçay [14] used two ATM transaction datasets for about two successive years 2017-2018 to forecast cash withdrawals on a weekly and monthly basis. The original dataset consists of 105 features. The features listed in the table 2-4 were added to the dataset. The authors investigated the use of different machine learning approaches for forecasting, including linear regression, and the traditional statistical analysis ARIMA on the same ATM data. They compared the results of these methods and showed that machine-learning methods in comparison with ARIMA have higher accuracy. In addition, it was revealed that among the machine learning models, long short-term memory LSTM gives the most accurate predictions and uses fewer features than other models.

Table 2-4 Features in [14]

Features in [14]	
Is the first working day	Is mid working day
Is the last working day	Average withdrawal amounts previous month
Amount from same day from the previous month	Amount from same day from previous 1 week
Amount from same day from previous 2 week	Amount from same day from previous 3 weeks
Amount from same day from previous 4 week	

Rajwani et al. [18] compared different algorithms for cash forecasting in ATMs in Karachi. They collected data from 86,129 records that represent withdrawal transactions of two and a half years from 10 ATMs, data processing and exploring analysis were conducted which showed the effect of the day of the month and of the week's importance in predicting cash forecasting. They fitted and tested several algorithms such as ARIMA, linear regression, Ridge Linear Regression, Least Absolute Shrinkage and Selection Operator LASSO Model Ridge Cross Validation CV Model, LASSO least-angle regression LAR, Bayesian Ridge Regression, Random Forest Regression, time series prediction, RRN, and RRN with time series. The RRN with the time series model achieved the highest forecasting accuracy of 98%.

Arabani and Komleh [19] used data from 1056 ATMs for one year to forecast the daily cash demand for 73 days, they utilized 3 models; Fully Connected Network (FCN), Support Vector Machines, and Convolutional Neural Networks (CNN). After evaluating the models, the SVM was excluded as it

provided a poor performance. The CNN and FCN achieved Root Mean Square Error (RMSE) of 4.74 and 5.01, respectively.

Authors in [20] used the XYZ bank dataset, which consists of 11,588 rows and 10 columns, to build four cash forecasting models: LR, ARIMA, LSTM, and Prophet. They used the MSE, MAE, and  $R^2$  to evaluate the four models. The LSTM has the least MAE and MSE of 20.686,91 and 710.590.544,24 respectively, and  $R^2$  of 0.71.

An innovative approach was proposed by [21], their data set was from VTB Bank, which owns the largest ATM network in Russia, the dataset covered 1000 days period and belonged to 500 ATMs, they coped with the Nonstationarities of the data set their preprocessing pipeline consists of:

1. Detection of the last correct history
2. Mass payment days' detention
3. History selection and replacement.
4. Stochastic outliers' replacement

They used two kinds of model: local and global depending on time series analysis, each time series had its model, and they generated 57 features from lag, rolling statistics with and without time-related grouping, and 30 of them were extracted from the industrial calendar, they utilized Random Forest, CNN, and MLP regressor.

The global approach depended on one model to forecast multiple time series, they used encoder-decoder LSTM, and the input to the model was restricted to features extracted from the industrial calendar and from days with mass payments.

They ran the models for one, fifteen, and thirty days ahead forecasting and compared the results to some traditional statistical time series models: Holt-Winters method, SARIMA, and Fbprophet, LSTM was the Competition winner with Weighted Absolute Percentage Error WAPE of 0.288 and 0.309 for one day and fifteen days respectively, RF won the thirty days forecasting with WAPE of 0.47.

The Summary of ATM Datasets in the Literature for Each ATM's transactions represent one-time series are included in Table 2-5 below.

Table 2-5 Summary of ATM Datasets in the Literature for Each ATM's transactions represent one-time series

Reference	Name of the Dataset	Number of ATMs	Country	Number of Records (rows)	Period of the collected	Number of features	List of features
[2]	---	6500	Turkey	----	2013-2018	29	<p>Is it a religious holiday? Is it Before a religious holiday?</p> <p>Week of Month/ year/ Month of year/ Day of month</p> <p>Year/ Yearly deviation/ Season</p> <p>Is Monday? Tuesday? Wednesday? Thursday? Friday? Saturday? Sunday?</p> <p>Is it Father's Day? Is it Mother's Day? Is Valentine's Day? Is Work Day?</p> <p>Is the First day of the month/ Is the First work day of the month</p> <p>Is Trx Amount of 1 /2/3 day earlier</p> <p>Is Middle of the Month? / Is Middle Workday of the Month</p> <p>Is it an exceptional salary day? / Is it half a day?</p>

[4]	---	1	Russia	--	2016-2019	23	The amount of demand for 1,2,3,4,5,6, and 7 days before the forecast point
							the minimum, maximum, standard deviation, mean, and median demand value for two days of the same past
							the minimum, maximum, standard deviation, mean, and median demand value for the past month
							the minimum, maximum, standard deviation, mean, and median demand value for the week
[3]	--	40	India	---	two consecutive years (21-11-2017 to 20-11-2019)	9	Daily withdrawals, Is Monday? Tuesday? Wednesday? Thursday? Friday? Saturday? Sunday?
[17]	NN5	111	UK	735 observations of each ATM	2 years, 18/3/1996 to 22/3/1998		111-time series of ATM withdrawals
[14]	---	2	Turkey	731 for ATM1 and for ATM2	2016-2018	31	City Salary Amount /Quarter Salary Amount /Amount from Same Day of Previous 4/3/2/1Weeks
							Amount From Same Day of Previous Month /Average Withdrawal Amount Previous Month
							Is the Last Working Day/ Is the Middle Working Day

							/Day of the Week /Is the First Working Day
							Is the Last Working Day/ Is the Middle Working Day /Day of the Week /Is the First Working Day
							Is In December / November /October / September / August / July / June /May / April / February/ March/ January
							Is Monday? Tuesday? Wednesday? Thursday? Friday? Saturday? Sunday?
[18]	----	10	Pakistan.	86,129	2013 -2015	60	transaction information, such as date, amount, time, terminal ID (ATM ID), ATM location, currency code, customer ID, card number, track information, transaction type, channel type, transaction ID, etc.
[8]	---	1056	---	365 column X 1056 ATM	1 year	2d matrix	Matrix of ATM transactions as rows and day of the year as column
[9]	XYZ	5	India	11. 588	2010-2014	8	ATM name, no of withdrawal, no of XYZ card, the amount withdrawn, amount withdrawn, weekday, Transaction date, and working day

The Summary of Models Used for ATMs' transactions represent a one-time series is included in Table 2-6 below.

Table 2-6 Summary of Models Used for ATMs' transactions represent one-time series

Reference	Forecasted value	Models used	Metric used to evaluate the model	The value of each metric
[2]	Cash demand for the next 15 days	three layers neural network	---	----
[4]	Daily cash demand	Random Forest and gradient boosting over decision trees	MAPE	18%
[3]	56 days of cash demand	GRNN	SMAPE	<b>14.713</b>
		MLFF		17.435
		GMDH		16.947
[16]	Forecasting the following 30-day cash withdrawals	ARIMA	SMAPE	25.4
		RF		<b>21.48</b>
		SVR		22.81
		MLP		23.63
		GMDH		23.91
		GRNN		24.06
		LSTM		22.56
		XGBOOST		23.63
		1D CNN		22.36
		[14]		Monthly and Weekly withdrawals
SVR	7959			
ANN	8037			

		ARIMA		8772
		LSTM		<b>7643</b>
[18]	Transaction Amount	LR	MSE	5%
		Ridged LR		3.0%
		Ridges Model		3.0%
		Lasso		3.0%
		Bayesian Ridge Regression		3.0%
		RNN (LSTM)		<b>2.8%</b>
		LSTM for Regression with Time Series		2.9%
[8]	to forecast the daily cash demand for 73 days	ANN	RMSE	5.01
		SVM		13.66
		CNN		<b>4.74</b>
[9]	56 days of cash demand	LR	MSE	0717121972.31
		Prophet		2480637970.01
		ARIMA		1744616066.79
		LSTM		<b>0710590544.24</b>
[10]	One day ahead forecasting	Holt-Winters	WAPE	0.341
		SARIMA		0.340
		FBProphet		0.339
		CNN		0.335
		MLP		0.401
		RF		0.307
		LSTM		<b>0.288</b>

Features included have high impact on its performance, (Catal et al., 2015) [11] emphasize that including 19 special days in the UK when analyzing the NN5 dataset, in addition to flag Saturday and

Sunday, and taking under consideration month and quarter features, and data were exposed to exponential smoothing, they found that enhancing features improve error term even they did not use complex computational intelligence techniques, as they achieved average SMAPE of 21.57 % in forecasting 56 days cash withdrawals with Exponential Smoothing and outperforming the other two models: Neural Networks and ARIMA.

Also, many other papers extracted a lot of features from the transaction date feature, some of these features are: Last Day Amount, Last Week's Average Amount, Last Week's Amount, Salary Day, Month, Day of Week, is it Weekend, is it Peak Period, like in [10], and is a holiday like in [12], authors of [13] extracted 29 features showed in table 2-3, feature extractions is significant, it enables researchers to depend on the most informative and essential features by feature selection, [14] added day of the month, is last day name within the month and which working day of the week, Experiments of [13] showed that cash withdrawals have a seasonal trend based on date time features and they are predictable, also indicated that unused balance can be decreased dramatically depending on accurate predictions.

### **2.3. Conclusion**

In this chapter, a comprehensive literature review was conducted, related studies with different approaches, datasets, and algorithms were presented and discussed, datasets belonging to different countries were utilized, while as noted, there is no study on ATM transactions in Palestine nor in the Arab countries.

Most of the previously mentioned related work has dealt with the issue in two ways: forecasting the value of the required cash to meet withdrawal transactions, or finding the most appropriate amount of refeeding ATMS [14], for this study the bank approves providing ATM transactions only.

And from the above literature review, it is so obvious that the dataset should at least include the data of 2 years or more, to make sure all periods are included (seasons, cyclic change, peaks, holidays...etc.), feature engineering played a vital role in enhancing the accuracy of the results, and Recurrent Neural Network RNN models like Long short-term memory LSTM [14, 22, 9, 20], which has the most accurate performance beside Random Forest RF [22, 16].

### 3. Chapter Three: Exploratory Data Analysis

#### 3.1. Introduction

In this chapter, the study dived deep into the data set, investigating to reveal hidden insights, which led us to figure out the relations, correlations, and behavior of the features included in the data set. This part of the data science process clears the way for the next stages of the process [23], this chapter contains the following sections: data collection, correlation, ATM transactions behavior analysis, time series analysis for ATM transactions, and conclusion.

#### 3.2. Data Collection

The dataset belongs to Bank of Palestine (BOP), the most popular bank in Palestine with the largest number of clients and branches in the West Bank and Gaza Strip, and with the biggest network of ATMs, which makes it the most suitable bank in Palestine for this research. The data was received in Excel file format, covering the period of 1/1/2017 to 31/12/2018. It contains 8,999,973 transactions for 172 ATMs located in the West Bank, Gaza Strip, and Jerusalem districts, and it is spread over 25 Palestinian cities, The nature of this data will be described in more detail in the next chapter.

#### 3.3. Correlation

It describes the contribution of each variable by measuring only how much the two variables are related, it is an indicator of how strongly two variables are related, positively or negatively [24], highly correlated feature is excluded in forecasting and predicting modeling, as can predict one of the two highly related from the other one, the correlation is calculated by the equation:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3.1)$$

Where:

$r$  = correlation coefficient

$x_i$  = values of the x-variable in a sample

$\bar{x}$  = mean of the values of the x-variable

$y_i$  = values of the y-variable in a sample

$\bar{y}$  = mean of the values of the y-variable

Heat map visualization is the best to visualize the correlation between each possible pair in the dataset [25], figure 3-1 shows the correlation between ATM transaction amount with many new features which is explained in section 4.5 which contains feature engineering done in this research, an example of positive correlation pairs (hour, Transaction\_Amount\_ILS) =0.33, (same\_days\_avg, Transaction\_Amount\_ILS) =0.5, and (same\_days\_avg, Transaction count) = 0.39, also there are negative correlation pairs( same\_days\_avg, Dys\_bet\_trans) = -0.47, (day of month, same\_days\_avg) -0.37, and (workday, day of weak) = -0.4, a larger size of this map is included in Appendix (1).

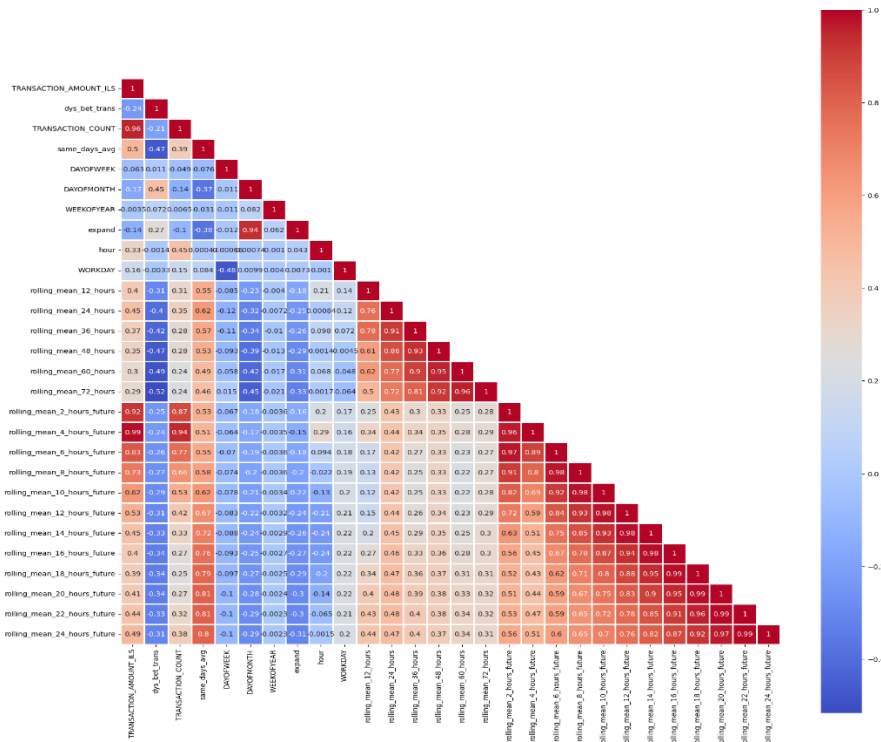


Figure 3-1 Heat map for correlation between the dataset's features

### 3.4. ATM Transaction Analysis

Several visualizations were drawn in this section; it helps to get a more accurate understanding of the behavior of the ATM transactions data set.

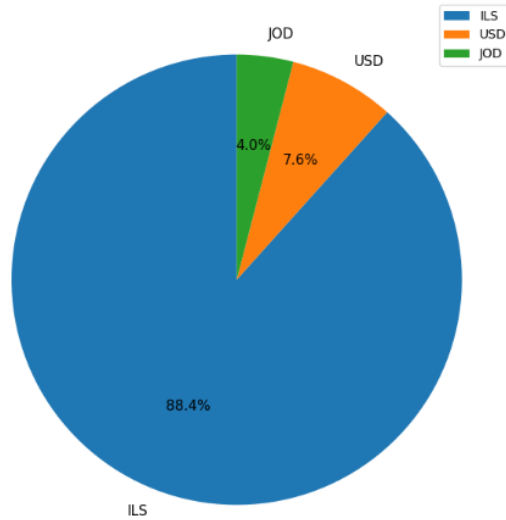


Figure 3-2 Currency percentages

Figure 3-2 shows that the number of transactions in ILS was 88.4% of the whole data set, followed by USD with 7.6 % and last JOD with 4%, which also explains that the salaries of PNA’s employees are in ILS, besides ILS is the currency in circulation in the West Bank, Jerusalem, and Gaza, these facts lead to the need to make ILS the only currency in forecasting.

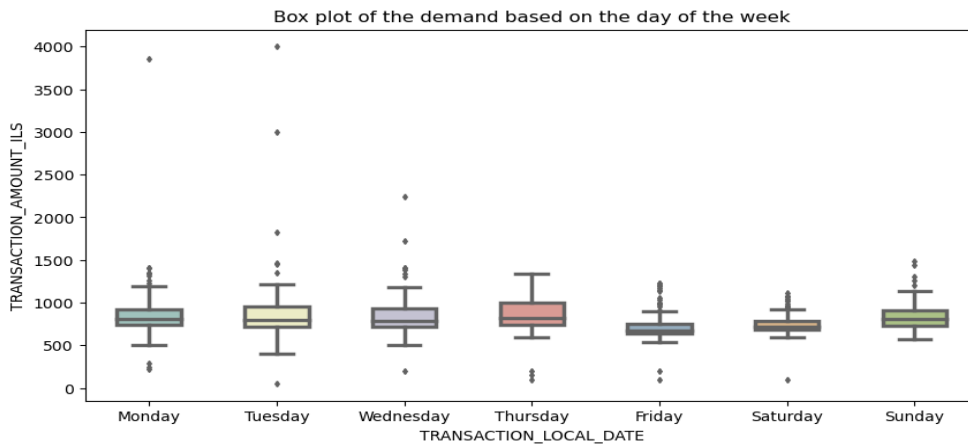


Figure 3-3 box plot for the day of the week demands

Figure 3.3 shows the distribution of the average transactions amount based on the day of the week, it reveals that Thursday has the highest average transactions amount as the last day in working days of the week, with  $Q1 = 733$  and  $Q3 = 962$  while Friday ( $Q1 = 628$  and  $Q3 = 730$ ) and Saturday ( $Q1 = 664$  and

Q3 =770) have the lowest average transactions amount for ATM transactions as they are the weekend in Palestine.

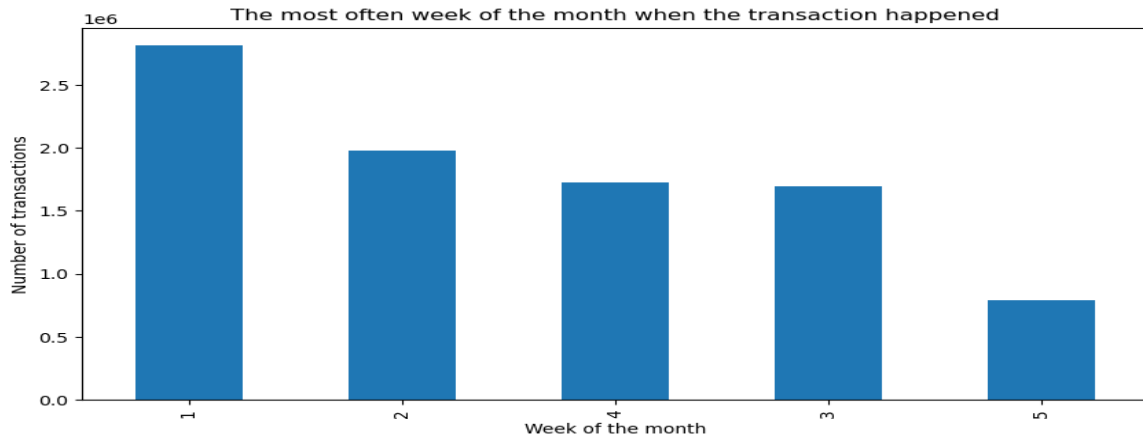


Figure 3-4 transactions number per week of the month

Figure 3-4 shows that the highest transaction count occurred in the first week of the month, which can be explained by the date of transferring the salaries of the employees of Palestine National Authority PNA, whose number reached 209,500 employees in the year 2017-2018 [26], it could also added that the last week before PNA salaries transfer.

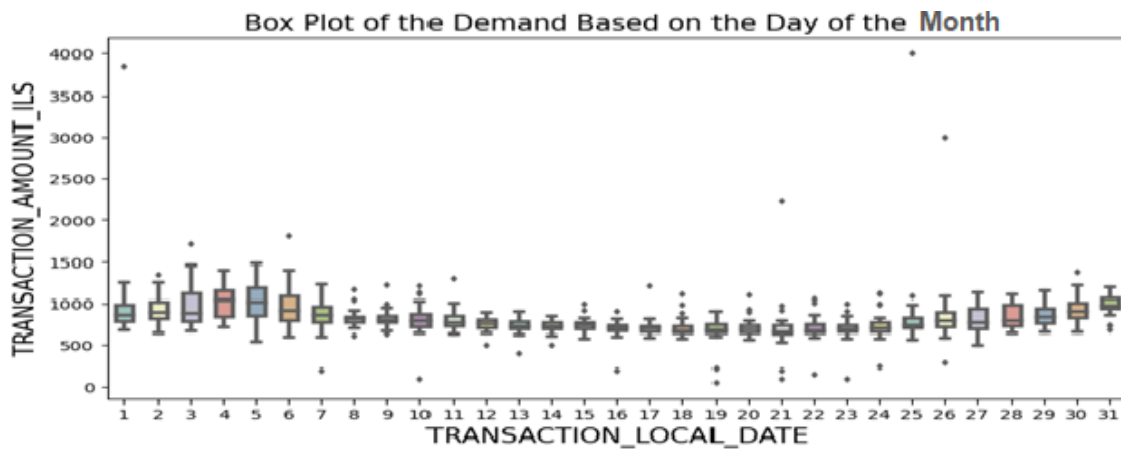


Figure 3-5 box plot for the day of the month demands

Figure 3-5 shows the distribution of the average transactions amount per day of the month, it can be divided into 3 periods according to IQR: the first is week number 1 of the month, when the PNA salaries are transferred, which can be recognized with a higher IQR of about 643 – 1025, the second occurs at

the last days of the month, which can be identified with 571 - 714, when private sector salaries are transferred, the third lays between them with IQR lies between 214 - 430.

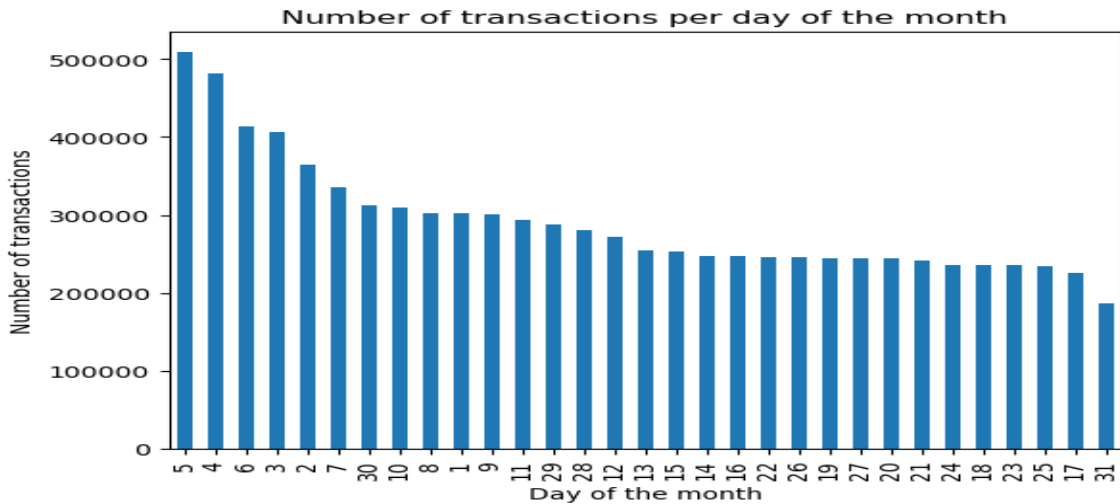


Figure 3-6 Number of transactions per day of the month

Figure 3-6 shows that the day emphasizes the findings in figures 3-4 and 3-5 about the effect of PNA transferring salaries, days 26-30 of the month have the highest count after the first week, which can be explained with the private sector salaries transfer and withdrawal of the remaining of PNA salaries.

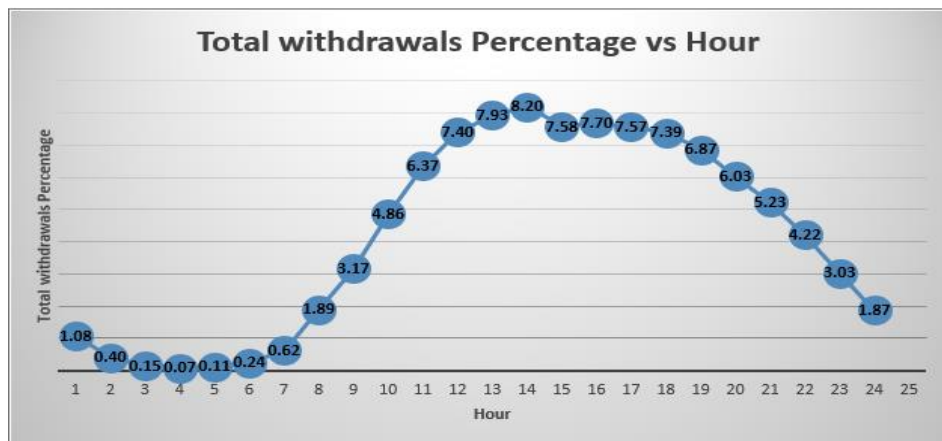


Figure 3-7 Total withdrawals Percentages vs Hour

Figure 3-7 shows the ATM transactions percentages per hour , they were very weak from 02:00- 07:00 with a range of [.07-.62], whereas, they start to increase gradually after that, until their peak at 14:00

with 8.2 % , then start to decrease gradually, d amount during 08:00 – 23:00, it also indicates that hours during 08:00 – 23:00 are more important to forecast ATM cash withdrawals during them.

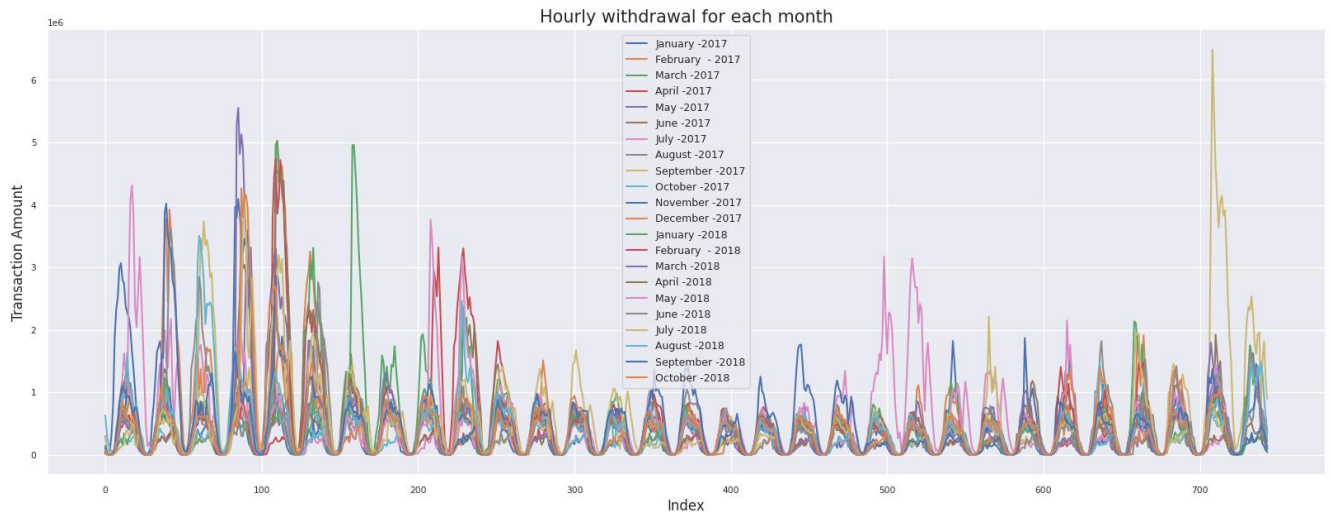


Figure 3-8 Hourly data for each month

Figure 3-8 shows the hourly amount as one series for each month, as shown in the legend, this figure reveals the existence of seasonality with a period of 24 hours, 31 curves can be seen and each curve represents one day, for more clarity; each month is plotted separately in Appendix (2).

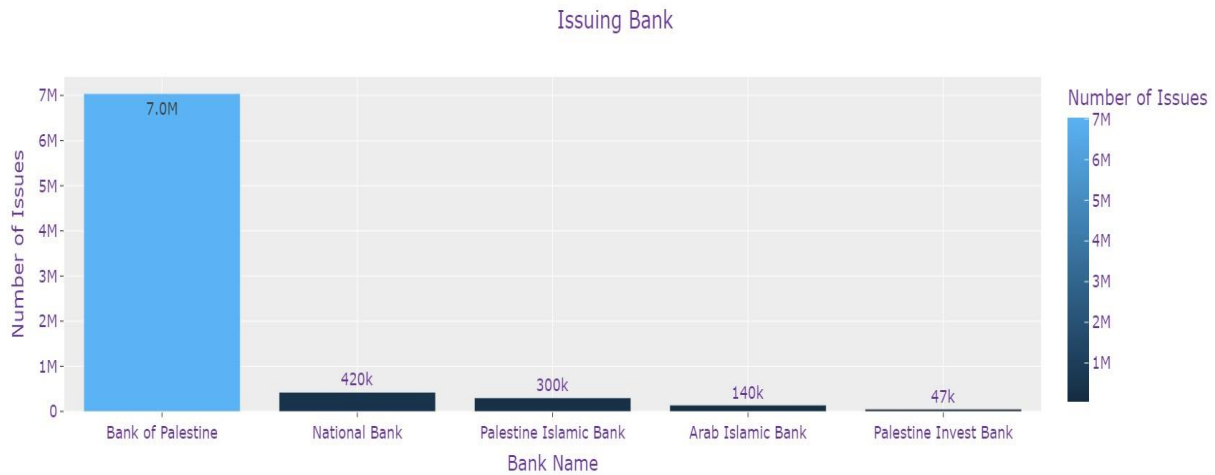


Figure 3-9 Issuing Bank vs ATM transactions count

Figure 3-9 shows the issuing bank with the count of transactions which had been made by the holder of the credit card issued by this bank, this will be explained in detail in section 4.2.

### 3.5. Time Series Analysis for ATM Transactions

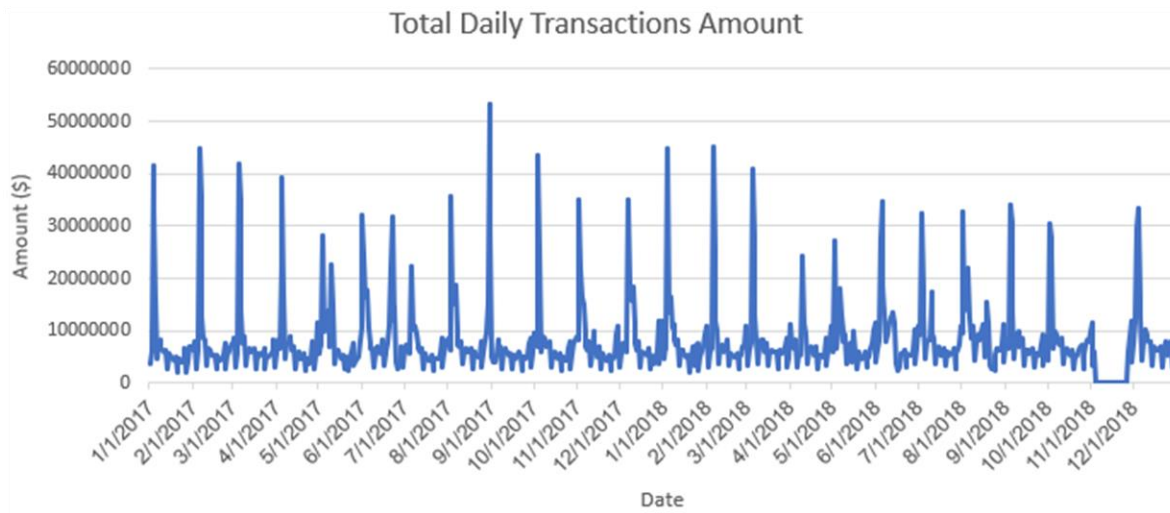


Figure 3-10 Total Daily Transactions Amount

Figure 3-10 shows a time series analysis for the total daily withdrawals through the period of 2017-2018, the time series decomposing proved that there is no obvious trend, and showed seasonal behavior due to what was mentioned in the previous figures 9 and 10, it is obvious that the seasonality variation is constant, which means it does not vary in frequency and amplitude, and it does not look like an increasing funnel, so additive model is assumed for dataset [27].

### 3.6. Conclusion

In this chapter an exploratory data analysis was carried out, which led to deeper insight into the dataset, the number of transactions was compared according to day of the month, week of the month, and region, and the currencies used in the transactions, it was found that ILS is the dominant, it was noticed that the transactions amount are highly affected by of the transferring PNA salaries and private sector salaries, maximum amount were on these dates, these amounts increase on Thursdays and minimum amount were on weekends.

## 4. Chapter Four: Proposed Methodology

### 4.1. Introduction

In this chapter author will illustrate the proposed methodology to answer the questions of this research, predicting the optimal cash amount replenishment, it starts with the description of the collected data set, the preprocessing and scaling, and feature selecting and engineering, depending on three machine learning methods, the models were used to forecast the Hourly cash demand amount depending on the cash withdrawals, the methodology is summarized in Figure 4-1 below.

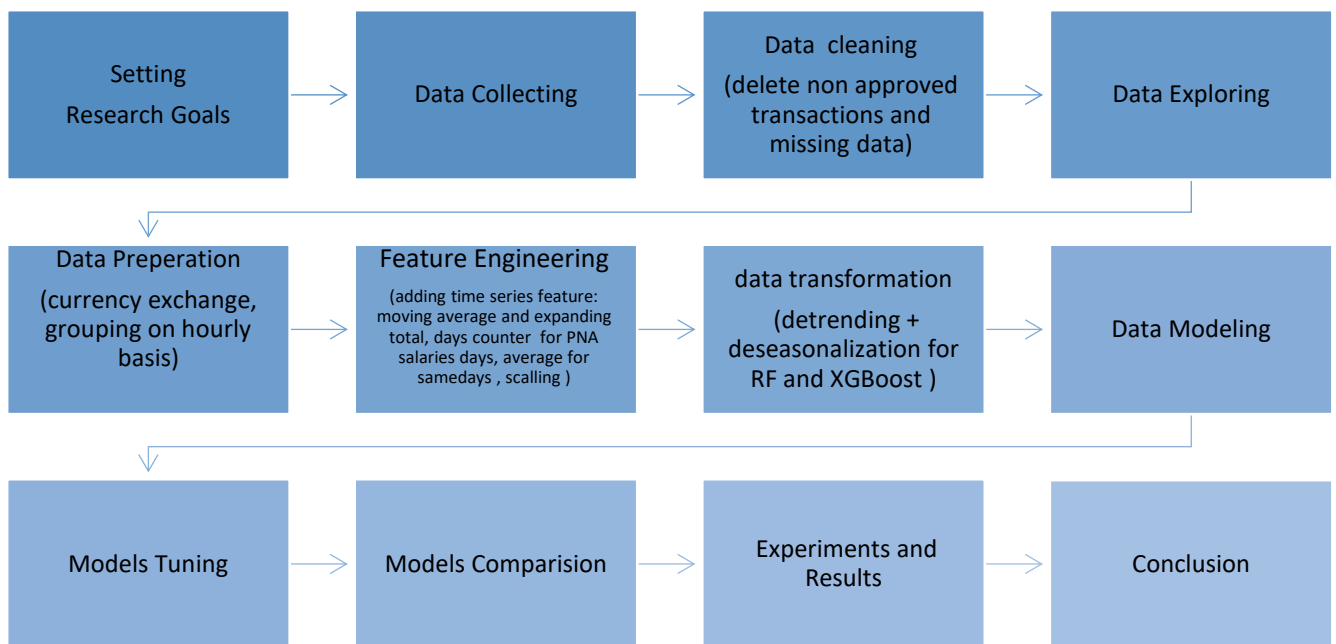


Figure 4-1 Proposed Methodology

### 4.2. Dataset

Banking sector data are recognized with high sensitivity, and banks are committed to meeting compliance requirements with the Payment Card Industry Data Security Standard (PCI DSS)

[28], which makes the process of sharing these data very complicated and restricted, in the other hand, banking sector like any other industry; spare no effort to retain its customer [12].

The dataset belongs to Bank of Palestine (BOP), the most popular bank in Palestine with the largest number of clients, and the biggest network of ATMs, as shown in table 4-1, which makes it the most suitable bank in Palestine for this research.

Table 4-1 BOP statistics 2017-2018 [32] [33]

Item	2017	2018
Clients	<b>917,076</b>	<b>About 1,000,000</b>
Branches	<b>91</b>	<b>94</b>
market share of credit facilities	<b>31.37%</b>	<b>32.97%</b>
market share of deposits	<b>31.47%</b>	<b>30.5%</b>

The raw data consists of 15 features, these features, and its description is included in Appendix (3) below.

Table 4-2 shows a statistical summary of the dataset:

Table 4-2 Statistical Summary

	<b>Information</b>
First transaction date	<b>1/1/2017</b>
Last transaction date	<b>31/12/2018</b>
Number of records	<b>8,999,973</b>
Number of ATMs	<b>172</b>
Approved Transactions	<b>7,947,351</b>
Nonapproved Transactions	<b>1,052,622</b>
Mean of Withdrawal Amount Daily	<b>8,073,021</b>
Standard Deviation of Withdraw Amount	<b>7,022,423</b>
Median of Withdrawal Amount Daily	<b>6,247,360</b>
Minimum Withdrawal Amount Daily	<b>0</b>
Maximum Withdrawal Amount Daily	<b>53,370,500</b>

A statistical summary was provided in Table 4-1, which shows that the dataset covers the period of 2017-2018, with 8,999,973 transactions. Some of these transactions were made by clients from banks other than BOP, banks with the agreement of facilitating are Arab Islamic Banak, Palestinian Islamic Bank, and Palestine Investment Bank, other banks' ATM holders used BOP ATMs through the National Switch System (194), this system operated by the Palestine Monetary Authority, which provides a unified network for all ATM banks and electronic points of working in Palestine, which enable any ATM cardholder to withdraw from any ATM within this network, or buy from any sale point [29].

### **4.3. Data Cleaning**

In this section the dataset was examined, the data observations that does not belong to this thesis scope were excluded.

#### **4.3.1. Drop nonsuccessful transactions**

Approved and successful transactions were included in this research, it contained 7947351 transactions, while 1,052,622 non-successful transactions were excluded due to many reasons since they are not related to the model forecasting scope.

#### **4.3.2. Missing values**

It was found that most of 11/2018 are missing; the series was not suitable for time series analysis [30], so the data of 11/2017 and 12/2017 were deleted from the dataset, as any time series data should be measured on a regular time interval basis.

#### **4.3.3. Outliers Detection**

According to Figure 3-5 and dates of PNA salaries transfer, outliers can be visually identified, they occurred on PNA salaries transfer days and 2-3 days after it, these outliers have obvious periodicity, which means it is part of the feature space and cannot be excluded [15], so outliers were included in the analysis.

### **4.4. Data Preprocessing**

#### **4.4.1. Currency Exchange**

As shown in Figure 3-2, BOP ATMS provides 3 currencies: USD, ILS, and JOD, most withdrawals are in ILS, so the exchange prices were taken from PMA, the official institution for issuing currency exchange rates in Palestine, then all amounts in USD and JOD were transferred into ILS according to the exchange price in the same day of the withdrawal.

#### 4.4.2. Grouping transactions on an hourly basis

Time series is data observations captured in equally-spaced time [31] , Withdrawals for all ATMs were collected at the hourly level, to preserve the property of time series data, which is recorded at equal intervals.

#### 4.5. Feature engineering

New features were added using the date of the transactions, moving average and expanding of transactions amount and expanding, and a counter for PNA salaries transfer days, these new features are illustrated in Table 4-3:

Table 4-3 New Features Description

#	Feature	Method	Description
1.	'Rolling_mean_12_hours'	Rolling Window Feature	To increase the accuracy of the models' moving averages were used as data preparation steps [32], and rolling average features for past hours data were added on 12-,24-,36-,48-,60-, and 72-hour basis.
2.	'Rolling_mean_24_hours'		
3.	'Rolling_mean_36_hours'		
4.	'Rolling_mean_48_hours'		
5.	'Rolling_mean_60_hours'		
6.	'Rolling_mean_72_hours'		
7.	'expand'	Expanding Window Feature	An accumulated summation of the transaction amount, starting when 'dys_bet_trans' = 1 and stopping with the last day before the salary transfer
8.	'hour'	Time-Based Features	Extracting the order of the hour within the day
9.	'DAYOFWEEK'		Extracting the order of the day within the week
10.	'DAYOFMONTH'		Extracting the order of the day within the month
11.	'WORKDAY'		Is it a working day?
12.	'WEEKOFYEAR'		Extracting the order of the week within a year
13.	'Dys_bet_trans'	Domain-Specific Features	Days between PNA salaries transfer days: which is a counter between every two days of salary transfer, these days were obtained from the Ministry of Finance, which is responsible for salary transfer
14.	'Same_days_avg'		Average transaction amount for the same counter of the day

15	TRANSACTION_COUNT'		The number of transactions of all ATMs per hour
----	--------------------	--	---

Appendix (4) includes a list of the date-based features that were tested and excluded as they were not informative in the data modeling stage and their correlation to target feature were weak .

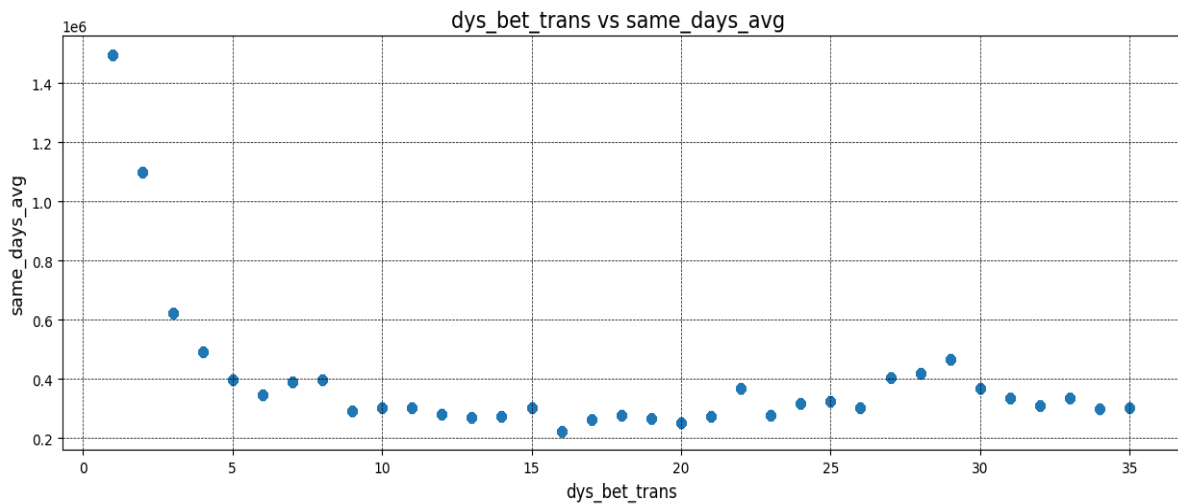


Figure 4-2 dys\_bet\_trans Vs same\_dys\_avg

Figure 4-2 shows the relation between same\_dys\_avg as the y-axis with dys\_bet\_trans as the x-axis in a scatter plot, which also emphasizes the strong relation between PNA salary transfer and the huge increase in transactions amount.

#### 4.6. Data Normalization

This preprocessing technique overcomes the different range of the scale of each feature, which unifies the weight of each feature in affecting the decision of the model, which minimizes high scale features effect and saves that of small scale features, which grantee equal weight for all features according to the model, there are many scalers in data science researches, Standardization, and min-max scalar are the most used for this purpose [33], in this study Min – Max scalar was used, as it keeps the importance of the outlier [34], as they represent the peak of transactions amount; which the bank wants to forecast the most

min-max scalar principle summarized in the equation 4.1 below:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.1)$$

$x'$  denotes the normalized value,  $x$  is the original feature value,  $x_{min}$  and  $x_{max}$  represents the minimum the maximum value of the feature respectively [35], this scalar is adequate for time series with stochastic and constant seasonal patterns [36].

#### 4.7. Time series decomposing

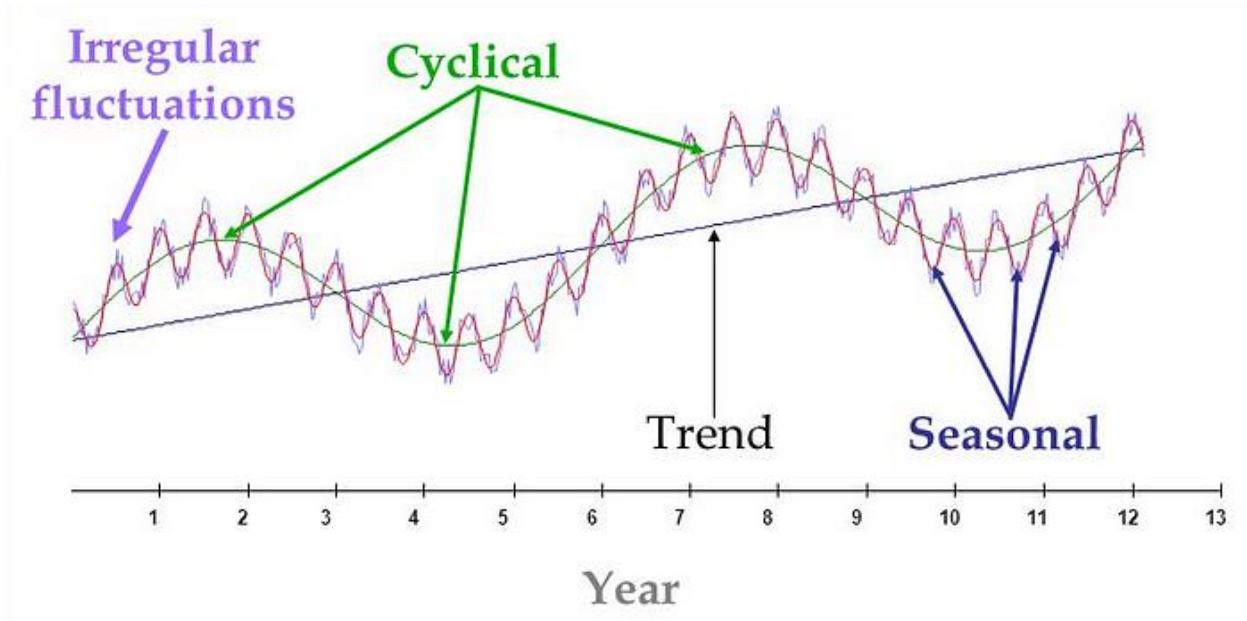


Figure 4-3 Time Series Components [41]

This process enables us to separate each of the time series into its components [37], as shown in Figure 4-3:

- Trend (T): the slow constant upward or downward direction of the data, which influences the data to increase or decrease gradually over the dataset period.
- Seasonality (S): it is repeated difficulties over a fixed period; it can be noticed visually as a repeated pattern.
- Cyclic (C): it is repeated ups and down over different period.
- Residual (R): it represents random error and all unexpected values or behavior of the time series that cannot be explained within the trend or the seasonality.

To define the relation between these components, it should be decided first if the time series belongs to an additive or multiplicative model. To choose what applies to the model in this study ; it should be distinguished between these two models:

**Additive model:** the value of the time series equals the summation of the components formerly mentioned, so the time series value  $Y$  at a certain time  $t$  will be:

$$Y(t) = T + S + C + R \quad (4.2)$$

The additive model has linear behavior, and linear seasonality with no change in amplitude or frequency [38].

**Multiplicative model:** the value of the time series equals the multiplication of the components formerly mentioned, so the time series value  $Y$  at a certain time  $t$  will be:

$$Y(t) = T * S * C * R \quad (4.3)$$

This equation yields nonlinear behavior, with decreasing or increasing seasonality characteristics like amplitude and frequency, having a graph resampling a funnel [39].

Removing one or more of the above-mentioned components will make the forecasting better and more accurate [40], as it will give the model clearer teaching and connect the inputs to the outputs in a more obvious relationship.

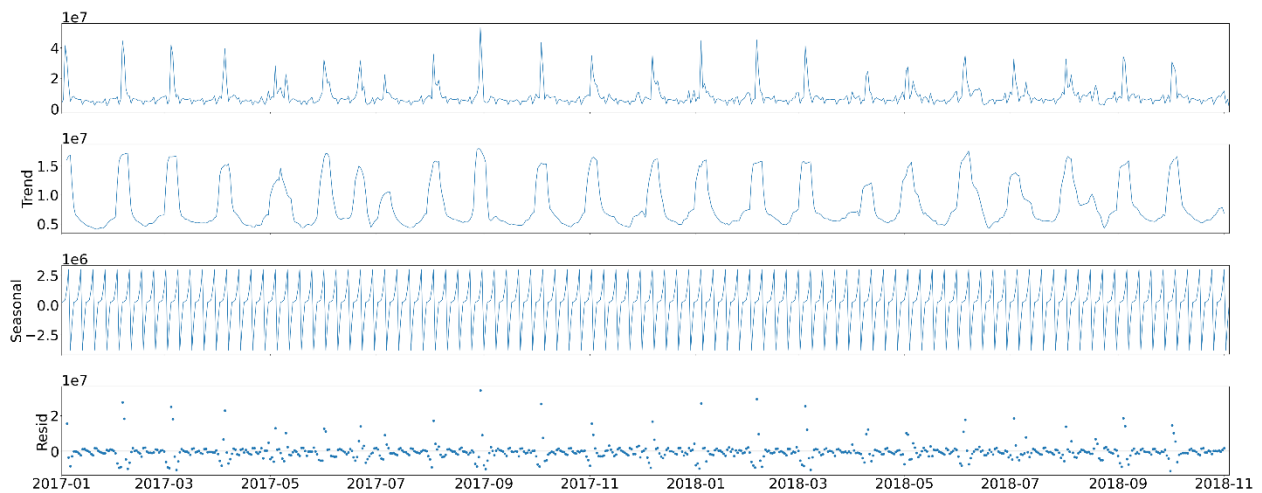


Figure 4-4 Transaction's amount series components

Figure 4-4 shows the decomposition of transaction daily amount, which shows a weak trend, and seasonal behavior due to what was mentioned in the previous figures in section 3.4, and it is easy to see

that the seasonality variation is constant, which means it does not vary in frequency and amplitude, and it does not look like an increasing funnel, so additive model is assumed for dataset [27].

Three-time series decomposing techniques were used:

#### 4.7.1. Deseasonalization

It is a time series transforming technique, which removes the effect of the seasonality of the dataset, it is recommended to take this step with RNN algorithms if the data set has not homogeneous seasonality [41], there are many ways to deseasonalize, moving average technique was utilized in this research.

As shown in Figure 3-7 and Figure 3-8 , there is hourly seasonality which lasts for 24 hours

#### 4.7.2. Detrending

Trend removal transformations by detrending (DT) by simple differencing (DIF), the DT consists of the detection and removal of the trend found in a time series behavior [42]. A deterministic function is used to handle the observed trend to calculate and define the time effect on it, which may be shown as a fixed component in a time series model [43]

$$\hat{x}_t = x_t - a_t \quad \mathbf{R} \quad (4.4)$$

Where  $\hat{x}_t$  is the detrended time series,  $x_t$  the original data,  $a_t$  is the time effect of the observed trend

### 4.8. Machine Learning Algorithms

In this section, three chosen machine learning algorithms are to be look over: Random Forest, XGBoost, and LSTM, according to the related work like this research, these two algorithms' performance is the best in many research and different cases, compared to other ML algorithms used in these researches.

Mapping between input and output values with a diverse scale is called regression, the input features are considered the dependent values that affect the values of the output, and the dependent variable the selected features of the data set will be utilized to teach the models the nature of the data set, teaching and testing process will share the date set according to percentages, the higher percentage goes to the training [44]

#### 4.8.1. Random Forest RF

The Random Forest model proposed by L. Breiman in 2001 [45], belongs to the decision trees family, it simply represents the “divide and conquer” principle, and it utilizes several random trees for making predictions from different rows and columns out of the training data set, using a bootstrapping process to train the trees. This process takes out any correlation between these trees and reduces variance. It does

this by using the bootstrapped random samples from the input training data, at run time, each tree produces a prediction and the model then takes the mean of each of the individual predictions to make an overall prediction result.

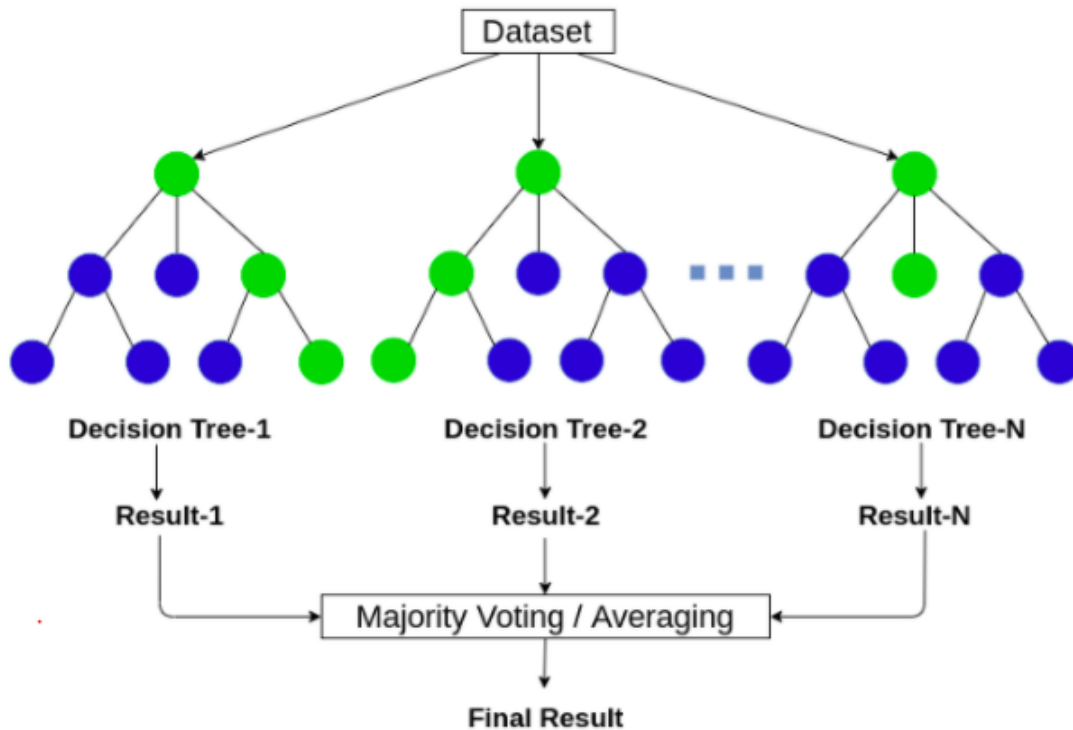


Figure 4-5 Random Forest regressor [46]

Random Forest has many hyperparameters, each one of effect and being affected by the other, besides the characteristics of the data set [47], all which control RF model performance, the most important are summarized as follows:

- **Maximum depth:** The largest number of steps or splits between the first (root) node and the furthest leaf node, overfit occurs when the maximum depth it is greater than the appropriate number for optimal accuracy.
- **The number of trees:** it has an impact on the model. The number of trees is one of the hyperparameters to be tuned, the optimal value is achieved when the balance between having an accurate, increasing number of trees with adequate depth can overcome high dimensionality issues [48].

- Minimum sample split: it is the minimum observation for each node to split, higher values can prevent overfitting by reducing splitting.

#### 4.8.2. eXtreme Gradient Boosting XGBoost

Is one of the tree-based ML models, it was first introduced by Chen and Guestrin in 2016 [49], it highly concentrates on scalability, parallelization, and distributed execution, it depends on building different regression models and training the one by one, each one will be trained with the residuals in the previous model, which means that it will correct the error the previous model, which yield a strong model which overcomes the weakness of each model in this group. [50]

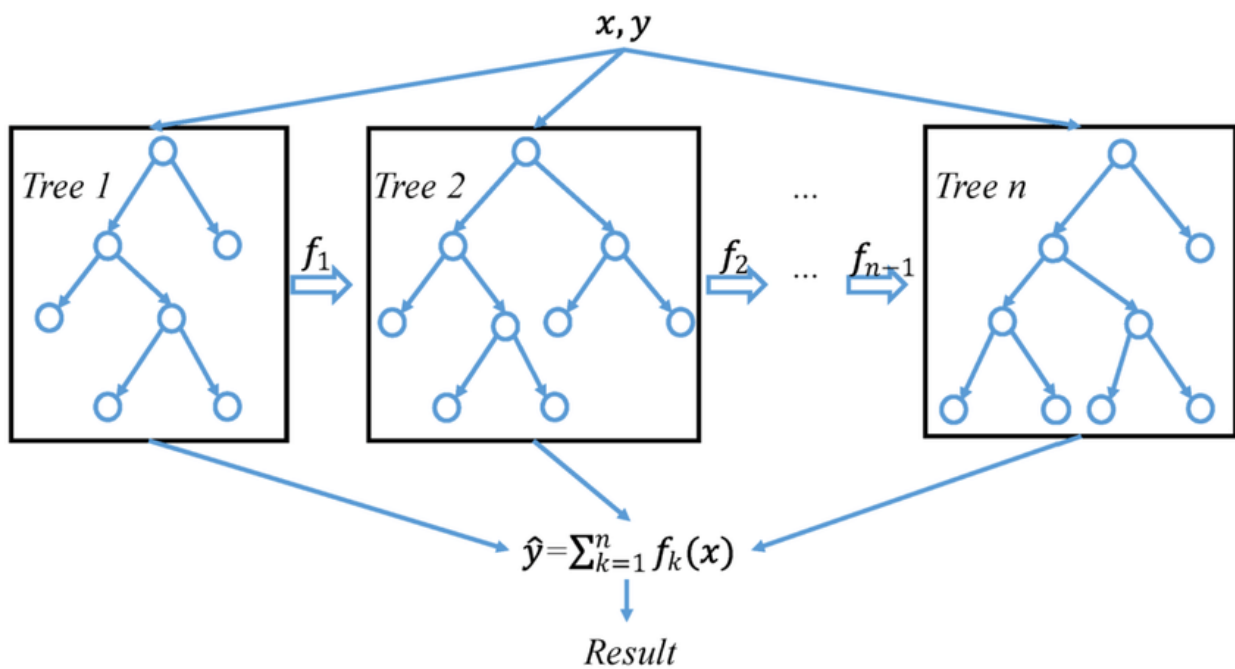


Figure 4-6 General Architecture of XGBoost [51]

The performance of the XGBoost is linked to a set of seven hyperparameters used; maximum depth of the trees and the number of trees like any decision tree model, the number of boosting rounds, learning rate, minimum weighted observations in a child node, minimum loss reduction, regularization parameters, and more [52, 53].

In this study, besides the maximum depth of the trees and the number of trees, the following hyperparameters were under focus:

- coal sample by tree: it refers to the percentage of columns to use for a random sample at each tree [54].

- Learning Rate: it is also called the shrinkage factor, which regulates the weight of each new tree contribution in the model performance, and it also determines the percentage of the model adaptation to the training data [55].

In this research, a finetuned XGBoost model is utilized to forecast the hourly ATM cash withdrawals, its performance was evaluated and compared to the other model, details are included in the next chapter.

#### 4.8.3. Long Short-Term Memory LSTM

LSTM belongs to the Recursive Neural Network RNN family, RNN was invented in 1990 by Elman [56] to enable ANN to learn temporal dependencies in sequence data, RNN has the advantage of sharing parameters over different parts of the model, although RNN has loops helps to preserve previous observations, it can't do that for long sequences, which can be explained by small weights decays exponentially, and leads to lose and forget corresponding observation [57], this situation leads to the need of RNN with abilities to keep long and short memory, which was first introduced in 1997 [58], and that was LSTM.

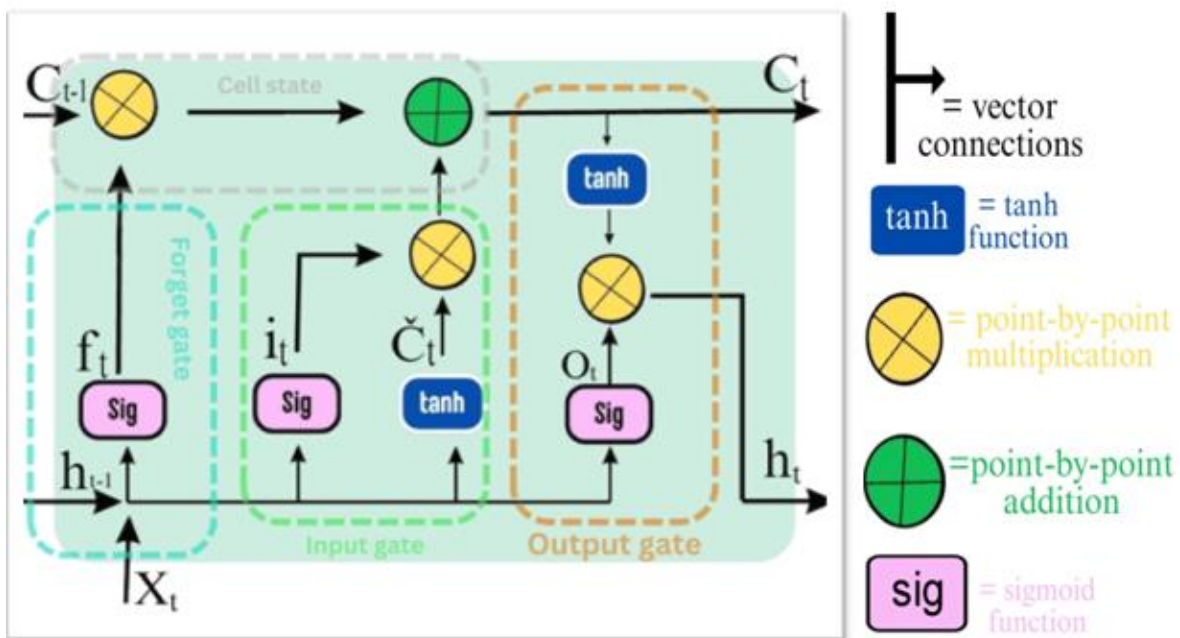


Figure 4-7 LSTM cell

As shown in Figure 4-7, LSTM consists of two paths and 3 gates: forget, input, and output, and it has three inputs: the new observation vector  $T_x$ , the long-term memory  $C_{t-1}$ , the short-term memory  $H_{t-1}$ , and it has two outputs: the new long-term memory  $C_t$  and the new short-term memory  $H_{it}$ ,

Forget Gate this gate's role is to decide what percentage of long-term memory to keep

$$f_t = \sigma(W_f |h_{t-1}, x_t| + b_f) \quad (4.5)$$

$t$  = indexes the time step

$f_t$  = forget gate's activation vector

$\sigma$  = sigmoid function.

$W_f$  = the weights of the forget gate

$h_{t-1}$  = hidden state vector, output vector of the LSTM unit,  $h_0 = 0$

$x_t$  = input vector to the LSTM unit

$b_f$  = bias vector parameters of forget gate

Input Gate: the role of this gate is to decide how to update the long-term memory

$$i_t = \sigma(W_i |h_{t-1}, x_t| + b_i) \quad (4.6)$$

$$\overline{C}_t = \tanh(W_c |h_{t-1}, x_t| + b_c) \quad (4.7)$$

$i_t$  = input/update gate's activation vector

$\overline{C}_t$  = cell input activation vector

$W_c$  = weights of the cell state

$W_i$  = weights of the input gate

$b_c$  = bias vector parameters of cell state

$b_i$  = bias vector parameters of input gate

Cell State is the path contains the long-term memory, and the most important part, which is updated through pointwise operations with the LSTM gates.

$$C_t = f_t C_{t-1} + i_t \overline{C}_t \quad (4.8)$$

$C_t$  = cell state vector,  $C_{t0} = 0$

Hidden state is the path for short-term memory, output gates is responsible of updating the short-term memory in the hidden state

$$o_t = \sigma(W_o |s_{t-1}, xt| + b_o) \quad (4.9)$$

$$s_t = o_t \tanh(c_t) \quad (4.10)$$

Output Gate: the role of this gate is to update the short-term memory

LSTM performance is enhanced by hyperparameter tuning, these parameters can include:

- **Optimizer:** optimizer controls the update process of the weights and biases of the model, Adaptive moment estimation (ADAM) were used in this research, since it can be taken as good default choice [44], and regarding the related work in the field of ATM cash analysis [59, 22], it dynamically updates learning rates based on individual adjusting of the weights [60], it also own the befits of both Adaptive Gradient algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp).
- **Learning rate:** this hyperparameter define the speed and size of the steps to reach the loss function minimum and reach the optimal weights of the deep learning model, it controls if the newly earned information can overwrite old ones [61]. A tradeoff should be made when tuning the learning rate, low values are very slow and can be stuck at an undesirable local minimum, while high learning rate values make the learning process jump in wide step; which may lead to skipping the optimal solution [62]. It is advisable to use a high learning rate at the initial iteration, while a low learning rate is kept for the rest of the iterations [63].
- **Number of epochs and batch size:** an epoch is when the algorithm is trained with the entire dataset for one time [64], The model is fed with data in the form of batches, and the batch size is another hyperparameter to tune, it is the number of observations that must pass through the algorithm before updating the inner parameter of the model [65], there is inherent tug-of-war to decide the size of the batch, big batch size leads to

converges slowly to the optimal solution with accuracy being negatively influenced [66], while although small batch size helps the model to converges faster as it starts learning before holding the whole dataset, it does not guarantee to convergence to global optima [67].

- Number of hidden layers: they are the intermediate layers between the input and output layer, they utilize the activation to extract the output from the input vector, finding the optimal number of hidden layers is confusing and always achieved by trial and error [68], there many advice in the starting default number; tow hidden layer are capable to solve most of the problems according to [44]; while three hidden layers are optimal for regarding the goal of achieving low time complexity and qualified accuracy in [68].
- Dropout: it's a generalization technique, that was first described in 2012 [69], it aims to prevent co-adaptation of feature detectors; which simply occurs when some neurons are dependent on other neurons, dropout fixes this by randomly choosing and eliminating the number of neurons (including their connections) in each observation presentation, it is recommended to set dropout to 0.2 in input layer and 0.5 for hidden layers [70, 44]. Connected network is exposed to overfitting, which Reduces performance in DN with many parameters, dropout overcomes this issue and improves the performance in very complicated tasks like vision and speech recognition [71].

#### 4.9. Metrics Selection

Models' performance comparison is based on globally defined error terms, these terms measure two values: performance and error [72], in this study three metrics were used for evaluation and comparison: time, R2, MAPE, RMSE, and MAE.

**Time:** The time for each model to fit the training data, was measured in seconds.

**The coefficient of Determination ( $R^2$ )** has a value between 0 and 1, it represents the percentage that the model was able to explain of response variable change, minus values mean that the prediction is worse than take mean values and prediction [44]. Equation 4.11 shows the  $R^2$  formula.

$$\text{coefficient of determination}(R^2) = 1 - \frac{SS_{residuals}}{SS_{total}} \quad (4.11)$$

Where:

- $SS_{residuals}$  – sum of squared residuals
- $SS_{total}$  – The sum of all square's observations deviation from the mean

**Mean Absolute Percentage Error (MAPE):** which express the quality of the prediction for forecasting:

$$MAPE = \frac{1}{t} \sum_{i=1}^t \left| \frac{A_t - F_t}{A_t} \right| \quad [73] \quad (4.12)$$

**Root Mean Square Error (RMSE)** is equal to the square root of the average squared error. Equation 4.13 shows how to calculate RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.13)$$

**Mean Absolute Error (MAE)** is the mean of the absolute value of the errors. Equation 4.14 shows how to calculate MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (4.14)$$

#### 4.10. Conclusion

This chapter summarizes the methodology of forecasting next hourly withdrawals; the methodology starts with data preparation and exploration, then the data preprocessing for machine learning; the third step is to utilize of different machine learning algorithms, and finally use different performance and error metrics to compare between RF, XGBoost, and LSTM performance and select the best model.

## 5. Chapter Five: Experiments and Result Analysis

### 5.1 Introduction

After RF, XGBoost, and LSTM models were deployed, in this chapter the error terms are used to evaluate the models' efficiency, an illustration is the forecasting results, the performance metrics that include  $R^2$ , RMSE, Time, MAPE, and MAE, for each algorithm.

### 5.2 Experimental Environment

Sktime Version 0.25.0 and Scikit-Learn library Version 1.2.2 libraries were used to machine learning function and environment Python 3.10.12 Development IDE is Google Colab.

The test server operating system is Windows 10 Pro 21H2 64-bit OS. The processor of the machine is Intel® Core® i5-8265U CPU @ 1.60GHz 1.80 GHz, and the installed memory RAM is 8 GB.

### 5.3 Dataset preparing for models

This section contained the clarification how data were prepared before start modeling, it contained these two steps:

#### 5.3.1 Train & Test Set Split

It is inevitable to use split into train and test, this split enables us to measure and estimate performance of the model on test data, the percentage were 80% for training to 20% for testing.

#### 5.3.2 Time-based cross-validation.

This approach includes an Expanding Window and Sliding Window. In **Expanding Window** Figure 5-1, the training set is being extended by a fixed number of data points in each run. This way creates multiple train-test subsets. The process takes place until the training set reaches a specified maximum size. In **Sliding Window** Figure 5-2, with a fixed size of the training set and move it across the data.

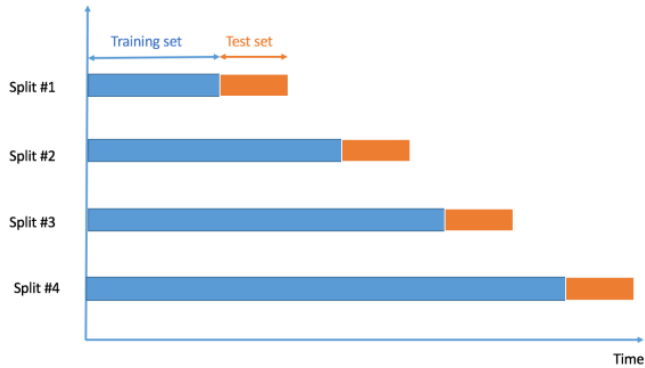


Figure 5-1 Expanding window [45]

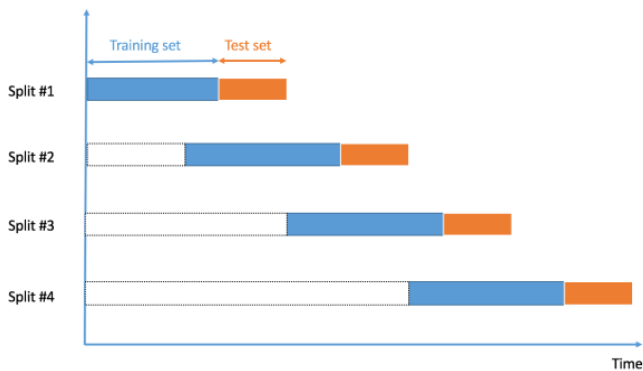


Figure 5-2 sliding window [45]

In this study expanding window approach was applied, as the data sets were grouped on an hourly basis which made the data set medium in size, which makes the expanding window better, and enables the researcher to balance between model performance and training time [74].

#### 5.4 Generating Models

In this study, 3 different machine learning models were fitted, Random Forest, XGBoost, and LSTM, they were trained and tested on the same data sets, which enabled us to hold fair comparison between them, each model forecasted the same target feature: the total transaction amount for the next hour, the rolling mean of total transaction for the next 2 hours, 4 hours, 6 hours, 8 hours, 10 hours, 12 hours, 14 hours, 16 hours, 18 hours, and 20 hours.

#### 5.5 Parameter Fine-Tuning

Each one of the 3 models, has its parameters which needs to be tuned, this tuning enables us to find the optimal performance and accuracy, to find out that, Grid Search were applied for RF

and XGBoost, which run and measure all possible combinations of selected parameters and defined the best of them according to predefined performance measure [75], while finding the optimal parameter for LSTM depended on trial and error with different combinations of : learning rate, number of hidden layer, and batch size .

The selected parameters for the models and the optimal parameter can be seen in Table 5-1:

Table 5-1 Parameter Fine-Tuning

Regression Models	Possible Parameters	Best Parameters
RF	n_estimators: range [100, 400] max_depth: range: [5, 20] min_samples_split: [2,3] min_samples_leaf = [ 1,2, 3]	n_estimators: [ 300] max_depth: [10] min_samples_split:[2] min_samples_leaf : [1]
XGBoost	n_estimators: range [100, 400] max_depth: range [5, 15] learning_rate: range [0.001, 0.1] colsample_bytree: range [0.3, 0.7]	n_estimators: [300 ] max_depth:[ 8] learning_rate: [0.01] colsample_bytree:[ 0.6]
LSTM	batch size: [4,5,6,7, 10, 15, 32] number of neurons in the hidden layer: [20, 30, 40] number of neurons in the input layer: [50, 60, 70, 80] learning rate: [0.01,0.001, 0.0001] dropout input layer: [0.5, 0.4, 0.3, 0.2] dropout hidden layer: [0.5, 0.4, 0.3, 0.2] number of hidden layers [1,2,3]	batch size: [4] input layer: [60 neurons] hidden layer: [30 neurons] learning rate: [0.001] dropout input layer [0.2] dropout in hidden layer [0.5] hidden layer: [1]

A different combination of learning rate, number of hidden layers, and batch size were tested to find the optimal value for each one of them, these combinations and their results are shown in table 5-2.

Table 5-2 LSTM trial and error combinations results

Learning Rate	H. layer	Batch	TIME:	R2:	MAE:	MAPE:	RMSE:
0.01	1	4	175.32	0.83	0.02	3.27	0.03
0.001			175.54	0.91	0.01	0.37	0.02
0.01	2		268.16	0.56	0.03	3.76	0.05
0.001			267.56	0.87	0.02	0.48	0.03
0.01	3		328.54	0.72	0.03	2.25	0.04
0.001			329.56	0.88	0.02	0.42	0.03
0.01	1	8	145.73	0.82	0.02	0.45	0.03
0.001			147.09	0.86	0.02	0.41	0.03
0.01	2		115.80	0.78	0.02	0.49	0.04
0.001			147.95	0.86	0.02	0.41	0.03
0.01	3		149.19	0.36	0.04	0.79	0.06
0.001			149.00	0.83	0.02	0.40	0.03

then a new model with the best combination (batch size = 4, one hidden layer, learning rate = 0.001) was examined under different numbers of epochs, as stated in Table 12, 25 epochs was the optimal number of epochs.

Table 5-3 Epoch number tuning results

Epoch	TIME:	R2:	MAE:	MAPE:	RMSE:
15	326.613	0.896	0.015	0.38	0.023
20	446.891	0.902	0.015	0.375	0.023
25	506.296	0.911	0.014	0.370	0.023
30	566	0.889	0.016	0.384	0.024

## 5.6 Results

This section will display and explain the results of each model, different presentation methods were used; table with best score in red, various plots were used to clarify the results, and to facilitate the comparison process.

### 5.6.1 Random Forest Results

Table 5-4 RF error terms value the best results for each model are shown in bold

RF model predicting target:	Training Time:	R2:	MAE:	MAPE:
transaction amount for the next 1h	<b>76.08</b>	0.87	85820.71	0.29
next 2h Rolling mean	79.35	0.88	82240.32	0.25
next 4h Rolling mean	77.97	<b>0.89</b>	<b>78001.25</b>	0.26
next 6h Rolling mean	85.71	0.88	82933.51	0.22
next 8h Rolling mean	78.41	0.87	82112.05	0.20
next 10h Rolling mean	78.38	0.87	81802.21	<b>0.19</b>
next 12h Rolling mean	78.30	0.85	81271.40	<b>0.19</b>
next 14h Rolling mean	79.06	0.84	81984.59	<b>0.19</b>
next 16h Rolling mean	78.65	0.82	82089.37	<b>0.19</b>
next 18h Rolling mean	79.23	0.79	83797.41	0.20
next 20h Rolling mean	79.39	0.75	89486.51	0.21

Table 5-4 shows that time for all the RF experiments ranges: from 76- 85 seconds which is very good, R2 is gradually increases and reaches its highest value at 4h rolling mean, then it changes into gradually decreasing when the forecasting hour increased, MAEs, performance matches that of R2, and MAPE Showed somewhat stability at 6h-20h rolling mean, which indicates good forecasting ability, and indicates that the average absolute percentage difference between the predictions and the actuals is 19-22%.

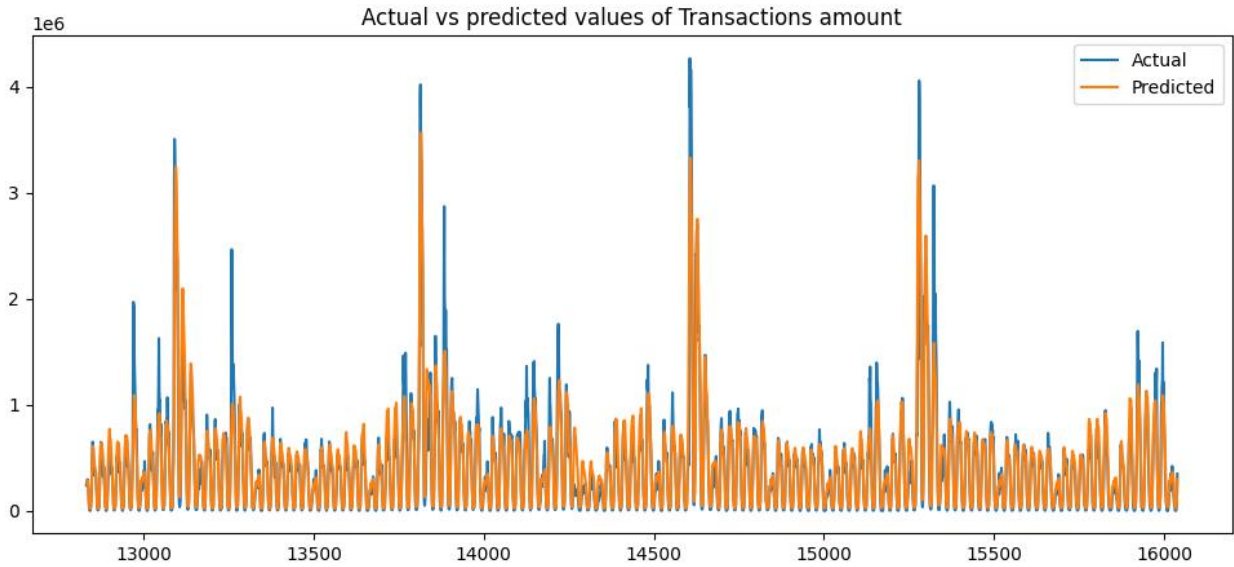


Figure 5-3 Actual vs predicted values of Transactions amount for RF

Figure 5-3 shows that RF forecasts the low values almost perfectly, while cannot forecast the higher values, which determines the source of error, which resembles RF in this fact.

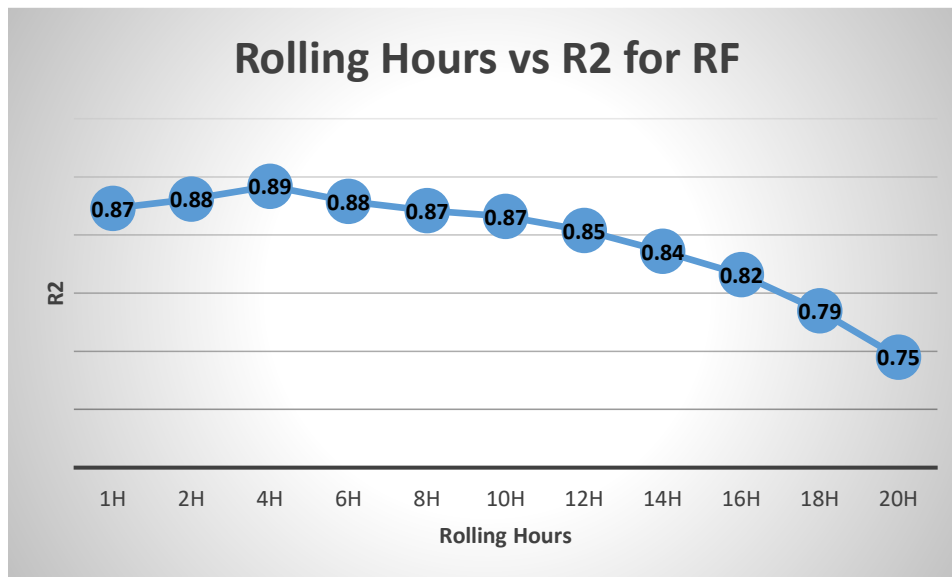


Figure 5-4 Rolling Hours vs R2 for RF

Figure 5-4 shows that R2 is gradually increasing and reaching its highest value at 4h rolling mean, then it changes into gradual decreasing when the furcating hour increases.

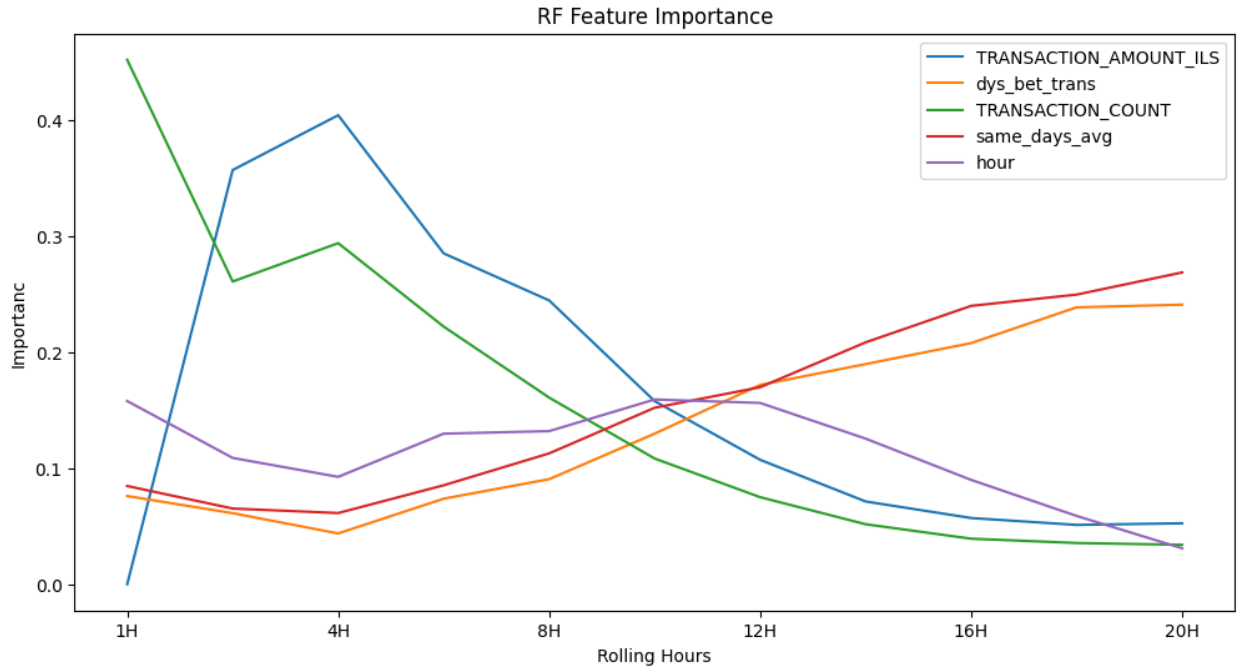


Figure 5-5 Highest 5 features importance

As shown in Figure 5-5, TRANSACTION\_AMOUNT\_ILS and TRANSACTION\_COUNT have higher values at low values of rolling hours and they decrease as rolling hours increase, while dys\_bet\_trans and same\_days\_avg act the opposite, they have low values at higher values of rolling hours. The Hour feature oscillates between 0.16 and 0.1.

### 5.6.2 XGBoost results

Table 5-5 XGBoost error terms values the best results for each model are shown in red.

XGBoost model predicting target:	Training Time:	R2:	MAE:	MAPE:
transaction amount for the next 1h	<b>67.159</b>	<b>0.886</b>	<b>94510.882</b>	0.345
next 2h Rolling mean	68.949	0.815	110023.006	0.313
next 4h Rolling mean	67.924	0.85	100600.056	0.485
next 6h Rolling mean	71.037	0.807	112536.029	0.31
next 8h Rolling mean	77.645	0.805	112611.809	<b>0.295</b>
next 10h Rolling mean	73.692	0.766	116956.523	0.296
next 12h Rolling mean	73.322	0.736	117582.145	0.297

next 14h Rolling mean	71.999	0.641	130892.333	0.318
next 16h Rolling mean	69.434	0.566	138603.005	0.34
next 18h Rolling mean	81.234	0.368	152171.951	0.379
next 20h Rolling mean	68.038	0.037	167074.567	0.46

Table 5-5 shows that time for all the RF experiments ranges: 67-81 seconds which is very good, and it is the least of the three models, but it does not intercede with poor performance in general, as R2 and MAE start in competitive value of 0.886 and 94510.882 respectively, which retreated until it arrived is gradual until it reached 0.37 and 167074.567 respectively, and MAPE had unstable performance with values range of 0.295 – 0.485.

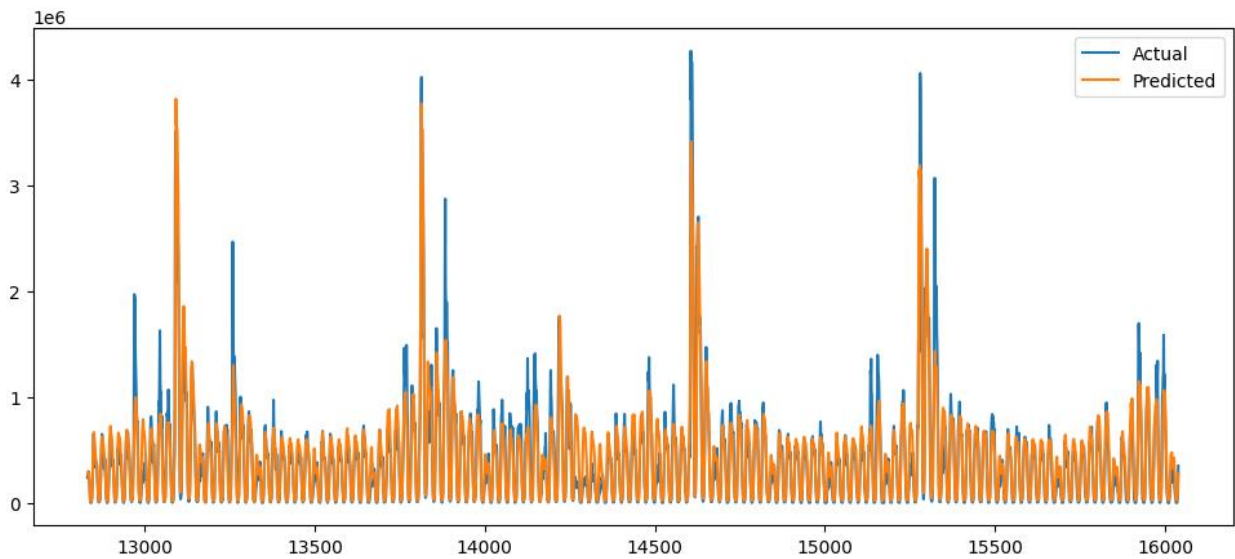


Figure 5-6 Actual vs predicted values of Transactions amount for XGBoost

Figure 5-6 shows that XGBoost forecasts the low values almost perfectly, while could not forecast the higher values, which determines the source of error.

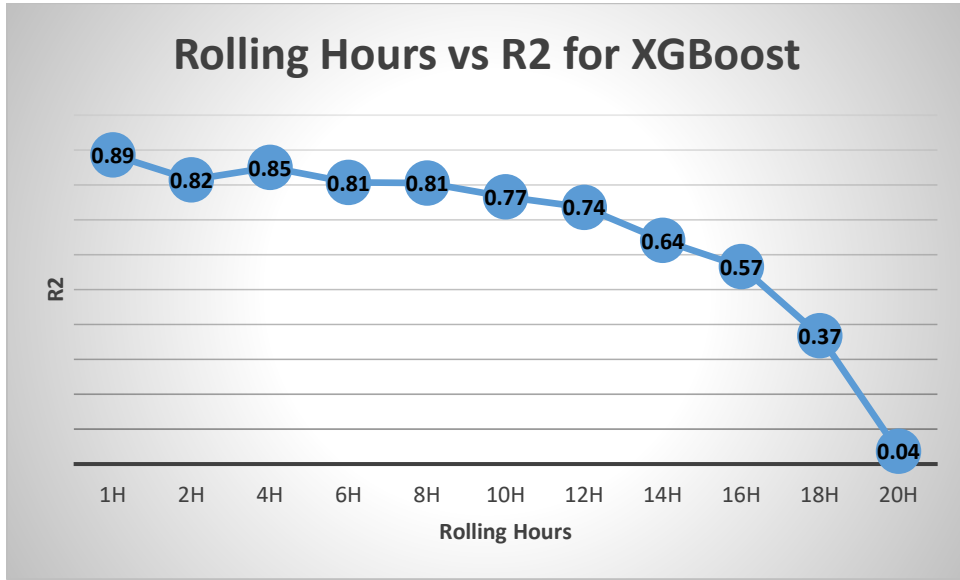


Figure 5-7 Rolling Hours vs R2 for XGBoost

Figure 5-7 shows that XGBoost starts with a high score of 0.89, then shows an unstable decrease of 2h- 12 hour rolling mean, then there is an accelerated decrease until it reaches the value of 0.04 at 20h rolling mean.

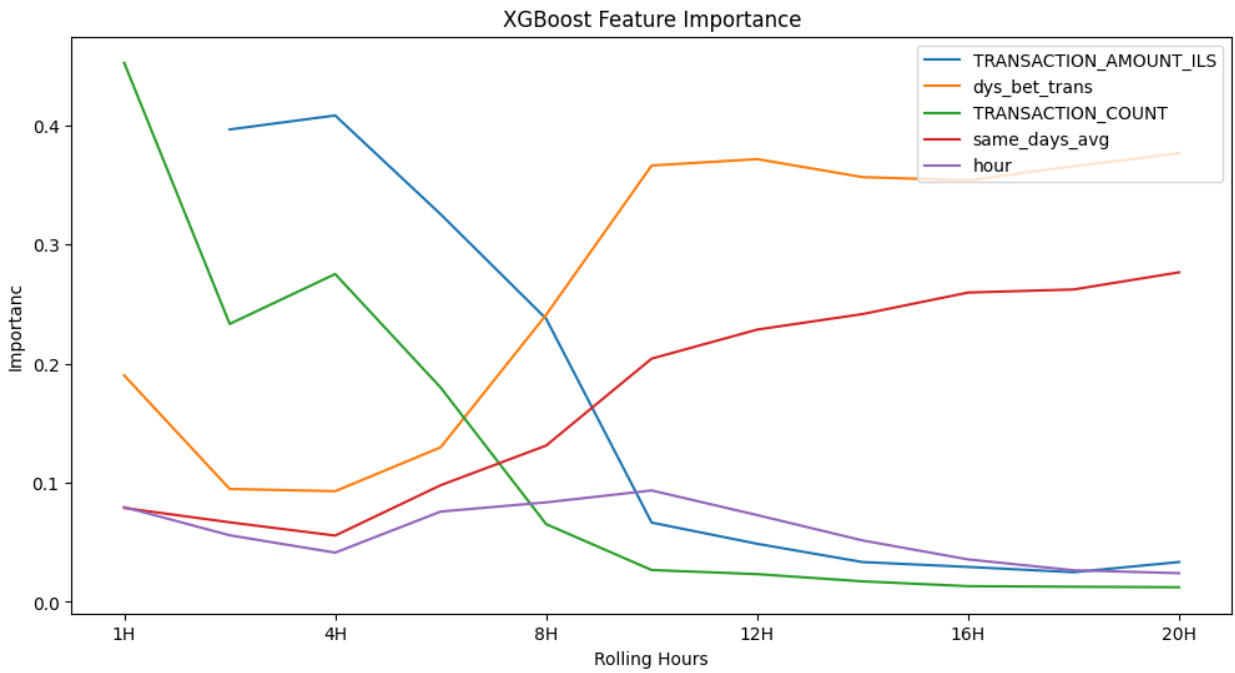


Figure 5-8 Highest 5 features importance

As shown in Figure 5-8, TRANSACTION\_AMOUNT\_ILS and TRANSACTION\_COUNT have higher values at low values of rolling hours and they decrease as rolling hours increase, while dys\_bet\_trans and same\_days\_avg act the opposite, they have low values at higher values of rolling hours. The Hour feature oscillates between 0.09 and 0.03

### 5.6.3 LSTM results.

Table 5-6 LSTM error terms value the best results for each model are shown in red.

LSTM model predicting target:	Training Time:	R2:	MAE:	MAPE:
transaction amount for the next hour	506.37	0.91	0.01	0.37
next 2h Rolling mean	505.57	0.96	0.01	0.28
next 4h Rolling mean	506.01	0.97	0.01	0.25
next 6h Rolling mean	505.94	0.98	0.01	0.21
next 8h Rolling mean	506.58	0.98	0.01	0.17
next 10h Rolling mean	506.22	0.98	0.01	0.15
next 12h Rolling mean	504.90	0.98	0.01	0.14
next 14h Rolling mean	506.71	0.98	0.01	0.13
next 16h Rolling mean	506.05	0.99	0.01	0.10
next 18h Rolling mean	505.56	0.99	0.01	0.09
next 20h Rolling mean	<b>479.81</b>	<b>0.99</b>	<b>0.01</b>	<b>0.08</b>

Table 5-6 shows the magnificent performance of LSTM, as it starts with competitive values at 1h in all error metrics, and increases to reach the highest values over the three model 20H rolling mean.

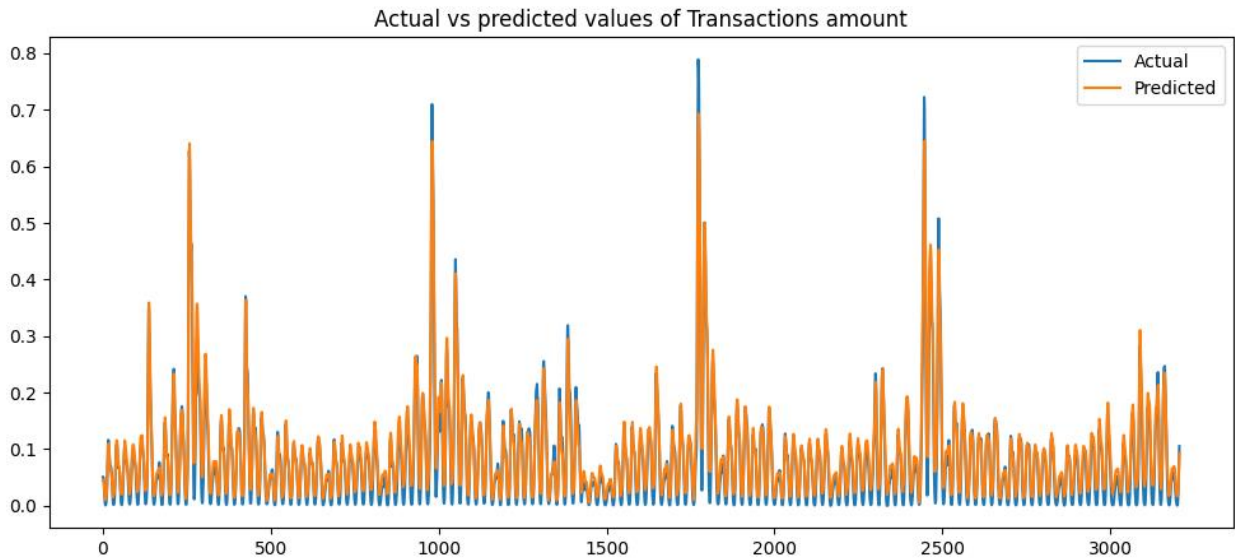


Figure 5-9 Actual vs predicted values of Transactions amount for LSTM

Figure 5-9 shows that LSTM has shortages in forecasting both low and high values, and since the low values are much more than high values, this reveals that the error term comes from low values forecasting.

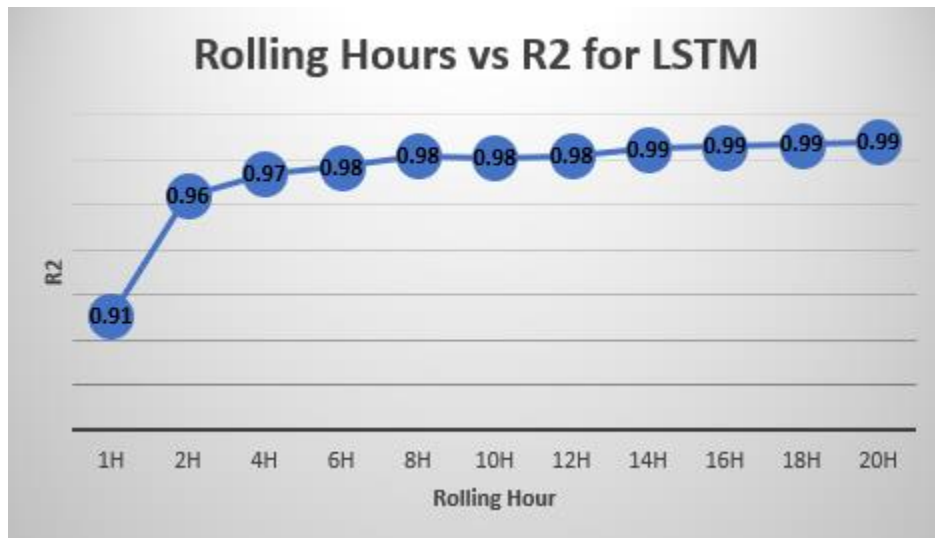


Figure 5-10 Rolling Hours vs R2 for LSTM

Figure 5-10 shows the magnificent performance of LSTM, as it starts with 0.911 at 1h, and rapidly increased to 0.96 at 2h rolling mean, then steadily increases to reach 0.99 at 20h rolling mean.

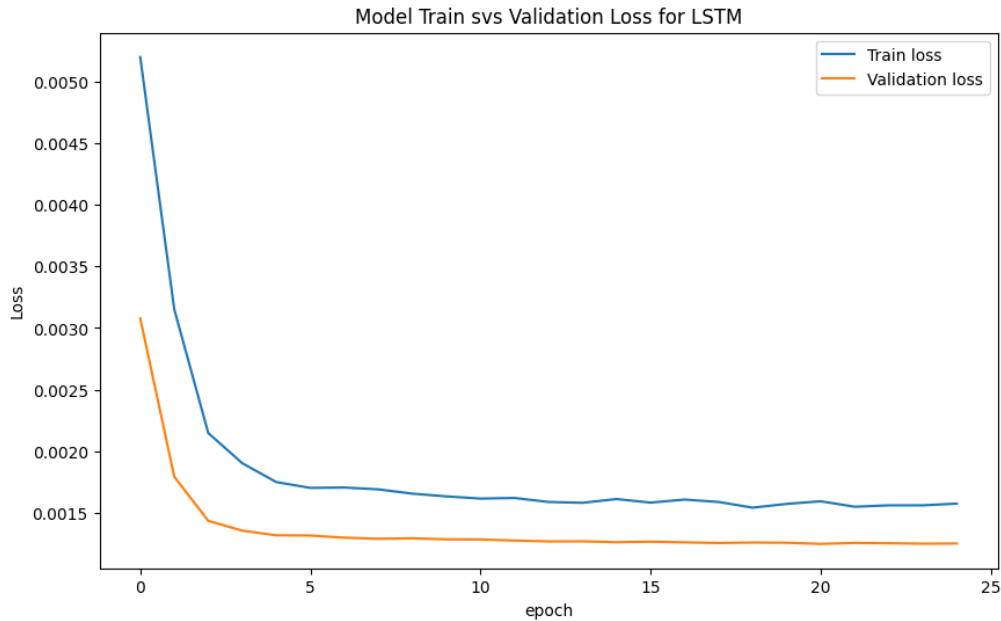


Figure 5-11 Loss vs validation for LSTM

Figure 5-11 shows the loss continues to decrease gradually and smoothly, and the validation loss stayed less than the testing loss, which shows the stable performance of the model and the non-existing of overfitting.

## 5.7 Discussion

After obtaining the results of the three models, a comparison between them will be on 3 axes:

### 5.7.1 Error term-based comparison

Table 5-7 Models error terms comparison

	LSTM			RF			XGBoost		
	TIM E:	R2 :	MAP E:	TIM E:	R2: R2:	MAP E:	TIM E:	R2: R2:	MAP E:
transaction amount for next 1h	506.37	0.91	0.37	76.08	0.87	0.29	67.16	0.89	0.35
next 2h Rolling mean	505.57	0.96	0.28	79.35	0.88	0.25	68.95	0.82	0.31
next 4h Rolling mean	506.01	0.97	0.24	77.97	0.89	0.26	67.92	0.85	0.49

next 6h Rolling mean	505.94	0.98	0.21	85.71	0.88	0.22	71.04	0.81	0.31
next 8h Rolling mean	506.58	0.98	0.17	78.41	0.87	0.20	77.65	0.81	0.30
next 10h Rolling mean	506.22	0.98	0.15	78.38	0.87	0.19	73.69	0.77	0.30
next 12h Rolling mean	504.90	0.98	0.14	78.30	0.85	0.19	73.32	0.74	0.30
next 14h Rolling mean	506.71	0.98	0.13	79.06	0.84	0.19	72.00	0.64	0.32
next 16h Rolling mean	506.05	0.99	0.10	78.65	0.82	0.19	69.43	0.57	0.34
next 18h Rolling mean	505.56	0.99	0.09	79.23	0.79	0.20	81.23	0.37	0.38
next 20h Rolling mean	479.81	0.99	0.08	79.39	0.75	0.21	68.04	0.04	0.46

Table 5-7 summarizes the results, where it is obvious that LSTM had won the competition on R2 values 0.91-0.99, which means it is the best that fits the data set and explains most of the variance in it.

On the MAPE level, RF has the best score when predicting the next 1h, 2h rolling mean, and 4h rolling mean, LSTM outperforms in the rest of the predicting hours.

### 5.7.2 Graph-based comparison

Regarding the results in Table 16, a comparison of the graph of RF in Figure 5-12 and LSTM in Figure 5-13 below

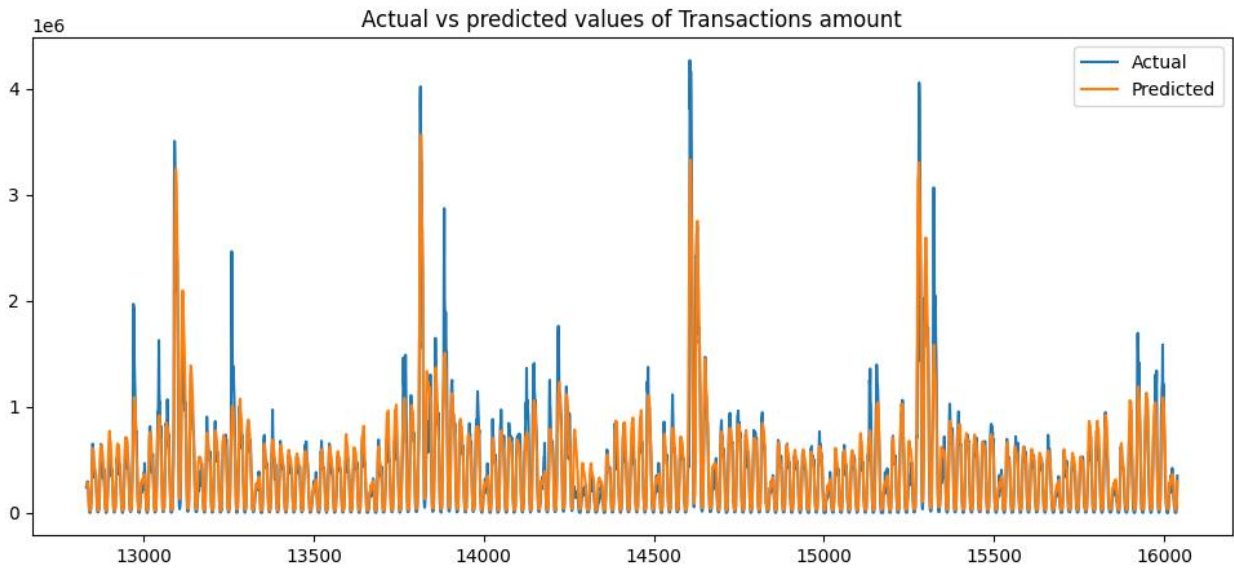


Figure 5-12 'Actual vs predicted values of Transactions amount for RF

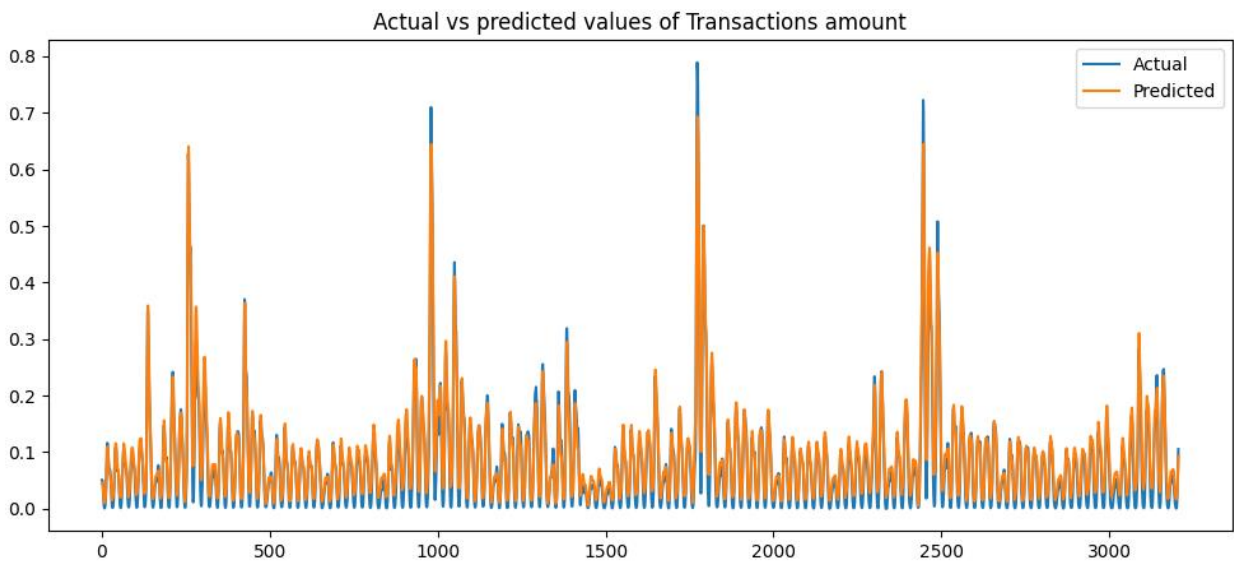


Figure 5-13 Actual vs predicted values of Transactions amount for LSTM

It is easy to see, LSTM shortages were in low values of the predicted values, RF did not have these shortages as it predicted low values better than LSTM, which is consistent with the researcher's study [14]. On the other side, bigger values were predicted better in LSTM than RF, these facts can be used efficiently in combining these two models, and take unique advantages of each of them [76, 77]; as they complement each other.

### 5.7.3 Time-based comparison

Training time is very important to compare one algorithm to another, even between models with different working principles, in this study , RF has 76-85 seconds, XGBoost has a range of 67-81 seconds and 480-507 seconds for LSTM, even LSTM has a longer period than RF and XGBoost it still the best, as the bank will run the model once every hour, LSTM execution time seems reasonable regarding the frequency of running it ta the bank.

### 5.7.4 Features-based comparison

As shown in Figure 5-5 and Figure 5-8 ,the features important the reflects over time, as it begins with high importance for both TRANSACTION\_AMOUNT\_ILS and TRANSACTION\_COUNT and low importance for both while dys\_bet\_trans and same\_days\_avg, then the importance of TRANSACTION\_AMOUNT\_ILS and TRANSACTION\_COUNT decrease and the importance of dys\_bet\_trans and same\_days\_avg increase, which leads us to the conclusion that the new features that were engineered, are important in helping RF and XGBoost in forecasting increases as forecasting rolling hours increase, it's obvious to see that the hour feature is one of the 5 highest features importance, which can be explained by hourly seasonality.

## 5.8 Conclusion

In this chapter hyperparameter tuning sets were illustrated, the results of implementing the proposed methodology were recorded, grouped, and discussed in different axes, and different metrics were used to describe and compare the performance of each model and among them, the best performance was by LSTM as stated above.

## **6. Chapter Six: Thesis Conclusion and Future Work**

### **6.1 Thesis Conclusion**

This thesis was based on the use of different machine-learning algorithms, to help the Bank of Palestine forecast the next hour's total withdrawal amount in its extensive ATM network, a dataset of about 9 million ATM transactions from 172 ATMs during 2017-2018 was collected, and provided to us by the Bank of Palestine, each bank needs to manage the cash amount in the ATMs, and forecasting the optimal amount replenishment is the first priority to achieve, as excess amounts should be avoided to lower expenses and insurance cost, and insufficient amounts to cover customers' needs lead to a decrease in their satisfaction and negatively affect the bank's reputation.

Raw data were cleaned and preprocessed, features were engineered, grouped on an hourly basis, scaled, and transformed using time series techniques, and then a sample of 80% of the data was used to train three different algorithms (RF, XGBoost, LSTM), a hyperparameter fine-tuning were performed to reach the best performance of each algorithm, these algorithms were utilized to forecast the next hour's total withdrawal amount and the rolling mean for the next different numbers of hours of all the ATMs belonging to the bank, LSTM obtained the highest accuracy score and the lowest error rate, with R2 value 0.91 for forecasting the first hour, and a range of 0.96 -0.99 for next 2-20 hours rolling mean rolling, and MAPE a range of 0.28-0.08 for next 2-20 hours rolling mean.

Exploratory data analysis reveals many facts about the reality of ATM transactions in Palestine, first, all outliers happened on PNA salary transfer days and a few days after, the last week before the transfer is also affected and has the least amount among the other weeks within the month, the effect of special holidays and religious Eid can't be noticed; as PNA delayed the salary until these days come, second, Thursday has the highest transaction count as it is the last day in working days of the week, while Friday and Saturday have the least demand for ATM transactions as they are the weekend in Palestine.

The use of the model in this thesis can contribute to the bank in achieving its strategic objectives, particularly, excellence in Customer Service and Financial Inclusion. The model can lead to an increase in successful transactions and a reduction in those rejected due to exceeding the daily limits by raising the Daily ATM Transaction limits.

Depending on the graph-based analysis of RF and LSTM, it is possible to find out that these two models can combine to complement each other in forecasting extreme values, which makes the results of the combined model to be better than the results of each of them alone.

## **6.2 Future Work**

ATM transaction analysis has a wide range of potential possibilities; the next few aspirations can be mentioned:

1. Including economic indexes or demographic factors, like inflation, and currency exchange rate.
2. Cluster ATMs according to the location, taking under concentration the population of commercial, entertainment, and governmental buildings.

## References

- A. Tahtamouni, "E-banking services and the satisfaction of customers in the Jordanian banks," *Journal of Science and Technology Policy Management*, 2022.
- A. Association of Banks in Palestine, "Banking Sector Statistics of 2022," *Banks in Palestine*, vol. 93, no. 4, pp. 100-105, 2023.
- E. W. Group, "Report of the ERPB Working Group on Access and Acceptance of Cash," ERPB Working Group on Access to and Acceptance of Cash , 2021.
- Canser BİLİR and Adil DÖŞEYEN, "Optimization Of ATM And Branch Cash Operations Using An Integrated Cash Requirement Forecasting And Cash Optimization Model," *BMIJ*, vol. 6, no. 1, pp. 237-255, 2018.
- Raza, Muhammad Rafi and Mohammad Taha Wahab and Muhammad Bilal Khan and Hani, "ATM Cash Prediction Using Time Series Approach," in *3rd International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2020*, Karachi, Pakistan, 2020.
- K. Venkatesh, V. Ravi, A. Prinzie, D. Van Den Poel, "Cash demand forecasting in ATMs by clustering and neural networks," *Eur. J. Oper. Res*, vol. 232, p. 383–392, 2014.
- Pankaj Kumar Jadwal, Sonal Jain, Umesh Gupta, Prashant Khanna, "K-Means Clustering with Neural Networks for ATM Cash Repository Prediction," *Springer International Publishing*, vol. 1, pp. 588 - 596, 2018.
- Yeliz Ekinci, Jye-Chyi Lu, Ekrem Duman, "Optimization of ATM Cash Replenishment With Group-Demand Forecasts," *Expert Systems with Applications*, vol. 42, no. 7, p. 3480–3490, 2015.
- Asad, M., Shahzaib, M., Abbasi, Y. and Rafi, M.,, "A Long-Short-TermMemory Based Model for Predicting ATM Replenishment Amount," in *International Arab Conference on Information Technology (ACIT)*, Giza, Eygpt, 2020.
- Ranade, Pushkar V. Dandekar and Ketki M., "ATM Cash Flow Management," *International Journal of Innovation, Management and Technology*, vol. 6, no. 5, pp. 343-347, 2015.

Cagatay Catal, Ayse Fenerci , Burcak Ozdemir, Onur Gulmez, "Improvement of Demand Forecasting Models with Special Days," . *Procedia Computer Science*, vol. 59, pp. 262-267, 2015.

Akber Rajwani, Tahir Syed, Behraj Khan, Sadaf Behlim, "ATM cash flow prediction: Exploratory Data Analysis and results from sequence-bases predictions," *Science Direct*, 2019.

Sefik Ilkin Serengil, Alper Ozpinar , "ATM Cash Flow Prediction and Replenishment Optimization with ANN," *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, vol. 11, pp. 402-408, 2019.

E. Gökçay, ATM Cash Stock Prediction Using Different Machine Learning Approaches, Istanbul: Sabancı University, 2020.

Olga Gorodetskaya, Yana Gobareva, Mikhail Koroteev, "A Machine Learning Pipeline for Forecasting Time Series in the Banking Sector," *Economies*, vol. 205, 2021.

Vangala Sarveswararao, Vadlamani Ravi, "ATM Cash demand forecasting in an Indian Bank with chaos and deep learning," *Expert Systems with Applications*, vol. 211, 2022.

Venkatesh Kamini, Vadlamani Ravi, D. Nagesh Kumar, "Chaotic Time Series Analysis with Neural Networks to forecast Cash Demand in ATMs," in *IEEE International Conference on Computational*, Coimbatore, India, 2014.

Akber Rajwani, Tahir Syed, Behraj Khan, Sadaf Behlim, "Regression analysis for ATM cash flow prediction," in *International Conference on Frontiers of Information Technology*, Karachi, Pakistan, 2017.

Soodabeh Poorzaker Arabani1 · Hosein Ebrahimpour Komleh1, "The Improvement of Forecasting ATMs Cash Demand of Iran Banking Network Using CNN," *Arabian Journal for Science and Engineering*, vol. 44, p. 3733–3743, 2018.

Puspa Pratiwi, Chandra Utomo, Muhammad Amartya, "Machine Learning Model using Times Series Analytics for Prediction of ATM Transactions," in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Jakarta, Indonesia, 2022.

A Riabykh, I Suleimanov, D Surzhko, M Konovalikhin, V Ryazanov, "ATM Cash Flow Prediction Using Local and Global Model Approaches in Cash Management Optimization," *Pattern Recognition and Image Analysis*, vol. 32, no. 4, pp. 803-820, 2022.

Ravi, Vangala Sarveswararao and Vadlamani, "ATM Cash demand forecasting in an Indian Bank with chaos and deep learning," *Expert Systems with Applications*, vol. 211, 2022.

Davy Cielen, Arno D. B. Meysman, and Mohamed Ali, "Exploratory data analysis," in *Introducing Data Science*, Manning Publications Co, 2016, pp. 43-47.

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, "Measures of Association Between Two Variables," in *Statistics for Business and Economics*, Cengage Learning, 2015, pp. 138-148.

Cansu Yalim and Holly A. H. Handley, "The Effectiveness of Visualization Techniques for Supporting," in *Modeling, Simulation and Visualization Student Capstone Conference*, California, USA, 2023.

"Palestinian Central Bureau of Statistics," Palestinian Central Bureau of Statistics, 14 5 2020. [Online]. Available: <https://www.pcbs.gov.ps/postar.aspx?lang=ar&ItemID=3713>.

Douglas C Montgomery, Cheryl L Jennings, Murat Kulahci, *Introduction to time series analysis and forecasting*, John Wiley & Sons, 2015.

V. Bandari, "Enterprise Data Security Measures: A Comparative Review of Effectiveness and Risks Across Different Industries and Organization Types," *International Journal of Business Intelligence and Big Data Analytics*, vol. 6.1, pp. 1-11, 2023.

P. M. Authority, "Palestine Monetary Authority," Palestine Monetary Authority, 15 November 2015. [Online]. Available: [https://www.pma.ps/ar/key\(arEG\)](https://www.pma.ps/ar/key(arEG)). [Accessed 15 November 2015].

Robert Alan Yaffee and Monnie McGee, *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*, Elsevier, 2000.

L. Leemis, *Statistical Modeling: Regression, Survival Analysis, and Time Series Analysis*, William & Mary, 2023.

Innocent MPAWENIMANA, Alain PEGATOQUET, Win Thandar SOE, Cécile BELLEUDY, "Appliances Identification for Different Electrical Signatures using Moving Average as Data Preperation," in *2018 Ninth International Green and Sustainable Computing Conference*, Pittsburgh, PA, USA, 2018.

C. C. Aggarwal, *Data Mining: The Textbook*, New York: Springer, 2015.

T. J. Sefara, "The Effects of Normalisation Methods on Speech Emotion Recognition," in *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, Vanderbijlpark, South Africa, 2019.

Samit Bhanja and Abhishek Das, "Deep Neural Network for Multivariate Time-Series Forecasting," *Advances in Intelligent Systems and Computing*, vol. 1255, pp. 267-277, 2020/11/24.

Hugo Siqueira, Yara de Souza Tadano, Thiago Antonini Alves, Romis Attux, and Christiano Lyra Filho, "Deseasonalization Methods in Seasonal Streamflow Series Forecasting," *New Trends in Computational Vision and Bio-inspired Computing*, p. 1551–1560, 2020.

M. Peireixo, *Time Series Forecasting in Python*, Simon and Schuster, 2022.

Fatoumata Dama and Christine Sinoquet, "Time Series Analysis and Modeling to Forecast: a Survey," *arXiv preprint arXiv:2104.00164*, 2021.

"Methods for choice of model in descriptive time series : A review with example," *International Journal of Advanced Statistics and Probability*, vol. 6, no. 1, pp. 10-17, 2018.

Renfei He, Limao Zhang, Alvin Wei Ze Chew, "Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning," *Knowledge-Based Systems*, vol. 251, no. 109125, 2022.

Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388-427, 2021.

Rebecca Salles, Kele Belloze, Fabio Porto, Pedro H. Gonzalez, Eduardo Ogasawara, "Nonstationary time series transformation methods: An experimental review," *Knowledge-Based Systems*, vol. 162, no. 0950-7051, pp. 274-291, 2019.

zhaohua Wu, Norden E. Huang, Steven R. Long, and Chung-Kang Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences*, vol. 104.38, pp. 14889-14894, 2007.

P. Dangeti, *Statistics for Machine Learning*, Birmingham: Packt Publishing Ltd., 2017.

G´erard Biau and Erwan Scornet, "A Random Forest Guided Tour," *TEST*, vol. 25, no. 2, pp. 197-227, 2016.

"IBM," IBM, 2022. [Online]. Available: <https://www.ibm.com/topics/random-forest>. [Accessed 2022].

Philipp Probst, Marvin Wright and Anne-Laure Boulesteix, "Hyperparameters and Tuning Strategies for Random Forest," *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, vol. 9, no. 3, p. e1301, 2019.

Abolfazl Nadi and Hadi Moradi, "Increasing the views and reducing the depth in random forest," *Expert Systems with Applications*, vol. 139, no. 112801., 2019.

Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," *J. Assoc.*, vol. 42, no. 8, pp. 785-794, 2016.

Jessica Pesantez-Narvaez , Montserrat Guillen, and Manuela Alcañiz, "Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression," *Risks*, vol. 7, no. 2, p. 16, 2019.

Y. Wang, Z. Pan, J. Zheng, L. Qian & M. Li, "A hybrid ensemble method for pulsar candidate classification," *Astrophysics and Space Science*, vol. 364, no. 8, p. 139, 2019.

Louise Bloch<sup>1,2(B)</sup> and Christoph M. Friedrich<sup>1,</sup>, "Using Bayesian Optimization to Effectively Tune Random Forest and XGBoost Hyperparameters for Early Alzheimer's Disease Diagnosis," in *International Conference, MobiHealth 2020*, online, 2020.

Johanna Sommer, Dimitrios Sarigiannis, Thomas Parnell, "Learning to Tune XGBoost with XGBoost," in *arXiv preprint arXiv*, online, 2019.

Utkarsh Singh and Shelly Vadhera, "Random Forest and Xgboost Technique for Short-Term Load Forecasting,," in *st International Conference on Sustainable Technology for Power and Energy Systems (STPES)*, Srinagar, India, 2022.

SayanPutatunda and Kiran Rama, "A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost," in *Proceedings of the 2018 international conference on signal processing and machine learning*, New York, USA, 2018.

J. L. Elman, *Finding Structure in Time*, Cognitive Science, 1990.

F. D. GONZÁLEZ, "Federated Learning for Time Series Forecasting Using LSTM Networks: Exploiting Similarities Through Clustering," KTH Royal Institute Of Technology, Stockholm, 2019.

Sepp Hochreiter, Jurgen Schmidhuber, "Long Short term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

Hans Weytjens · Enrico Lohmann · Martin Kleinstauber, "Cash Flow Prediction: MLP and LSTM compared to ARIMA and Prophet," *Electronic Commerce Research*, vol. 21, no. 2, pp. 371-391, 2021.

Nivethitha Somu, Gauthama Raman M R, Krithi Ramamritham,, "A hybrid model for building energy consumption forecasting using long short term memory networks,," *Applied Energy*,, vol. 261, no. 114131, 2020.

Bernard Delyon, "Stochastic approximation with decreasing gain : Convergence and asymptotic theory," *Unpublished lecture notes*, vol. 26, pp. 1-36, 2000.

Nikhil Buduma , Nicholas Locascio, *Fundamentals of Deep Learning : Designing Next-Generation*, California, USA: O'Reilly, 2017.

"Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting," *Water*, vol. 12, no. 5, p. 1500, 2020.

Marleen C. Tjepkema-Cloostermans, Rafael C.V. de Carvalho, Michel J.A.M. van Putten,, "Deep learning for detection of focal epileptiform discharges from scalp EEG recordings,," *Clinical Neurophysiology*,, vol. 129, no. 10, pp. 2191-2196, 2018.

J. Brownlee, "What is the Difference Between a Batch and an Epoch in a Neural Network," 15 8 2022. [Online]. Available: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>.

"Enhancing the Performance of an Image Steganalysis Approach Using Variable Batch Size-Based CNN on GPUs," in *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Metz, France, 2019, 2019.

"Small Batch or Large Batch? Gaussian Walk with Rebound Can Teach," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, NS, Canada, 2017.

Muhammad Uzair, Noreen Jamil, "Effects of Hidden Layers on the Efficiency of Neural networks," in *2020 IEEE 23rd international multitopic conference (INMIC)*., Bahawalpur, Pakistan, 2020.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv*, 2012.

Pierre Baldi and Peter Sadowski, "Understanding Dropout," *Advances in neural information processing systems*, vol. 26, 2013.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

Vagelis Plevris, German Solorzano, Nikolaos Bakas, Mohamed Ben Seghier, "investigation of performance metrics in regression analysis and machine learning-based prediction models," in *The 8th European Congress on Computational Methods in Applied Sciences and Engineering*, Oslo, Norway, 2022.

Arnaud De Myttenaere, , Boris Golden, Bénédicte Le Grand, Fabrice Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, no. 0925-2312, p. 38–48, June 2017.

R. Yang, "Omphalos, Uber's Parallel and Language-Extensible Time Series Backtesting Tool," 24 1 2018. [Online]. Available: <https://www.uber.com/en-JO/blog/omphalos/>.

Petro Liashchynskyi and Pavlo Liashchynskyi, Grid search, random search, genetic algorithm: a big comparison for NAS, Ternopil, Ukraine: arXiv preprint arXiv:1912.06059, 2019, 2019.

G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

Ratnadip Adhikari and R. K. Agrawal , "An introductory study on time series modeling and forecasting," *arXiv:1302.6613*, 2013.

A., Bilir C. & Doseyen, "Optimization Of ATM And Branch Cash Operations Using An Integrated Cash Requirement Forecasting And Cash Optimization Model," *BMIJ*, vol. 6, no. 1, pp. 237-255, 2018.

Olga Gorodetskaya, Yana Gobareva , Mikhail Koroteev, "A Machine Learning Pipeline for Forecasting Time Series in the Banking Sector," *Economies*, vol. 205, 2021.

Venkatesh Kamini ,Vadlamani Ravi , D. Nagesh Kumar, "Chaotic Time Series Analysis with Neural Networks to forecast Cash Demand in ATMs," in *IEEE International Conference on Computational* , Coimbatore, India, 2014.

Genevois, Michele Cedolin and Müjde Erol, "A SYSTEMATIC LITERATURE ANALYSIS ON CASH FORECASTING problems in ATMS," *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, vol. 20, no. 40, pp. 287-309, 2021.

T.V.Balakrishnan ,R Kalaiarasi, "A Literature Survey On Automated Teller Machine Cash Demand Analysis And Prediction In Financial Sector," *Ilkogretim Online - Elementary Education Online*, vol. 20, pp. p. 2146-2161, 2021.

Ranade, Pushkar V. Dandekar and Ketki M., "ATM Cash Flow Management," *International Journal of Innovation, Management and Technology*, vol. 6, no. 5, pp. 343-347, 2015.

Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Mathematics for machine learning., Cambridge: Cambridge University Press, 2019.

"Fast Fourier Transformation FFT - Basics," NTi Audio, [Online]. Available: <https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft>.

B. o. Palestine, "Annual report and financial statements," Bank of Palestine , Ramallah - Palestine, 2017.

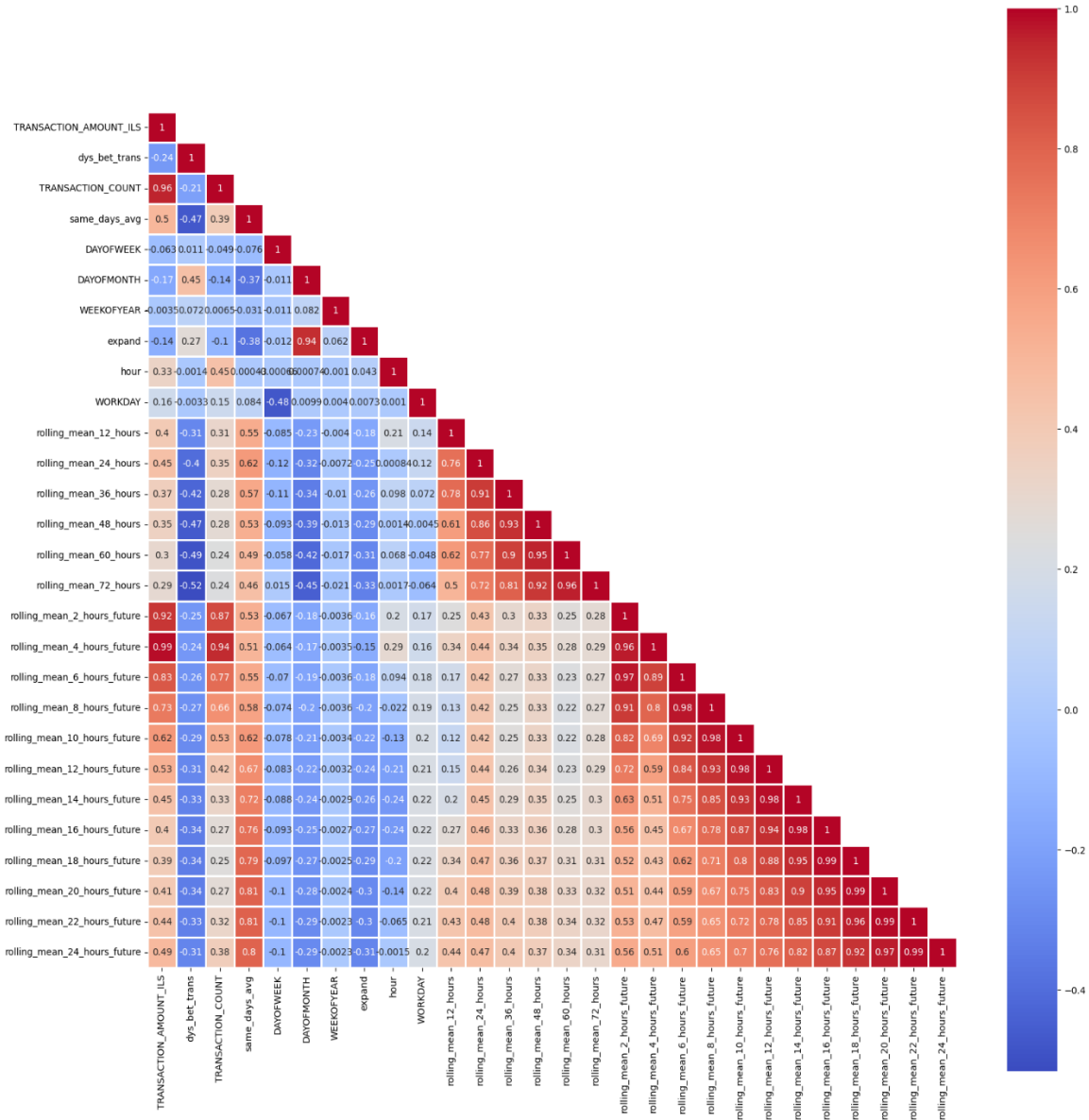
B. o. Palestine, "Annual report and financial statements," Bank of Palestine , Raamallah - Palestine, 2018.

Donato\_TH, "The Journey into Time Series: Dancing with Trends and Seasons," 7 2023. [Online]. Available: <https://medium.com/donato-story/the-journey-into-time-series-dancing-with-trends-and-seasons-d8b635f8fa3c>.

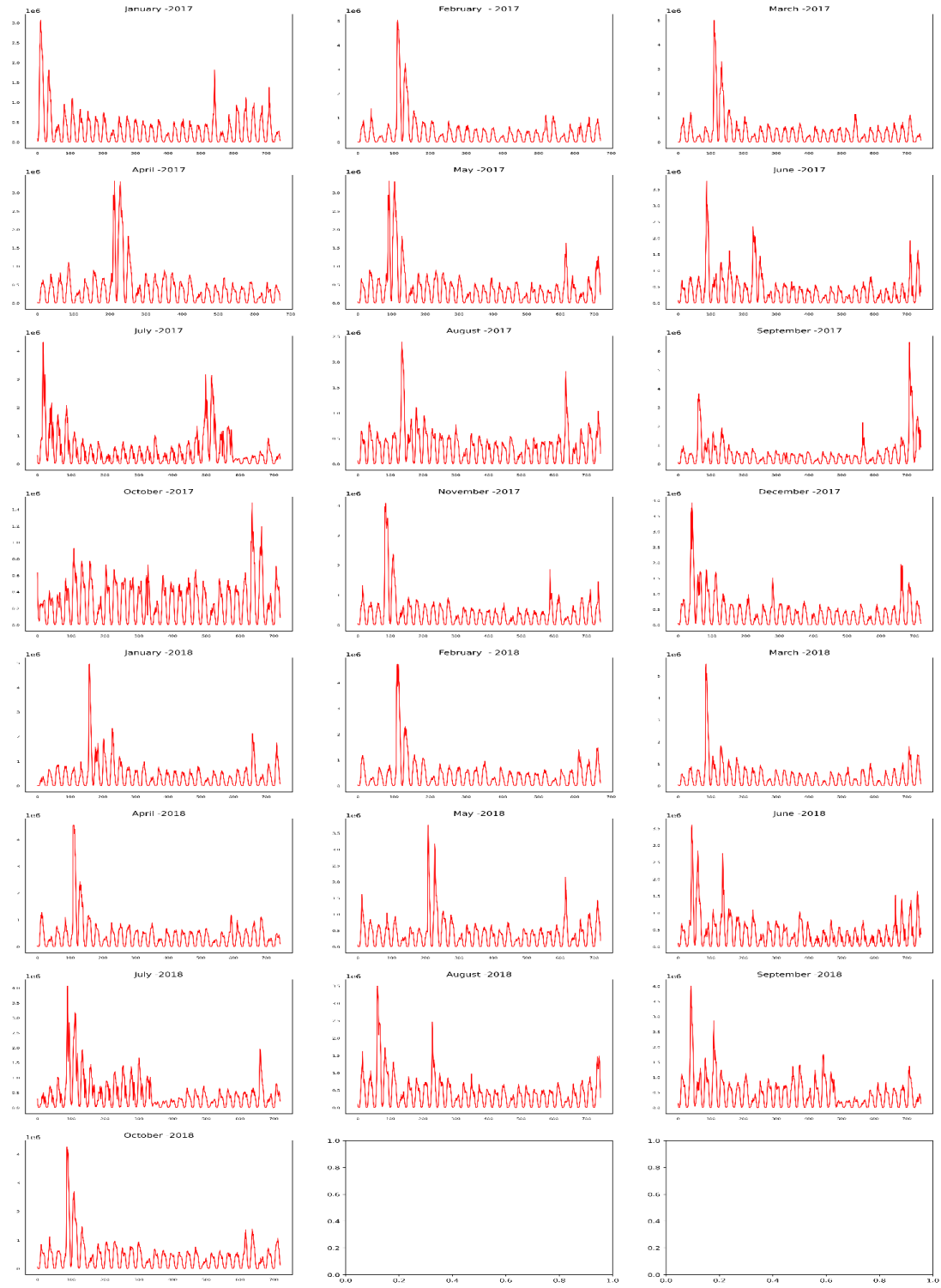
Dena Markudova Politecnico, Elena Baralis Politecnico, Luca Cagliari Politecnico, Marco Mellia, Luca Vassio Politecnico, Elvio Amparore, Riccardo Loti Tierra, Lucia Salvatori Tierra , "Heterogeneous industrial vehicle usage predictions: A real case.," in *Proceedings of the Workshops of the {EDBT/ICDT} 2019 Joint Conference*, Lisbon, Portugal., 2019.

## Appendices

### Appendix (1): Correlation heat map



### Appendix (2): Hourly seasonality



## Appendix (3): Raw data fields description

TRANSACTION_LOCAL_DATE	Date and Time of Withdrawal
WORDING	Transaction Type
ABRV_WORDING	Transaction Status (if successful or reason for lack of success)
ISSUING_BANK	Card Issuer/Network (89 – Bank of Palestine, 2020 – National Switch, 73 – ARAB ISLAMIC BANK 76 – PALESTINE INVEST BANK 81 - PALESTINE ISLAMIC BANK
TRANSACTION_AMOUNT	Transaction Amount (in transaction currency)
CURRENCY_CODE_ALPHA	Transaction Currency
TERMINAL_ID	Terminal ID
CARD_ACC_NAME_ADDRESS	Terminal (ATM) Full Description
ACRONYM	Terminal (ATM) Sort Description
LOCATION	Terminal Location
CITY	Terminal City
REGION	Terminal Region
REVERSAL_FLAG	If transaction was reversed (not applicable)
MODEL	ATM Model
CASH_DEP_FLAG	Indicates if ATM allows cash deposit (1 = Yes, 0 = No)

## Appendix (4): date- based features that were tested and excluded

#	Feature	Description
1.	day of week	Extracting the day of the week
2.	Month	Extracting the month
3.	Year	Extracting the year
4.	Quarter	Extracting the quarter
5.	Day of the Year	Extracting the day of the year
6.	Week of month	Extracting the week of the month
7.	Is weekend	Extracting the weekend
8.	Is the end of the month	Extracting is end of the month
9.	Is Christmas	Extracting is Christmas
10.	Is New Year	Extracting is New Year
11.	Is Easter	Extracting is Easter
12.	Is Black Friday	Extracting is Black Friday
13.	Is Valentine's Day	Extracting is Valentine's Day
14.	Is Mother's Day	Extracting is Mother's Day
15.	Is Father's Day	Extracting is Father's Day
16.	Is Halloween	Extracting is Halloween
17.	Is thanksgiving	Extracting is Thanksgiving
18.	Is Eid al-Fitr	Extracting is Eid al-Fitr
19.	Is Eid al-Adha	Extracting is Eid al-Adha
20.	Is religious holiday	Extract is Religious Holiday

## الملخص

تعد ماكينة الصراف الآلي الخدمة المصرفية الأكثر انتشارًا، كما أنها الأكثر استخدامًا بين الخدمات المصرفية الإلكترونية، وتمثل أداة مهمة للبنك للانتشار والترويج وجذب العملاء، وهناك عدة عوامل يجب مراعاتها في إدارة النقد لأجهزة الصراف الآلي، والتنبؤ بمبلغ السحب هو الأولوية الأولى لأي بنك، وتجديده بالمبلغ الأمثل مما يوفر النفقات ويعظم الربح.

بناءً على بيانات حركات أجهزة الصراف الآلي الأولية من بنك فلسطين للأعوام 2017-2018، تمت معالجة هذه البيانات مسبقًا وهندسة الميزات الجديدة باستخدام أساليب مختلفة، وتم استخدام 3 نماذج (Random Forest، XGBoost، وLSTM)، وحصلت LSTM على أفضل نتيجة في مهمتين، أولاً التنبؤ بعمليات السحب في الساعة التالية لـ 172 جهاز صراف آلي مجمعة في سلسلة زمنية لكل ساعة وتعتبر بمثابة ماكينة صراف آلي واحدة، مع  $R^2$  بنسبة 91% وMAPE قدره كانت المهمة الأخرى هي التنبؤ بمبلغ متوسط الحركات للساعات الزوجية التالية (2-20)، مع أفضل أداء مع  $R^2$  96% و99% MAPE من (0.08-0.28).

توصل هذا المشروع إلى أن مبالغ السحب النقدي تتأثر بشكل كبير بأيام تحويل رواتب السلطة الوطنية الفلسطينية وساعات الحركات. وأثبت أن استخدام الميزات المصممة وخوارزميات التعلم الآلي المناسبة يسمح بالتنبؤات الدقيقة لعمليات السحب من هذه الأجهزة.