MEF UNIVERSITY

PREDICTING TRANSACTION NUMBERS IN ATM

Capstone Project

Ahsen Ceren Karasu

İSTANBUL, 2018

MEF UNIVERSITY

PREDICTING TRANSACTION NUMBERS IN ATM

Capstone Project

Ahsen Ceren Karasu

Advisor: Asst. Prof. Dr. Ahmet Serdar Tan

İSTANBUL, 2018

MEF UNIVERSITY

Name of the project: Predicting Transaction Numbers in ATM
Name/Last Name of the Student: Ahsen Ceren Karasu
Date of Thesis Defense: 08/09/2018

I hereby state that the graduation project prepared by Ahsen Ceren Karasu has been completed under my supervision. I accept this work as a "Graduation Project".

08/09/2018

I hereby state that I have examined this graduation project by Ahsen Ceren Karasu which is accepted by her supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

08/09/2018

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of __________ and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1.	………………………..
2. …………………………..	………………………..

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Name Date Signature

EXECUTIVE SUMMARY

PREDICTION TRANSACTION NUMBERS IN ATM

Ahsen Ceren Karasu

Advisor: Asst. Prof. Dr. Ahmet Serdar Tan

JULY, 2018, 15 pages

ATMs continue to be one of the most important channels for banks to touch their customers. They play an active role in life in terms of cash access and banking experience. The ability of a bank to predict the number of transactions that will occur from ATMs is crucial for the proper control of the budgetary source. When cash is loaded into ATMs, the average transaction made from that ATM is taken into consideration and alarm mechanisms can be activated when a decreasing trend is observed on transaction basis.Before a new ATM is set up, the banks investigate how often customers in that area use other bank ATMs and calculate the commission costs incurred from those uses. As a result, the number of transactions made from ATMs is one of the most monitored KPIs of a bank and has important place in the cash management of the bank.The aim of this study is to estimate the number of future transactions with Auto Regressive Moving Average (ARIMA) method based on the number of transactions that occurred from ATMs.

Key Words: Transaction, Regression, ARIMA, Prediction, Time series, Exploratory Data Analysis.

ÖZET

ATM İŞLEM ADEDİ TAHMİNLEME

Ahsen Ceren Karasu

Tez Danışmanı: Doç. Dr. Ahmet Serdar Tan

TEMMUZ, 2018, 15 sayfa

ATM’ler bankalar için müşterilerine dokunabildikleri en önemli kanallardan biri olma görevini sürdürmektedir. Nakite ulaşma ve bankacılık deneyimini yaşamada etkin rol oynarlar.Bir bankanın ATM'lerden yapılan işlem sayısını tahmin etme yeteneği, bütçe kaynağının uygun şekilde kontrolü için çok önemlidir. ATM'lere nakit ikmal edildiğinde söz konusu ATM'den yapılan ortalama işlem dikkate alınır ve işlem bazında azalan bir eğilim gözlendiğinde alarm mekanizmaları etkinleştirilebilir. Yeni bir ATM kurulmadan önce, bankalar o lokasyondaki müşterilerinin diğer banka ATM'lerini ne sıklıkla kullandığını ve bu kullanımlardan kaynaklanan komisyon masraflarını hesaplar. Sonuç olarak, ATM'lerden yapılan işlem sayısı, bir bankanın en çok izlenen KPI'larından biridir ve bankanın nakit yönetiminde önemli bir yere sahiptir. Bu çalışmanın amacı, ATM'lerden kaynaklanan işlem sayısına dayanarak gelecekteki işlem sayısını oto regresif hareketli ortalama (ARIMA) ile tahmin etmektir.

Anahtar Kelimeler: Regresyon, Tahminleme, Zaman serileri, Data Analizi, ARIMA

TABLE OF CONTENTS

1. INTRODUCTION

Nowadays, customer satisfaction oriented banks make many innovations to provide a banking experience that can be reached to customers 24/7. In addition to branch banking, which provides customers with limited hours, they invest a lot in attracting customers to ATM, internet and mobile banking channels. These channels, which are called direct banking, branchless banking or self-service banking, are also important in terms of enabling customers to complete their transactions quickly, apart from being channels that can be accessed at any time.

ATMs are the channels through which customers can quickly meet their cash needs. Today, many banks serve their customers in ATMs as well as serve different bank customers. The Banking Regulation and Supervision Agency is issuing an arrangement to allow customers to receive joint services from many ATMs, whether or not the bank is a self-employed bank or not.

Turkey has 50.286 ATMs in March 2018. [1] The vast majority of these ATMs are called common ATMs and serve to all bank customers. Certain transactions within the services provided not only generate income for the service provider bank but also provide the customer's own bank's income.As a result, it is important for all banks to be able to monitor the ATM usage frequency whether or not the ATM is used by its own customers.

To set up a new ATM, the banks simulate the customer's ATM usage frequency and take action according to the result.The most important thing that needs to be done before a new ATM is established is to estimate the total number of transactions to be made at that location.

With this study, the number of transactions that will be realized in the future will be tried to be estimated by using 3-year ATM transaction numbers of a bank. Due to data includes time series, ARIMA function will be used to prediction and correction of data.

1.1. Literature Review

Numerous authors had done work to alleviate ATM cash problem, prediction of number of transaction, stock, price and service analysis but not a specific experiment has done to predict ATM transaction numbers. With this paper several papers have been examined and research techniques, methodologies have been adapted.

Erol Genevois et al. [2] focused the problem of ATM location and cash management in ATM in 2015. In their research they discussed two problems which bank are facing one is finding suitable location for ATM and other is cash management strategy. Estimating the number of transactions and the cash management of ATMs are two parallel issues and related to each other. Transaction numbers directly refers to cash demands of ATM so methodology of calculate the cash demand also serve as a key role in prediction of future.

Boufounou et al. [3] presents a model to help management in setting branch, evaluating branch performance and planning new locations. Paper classifies the relevant factors of decision making with respect to branch location into four categories. These categories are; location features, trade area characteristics, competitive situation features and internal branch characteristics. Boufounou's study may be utilized in locating ATM problem. Same problems arise when calculating the number of ATM transactions. The seasonal effect, the choice of faulty location, and even the customer segment directly affect the number of transactions from ATMs.

Brentnall et al. [4] construct a method for predicting the daily amounts of transactions from ATMs. The data which they have used consisted of information of two years of 190 ATMs of one bank. In paper time series regression method have been used and time series are explained: "It takes the decision by working on time that is (minutes, years, days, hours) and find out hidden insight in data. It works well when the data is correlated. It is basically a set of data points gathered at constant time interval. There are two things which makes time series special and different from linear regression. First it is time dependent, unlink linear regression which says that observations are independent, second, it identify the sessional trends in the data for example transactions occur most before gazette holiday etc. In order to run time series model we have assumed that time series (TS) is stationary and its statistical properties such as variance, mean remain same over period. It is important to because there is very high probability that series will follow same pattern in future also."

Ayodele A. Adebiyi et al. [5] make a prediction with using ARIMA on stock price.In paper extensive process of building ARIMA models for short-term stock price prediction was presented. The results obtained from real-life data demonstrated the potential strength of ARIMA models to provide investors short-term prediction that could aid investment decision making process. this work is inspired by this article in the context of the work and the results are shared in a similar way.

2. DATASET

In the first part of this section, data structure and properties will be explained. In the second part, exploratory data analysis will be done.

2.1. About the Data

Predicting ATM transaction numbers is challenging due to unpredictability of customer behavior, seasonal effect and duration of intervention with technical insufficiencies. There are several ways to do this but it will be explained in methodology part. Before this, we will look closer to data and their features.

There are 50.286 ATMs in Turkey according to BKM's 2018 Report "Number of POS, ATM and Cards". In this project one of the bank ATMs will be examined due to number of transactions from January 2014 to December 2016.

Data includes 82 uniq ATM's transaction date, ATM ID, city code, city name and transaction numbers by monthly. Because this paper explains predicting transaction numbers between 2014 and 2016 ATMs which was opened in 2015, early months of 2014 and 2016 was removed so uniq ATM number decreased to 36.

DATA TYPES	
Date	object
ID	int64
Citycode	int64
City	object
Numberoftransaction	int64
Dtype	object

DATAFRAME SHAPE
(2195, 5)

2.2. Exploratory Data Analysis

Data contains number of transactions made from the Model Bank's 82 ATMs on a monthly basis.A sample table of data is shown below.

Date	ID	Citycode	City	# of Transaction
2014-01	11	34	İstanbul	332
2014-01	12	34	İstanbul	737
2014-01	1011	34	İstanbul	872
2014-01	1012	34	İstanbul	951
2014-01	1021	34	İstanbul	2121

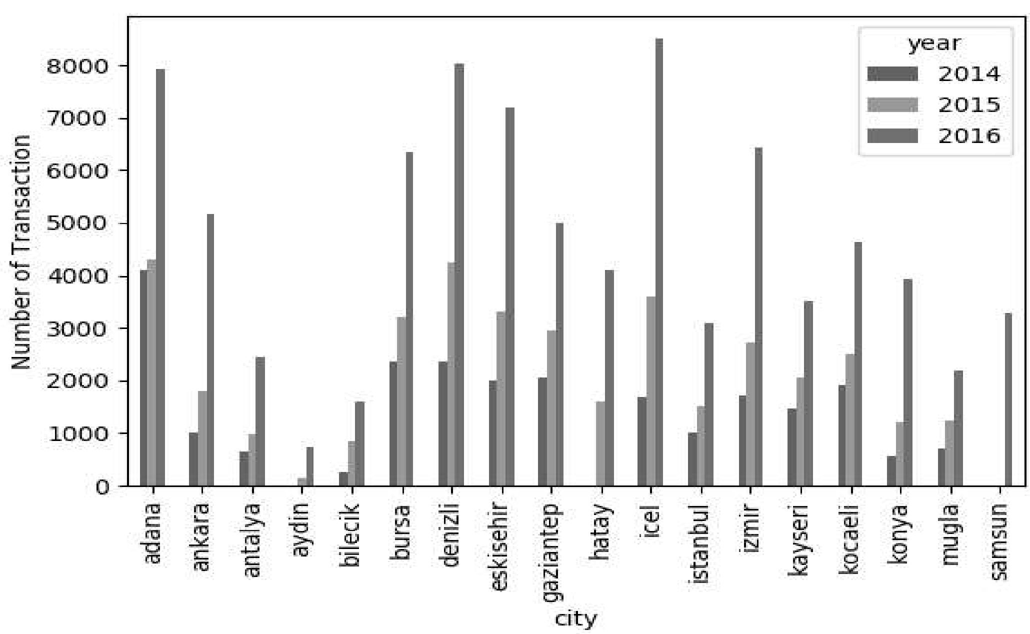
82 ATMs of model bank in several cities of Turkey. Here is a table of ATMs by locations. Most of the ATMs of bank in İstanbul.

City	# of ATM
İstanbul	40
Ankara	7
İzmir	6
Kocaeli	4
Adana	3
Antalya	3
Bursa	3
Kayseri	3
Bilecik	2
Gaziantep	2
Konya	2
Aydin	1
Denizli	1
Eskisehir	1
Hatay	1
İcel	1
Mugla	1
Samsun	1
Grand Total	**82**

Number of transactions was change in years. Here is a table to show transaction numbers in years.

Year	Numberof Transaction
2014	872.014
2015	1.495.128
2016	3.271.492

The chart for the annual transaction numbers by cities is as follows. As can be seen below, the number of trades has been on an increasing trend over the years. Mersin has one of the biggest customer portfolio of banks and number of product by customer is higher than any other cities so number of transaction is higher in this city. Some ATM's was opened in 2015 and 2016. In analysis part this ATM's were removed. Some ATMs like Aydın and Bilecik were opened for special contracts with companies. Their number of transactions lower than others because customer numbers to serve is lower than others. (Setting up in a factory etc.)



3. PROJECT DEFINITION

There are two channels to meet the cash needs of the customers of the banks: Branches and ATMs. With the closing of the branches on holidays, customers can only access cash via ATMs. The profit of a bank obtained from an ATM is calculated by means of three ways; income from the customers, income from the other banks and the commission paid to the other banks that mediate their customers.

This key role of ATMs' cash needs is directly related to the number of transactions. By multiplying the number of transactions by the commission fee from the related transaction, the income from that transaction is calculated.

The number of transactions is important not only for calculating income but also for predicting ATM's cash requirement and cash capacity.

Cash-on-delivery vehicles are cash-dispatched to ATMs with the triggering of banks, thus directing the relevant service to the correct ATM is a manual process when there is a diminished trend in cash-flow services and cash volume.

With all these reasons, an automated process is needed to predict the ATM revenue, anticipate the need for cash, and retrieve an optimized road map for ATM-cash-carrying vehicles. Predicting the transaction to be made from ATMs with a high degree of accuracy will play a facilitating role for the next steps.

In this project, it is aimed to create a model that can adapt to any data's even though a limited data's have been applied. Within the scope of the project, the number of transactions in 2016 will be estimated using the number of transactions made from 36 ATMs in 2014 and 2015 and the results will be compared with the actual data for 2016 to measure the accuracy of the model.

4. METHODOLOGY

This paper suggests to predict future transaction numbers due to historical values. Data understanding, cleaning, visualization and exploratory data analysis parts are done with Python and some tables are made in Excel.

In literature lots of the methods are used for forecasting like Bayesian Ridge Regression, Linear Regression and Support Vector Machine. Due to data contains time series and unstable, ARIMA model will be used to prediction.

To apply the ARIMA model, a lot of resources have been scanned and most have directly influenced at the end of the model. Machine Learning Mastery web site [6] explains time series clearly and helps to understand what time series is, how can we handle with unstable data and explanation of ARIMA models for beginners. Also kernels in Kaggle like Bitcoin Prediction [7] there are useful resources to understand how ARIMA is implemented and its data structure even though it is not related to the selected topic. Kernels also help to understand the necessary packages for python and why they are needed. In DigitalOcean web site "A Guide to Time Series Forecasting with ARIMA in Python" [8] article has been used to clarify results.

One of the things that needs to be done before understanding the structure of the model and how to write it is to understand the logic of the model. Ina Khandelwal et al [9] explains ARIMA models as most popular and effective statistical model for prediction. In this paper, prediction of time series are explained with calculations and fundamental principles are explained with mathematical formulas.

Graphs are an effective tool to summarize the whole work as a result of the analysis made. Inspired by the Analytics Vidhya [10] website to visualize the results of the analysis.

In this work, the number of transactions is estimated for 36 individual ATMs using Python's 'statsmodels' package to apply ARIMA.

5. RESULTS

In this study, the number of transactions in 2016 was estimated by ARIMA using the trading numbers of 36 ATMs between 2014 and 2015, and the results were compared with the real values for 2016.

R square results of 36 ATMs are as follows. As shown in the table, the results showed ATM-based variability. Charts for monthly forecasts are listed below.

Factors such as seasonal effects, serviced customers' segment and the bank's growth or downsizing strategy have reduced the model's ability to predict. On the other hand, it is seen that the r2 values of ATMs that work efficiently and serve with maximum capacity are high.

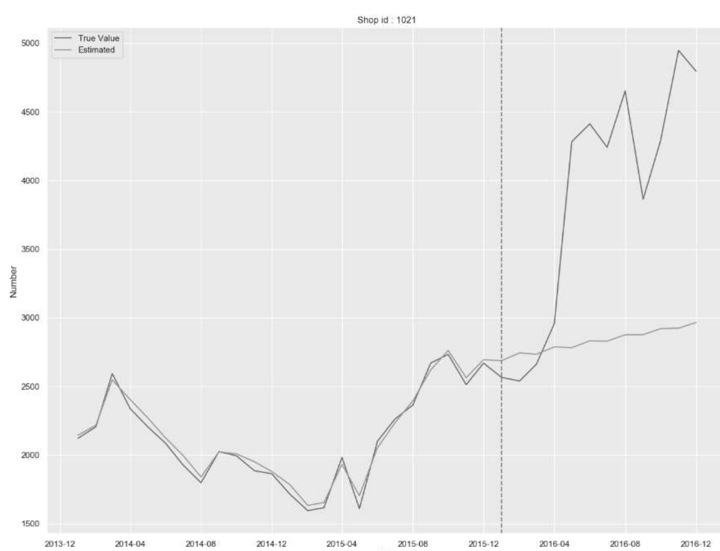
ID	City	r2_score
1091	İstanbul	0,66
2221	Kocaeli	0,66
1161	Antalya	0,52
1112	İstanbul	0,50
2161	Kayseri	0,50
1151	Kayseri	0,48
1131	Denizli	0,46
2033	İstanbul	0,45
21	İstanbul	0,42
12	İstanbul	0,41
1012	İstanbul	0,40
1191	Ankara	0,36
1072	Gaziantep	0,33
1231	İstanbul	0,31
1181	Mugla	0,27
1211	İstanbul	0,27
2021	İstanbul	0,23
1202	Konya	0,20
1032	Ankara	0,19
2201	İzmir	0,13
2032	İstanbul	0,12
1071	Gaziantep	0,03
1061	Adana	0,01
2051	Ankara	0,01
1051	Bursa	0,00
1052	Bursa	0,00
2191	İstanbul	0,00

2121	İstanbul	-0,10
1081	Kocaeli	-0,20
2211	İstanbul	-0,20
1041	İzmir	-0,30
1062	Adana	-0,30
2011	İstanbul	-0,50
2162	Kayseri	-0,50
2041	İzmir	-0,90
1021	İstanbul	-1,21

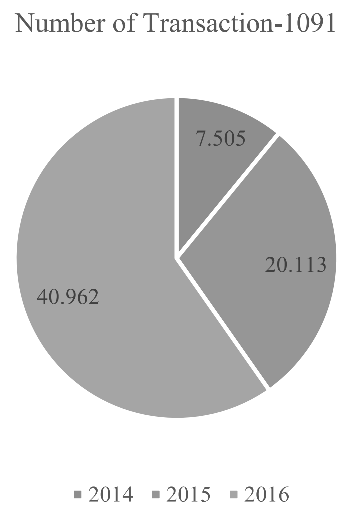
5.1.Comparison

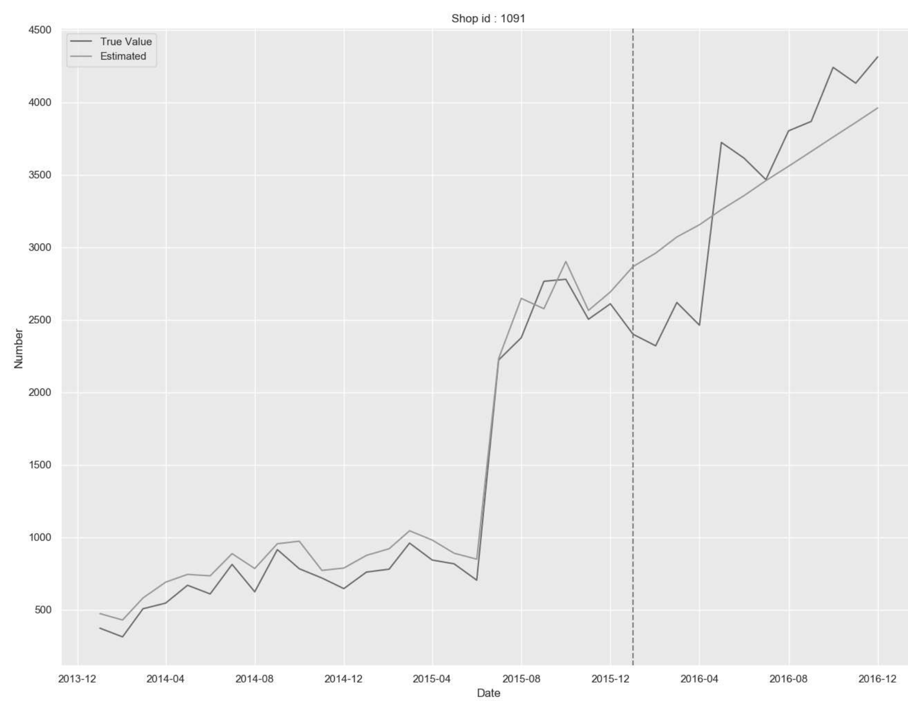
Two ATMs with the lowest (1021) and highest (1091) r2 values were compared.

ATM- 1021 is a branch ATM, and by 2016, March there were 2 ATMs in the same branch, so the number of transactions was shared between 2 ATMs. With the closure of the other ATM, the number of transactions of ATM-1021 increased. Since the model can not predict this situation, the r2 value is low.



ATM- 1091 is a offsite (not on a branch) ATM and this types of ATMs have usually stable number of transactions unlike branch ones. Due to its stability r2 value is high. In 2014 almost all ATMs' number of transaction value is lower than other years. But there is a regular increase in years and model captures it.


Number of Transaction-1091
7.505
20.113
40.962
2014
2015
2016


Shop id : 1091
True Value
Estimated
Number
Date
4500
4000
3500
3000
2500
2000
1500
1000
500
2013-12
2014-04
2014-08
2014-12
2015-04
2015-08
2015-12
2016-04
2016-08
2016-12

6. SOCIAL AND ETHICAL ASPECTS

No personal data were used in this project and a scope to influence ethical values was not included in the project. ATM ID columns have been anonymized and shared.

7. CONTRIBUTION

The transactional model of estimation discussed in this study is of great importance for reducing costs for all sectors, whether banking or not. This study can be used in annual budget plans, including income calculations.

Estimating the number of people to use ATMs can be used to optimize ATM location by combining the heat maps of the regions, as well as can be used to draw real-time root plans for cash-delivery cars to ATM.

This study was done with the aim of providing a basis for solving all these optimization problems that could be done in the future.

APPENDIX

```
import os
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import r2_score

sns.set()
plt.rcParams["figure.figsize"] = (16, 12)
fontsize = 16

filename = "arima_data.csv"
df = pd.read_csv(filename, parse_dates=True, infer_datetime_format=True,
index_col=0)

group_by_id = df.groupby("ID")
shop_map = dict()
for name, group in group_by_id:
  if group.shape[0] >= 36:
    shop_map[name] = group

param_dict = dict()
default_param = (3, 1, 0)
param_dict[1021] = (2, 1, 1)
param_dict[1051] = (1, 1, 0)
param_dict[2041] = (3, 1, 1)
param_dict[2121] = (4, 1, 0)
param_dict[2191] = (3, 1, 1)

plot_folder = "plot"
if not os.path.exists(plot_folder):
    os.makedirs(plot_folder)

r2_score_list = list()
for shop_id in shop_map:
  x = shop_map[shop_id].index
  y = shop_map[shop_id]["num"]
  param = default_param
  if shop_id in param_dict:
    param = param_dict[shop_id]

  mdl = ARIMA(y.values[0:24], order=param)
  mdl_fit = mdl.fit(disp=False)

  y_predict = mdl_fit.predict(start=1, end=36, typ="levels")

  # Plot prediction and true values
  plt.figure()
  plt.plot(x, y, label="True Value")
  plt.plot(x, y_predict, label="Estimated")
  plt.axvline(x=datetime.date(2016, 1, 1), color="gray", linestyle="--")
  plt.legend(fontsize=fontsize)
  plt.tick_params(labelsize=18)
```

```
  plt.xlabel("Date", fontsize=fontsize)
  plt.ylabel("Number", fontsize=fontsize)
  plt.title("Shop id : %s" % shop_id, fontsize=fontsize)
  plt.savefig("%s/%s" % (plot_folder, shop_id))
  plt.close()

  r2 = r2_score(y[24:], y_predict[24:])
  r2_score_list.append(r2)
result_dict = dict()

result_dict["shop_id"] = list(shop_map.keys())
result_dict["r2_score"] = r2_score_list

result_df = pd.DataFrame(result_dict)
result_df = result_df[["shop_id", "r2_score"]]
result_df.to_csv("result.csv", index=False)
```

8. REFERENCES

[1] POS, ATM, Kart Sayıları. (2018, March). Retrived from https://bkm.com.tr/pos-atm-kart-sayilari/

[2] Erol Genevois, M., & Celik, D., & Ulukan, H. Z., (2015) “ATM Location Problem and Cash Management in Automated Teller Machines”. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial EngineeringVol: 9, No: 7.*

[3] Boufounou, P. V. (1995). Evaluating bank branch location and performance: *A case study. European Journal of Operational Research, 87(2).* 389-402.

[4] Brentnall,A.R., & Crowder, M.J.,& Hand, D.J., (2010) "Predictive-sequential forecasting system development for cash machine stocking", *International Journal of Forecasting, vol. 26,* 764-776

[5] Adebiyi, A. A., & Adewumi, A. O., & Ayo, C.K. (2014), Stock Price Prediction Using the ARIMA Model, *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.* 105-111

[6] Brownlee, J. (2018, August 8). Taxonomy of Time Series Forecasting Problems. Retreived from https://machinelearningmastery.com/taxonomy-of-time-series-forecasting-problems/

[7] Артём. (2017, December). Bitcoin Price. Prediction by ARIMA. Retrieved from https://www.kaggle.com/myonin/bitcoin-price-prediction-by-arima

[8] Vincent, T. (2017, March 23). A Guide to Time Series Forecasting with ARIMA in Python. Retreived from https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3

[9] Khandelwal, I., & Adhikari, R., & Verma, G. (2015). Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition. *Procedia Computer Science, Volume: 48*. 173 – 179

[10] Jain, A. (2016, February 6). A comprehensive beginner’s guide to create a Time Series Forecast. Retreived from https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/