ANALYSIS AND OPTIMIZATION METHOD FOR CASH WITHDRAWAL OF ATM IN INDIA

Pannawit Kongmuangpak

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied Mathematics

Suranaree University of Technology

Academic Year 2020

การวิเคราะห์และวิธีการหาค่าที่ดีที่สุดสำหรับการถอนเงินสด ของตู้เอทีเอ็มในประเทศอินเดีย

นายปัณณวิชญ์ กองเมืองปัก

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2563

ANALYSIS AND OPTIMIZATION METHOD FOR CASH WITHDRAWAL OF ATM IN INDIA

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee

(Assoc. Prof. Dr. Eckart Schulz)

Chairperson

J.Tanthanuch

(Asst. Prof. Dr. Jessada Tanthanuch)

Member (Thesis Advisor)

Benjawan Rodjanadid

(Asst.Prof.Dr.Benjawan Rodjanadid)

Member

Surattana Sungnul

(Assoc. Prof. Dr. Surattana Sungnul)

Member

Chatchai Jothit

(Assoc. Prof. Dr. Chatchai Jethityangkoon)

Vice Rector for Academic Affairs

and Quality Assurance

(Assoc. Prof. Dr. Worawat Meevasana)

Dean of Institute of Science

ปัณณวิชญ์ กองเมืองปัก: การวิเคราะห์และวิธีการหาค่าที่ดีที่สุดสำหรับการถอนเงินสดของตู้เอทีเอ็มในประเทศอินเดีย (ANALYSIS AND OPTIMIZATION METHOD FOR CASH WITHDRAWAL OF ATM IN INDIA) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.เจษฎา ตัณฑนุช, 43 หน้า

ตู้เอทีเอ็ม/การถดถอยเชิงเส้นพหุคูณ/ขั้นตอนวิธีการจัดกลุ่ม

งานวิจัยนี้สร้างตัวแบบที่เหมาะที่สุดสำหรับการทำนายความต้องการในการเบิกเงินของลูกค้าจากตู้เอทีเอ็มในประเทศอินเดีย จำนวนข้อมูลประกอบด้วย 910 ชุดข้อมูล ของตัวแปรทั้ง 7 ได้แก่ ปริมาณของเงินที่ถอน วัน (อาทิตย์-เสาร์) วันทำงาน วันที่ เดือน ค่าเฉลี่ยของการปริมาณเงินที่ถอนในสัปดาห์ก่อนหน้า ค่าเฉลี่ยของปริมาณเงินที่ถอนรายวันในสัปดาห์ก่อนหน้า โดยข้อมูลดังกล่าวเป็นข้อมูลการถอนเงินของตู้เอทีเอ็มธนาคารเมาท์โรด ประเทศอินเดีย ข้อมูลที่นำมาใช้เป็นข้อมูลตั้งแต่ปี 2011 ถึงปี 2013 โดยเป็นข้อมูลที่เปิดเผยจากเว็บไซต์ www.kaggle.com/nitsbat/data-of-atm-transaction-of-xyz-bank โปรแกรมที่ใช้ในการดำเนินการวิจัยได้แก่ RStudio RapidMiner และ Python ทั้งนี้ข้อมูลทั้งหมดได้ถูกแบ่งสำหรับใช้เป็นชุดข้อมูลฝึกฝนจำนวนร้อยละ 80 และที่เหลืออีกร้อยละ 20 ถูกใช้เป็นชุดข้อมูลทดสอบ การดำเนินการวิจัยในขั้นแรกใช้การถดถอยเชิงเส้นพหุคูณแบบชั้นเอกวิเคราะห์หาปัจจัยที่มีนัยสำคัญ พบว่ามี 3 ปัจจัยหรือตัวแปรที่มีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ วันทำงาน วันที่ และค่าเฉลี่ยของปริมาณเงินที่ถอนรายวันในสัปดาห์ก่อนหน้า ขั้นถัดมาใช้ 3 ตัวแปรดังกล่าวสร้างตัวแบบเพื่อใช้ทำนาย วิธีการสร้างตัวแบบที่พิจารณาได้แก่ ตัวแบบของการถดถอยเชิงเส้นพหุคูณ ตัวแบบของขั้นตอนวิธีการจัดกลุ่มร่วมกับตัวแบบของการถดถอยเชิงเส้นพหุคูณ และตัวแบบของขั้นตอนวิธีการจัดกลุ่มร่วมกับตัวแบบการถดถอยพหุนาม ทั้งนี้ผลการดำเนินการสร้างตัวแบบโดยวิธีการดังกล่าวพบว่าแต่ละตัวแบบมีค่าความคลาดเคลื่อนรากที่สองเฉลี่ยเท่ากับ 1.0360 1.0137 และ 0.8318 ตามลำดับ และมีค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์เท่ากับ 0.7485 0.7498 และ 0.6305 ตามลำดับ สำหรับงานวิจัยครั้งนี้ตัวแบบของขั้นตอนวิธีการจัดกลุ่มร่วมกับตัวแบบการถดถอยพหุนามเป็นตัวแบบให้ผลในการทำนายดีที่สุด

สาขาวิชาคณิตศาสตร์ ลายมือชื่อนักศึกษา Pannawit

ปีการศึกษา 2563 ลายมือชื่ออาจารย์ที่ปรึกษา

PANNAWIT KONGMUANGPAK : ANALYSIS AND OPTIMIZATION METHOD FOR CASH WITHDRAWAL OF ATM IN INDIA. THESIS ADVISOR : ASST. PROF. JESSADA TANTHANUCH, Ph.D. 43 PP.

Auto Teller Machine/Multiple Linear Regression/Clustering Algorithm

This research constructed an optimal model to forecast the customers' demand of the ATM in India. The data consisted of 910 sets of 7 variables each, namely the amount of withdrawal, weekday, working day, date, month, mean withdrawal of the previous week and average daily withdrawal in the previous week, collected from the ATM of the Mount Road bank branch during 2011 to 2013. This data is open source and provided by www.kaggle.com/nitsbat/data-of-atm-transaction-of-xyz-bank. The software used in the construction of the model was RStudio, RapidMiner and Python programs was used as 80% of the data set was cased as to a training data and the remaining 20% of the data was used as a test set. Classical multiple linear regression was first used for analysis of the significant factors. It was found that working day, holiday, and same weekday withdrawal of the previous week affected to the amount of withdrawal with 5 statistical significance. The same factors were used to construct 3 forecasting models, which were a multiple linear regression model, a clustering algorithm with multiple linear regression model and a clustering algorithm with polynomial regression model. The root mean square error of the models were 1.0360, 1.0137, and 0.8318, respectively and the mean absolute error of the models were 0.7485, 0.7498 and

0.6305, respectively. In this research, the clustering algorithm with polynomial regression model performed the best forecast.

School of Mathematics Student's Signature Pannawit

Academic Year 2020 Advisor's Signature J. Tanthanuch

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my thesis advisor Asst. Prof. Dr. Jessada Tanthanuch for giving me to opportunity to do this research. He was kindly guiding, encouraging, and explaining to me whenever I did not understand some details in this thesis work. In addition, I am grateful for professional support received from Asst. Prof. Dr.Benjawan Rodjanadidand, and all professors in the School of Mathematics, Institute of Science, Suranaree University of Technology.

Moreover, I would like to express big thanks to the Development and Promotion of Science and Technology Talents Project scholarship for financial support along all my degrees.

Finally, I most gratefully acknowledge my family and my friends for their good support and encouragement.

Pannawit Kongmuangpak

CONTENTS

Page

CHAPTER

CONTENTS (Continued)

Page

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER I

INTRODUCTION

An usual automated teller machine (ATM) is an electronic machine for servicing customers banking transactions with individual pin, which are withdrawing, depositing, bill paying and money transfer. In 1967, 6 ATMs were installed by Barclays Bank in England with 4-digit pin code for withdrawing. The two years later, magnetic ships were added to ATM cards, and another two year later these machines could be account cash deposits instead of cheques. Recently, in January 2020 in Thailand, The Thai National Bank increased higher security to prevent fraud, counterfeit card fraud and skimming, for any transactions on the ATMs. The risk in ATM transactions was reduced by using ship cards instead of the magnetic ships. Nowadays, ATMs have become less popular because of mobile banking, as transactions can be made conveniently at any place all the time.

The aim of this research is to propose an optimization method to effective management of ATMs. An operation of the replenishment concerns about the treasury's administration such as cash management, which leads to replenish with an appropriate quantity and time for the customers' demand. The cash replenishment lets some amount of cash held in ATM which makes an opportunity cost for losing some benefit of higher return investment. When replenishment cash exceeds customers' demand, the bank loses interest in any investment opportunity. On the other hand, if customers' demand exceeds the replenishment cash, then it causes more dissatisfaction of customers.

In this thesis, we use regression and clustering classification, to forecast cash demand. The cash demand data set used in this research would be adopted from the ATM in India, obtained from Mount Road branch of ATM in India collected from 2011 to 2017. The data was an open source provided by https://www.kaggle.com/nitsbat/data-of-atm-transaction-of-xyz-bank.

Moreover, in 2018 there is research about cash replenishment with neural network algorithm to predict the amount of cash deposited and they use K-means clustering to classify the group of data (Jadwal P. K., Jain S., Gupta U. and Khanna P., 2018).

1.1 Research Objectives

To study the factors, which affect to the cash withdrawal of customers of ATM in India and construct the models to predict the cash withdrawal of customers of ATM in India.

1.2 Scope and Limitations

The data is obtained by www.kaggle.com/nitsbat/data-of-atm-transantion-of xyz-bank, and we manage the data with technique consisting of K-mean algorithm, multiple linear regression and polynomial regression.

1.3 Research Procedure

The research procedure follow these steps:

1. Studying the significant factors by multiple linear regression.

2. Constructing three models of multiple linear regression, polynomial regression and K-mean algorithm from the significant factors.

3. Measuring the accuracy of our models by using root mean square error and mean absolute error

1.4 Expected Result

We can apply our model with best performance to forecast the cash demand of customer of ATM in the future.

CHAPTER II

LITERATURE REVIEW

2.1 Mean

There are several kinds of mean in mathematics, we will show some kinds of mean in statistics:

The arithmetic mean(AM) is the sum of all of the members in the finite set and the set is divided by the number of members in that set. The arithmetic mean of a number set $\{x_1, x_2, ..., x_n\}$ is denoted by $\bar{x}$.

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + ... + x_n}{n}$$

Remark: if the data sets are collected as observations, the mean value is called population mean and is denoted by μ.

The geometric mean(GM) ,which is useful for sets of positive numbers is the operation of multiplication of the elements in a set and the multiplication keeps rooted by a number of the element in that set:

$$\bar{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = (x_1 x_2 \cdots x_n)^{\frac{1}{n}}$$

The harmonic mean(HM) is a kind of average value which is useful for sets of number which are defined in relation of some physical such as speed(i.e. distance per a unit of time)

$$\bar{x} = n\left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1}$$

2.2 Standardization

The standardization method has the purpose to rescale the data, so that the rescaled data has the mean and standard deviation equal to 0 and 1 respectively. Let A consisting of numbers $x_1, x_2, ..., x_n$ be a finite set. The equation for the standardization method is as follows,

$$z_i = \frac{x_i - \bar{x}}{s},$$

where z_i is the rescaled value of x_i, $i = 1, 2, ..., n$, $\bar{x}$ is a mean of these elements in A, and s is the standard deviation.

2.3 K-Means Clustering

The two popular clustering algorithms are partitional and hierarchical clustering. These algorithms have been used in many applications because of their simplicity and flexibility in adoption to other clustering algorithms. Partitional clustering algorithms have the main focus to discover the grouping of data by optimizing a distinct objective function and improving the quality of the partitions. These algorithms normally require parameters with its stability to choose the root of points that indicate each cluster.

K-means clustering is mostly used as a part of the partition clustering algorithm. Its process starts by choosing K representatives for the number of groups and points as the initial centroids (center points in each cluster). Each data point is compelled to stay in the same cluster with the closet centroid depending on choosing a particular measure. When the clusters are formed, the centroids for each cluster are updated. The algorithm then repeats these steps until the centroids do not change the position or any alternative replaced convergence criterion is met. K-means clustering is an algorithm which is ensured to converge to a local minimum but the minimization of its score function is known to be NP-hard optimization problem. Typically, the convergence condition is flexible and a weakness of condition used. In practical, it follows the rule that iterative process must be continued until all data stay still or 1% of the points change their cluster. A proof of the mathematical convergence of K-means can be found (S.Z. Selim and M.A. Ismail, 1984).

```
Algorithm K-Means Clustering
Randomly select $K$ points as initial centroids.
Repeat
    From $K$ clusters by assigning each point to its closest centroid.
    Recompute the centroid of each cluster
Until the convergence criterion is reach (1% of point change)
```

Given data set $A = \{x_1, x_2, ..., x_n\}$ consists of n points and indicate the cluster after applying K-Means algorithm by $C = \{C_1, C_2, ..., C_k\}$. The sum of square errors (SSE) for the clustering is defined in the Equation (2.3.1), where c_k is centroid of cluster C_k. The aim of SSE is to find a centroid, which makes SSE to have minimum value. The reiteration and renewal of the centroid value of the K-means algorithm aim to minimize the SSE value for the updated centroids.

$$SSE(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - c_k||^2 \qquad (2.3.1)$$

$$c_k = \frac{\sum_{x_i \in C_k} x_i}{|C_k|} \qquad (2.3.2)$$

Minimization of Sum of Square Error

K-Means clustering is crucially an optimization problem with the goal of minimizing the SSE objective function. We will principally prove the reason why we choose the centroid as a mean of data in its cluster as the root representative for a cluster in K-Means algorithm. Denoting C_k as the k^{th} cluster, any x_i is an element in C_k and c_k is the centroid of the k^{th} cluster. We solve for the representative of C_j which minimizes the SSE by differentiating the SSE with respect to c_j and setting it equal to zero.

$$SSE(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} (c_k - x_i)^2 \qquad (2.3.3)$$

$$\begin{aligned} \frac{\partial}{\partial c_j} SSE &= \frac{\partial}{\partial c_j} \sum_{k=1}^{K} \sum_{x_i \in C_k} (c_k - x_i)^2 \\ &= \sum_{k=1}^{K} \sum_{x_i \in C_k} \frac{\partial}{\partial c_j} (c_k - x_i)^2 \\ &= \sum_{x_i \in C_j} 2 * (c_j - x_i) = 0 \end{aligned}$$

$$\sum_{x_i \in C_j} 2 * (c_j - x_i) = 0 \Rightarrow |C_j| \cdot c_j = \sum_{x_i \in C_j} x_i \Rightarrow c_j = \frac{\sum_{x_i \in C_j} x_i}{|C_j|}$$

Then, the best value of a centroid for minimizing the SSE of a cluster is the mean of the points in that cluster. In K-means, the SSE monotonically decreases with each step of repetition. This behaviour of one-way decrease will finally converge to a local minimum.

Factors affecting K-Means algorithm

The main factors that can impact the performance of the K-means algorithm are following:

1. Initial centroids(first picked points) affects the repetitions of the algorithm to find centroids.

2. The number of clusters K. The suitability of a number of clusters is characteristic for distinct data.

2.3.1 The popular Initialization Methods

In the history of the initialization methods for the Macqueen(1967) published the classical paper which is a simple method by choosing any initial point randomly. This process is worldwide used. Moreover, there are popular K-means initialization methods which improve the performance of its former method. They are shown below.

Hartigan and Wong [1979]: The concept of this method will be good performance for well separated points and the large size of data within surrounding high-dimension sphere for initial points. This method use the Euclidean distance for calculating the distance between points in (2.3.4) and the next centroids are picked by order of decreasing density and maintaining the separation of d_1 from all former centroids.

$$d_1 = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} ||x_i - x_j|| \tag{2.3.4}$$

Milligan [1981]: Milligan uses the result of agglomerative heirarch from Ward's method. Ward's method uses the sum of square errors to calculate the distance between two clusters for the initial centroid and this method keeps and approach the agglomerative growth for smallest value as possible as it can.

Bradley and Fayyad[1998]: This method constructs a subsample from data and applies K-means to calculate the centroids by choosing randomly initial points in Macqueen classical method from the subsample constructed. The final centroids from the subsample will be the initial points for the whole point in data.

D.Arthus and S. Vassilvitskii [2007]: This method choose the first centroid as a random point of data and the next data is chosen by the farthest distance from previous centroids. This method depends on a weight probability score and continues the selection until one obtains the amount of the clusters required.

2.3.2 Estimating the Number of Cluster

Another big problem of the K-means algorithm is the number of clusters (K). One tries to find the new methods for optimizing this challenging problem. These distinct methods are shown below

Calinski-Harabasz Index: The Calinski-Harabasz Index is defined by Equation (2.3.5):

$$CH(K) = \frac{\frac{B(K)}{(K-1)}}{\frac{W(K)}{N-K}} \tag{2.3.5}$$

where N is the number of elements in data or size of data. The number of clusters is considered by evaluating the maximum value of the given function as Equation (2.3.5). Here $B(K)$ and $W(K)$ are the between and within cluster sum of squares, respectively (with K clusters).

Akaike Information Criterion (AIC): This process is improved by the loglikelihood and adding additional constraints of Minimum Description Length (MDL) to calculate the K.M. where K.M. is the dimensionality of the data. K-

means uses an improved AIC as below.

$$KMeans_{AIC} : K = argmin_K[SSE(K) + 2MK] \tag{2.3.6}$$

Bayesian Information Criterion(BIC): The Bayesian method gives approximation to a transformation of the Bayesian probability. This method is similar to AIC, and its process also depends on the loglikelihood. N is the number of data, the value K minimizes the function of BIC as below.

$$BIC = \frac{-2 \times \ln(L)}{N} + \frac{K \times \ln(K)}{N} = \frac{1}{N} \times \ln \frac{N^K}{L^2} \tag{2.3.7}$$

Silhouette Coefficient: The Silhouette Coefficient method provides an appropriate number of clusters. The Silhouette Coefficient has range from -1 to 1, where a high value indicates that clusters are well separated. On the other hand, a low or negative value indicates that it is not a well separated cluster. The Silhouette Coefficients are calculated as Euclidean distance. For a given point i, the average inner distance of its cluster is a_i and the average outer distance of nearest cluster is b_i. Then the Silhouette value of i is $s(i)$ as follow:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Moreover, the Silhouette Coefficient is the average Silhouette value of all data points i in dataset.

2.3.3 The variation of K-means

The proper scope of the K-means makes it very flexible to adapt and construct better algorithms on top of this kind of these algorithms. Some of the diversity of the algorithm tried to the K-mean algorithm are depended on 1) To choose the distinct representative roots for the clusters 2) To choose better initial centroid evaluations, 3) applying some technique of feature transformation.

The most notable variants of K-mean clustering have been implemented as components of partitional clustering.

K-Medoids Clustering: K-medoids and K-means are also partitional clustering methods which try to minimize the value of distance between labeled points in the same clusters and the center point of that cluster, but K-medoids' centroid is actual a data point.

K-Medians Clustering: The K-median computes the median for each cluster as compared to calculate the mean of the cluster (K-means). The K-medians clustering algorithm choose a number of cluster as K that aim to minimize the sum of distance measure between each point and the closest cluster center. The distance measure used in the K-medians algorithm is the L_1-norm, while the square of the L_2-norm used in the K-mean algorithm. The error value for the K-medians algorithm is defined as follow:

$$S_{error} = \sum_{k=1}^{K} \sum_{x_i \in C_k} |x_{ij} - med_{kj}|$$

where x_{ij} represents the j^{th} attribute of the element x_i and med_{ij} represents the median for the j^{th} attribute in the k^{th} cluster C_K.

K-Modes Clustering: One big problem of K-means is its inablity to do with nonnumerical attributes of the variable. Using transformation methods, categorical data can be transformed into new feature spaces, and then the K-means algorithm can be applied to this newly transformed space to obtain the final clusters. However, this method has proven to be very ineffective and does not produce good clusters. It is observed that the SSE function and the usage of

the mean are not appropriate when dealing with categorical data. Hence, the K-modes clustering algorithm has been proposed to tackle this challenge. K-modes is a nonparametric clustering algorithm suitable for handling categorical data and optimizes a matching metric without using any explicit distance metric. The loss function here is a special case of the standard L_p norm where p tends to zero. As opposed to the L_p norm which calculates the distance between the data point and centroid vectors, the loss function in K-modes clustering works as a *metric* and uses the number of mismatches to estimate the similarity between data points.

2.4 Multiple Linear Regression Analysis

Regression analysis is a statistical technique for predicting and investigating the relationship between variables. Multiple Linear Regression(MLR) is a kind of a linear regression model with one response(dependent) variable and more than one explanatory(independent) variables. The general form of the Multiple Linear Regression with k independent is as the follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon, \tag{2.4.1}$$

where Y is the response variable, X_i are independent variable, β_i are the regression coefficients for $i = 1, 2, 3, ..., k$ and ε is the random error component.

2.4.1 The Assumption of the Multiple Linear Regression

We construct some basic assumption on the model as follow:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, 3, ..., n, \tag{2.4.2}$$

with these conditions

1. ε_i and ε_j are uncorrelated for $i \neq j$, that is $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$;
2. ε_i is a normally distributed random variable with zero mean and variance σ^2, that is $\varepsilon_i \sim N(0, \sigma^2)$;
3. There is no a perfect linear relationship between the independent variables, there is no multicolinearity.

2.4.2 Estimation of the Model Parameters

The Multiple Linear Regression in equation (2.4.2), can be written in matrix form as below:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{1k} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

It can be rewritten as:

$$Y = X\beta + \varepsilon, \tag{2.4.3}$$

where Y is the n column vector of the observation,

β is a $k+1$ column vector of the regression coefficient,

X is an $n \times (k+1)$ matrix of the level of the regression variables,

ε is an n column vector of random errors.

Definition 2.1 (Residual). *The difference between the observed(cumulative) value Y_i and the corresponding fitted value $\hat{Y}_i$*

$$\varepsilon_i = Y_i - \hat{Y}_i, \tag{2.4.4}$$

Least Square Estimation of the Regression Coefficient

The Least Square method of the Regression coefficient is to estimate the regression coefficient in equation (2.4.2). Now, we assume $b_0, b_1, b_2, ..., b_k$ to be the estimators of $\beta_0, \beta_1, \beta_2, ... \beta_k$ respectively. Then we get the equation of the corresponding value $\hat{Y}_i$ as follow:

$$\hat{Y} = Xb.$$

Then, the least square error(S) of the regression is

$$\begin{aligned} S = \sum_{i=1}^{n} \varepsilon_i^2 &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} (Y_i - b_0 - \sum_{j=1}^{k} b_j X_{ij})^2. \end{aligned}$$

The least square error function S must have minimum value because we need the least error value, then we take partial derivative and take the partial derivative equal to 0 as below:

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^{n} \left[Y_i - b_0 - \sum_{j=1}^{k} b_j X_{ij} \right] = 0 \qquad (2.4.5)$$

$$\frac{\partial S}{\partial b_j} = -2 \sum_{i=1}^{n} \left[Y_i - b_0 - \sum_{j=1}^{k} b_j X_{ij} \right] X_{ij} = 0, \quad j = 1, 2, ..., k \qquad (2.4.6)$$

From equation (2.4.5) and (2.4.6), we obtain the expressed equation of the least square equations

$$\begin{aligned} nb_0 + b_1 \sum_{i=1}^{n} X_{i1} + b_2 \sum_{i=1}^{n} X_{i2} + b_3 \sum_{i=1}^{n} X_{i3} + \cdots + b_k \sum_{i=1}^{n} X_{ik} &= \sum_{i=1}^{n} Y_i \\ b_0 \sum_{i=1}^{n} X_{i1} + b_1 \sum_{i=1}^{n} X_{i1}^2 + b_2 \sum_{i=1}^{n} X_{i1} X_{i2} + \cdots + b_k \sum_{i=1}^{n} X_{i1} X_{ik} &= \sum_{i=1}^{n} X_{i1} Y_i \qquad (2.4.7) \\ \vdots \qquad\qquad &\qquad \vdots \end{aligned}$$

$$b_0\sum_{i=1}^{n}X_{ik}+b_1\sum_{i=1}^{n}X_{ik}X_{i1}+b_2\sum_{i=1}^{n}X_{ik}X_{i2}+\cdots+b_k\sum_{i=1}^{n}X_{ik}^2=\sum_{i=1}^{n}X_{ik}Y_i$$

We have obtained $k+1$ equations and the $k+1$ unknown regression coefficients. Now, we can find the solution (coefficients $b_0, b_1, ...b_k$) with the least square estimators.

From the least square equation (2.4.7) as below:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & \dots & X_{nk} \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}=\begin{bmatrix} n & \sum_{i=1}^{n}X_{i1} & \dots & \sum_{i=1}^{n}X_{ik} \\ \sum_{i=1}^{n}X_{i1} & \sum_{i=1}^{n}X_{i1}^2 & \dots & \sum_{i=1}^{n}X_{i1}X_{ik} \\ \sum_{i=1}^{n}X_{i2} & \sum_{i=1}^{n}X_{i1}X_{i2} & \dots & \sum_{i=1}^{n}X_{i2}X_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}X_{ik} & \sum_{i=1}^{n}X_{i1}X_{ik} & \dots & \sum_{i=1}^{n}X_{ik}^2 \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

It can be written in the another way as below

$$X^TY=(X^TX)b, \qquad (2.4.8)$$

where $b=[b_0,b_1,...,b_k]^T$. Our purpose is to find the solution of the least square estimators for the least error, Then we solve the above equation by multiplying $(X^TX)^{-1}$ on the both sides of the equation s

$$(X^TX)^{-1}X^TY=(X^TX)^{-1}(X^TX)b$$

$$(X^TX)^{-1}X^TY=Ib,$$

where I is the identity matrix. Now, we can conclude that $b=(X^TX)^{-1}X^TY$ has the least error for the regression coefficients.

2.4.3 Test for Significance of Regression

The test for significance of regression is kind of consideration test between response variable and regressor variables in term of linear relationship.

Assume that the estimator parameter $\beta_0, \beta_1, ..., \beta_k$ are $b_0, b_1, ..., b_k$ respectively.

The statement for the hypotheses is:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = ... = \beta_k = 0;$$

$$H_1 : \beta_i \neq 0 \text{ for at least one } i.$$

The statistic, which can be used for testing H_0 against H_1 is the t test statistic in the following form:

$$t^\star = \frac{b_i}{\mathrm{se}(b_i)} \sim t_{\frac{\alpha}{2}, n-(k+1)}, \qquad \forall i = 1, 2, 3, ..., k,$$

where $\mathrm{se}(b_i) = \sqrt{\sigma^2 C_{ii}}$, C_{ii} are the elements on the main diagonal of $(X^T X)^{-1}$ and we reject H_0 if $t_{\frac{\alpha}{2}, n-(k+1)} < |t^\star|$

2.5 Polynomial Regression

The polynomial regression is a kind of statistical technique similar to linear regression. The polynomial regression can be used, when the relationship between response variable and explanatory(independent) variables is curvilinear. The polynomial regression model should keep the order of independent variables as low as possible for avoiding the situation of over-fitting of the data.

2.5.1 Polynomial models

Suppose 2 is the highest degree as order of the polynomial regression in two variable given by

$$Y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 + \beta_6 x_2^2 + \varepsilon$$

where Y is the response variable, x_1 and x_2 are an independent variables, β_i are the regression coefficient for $i = 1, 2, ..., 6$ and ε is the random error component. Moreover, the solution for polynomial's coefficients can be considered as Multiple Linear Regression and also solved by least square estimation.

2.5.2 Mean Absolute Error

Mean Absolute Error (MAE) is a commonly used to measure the difference between the observed value Y_i, and the corresponding fitted value $\hat{Y}_i$. Its form is as follow:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |Y_i - \hat{Y}_i|,$$

where N is a number of observations.

2.5.3 Root Mean Square Error

Root Mean Square Error is another commonly used measure of the difference between the observed value Y_i, and the corresponding fitted value $\hat{Y}_i$. Its form is as follow:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N}},$$

where N is a number of observations.

2.5.4 Related Research

Tzortzis, G. and Likas, A. (2008) developed kernel K-means by extension from normal K-means. Their aim is to approach to kernel-based clustering by the global K-mean algorithm. Moreover, they can solved the initialization problem with the kernel K-mean and reduce the computation cost.

Liu, Y. and Jiang, K. (2014) first applied an ANN-based bagging algorithm to forecast the daily cash demand for the next few days, then constructed the optimal integer programming model by significant factors. This method is also suitable for any cash circulation domain such as bank cash inventory management.

Venkatesh, K., Ravi, V., Prinzie, A. and Van Den Poel, D. (2014) first clustered ATM centers into ATM clusters having similar day-of-the week withdrawal patterns. They built a time series model for each ATM. For each cluster of ATMs, four neural networks viz., general regression neural network (GRNN), multi layer feed forward neural network (MLFF), group method of data handling (GMDH) and wavelet neural network (WNN) are built to predict an ATM center's cash demand. They observed that GRNN yielded the best result of 18.44% symmetric mean absolute percentage error (SMAPE), which is better than the result of Andrawis, Atiya, and El-Shishiny (2011).

Ekinci, Y., Lu, J.-C. and Duman, E. (2015) The bank want to take less resource for the fluctuated demand of customer. They group ATMs into nearby-location clusters and also optimize the aggregates of daily cash withdraws in the forecasting process.

Jadal, P.K., Jain, S. and Khanna, P. (2018) forecasted the cash demand forecasting of NN5 data for ATMs with Neural network. The NN5 reduced dataset is a subsample of 11 series time of 111 complete dataset of 111 daily times series. In their second model, they apply the clustering algorithm to their ATMs, then applying Neural Network. The root mean square error is calculated for 2 models. They can conclude that the second model gives the least root mean square error.

CHAPTER III

RESEARCH METHODOLOGY

This part will show the process, which consists of 6 parts, used in this research.

3.1 The Collected Data and Determined Variables as Statistics

This research used secondary data collected from 2011 January until 2013 May, 910 data set in total. Table (3.1) below shows the determined roles of each variables in our data provide by https://www.kaggle.com/nitsbat/data-of-atm-transaction-of-xyz-bank.

Table 3.1 A table with roles of collected data.

Variables	Types
The amount of Withdrawal	Response
Weekday	Independent
Working Day	Independent
Date	Independent
Month	Independent
Mean Withdrawal of Previous Week	Independent
Withdrawal of Previous Same Day	Independent

The data in Table 3.1 is saved in form of a CSV file because CSV is available for program R. Moreover, our data is analyzed by an R-program and Rapidminer program.

3.2 Program

We use the free software environment for analysis and computation. Our program consist of R Studio program with version 1.3.1093, Rapidminer with version 9.8.001 and Python 3 version 3.8.6, working on MS Windows 10 operation system.

3.3 Study and Analyze the Relation between Independent and response Variables

We construct the Multiple Linear Regression Model of all variables to analize the significance of independent variables which affect to the amount of withdrawal by following equation:

$$\begin{aligned}\mathbf{Total} = \beta_0 + \beta_1\mathbf{Weekday} + \beta_2\mathbf{WorkingDay} + \beta_3\mathbf{Date} + \beta_4\mathbf{Month}\\ + \beta_5\mathbf{MeanofPrev} + \beta_6\mathbf{SamedayPrev},\end{aligned}$$

where

1. **WorkingDay** means that day is holiday or working day;

2. **MeanofPrev** is Mean Withdrawal of Previous Week;

3. **SamedayPrev** is Withdrawal on the Day in the Previous week.

Test for Significance of Regression

We used this test for investigating the independent variables which most affect the amount of withdrawal in model of Multiple Linear regression.

Assume that the estimator parameter $\beta_0, \beta_1, ..., \beta_6$ are $b_0, b_1, ..., b_6$ respectively.
The statement for the hypotheses is:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = ... = \beta_6 = 0;$$

$$H_1 : \beta_i \neq 0 \text{ for at least one } i.$$

The statistic, which can be used for testing H_0 against H_1 is the t test statistic in the following form:

$$t^{\star} = \frac{b_i}{\mathrm{se}(b_i)} \sim t_{\frac{\alpha}{2},n-(k+1)}, \qquad \forall i = 1, 2, 3, ..., 6.$$

where $\mathrm{se}(b_i) = \sqrt{\sigma^2 C_{ii}}$, C_{ii} is an element on the main diagonal of $(X^T X)^{-1}$ and we reject H_0 if $t_{\frac{\alpha}{2},n-(k+1)} < |t^{\star}|$

In term of technical method, we used 5% of a significant level ($\alpha = 0.05$), and a number of training data equal to 729 ($n = 729$). Then the term of t-distribution $t_{0.025,727}$ is estimated in t-Distribution table approximately to 1.965.

3.4 Creating a Forecasting Model

After we had obtained the significant factors affecting the amount of withdrawal, we constructed the appropriate model for forecasting the value of the amount of withdrawal by

1. Multiple Linear Regression;

2. K-mean Clustering together with Multiple Linear Regression;

3. K-mean Clustering together with Polynomial Regression.

The K-mean Clustering is a kind of classification in (unsupervised) Machine Learning to separate things into the same cluster (group) with an appropriate number of cluster.

Method for selecting an appropriate number of cluster

- Silhouette Coefficient.

3.5 Predicted Data

The predicted data is used for the above model from 2013 January until 2013 May, using the remaining 181 data.

3.6 Measurement of Forecasting Model

When we try to forecast the future value, we must have an indicator of the error to show the accuracy of the forecasting model

The indicator of error value for forecasting model

1. Root Mean Square Error

2. Mean Absolute Error

CHAPTER IV

RESULTS

This part will show the result of the process from previous Chapter. Our goals is to show the consequence computed in 3 models. Moreover we could gain the knowledge thoroughly about the process of finding the significant factors affect to the amount of withdrawal, choosing a number of cluster by silhouette coefficient, concluding the best model among them for this data.

4.1 Descriptive Statistics of Training Data

Descriptive statistics is a summary statistics that quantitatively describes or summarizes features from the collection of the training data that is essential in part of Machine Learning

Table 4.1 A details of the training data.

Variables	**Minimum**	**Maximum**	**Average**
Total	59700	1296600	546177.709
Weekday	1	7	3.996
Working Day	1	2	1.357
Date	1	31	15.778
Month	1	12	6.535
MeanofPrev	352272	937129	545595.498
SamedayPrev	59700	1296600	547260.115

Remark: Since, the Weekday is from Sunday to Saturday and we substitute from number 1 till number 7 i.e. Sunday = 1, Monday = 2, ... , Saturday = 7. Working is also substituted as Working Day = 1 and Holiday = 2 and The factor Month is also substituted from 1 to 12 with initial Month as January.

4.2 The Significant Factors

This step is an analysis of these independent factors affecting the amount of withdrawal by using Multiple Linear Regression with 5% of significant level.

Before analysing the factors, we have to rescale the data into the same standard by the standardization method. the rescaled data has the mean and standard deviation equal to 0 and 1 respectively. The equation for the standardization method is

$$z_i = \frac{x_i - \bar{x}}{\sigma},$$

Next, factors resulting in the lowest root mean square error singled out to be used in the model. Then, we get 3 significant factors affecting the amount of withdrawal namely Working Day, Date, Withdrawal of Previous the Day in previous week. Their coefficients are shown in Table 4.2 commputed by R program with the following command:

```
> fit = lm(Total ~ Weekday + Workingday + Date + Month + MeanofPrev + SameDayPrev, data = Data)
```

Table 4.2 The coefficients and statistical information of the training data with 5% of a significant level.

Variables	**Coefficient**	**std.error**	**t value**	**p-value**
Intercept	-2.216e-10	3.312e-02	0.000	1.00000
Weekday	-5.455e-03	4.417e-02	-0.123	0.9017
Working Day	-1.004e-01	-4.403e-02	-2.279	0.0229
Date	-0.418	3.356e-02	-12.704	< 2e-16
Month	5.969e-02	3.324e-02	1.796	0.0729
MeanofPrev	-6.822e-02	4.035e-02	-1.691	0.0913
SamedayPrev	1.295e-01	4.010e-02e-02	3.228	0.0013

4.3 Constructing the forecasting model with the significant factors

Having obtained the three significant factors, we took these factors to create the model by using the Rapidminer program and Python. We constructed 3 models for forecasting the amount of withdrawal.

Model I: Multiple Linear Regression Model

We create the Multiple Linear Regression model by using Rapidminer

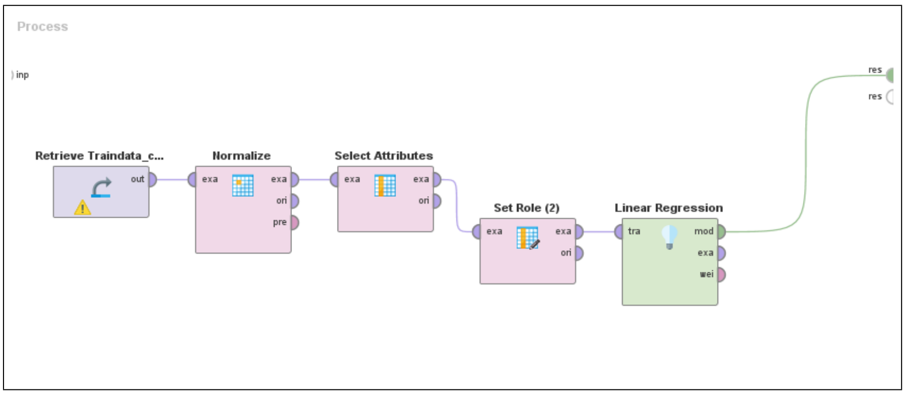


Figure 4.1 Constructing Multiple Linear Regression model in Rapidminer.

Table 4.3 The coefficients and statistical information of the significant training data.

Variables	**Coefficient**	**std.error**	**t value**	**p-value**
Intercept	-2.232e-10	3.319e-02	0.000	1.00000
Working Day	-1.044e-01	3.321e-02	-3.144	0.00173
Date	-0.4181	3.331e-02	-12.554	$<$ 2e-16
SamedayPrev	9.597e-02	3.330e-02	2.881	0.00408

Then the equation for forecasting in model I as follow:

$$\textbf{Total} = -0.1044\times\textbf{Working Day}-0.4181\times\textbf{Date}+0.09597\times\textbf{SamedayPrev}-2.232e-10$$

In the next step, we push the test data against the above model. We can get the consequence by Rapidminer in Figure 4.2, and computing the accuracy of the model by Root Mean Square Error(RMSE) and Mean Absolute Error(MAE). The RMSE and MAE of this model is equal to 1.0360 and 0.7485 respectively.

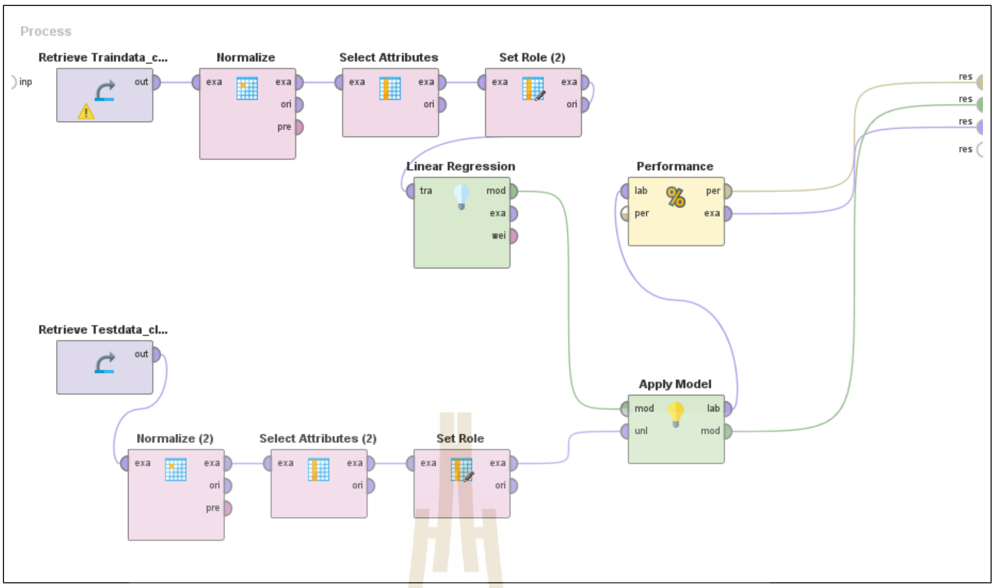


Figure 4.2 Test data against Multiple Linear Regression model in Rapidminer.

In the remaining two models, a clustering part is inserted before constructing both models. So, we express the clustering algorithm how it works for our data. Of concern are the initial point and estimating the number of clusters. In our process, we ignored the problem of picking the initial point because we are not interested in the number of iterations of K-mean clustering. We are only concern with the problem of a number of cluster and we solve this problem by Silhouette Coefficient method mentioned below.

Silhouette Coefficient Method

The Silhouette Coefficient method provides appropriate number of cluster. Silhouette Coefficient has range from -1 to 1, with a high value indicating that clusters are well separated. On the other hand, a low or negative value indicates that it is not a well separated cluster. The Silhouette are calculated as Euclidean distance.

First, assume the data are clustered by K-means technique into k clusters

For any point $x \in C_i$ (data point in its cluster) let $a(x)$ denote its inner average distance (average distance between point i and other point in same cluster)

$$a(x) = \frac{1}{|C_i| - 1} \sum_{y \in C_i, x \neq y} ||x - y||_2$$

where $|| \cdot ||_2$ is Euclidean distance and $|C_k|$ is a number of elements in cluster k. $b(x)$ is an average distance of point x to all points of nearest cluster. Suppose cluster k is nearest cluster to point x, then we get average distance $b(x)$ as equation:

$$b(x) = \frac{1}{|C_k|} \sum_{y \in C_k} ||x - y||_2$$

Now, we define a silhouette value $s(i)$ of point i by

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad |C_i| > 1 \qquad (4.3.1)$$

In case $|C_i| = 1$, then $s(i) = 0$. In description of $s(i)$, equation (4.3.1) shows the range of $s(i)$ less than 1 and more than -1 ($-1 \leq s(i) \leq 1$). It can be written as:

$$s(i) = \begin{cases} 1 - a(i)/b(i), & a(i) < b(i) \\ 0, & a(i) = b(i) \\ b(i)/a(i) - 1, & a(i) > b(i) \end{cases}$$

The average of Silhouette value of the entire dataset for specific k is denoted by $s(k)$. Then, the Silhouette Coefficient(SC) for maximum $s(k)$ is

$$SC = \max_k\{s(k)\}$$

We calculated the Silhouette Coefficient via Python 3.8.6. First, we show the data in form of Standardization in Figure 4.3

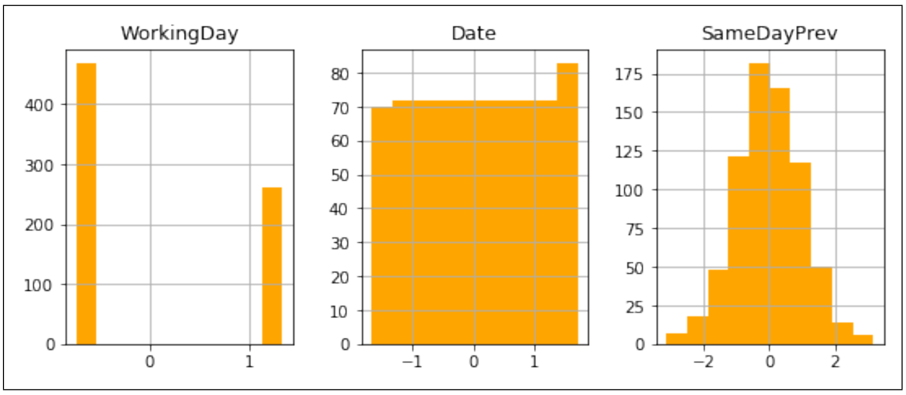
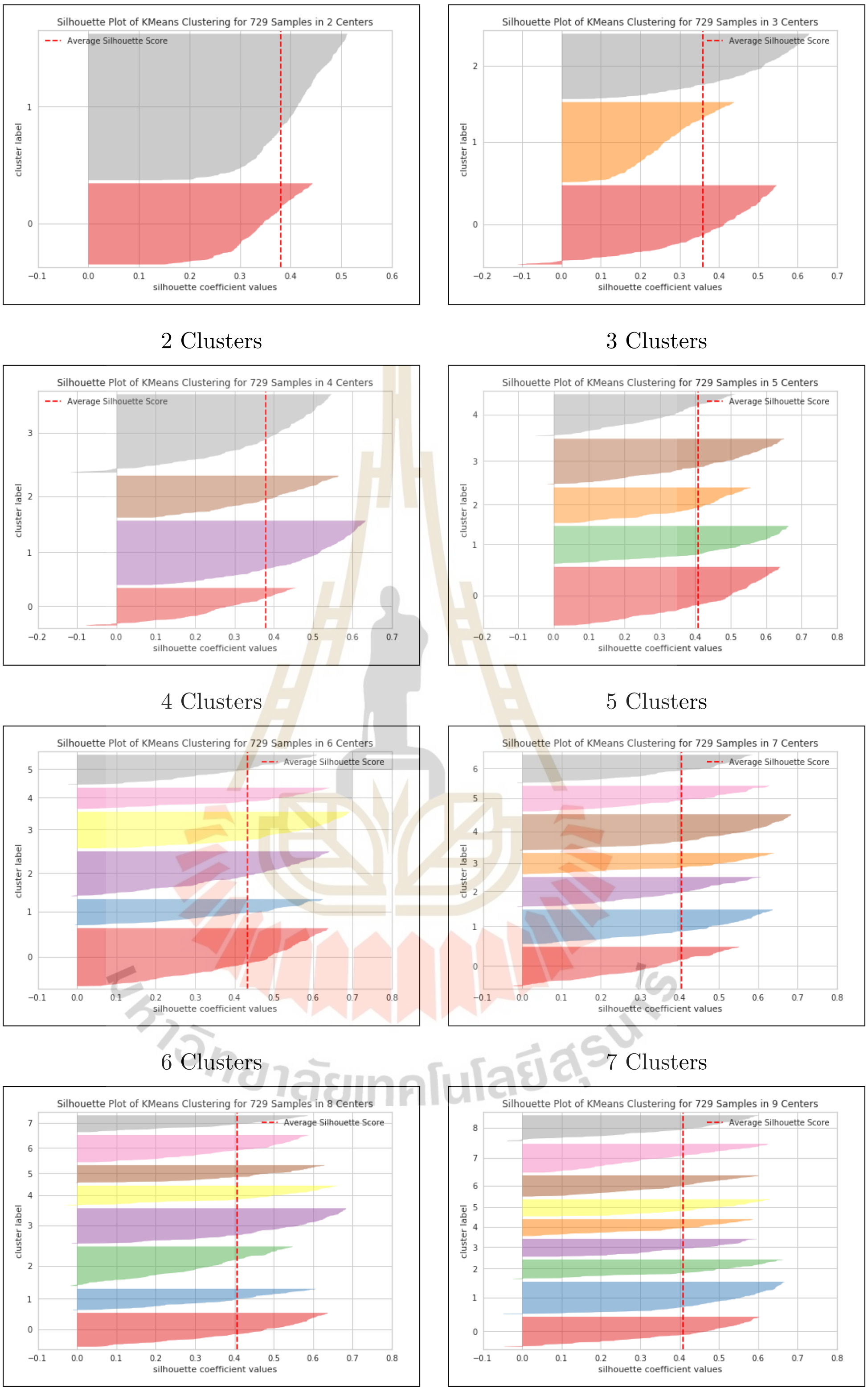


Figure 4.3 Training Data Standardization for the 3 significant variables.

Next, we find the appropriate number of clusters with Silhouette Coefficient method by supposing a number of clusters starting from 2 clusters up to 10 clusters. Then, we obtained the result as shown in Figure 4.4



2 Clusters

3 Clusters

4 Clusters

5 Clusters

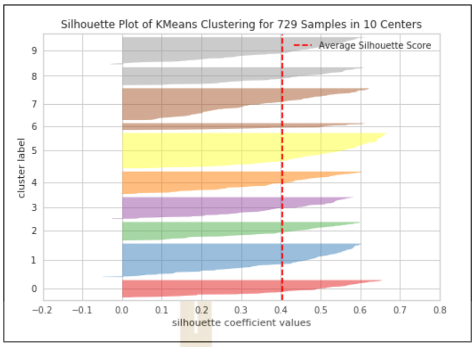
6 Clusters

7 Clusters

8 Clusters

9 Clusters

Figure 4.4 The Silhouette Coefficients from 2 clusters to 10 clusters.



10 Clusters

Figure 4.4 (Continued) The Silhouette Coefficients from 2 clusters to 10 clusters.

Table 4.4 Silhouette Coefficients of 9 cluster forms.

Number of Clusters	**Silhouette Coefficients**
2	0.379906
3	0.358355
4	0.379309
5	0.407054
6	0.434109
7	0.405206
8	0.406324
9	0.409357
10	0.404410

We can detect that the best Silhouette Coefficient value is 0.434109 from data with 6 clusters. Thus, we conclude that the best number of clusters for this data is 6 clusters.

K-mean Clustering

Next, we use K-mean clustering to separate the data into 6 clusters. The number of data and their centroids in each cluster after clustering are shown in Figure 4.5 and Table 4.5.

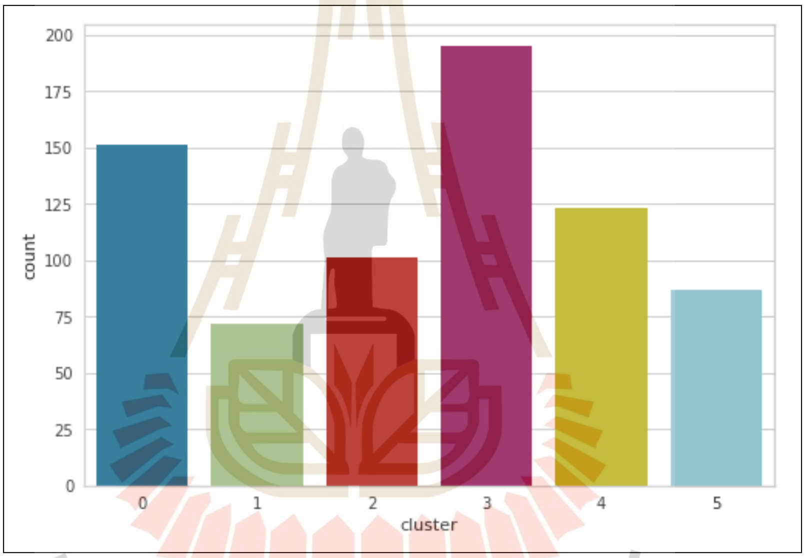


Figure 4.5 A number of data in each clusters.

Table 4.5 Table shows the centroids of each clusters.

Cluster	Working Day	Date	SamedayPrev
0	-0.7446	-0.2604	0.999
1	1.3431	-1.1058	-0.7296
2	1.3431	0.9314	-0.571
3	-0.7446	0.9627	-0.3906
4	-0.7446	-1.1851	-0.5461
5	1.3431	-0.1964	1.1803

Moreover, we show a violin plot in Figure 4.6 from these factors to understand the value of variables in each cluster.

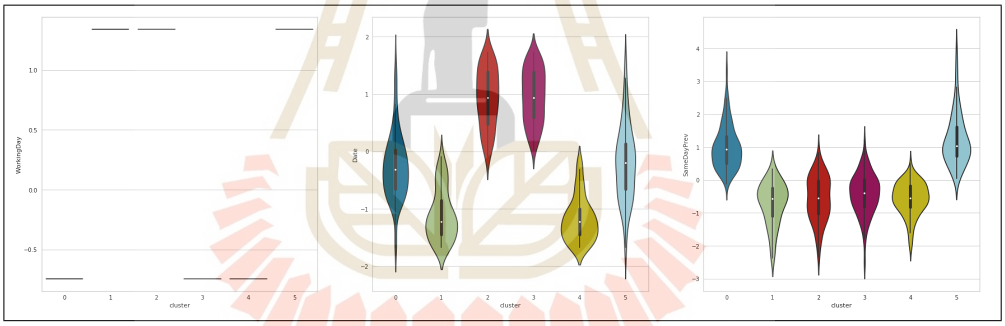


Figure 4.6 Value of variables in each cluster as violin plot.

Model II: Clustering + Multiple Linear Regression Model

After separating the data by K-mean clustering, we construct the Multiple Linear Regression model in each cluster from training data by Rapidminer. This means these models are unique(distinct among them).

The equation of the model in **Cluster 0**:

$$\textbf{Total} = -0.588 \times \textbf{Date} + 0.05 \times \textbf{SamedayPrev} - 0.035.$$

The equation of the model in **Cluster 1**:

$$\textbf{Total} = -1.264 \times \textbf{Date} + 0.33 \times \textbf{SamedayPrev} - 1.012.$$

The equation of the model in **Cluster 2**:

$$\textbf{Total} = 0.1831 \times \textbf{Date} + 0.2472 \times \textbf{SamedayPrev} - 0.6188.$$

The equation of the model in **Cluster 3**:

$$\textbf{Total} = -0.1212 \times \textbf{Date} + 0.1046 \times \textbf{SamedayPrev} - 0.2939.$$

The equation of the model in **Cluster 4**:

$$\textbf{Total} = -0.552 \times \textbf{Date} - 0.117 \times \textbf{SamedayPrev} - 0.071.$$

The equation of the model in **Cluster 5**:

$$\textbf{Total} = -0.398 \times \textbf{Date} + 0.136 \times \textbf{SamedayPrev} - 0.065.$$

In the next step, we classify the test data into a cluster of a nearest centroid i.e. suppose data point i is nearest to the centroid of cluster k, then we classify the data point i into Cluster K. And we found that 37, 18, 34, 43, 32, 17 data points are in cluster 0, 1, 2, 3, 4, 5 respectively. Now, we push the test data against the model of its cluster. We obtain the result by Rapidminer in Figure 4.7, and compute the accuracy of all model by Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The RMSE and MAE of this model are equal to 1.0137 and 0.7498, respectively.

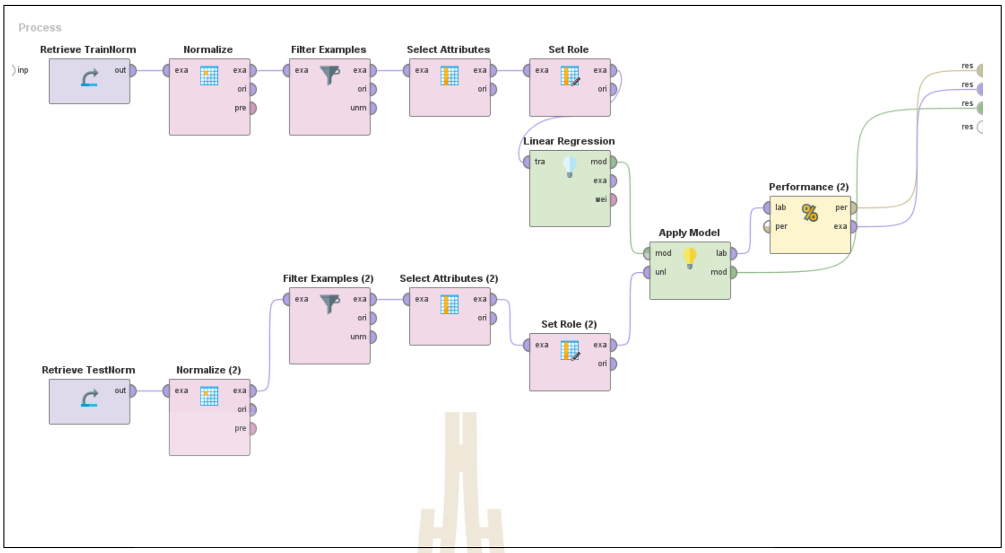


Figure 4.7 Test data against Multiple Linear Regression model after clustering in Rapidminer.

Model III: Clustering + Polynomial Regression Model

In this part, we perform every step similar to model II but replacing Multiple Linear Regression by Polynomial Regression.

Having separated the data by K-mean clustering, we construct the Polynomial Regression model in each cluster from the training data by Rapidminer. This means these model are also unique.

The equation of the model in **Cluster 0**:

$$\textbf{Total} = -0.365 \times \textbf{Date}^3 + 0.113 \times \textbf{SamedayPrev} - 0.008.$$

The equation of the model in **Cluster 1**:

$$\textbf{Total} = 0.576 \times \textbf{Date}^2 + 0.017 \times \textbf{SamedayPrev}^4 - 0.622.$$

The equation of the model in **Cluster 2**:

$$\textbf{Total} = 0.203 \times \textbf{Date} - 0.04 \times \textbf{SamedayPrev}^4 - 0.76.$$

The equation of the model in **Cluster 3**:

$$\mathbf{Total} = -0.137 \times \mathbf{Date} + 0.022 \times \mathbf{SamedayPrev}^2 - 0.234.$$

The equation of the model in **Cluster 4**:

$$\mathbf{Total} = 0.045 \times \mathbf{Date}^4 - 0.142 \times \mathbf{SamedayPrev} + 0.423.$$

The equation of the model in **Cluster 5**:

$$\mathbf{Total} = -0.297 \times \mathbf{Date}^3 + 0.154 \times \mathbf{SamedayPrev} - 0.051.$$

In the next step, we classify the test data into a clusters of a nearest centroid, and push the test data against the model of its cluster. The result obtained by Rapidminer in Figure 4.8. The accuracy of all models is computed by Root Mean Square Error(RMSE) and Mean Absolute Error(MAE), are equal to 0.8318 and 0.6305, respectively.

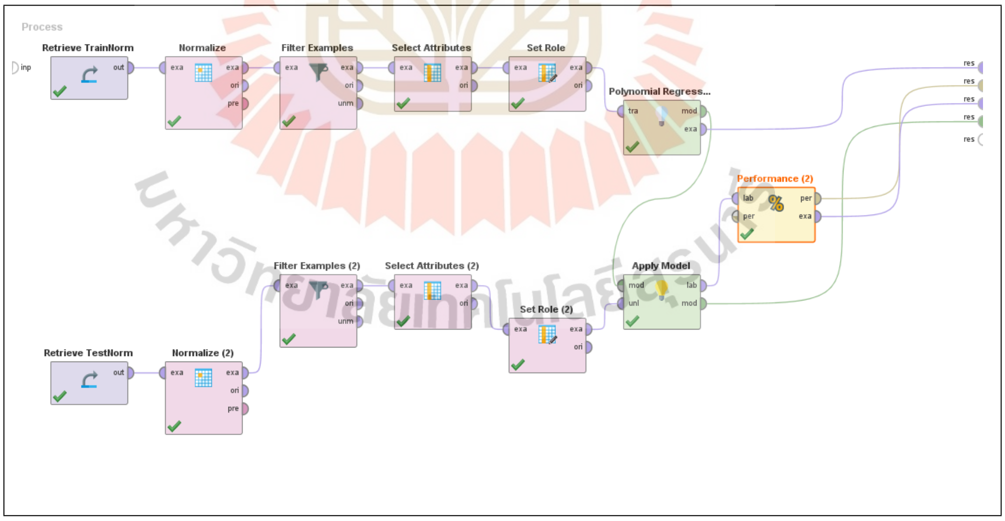


Figure 4.8 Test data against Multiple Linear Regression model after clustering in Rapidminer.

Table 4.6 shows a comparison of the RMSE and MAE errors obtained by the three models:

Table 4.6 Root Mean Square Error and Mean Absolute Error of three proposed model.

Attribute	MLR	Clutering and MLR	Clustering and Polynomial
RMSE	1.0360	1.0137	0.8318
MAE	0.7485	0.7498	0.6305

Finally, the graph of prediction in three model in term of rescaled data as standardization method is shown in Figure 4.9.

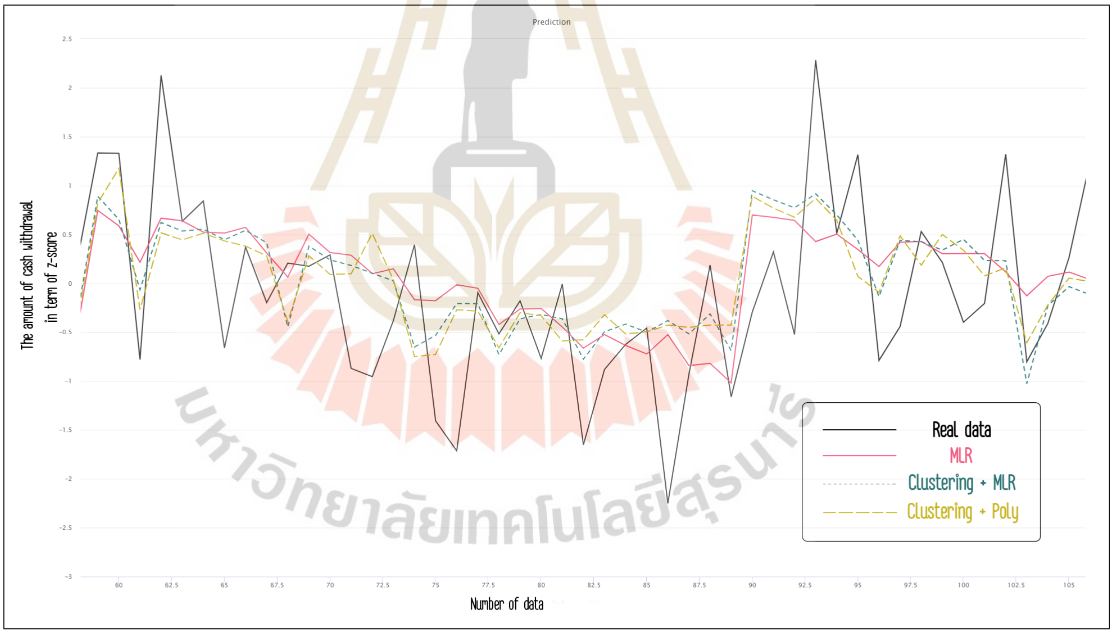


Figure 4.9 The graph of the forecasted values by the three models and the real value of the amount of withdrawal of the customers' demand.

CHAPTER V

CONCLUSION

This research constructed an optimal model to forecast the amount of withdrawal of customer's demand by Polynomial regression and Clustering algorithm. The factors' analysis shows that there were 3 factors which affected to the working day, date, withdrawal of previous same day, with 5% statistical significance. The same direction factors were working day and date, but the opposite direction factor was withdrawal of previous same day. The verification by using data of the amount of withdrawal of customer's demand from 2013 January until 2013 May showed that our polynomial model provided the least error of both RMSE and MAE, 0.8318 and 0.6305 respectively, only whereas the classical multiple linear regression model provided RMSE and MAE 1.0360 and 0.7485, respectively. Therefore, we claim that the model using the polynomial regression and clustering algorithm got better RMSE and MAE values than the classical multiple linear regression.

REFERENCES

REFERENCES

Bisht, N. (2019). **Bank Transactions Data of ATM transaction of XYZ bank**. (online) Available: https://www.kaggle.com/nitsbat/data-of-atm-transaction-of-xyz-bank.

Bradley, P.S. and Fayyad, U.M. (1998). Refining Initial Points for K-Means Clustering. **Proceedings of the Fifteenth International Conference on Machine Learning**. Madison, WI, 24-27 July 1998, Pages 91-99.

David, A. and Sergei, V. (2007). k-means++: the advantages of careful seeding. **Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms**. January 2007, Pages 1027–1035.

Ekinci, Y., Lu, J.-C. and Duman, E. (2015). Optimization of ATM cash replenishment with group-demand forecasts. **Expert Systems with Applications**. 42(7): 3480-3490.

Jadwal, P.K., Jain, S., Gupta, U. and Khanna, P. (2018). K-means Clustering with Neural Networks for ATM Cash Repository Prediction. **Conference: International Conference on Information and Communication Technology for Intelligent Systems (ICTIS 2017)**. 1; 588-596.

Jong, P.D., and Heller, G.Z. (2008). Geralized Linear Models for Insurance Data. **Cambridge University Press**.

Liu, Y. and Jiang, K. (2014). An optimal ATM cash replenishment solution using ANN-based bagging algorithm. **9th International Symposium on Linear**

Drives for Industry Applications. Hangzhou, 7 July 2013 - 10 July 2013. 270(1): 217-224.

Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. **In proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability**. 1: 281-297, Berkeley, CS, USA, 1967.

Manning, C.D., Raghavan, P. and Schutze, H. (2007). Rough set based generalized fuzzy c-means algorithm and quantitative indices. **IEEE Transaction on Systems, Man, and Cybernetics**, 37(6): 1529-1540.

Milligan Glenn, W. (1981). A Review Of Monte Carlo Tests Of Cluster Analysis. **Multivariate Behavioral Research**, 16(3); 379-407.

Selim, S.Z. and Ismail, M.A. (1984). K-means-type algorithm: A generalized convergence theorem and characterization of local optimality. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. 6(1): 81-87.

Tzortzis, G. and Likas, A. (2008). The Global Kernel k-Means Clustering Algorithm. **2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)**. Hong Kong, China, June 1-8, 2008, Pages 1977-1984.

Venkatesh, K., Ravi, V., Prinzie, A. and van den Poel, D. (2014). Cash demand forecasting in ATMs by clustering and neural networks. **European Journal of Operational Research**, 232(2): 383-392.

CURRICULUM VITAE

NAME : Pannawit Kongmuangpak **GENDER :** Male

EDUCATION BACKGROUND:

- Bachelor of Science (Mathematics), Suranaree University of Technology, Thailand, 2018

SCHOLARSHIP:

- Development and Promotion of Science and Technology Talents Project Scholarship

CONFERENCE:

- The 1st International Virtual Conference on Science and Technology (SUT-IVCST 2020) August 28th, 2020
- The 25th Annual Meeting in Mathematics (AMM2021) May 28th, 2021

EXPERIENCE:

- Teaching assistant in course of Calculus at Suranaree University of Technology, semester from 2/2019 to 1/2020
- Staff of DPST summer camp at Khaoyai June, 2019