



ST. MARY'S UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF INFORMATICS

MSc PROGRAM IN COMPUTER SCIENCE

**CONSTRUCTING A MODEL FOR BANK CUSTOMER SEGMENTATION: THE CASE
OF AWASH BANK**

BY:-Abebe Zeleke

ID No.SGS/0438/2010A

Advisor: Million Meshesha(Ph.D.)

**A Thesis Submitted to St. Mary's University as partial fulfillment of the requirements for
Master of Science in Computer Science**

January 2024

ADDIS ABABA, ETHIOPIA

ST. MARY'S UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Faculty of Informatics

Board of Examiners

As members of the Examining Board of the final MSc, open defense, we certify that we read and evaluated the thesis prepared by Abebe Zeleke and recommended that it be accepted as fulfilling the thesis requirement for the Degree of Master of Computer Science

1. _____
Name of Chairman
Signature and Date

2. _____
Name of Advisor
Signature and Date

3. _____
Name of External Examiner
Signature and Date

4. _____
Name of Internal Examiner
Signature and Date

Declaration

I, Abebe Zeleke, declare that this thesis is my original work, prepared under the guidance of Million Meshesha (Ph.D.) all sources of materials used for the thesis have been duty knowledge. I further confirm that the thesis has not been submitted either in part or full to any Other higher learning institution for the purpose earning any degree .

Name: Abebe Zeleke

Signature _____ Date _____

Endorsement

This thesis entitled “Constructing a Model for Bank Customer Segmentation in the Case of Awash Bank in Addis Ababa” has been submitted to St. Mary’s University, School of graduate students for examination with my approval as a university advisor.

Name of Advisor, Million Meshesha (PhD)

Signature_____Date_____.

Acknowledgements

Above all, I praise Almighty God who has blessed my work and for giving me the health, strength and endurance until this time. In the preparation of this thesis many have contributed priceless data, idea, resource books as well as moral support. First and foremost, my deepest gratitude goes to my advisor, Million Meshesha (Ph.D.), for his valuable and constructive ideas as well as all his precious time spent with high collaboration in reviewing and improving the quality of the thesis work.

I am deeply grateful to all including my wife who has given me assistance in obtaining the information and data related to this work. Particular thanks also go to the experts and staff at the Awash Bank for their willingness to provide me with all the necessary data and information so that the research work could be carried out. I am also thanks to the moral support from my family abroad, in Ethiopia and friends will never be forgotten and this work is dedicated to them.

Table of Contents

Acknowledgements.....	1
List of Figures.....	4
List of Tables	5
Abstract.....	6
CHAPTER ONE	Error! Bookmark not defined.
INTRODUCTION.....	Error! Bookmark not defined.
1.1Background of the Study	Error! Bookmark not defined.
1.3. Statement of the problem.....	Error! Bookmark not defined.
1.3.1 Research Questions.....	Error! Bookmark not defined.
1.4 Objective of the study	Error! Bookmark not defined.
1.4.1 General Objective	Error! Bookmark not defined.
1.4.2 Specific Objectives	Error! Bookmark not defined.1
1.5 Scope of the Study	Error! Bookmark not defined.1
1.6 Significance of the Study	Error! Bookmark not defined.
1.7 Methodology of the study	Error! Bookmark not defined.2
1.7.1. Research design	Error! Bookmark not defined.3
1.7.2 Method of data collection and preparation	Error! Bookmark not defined.3
1.7.3 Implementation Tools	Error! Bookmark not defined.3
1.7.4 Evaluation methods.....	Error! Bookmark not defined.3
1.7.5 Definition of key terms	Error! Bookmark not defined.4
1.7.6 Organization of the Study	Error! Bookmark not defined.4
CHAPTER TWO	Error! Bookmark not defined.5
LITERATURE REVIEW	Error! Bookmark not defined.5
2.1 Overview of Customer Segmentation.....	Error! Bookmark not defined.5
2.2 Customer segmentation models	Error! Bookmark not defined.6
2.3 Types of customer Segmentation.....	Error! Bookmark not defined.7
2.4.1 Benefits of customer Segmentation	19
2.5 Data mining methods for identifying clients.....	Error! Bookmark not defined.0
2.6 Related Works.....	21

CHAPTER THREE	Error! Bookmark not defined.	3
APPROACHES AND TECHNIQUES	Error! Bookmark not defined.	3
3.1 Overview	Error! Bookmark not defined.	3
3.2. Data Mining in Awash Bank.....	Error! Bookmark not defined.	3
3.3 Data Mining and Knowledge Discovery in Database	Error! Bookmark not defined.	3
3.4 Learning Approaches	Error! Bookmark not defined.	5
3.4.1 Supervised learning.....	Error! Bookmark not defined.	5
3.5.1. Filtered Clustering Methods	Error! Bookmark not defined.	7
3.5.2 Density-based Clustering method		28
3.5.3 K-means algorithm.....		29
3.6 Collection and Preparation of Data	Error! Bookmark not defined.	0
3.6.1 Dataset collection.....	Error! Bookmark not defined.	0
3.6.2 Customer Details Sample Record	Error! Bookmark not defined.	0
3.7. Data Preprocessing.....	Error! Bookmark not defined.	2
3.7.1 Data Cleaning.....	Error! Bookmark not defined.	2
3.7.2 Data format conversion.....	Error! Bookmark not defined.	2
3.7.5. Transaction date	Error! Bookmark not defined.	5
CHAPTER FOUR	Error! Bookmark not defined.	7
RESULT AND DISCUSSION		377
4.2.1 Experimenting K-means Clustering.....		38
4.2.2 Experimenting Filtered Clustering.....		40
4.2.3 Experimenting Density Based Clustering		401
4.3 Evaluation using user acceptance testing.....		433
4.4. Discussion of result.....		433
CHAPTER FIVE		455
CONCLUSIONS AND RECOMMENDATION		455
5.1 Conclusion.....		455
5.2 Recommendations		456
Reference		477

List of Figures

Figure 3-1, Basic representation of KDD Process	Error! Bookmark not defined.	5
Figure 4-1, Result of K-Means Algorithm.....		39
Figure 4-2 , Filtered Cluster		40
Figure 4.3 Density Based cluster		422

List of Tables

Table 3-1 List of Attributes used in this study.....	Error! Bookmark not defined.	0
Table 3-2 Sample Customer Transaction Records.....		31
Table 3-3 CSV format table		34
Table 3-4 clients DOB	Error! Bookmark not defined.	
Table 3-5 Customer Gender	Error! Bookmark not defined.	4
Table 3-6 Customer Location	Error! Bookmark not defined.	4
Table 3-7 Customer Balance.....	Error! Bookmark not defined.	5
Table 3-8 Customer Date	Error! Bookmark not defined.	5
Table 3-11 Transaction Amount	Error! Bookmark not defined.	6
Table 4.1 Attribute Interface.....		36

Abstract

The global banking industry is facing more competition than ever before. Banks must attend to fulfill the wants and desires of their clients to maintain a competitive advantage in the market.

Segmenting customers is one of the best ways to interact with them. Because of applying of data mining clustering technologies, customer segmentation can assist banks in identifying more effective marketing tactics for the segments.

By treating the customer segments according to their transactional events (such as customer location, customer dob, customer gender, customer account balance, transaction date, and transaction amount), the unsupervised clustering technique is used in this thesis to segment Awash Bank customers to retain current customers, attract potential customers and improve customer service delivery processes.

The beging of clustering segments, the acquire data is clean and preprocess in an aggregated data style. WEKA knowledge discovery software is utilize for data mining using clustering algorithms, such as k-means, density-base, and filter clusters. Experimental result shows k-means, filtered, density-based clustering algorithms separate the algorithmic result into two clusters based on the given attributes, such as customer gender, transaction amount, and customer account balance. The finding of the study is putting the whole bank customers into two clusters base on the given attributari

This study, the performance is excute to apply customer segmentation for Awash Bank is an input for the bank customer prediction. Further study can create a customer prediction model that improve customer relationship management based on the clustering result.

Keywords: Awash Bank, Data Mining, Clustering Algorithms, Customer Segmentation

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

An individual or any corporate organization that purchases goods or services is referred to as a customer[1]. A bank that wishes to please its customer may define its services by their needs or wants [2]. The customer is often the most significant asset of any business organization [1]. A set of variables or characteristics used to assign customers to a homogeneous group refer to as customer segmentation[3].

Recognizing the value of customers, who are the best asset to any firm in any industry belongs or operates, represents a paradigm shift in how businesses are conducted [4]. By taking steps to determine who these customers are and how they might be classified, customer segmentation aims to improve relationships with them. When a bank looks for existing and potential customers with similar or identical traits, the idea of customer segmentation emerges[4]. The customer segmentation strategy depends on determining the crucial characteristics that can be used to divide customers into categories[5].

Different treatment of customer segments can lead to increased revenues [6]. So that result in a firm making a breakthrough in the market competition. Since the impact of different preferences on profitability is dwindling, establishing customer segmentation is the best course of action to achieve the Bank's objectives. Customer segmentation is typically accomplished through statistics-based methods, which compute several statistical measures from the customer data and then divide the customer population into several segments by utilizing clustering algorithms in the space of these statistics [6].

In the actual world, customer segmentation is very widespread. For instance, many commercial Banks distinguish between members and non-members among their clients. Additionally, many businesses offer several service levels for various customer classes. Customers, for instance, can be split up into a few classescatagories like individuals, partnership firms, Trusts, and Banks[7].

A bank's marketing expenses can be significantly reduced by the use of customer segmentation, which will also aid the bank in gaining a more noticeable and lucrative market share [7]. It enables businesses to create and implement different tactics wh to maximize the value of their clients [8]. The world we live in today is dynamic. Understanding the shifting environment,

comprehending the altered case, and comprehending the procedure is crucial for various banks. The arrangement of Awash Bank's attribute list model for customer happiness is the optimum method for client segmentation. Furthermore, because many banks operating in the financial sector nowadays, there is aggressive market competition. As a result, banks should put a lot of effort into improving customer happiness for each customer segmentation model, which is crucial to mining customer data for the bank's goal-achieving needs. However, for the Bank to compete with other fiercely competitive Banks, it is often required to establish the best completion plan. Many banks used to categorize their customers in this way. However, for automatic and comprehensive information in Awash Bank, client segmentation utilizing data mining and machine learning is crucial.

Data mining is searching and analyzing a large batch of raw data to identify patterns and extract useful information. The data mining algorithms categories classification, prediction, time-series analysis, association, clustering, dimensionality reduction [8]. All these tasks are either predictive data mining tasks or descriptive data mining tasks. Descriptive data mining is a technique that analyzes past data to provide the latest information on past events. So that so predictive data mining is a data mining technique uses to analyze past data and provide answers to future queries. The advantages of data mining are enhanced organizational efficiency, customer satisfaction, and quality decision-making. It's crucial because it enables users to examine the data in fresh ways or unearth trends they were unaware of.

1.2. Background of the Organization

After the socialist system overthrown on November 10, 1994[7]. Ethiopia's first private bank, Awash Bank, was founded. The Bank began operating as a bank on February 13, 1995. The Bank formed by 486 founding shareholders with a paid-up capital of Birr 24.2 million. Since the beginning operations, the Bank has had impressive expansion. Awash Bank has demonstrated exceptional operational and financial performances among Ethiopia's private banks eventhough the national and international obstacles. Awash Bank objective is to improve its capital structure, technological prowess, human capital, and client bank [8].

The Awash River, which is widely used in Ethiopia for small- to large-scale irrigation programs, hydroelectricity generation, and industrial operations, is the source of our name. The slogan "nurturing like the river" alludes to Awash Bank's significant contribution to the socioeconomic advancement of the nation. By promoting the practice of saving, offering credit options, and facilitating effective and quick payment systems, we also provide services to the populace[8].

Accessibility is a fundamental principle at our bank. By using several channels for service delivery, the bank constantly works to increase its accessibility. Due to the enormous branch networks and wide footprint, the bank is currently the most accessible private bank in the nation. Along with its branch networks, the bank also offers our customers the convenience of 24/7 service through Automatic teller machines, point-of-sale machines, internet, mobile, and agency banking[8].

Since its inception, Awash Bank has to place a high priority to corporate social responsibility. money investing of in the social and environmental well-being of the underprivileged strata. Moreover, their education and health, the Bank hopes to enhance the socioeconomic circumstances of the areas in which it operates. The number of elementary schools created in partnership with NGOs, the upgrades made to healthcare facilities, the planting of trees around the nation, and other initiatives show the positive influence of bank efforts.

Realizing the organization's goals and the targets outlined in its strategy is a typical indicator of the bank's performance. With the topic "Transforming AIB: Vision 2025," the bank has created a 10-year strategy roadmap for its performance. It has challenging financial and non-financial goals. However, based on its performance, the Bank emerged to be on track to meet those goals well in advance to 2025. In most case performance metrics during the past ten years, Awash Bank has had growth rates above the normal for the banking sector. In fact, among the nation's private banks, Awash Bank has experienced the quickest growth. We credit this outstanding accomplishment to our Board of Directors' innovative leadership, as well as to the hardworking management team, motivated employees, and devoted clients.

The aspirations of the vision and mission serve as the foundation for Awash Bank's strategic aims and objectives. On the one hand, the Awash bank's vision is 'being the best and most favored provider of financial services where clients are to be treated with respect'. On the other hand, the bank mission is 'to offer the most innovative and easily accessible banking services to its clients, utilizing the most current technology, and talented and driven employees to maximize stakeholder value and positively impact society'. As stated by the Awash bank, the core values of the bank are accessibility, wisdom, accountability, social responsibility, and honesty.

To achieve the bank vision, mission, and core values, the need to know its customers. The customer segmentation to improve its customer relationship management. Hence, the study attempts to apply customer segmentation of the bank using data mining technology.

1.3. Statement of the problem

Nowadays, customer segmentation is critical for different reasons, including the following: Customers must receive optimal services. These services must, however, be of higher quality [8][4]. Customer segmentation is often too necessary to increase service quality, since low-quality services are not good action for the banking industry's expansion [8].

A lack of specialize banking services, can cause customers to stop being loyal. Therefore, to improve customer happiness and increase customer loyalty, customer segmentation is crucial for the processes used to deliver banking services [8][1].

Profitability is the third factor driving our client segmentation strategy. The objective of many commercial institutes like banks are to increase profits. Segmenting bank customers is to retain customers and to profitability increasing of the bank[9].

The Bank's strategic goal is the fourth justification for customer segmentation. As a result, the strategic goal of the Bank has to accomplished. One element needed to achieve the bank's strategic goal has been customer segmentation[9].

However, banks can discern between the following due to a lack of customer segmentation. The first one is the Bank will be unable to separate traditional or regular customers from IFB or interest-free Banking. Considering that the conventional is an interest-bearing category, the categories of wholesale (premier or high-value customers) includes Banks, PLCs, share banks, and other entities. The application is for both the normal and the IFB. Moreover, most of them are ordinary retailers out of bank clients.

For Awash Bank, customers divide into loan, deposit, foreign currency, and turnover categories. Due to incorrect client handling caused by the Bank's lack of the a fore mentione segmentation procedure, performance issues that can be time and money-consuming, as well as customer complaints, may result.

Lack of customer segmentation results in reduce effectiveness and efficiency, which generates issues and shortcomings for Awash Bank. To gain a competitive edge in the banking industry market, a bank should categorize and act toward its customers in advance [9]. Customer service is challenging since every customer is the same, which results in time and expense overruns for the Bank. These are holes, issues, or gaps that need to be fill.

Another issue is the movement of too many customers to banks or other financial institutions to receive better and faster service from those institutions. Therefore, the Bank must implement strategic measures to address these issues. Customer segmentation techniques based on data mining are more advantageous in the following ways [10][5]: The objectivity of the data affects how segmentation based on data mining turns out. Therefore, this study uses data mining for Awash Bank customer segmentation.

1.3.1. Research Questions

The following research issues are therefore explore and address in the current study.

1. What are the data mining clustering algorithms suitable for customer segmentation?
2. Which clustering algorithm is efficient in bank customer segmentation?

1.4 Objective of the study

1.4.1 General Objective

The main objective of this study is to apply data mining clustering algorithms for constructing a descriptive model for Bank customer segmentation.

1.4.2 Specific Objectives

The following particular objectives are at accomplishing the overall goal of the study.

- To select suitable methodologies and approaches for the study after reviewing the literature.
- To gather and get ready data for experiments.
- To choose appropriate clustering methods for customer segmentation.
- To apply the chosen clustering algorithms and propose a descriptive model.
- To evaluate the effectiveness of the proposed model.

1.5 Scope of the Study

This study attempts to investigate and propose Awash Bank customer segmentation base on their withdrawal transactions. Some clients make transactions sometimes or frequently. Knowing these clients will assist the Bank in developing targeted marketing. The secondary source at Awash Bank uses to acquire the necessary data for the experiment. The information gathered

includes financial statements and other information (such as annual reports from the organizational website) about customers. Once the acquired data is well prepared, data mining clustering algorithms used to construct a descriptive model for customer segmentation.

1.6 Significance of the Study

The study will help the bank's clients by enhancing customer service and encouraging continue transactions. The study will increase customer happiness.

The research's main contribution is the ability to predict changes in the target market's bank performance and quickly respond with new products or events. Additionally, because target segments see that the bank values them, it encourages and makes it easier for customers to stay for very long periods.

It is the primary factor influencing the bank market's efforts to reduce expenses. The impact of the research is that it helps each segment's profit be optimized.

When the study comes up with clustering, the customer segmentation's adoption will save Awash Bank users time and effort. That is by providing special attention to clustered customers.

- The cost of the Awash Bank's working time is to maintain by distinguishing between cluster-oriented categories, which can decrease customer service time uses of customer segmentation.
- Based on these specific categories, the customer's personal information will be store and retrieve to quickly as possible.
- The study looks at the withdrawal of money from the bank by dividing two categorical groups or two clusters, since they are only two clusters that teach many customers to expect to be serve in one of the two categories in the banking system.
- It makes it easier to separate product users based on their rights and roles.

1.7 Methodology of the study

Research methodology is the specific procedures or techniques used to identify, select, process, and analyze information about a topic. It is needed because, in a research paper, the methodology allows the reader to evaluate a study's overall validity and reliability.

1.7.1. Research design

The study follows quantitative experimental research. The scientific method is in use in experimental quantitative research design. It sets protocols that enable the researcher to test within the laboratory and to methodically and objectively investigate causal links between variables [10]. Towards an extensive experiment, the study applies three steps: data collection and preparation, implementation using the selected tool, and evaluation to measure the performance of the models.

1.7.2 Method of data collection and preparation

To conduct data analysis is one part of taking data and gathering the require data from organizational databases, data warehouses, manuals, etc. In this study, the data for mining is collect from Awash Bank. Then can be prepare to come up with suitable data for analysis. Data preparation is a process that is transforming raw data into something that can be used for analysis and subsequent processing [10]. The stages of clearing, integrating, reducing, and transforming raw data, then giving to data mining algorithms is done by the tool, follow by data exploration and visualization, are crucial. Accordingly, the tool applies preprocessing techniques for preparing a quality data set for training.

1.7.3 Implementation Tools

The clustering, classifying, and identification of association rules are the algorithms types, the tasks that the data mining techniques are in use of them for designing and using. WEKA's knowledge discovery technology is in use for this investigation. Weka includes a selection of graphical user interfaces for quick access to these features, visualization tools and algorithms for data analysis, including predictive and descriptive modeling. In this study, different clustering algorithms are experimented with for descriptive modeling employing WEKA knowledge discovery software.

1.7.4 Evaluation methods

One of the most essential steps in any data mining process is evaluation. Evaluation methods are the steps in any data mining process of clustering results. The evaluation methods fulfill two functions: it predicts the future performance of the final model (or even whether it should be employed at all), and it is a crucial component of many learning techniques that aid in

identifying the model that best reflects the training data[11]. Out of the three evaluation methods in data mining (goal-based, process-base, and outcome-base evaluations), outcome-base evaluation is use in this study.

Outcome-based evaluation works base on determining what information the evaluation must provide, defining the data that needs to be collected, deciding on data collection methods, developing and pretesting data collection instruments, collecting data, processing data, analyzing the data to answer the evaluation questions and write an evaluation report [11].

1.7.5 Definition of key terms

Customer segmentation is the process of separating a customer base into groups of people share traits to marketing, such as age, gender, interests, and purchasing patterns [1].

Knowledge discovery in data (KDD), another name of data mining, is the process to uncovering patterns and other important information from big data sets [23].

Clustering techniques known as unsupervise learning process extracts references from dataset make up of input data without labeled responses [27].

1.7.6 Organization of the Study

There are five chapters in this essay. The purpose of the study, introducing the problem, scope, significance, and methodology of study are outline in the opening section. The concept, goal, importance, and many methods of customer segmentations are review and explore in the second chapter. Technology related to machine learning and how it pertains to client segmentation and an experiment analysis that displays the whole process of the experiment performed in this study are covere in Chapter Three. Chapter Four discusses detail concerning the experimental results with a discussion of the results. Finally, Chapter Five covers the Summary, conclusion, and recommendation part of the thesis.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview of Customer Segmentation

The technique of splitting the population of potential customers into different groups is known as customer segmentation. Customers of the same segment have similar traits that might aid a business in effectively targeting and marketing to them [3]. One of the most crucial marketing principles is segmentation. The capacity of the bank to cater to different types of clients varies greatly. Banks should categorize the market rather than try to compete in it as a whole. Businesses will determine which market they get service through customer segmentation [8].

The banking sector offers a wide range of services to a sizable number of clients worldwide of businesses experiencing market competition due to the rising number of banks, which has led some customers to migrate to competitors for several reasons. The marketing researchers have been developing models and methods for customer segmentation base on data mining and machine learning techniques. This review concentrates on studies pertinent to this investigation on customer segmentation [11].

The customer segmentation is so well-liked. Because it enables organizations to decide which new products and services to develop next, together with how to market and sell the current offers more successfully. Because of an improved comprehension of the requirements and preferences of customers of a given organization [11].

By segmenting a market into groups of customers that share are thought to have common wants, interests, and priorities, businesses can then create and implement strategies group of customers. Understanding traits share by their customer base, which can be the difference between an unused offering and one that resonates with customers, is how segmentation, a crucial technique, helps businesses serve low-income markets[12].

We like to emphasize that numerous other customer segmentation techniques exist in the literature but have not been discuss here because their applications are fundamentally different from those of our study. One of them, for instance, can divide customers into groups base on the order in which they make purchases online[13].

2.2 Customer segmentation models

The seven primary types of customer segmentation models are demographic, psychographic, behavioral, geographic, technographic, needs-base, and values-base segmentation. There are many additional types of segmentation, including numerous variations on the seven primary types [8]. Table 2.1 below provides a customer segmentation model with a description of tells how to segment customers of a given organization.

Demographic segmentation is a precise form of audience identification based on data points like age, gender, marital status, family size, income, education, race, occupation, nationality, religion.

Geographic segmentation is the practice of dividing organizational customers based on geographic location, from country right down to zip code. That is for products, services, or marketing messages at people who live in, work in, or shop at a particular location.

Psychographic segmentation is a market research method a market or customer group into segments based on their beliefs, values, and lifestyles.

Technographic segmentation is a marketing strategy that categorizes customers into groups based on their technology usage and preferences. This type of segmentation can provide valuable insights into the interests and needs of customers and can inform marketing and sales efforts.

Behavioral segmentation is the process of sorting and grouping customers based on the behaviors they exhibit. These behaviors include the types of products and content they consume and the cadence of their interactions with an app, website, or business.

Need-based segmentation is grouping people based on their shared needs instead of grouping them based on data like age or location. Once a common need is identify, the organization can learn more about why customers make buying decisions based on that need.

Value-based segmentation evaluates groups of customers in terms of the revenue they generate and the costs of establishing and maintaining relationships with them. It also helps Banks determine which segments are the most and least profitable so they can adjust their marketing budgets accordingly.

Segmentation Model	How to Segment Customers
Demographic Segmentation	Age, gender, income, education, marital status
Geographic Segmentation	Country, state, city, town
Psychographic Segmentation	Personality, attitude, values, and interest
Technographic Segmentation	Mobile use, desktop use, apps, and software
Behavioral Segmentation	Tendencies and frequent actions, feature or product use, habits
Needs-Based Segmentation	Product or service must-haves and needs of specific customer groups
Values-Based Segmentation	Economic value of specific customer groups on the business

Table 2.1. Customer segmentation model with description

2.3 Types of Customer Segmentation

Banks assess the effectiveness of customer segmentation solutions by verifying them using segmentation criteria because market conditions are changing the different types of segmentation strategies used in the banking industry.

One of the most common kinds of segmentation is by demographics. This phrase relates to categorizing clients based on observed differences in people. Some are age, sex, marriage status, the size of the family segmented client base, occupation, education level, income, race, and religion. The benefit of this type of segmentation is that it is both reasonably easy to obtain and reasonably price. The target segment that is most likely to purchase or use a Bank's products or services is usually develop using demographics in combination with other segmentation. But today, demographic factors are in client segmentation due to market globalization and banking sector improvements [17].

The most prevalent sort segmentation is behavioral, which focuses on the customer's purchasing patterns, including brand loyalty, level of involvement, and service usage. Customers' attitudes, viewpoints, hobbies, and other related factors are in behavioral segmentation, which is the most

successful [8]. Behavioral segmentation, there is a need to be aware of customers' actions. The customer engages with an organizational brand or to actions elsewhere. Effective customer segmentation can be get by concentrating on purchasing patterns and a few other behavioral markers. Among the necessary data are product ownership and usage statistics, which often kept up to date and available in the organization's databases. Base on the observed behavior and usage of the customers

Using behavioral segmentation, banks create groups of customers base on their familiarity with, attitude toward, use of, or responsiveness to services. Behavioral segmentation is frequently use to categorize customers into heavy, medium, and light users. Heavy users are a bank's most important customers, and they should always take them seriously. Marketing professionals commonly utilize strategies to increase both heavy and light users' product consumption. One of the best ways to employ behavioral segmentation is to divide markets based on various customer advantages [8].

Customers are divided up based on their locations in a process known as geographic segmentation of the customer base. A country, region, or city, particular properties in particular cities, might all be considere customer segments in this scenario. These types of customer segmentations are use by both large and small businesses.

Banks divide their client bank into geographic groups for reasons, including assessing the varying investment needs in various regions and accounting for regional differences in language, hobbies, customs, and other traits. Furthermore, the banking sector benefits greatly from it.

Regional banks target particular customer's area by using geographic segmentation. It helps small businesses set up their budgets, curb unnecessary spending, and avoid having their products or services misunderstood by customers.

The benefit of geographic segmentation is that each bank can easily be categorized into a geographic area. As a result, the targeted geographic groups for geographically divided banks are straightforward. The biggest drawback is that because people reside in the same nation or area, it does not indicate that they have identical characteristics crucial to businesses [8].

The term "psychographic segmentation" refers to customers into groups base on more cerebral and emotional, such as personality traits, passions, worldviews, attitudes, and lifestyles, habits, interests, and social standing [8]. By fusing psychological research with demographic data, the field of psychographics aims to understand customer behavior. Psychographic segmentation

divides customers into groups according to their personalities, lifestyles, or values. Individuals may exhibit varying psychographic traits even when they belong to the same demographic group [8]. Categorize customers based on psychographic trends, banks must perform more in-depth analysis.

2.4.1 Benefits of Customer Segmentation

There are many benefits to segmentation. Fundamentally, segmentation makes businesses consider where they are now and where they want to be in the future. Banks are on the process of assessing their competitive advantages and make an effort to comprehend customer needs. In addition to examine and revise their opinions, segmentation gives significant new perspectives and insights [15].

Segmentation is one of the tools businesses may use to better utilizing their marketing budgets, gain an edge over rivals, and, most importantly, develop a deeper understanding of their target audience's wants and desires [16].

Below are some advantages of customer segmentation [16].

Market Concentration tactic: its competitiveness in the market since the category at which right segment is targeted with the right product, producing the best return on investment. Once specific market segments have defined, it is simpler for a Bank to concentrate on the groups they wish to target services and marketing. When Banks give attention to segments, The bank increases the keeping of customers: Through customer segmentation, it is possible to determine which customers need extra care, which frequently churn, and which ones have the potential value. Additionally, it can help in focused activities that draw customers' attention and provide positive, valuable interactions.

Superiority over rivals: a combination of market segmentation strategies may help to protect income due to the intense rivalry in the banking sector. In focusing geographical segmentation and behavioral segmentation, for instance, businesses can gain insights into customers behavioral habits across different geographic regions.

Customer Satisfaction: is with well-defined categories, businesses may cater to a range of customer needs by providing various bundles and incentives. Different forms of promotional and advertising activities will be employed, depending on needs, desires, and trait of each group .

Customer-specific product Customers are more engaged and satisfied when offered and communicate with in a more personalized way.

The more effective distribution tactics Knowing when and where customers purchase can help distributors make decisions about the kind of goods to stock in various places.

Because the objective is to increase the value of each customer, a bank must be aware of how every specific marketing effort will affect the customer in advance.

2.5 Data mining methods for identifying customers

To discover previously undiscovered, reliable patterns and links in enormous data sets. Data mining is using complicated analysis methods. These instruments may include mathematical algorithms like neural networks or decision trees, machine learning methods, and statistical models. There are both analysis and prediction in data mining. Data mining specialists have dedicated their lives to better understanding how to process and draw conclusions from the enormous amount of data, relying on techniques and technologies from the convergence of machine learning, database administration, and statistics. Association, classification, clustering, prediction, sequential patterns, and regression are just a few of the data mining techniques created and put to use[21].

In customer segmentation, customer clustering analysis uses a mathematical model to discover groups of similar customers based on finding the smallest of variation among customers within each group.

The clustering algorithm is a technique that assists customer segmentation process of classifying similar customers into the same segment. The clustering algorithm is helping for better understanding of the customers. That is in terms of both static demographics and dynamic behaviors. Customers with comparable characteristics often interact with the business similarly in the same way a business can benefit from this technique by creating tailored marketing strategies for each segment.

The clustering algorithm is an unsupervised machine learning algorithm that discovers groups of data points that are closely related. The fundamental difference between supervised and unsupervised algorithms is that:

- supervised algorithms : it requires partitioning the dataset into train and test sets, and the algorithms learn based on the output/label of the train dataset and generalized it to unobserved data. For instance, decision trees, regression, and neural networks.
- unsupervised algorithm: it is discover hidden patterns when there is define output/label from the dataset. For instance, clustering, association rule mining, and dimension reduction.

Clustering methods. These methods allow grouping the data sets into clusters, such that the data sets within one cluster are more related to each other than the data sets in different clusters.

The term clustering is use to refer to the act of grouping information, which can be data from customers, products, and computers, among others. Separating into groups, categorizing, and segmenting is a way of gathering information or data based on common characteristics [21].

2.6. Related works

The purpose of the literature review is to learn more about the subject matter, identify the kinds of studies relevant to the one under consideration, and learn about the techniques and algorithms that in previous studies.

Anta m. sundjaja. The banking industries fierce competition has made it harder to raise loanable capital as customers needs become more complicated. In this research, unstructured interviews with the management team and bi-department at the head office of a large private bank are use to define customer needs, customer profiles, preferred transactions, and channels from each segment using data mining techniques.

Data warehouses in Indonesia provided the data, which Microsoft Excel uses to evaluate.

This research uses demographic segmentation to find insight into the products and services already provide to current clients, even though more sophisticated segmentation approaches available. To enhance the development of product and service market oriented promotion results are further to be evaluated [21].

YongpingLiu, a great change in banking services is because of the application of data mining. Specifically in retail banking case analysis China merchant bank takes just as an example, data environment evaluation analysis, operation efficiency, and profitability as a model the application for performance of china merchant bank deeply seen. Eventually, we give for pieces of advice on future development to China merchant banks [21].

Jared O. Onyuna, Studies of consumers will need to concentrate not only on understanding product choice and the reasons behind channel choice but also on understanding the frequency, currency, type of transaction, and recency of the customers' transactions, give the rapid increase in mobile phone access and the growing prevalence of multi-channeling. The accurate patterns for segmentation and profiling can be obtain by combining the frat version of the rim with data mining analysis and demographic characteristics (gender, age, and location).

Businesses can use customer lifetime value on three factors: customer churn, potential value, and present worth of the customer. Cross-selling opportunities for existing customers are the main focus of potential customer value. As a result, cross-selling models are base on the entire database of uninteresting customers. To combat this, we introduced a framework that computes the customers past, present, and churn likelihood. We then divide the customers into groups according to these components, and categorize them according to their lifetime value and demographic characteristics.

While various researchers have propose techniques, such as CRM, customer lifetime value, regency frequency and monetary, and few studies have explored the inculcating of different qualities in the models.

In this work, we use a case study on a banking database to describe customer behavior base on factors relate to their lifetime value and demographics. To determine the efficiency of and combining of customer demographic data with variables for mobile banking customer segmentation and profiling, a descriptive study is presented in the research proposal [22].

Jianwen Ye¹, Jianhua Zhang¹, *, and Dapeng Dong¹. Customer relationship management is currently the center of attention for customer-centric management in the commercial banking industrial centers. Success will go to the person who can successful in classify, mine and process commercial bank customer information using data mining technology. The next step is to segment your consumer base so that you can promptly make enhancements and adjustments to meet changing market conditions and pursue new business opportunities. This study examines the function of data mining and customer relationship management and their recent advancements. Customers can access a wealth of information in the k-means data mining algorithm analysis and other customer information summary analyses. Subdivisions' objective [23].

CHAPTER THREE

APPROACHES AND TECHNIQUES

3.1 Overview

In this chapter, the clustering algorithms used are Fitere cluster, Densitry-base cluster, and the k-means cluster methods.

3.2. Data Mining in Awash Bank

The banking sector generates a lot of data due to the high volume of transactions. Thus, by enhancing services and providing essential customer information, data mining tools in business benefits both the Bank and the customers [22][23].

In the customer databases of the Awash Bank there is big data. The datasets are collect and take, and data mining techniques are apply in identifying hidden patterns and unanticipate trends. To better understand the issue, data mining is the that evaluating and methodically extracting significant and previously unrecognized links and associations between data. Data mining employs tried-and-true methods for examining recent and old data to forecast new patterns [23]. Solving issues entails analyzing data that has already in databases.

Data mining techniques can be use by banks to strengthen their position as market leaders and raise customer value. The primary in data mining include genetic algorithms, decision trees, association rules, classification, clustering, rule induction, and neural networks [21][23].

3.3 Data Mining and Knowledge Discovery in Database

The idea of analyzing enormous amounts of data is apply, and pulling out pertinent information results in the knowledge discovery process, which pulls out patterns, rules, and models from unstructured data and clarifies patterns that have found. KDD is an analytical method for 'knowing' or extracting 'knowledge' from that data . Data mining is sometimes used interchangeably with the concept of knowledge discovery from data by academics, but it is also sometimes seen as an essential step in the process [21][23].

According to Jiawei, Kamber, and Pei[23], this list contains the key actions that make up the data discovery process (see Figure 3.1).

- **Data Cleaning:** To increase a data set's dependability and effectiveness, it involves looking for missing data and eliminating noisy, irregular, redundant, and low-quality data.
- **Data Integration:** Various data sources can be combine and arrang into sets according to their usefulness, importance, and quality. A preprocessing step that produces data in a database includes cleaning and integrating the data.
- **Data Selection:** Finding and retrieving usable data from a database that is pertinent to the analysis activity is what it entails.
- **Data Transformation:** The data are prepare for entry into the data mining algorithms in this step. As a result, summary analysis and aggregator activity are use to modify and combine the data into mining-ready formats.

.Data Mining Is a process of algorithm usage to draw out data patterns from the modify data is a crucial step.

- **Pattern Evaluation:** It involves finding intriguing patterns that indicate knowledge base on patterns that are valid, understandable, maybe valuable, and novel, and it supports a hypothesis that the user was trying to verify.
- **Knowledge Representation:** The knowledge obtaine from the pattern assessment stage must now be apply to a particular domain utilizing visualization and knowledge representation techniques and formats, such as trees, tables, graphs, reports, and matrices. That is the last step in the knowledge discovery process.

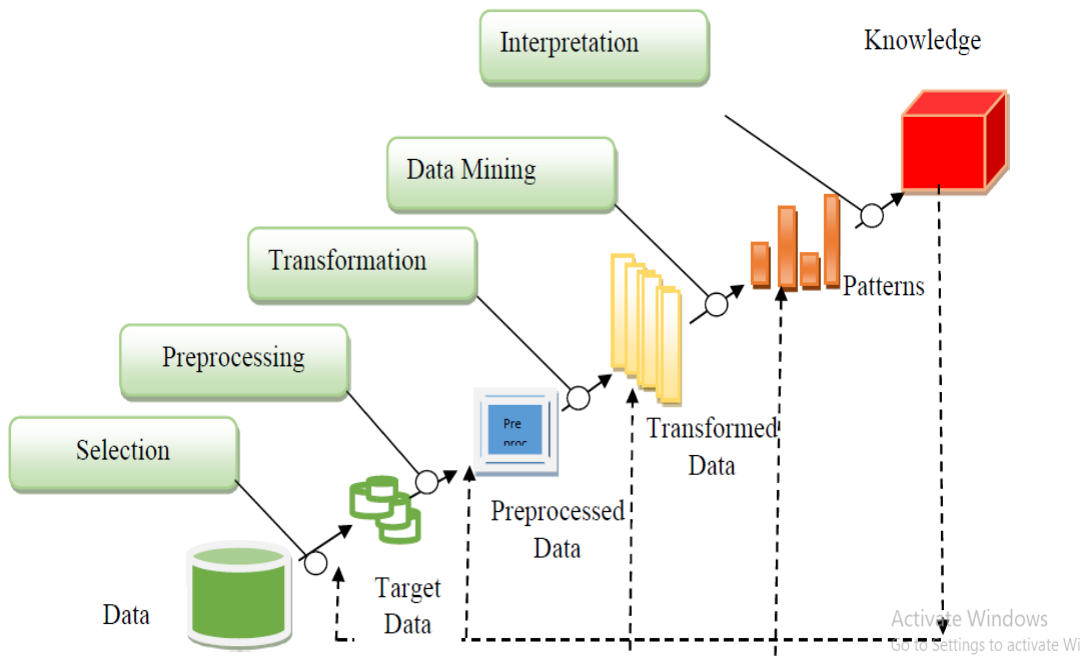


Fig 3-1, Basic representation of KDD Process[23].

3.4. Learning Approaches

The Unsupervised learning and the supervised learning approaches are the two main learning strategies used in machine learning. The difference among these two learning approaches is discussed as follows.

3.4.1 Supervised learning

Using labeled datasets, supervised learning trains algorithms that can accurately categorize data or forecast events. The model may evaluate its accuracy and develop over time by using inputs and outputs that have labels on them. When both the input and output values are known, this technique is then applied [24]. For each incoming data sample, classification is uncertain, supervised learning aims to enable machine learning algorithms to anticipate the output class.

Supervised learning is utilized in data mining to address classification and regression challenges. The classification task entails guessing a class label. To accurately identify test data, many techniques are employed. It locates specific things in the dataset that attempts are making to draw judgments about how to label or characterize them.

Regression uses an algorithm to forecast a real value for each item in a numerical label to examine the relationship between dependent and independent variables. Instead of determining whether

these two values are equal or unequal, the regression error is determined by the size of the discrepancy between the real-value predict label and the true or accurate one.

3.4.2 Unsupervised learning

Unsupervised learning is a type of machine learning where models are not labeled data and then allowed to act on it unmonitored. Only unlabeled training data are provided to the algorithm, which uses them to uncover patterns and insights hidden in the data before producing predictions for all concealed points. Without any outside assistance, the algorithms can discover hidden patterns in data. Given that labeled data are typically not available, it uses objectively assess an algorithm's effectiveness.

Clustering and association problems are subcategories of unsupervised learning. During clustering, unlabeled data are grouped into homogeneous groups using the data mining approach of clustering, which looks for patterns of similarity or difference. Analysis of big data sets commonly uses clustering.

Association helps to identify relationships between variables in a sizable database. It is the one that an unsupervised learning technique called an association rule is used. It indicates the collection of related data set components. [25].

A clustering unsupervised learning strategy is taken into consideration for this thesis study. We like to utilize the Weka knowledge discovery tool because it is straightforward, has an easy-to-use interface, and offers several capabilities for data preparation, classification, regression, clustering, association rules mining, and visualization [26].

3.5 Clustering method

An unsupervised learning approach used in machine learning is clustering. It is a strategy for organizing a collection of data objects into clusters with qualities that are similar across objects within a cluster but very different from those in other clusters. The conclusions are reached using unsupervised learning techniques on data sets without labeled output variables. This method of partitioning the data is in use. Clustering can result in the identification of data patterns that were previously undetected [21]. We can examine collections of multivariate data using this exploratory data analysis.

To characterize the distribution of the data clustering algorithm is typically important. Clustering algorithm is a technique for finding instances of related objects in a data set. The four most

popular clustering methods are filtered cluster, K-means, hierarchical-based, and density-based [27] [28].

3.5.1. Filtered Clustering Methods

The filtered clustering method helps to select clusters and records so that one can evaluate whether all records in a cluster are homogeneous and represent the same real-world cases, and that no other records are missing from the cluster.

The filtering process follows three steps. First, there is a need to perform several clustering trials. It is followed by scoring each trial and finally choosing the trial with the best score. Figure 3.2 below presents a framework for similarity and filtered algorithms.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right] \left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

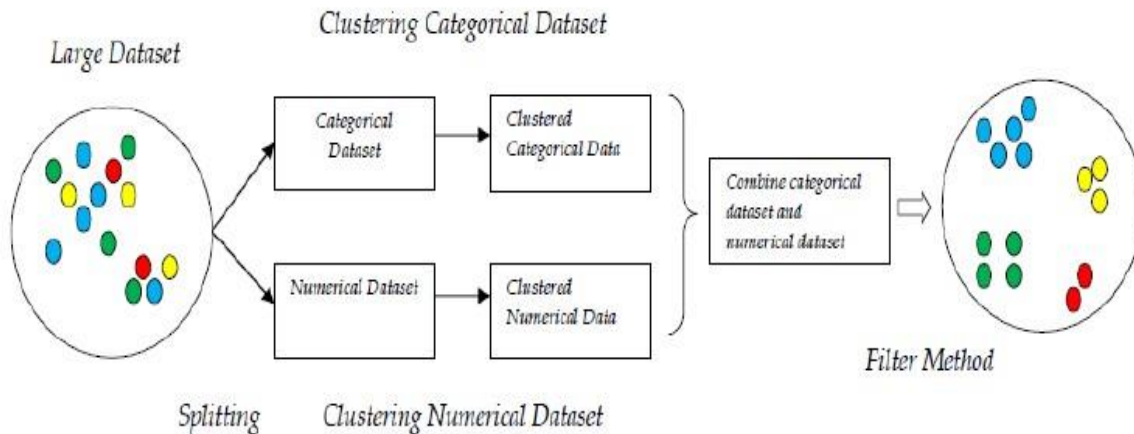


Figure 3.2. Overview of Similarity and Filtered Algorithm Framework

3.5.2 Density-based Clustering Method

The density-base Clustering method is one of the clustering methods base on density (local cluster criterion), such as density-connect points. Clusters are segregate from each other by low-density regions according to techniques for density-based clustering. The number of each object shares in a radius determines the density value it is assigned to it by this algorithm. According to the definition, a dense item is initially create as a clustering density. That density is large in size than a specific threshold. A merger occurs when two clusters share a denser neighbor. Following are several density-based clustering algorithms: DBSCAN, OPTICS, DENCLUE, and DBCLASD [27].

Out of the four types of density-based algorithms, the selected one is DENCLUE. DENCLUE is one of the density-based clustering algorithms that works effectively in a dataset. It uses grid cells but only keeps information about grid cells that contain data points and manages these cells in a tree-based access structure.

Influence function This describes the impact of a data point within its neighborhood. The Overall density of the data space is to be in calculation in the sum of the influence function of all data points.

The Clusters are determined mathematically by identifying density attractors. The Density attractors are the local maxima of the overall density function.

3.5.3 K-means Algorithm

Numerous data mining methods are utilize in the banking sector. That is to process the enormous volume of data are produce in their network databases. However, the most effective and popular data processing approach to cluster data mining tasks is k-means clustering [22].

By creating clusters iteratively, it utilizes to tackle clustering difficulties. At the beginning of the K-means algorithm, we must provide the number of Clusters. Next, specify the locations at random as the cluster centers, and for each point, get the distance between the centroid and the data point. Based on this distance, each instance is locate to the cluster center that is closest to it. The centroid serves as the reference point for measuring the separation between each data point. There are various ways to determine the distance, including the Manhattan distance computation, Calculations of Euclidean distance, cosine similarity, and so forth. The K centroids are once more calculate following the placement of all the data points. The centroid of a cluster is the average value of all the points in that cluster. Then, all data points are redistribute to clusters concerning the new centroids using the centroid-data point distance. The process iterates until the samie coordinates are repeatedly allocate to each cluster; at this time, the cluster centers have stabilized and will not change [28] [29].

According to Sanse and Sharma[28], the k-means clustering algorithm is describe as follows:

1. Set K, the size of clusters.
2. Select a similar number of data points to the centroids.
3. Find the separation between each centroid and the given location. The cluster with the least centroid-point distance should receive the point.
4. To determine the new centroid of a cluster, compute the mean of all the points in the cluster after completing step 3 for each points.
5. Till the centroids remain unchanged, keep iterating starting step 2.

3.6. Collection and Preparation of Data.

In this study, to conduct a data mining experiment to choose the kind of data that would be useful for the experiment and preparing a quality dataset is the first stage before constructing a descriptive model using clustering algorithms.

3.6.1 Dataset collection

The process of methodological acquiring pertinent data for a given objective is data collection. We issued a formal data request to the appropriate section of the Bank by the management agreement, whereby we agreed that the data would not be use for any other purpose than research. We started collecting data after receiving the security section's consent.

The information about customers is store in the information technology system databases. The primary data obtained from the Awash bank contains the customer's name, customers' account number, customer transaction, customer transaction date, customer balance, customer location, and others. The following table shows a list of attributes with their description that uses to collect the data required for the experiment.

No	Attributes	Data type	Description
1	Customer DOB	number	Customer date of birth
3	Customer Gender	text	Customer Gender
3	Customer location	text	Customer location
4	Customer account balance	currency	Customer account balance
5	Transaction date	number	The date of transaction
6	Transaction amount	currency	The amount of transaction

Table 3.1: List of attributes used in this study

3.6.2 Customer Details Sample Record

Any bank's computerize record of transactions transfer to its database for later use. A set of attributes allows for the networking of the data stored in the database with a particular data set[31].

Customer DOB	Customer Gender	Customer Location	Customer Account Balance	Transaction Date	Transaction Amount
10/1/1994	F	North A.A R.	17819.05	1/1/2014	9125
26/11/96	F	Awelia	17874.44	1/1/2014	7459
14/9/73	F	South A.A R.	866503.21	1/1/2014	2060
24/3/88	F	Ameen	6714.43	1/1/2014	1762.5
8/10/1972	F	Ihsan	53609.2	1/1/2014	7676
26/1/92	F	West A.A R.	973.46	1/1/2014	2566
19/4/88	F	Amal	14906.96	1/1/2014	1833
22/6/84	M	Khidma	4279.22	1/1/2014	289.11
22/7/82	M	Manal	48429.49	1/1/2014	3259
13/6/78	M	Yusara	32274.78	1/1/2014	12300
24/3/78	M	Kuaser	10100.84	1/1/2014	6338
24/6/85	M	Berkah	10643.5	1/1/2014	2945
1/10/1986	M	Ashura	12868.42	1/1/2014	1291
17/5/91	F	Rayan	2951.1	1/1/2014	1892
24/2/93	M	Barwako	3297.74	1/1/2014	3242
21/1/67	F	Meshreq	298461.05	1/1/2014	4500
26/6/91	F	Nahar	6384.54	2/1/2014	5610
25/2/85	F	Dara	57791.69	2/1/2014	3465
25/7/79	F	Billal	123781.9	2/1/2014	6800
23/4/77	M	Taajir	39.73	2/1/2014	9110
25/10/84	F	Afelah	213.06	2/1/2014	1081
30/12/97	F	Akrem	44627.73	2/1/2014	2999
25/6/85	M	Mabrook	2477.42	2/1/2014	1683
13/4/83	M	Kebir	58406.7	2/1/2014	8750

Table 3-2 Sample Customer Transaction records

3.7. Data Preprocessing

Data preprocessing is a phase in the data mining and data analysis process that takes raw data and turns it into a format that machine learning tools and algorithms can understand and evaluate.

Before finding patterns and learning new information, raw data needs to be cleaned and sorted. Every data mining study depends on the data preparation procedure to be successful and register good performance [21].

The main tasks of data preprocessing are feature selection, data cleansing, integration, reduction, and transformation.[21]. In this study, data cleaning is considered to improve the quality of the data set.

3.7.1 Data Cleaning

There could be errors and abnormalities in the data when a large amount of data from many sources involve. The primary goal of this stage is to clean the data by removing outliers, restoring missing numbers, minimizing noisy data, and adjusting inconsistent data. Data accuracy and relevance are ensured by the cleaning process, which also helps to remove duplications, prevent inconsistencies, and avoid them [31].

There were numerous problems with the customer information and detail records that are gathered. Inaccurate data was recorded in fields. Since cleaning is one of the preprocessing of WEKA has cleaned inaccurately. Then, data has been processed.

3.7.2 Data format conversion

The Excel file contains data from a year's worth of Bank customer transactions on withdrawal of customer information in large formats and data kinds. To alter the data, convert it to CSV format. The Excel program was used to perform the first data preparation.

The data set file is aggregated and prepared in numerical form once the data have been cleaned to get it ready for constructing a descriptive model using clustering algorithms. Here, under sample data in CSV format is present.

Table 3-3 CSV format

Customer dob	Customer Gender	Customer Location	Customer Account Balance	Transaction Date	Transaction Amount
10/1/1994	F	North A.A R.	17819.05	1/1/2014	9125
26/11/96	F	Awelia	17874.44	1/1/2014	7459
14/9/73	F	South A.A R.	866503.21	1/1/2014	2060
24/3/88	F	Ameen	6714.43	1/1/2014	1762.5
8/10/1972	F	Ihsan	53609.2	1/1/2014	7676
26/1/92	F	West A.A R.	973.46	1/1/2014	2566
19/4/88	F	Amal	14906.96	1/1/2014	1833
22/6/84	M	Khidma	4279.22	1/1/2014	289.11

3.8. Statistical analysis of the data set

In this section statistical analysis of the data set is present. The data visualization is present for data of birth, customer gender, customer location, Customer balance, Transaction date, and Transaction amount. The analysis from WEKA, and table samples from the data set in number amount of ten from each attribute.

3.8.1. Date of birth

Table 3-4 customers' date of birth, Below is a summary of banks' customers' date of birth which is take from customers' age attribute on the WEKA interface. The table is a sample. Most of the customers are about the age of thirty-six.

Customer date of birth(dob)	
1	26/11/96
2	14/9/73
3	24/3/88
4	26/1/92
5	27/1/82
6	19/4/88
7	22/6/84
8	22/7/82
9	13/6/78
10	24/3/78

Table 3-4 customers date of birth

3.7.2. Customer gender

Figure 3-7 Depicts a summary of bank Customer Gender which is take from customers' Gender attributes on the WEKA interface. The table is a sample. The gender of most of the customers are male.

Customer Gender	
1	F
2	F
3	F
4	F
5	M
6	F
7	M
8	M
9	M
10	M

Table 3-5 Customer Gender

3.7.3. Customer location

Table 3-8Tells Customer Location which is take from customers' Location Attribute on the WEKA interface. The table is a sample. Most of the customers are from A.A.R

Customer Location	
1	Awelia
2	South A.A R.
3	Ameen
4	West A.A R.
5	Hayahu
6	Amal
7	Khidma

8	Manal
9	Yusara
10	Kuaser

Table 3-6 Customer Locations

3.7.4. Customer balance

Table 3-7 Tells Customer Balance which is taken from customers' Balance Attribute on the WEKA interface. The table is as a sample, Most of the customers are with a balance of around one hundred eleven thousand seven hundred seventy-seven point eight nine.

Customer Account Balance	
1	17874.44
2	866503.21
3	6714.43
4	973.46
5	95075.54
6	14906.96
7	4279.22
8	48429.49
9	32274.78
10	10100.84

Table 3-7 Customer Balance

3.7.5. Transaction date

Table 3-8 Tells the transaction date which is take from the transaction date attribute on the WEKA interface. The table is a sample. Most transactions initiate by customers are on the date 2/1/2014.

Transaction Date	
1	1/1/2014
2	1/1/2014
3	1/1/2014
4	1/1/2014
5	1/1/2014
6	1/1/2014
7	1/1/2014
8	1/1/2014
9	1/1/2014
10	1/1/2014

Table 3-8 Transaction Date

3.7.5. Transaction amount

Table 3-9 Tells the Transaction Amount which is take from the transaction amount attribute on the WEKA interface. Most transaction amounts of the customer are around one thousand six hundred three point nine eight.

Transaction Amount	
1	459
2	2060
3	1762.5
4	566
5	148
6	833
7	289.11
8	259
9	12300
10	338

Table 3-9 Transaction Amount

CHAPTER FOUR

RESULT AND DISCUSSION

The study's results are presented in this chapter along with a discussion of how they relate to the study's objective. The clustering outcome of the aggregate customer transactions and customer status data by k-means cluster, filtered cluster, and density based cluster are describe in the first sections. Moreover, all the dataset is input to the training data. Then, It follows by a discussion of the experimental results.

4.1The dataset

Attributes use in this experiment for customer segmentation using a machine learning algorithm are customers and transaction-relate data (see Table 4.1). The effectiveness of the suggest model is base on the execute customer transactions event. That data gathers customer transaction from the customers' relationship management and IT expert of Awash Bank Database. A form of table is also shown below.

1	Customer date of birth
2	Customer gender
3	Customer location
4	Customer account balance
5	Transaction date
6	Transaction amount

Table 4.1 List of Attributes

A total of 511936 customer data with eight attributes is used for running clustering algorithms for customer segmentation. These clustering algorithms include k-means clustering, density-based clustering, and Filtered clusters.

4.2.Experimental result

Base on the attributes choice from the customer transactions and customer status data, that describes in detail in Chapter 3, the experimentation is conduct. The outcome is groupe into two categories base on the optimal clustering result. These are after each group's data on a year's worth of customer transactions is examine.

4.2.1. Experimenting k-means clustering algorithm

The total of 511, 943 sample customers chosen for the experiment are divid into two groups by the k-means clustering method, as seen in Figure 4-1. The attributes customer dob, customer gender, customer location, customer account balance, transaction date, and transaction amount. Base on the class-to-clusters evaluation, the correctly cluster instances for k-means take a time of 2.4 seconds to fully train data and construct the descriptive model.

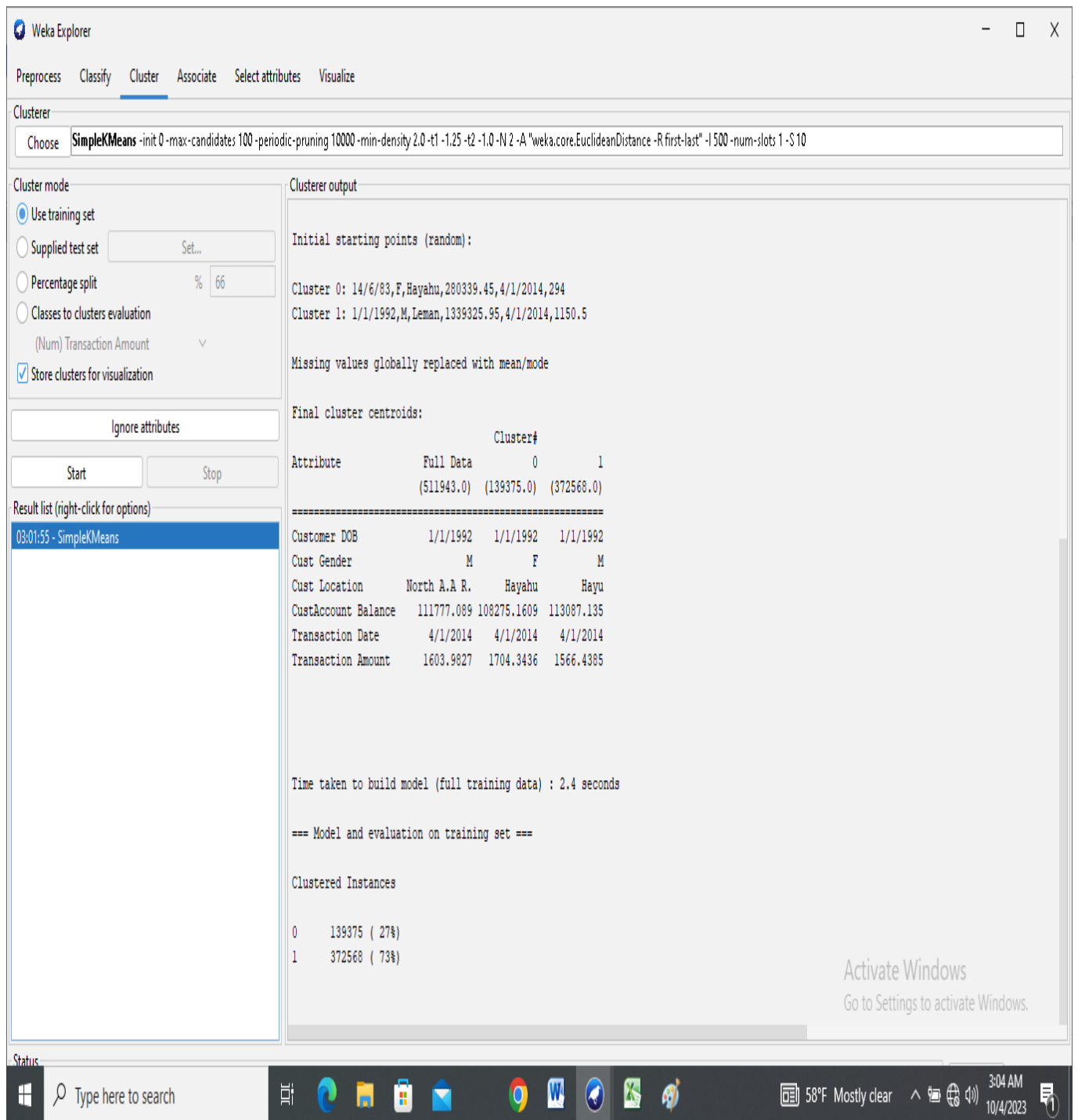


Figure 4-1, Result of K-Means Algorithm

As shown in Figure 4-1, among the two cultures (cluster 0 and cluster1) These two clusters are the bank clustering models that represent the bank’s customers. In k-means clustering algorithm groups, most of the transactions are accounts for the second group. This means Cluster One customers account for 27% of the results and Cluster Two customers account for 73% of the results. Base on Customer Gender twenty-seven percent of the customers are females and seventy-three percent of the customers are males, base on Customer Balance

cluster one is 108275.1609, and cluster two is 113087.135. Based on Transaction Amount cluster one is 1704.36 and cluster two is 1566.4385.

4.2.2 Experimenting Filtered Cluster

The total of 511, 943 sample customers chosen for the experiment are divided into two groups by the Filtered clustering method, as seen in Figure 4-2. The attributes are customer dob, customer gender, customer location, customer account balance, transaction date, and transaction amount. Based on the class clusters evaluation the correctly clustered instances for Filtered cluster take a time of 2.92 seconds to fully train data and construct the descriptive model.

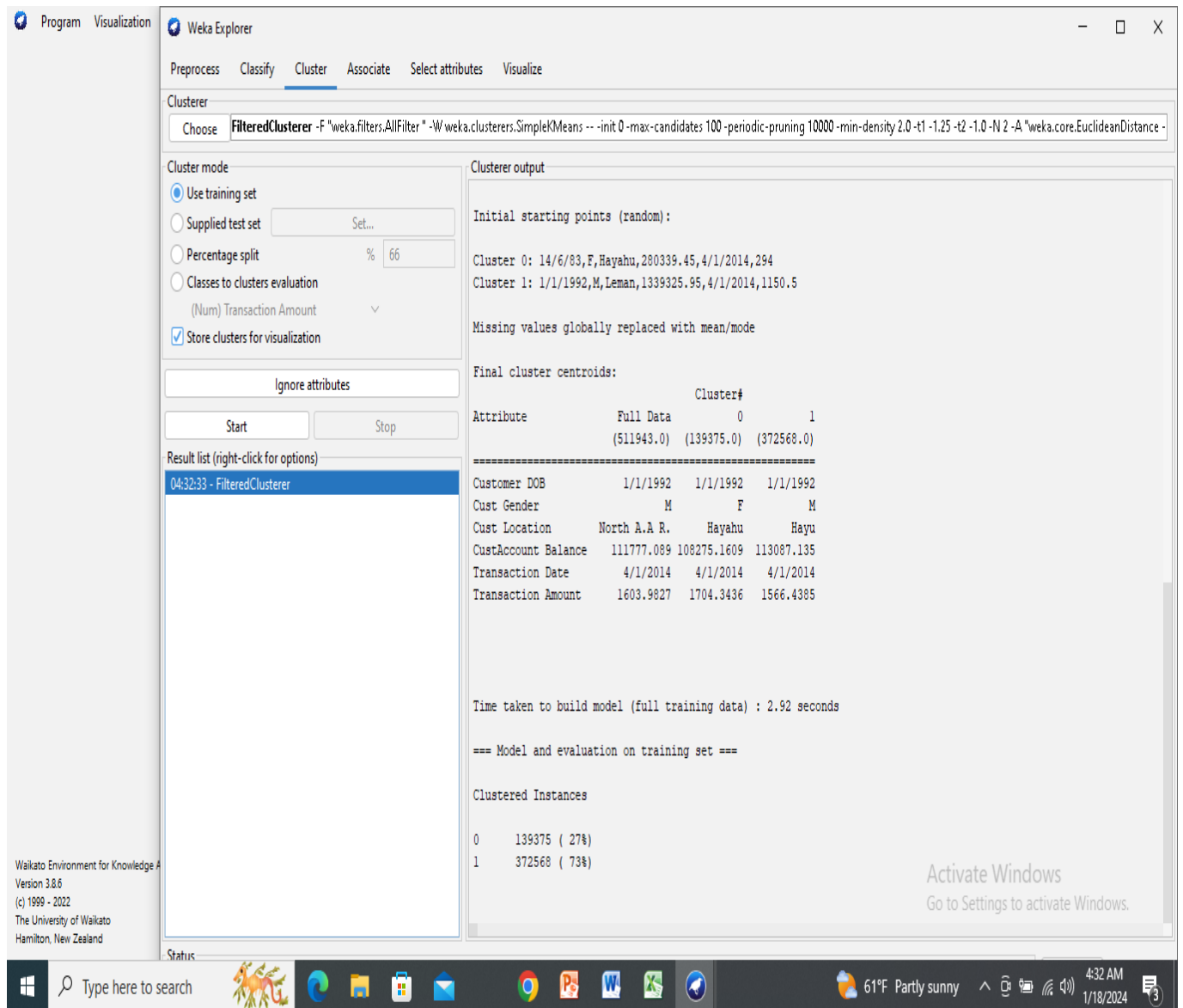


Figure 4-2 Filtered Cluster

As shown in Figure 4-2, among the two cultures (cluster 0 and cluster1) These two clusters are the bank clustering models that represent the bank's customers. In Filtered clustering algorithm

groups, most of the transactions are accounts for the second group. This means Cluster One customers account for 27% of the results and Cluster

Two customers account for 73% of the result. Based on Customer Gender twenty-seven percent of the customers are females and seventy-three percent of the customers are males, based on Customer Balance cluster one is 108275.16, and cluster two is 113087.14 Based on Transaction Amount cluster one is 1704.34, and cluster two is 1566.44.

4.2.3 Experimenting Density Based Clustering

The total of 511, 943 sample customers chosen for the experiment are divided into two groups by the Density-based clustering method, as seen in Figure 4-3. The attributes are customer dob, customer gender, customer location, customer account balance, transaction date, and transaction amount. Based on the class to clusters evaluation, the correctly clustered instances for the Density-based cluster take time of 2.76 seconds to fully train data and construct the descriptive model.

Provide that the algorithm outcome out of which 27% of the customers are grouped into cluster one, and 73% are grouped into cluster two. This indicates that a majority of the customer's results fall into cluster two which is similar to the clustering result of k-means and filtere cluster, however, it is taking more time than k-means clustering.

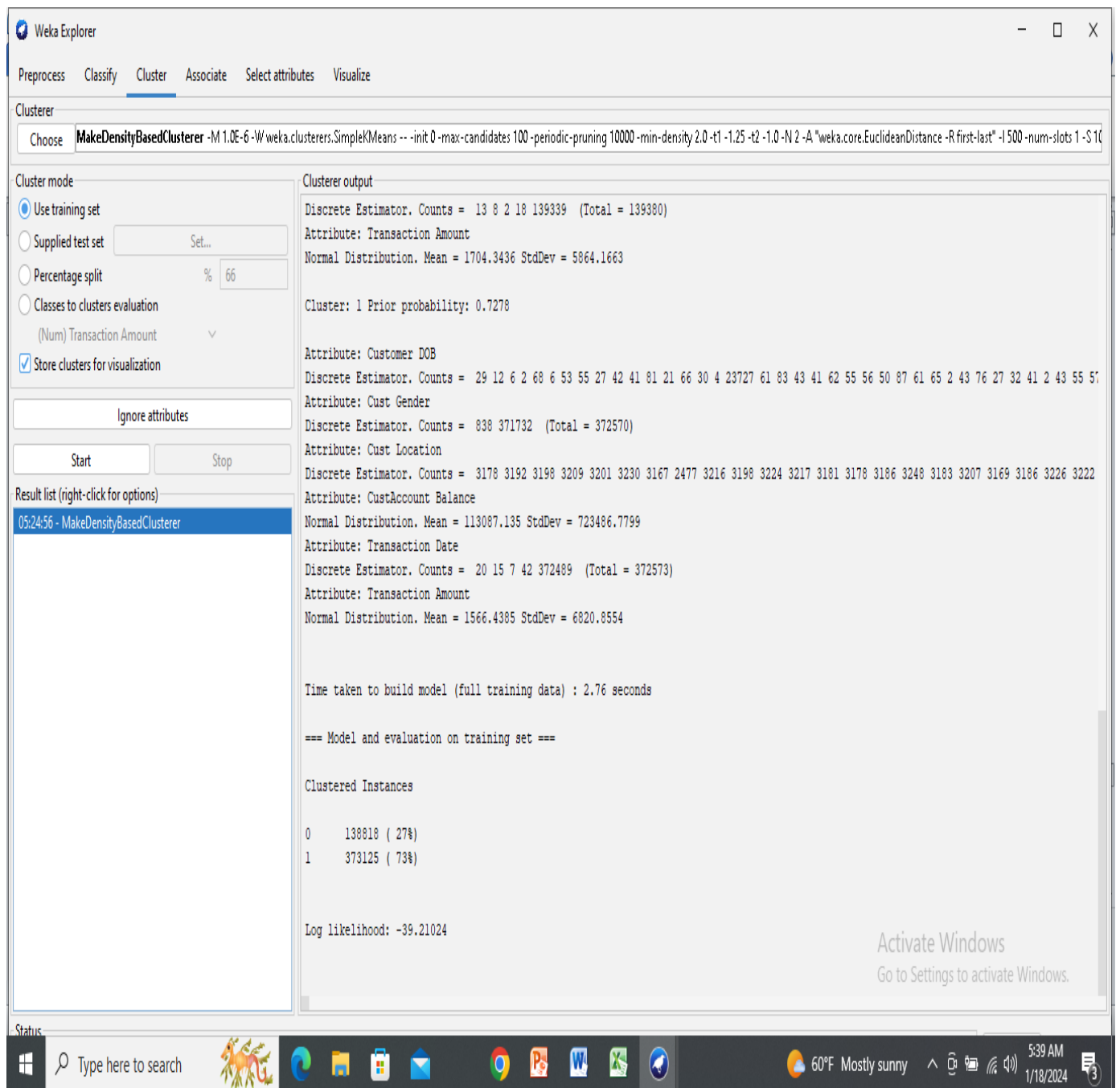


Figure 4.3 Density Based cluster

4.3. Evaluation using user acceptance testing

Based on the professional selection and comment clustering result is relevant to managing customers with due care service. It is also appropriate for customer segmentation to serve those customers with the best speed and to dedicate who does for the customers. Applicability of the result is highly best fit for customer retention and for special treatment. Still, the clustering is a reason for cross-selling or a customer that calls other customers. Finally, based on the user acceptance testing the method is effective, efficient, easy to use, and easy to remember.

4.4. Discussion of result

Understanding various customer behaviors and identifying which customer groups are significant to the Bank needs the use of customer segmentation. However, the size of the groupings and the number of clusters play a significant role. To improve customer service and retention rates, the Bank should create a special marketing plan to target each group of customers after categorizing them.

In order to grow its market share, Awash Bank should think about creating marketing plans that are specially customized to various customer segments.

The major goal of this thesis is to use data mining to separate Awash Bank customers depending on their characteristics and the state of their transactions. Unsupervised machine learning algorithms like K-means are straightforward and popular. K-means the technique is employed for clustering. Filtered clustering and Density-based clustering also have been used for KDD. The methods divided the clustering result into two groups, and a clustering analysis was conducted using the results. Out of the three algorithms, k-means is better in speed. The k-means algorithm speed accounts for 2.4 seconds only. However, the density-based and filtered clusters are 2.76 and 2.92 respectively. In order to boost customer service and keep existing customers (finding new customers is more expensive than keeping existing ones), The results should be addressed, and the Bank needs to set up a way for customers to communicate in that per cluster group. It is not advisable to create a marketing plan prior to market segmentation. Furthermore, as was already said, there is a certain strategy that gives banks more chances to stand out from the competition and win over particular demographics. After consulting with marketing specialists and learning from several relevant literature sources, the marketing recommendation is finally place.

As mentioned already; pertinent to the number of withdrawal transactions, the clusters are divide into groups by customer dob, customer gender, customer location, customer account balance, transaction date, transaction amount, the whole withdrawal transactional groups (27% cluster 1, 73% cluster 2) is provided by the three algorithms. Then, because k-means is more focused on data variance faster in execution time than the other two algorithms, the k-means algorithm is the most selective.

The research questions are answered based on the selected algorithms; those are filtered clustering algorithm, density-based clustering algorithm, and simple k-means clustering algorithm. Then, they clustered all the given data sets into two clusters that hold six attributes. The experiment result shows the selected algorithm is suitable for clustering bank customers.

Out of the three clustering algorithms such as filtered, density, and simple k-means; although they have equal performance in grouping the data into two clusters, the k-means clustering algorithm is most efficient based on the execution speed in Weka object-oriented software. In addition, the identified centroid values are well expressing the difference between the two clusters which can be used for representing the two clusters for grouping other customers of the bank.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATION

5.1 Conclusion

The Banking sector can now provide more services as technology develops. This implies that there will be more competition, and if organizations want to compete, they must comprehend the offerings and actual needs of customers.

Understanding different customer services and figuring out which customer groups are most crucial to the Bank and which need more attention are both made possible through customer segmentation. To improve customer service and retention rates after categorizing its customers, the Bank should create a specific marketing plan to target the cluster groups.

The market will see new and aggressive competition for Awash Bank as a competitive bank. It can learn more about the activities of its customers and enhance the caliber of its services by using the transactions with and information about them. In order to grow their market share, banks should think about creating marketing plans that are specifically suite to various customer segments.

In this study, an attempt is make to apply clustering algorithms towards Awash Bank customer segmentation. To this end, customer and transaction data are collect from Awash Bank. The data is prepare to make it ready for constructing a descriptive model using k-means, density-base, and filtering clustering algorithms. The clustering result is divide into two groups using the algorithm, and the clustering analysis is do base on the results and score the value to map with the customer number (%). The majority of clients accounted for 73% of the total data, while the others made up 23%. Clustering is therefore crucial for keeping clients.

In this study, customer clustering base on transactional data is do using data mining techniques. After being preprocess, the database is then fille with the dataset that is obtain from the IT department. By using the unsupervise k-means clustering algorithm, filtered clustering algorithm, and density-based clustering algorithm techniques, the aggregated data is group in the Weka preprocessing and final tool. Because of this, building clusters will improve customer service to the point of competitive advantage. It seems possible to construct an environment that will draw customers. Finally, a marketing plan is suggest for the cluster groups base on the findings of the study in order to retain clients and gain a competitive advantage in the business market.

5.2 Recommendations

The transactional data generate by Awash Bank's customers is extensive. In order to uncover customer groups and trends in the collected customer data, data mining technologies can help the banking sector with marketing, gaining a competitive edge, and client retention. However, there is a challenge of acquiring data for analysis. So we recommend that financial institutions like banks collect and retain standardize data sets for experimentation. In addition, there is also a need for customer prediction base on customer segmentation using classification algorithms.

This study largely focuses on using customer transactional data or withdrawal-orient data to gain insights for segmenting customers base on their transactional state and providing a business strategy for appropriately treating and maintaining existing customers. In the future, there is a need to be able to use information from several banks in a variety of banking services, which will enable us to get around the restriction of only utilizing information from one bank. Additionally, Awash Bank has a wide range of data mining tools at its disposal that can be used in marketing in order to take advantage of the enormous volumes of transactional client data. Finally, what is recommend is that the bank make clustering models on deposit, loan, foreign exchange, and turnover.

Reference

- [1] Alizon, F., Shooter, S. B., and Simpson.(2009). Lessons from Henry Ford and the Model T for mass customisation and product platforms.30(5), pages 588–605, Design Studies.
- [2] Almotairi. a plan for effective CRM adoption. (2009) Sixth European and Mediterranean Conference on Information Systems (pages 1–14). Brunel and DokuzEylul Universities are located in Izmir.
- [3] Anderson, J. C., and Narus, J. A. (2004). 2nd edition of Business Market Management: Understanding, Creating, and Delivering Value. Publisher in New Jersey.
- [4] P. Andritsos.(2002) Strategies for data clustering. Computer science division of the University of Toronto.
- [5] L. Araujo, M. Spring, and others.(2006) the institutional framework of production, services, and goods. Journal of Industrial Marketing Management, 35(7), 797-805.
- [6] (n.d.) B2B International. Market division in B2B environments.From <http://www.b2binternational.com/publications/b2b-segmentation-research/>, retrieved November 5, 2013.
- [7] Baço, Lobo, and Painho (2005). Alternatives to k-means clustering using self-organizing maps. Computational Science-ICCS 2005 (pp. 476-483). Springer. Berlin.
- [8] Wilson, H., Bailey, C., and Clark, M. (2009). Customers no longer need to be grouped together, according to segmentation and customer intelligence in modern services marketing. Journal of Marketing Management, 25(3/4), pp. 227–252.
- [9] A. Weinstein and J. Barry (2009). Review of business psychographics: from segmentation theory to effective marketing techniques. Journal of Marketing Management, 25(3/4), pp. 315–340.
- [10] Levine, R., Caprio, G., and Barth, J. R. (2003).Lessons from a new database for banking regulation and oversight.macroeconomic stability, the development of the financial system, and economic growth. Banco de Mexico, a bank in Mexico City.
- [11] J. R. Barth, G. Caprio, and D. E. Nolle (2004).aspects of banking that are comparable internationally. The currency controller is located in Washington.
- [12] Beck, T., Demirgüç-Kunt, A., & Levine, R. (2006). Bank concentration, competition, and crises: First results. Journal of Banking & Finance, 30(5), 1581-1603.

- [13] F. Abdi and S. Abolmakarem, "Customer Behavior Mining Framework (CBMF) using clustering and classification techniques," *J. Ind. Eng. Int.*, vol. 15, no. s1, pp. 1–18, 2019, doi: 10.1007/s40092-018-0285-3.
- [14] Gordon S. Linoff and Michael J. A. Berry, *Data Mining Techniques. Third Edition For Marketing, Sales, and Customer Relationship Management, Third Edit.*, vol. 53, no. 9. Wiley, 2013.
- [15] S. Dolnicar, B. Grün, and F. Leisch, *Market segmentation analysis in Management for Professionals Being Aware of It, Doing It, and Making Use of It*. 2011.
- [16] E. Mattilla and S. Arnborg, "Behavioral Segmentation of Banking Communication Customers: Behavioral Segmentation of Banking Communication Customers," vol. fgg, 2008.
- [17] Z. Wang, L. Tu, Z. Guo, L. T. Yang, and B. Huang, "Analysis of user behaviors by mining large network data sets," *Future Generation Computer Systems*, vol. 37, no. 4, 2014, pp. 429–437, doi: 10.1016/j.future.2014.02.015.
- [18] A. S. M. S. Hossain, "Customers can be segmented using centroid-based and density-based clustering techniques," *3rd International Conference on Electronic Information and Communications Technology, EICT 2017*, volume 2018-Janua, no. December, pages 1-6, 2018, doi: 10.1109/EICT.2017.8275249.
- [19] O. C. Ferrell and M. D. Hartline, *South-Western Cengage Learning, Mason, Ohio, 2011. Marketing Strategy, SIXTH EDIT.*
- [20] K. Philip and K. Kevin Lane, *Marketing Management*, 14th ed. 2012. Prentice Hall.
- [21] H. Jiawei, M. Kamber, and J. Pei, (2011), *Data mining: principles and techniques*, 3rd edition, edition from Morgan Kaufman; *Analyzing of customer segmentation in Bank xyz using data mining techniques* Anta m. sunja, 2013.
- [22] A case study from South China University of Technology's Yiping Liu Applied Mathematics Department examined the use of data mining in the banking sector (2003).
- [23] "Using data mining for mobile communication clustering and characterization," A. Bascacov, C. Cernazanu, and M. Marcu. *8th IEEE International Symposium on Applied Computing, Intelligence, and Informatics, Proceedings, 2013*, pp. 41–46, doi: 10.1109/SACI.2013.6609004.
- [24] *Customer segmentation and profiling by data mining using the FRAT-RFM analysis approach: a case study of the cooperative bank of Kenya (2017)*

- [25] P.-S. Gregory, and S. Padhraic, Knowledge Discovery in From Data Mining to Databases, AI, vol. 17, no. 3, 1996, doi: 10.1609/aimag.v17i3.1230. Research on Commercial Bank's Data-Mining-Based Customer Segmentation Method
- [26] China's Henan province's Zhengzhou University's Management Engineering Department (2017)
- [27] M. Mohri, A. Rostamizadeh, and A. Talwalkar's book Foundations of Machine Learning. 2012 edition
- [28] A Comparative Study of Methods to Predict Customer Churn in the Banking Industry by M. Kaur, 2017.
- [29] D. Baumgartner and G. Serpen, "Large Experiment and Evaluation Tool for WEKA Classifiers," no. April 2017, 2009.
- [30] "A Grid-based Clustering Algorithm using Adaptive Mesh Refinement," 7th Work.Min.Sci. Eng.Datasets SIAM Int.Conf.Data Min, 2004, [Online]. LiaLiu04A.pdf is accessible at: users.eecs.northwestern.edu/choudhar.
- [31] K. Sanse and M. Sharma, "Clustering methods for Big dataanalysis," International Journal of Advanced Research in Computer Engineering Technology, vol. 4, no. 3, 2015, pp. 642–648.
- [32] Sanatkumar and IyerAurobindVenkatkumarJayantibhai"Comparative study of Data Mining Clustering algorithms," KondholShardaben, 2015, [Online]. Discreetly accessible at: <http://bdbib.javerianacali.edu.co:2208/document/7823946>.
- [33]. Ethio-Banking, "<https://www.ethioBanking.et/>," OUR STATISTICS, 2021
- [34] D. Steinberg, Chair for Session A of the Advanced Data Mining Tour.
- [35] H. Osaka U. Motoda, H.National U. of S. Liu, R. Setiono, and Z. Zhao, "Feature Selection: An Ever Evolving Frontier in Data Mining," J. Mach. Study. Research.Work.Conf. Proceedings. 10 Fourth Work.Featur.Sel.Data Min., pp. 4–13, 2010.
- [36] A. Vysala and D. J. Gomes, "Evaluating and Validating Cluster Results," 2020, pp. 37–47, doi: 10.5121/csit.2020.100904.

[37] W. Xiong, L. Chen, Z. ZhanG, and Z. Qiu, "RFM value and grey relation based customer segmentation model in the logistics market segmentation," Proc.- Int. Conf. Computer Sci. Softw. Eng. CSSE 2008, vol. 5, no. 1, pp. 1298-1301, 2008, doi: 10.1109/CSSE.2008.79.