



---

## **Privacy and Customization of Clusters with Application Development**

*<sup>1</sup>Chandana G R, <sup>2</sup>Pallavi G N, <sup>3</sup>Rekha Sahithi V, <sup>4</sup>Shwetha C, <sup>5</sup>Dr. Sumithra Devi KA*

<sup>1 2 3 4</sup>BE Students, <sup>5</sup> Project Guide Department of Information Science and Engineering

<sup>1</sup> Dayananda Sagar Academy of Technology and Management Bangalore, Karnataka, India.

<sup>1</sup>grchandana63@gmail.com, <sup>2</sup>gnpallavi6@gmail.com, <sup>3</sup>rekhasahithi003@gmail.com, <sup>4</sup>shwethacmalkhedkar@gmail.com

---

### **ABSTRACT -**

This paper presents a high level view of how clusters are being used in large number of domains for preserving and protecting the data. Because of these , clusters are being exposed to many attacks coming from open network.. Hence there are many methods to design a privacy preserving clusters. As the nations are connected to each other, so is the banking sector. Due to this globalization, the banks struggle to gain an edge over others. To overcome this, banks has rapidly increased the ability to generate, manipulate and storage of data of the customers because the information contained in this data can be highly useful[2]. The Ebanking industry around the globe has undergone a remarkable change in the way business is conducted. The banking industry has realized the need of the techniques like data mining, web mining, etc which can help them to compete in the market and to make the prediction and decision for the industry for growth. Banks can use data mining tools like clustering on customer's profile. This paper provides an overview of the concept of Kmean clustering algorithm and highlights the applications of e-banking to increase the customer's satisfaction Keywords K-mean algorithm, Clusterin security model for these banking applications. Overall this papers is all about banking applications which will let users customise their data clusters which has to be protected and different level of privacy is introduced in other to access this private data

**Key words:** Clusters, K-means algorithm, Open network, Data Mining, Protocols, privacy preserving, Sensitive data, Fraudulent activities, Data mining tools.

---

### **INTRODUCTION**

Data mining is becoming strategically important area for many business organizations including banking sector. It is a process of analyzing the data from various perspectives and summarizing it into valuable information. Data mining assists the banks to look for hidden pattern in a group and discover unknown relationship in the data. Today, customers have so many opinions with regard to where they can choose to do their business. Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations for the banking sector to avoid customer attrition. Customer retention is the most important factor to be analyzed in today's competitive business environment. And also fraud is a significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection. This paper analyzes the data mining techniques and its applications in banking sector like fraud prevention and detection, customer retention, marketing and risk management. Data to producefour separate clusters. IT has aided the bankingbusiness in dealing with a variety of economicdifficulties. Banks have realised that client relationships are a critical component of their success in recent years[1]. Customer relationship management (CRM) is a technique that will assist them in developing long-term ties with customers while also increasing revenues and profits. CRM is very important in the banking industry. In this research, we will use TwoStep Cluster to identify bank clients' profiles, starting with a public dataset provided by a German bank. This method has the advantage of determining the correct number of clusters; thus, the goal is to find the appropriate number of profiles in order to effectively managecurrent and potential clients.

---

### **LITERATURE SURVEY**

Dr. K. Chitra1, B. Subashini, presented the Data mining is becoming strategically important area for many business organizations including banking sector. It is a process of analyzing the data from various perspectives and summarizing it into valuable information. Data mining assists the banks to look for hidden pattern in a group and discover unknown relationship in the data. Today, customers have so many opinions with regard to where they can choose to do their business. Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful interpretations for customer attrition. Customer retention is the most important factor to be analyzed in today's competitive business environment. And also fraud is a significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop

new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection[3]. This paper analyzes the data mining techniques and its applications in banking sector like fraud prevention and detection, customer retention, marketing and risk management.

Kazi Imran Moin, Dr. Qazi Baseer Ahmed, presented that, In today's globalization and cut throat competition the banks are struggling to gain a competitive edge over each other. Apart from execution of business processes, the creation of knowledge base and its utilization for the benefit of the bank is becoming a strategy tool to compete. In recent years the ability to generate, capture and store data has increased enormously. The information contained in this data can be very important. The wide availability of huge amounts of data and the need for transforming such data into knowledge encourage IT industry to use data mining. The banking industry around the world has undergone a tremendous change in the way business is conducted. The banking industry has started realizing the need of the techniques like data mining which can help them to compete in the market. Leading banks are using Data Mining (DM) tools for customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. This paper provides an overview of the concept of DM and highlights the applications of data mining to enhance the performance of some of the core business processes in banking industry.

Aslı Çalış, Because of the phenomenal rise in information, future forecasting systems about strategy development were needed in each area. Therefore, data mining techniques are used extensively in banking area such as many areas. In this study, conducted in banking sector, it was aimed to reduce the rate of risk in decision making to a minimum via analysis of existing personal loan customers and estimate potential customers' payment performances with k-means method is one of the clustering techniques and the decision trees method which is one of the models of classification in data mining. In the study, SPSS Clementine was used as a software of data mining and an application was done for evaluation of personal loan customers.

Kanika Bansal, Anjali Bohra, As the nations are connected to each other, so is the banking sector. Due to this globalization, the banks struggle to gain an edge over others. To overcome this, banks has rapidly increased the ability to generate, manipulate and storage of data of the customers because the information contained in this data can be highly useful. The Ebanking industry around the globe has undergone a remarkable change in the way business is conducted. The banking industry has realized the need of the techniques like data mining, web mining, etc which can help them to compete in the market and to make the prediction and decision for the industry for growth. Banks can use data mining tools like clustering on customer's profile. This paper provides an overview of the concept of Kmean clustering algorithm and highlights the applications of e-banking to increase the customer's satisfaction.

Majid Sharahi, Mansoureh Aligholi, The purpose of this research is segmentation of bank customers using clustering techniques and is providing marketing strategies for each cluster of customers. Nowadays due to the large amount of customer data entities in the banking industry, the analysis of data obtained from the data bases of customers can provide useful information to detect hidden patterns in the data and can improve the level of banking services to each group of customers. In this study, we use the Excel file contains the information of 60 companies from customer of Bank Sepah in database of this bank and analyze it after got approval from the relevant authorities. Data analysis is performed using the Two-steps and K-means clustering algorithms which are from data mining methods. The file contains 14 data fields and after processing of data 7 fields are selected as a final variable to enter K-means algorithm. These fields are "Type of Company", "Life time of Company", "Type of Activity", "Time of Collaboration with Bank", "Credit history", "Type of Credit", "Amount of credit". Data were categorized into two clusters based on these even fields which 17 companies were classified into the first cluster and 43 companies were classified into the second cluster. First cluster companies often use the credit of civic participation in Bank and have the credit in the bank more than 10 years but the second cluster companies have the credit in the bank less than 5 years and use which could include 1- reduce services costs, 2- providing better conditions for the use of electronic service systems, 3- development of customer-oriented culture, 4-support of customer of production division as constant customers bank credit less in compared with the first cluster. We offer some suggestions about marketing strategy for each cluster after analyzing information of clusters. One of these suggestions is satisfaction strategy of Bank. These approaches while increasing profitability may also affect economic prosperity and employment. Banks can use data mining tools like clustering on customer's profile. data mining techniques are used extensively in banking area such as many areas.

---

## PROPOSED SYSTEM

Cluster analysis is a technique that identifies the complex relationships between variables, without imposing any restriction. Therefore, the input dataset doesn't need the distinct specification of an explanatory variable (the dependent variable) and respectively, of predictor ones (independent variables). There is no difference between the level of importance of the variables, the aim of the analysis is not to predict a certain value, but, to provide some clear view for the presence of specific patterns or correlations among variables, to include the different variables or cases into more homogenous groups. Cluster analysis can be used to explore the hierarchical structure of a system and that does not only provides an intuitive picture of the linkages of the system, but also displays meaningful cluster. Cluster analysis which groups (clusters) so that objects from the same cluster are more similar, with respect to a given attribute, to each other than objects from different clusters is a common technique for statistical data analysis. We propose a k-means clustering algorithm computes centroids and repeats until the optimal centroid is found. It is presumptively known how many clusters there are. It is also known as the flat clustering algorithm. The number of clusters found from data by the method is denoted by the letter 'K' in K-means. In this method, data points are assigned to clusters in such a way that the sum of the squared distances between the data points and the centroid is as small as possible. It is essential to note that reduced diversity within clusters leads to more identical data points within the same cluster.

In clustering applications we end up performing three tasks.

1. Separation of data sets into sections within the clusters,
2. Verification of results of clustering,

Interpreting the clusters. In the K-means algorithm, K value can be determined according to problem or it can not be determined. Like squared error criterion, there is need to have a clustering criterion. The K-means algorithm starts with random selection of an object that will represent every cluster. Each of remaining objects is assigned to a cluster and the clustering criterion is used to compute average of the cluster. These averages are used as new cluster averages and each of the objects are assigned again to the cluster that resembles itself most. These clusters are computed and until no change is observed in the clusters and no change fallunder the desired error level, this cycle is continued.

---

## IMPLEMENTATION

### Modules

1. Data Collection
2. Data Transmission
3. Data Pre-Processing
4. Data Clustering
5. User Module

### Modules Description

#### Data Collection

Data collection is the process of gathering and measuring data, information or any variables of interest in a standardized and established manner that enables the collector to answer or test hypothesis and evaluate outcomes of the particular collection. Within the scope of the study, data containing customer numbers and information about the status of paybacks belonging to the customers transaction of the branch where the application is going to be made were secured from the operating within the structure of General Directorate. Information about gender, marital status, age, monthly income, emailed, cvv, transaction date, transaction time., status of education, transaction amount, mode to transaction etc., the way of work were reached by using customer numbers and the existing system in the bank for making examinations.

#### Data Transmission

Data transformation is a technique used to convert the raw data into a suitable format that efficiently eases data mining and retrieves strategic information. Data transformation includes data cleaning techniques and a data reduction technique to convert the data into the appropriate form. Data transformation is an essential data preprocessing technique that must be performed on the data before data mining to provide patterns that are easier to understand. Data transformation changes the format, structure, or values of the data and converts them into clean, usable data. In our project we are converting gender, marital status & mode of payment raw data into a suitable format for data mining.

#### Data Pre-Processing

Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance, and is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running any analysis. Often, data preprocessing is the most important phase of a machine learning project. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult.

#### Data Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique [5]. Clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. In the K-means algorithm, K value can be determined according to problem or it can not be determined. Like squared error criterion, there

is need to have a clustering criterion. The K-means algorithm starts with random selection of an object that will represent every cluster. Each of remaining objects is assigned to a cluster and the clustering criterion is used to compute average of the cluster. These averages are used as new cluster averages and each of the objects are assigned again to the cluster that resembles itself most. These clusters are computed and until no change is observed in the clusters and no change fall under the desired error level, this cycle is continued The algorithm is composed of the following steps: Step 1: Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.

Step2: Assign each object to the group that has the closest centroid.

Step 3: When all objects have been assigned, recalculate the positions of the K centroids.

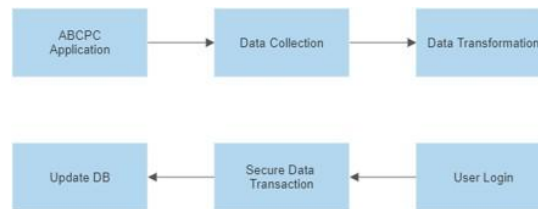
Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

### User Registration

The user module supports user for below operations: User can register and login with the application User can do the amount transaction

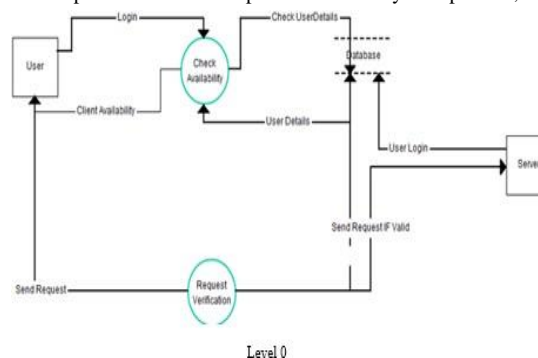
User can view the transaction details by providing the OTP send to registered email ID.

### System Architecture



### Data Flow Diagram

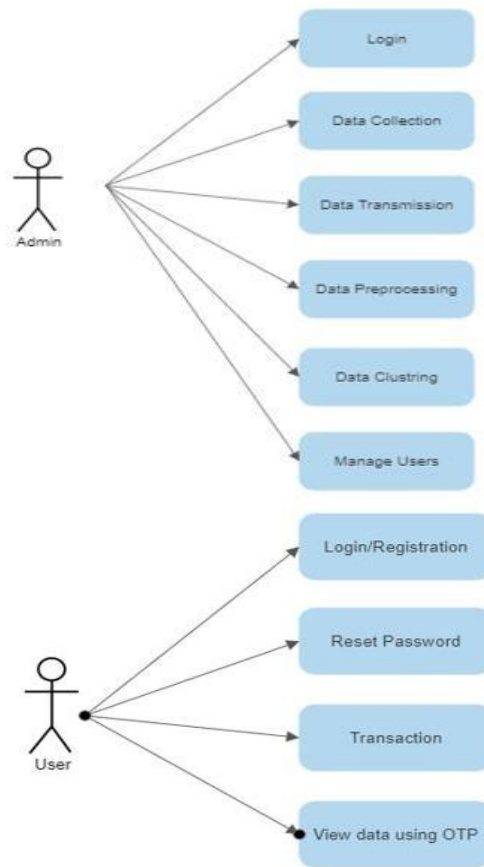
The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity



that interacts with the system and the information flows in the system. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

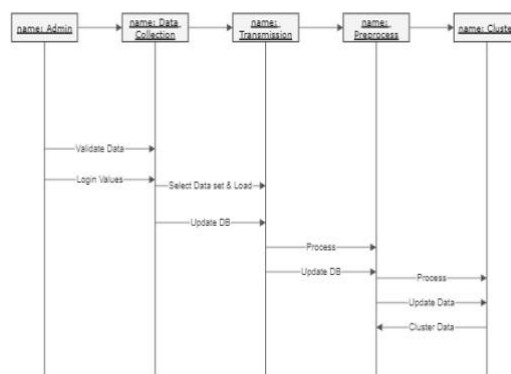
### Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



**Sequence Diagram**

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



**CONCLUSION AND FUTURE ENHANCEMENT**

Complementary security measures for clusters are discussed in this study, ranging from carrier-class clusters to High-Performance Computing clusters. Clusters at the carrier-class end of the cluster environment spectrum must be locked down to the greatest extent possible, with a focus on production dependability. Most clustering algorithms' final results are sensitive to accuracy factors in data mining, leading to algorithms being referred to as mature and practically intelligent machine learning algorithms. The approach detects issues in applications, including both unknown and known harmful

programmes. The security of applications is ensured by the protection technique. Two privacy-preserving k-means algorithms were presented. We also developed these algorithms and conducted extensive evaluations of them[2,4]. There are a few places where more research is needed. To reduce execution and bandwidth overheads, we need to make more improvements to our tools. Other clustering techniques with privacy-preserving variants are something we'd want to investigate. Hierarchical clustering methods are of great importance to us. Data warehousing is used to combine data from diverse data sources into a usable format so that the data may be mined. The data has been analysed, and it is now being used throughout the organisation to aid decision-making. Our project can be integrated to cloud technology. And also we can use our application for fraud detection data mining. We can integrate multiple data clustering algorithm, so that we can more filter the large amount of data.

---

## REFERENCES

---

- [1]. Dr.K.Chitra, B.Subashini(2019).Data mining techniques and its applications in banking sector.
- [2]. Fatemeh Amiri, Gerald Quirchmayr, Peter Kieseberg (2020).Sensitive data anonymisation using algorithms for SOM based clustering.
- [3]. Daniela Schiopu(2017). Applying to step clusteranalysis for identifying Bank customer profile.
- [4]. Francisca, Nonyelum, Ogwueleka (2019).Data mining applications in credit card fraud detection system.
- [5]. Makan Pourzandi ,David Gordon, William Yurchik (2018).Clusters and security in distributed security for distributed systems.
- [6]. Dingqi Yang, Daqing Zhang, Bingqing Qu (2020)Priv check privacy preserving check in data publishing for personalized location based security.
- [7]. Jialiu Lin, Bin Liu, Norman Sadeh (2017.)Modelling uses mobile application privacy preferences restoring usability in a sea of permission settings.
- [8]. Ferenc Kovacs, Csaba Legany,Attila Babos (2017)Cluster validity measurement techniques.
- [9]. Ali Inan, Yucel Saygm, Erkay Savas(2019).Privacy preserving classes onhorizontally partitioned data.
- [10]. S.Jha, L.Kruger, P.McDaneil (2019).Privacy- preserving clusters.
- [11]. Nils Urbach, Alexander Rieger, Gilbert Fridgen. (2019). Building a blockchain application that comprises with the general data protection regulation.
- [12]. Chunqiong WU. (2018). Research on cluster algorithm based on big data background.
- [13]. Adrian Holzer, Jan Ondrus. (2020). Trends in Mobile application Development.
- [14]. Hang Dong, Chengze Li, Ting Li, Yuejin Du, Guoai Xu. (2017). Research on security model of Mobile Applications.