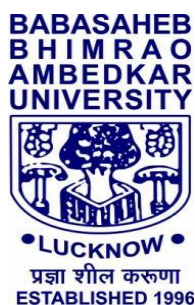


ANALYSIS AND DESIGN OF DATA MINING APPLICATIONS FOR BANKING INDUSTRY

**A Summary of Thesis
Submitted to the
Babasaheb Bhimrao Ambedkar University, Lucknow
in Fulfillment of Requirement for the Award of Degree of**

Doctor of Philosophy **IN** **COMPUTER SCIENCE**



BY
Banshi Dhar Choudhary
ENROLLMENT NO.- 1534/19

UNDER THE SUPERVISION OF
Prof. Vipin Saxena

**DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY) (NAAC-A++ AECREDITED)
LUCKNOW-226025, UTTAR PRADESH(INDIA)**

2024

SUMMARY

In the recent days, online business is gaining popularity, and everything is moving towards the digitization. The impact of coronavirus is also one of the reasons to expand the business in the online mode. Due to this, tremendous amount of data is increasing in exponential manner especially in the banking sector and banking industry is facing problems of managing and analyzing the vast amounts of data. Extraction of the desired amount of data in the minimum response time is also a big problem in the banking industry. For this purpose, and to address said challenges, data mining techniques are extensively upgraded by the researchers and scientists and the main aim of the present work is to explore the applications of data mining techniques in the cloud computing environment by considering the data of banking sector. Several data mining techniques have been investigated, including cube technology, clustering of data through K-means, C-means, fuzzy C-means and finally the utilization of ElGamal and blockchain techniques for securing data over the cloud are proposed for enhancing the services of the banking sector.

The cube technology enables the efficient analysis by summarizing and organizing data in a multidimensional format. K-means and fuzzy based C-means clustering algorithms help to identify the patterns and segments of the customers, enhancing marketing strategies and the management of the risks. Further, the implementations of ElGamal and blockchain techniques ensure the security and privacy of sensitive customer data during transmission and storage.

The finding of the proposed research emphasizes the importance of secure data mining techniques in the banking industry. The main aim of the proposed work is to protect the confidentiality and integrity of customer data with utmost importance and sensitive nature

of the information. The proposed techniques provide valuable solutions to address the security concerns and to facilitate the responsible use of data in the banking sector.

Overall, the proposed work contributes to the advancement of data mining techniques in the banking industry, providing valuable insights and practical applications. The presented findings serve as a guide for banking professionals, enabling to leverage data mining techniques to improve decision-making processes, enhance customer services and ensure the security of sensitive information. The chapter-wise summary of the proposed work is given below:

CHAPTER 1 INTRODUCTION

In the first chapter, the essential features that are crucial in addressing the challenges faced by data mining applications in the banking industry are discussed with the significance of ensuring data security during the transmission process over cloud servers. This chapter offers a comprehensive overview of the various tools, techniques and steps that are currently employed in the field to tackle the said issues effectively.

CHAPTER II REVIEW OF LITERATURE

This chapter presents a comprehensive review of the existing research conducted in the field of data mining applications for the banking industry. A meticulous analysis of various research publications spanning over the past eleven years has been conducted, with a particular emphasis on discussing and addressing key issues and challenges associated with the utilization of data mining techniques in the banking sector. The review encompasses an extensive survey of the literature to provide a comprehensive understanding of the current research landscape in this domain.

CHAPTER III DATA CUBE TECHNOLOGY

This chapter explores the utilization of data cube technology, a multidimensional storage and querying method. In the recent years, Online Analytical Processing (OLAP) has gained significant popularity as a powerful tool for accurate data analysis, particularly in the context of complex server architectures. The work presented in this chapter aims to address the challenges by employing the data cube technology and demonstrated its efficiency in terms of the reduction in the processing time, particularly for large databases of the credit cards. Additionally, computation of the correlation coefficient represents the excellent bindings among the data.

CHAPTER IV CLUSTER ANALYSIS FOR CLOUD DATA

In the current era, many of the organizations are migrating the large databases from the traditional servers over the cloud servers which may in public or private cloud environment. This is necessary to accommodate the exponential growth of the volume of the data as data contains text, audio and video files. The present chapter explores the possibility to create a different various cluster within the database and the K-means method is employed to identify duplicate patterns within the clusters. To demonstrate the effectiveness of the proposed approach, a sample database of credit cards is utilized for analysis, and the results are presented through tables and graphs. The research aims to provide comprehensive solutions for identifying and managing duplicate patterns in the large databases, offering potential benefits to organization in terms of data organization and optimization.

CHAPTER V DATA CLUSTERING THROUGH C AND K-MEANS TECHNIQUES

C and K-Means algorithms are widely recognized and employed the clustering techniques within the realm of the data mining. C-Means operates as a soft clustering algorithm,

enabling data points to be assigned to multiple clusters based on the similarity degree. Conversely K-Means function is a hard clustering algorithm, assigning each data points to a single cluster. This chapter presents a technique specifically designed for fuzzy C-Means clustering for database of banking and agriculture fields. The proposed technique holds a promise in assisting banks, farmers and service providers in effectively identifying and analyzing the data patterns. The enhanced accuracy of pattern recognition achieved through proposed technique which can contribute to improve the decision-making process and ultimately yields more favorable business outcomes.

CHAPTER VI CLOUD DATA CLUSTERING THROUGH WEKA AND R LANGUAGE

In the banking industry, Weka and R language stand out as prominent. Open-source data mining tools renowned for versatility and extensive utilization. Weka encompasses a collection of machine learning algorithms tailored for various data mining tasks, whereas R language serves as programming language and software environment primarily focused on statistical computing and graphics. The tools prove invaluable to banks and facilitate a wide array of data mining tasks, including, regression and association rule mining. Leveraging Weka and R language empowers the banking industry to efficiently process and analyze vast volumes of data. In the present chapter, the K-means clustering algorithm for cloud data using Weka and R programming language is proposed and it holds significant potential in assisting cloud service providers in achieving more accurate and efficient data clustering. Thus, in turn, it enhances the overall data management and analytical capabilities, contributing to improved decision-making processes.

CHAPTER VII FUZZY TECHNIQUE FOR SECURING THE CLOUD DATA

Cloud computing has emerged as the predominant computing paradigm, attracting significant attention from scientists and engineers. The continuous exchange of data between users and cloud servers, facilitated by high-speed internet service, necessitates the development of robust security techniques. This is particularly crucial for safeguarding sensitive information, such as digital currency-related data, from potential hackers. In this chapter, a secure approach is proposed by incorporating fuzzy logic into the ElGamal encryption technique, enabling secure data transfer during uploading and downloading processes via cloud servers. The effectiveness of the proposed approach is evaluated through a comprehensive case study, with the result presented in the form of tables and graphs, providing insights into its performance and security capability.

CHAPTER VIII DATA BASE ACCESSING THROUGH BLOCKCHAIN TECHNIQUE

In the recent times, distributed computing has emerged as a popular approach, allowing users to access databases at any time throughout the year. The banking industry, with its vast customer base, deals with a substantial amount of data. In this chapter, the application of blockchain is stored on cloud server's technology to enhance the accessibility of information stored on cloud servers, including text, audio and video files.

By leveraging the inherent features of blockchain, a novel approach is proposed for creating blocks of the database. The blocks enable faster and more efficient access to information, providing enhanced usability and convenience for users. The computer results obtained from implementing the approach are presented in the form of tables and graphs, showing the effectiveness of the proposed solution.

CHAPTER IX CONCLUSIONS AND FUTURE SCOPE OF WORK

Based on the comprehensive analysis presented in the preceding chapters, it can be concluded that data mining plays a crucial role in the banking industry. The application of data mining techniques facilitates the identification of valuable patterns in customer behavior, enables effective credit analysis and enhance fraud detection capabilities. The outcomes are achieved through the utilization of various tools and methodologies such as data cube technology, K-Means clustering, Weka and R programming languages, as well as fuzzy ElGamal encryption of data over cloud. Moreover, techniques like C-Means clustering, K-Means clustering and blockchain technology have also been recognized as valuable data mining approaches within the banking sector.

It is worth noting that the research conducted in the banking industry can be extended to other sector that also face the challenges of managing of the vast amount of data. The agriculture and health sectors may be benefited from similar data mining approaches to extract valuable insights, optimized operations and improved the decision-making processes.