Turkish Journal of Computer and Mathematics Education *Vol.10 No.01 (2019), 568-574*
DOI: https://doi.org/10.17762/turcomat.v10i1.13550

Research Article

Data Mining for Risk Assessment in Banking and Finance: A Review

Amit Gupta
Department of Computer Science & Information Technology, Graphic Era Hill University, Dehradun Uttarakhand India 248002

Abstract
A powerful analytical tool, data mining has become a common method utilized by financial institutions to identify and manage risks. The paper aims to provide a comprehensive analysis of the various techniques used in the mining of data for financial and banking organizations. It also explores the applications of these techniques in the risk assessment process. Risk assessment is a vital part of the operations of financial institutions. It involves identifying, measuring, and mitigating the risks that can affect an institution's financial health. The paper covers the various types of data mining tools that are commonly utilized for assessing financial and banking risks, such as clustering, association rule mining, and classification. It also provides a review of the limitations and challenges associated with using such techniques. The paper also reviews the literature on the various applications of data mining in the financial and banking sectors. These include operational risk assessment, credit risk assessment, and fraud detection. The review provides an overview of each application's aspects, such as data sources, preprocessing techniques, algorithms, and the results. The paper then explores the future directions for the research on the use of data mining for the assessment of financial and banking risks. It covers the latest trends in the field of data analysis, such as the incorporation of artificial intelligence, machine learning, big data analytics, and more
Keywords: Risk assessment, Banking, Finance, Fraud Detection, Portfolio management.

Introduction

In risk management, data mining is a process utilized to identify potential threats associated with financial activities. It involves analyzing vast amounts of information to identify trends and anomalies that could indicate a potential fraud or risk. This type of analysis can help financial institutions improve their risk assessments and prevent losses by identifying potential irregularities. The process of risk assessment is utilized by financial institutions to evaluate the potential rewards and risks associated with certain types of investments and transactions. It aims to identify the factors that can affect a particular investment or transaction and to recommend the appropriate steps to minimize these risks. It helps financial firms make sound decisions and improve their efficiency[1], [2].

Due to the increasing popularity of data mining in the financial and banking sectors, it has become a widely used tool for assessing risk. This process involves extracting meaningful and valuable information from vast amounts of data. Data mining can help identify relationships and patterns among various factors, which can then be utilized to make more informed decisions[3], [4]. In the finance and banking industry, it has been used for fraud detection and customer segmentation. The process of risk assessment is a vital part of the operations of financial institutions. It involves identifying and evaluating the various risks associated with certain financial transactions and lending. It helps banks and other financial firms minimize their losses and maximize their profits.

Traditionally, the process of risk assessment was carried out by financial experts. This can be time-consuming and error-prone. Due to the emergence of data mining technology, the method of risk assessment has become more streamlined. This paper aims to provide a comprehensive analysis of the latest data mining techniques that are used in the risk assessment process in the finance and banking sectors. It will cover various techniques such as association rule mining and decision trees. The paper will also cover the limitations and challenges of using data mining techniques in the finance and banking industry. It provides a wealth of knowledge and insight for practitioners, researchers, and stakeholders who are interested in learning more about how to utilize such techniques in risk assessment[1], [5].

Literature Review

The goal of this review is to provide an extensive analysis of the literature on the subject of data mining techniques for financial data analysis and fraud detection. The scope of the literature review indicates that data mining techniques

have gained widespread application in the financial sector. It explores the various uses of this technology, such as fraud detection and cybercrime prevention. Through data mining techniques, banks can identify and analyze various patterns and anomalies in their operations, which can help improve their decision-making and performance. It can also help them develop effective risk management strategies.

Rambola et al.[6] discuss the various aspects of fraud detection and the advantages of data mining techniques. They argue that these techniques can help financial institutions identify fraud. They highlight the different types of data mining techniques that are used in this field and their performance. The authors of this review state that decision trees are more effective than other techniques when it comes to detecting fraud in the financial sector. They also suggest that they can be utilized as a primary tool for this purpose.

Kadam et al.[7] talk about the importance of data mining and how it can be used to analyze financial data. They present an overview of various techniques, including clustering, decision trees, neural networks, and association rules. They also discuss their applications in the field of financial analysis. They conclude that financial institutions can make better decisions by utilizing data mining techniques. They also recommend that such methods be utilized to gain a competitive edge.

Zakirov et al.[8] covers the topic of data mining for banks and discusses the various types of financial information that can be mined using such techniques. He provides a brief introduction to some of the most common methods used in this field, such as clustering and decision trees. He concludes that financial institutions can make better decisions by utilizing data mining techniques. He also suggests that these methods should be utilized to gain a competitive edge.

Chawan et al.[9] talk about the significance of data mining and the various kinds of financial data that are available through this process. They detail the different techniques involved, such as clustering, association rules, decision trees, and neural networks. They state that financial institutions can gain a competitive edge by utilizing data mining techniques.

Moin et al.[10] talk about the significance and application of data mining in the financial sector. They present an extensive overview of the various techniques involved and their applications in the analysis of financial data. They conclude that financial institutions can gain a competitive advantage by using data mining techniques.

Hariharan et al.[11] noted that data mining could help financial institutions make informed decisions and improve their efficiency by analyzing vast amounts of data collected from various sources. The authors also emphasized the need for data preprocessing techniques and the use of feature selection methods to improve the models' accuracy.

Farooqi et al.[12] revealed that data mining could help improve the efficiency of banks by reducing the risk of loan defaults and improving the loan approval process. The study also noted that it could help improve the management of customer relationships by identifying their specific needs and preferences.

Raju et al.[13] conducted a study on the use of data mining techniques in retail and banking industries. They then proposed a framework that aims to integrate data mining with the management of customer relationships. The four phases of the framework are data collection, data processing, data mining, and management.

Lekha et al.[14] looked into the use of data mining methods to predict and detect cyber crimes in the financial sector. The author utilized various approaches, such as neural networks and decision trees, to analyze the incident data. The findings showed that these techniques could help financial institutions identify patterns and prevent future attacks.

S. Lahmiri et al.[15] analyzed the various techniques used in data mining for financial risk classification. He used six classification algorithms to analyze the data. Some of these include logistic regression, decision tree, support vector machines, neural networks, and random forest. The study was conducted on financial statements of companies that filed for bankruptcy from 1990 to 2012. The findings of the study revealed that feature selection is very important in improving the accuracy of classification systems. It can help reduce the number of irrelevant entries and improve the performance of the algorithms. The authors noted that support vector machines and random forest are the most accurate of the various classification algorithms.

The reviewed papers indicated that data mining techniques are commonly used in the finance industry. They can help banks identify potential risks and improve their customer relationships. The accuracy of data mining techniques

depends on the quality of the data collected and the features and preprocessing techniques utilized. This is why it is important that financial institutions thoroughly implement the necessary steps to improve the accuracy of their models. Although this technology has been widely used in the financial sector, further research is needed to examine its potential in other areas such as insurance and investment management. The review emphasizes the crucial role that data mining plays in the financial sector, and how it can help drive innovation and growth.

Various steps of risk assessment

The process of assessing risk in financial services is a crucial step in optimizing returns and managing exposure. It involves identifying and mitigating various risks associated with investments and transactions as shown in figure.1.

Figure 1 Steps in risk assessment

- Identification of Risks: The first step in identifying risks involves conducting a comprehensive analysis of the potential effects of a transaction or investment. This process can be carried out through various techniques such as historical data analysis and expert judgment.
- Risk Analysis: After identifying the potential risks, the next step involves assessing their likelihood and impact. Financial institutions utilize various analytical tools to analyze the risks, such as stress testing and statistical analysis.
- Risk Evaluation: After identifying the possible risks, the next step involves assessing their likelihood and impact. This process can be carried out by analyzing the multiple factors that affect a transaction or investment.
- Risk Mitigation: After carrying out the assessment, the next step is to implement the necessary actions to minimize the exposure of financial institutions. These include establishing risk limits, hedging strategies, and diversification.
- Risk Monitoring: The last step in the assessment process is to continuously monitor the potential risks associated with certain investments or transactions. This step ensures that the mitigation strategies are working.

Risk management and the optimization of returns are dependent on the efficient assessment of risk. Effective risk management involves knowing the various risks associated with an investment or transaction and utilizing suitable analytical methods. Thorough risk assessment techniques can help financial institutions reduce their exposure and increase their returns.

Crucial areas where risk assessment is useful in Banking and Finance

Risk assessment in the financial services industry can be carried out using data mining. This process can be used in various areas such as fraud detection and credit risk management. Data mining techniques are used in credit risk assessment to identify borrowers with high risk of default and to determine their likelihood of failing. In fraud detection, they can analyze the patterns of activity that lead to losses. Data mining can help improve the customer relationship management process by identifying the various patterns and preferences of customers. It can also help asset managers identify potential opportunities by analyzing market trends[2], [16], [17].

- Fraud detection: Data mining techniques can help identify fraudulent activities by analyzing customer transactions and behavior. This method can then flag unusual purchases or activities. For instance, it can analyze credit card transactions to determine if there was a sudden spike in activity or if the customer's spending habits had changed.
- Risk management: Mining data can help financial institutions identify potential risks and improve their efficiency. It can then inform their risk management strategies by analyzing customer data. For instance, by analyzing a customer's credit history, data mining can predict how likely it is that they will default on a loan.

- Customer relationship management: Financial institutions and banks can then use data mining techniques to identify patterns in their customers' behavior, such as their interests and preferences, to improve their customer service and marketing efforts.
- Portfolio management: Mining customer data can also help financial institutions identify opportunities to improve their asset allocation and diversify their portfolios. For instance, by analyzing the purchase history of customers, data mining can identify which services and products are most popular.
- Loan approval decision: When it comes to making a loan decision, data mining can help financial institutions identify which borrowers are most likely to default and adjust the terms of their loan. This method can analyze various factors such as income, employment status, and credit history.

Approach for Data Mining in Banking and Finance

Several steps are involved in implementing a data mining approach for risk assessment in finance and banking as shown in figure.2.

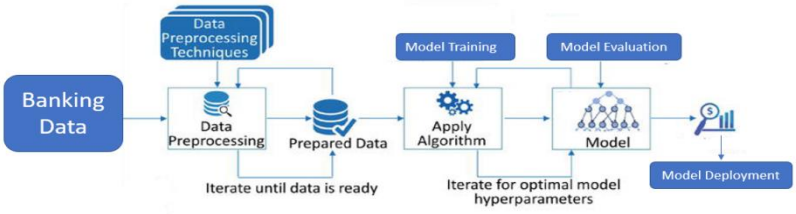

Figure 2 Data mining approach in banking and finance

- Data Collection: The first step is to collect and interpret the various types of banking data collected from various sources. This data can be used to perform risk assessments and improve the efficiency of the bank.
- Data Preprocessing: The preprocessing of the collected data involves various steps. These include data integration, which involves merging multiple sources, and data selection, which involves choosing the appropriate features from the data.
- Model Training: Once the data has been processed, a machine learning model is trained to map the inputs to the desired outputs. Various methods are used for this process, such as neural networks, logistic regression, or decision trees. The model's performance is then evaluated by validating the results.
- Model Evaluation and Improvement: The model's accuracy and performance are evaluated by validating the results. This process involves using metrics such as recall, precision, and F1 score. Its improvement can be achieved by modifying its parameters, performing different training steps, or by implementing a new algorithm.
- Model Deployment: After the model has been validated and trained, it can then be used to predict the likelihood of a customer experiencing a financial transaction. This process involves providing new data to the model.

  The goal of a data mining approach for assessing financial transactions and customers is to train a machine learning framework that can identify the risks associated with certain types of transactions. This method can then be used to perform risk assessments on other data sets.

Major data mining techniques to implement for risk assessment

- **Decision trees**: Decision trees are a type of supervised learning algorithm that can be used for classification or regression tasks. They work by recursively splitting the data into subsets based on the most informative features until a stopping criterion is met. The output is a tree-like structure that can be used to make predictions on new data. The formula for calculating the information gain of a feature is given by eq.1 :

  Information Gain= $H(T) - \sum_{i=1}^{n} \left|\frac{T_i}{T}\right| H(T_i)$ … (1)

  where T is the entire dataset, T_i is the subset of T that corresponds to the ith value of the feature, and H is the entropy of the dataset or subset is denoted by eq.2 :

$H(T) = -\sum_{i=1}^{c} p_i log_2(p_i)$ … (2)

where c is the number of classes and p_i is the proportion of samples in class i.

- **Logistic regression**: Logistic regression is a type of supervised learning algorithm that can be used for binary classification tasks. It works by modeling the probability of the positive class as a function of the input features using the logistic function is given by eq.3:

  $\sigma(z) = \frac{1}{1+e^{-z}}$ …(3)

  where $z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_mx_m$ is linear combination of the input features and weights, and σ is the logistic function. The weights w_i can be learned using maximum likelihood estimation.
- **Support vector machines** (SVMs): SVMs are a type of supervised learning algorithm that can be used for classification or regression tasks. They work by finding the hyperplane that maximally separates the positive and negative examples in the input space. The hyperplane is defined by the equation.4:

  $\vec{w} . \vec{x} - b = 0$ … (4)

  where $\vec{w}$is the weight vector, $\vec{x}$ is the input vector, and b is the bias term. The weight vector and bias term can be learned using an optimization algorithm such as gradient descent.
- **Random forests**: Random forests are an ensemble learning algorithm that combines multiple decision trees to improve accuracy and reduce overfitting. They work by randomly selecting subsets of the data and features to train each tree, and then combining the predictions of all the trees. The output is the most frequent prediction among the trees.
- **Naive Bayes**: Naive Bayes is a type of supervised learning algorithm that can be used for classification tasks. It works by modeling the conditional probability of the class given the input features using Bayes' theorem is given by eq.5:

  $P(y|x_1, x_2, x_3, \ldots, x_m) = \frac{P(x_1, x_2 \ldots\ldots, x_m|y)P(y)}{P(x_1, x_2, x_3, \ldots, x_m)}$ …(5)

  where y is the class$x_1, x_2, x_3, \ldots, x_m$are the input features, $P(y|x_1, x_2, x_3, \ldots, x_m)$is the posterior probability of the class given the features, $P(x_1, x_2, x_3, \ldots, x_m|y)$ is the likelihood of the features given the class, $P(y)$ is the prior probability of the class, and $P(x_1, x_2, x_3, \ldots, x_m)$is the marginal probability of the features.
- **K-nearest neighbors** (KNN): KNN is a type of lazy learning algorithm that can be used for classification or regression tasks. It works by finding the k closest examples to a new input and taking the majority class or average value of their labels as the prediction. The distance between examples can be calculated using different metrics such as Euclidean distance or cosine similarity.

  The formula for calculating the Euclidean distance between two examples $\vec{x}, \vec{y}$ is given by eq.6:

  $d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$ … (6)

  where $x_1, x_2, x_3, \ldots, x_m$ is the number of features.

  These data mining techniques can be used to analyze and model various risk factors in banking and finance such as credit risk, market risk, and operational risk. By leveraging the power of machine learning, financial institutions can make more informed decisions and minimize the impact of potential risks on their operations and customers.

Issues and challenges associated with risk assessment

Risk assessment is a process that involves analyzing and identifying potential risks in financial transactions. This process can be very useful in identifying potential problems and preventing financial losses. However, it can also be very challenging. This article will discuss some of the obstacles that prevent data mining from being used effectively in this field.

- Data quality and accuracy: The quality and accuracy of data collected for risk assessment are two of the most critical factors that can be used to make informed decisions. Unfortunately, it can be very challenging to maintain and obtain the necessary data due to various factors such as data bias and missing values. This can lead to inaccurate and potentially damaging predictions.

- Complexity of data: The complexity of the data collected by financial firms can make it difficult to extract meaningful insights from it. The data may come in various formats and sources, which can make it hard to identify meaningful correlations. Moreover, the data can get very voluminous, which can result in issues with processing and storage.
- Regulatory compliance: Due to the complexity of data mining, it can be very challenging to comply with the regulations set by financial institutions. This is why it is important that the data is analyzed and used in a way that is in compliance with these regulations.
- Lack of skilled personnel: In order to effectively carry out risk assessment, financial institutions need skilled individuals who can interpret and analyze data. Unfortunately, there are not enough people with this type of expertise. This can make it difficult for financial firms to perform data mining.
- Privacy and security concerns: When it comes to conducting risk assessment, data mining requires the access to sensitive information such as financial and personal data. However, this process can expose financial institutions to various security and privacy issues. To minimize these risks, financial institutions should regularly monitor and implement measures to protect their data.
- Data interpretation and decision-making: Due to the complexity of the data collected by mining for risk assessment, it can be hard to interpret and analyze. This process can also prevent financial institutions from making effective decisions.

  Data mining is an essential part of the risk assessment process in financial transactions to identify potential risks and prevent financial losses. It can also be very challenging to use effectively due to the complexity of the data and the lack of skilled personnel. To overcome these issues, financial institutions should consider investing in the necessary resources and implementing effective strategies.

Future directions

For many years now, data mining has been a vital part of the financial and banking sectors' risk assessment efforts. Due to the technological advancements that have occurred in this area, there are numerous new directions that researchers can pursue.

- Incorporation of Artificial Intelligence: With the help of AI, risk assessment in financial services and banking can be improved by identifying trends and patterns that are not apparent with traditional methods. Machine learning can then analyze historical data and provide predictions based on their findings.
- Through the use of AI, financial firms can now gain a deeper understanding of their data and identify potential risks. It can also help them automate the process of performing risk assessments, which can reduce the time and effort involved in the analysis.
- Big Data Analytics: Due to the increasing amount of data that banks and financial institutions collect, big data analytics is becoming an integral part of their risk assessment efforts. This technology can analyze large datasets and identify trends and patterns that can't be observed through traditional methods. In addition, it can provide banks and financial firms with a comprehensive view of their risks by integrating information from various sources.
- Social Media Analytics: Social media platforms collect a lot of information about their users. This data can be used to analyze and identify potential risks associated with certain behaviors or preferences. For instance, financial institutions can use social media analytics to monitor the activities of their customers.
- Visualization Techniques: Big data analytics can also be used to visualize the results of its mining process, making it easier for financial firms to understand the findings. This process can help them make informed decisions and improve the efficiency of their risk assessment efforts.
- Integration of Different Data Sources: An organization can gain a more holistic view of its risks by integrating different sources' data. For instance, by combining structured and unstructured information, such as social media data, it can provide a more accurate picture of its risks.
- Focus on Cybersecurity: Due to the increasing number of cyber attacks being carried out against financial firms, it is important that they adopt effective strategies and procedures to prevent these types of threats. One of the most common

ways that financial institutions can identify these risks is by conducting data mining. This process can analyze the data collected from various sources to identify anomalous behavior.

The field of data mining for financial and banking organizations is constantly evolving. With the incorporation of artificial intelligence, social media, and big data analytics, they can now gain a more comprehensive understanding of their risks. Doing so can help prevent financial losses and make informed decisions.

References

[1] A. Salehi, M. Ghazanfari, and M. Fathian, "Data mining techniques for anti money laundering," *Int. J. Appl. Eng. Res.*, vol. 12, no. 20, pp. 10084–10094, 2017, doi: 10.5120/ijca2016910953.

[2] S. Khemakhem and Y. Boujelbene, "Predicting credit risk on the basis of financial and non-financial variables and data mining," *Rev. Account. Financ.*, vol. 17, no. 3, pp. 316–340, 2018, doi: 10.1108/RAF-07-2017-0143.

[3] S. Bhattacharya, "Intelligent Frequent Pattern Analysis in Web Mining," vol. 2, no. 3, 2013.

[4] S. Bhattacharya, S. Rungta, and N. Kar, "International Journal of Digital Application & Contemporary research Software Fault Prediction using Fuzzy Clustering & Genetic Algorithm," vol. 2, no. 5, 2013, [Online]. Available: http://mdp.ivv.nasa.gov.in.

[5] S. SRIVASTAVA and G. ANCHAL, "Data Mining for Credit Card Risk Analysis: a Review," *Int. J. Comput. Sci.*, vol. 3, no. 2, pp. 193–200, 2013, [Online]. Available: http://pakacademicsearch.com/pdf-files/com/244/193-200 Vol. 3, Issue 2, Jun 2013.pdf.

[6] R. Rambola, P. Varshney, and P. Vishwakarma, "Data mining techniques for fraud detection in banking sector," *2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018*, pp. 1–5, 2018, doi: 10.1109/CCAA.2018.8777535.

[7] S. Kadam and M. Raval, "Data Mining in Finance," *Int. J. Eng. Trends Technol.*, vol. 16, no. 8, pp. 377–381, 2014, doi: 10.14445/22315381/ijett-v16p275.

[8] D. Zakirov, "Application of Data Mining in the Banking Sector," vol. 4, no. 1, pp. 13–16, 2015.

[9] A. A. S. Chawan, Pramila M, "Study of Data Mining Techniques used for Financial Data Analysis," *Int. J. Eng. Sci. Innov. Technol.*, vol. 2, no. 3, May 2013, 2017.

[10] K. I. Moin and Q. B. Ahmed, "Use of Data Mining in Banking," *Int. J. Eng. Res. Appl.*, vol. 2, no. 2, pp. 738–742, 2012, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.7821&rep=rep1&type=pdf.

[11] N. K. Hariharan and U. States, "Applications of data mining in finance," *... Appl. Data Min. ...*, vol. 5, no. 2, pp. 72–77, 2018, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3914436%0Ahttps://www.researchgate.net/profile/Naveen-Kunnathuvalappil-Hariharan/publication/354224201_APPLICATIONS_OF_DATA_MINING_IN_FINANCE/links/612d4f8138818c2eaf702bac/APPLICATIONS-OF-DATA-MINING-IN.

[12] R. Farooqi, "Available Online at www.ijarcs.info Effectiveness of Data mining in Banking Industry : An empirical study," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 827–830, 2017.

[13] P. S. Raju, V. R. Bai, and G. K. Chaitanya, "Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2014, pp. 2650–2657, 2014.

[14] K. C. Lekha, "Data M ining Techniques in D etecting and P redicting Cyber C rimes in Banking S ector," pp. 1639–1643, 2017.

[15] S. Lahmiri, "Features selection, data mining and finacial risk classification: a comparative study," *Intell. Syst. Accounting, Financ. Manag.*, vol. 23, no. 4, pp. 265–275, 2016, doi: 10.1002/isaf.1395.

[16] A. Keramati and N. Yousefi, "A Proposed Classification of Data Mining Techniques in Credit Scoring," *Techniques*, no. September, pp. 416–424, 2011.

[17] S. Khemakhem, F. Ben Said, and Y. Boujelbene, "Credit risk assessment for unbalanced datasets based on data mining, artificial neural network and support vector machines," *J. Model. Manag.*, vol. 13, no. 4, pp. 932–951, 2018, doi: 10.1108/JM2-01-2017-0002.