NUML International Journal of Engineering and Computing
Volume: 2 Issue: 1
https://numl.edu.pk/journals/nijec
Print ISSN: 2788-9629
E-ISSN: 2791-3465
DOI: 10.52015/nijec.v2i1.39

Data Mining Techniques in Decision Making

Amna Sajid[a, *], Basit Amin[b]

[a, *] *National University of Computer and Emerging Sciences, Islamabad*
School of Computing (amnasajid1996@gmail.com)

[b] *Riphah International University, Islamabad*
Faculty of Computing Department (basit.amintk@gmail.com)

Submitted	Revised	Published
17-June-2023	30-June-2023	26-July-2023

Abstract

Data mining aims to find relevant information for decision-making, forecasting, optimising, and other business or research reasons. In this study, data mining is used in the domain of decision-making. Various problems for this domain exist in the literature, including selection problems, considering unnecessary attributes, and removing irrelevant attributes in the dataset by applying different preprocessing techniques. For this purpose, a smartphone dataset is used, and different machine learning classifiers are applied. Decision Tree, Naive Bayes, SMO, bagging, and Random Forest were chosen for precision, recall, and F-measure. Results demonstrate that Random Forest obtains 57% accuracy and performs better on average than the other algorithms. The class "Saving Account" was classified with 52% accuracy as of the other attributes, but it has the fewest errors depending on the attributes. This study can be extended by applying the proposed methodology to a rich volume dataset and deep-learning techniques.

Keywords: *Data mining, Classification, Decision Trees, Random Forest, Decision Making, Marketing*

1. Introduction

In Data mining, mining means digging or extracting, so data mining is intended to extract some data. In this modern era, data mining is used in almost every domain widely. Data mining extracts useful and unknown information and patterns from the given data or datasets. Data mining is also known as the Knowledge Discovery Process (KDD) in which useful information is mined to make some decisions [1].

* Corresponding Author: Amna Sajid (amnasajid1996@gmail.com)

Data mining is used in e-commerce to increase productivity and analyze human buying nature. Data mining is also used in social networking such as Netflix, Facebook, Instagram, and all other social sites for recommendations of friends, products, movies, and much more [2]. There are four basic data mining goals: prediction, identification, classification, and optimization.

The data mining concept covers various aspects such as data preprocessing, data mining tasks (classification, clustering, association rules, etc.), evaluation metrics, and popular algorithms. The importance of data mining in extracting valuable knowledge and patterns from large datasets and their applications in various domains is relatively high and valuable [3].

Database management systems have been extensively utilized, and database technology has developed steadily in recent years. As a result, significantly more data is saved across many databases. Nevertheless, a significant portion of the information is obscured by these vast volumes of data. Businesses may make a lot of profits if the database's concealed information could be accessed; in addition, rivalry in many industries is already intense, growing more aggressive [4].

Data mining has different techniques used in different processes when using data mining approaches such as classification, clustering, regression, outlier detection, sequential patterns, predictions, and others [3]. Various problems exist in this domain which include:

- Selection problem
- Considering the critical attributes
- Removal of unnecessary attributes

2. Literature Review

Asdi, A. [5] works on the pricing strategies, product factors, sales locations, and promotional factors that affect customers' buying decisions for Samsung smartphones. These strategies are applied in Makassar City, and after the implementation, these factors positively impact the customers' buying decisions.

Nguyen, N. L. [6] carried out research in n Ho Chi Minh City by considering factors like lifestyle, brand image, and personality on the customer's buying decision. To evaluate the proposed methodology, 250 participants were interviewed to analyze their buying decisions.

Pradana, M. [7] researched humans' psychological and behavioral nature. This research proves that human nature can also affect customers' buying decisions. For the evaluation of this research, 385 subjects participated in the questionnaire survey, and linear regression was applied to the results for its results.

Mengash, H.A. [8] explores the application of data mining techniques to predict student performance and enhance decision-making in university admission systems. The study focuses on using machine learning algorithms and data analytics to analyze large datasets and extract patterns that can be used to predict student outcomes. The author highlights the importance of such predictive models in aiding universities in making informed decisions during admission. The paper concludes by emphasizing the potential benefits of utilizing data mining techniques in improving the effectiveness and efficiency of university admissions.

Kumar, T.S. [9] presents a study on developing a marketing decision support system using data mining techniques and a hybrid machine learning algorithm. The research focuses on leveraging data mining to analyze

marketing data and extract valuable insights for decision-making in marketing strategies. The author introduces a hybrid machine learning algorithm that combines multiple algorithms to enhance the accuracy and efficiency of the system. The paper emphasizes the significance of data mining in assisting marketers in making informed decisions and optimizing their marketing campaigns. Overall, the study highlights the potential of data mining and hybrid machine learning algorithms in enhancing marketing decision support systems.

Yun, Y. [10], in 2021, explores the development of a decision support system that integrates human-computer interaction techniques with data mining applications. The study emphasizes the importance of incorporating user-friendly interfaces and interactive features in decision support systems to enhance user engagement and decision-making processes. The authors highlight the potential of data mining techniques in extracting valuable insights from large datasets and utilizing them to support decision-making. The paper presents various applications of the proposed system in healthcare, finance, and marketing domains. Overall, the study underscores the significance of human-computer interaction and data mining in developing effective decision support systems.

Guo, Y. [11] focuses on developing a decision support system for intelligent manufacturing based on the Internet of Things (IoT) and data mining technology. The study emphasizes the integration of IoT devices and data mining techniques to facilitate information processing in manufacturing systems. The authors highlight the potential benefits of data mining technology in analyzing large volumes of manufacturing data to extract valuable insights and support decision-making processes. The paper discusses the application of the proposed system in various aspects of intelligent manufacturing, such as quality control and resource management. Overall, the study underscores the significance of IoT and data mining in improving decision support systems for smart manufacturing. Table 1 shows the comparative analysis of the literature review performed.

Table 1: Comparative Analysis of Literature Review

#	Contribution	Dataset Used	Technique	Results
[5]	Examines the effect of the marketing mix (4P) on buying decisions in the context of Samsung smartphone products.	Data on consumers' buying decisions and the marketing mix (4P) of Samsung smartphone products.	Statistical analysis and regression models.	Findings suggest that the marketing mix factors have a significant impact on consumers' buying decisions for Samsung smartphones.
[6]	Investigates the influence of lifestyle, brand image, and personalities on smartphone purchase decisions among consumers in Hochiminh City.	Data on consumers' lifestyles, brand image perception, personalities, and smartphone purchase decisions.	Structural equation modeling.	Results indicate that lifestyle, brand image, and personalities significantly affect consumers' smartphone purchase decisions in Hochiminh City.

[7]	Explores threat emotions and price impacts on Indonesians' smartphone purchasing decisions.	Data on Indonesians' smartphone purchasing decisions, threat emotions, and price perceptions.	Regression analysis and structural equation modeling.	Findings reveal that threat emotions and price perceptions significantly influence Indonesians' smartphone purchasing decisions.
[9]	Presents a data mining-based marketing decision support system that utilizes a hybrid machine learning algorithm.	Marketing data for decision support and analysis.	Hybrid machine learning algorithm.	The system demonstrates improved accuracy and efficiency in decision-making processes for marketing strategies.
[11]	Discusses an IoT-based decision support system for information processing in intelligent manufacturing.	Manufacturing data and IoT devices for information processing.	Data mining technology and IoT integration.	The system enhances information processing in intelligent manufacturing through the integration of IoT devices and data mining technology.

3. Research Methodology

In datasets, data mining techniques can find anomalies and outliers. These anomalies might indicate fraud, malfunctioning systems, or other strange behavior that needs prompt attention. Organizations can reduce risks, increase operational efficiency, and make better decisions by quickly identifying and addressing anomalies. Below figure 1 explains the flow of the process that is conducted for the data mining process. How the data is divided for training and testing, and how the data travels through data mining.

First, the data is downloaded from an online source and then sent for preprocessing. Preprocessing includes removing redundancy, removing duplications, removing the null value, and other related issues. After the preprocessing of the data is cross-fold by applying cross-validation techniques, the data is sent for classification. Different machine learning algorithms are applied for classification, such as random forest, naïve Bayes, and decision trees. After classification, the results are obtained and sent for testing and evaluation. For problems involving decision-making, these classifiers are frequently employed in machine learning. The data's nature, the task's difficulty, the need for interpretability, and performance concerns all play a role in the classifier of choice. Before choosing the best classifier for a certain decision-making scenario, testing out many classifiers and evaluating their performance is frequently advantageous. Classifiers play a significant part in decision-making

by assisting in the creation of predictions or classifications based on input data. The patterns and linkages identified from past data allow decision-makers to automate the labeling or decision-making process for new instances.

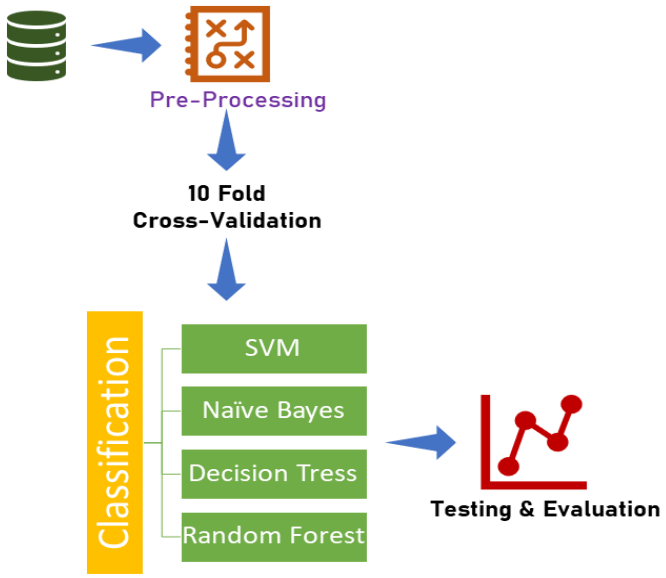

Figure 1: Research Methodology

Machine learning classifiers improve attribute accuracy in the dataset by efficiently handling missing data, outliers, imbalance, and noise. Additionally, they support feature engineering and selection, model optimization, and the use of information from related fields, all of which improve the accuracy of attribute representations.

3.1 Dataset

The dataset chosen for this research is "Smart Phone Buying Decision." The dataset "Smart Phone Buy Decision" was in CSV format. The dataset consists of five attributes: age, income, student, saving account, and buying a smartphone. The dataset contains 34 instances or rows. The dataset "Smart Phone Buy Decision" was downloaded from: https://sites.google.com/site/labitis462/lab-tutorials. The chosen dataset is publicly available, open source, and freely available for research. The dataset "Smart Phone Buying Decision" consists of five attributes: Age, Income, Student, saving account, and Buy smartphone. The dataset "Smart Phone Buying Decision" data type belongs to nominal. No numeric value exists in the entire dataset. No missing values are in the entire data set, and no unique values are identified in the dataset. So, there are no values that are no null values present in the dataset.

3.2 Implementation

This study was implemented using WEKA (Waikato Environment for Knowledge Analysis) tool. WEKA is a popular open-source software tool for data mining and machine learning. It provides a comprehensive set of algorithms and tools for data preprocessing, classification, regression, clustering, association rules mining, and more. Developed by the University of Waikato in New Zealand, WEKA offers a user-friendly graphical interface and a flexible framework for analyzing and modeling datasets.

4. Results

Data mining is a major part of creating decision support systems (DSS). These systems help decision-makers in difficult decision-making processes by combining user-friendly interfaces and data mining techniques. Decision-makers are equipped with actionable insights through the interactive visualization, data exploration, and scenario analysis capabilities offered by DSS [12].

Classification is applied to the dataset after the preprocessing of the dataset. In the selected dataset "Smart Phone Buying Decision," preprocessing techniques are applied, including handling missing values, removing redundancy, and others. After the implication of preprocessing techniques, five classification algorithms are applied to the dataset. The results after applying the classification algorithms are given below in the table.

Table 2: Results for Classifiers

Attributes	Decision Tree	Naïve Bayes	SMO	Bagging	Random Forest
Age	Precision 32.9%	Precision 44.2%	Precision 34.3%	Precision 44.2%	Precision 53.9%
	Recall 35.9%	Recall 44.1%	Recall 35.3%	Recall 44.1%	Recall 52.9%
	F-measure 32.9%	F-measure 44.1%	F-measure 33.5%	F-measure 43.6%	F-measure 52.7%
Student	Precision 44.5%	Precision 41.2%	Precision 40.6%	Precision 58.7%	Precision 59.9%
	Recall 44.1%	Recall 41.2%	Recall 41.2%	Recall 58.8%	Recall 58.8%
	F-measure 44%	F-measure 41.2%	F-measure 40.8%	F-measure 57.9	F-measure 58.5%
Saving Account	Precision 38.4%	Precision 38.4%	Precision 40.7%	Precision 77.9%	Precision 60.6%

	Recall 47.1%	Recall 47.1%	Recall 50%	Recall 64.7%	Recall 61.8%
	F-measure 40.9%	F-measure 40.9%	F-measure 42.7%	F-measure 55.5%	F-measure 58.2%

Now, after the classification, we calculate the error rate for every attribute present in the attributes. The error rate is calculated for every class present inside the dataset. Error rate means accurately classifying the data into their respective or desired classes. The class with the lowest error rate shows the highest classification accuracy while using specific machine learning or data mining technique or algorithm.

Decision-making benefits from machine learning and data mining models with decreased mistake rates. Higher accuracy and reliability of the models are indicated by lower error rates, which promotes more confident and well-informed decision-making. A lower error rate helps in decision-making by improving predictions, fewer risks, enhancing performance, and improving customer satisfaction. A decreased error rate in machine learning models dramatically improves decision-making, which can result in more successful and effective outcomes across various applications.

Ensemble methods, such as Random Forest and Gradient Boosting, combine multiple classifiers to make more accurate predictions. By aggregating the outputs of multiple models, ensemble methods reduce bias and variance, improving attribute accuracy.

Table 3: Error Rate

Attribute	Classification Ratio	Classification rate
Age	Youth = 5/12	15/34
	Middle = 6/19	
	Senior = 4/13	
Student	Yes = 4/16	12/34
	No = 8/18	
Saving Account	Yes = 9/14	13/34
	No = 4/20	
Smart Phone	SamsungS6 = 4/23	7/34

	iPhone = 3/11	

After implementing this dataset on WEKA, the results were obtained by applying different data mining and machine learning classifiers. Decision Tree, Naïve Bayes, SMO, Bagging, and Random Forest were selected for Precision, Recall, and F-measure. Results show that Random Forest performs better than the other algorithms and achieves 57% accuracy on average. While based on the attributes, the class "Saving Account" has a minimum number of errors and was classified with 52% accuracy as to the other attributes.

Data mining makes Predictive analytics possible, enabling decision-makers to make well-informed forecasts and projections based on past data. Organizations can predict future results, market demand, or customer behavior using regression, time series analysis, or machine learning techniques that aid in improving resource planning, sales forecasting, and inventory management [11].

5. Conclusion

Data mining techniques are useful in almost every known domain. These techniques are helpful in decision-making, finding trends, making predictions, and various other mechanisms. In this research, data mining techniques are used for the decision-making process. For this purpose, a smartphone dataset was chosen, which is publicly available and online. Different data mining techniques were applied to this dataset to test their precision-recall and F-measure scores. The result shows that the Random Forest classifier outclasses the other classifiers. And the attribute name "Saving account" has the minimum number of error rate present. This study can be enhanced by applying a rich volume dataset with more attributes and different deep-learning techniques.

References

[1] Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. International Journal of Information Technology, 12(4), 1243-1257.

[2] Hovale, S. V., & Poonam, G. (2016). Survey paper on recommendation systems using data mining techniques. International Journal Of Engineering And Computer Science ISSN, 2319-7242.

[3] Nivethithaa, K. K., & Vijayalakshmi, S. (2021, June). Survey on Data Mining Techniques, Process, and Algorithms. In Journal of Physics: Conference Series (Vol. 1947, No. 1, p. 012052). IOP Publishing.

[4] S. Li, W. Chen, J. Hu, J. Hu, ASPIE: a framework for active sensing and processing of complex events in the internet of manufacturing things, Sustainability 10 (3) (2018) 692.

[5] Asdi, A., & Putra, A. H. P. K. (2020). The Effect of Marketing Mix (4P) on Buying Decision: Empirical Study on Brand of Samsung Smartphone Product. Point Of View Research Management, 1(4), 121-130.

[6] Nguyen, N. L., Tran, T. T., & Vo, M. P. (2020). The Effect of Lifestyle, Brand Image and Personalities on Smartphone Purchase Decision of Consumers in Hochiminh City. International Journal of Innovative Science and Research Technology, 5(3).

[7] PRADANA, M., & WISNU, A. (2021). The Impacts of Threat Emotions and Price on Indonesians' Smartphone Purchasing Decisions. The Journal of Asian Finance, Economics, and Business, 8(2), 1017-1023.

[8] Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision-making in university admission systems. IEEE Access, 8, 55462-55470.

[9] Kumar, T. S. (2020). Data mining-based marketing decision support system using a hybrid machine learning algorithm. Journal of Artificial Intelligence, 2(03), 185-193.

[10] Yun, Y., Ma, D., & Yang, M. (2021). Human–computer interaction-based decision support system with applications in data mining. Future Generation Computer Systems, 114, 285-289.

[11] Guo, Y., Wang, N., Xu, Z. Y., & Wu, K. (2020). The Internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. Mechanical Systems and Signal Processing, 142, 106630

[12] Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. Morgan Kaufmann