*International Journal of Basic and Applied Sciences*, 14 (5) (2025) 700-713

**International Journal of Basic and Applied Sciences**

*Website: www.sciencepubco.com/index.php/IJBAS*
*https://doi.org/10.14419/npfq8f76*
***Research paper***

# Improved Classification Approach for Business Intelligence Using Data Specific and Feature Oriented Min-Max Normalization

**Syed Ziaur Rahman [1] *, Bhuvan Unhelkar [1], Prasun Chakrabarti [2], Mohammed Ali Hussain [3]**

[1] *University of South Florida, United States*
[2] *Department of Computer Science and Engineering, Sir Padampat Singhania University, Udaipur Rajasthan, India.*
[3] *Professor of CSE and Dean R & D, Sreenidhi Institute of Science and Technology, Hyderabad, India*
**Corresponding author E-mail: profziapd@gmail.com*

*Received: July 27, 2025, Accepted: September 2, 2025, Published: September 18, 2025*

**Abstract**

Business intelligence is becoming more and more important to managers in today's corporate setting as a crucial aspect of information technology. Many companies are eager to use intelligent technologies to enhance their decision-making processes in corporate operations. Accordingly, intelligence is typically linked to human-like traits, including the ability to assimilate new information, learn from mistakes, and pursue objectives similar to those of a human. One of the main objectives of business intelligence integration across businesses is to generate reports through management dashboards that use key indicators to facilitate informed decision-making. The objective of this paper is to develop a practical model that uses machine learning indicators and algorithms to optimize the product sales system using classification techniques. Several factors are used by the model to enhance client classification techniques. Additionally, the study uses association rules to look at clients' shopping carts, find links between the items they have bought, and generate personalized offers based on the rules they have discovered. Different classification techniques, including C4.5, random forest, and reduced error pruning tree, are used to compare with the suggested methodology before being used to evaluate and improve the results. The findings show that the best results should be obtained using the recommended methodology in conjunction with the indicated model.

***Keywords***: *Business Intelligence; Management Dashboards; Machine Learning; Min-Max Normalization; Classification Methods.*

## 1. Introduction

Business intelligence, or BI, is a technology-driven process that analyzes data and generates information to help executives, managers, and staff make informed and effective business choices. The "thing" in question is not just one. Business analytics, data mining, data visualization, tools, and infrastructure are among the fields that are integrated to assist firms in making more data-driven decisions. It is very different from traditional business intelligence, which originally surfaced in the 1960s and was only a way for businesses to share information. As computer models were developed throughout time, they also changed. This illustrates how much it has evolved to satisfy commercial needs over the past few decades. marketing, education, and information technology. and healthcare is the primary industry that has used business intelligence. A few real-world examples of business intelligence platforms in action are as follows: HelloFresh, a meal kit company, employed business intelligence (BI) to solve its ineffective digital marketing reporting problem. Coca-Cola bottling had a problem where access to real-time sales data was restricted by manual reporting procedures. This technological advancement has also enabled the use of business intelligence (BI) and machine learning (ML) techniques [[1]]. Online analytical processing, which is necessary for multidimensional reporting, forecasting, dashboard reporting, and machine learning (ML)-based recommendation systems, is all included in the broad category of business intelligence operations. Better management choices are made possible by the additional insights this evolution provides into important business characteristics, like clients and rivals [2]. Large amounts of data are routinely generated and managed by many organizations, especially companies and medium-sized enterprises [3]. Machine learning-powered business intelligence (BI) is essential for tech applications such as demand forecasting. Beyond these uses, business intelligence (BI) boosts an organization's worth, forecasts future events, simplifies processes, speeds up product development, strengthens ties with suppliers and consumers, and facilitates quicker and more accurate decision-making [4]. Additionally, BI helps companies find growth obstacles, operational issues, and new market trends [5]. Business intelligence (BI) is used by companies in a variety of industries, such as sales, production, payroll, and warehousing, to track performance in relation to corporate goals and store data in databases [5]. This includes tracking the advancement of objectives, assessing research that identifies the company's advantages and disadvantages, and utilizing BI tools to increase operational effectiveness. The notion that computers would be able to learn and do things without programming was born out of recognition. Numerous basic and complex algorithms are needed to finish the operation on different data formats and get the desired result. The two most often used machine learning approaches, among others, are supervised learning and unsupervised learning.

Amazon is an illustration of a business that makes the most of machine learning (ML) in business intelligence technologies. It helps them to generate precise data from a large quantity of raw data and support product customization according to customer requests. Work will be substantially improved by using machine learning, but there are also a number of disadvantages.

Business Intelligence: Business intelligence (BI) has developed over the past few decades to become a fundamental component of organizational decision support. BI, or business intelligence, is a method that is technologically advanced to assist us in data analysis and provide useful information regarding the performance of the specific business. It facilitates the process of making wise decisions that will lead to better results. This technological idea emerged soon after the Multiway Data Analysis Consortium, an international meeting that took place in Rome in 1988 [15]. They concluded that BI analysis should be made easier to understand and more straightforward. In addition, a lot of BI companies provided new BI products. Business intelligence software has helped organizations plan and analyze more quickly, better, and more accurately; enhance data quality; increase operational efficiency; improve customer and employee happiness; lower costs; raise revenues; and make better business decisions. For business intelligence projects to be successful, vendors, users, and information specialists must all be involved.

Applications of Machine Learning in Business: To gain a deeper understanding of their clients and internal processes, most businesses in the modern world rely on machine learning algorithms. The procedure makes use of a wide range of machine learning methods. The following are a few of the most popular uses of machine learning in business: [32]

Customer lifetime value models are crucial to e-commerce because they can be used to discover, comprehend, and acquire the most valuable clients for the business. They may also be used to forecast the potential revenue generated from a specific customer over a given time frame.

Churn modeling: It assists in determining which clients are most likely to discontinue doing business with the company. This enables the business to develop tactics that aid in drawing in new clients while retaining existing ones.

Dynamic pricing is the process of setting prices for products based on several variables, such as the target customer's level of interest and demand. Engines for recommendations: One of the most crucial elements that goes into an online business's success is recommendation engines. It remembers what clients prefer and makes recommendations for related products to entice them to buy the item. It lowers churn and improves the customer experience.

In this research, we present an effective model based on machine learning indicators and algorithms that employ categorization approaches to improve product sales systems. To analyze consumer purchasing patterns, identify product associations, and tailor offers according to established standards, the methodology makes use of association rules. Strict assessment standards for consumer classification algorithms are also incorporated. To enhance and validate the outcomes, a range of classification methods is applied.

Accuracy, AUC, TP Rate, TN Rate, FP Rate, FN Rate, Precision, Recall, and F-measure techniques are among the evaluation metrics that are beneficial in marketing for consumer segmentation based on purchase behavior and are the topic of this study [6]. By classifying customers into Platinum, Gold, Silver, and Bronze groups, businesses can increase customer loyalty and revenue. This increases consumer loyalty and promotes bigger purchases. Among the significant classification methods that are employed are C4.5, Random Forest, and Reduced Error Pruning (REP) trees. Maximizing client segmentation accuracy and identifying successful product pairings to boost sales performance are their objectives. This study contributes significantly in two ways: it maximizes the efficiency of product sales and determines the optimal algorithms for extremely accurate consumer segmentation.

This paper's remaining sections are arranged as follows: Section 2 examines relevant research in the area. The suggested BI approach, datasets, and implementation tools are presented in Section 3. Results and analysis of several machine learning methods are presented in Section 4 to identify the most successful strategy. Results and implications are discussed in Section 5. The paper is finally concluded with general observations and recommendations for further work in Section 6.

## 2. Background

In recent years, businesses and organizations have made extensive use of business intelligence, a data-driven technology. Contextualizing corporate data and accelerating centralized decision-making processes are critical functions of business intelligence that have grown in significance because of the growing competitiveness in the commercial and economic sectors. IT and management experts have taken on these responsibilities. Business intelligence is used in many different domains and is an essential tool for analysis and decision-making. It helps with client retention, hiring, and sales growth in addition to reducing expenses and losses and increasing corporate efficacy and efficiency. Businesses can also foresee market trends with the use of business intelligence. In traditional operating systems, it used to take a lot of time and effort to create multidimensional reports due to the extensive code integration and database linkages. Consider the difficulty of creating reports using data from several operating systems. For these kinds of reports, creating a data warehouse and compiling pertinent data into it became crucial, even though it was time-consuming and sometimes impractical. The cornerstone of creating a high-quality data repository is accurate identification, which can be a difficult process in operational settings where operating system knowledge is necessary.

Reporting errors and difficulties are directly caused by a lack of a sufficient data repository. On the other hand, a well-organized data repository facilitates the development of machine learning algorithms, improves operational effectiveness, and facilitates future planning for the company. This study provides a straightforward approach to enhancing the product sales system using classification approaches in conjunction with machine learning measurements and algorithms.

1) Access to Actionable Insights: BI provides decision-makers with timely, relevant data in an approachable manner. It assists them in gaining a comprehensive understanding of the company's performance, customer behavior, market trends, and other crucial factors that influence decision-making [33]. With this access to actionable information, decision-makers are better equipped to make choices that advance organizational objectives.
2) Data-Driven Decision-Making: Decisions can now be based on facts rather than merely intuition or guesswork thanks to business intelligence (BI). Decision-makers can use data analytics and visualization to evaluate historical performance, identify patterns, and predict future outcomes. This data-driven approach lessens subjective biases while increasing the accuracy and effectiveness of decision-making processes [34].

3) Increased Decision Agility and Speed: BI systems facilitate faster decision-making by providing users with real-time or almost real-time access to data and insights. When decision-makers have immediate access to relevant information, they can respond swiftly to changing customer demands, market conditions, and competition issues [35]. This flexibility allows organizations to stay ahead of the curve and seize opportunities fast.
4) Improved Strategic Planning: Decision-makers may extensively analyze customer preferences, market trends, and the competitive environment thanks to business intelligence (BI). Strategic planning is made feasible by this in-depth understanding of the business environment. Decision-makers can identify growth opportunities, optimize resource allocation, reduce risks, and align business goals with market expectations [36].
5) Performance Monitoring and Assessment: Real-time tracking and evaluation of metrics and key performance indicators (KPIs) is made feasible by BI. Decision-makers may track the progress of goals, identify areas that require improvement, and proactively address issues. Organizations can closely monitor performance and make data-supported adjustments to their operations, processes, and plans.
6) Enhanced Operational Efficiency: BI helps identify bottlenecks, inefficiencies, and holes in a company's processes. Decision-makers can use BI insights to streamline processes, optimize resource allocation, and identify areas for automation or improvement. This leads to increased operational efficiency, reduced costs, and increased productivity [35].
7) Improved Customer Understanding: Deeper understanding of consumer behavior, needs, and preferences is made feasible by BI. Consumer insights can be used by decision makers to develop customized marketing efforts, improve client experiences, and produce goods and services that are specifically targeted. Organizations can improve customer relations, encourage loyalty, and obtain a competitive advantage by coordinating choices with consumer needs.
8) Risk Mitigation and Compliance: Decision-makers can benefit from BI systems' ability to recognize possible hazards, spot irregularities, and handle compliance concerns. Decision-makers can proactively manage risks, put in place suitable controls, and guarantee adherence to laws and industry standards by examining data and keeping an eye out for trends [37].

Machine Learning (ML) Overview:

Three general categories can be used to classify ML algorithms:

1) Supervised Learning: In supervised learning, where each input label is connected to a matching output or target label, machine learning algorithms are trained on labeled data. The algorithm learns to convert the input data into the desired output by spotting patterns and connections in the training data. Following training, the model can make predictions or classify new, unseen data using the patterns it has found.
2) Unsupervised Learning: Unsupervised learning involves training machine learning algorithms on unlabeled data to identify patterns. connections or patterns in the data that lack labels [36]. Unsupervised learning methods are frequently applied to issues including clustering, dimensionality reduction, and anomaly detection. The computer learns to recognize outliers, cluster data points, or find commonalities based on inherent patterns in the data.
3) Reinforcement Learning: This method instructs an agent to interact with its environment and learn from the feedback it receives in the form of rewards or penalties. By exploring its environment, acting, and receiving feedback, the agent gradually learns and improves its decision-making techniques. Reinforcement learning is widely used when an agent has to make successive decisions, such as in autonomous systems or gaming [38]. Machine learning algorithms employ several techniques, such as decision trees, neural networks, support vector machines (SVMs), clustering techniques, and deep learning models, to tackle a range of problems. The choice of algorithm depends on the kind of data, the problem, and the desired outcome.

Developing algorithms that can identify patterns in data, learn from them, and make predictions or judgments is the aim of the AI subfield of machine learning [39]. It combines supervised, unsupervised, and reinforcement learning techniques and has multiple applications in a variety of fields. Machine learning (ML) is a powerful technique for tackling difficult problems and fostering creativity by enabling computers to learn and develop from experience.

Business Intelligence using Machine Learning

In the realm of business intelligence (BI), machine learning (ML) has become a game-changing technology that is transforming how businesses get insights from data and arrive at well-informed judgments. By facilitating more precise forecasts, sophisticated data analysis, and automated decision-making procedures, machine learning techniques improve business intelligence capabilities [41].

1) Preparing and Integrating Data: The BI data integration and preparation procedures can be automated and optimized with ML algorithms. ML may be used more effectively to carry out processes like feature engineering, data transformation, and data cleansing, which will cut down on the time and manual labor needed to get data ready for analysis.
2) Predictive Analytics: Predictive analytics in business intelligence (BI) is made feasible by machine learning (ML), which uses past data to anticipate future events. By employing machine learning (ML) algorithms to identify patterns and relationships in data, businesses may forecast consumer behavior, estimate demand, anticipate market trends, and make data-driven decisions.
3) Customer Segmentation: Machine learning algorithms can categorize consumers into distinct categories based on their characteristics, inclinations, or actions by evaluating customer data. Because of this segmentation, businesses can target specific consumer segments, tailor marketing campaigns, and alter products and services to meet the unique needs of different clientele groups [41].
4) Anomaly Detection: Machine learning algorithms are adept at identifying data abnormalities or outliers that may indicate fraudulent activities, abnormal behavior, or operational inefficiencies. By identifying these irregularities, organizations can proactively address potential risks, bolster security, and enhance operational processes.
5) Machine learning's Natural Language Processing (NLP): Branch enables machines to understand and assess human language. NLP can be used in business intelligence for sentiment analysis, text mining, and automatic text summarizing. It helps companies to extract insights from unstructured textual data, such as customer reviews, social media data, and survey responses.
6) Recommendation Systems: Machine learning-powered recommendation systems can analyze user behavior and preferences to provide personalized recommendations. By recommending products, content, or actions to users using techniques like content-based filtering and collaborative filtering, these systems enhance user experience and boost customer engagement.
7) Data Visualization and Reporting: ML algorithms in BI data visualization and reporting can be used to create interactive dashboards, real-time reports, and visual representations of data. Machine learning techniques such as dimensionality reduction and clustering can help visualize complex datasets and identify important patterns or relationships.
8) Augmented Analytics: Machine learning is transforming business intelligence (BI) by enabling augmented analytics, which combines ML capabilities with traditional BI tools. Through the automation of data analysis, the disclosure of hidden insights, and the provision of proactive insights and suggestions, augmented analytics empowers users to make better-informed decisions.

9) Real-time Decision-Making: Machine learning algorithms help organizations make data-driven decisions faster by processing and analyzing data in real-time. Real-time ML-powered analytics enable businesses to monitor operational indicators, spot anomalies, and respond swiftly to changing conditions or emerging patterns.
10) Continuous Improvement: In reaction to new data and human input, machine learning algorithms can learn and adapt over time, improving their functionality. This iterative learning process can help organizations refine their BI models, adapt to shifting business situations, and enhance decision-making processes.
11) The Business Intelligence (BI) process mainly depends on data collection and preparation, which Machine Learning (ML) techniques can significantly enhance. The primary components of BI data collection and preparation are as follows, along with how ML may be useful:
12) Data Collection: To create a comprehensive dataset for analysis, data collection entails obtaining pertinent information from a variety of internal and external sources. Machine learning (ML) can assist with a variety of tasks, including web scraping, automated data extraction from structured and unstructured sources, and data integration from numerous databases or systems [39]. Machine learning algorithms can automate data extraction, filtering, and aggregation, increasing process efficiency and accuracy.
13) Data Cleaning and Preprocessing: Data cleansing is crucial to ensuring the reliability, correctness, and consistency of the collected data. Machine learning techniques can automate and speed up the data cleaning process by identifying and controlling duplicates, outliers, missing values, and inconsistencies in the data. To improve the overall quality of the data, machine learning algorithms can recognize patterns in the data and make informed decisions about how to impute missing values or find and remove anomalies.
14) Data Transformation and Feature Engineering: Machine learning (ML) can help convert unprocessed data into an analysis-ready state. This entails activities like standardizing and normalizing data, encoding categorical variables, and developing new derived features. By recognizing patterns in the data and establishing correlations between variables, machine learning algorithms can propose beneficial changes or feature engineering strategies that improve the data's quality and predictive ability [38], [39].

## 3. Related Works

Hans Peter Luhn, a researcher at IBM, first used the term "business intelligence" in 1958. He described it as "the ability to understand and interpret data in a way that facilitates effective decision-making." Between the middle of the 1960s and the middle of the 1980s, there was a discernible increase in decision support systems. To describe "concepts and methods for improving business decision-making through fact-based support systems," Howard Dresner coined the term "business intelligence" in 1989. The word gained widespread usage starting around 1990 [7].

In essence, business intelligence refers to computer processes and strategies that transform data into actionable knowledge that might improve an organization. Business intelligence, which has its roots in data analysis, gives managers and end users useful information that enables them to apply machine learning algorithms to make strategic choices. Among the primary advantages of business intelligence technology are improved decision-making speed and accuracy, process optimization, operational efficiency, and gaining a competitive advantage over competitors [8]. Recent advancements have made it possible to successfully use data mining and machine learning to solve a variety of commercial challenges. After extensive development procedures, business intelligence employs a range of machine learning approaches to forecast organizational futures [9 - 11]. By highlighting the importance of machine learning methods in big businesses, Rath et al. [10] set the foundation for the development of business intelligence. While Gutnik et al. investigated how data mining and machine learning techniques boost the effectiveness of digital marketing tactics [9], Whig et al. examined the roles of cloud computing, machine learning, and artificial intelligence in corporate operations [11]. Somayaji et al. [12] suggested a deep neural network and blockchain method to control and predict battery life in Internet of Things (IoT) devices. The proactive scheduling of battery changes is made easier by this technique. To identify network intrusions, Khare et al. [13] used deep neural networks with spider monkey optimization.

The process of educating computers to learn from their past experiences and gradually get better at making decisions is known as machine learning. By introducing a novel deep learning model for spatiotemporal-based sentiment analysis on tweets, Parimala et al. [14] assisted governments in taking preventative measures following tragic incidents. Using deep neural networks to anticipate cyberattacks, Swarna et al. [15] created an efficient intrusion detection system for the Internet of Medical Things (IoMT) environment. A deep neural network model was presented by Reddy et al. [16] to continuously monitor the maritime environment and follow the effects of climate change. They also introduced an ensemble-based machine learning approach to diabetic retinopathy classification, which aids doctors in diagnosing various ailments [17]. Advanced advancements in business information across various domains are available to interested clients [28 – 29].

## 4. Motivation

Business Intelligence Machine Learning Implementation: A small number of firms have attempted, but failed, to integrate machine learning into business intelligence. They made use of the IBM H2O platform. An enterprise version of the machine learning software was developed in 2018 by IBM and H2O.ai [23]. The primary goal of the next-generation software H2O.ai is to use software to introduce artificial intelligence into the corporate world. IBM Power Systems, which are built to manage massive volumes of data and have high memory requirements for machine learning, are the ideal platform for the H2O autonomous AI. Why did the machine learning implementation fail? Relevant data is necessary for machine learning to succeed. Significant non-zero prediction bias exists in this instance [24]. Strong regularization, the lack of outside data affecting the result set, and the training data set's inadequate representation of specific data space subsets were the causes of this. In machine learning, a bias is a type of prediction error. Prediction bias is a metric that quantifies the difference between two averages. It can be calculated by deducting the average of the labels in a data set from the average of the forecasts. By searching for significant non-zero prediction bias, which indicates that the model is inaccurate with relation to the frequency of positive labels, we can identify any flaws in our model. Applying steganalysis [26] techniques and other data protection procedures, such as watermarking [25], can help address the serious problem of data security that arises in this situation. Instead of increasing the lessons learnt from certain biases, it is crucial to prevent bias so that the models can operate smoothly on any data. which machine learning algorithms would be most effective? [10]. Given the challenges encountered in this instance, where the company failed to successfully integrate machine learning into the system, the decision tree approach may be applied. Based on the idea of probability, this supervised machine learning method is an excellent choice due to its simpler interpretation and implementation. Numbers ranging from 0 to 1 that can be understood as probabilities are the outputs. Regularizing strategies and feature removal are two more ways to enhance decision tree models. Any type of corporate organization must have a solid understanding of sales forecasting. It informs us of the elements that are crucial to making money and the ones that should be avoided. Predictions can be made using an algorithm such as C4.5 decision tree. We can utilize the model-building approach since it enables us to forecast a variable's value by using the value of another variable.

Even when used in conjunction with other methods to achieve more precise results, it produces an accurate output and requires little training time. It is simple to understand. In a corporation, this algorithm can be used to forecast and assess trends. This algorithm, as its name implies, solves all problems using a decision tree model building. Nevertheless, the decision tree algorithm is limited to problems with simpler solutions, which may not be the case in many real-world situations. For instance, using information from past monthly sales, a business can forecast its sales for the upcoming months. Random forest regression is an additional algorithm that can be applied [27]. One of the most used supervised learning algorithms, it enables the quick extraction of important information from big data sets. To arrive at a single solution, it mostly uses the construction of several decision trees. This algorithm has a low risk of overfitting and performs well with non-linear data.

Machine learning is becoming more important in business intelligence (BI), according to recent academic research. For example, in their all-encompassing framework for machine learning in business intelligence analytics, Shmueli et al. (2023) highlight the possibilities for automated decision assistance and predictive modelling. But they do point to problems with understudied aspects of many BI solutions, such as dealing with different datasets and making sure they are interpreted. Furthermore, Khan et al. (2020) show how ML-enabled BI-driven demand forecasting models may be useful, but they also admit that there are problems with scalability and feature selection in real-time settings.

In a more recent study, Al-Quhfa et al. (2024) examine ML models used for talent recruitment and draw attention to the challenge of finding a balance between accuracy and fairness. On the other hand, Mohammed and Shihab (2024) apply deep learning to business intelligence (BI) based on social media and draw attention to the absence of techniques for normalizing diverse and noisy data sources. Gaps persist despite these advancements. While many studies have looked at algorithms like random forests and deep neural networks, few have taken a close look at preprocessing techniques like feature-oriented min-max normalization, which have a direct impact on classification accuracy and bias reduction. Another area where research is lacking is in personalized suggestion generation, specifically in the area of customer segmentation via the use of association rules and classification.

## 5. Proposed Method

The study uses machine learning algorithms to enhance decision-making processes aimed at boosting revenue and attracting new customers. In order to develop successful tactics for future legislation, the project intends to develop precise learning models and examine existing data trends. These frameworks for decision-making will be supported by modeling and statistics. Maintaining strong customer relationships is more challenging in today's competitive market environment. Businesses need to communicate with their customers every day in order to comprehend and meet their evolving needs. This will guarantee the company's prosperity. Because it encourages profitability and repeat business, which both support organizational operations, customer satisfaction—which is influenced by their perception of a product's value—is crucial [18]. Managers attempt to predict future consumer behavior by using historical data, but they encounter difficulties when attempting to develop practical sales strategies. Machine learning algorithms provide a solution by spotting instructive sales trends. Clients are first categorized according to predetermined standards, and then the policies for each class are tailored. The purchasing habits of each group are then examined using association rules.

### 5.1. The Proposed Method of Business Intelligence

The steps of the proposed method are presented in Figure 1. The detailed explanation of the different phases is presented in this section.

1) Data collection: The first step in BI is data collection, which includes obtaining information from both internal and external sources. Spreadsheets, enterprise systems, transactional databases, and other internal data repositories are examples of internal data sources. Public data sources, industry databases, social media platforms, and market research studies are examples of external data sources. Data collection may include processes for data loading, extraction, and transformation. Business intelligence connects supermarket data from many sources, which is a big advantage. In this part, the dataset used is from the data warehouse's (DW) sales system. The data from the store served as the foundation for every method and analysis used in this investigation.
2) Data Integration: Data must be collected and then consolidated into a single source. Combining data from several sources, standardizing data formats, settling disagreements, and guaranteeing data quality are all part of data integration. The goal of this approach is to develop a logical data perspective for reporting and analysis [4].
3) Data cleansing: Data cleansing is an essential BI step to ensure the accuracy and reliability of data. It comprises adding missing figures, standardizing data formats, removing duplicates, and identifying and correcting errors. By enhancing the data's quality and integrity, data cleansing yields more accurate insights.

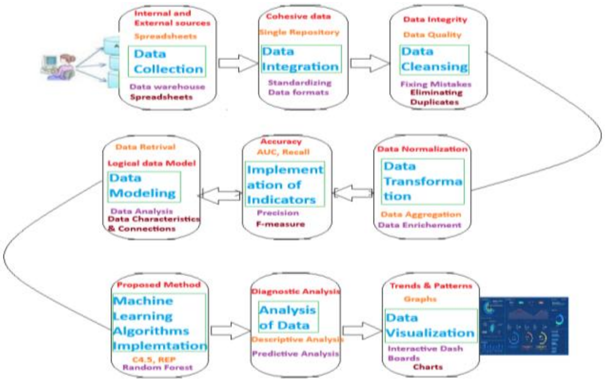


**Fig. 1:** Framework of the Proposed Method.

1) Data Transformation: This process entails transforming unprocessed data into a uniform format for reporting and analysis. Data normalization, data aggregation, data enrichment (e.g., combining external data with internal data), and the development of computed

measures or derived variables are a few examples of this procedure. For better transformation in this particular case, the min-max normalization technique is employed.

Min-Max Normalization

Min-max normalization, sometimes referred to as feature scaling, is used to linearly alter the original data. This method is used to retrieve all scaled data in the 0–1 range. Equation 1 provides the following formula to achieve this goal:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Min-max normalization is used to preserve the relationships between the initial data values. The disadvantage of this limited range is that it leads to smaller standard deviations, which could mitigate the effect of outliers.

2) Implementation of Indicators: During this stage, each firm must establish the criteria for monitoring its key performance indicators (KPIs) [26]. These indicators are crucial components of the assessment-based sales systems. The indicators used in this study are Accuracy, AUC, Precision, Recall, and F-measure, etc.
3) Data Modeling: Creating a BI system's data pieces' relationships and structure is known as data modeling. This entails establishing a logical data model that outlines the data's elements, characteristics, and connections. Efficient data retrieval, analysis, and reporting are made possible by data modeling.
4) 7.Machine learning algorithm implementation: Following the installation of the indicators, machine learning methods are used to run the model. The feature selection process is carried out in the first stage and is determined by the kind of indicators in these numerous features and studies that might have been present in the sales system, but had no bearing on the system's operation, have been eliminated. First, in order to eliminate irrelevant variables from the sales system, characteristics are chosen based on indicators pertinent to this study. Following that, a categorization technique [19–22] tailored to that specific index type is used to proceed with model construction. The important supervised learning algorithms C4.5 [23], random forest [24], and REP [25] are used in the study.
5) Analysis of Data: In BI, data analysis refers to a variety of approaches and procedures used to extract insights from the data. It includes descriptive analytics, which provides summary statistics and visualization of historical data, and diagnostic analytics, which seeks to understand the reasons behind certain outcomes or trends [6]. In order to predict future trends and outcomes, predictive analytics also uses statistical models and machine learning algorithms.
6) Data Visualization: Data visualization is essential to corporate intelligence because it makes difficult data visually appealing and understandable. To illustrate important measurements, trends, and patterns, it entails producing interactive dashboards, maps, graphs, and charts. Users may easily evaluate information, spot outliers, and get insightful knowledge with the use of data visualization [7].

Correlations between various dataset attributes are then found by looking into data relationships. This includes algorithms like FP-growth and Apriori that provide rules that are helpful for market analysis. Following data preparation based on evaluation indices, all modeling and classification techniques are used. To ascertain the optimal number of customer classes, an analysis is then carried out using root mean square error as the classification criterion. On the WEKA (Witten and Frank 2005) [27] workstation, the recommended approach is put into practice using an Intel(R) Core (TM) i5-10400 CPU system unit running at 2.90 GHz with 8.0 G of RAM and Windows 11. Additionally, the effectiveness of these methods is assessed for three different algorithms that are being considered for implementation.

## 5.2. Data Set

There are 1,000 instances in the supermarket dataset with 16 variables. Some of the attributes are numerical (e.g., quantity, unit price, total, gross revenue, rating), while others are more often used as categories (e.g., branch, location, client type, gender, product line, payment method). With the addition of sequential patterns made possible by temporal data like time and date, it is possible to analyse purchasing behaviour over multiple time periods. We find that there is an imbalance in several parameters, including gender and customer type, and that if we do not fix it, it can compromise the categorisation results. For instance, there may be bias in rule generation and segmentation due to the dominance of specific product lines and payment methods in transactions.

There are 6,435 occurrences in the Walmart dataset with 8 features. In most cases, the attributes are numerical (such as temperature, petrol price, CPI, unemployment, and weekly sales), but in a few cases, they are categorical (such as holiday flag and shop ID). There is a significant income imbalance in the weekly sales distribution due to the disproportionate impact of specific retailers and holiday weeks. The inability to generalise the model to non-peak weeks may be due to this skewness's tendency to favour times of high activity. Feature normalisation must also take into consideration the possibility that macroeconomic variables, such as unemployment and the consumer price index, create hidden correlations that vary by area and by time.

The supermarket dataset is city- and branch-specific, but the Walmart dataset is representative of the retail market in the United States as a whole, and both datasets display domain-specific biases. Despite the suggested approach's outstanding performance within the datasets, these constraints highlight the need for further validation on various datasets from other retail settings to prove its wider application.

**Table 1:** Details of the Experimental Datasets

| Dataset | Attribute: | | | |
|---|---|---|---|---|
| Supermarket | Branch | City | customer type | Gender |
| | product line | Tax | Time | gross margin percentage |
| | unit price | Total | Payment | gross income |
| | Quantity | Date | Cogs | rating |
| Walmart | holiday_flag | temperature | fuel_price | cpi |
| | unemployment | weekly_sales | Store | |

To enhance analysis and determine the best business principles, this study looks at the sales dataset from a few perspectives. The perspective of view can be assessed using the attributes and features of the dataset. Eight of the 16 attributes/features in the dataset are common and have been selected for additional analysis. Eight features were selected, including city, gender, branch, customer, date, payment, product, and time.

## 6. Results Evaluation

In this study, we validated the proposed algorithms using several popular and practical criteria. It is well recognized that the error rate should not be employed as an evaluation criterion in cases of unequal cost distribution or class imbalance. False positives (FP), which are negative but categorized as positive, true positives (TP), true negatives (TN), and false negatives (FN), which are truly positive but classed as negative, were counted in order to evaluate the classification results. In real-world applications, a variety of advanced and suitable metrics are used to evaluate unbalanced datasets.
Equations 2, 3, and 4 provide the formulas for some helpful evaluation metrics that we might derive:

$$\text{Precision} = TP / (TP + FP) \tag{2}$$

$$\text{Recall} = TP / (TP + FN) \tag{3}$$

$$\text{F-measure} = 2RP / (R + P) \tag{4}$$

Where R and P refer to Recall and Precision, respectively.
In assessing the effectiveness of classification, the F-measure, Precision, and Recall metrics are essential. Precision, also known as positive predictive value, is the percentage of relevant instances that are recovered, while recall, also known as sensitivity, is the percentage of relevant instances that are retrieved. According to the calculations, lowering FP encourages precision and increasing TP enhances recall. However, there is a trade-off between these two indicators, making it challenging to optimize both simultaneously. Therefore, to provide a fair measurement that accounts for Precision and Recall, the F-measure is used. When both Precision and Recall are high, a greater F-measure indicates better performance. To evaluate the performance of the proposed approach in this study, we compare it with several other algorithms from related literature, including C4.5, Random Forest, and Reduced Error Pruning Tree. The validation technique we use is 10-fold cross-validation. The results of the comparative tests on the accuracy, AUC, FP rate, TN rate, FN rate, precision, recall, F-measure, and RMSE of each method are shown in Tables 2 through 11.
Figures 2 to 11 display the average metrics of several indicators for each algorithm so that the performance of the three algorithms can be assessed using the recommended methodology. Tables 2 through 11 demonstrate how the suggested method performs better on the bulk of the datasets than the other three techniques. Tables 2 and Figures 2 to 11 illustrate the performance of the proposed method in comparison to other methods. This is because the suggested technique skews toward the minority samples because the degree of imbalance in the training set influences how best features are initialized. As a result, the algorithm's RMSE fluctuates. However, the proposed method performs better than C4.5, RF, and REP in terms of accuracy, TP Rate, TN Rate, Recall, F-measure, and AUC.
More precisely, the average precision of the recommended method is 1, but the values of the other algorithms are less than 1. The average recall increases to 1 while the other values remain below 1. While the other comparative algorithm values are less than 1, the average F-measure value is 1. Additionally, the average TP Rate and AUC values of the recommended method rise to 1, respectively, whilst the values of the other techniques fall below 1.
Overall, the experimental results show that the recommended approach outperforms all other approaches. This is explained by the employment of the best feature set, which effectively processes dubious data, especially in the hyperedge set's boundary region, by employing the noise instance detection technique. Furthermore, the neighborhood strategy, which incorporates more hyperedges in the class choosing process and improves accuracy, is used to locate noise examples to compute weights. These two elements contribute to the improved categorization outcomes of the proposed method. The accuracy results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 2. We can see that our recommended approach produces better outcomes. The Walmart sales results are shown in the table's last column.

**Table 2:** Accuracy Results on the Super Market dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 100 | 99.68 | 100 | 100 |
| Super Market Gender | 51.62 | 48.87 | 49.7 | 100 |
| Super Market Branch | 100 | 99.74 | 100 | 100 |
| Super Market Customer | 49.54 | 49.02 | 49.55 | 100 |
| Super Market Date | 1.02 | 0.97 | 1.97 | 4.87 |
| Super Market Payment | 33.48 | 33.29 | 33.74 | 100 |
| Super Market Product | 18.92 | 18.58 | 17.77 | 49.1 |
| Super Market Time | 0 | 0.02 | 0.49 | 10.76 |
| Walmart_Sales | 79.2 | | 57.6 | 71.6 |

The AUC results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 3. We can see that our recommended approach produces better outcomes. Walmart sales data is shown in the last column of Table 3, and the recommended method yields competitive results.

**Table 3:** AUC Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 1 | 1 | 1 |
| Super Market Gender | 0.516 | 0.484 | 0.49 | 1 |
| Super Market Branch | 1 | 1 | 1 | 1 |
| Super Market Customer | 0.494 | 0.487 | 0.494 | 0.514 |
| Super Market Date | 0.475 | 0.368 | 0.496 | 0.512 |
| Super Market Payment | 0.492 | 0.497 | 0.489 | |
| Super Market Product | 0.523 | 0.526 | 0.511 | 0.762 |
| Super Market Time | 0.497 | 0.494 | 0.5 | 0.572 |
| Walmart_Sales | 0.974 | | 0.961 | 0.993 |

The TP Rate information for the eight features—city, gender, branch, customer, date, payment, product, and time—is shown in Table 4. We can see that our recommended approach produces better outcomes. The Walmart sales data is shown in the last column of Table 4, and the recommended algorithm yields results that are competitive.

**Table 4:** TP Rate Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 0.998 | 1 | 1 |
| Super Market Gender | 0.814 | 0.501 | 0.903 | 1 |
| Super Market Branch | 1 | 0.998 | 1 | 1 |
| Super Market Customer | 0.924 | 0.496 | 0.93 | 1 |
| Super Market Date | 0 | 0 | 0 | 0 |
| Super Market Payment | 0.509 | 0.347 | 0.458 | 0 |
| Super Market Product | 0.214 | 0.186 | 0.114 | 0.97 |
| Super Market Time | 0 | 0 | 0 | 0 |
| Walmart_Sales | 0.944 | | 0.712 | 0.23 |

The FP rate information for the eight features—city, gender, branch, customer, date, payment, product, and time—is shown in Table 5. We can see that our recommended approach produces better outcomes. The Walmart sales data is shown in the last column of Table 5, and the recommended algorithm yields results that are competitive.

**Table 5:** FP Rate Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 0 | 0.003 | 0 | 0 |
| Super Market Gender | 0.783 | 0.524 | 0.91 | 0 |
| Super Market Branch | 0 | 0.001 | 0 | 0 |
| Super Market Customer | 0.934 | 0.515 | 0.941 | 0 |
| Super Market Date | 0.018 | 0.012 | 0.001 | 0 |
| Super Market Payment | 0.526 | 0.36 | 0.469 | 0 |
| Super Market Product | 0.171 | 0.161 | 0.105 | 0.63 |
| Super Market Time | 0.008 | 0.003 | 0 | 0.143 |
| Walmart_Sales | 0.001 | 0.002 | 0.003 | 0.003 |

The TN rate results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 6. We can see that our recommended approach produces better outcomes. Walmart sales data is shown in the last column of Table 6, and the recommended method yields competitive results.

**Table 6:** TN Rate Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 0.997 | 1 | 1 |
| Super Market Gender | 0.217 | 0.476 | 0.09 | 1 |
| Super Market Branch | 1 | 0.999 | 1 | 1 |
| Super Market Customer | 0.066 | 0.485 | 0.059 | 0 |
| Super Market Date | 0.982 | 0.988 | 0.999 | 1 |
| Super Market Payment | 0.474 | 0.64 | 0.531 | 1 |
| Super Market Product | 0.829 | 0.839 | 0.895 | 0.37 |
| Super Market Time | 0.992 | 0.997 | 1 | 0.857 |
| Walmart_Sales | 0.999 | 0.992 | 0.997 | 0.997 |

The FN rate results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 7. We can see that our recommended approach produces better outcomes. Walmart sales data is shown in the last column of Table 7, and the recommended method yields competitive results.

**Table 7:** FN Rate Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 0 | 0.002 | 0 | 0 |
| Super Market Gender | 0.186 | 0.499 | 0.097 | 0 |
| Super Market Branch | 0 | 0.002 | 0 | 0 |
| Super Market Customer | 0.076 | 0.504 | 0.07 | 0 |
| Super Market Date | 1 | 1 | 1 | 1 |
| Super Market Payment | 0.491 | 0.653 | 0.542 | 0 |
| Super Market Product | 0.786 | 0.814 | 0.886 | 0.03 |
| Super Market Time | 0.3 | 1 | 1 | 0.3 |
| Walmart_Sales | 0.056 | | 0.288 | 0.769 |

The precision results of eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 8. We can see that our recommended approach produces better outcomes. The Walmart sales data is shown in the last column of Table 8, and the recommended algorithm yields results that are competitive.

**Table 8:** Precision Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 0.995 | 1 | 1 |
| Super Market Gender | 0.512 | 0.491 | 0.498 | 1 |
| Super Market Branch | 1 | 0.999 | 1 | 1 |

| | | | | |
|---|---|---|---|---|
| Super Market Customer | 0.495 | 0.491 | 0.497 | 1 |
| Super Market Date | 0 | 0 | 0 | 0 |
| Super Market Payment | 0.217 | 0.335 | 0.333 | 0 |
| Super Market Product | 0.204 | 0.193 | 0.181 | 0.325 |
| Super Market Time | 0 | 0 | 0 | 0 |
| Walmart_Sales | 0.877 | 0.889 | 0.638 | 0.296 |

The recall results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 9. We can see that our recommended approach produces better outcomes. The Walmart sales data is shown in the last column of Table 9, and the recommended algorithm yields results that are competitive.

**Table 9:** Recall Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 0.998 | 1 | 1 |
| Super Market Gender | 0.814 | 0.501 | 0.903 | 1 |
| Super Market Branch | 1 | 0.998 | 1 | 1 |
| Super Market Customer | 0.924 | 0.496 | 0.93 | 1 |
| Super Market Date | 0 | 0 | 0 | 0 |
| Super Market Payment | 0.509 | 0.347 | 0.458 | 0 |
| Super Market Product | 0.214 | 0.186 | 0.114 | 0.97 |
| Super Market Time | 0 | 0 | 0 | 0 |
| Walmart_Sales | 0.944 | 0.523 | 0.712 | 0.23 |

The F-measure results for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 10. We can see that our recommended approach produces better outcomes. Walmart sales data is shown in the last column of Table 10, and the recommended method yields competitive results.

**Table 10:** F-Measure Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 1 | 0.997 | 1 | 1 |
| Super Market Gender | 0.622 | 0.495 | 0.638 | 1 |
| Super Market Branch | 1 | 0.998 | 1 | 1 |
| Super Market Customer | 0.636 | 0.492 | 0.645 | 1 |
| Super Market Date | 0 | 0 | 0 | 0 |
| Super Market Payment | 0.296 | 0.34 | 0.432 | 0 |
| Super Market Product | 0.206 | 0.187 | 0.184 | 0.486 |
| Super Market Time | 0 | 0 | 0 | 0 |
| Walmart_Sales | 0.899 | | 0.649 | 0.241 |

The RMSE findings for the eight features—city, gender, branch, customer, date, payment, product, and time—are shown in Table 11. We can see that our recommended approach produces better outcomes. The Walmart sales data is shown in the last column of Table 11, and the recommended algorithm yields results that are competitive.

**Table 11:** RMSE Results on Super Market Dataset

| | C4.5 | RF | REP | Proposed |
|---|---|---|---|---|
| Super Market City | 0 | 0.0 | 0 | 0 |
| Super Market Gender | 0.5 | 0.04 | 0.54 | 0 |
| Super Market Branch | 0 | 0.04 | 0 | 0 |
| Super Market Customer | 0.5 | 0.140 | 0.526 | 0 |
| Super Market Date | 0.13 | 0.249 | 0.106 | 0.104 |
| Super Market Payment | 0.471 | 0.034 | 0.51 | 0 |
| Super Market Product | 0.456 | 0.041 | 0.422 | 0.037 |
| Super Market Time | 0.054 | 0.062 | 0.044 | 0.042 |
| Walmart_Sales | 0.049 | 0.054 | 0.062 | 0.049 |

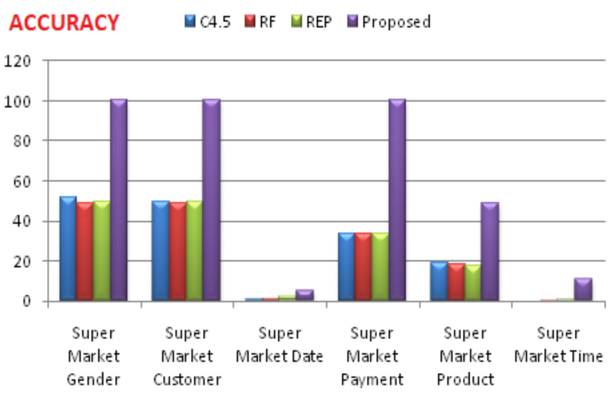


**Fig. 2:** Accuracy Metric Performance Evaluation on the Compared and Proposed Approach.

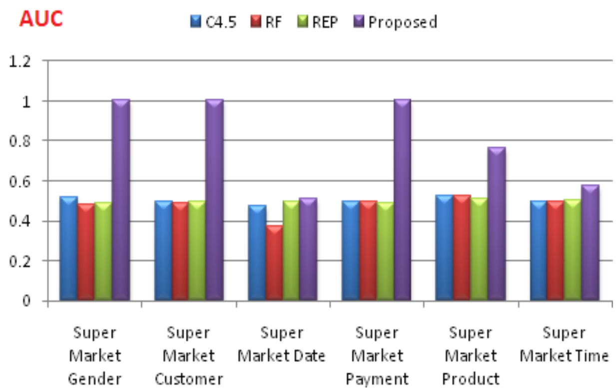


**Fig. 3:** AUC Metric Performance Evaluation on the Compared and Proposed Approach.

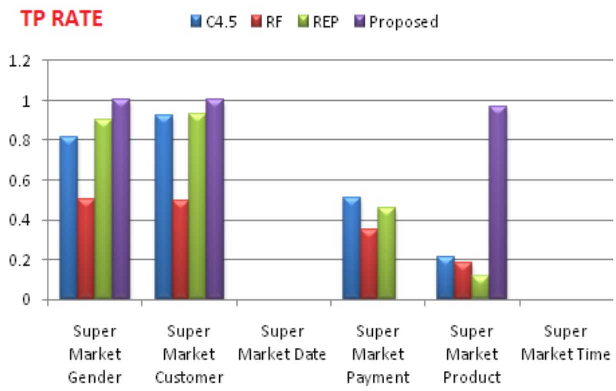


**Fig. 4:** TP Rate Metric Performance Evaluation on the Compared and Proposed Approach.

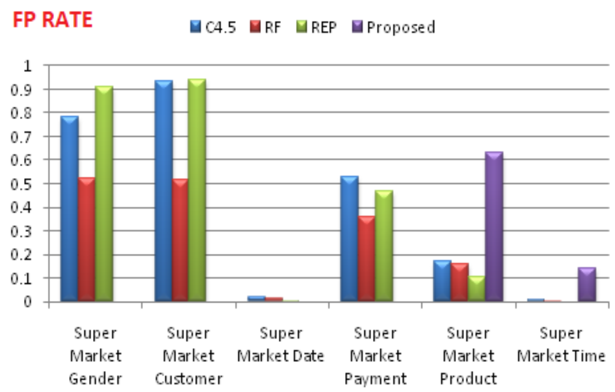


**Fig. 5:** FP Rate Metric Performance Evaluation on the Compared and Proposed Approach.

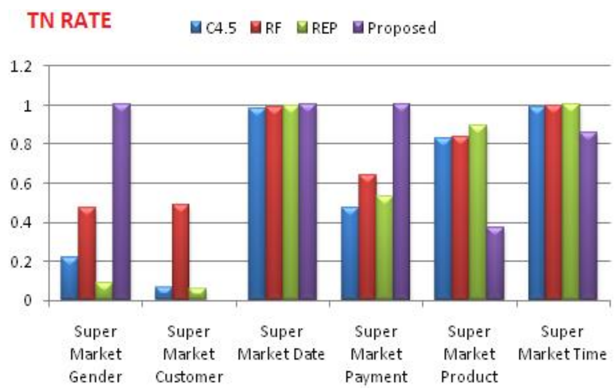


**Fig. 6:** TN Rate Metric Performance Evaluation on the Compared and Proposed Approach.

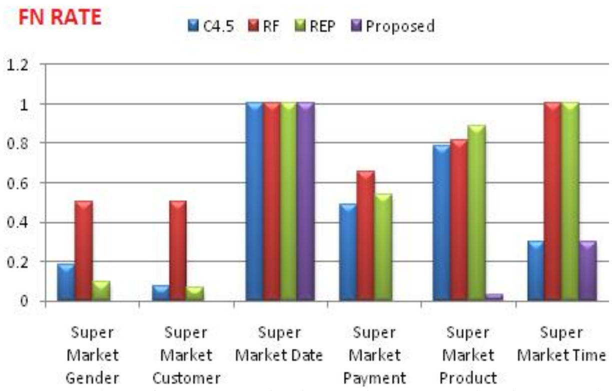


**Fig. 7:** FN Rate Metric Performance Evaluation on the Compared and Proposed Approach.

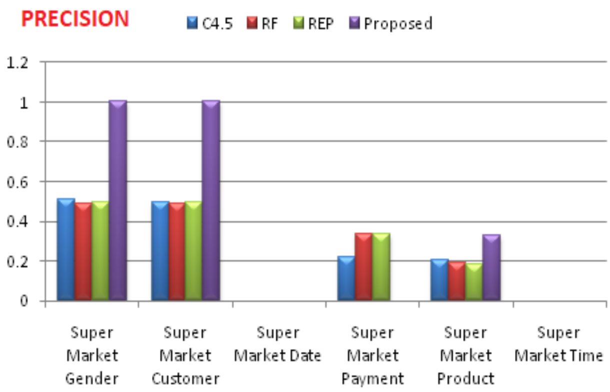


**Fig. 8:** Precision Metric Performance Evaluation on the Compared and Proposed Approach.

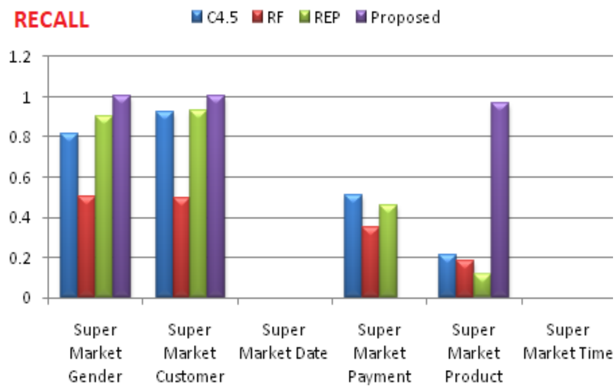


**Fig. 9:** Recall Metric Performance Evaluation on the Compared and Proposed Approach.

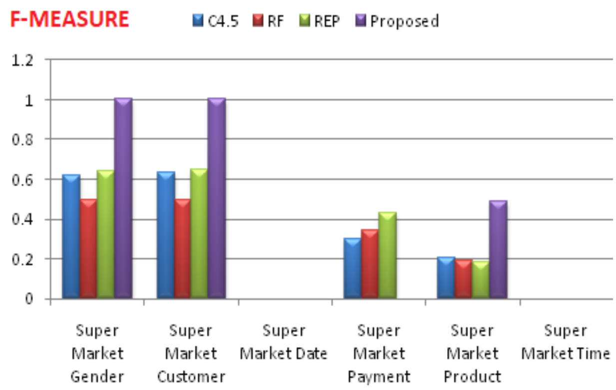


**Fig. 10:** F-Measure Rate Metric Performance Evaluation on the Compared and Proposed Approach.

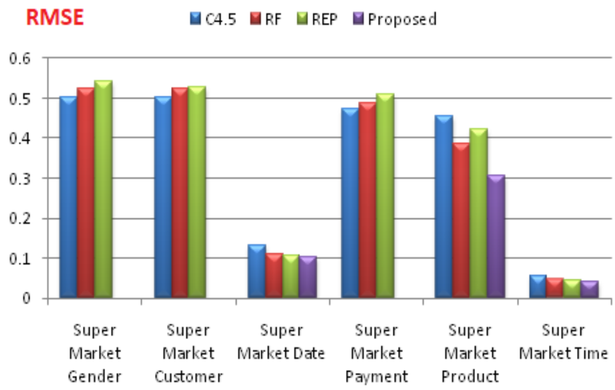

**Fig. 11:** RMSE Metric Performance Evaluation on the Compared and Proposed Approach.

## 7. Discussion

The current study looked for trends in the people and items in a supermarket using classification algorithms and association rules. The C4.5, RF, and REP algorithms were used to classify users and achieve the optimal number of rules. The suggested approach was then used to extract the rules for various features in each class. Finally, suitable methods were taken into consideration to classify the clients in each group based on the characteristics of the clients in each class.

As it comes to using business intelligence and machine learning for consumer segmentation, ethical concerns are paramount. Compliance with rules (e.g., GDPR, CCPA) necessitates handling issues like data protection, and algorithmic bias in features like gender or consumer type might result in unjust results. The suggested model should include privacy-preserving methods, learning that is fair, and explainability to keep decision-making transparent to be in line with ethical AI practices. There has to be a balanced and prudent deployment of personalized offers since, although they do increase engagement, over-targeting poses the danger of manipulating consumers.

Based on the results obtained, four classes are established with the optimal number of regulations for the clients. The following queries are addressed in this study.

1) For business intelligence applications, which subset of machine learning algorithms is most appropriate?
2) How many rules is the optimal number for customer segmentation, and what is the average client for each class?
3) Which products do customers usually buy within each category, and which products are usually offered together?

Sales growth is significantly impacted by the results obtained by using different strategies and the proposed model. The findings show that the proposed strategy performs the best in terms of classification among all the models stated above. The following should also be considered to make better decisions when firms and organizations deploy business intelligence solutions. Obtaining precise information and considering its appropriate relationships with every unit; - identifying the organization's objectives from the project's execution; - coordinating and collaborating amongst organizational units to advance the project. Before choosing to establish and oversee the required culture and training to motivate users, it is crucial to consider the benefits and viability for the company.

One major benefit of using business intelligence in industrialized nations is that it helps improve product sales and distribution. Given how much business intelligence is employed in today's companies, it is recommended that machine learning algorithms be used to enhance product supply and sales performance to increase marketing and profit. Without a doubt, machine learning algorithms and KPIs like AUC, Recall, and Precision can teach managers a lot about the sales sector. It goes without saying that the most effective strategy for every class can significantly boost the company's growth. By user shopping cart evaluation and association rule algorithms, the business also obtains a deeper understanding of each product and product group. The literature that has been published can be used to analyze in-depth the current routes that many scholars in the field of business intelligence have taken [30–31].

There are problems regarding possible overfitting or dependence on dataset-specific properties, despite the suggested model demonstrating remarkable performance across numerous parameters. Precision, recall, and F-measure values are either close to or equal to 1. Perfect scores are uncommon because of factors such as class imbalance, noise, and intrinsic data unpredictability (Shmueli et al., 2023). The model seems to have been fine-tuned for the datasets utilised, which restricts its usefulness outside of the store and Walmart scenarios. To overcome this constraint and guarantee robustness and generalisability, future validation should use datasets from other domains, such as retail, logistics, and finance. Furthermore, external benchmarking and layered cross-validation might mitigate the possibility of overly optimistic performance estimations. Additional methods to reduce bias and avoid overfitting include regularisation, dropout (in neural networks), and ensemble diversity tests. In the end, it is important to conduct longitudinal tests before releasing the model to the public to make sure it can keep up with changing consumer habits and market conditions. Critically recognising these dangers enhances the study's credibility and lays out a plan to expand the use of the recommended BI technique.

## 8. Conclusion

One of the factors affecting how economic and commercial activities develop in emerging countries is the creation, assessment, and exploration of data that affects a company's product sales and distribution. Examining how business intelligence affects business enhancement is the general goal of this study. Using machine learning signs and algorithms, this study suggests a novel model to optimize the product sales system using categorization techniques. Several factors are used by the model to enhance client classification techniques. Additionally, the study uses association rules to look at clients' shopping carts, find links between the items they have bought, and generate personalized offers based on the rules they have discovered.

## References


[1] Hlaváč, J., & Štefanovič, J. (2020, January). Machine learning and business intelligence or from descriptive analytics to predictive analytics. In *2020, Cybernetics & Informatics (K&I)* (pp. 1-4). IEEE. https://doi.org/10.1109/KI48306.2020.9039874.

[2] Bayer, H., Aksogan, M., Celik, E., & Kondiloglu, A. (2017). Big data mining and business intelligence trends. *Journal of Asian Business Strategy, 7*(1), 23. https://doi.org/10.18488/journal.1006/2017.7.1/1006.1.23.33.

[3] Howson, C. (2007). *Successful business intelligence: Secrets to making BI a killer app*. McGraw-Hill Osborne Media.

[4] Tripathi, A., Bagga, T., & Aggarwal, R. K. (2020). Strategic impact of business intelligence: A review of literature. *Prabandhan: Indian Journal of Management*, *13*(3), 35-48. https://doi.org/10.17010/pijom/2020/v13i3/151175.

[5] Cheng, C., Zhong, H., & Cao, L. (2020). Facilitating speed of internationalization: The roles of business intelligence and organizational agility. *Journal of Business Research*, *110*, 95-103. https://doi.org/10.1016/j.jbusres.2020.01.003.

[6] Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(1), 470-477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477.

[7] Tutunea, M. F., & Rus, R. V. (2012). Business intelligence solutions for SME's. *Procedia economics and finance*, (3), 865-870. https://doi.org/10.1016/S2212-5671(12)00242-0.

[8] Masa'Deh, R. E., Obeidat, Z., Maqableh, M., & Shah, M. (2021). The impact of business intelligence systems on an organization's effectiveness: the role of metadata quality from a developing country's view. *International Journal of Hospitality & Tourism Administration*, *22*(1), 64-84. https://doi.org/10.1080/15256480.2018.1547239.

[9] Gutnik, S. (2021). Application of data mining and machine learning methods to enhance the effectiveness of digital marketing strategies. In *Digital Strategies in a Global Market: Navigating the Fourth Industrial Revolution* (pp. 131-144). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58267-8_10

[10] Rath, M. (2021). Realization of business intelligence using machine learning. *Internet of Things in Business Transformation: Developing an Engineering and Business Strategy for Industry 5.0*, 169-184. https://doi.org/10.1002/9781119711148.ch10.

[11] Whig, P. (2019). Artificial intelligence and machine learning in business. *International Journal on Integrated Education*, *2*(2), 334128.

[12] Somayaji, S. R. K., Alazab, M., Bucchiarone, A., Chowdhary, C. L., & Gadekallu, T. R. (2020, December). A framework for prediction and storage of battery life in IoT devices using DNN and blockchain. In *2020 IEEE Globecom Workshops (GC Wkshps* (pp. 1-6). IEEE. https://doi.org/10.1109/GCWkshps50303.2020.9367413.

[13] Khare, N., Devan, P., Chowdhary, C. L., Bhattacharya, S., Singh, G., Singh, S., & Yoon, B. (2020). Smo-dnn: Spider monkey optimization and deep neural network hybrid classifier model for intrusion detection. *Electronics*, *9*(4), 692. https://doi.org/10.3390/electronics9040692.

[14] Parimala, M., Swarna Priya, R. M., Praveen Kumar Reddy, M., Lal Chowdhary, C., Kumar Poluru, R., & Khan, S. (2021). Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach. *Software: Practice and Experience*, *51*(3), 550-570. https://doi.org/10.1002/spe.2851.

[15] RM, S. P., Maddikunta, P. K. R., Koppu, S., Gadekallu, T. R., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. *Computer Communications*, *160*, 139-149. https://doi.org/10.1016/j.comcom.2020.05.048.

[16] Reddy, T., RM, S. P., Chowdhary, C. L., Hakak, S., & Khan, W. Z. (2020). A deep neural networks based model for uninterrupted marine environment monitoring. *Computer Communications*, *157*, 64-75. https://doi.org/10.1016/j.comcom.2020.04.004.

[17] Reddy, G. T., Bhattacharya, S., Ramakrishnan, S. S., Chowdhary, C. L., Hakak, S., Kaluri, R., & Reddy, M. P. K. (2020, February). An ensemble based machine learning model for diabetic retinopathy classification. In *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)* (pp. 1-6). IEEE. https://doi.org/10.1109/ic-ETITE47903.2020.235.

[18] Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications*, *36*(3), 4176-4184. https://doi.org/10.1016/j.eswa.2008.04.003.

[19] Mahmood, A. M., & Kuppa, M. R. (2010, December). Early detection of clinical parameters in heart disease by improved decision tree algorithm. In *2010 second Vaagdevi international conference on information technology for real world problems* (pp. 24-29). IEEE. https://doi.org/10.1109/VCON.2010.12.

[20] Mahmood, A. M., Rao, K. M., & Reddi, K. K. (2010). A novel algorithm for scaling up the accuracy of decision trees. *International Journal on Computer Science and Engineering*, *2*(2), 126-131.

[21] Mahmood, A. M., Kuppa, M. R., & Reddi, K. K. (2010). A new decision tree induction using composite splitting criterion. *Journal of Applied Computer Science & Mathematics*, *9*(4), 69-74.

[22] Reddi, K. K., Mahmood, A. M., & Rao, K. M. (2010). Generating optimized decision tree based on discrete wavelet transform. *(IJEST) International Journal of Engineering Science and Technology*, *2*(3), 157-164.

[23] Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

[24] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324.

[25] Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, *27*(3), 221-234. https://doi.org/10.1016/S0020-7373(87)80053-6.

[26] Mahmood, A. M., & Kuppa, M. R. (2012). A novel pruning approach using expert knowledge for data-specific pruning. *Engineering with Computers*, *28*(1), 21-30. https://doi.org/10.1007/s00366-011-0214-1.

[27] Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005, June). Practical machine learning tools and techniques. In *Data mining* (Vol. 2, No. 4, pp. 403-413). Amsterdam, The Netherlands: Elsevier.

[28] Al-Quhfa, H., Mothana, A., Aljbri, A., & Song, J. (2024). Enhancing talent recruitment in business intelligence systems: a comparative analysis of machine learning models. *Analytics*, *3*(3), 297-317. https://doi.org/10.3390/analytics3030017.

[29] Mohammed, R., Shihab, I., & Musa, M. (2024). Social Media Data-Based Business Intelligence Analysis Using Deep Learning. *International Journal of Intelligent Engineering & Systems*, *17*(1). https://doi.org/10.22266/ijies2024.0229.29.

[30] Sarkar, M., Jewel, R. M., Chowdhury, M. S., Al-Imran, M., Sawalmeh, R. S., Ray, R. K., & Ghosh, S. K. (2024). Revolutionizing organizational decision-making for stock market: A machine learning approach with CNNs in business intelligence and management. Journal of Business and Management Studies, 6(1), 230. https://doi.org/10.32996/jbms.2024.6.1.16.

[31] Gazi, M. S. (2024). Optimizing regional business performance: leveraging business and data analytics in logistics & supply chain management for USA's sustainable growth. *Journal of Business and Management Studies*, 6(2), 144. https://doi.org/10.32996/jbms.2024.6.2.14.

[32] Bharadiya, J. P. (2023). Artificial intelligence and the future of web 3.0: Opportunities and challenges ahead. *American Journal of Computer Science and Technology*, *6*(2), 91-96.

[33] Shmueli, G., Bruce, P. C., Deokar, K. R., & Patel, N. R. (2023). *Machine learning for business analytics: Concepts, techniques, and applications with analytic solver data mining*. John Wiley & Sons.

[34] Bharadiya, J. P., Tzenios, N. T., & Reddy, M. (2023). Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches. *Journal of Engineering Research and Reports*, *24*(12), 29-44. https://doi.org/10.9734/jerr/2023/v24i12858.

[35] Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Ríos Prado, A. (2021). Business intelligence: business evolution after industry 4.0. *Sustainability*, *13*(18), 10026. https://doi.org/10.3390/su131810026.

[36] Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., ... & Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE access*, *8*, 116013-116023. https://doi.org/10.1109/ACCESS.2020.3003790.

[37] JasminPrafulBharadiya, *The Future of Cybersecurity: How artificial Intelligence Will Transform the Industry.*

[38] Bharadiya, J. (2023). A comprehensive survey of deep learning techniques natural language processing. *European Journal of Technology*, *7*(1), 58-66. https://doi.org/10.47672/ejt.1473.

[39] Chu, M. K., & Yong, K. O. (2021). Big data analytics for business intelligence in accounting and audit. *Open Journal of Social Sciences*, *9*(9), 42-52. https://doi.org/10.4236/jss.2021.99004.

[40] Hindle, G., Kunc, M., Mortensen, M., Oztekin, A., & Vidgen, R. (2020). Business analytics: Defining the field and identifying a research agenda. *European Journal of Operational Research*, *281*(3), 483-490. https://doi.org/10.1016/j.ejor.2019.10.001

[41] Bharadiya, J. (2023). Convolutional neural networks for image classification. *International Journal of Innovative Science and Research Technology*, *8*(5), 673-677.

[42] Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: a survey. *IEEE Transactions on Industrial Informatics*, *8*(3), 679-687. https://doi.org/10.1109/TII.2012.2188804.