Application of Data Mining Techniques in Enterprise Decision Support Systems

Dawei Zhao[1,*], Nianchang Yu[2,a]

*dawei.zhao@lpunetwork.edu.ph, [a]yu_nc@hdec.com

[1]Lyceum of the Philippines University, Manila
[2]Powerchina Huadong Engineering Corporation Limited, Hangzhou, China

Abstract: This paper delves into the application of data mining techniques in enterprise decision support systems and methods for evaluating their effectiveness. The article first provides a comprehensive theoretical framework by introducing the fundamental theories and background knowledge of data mining. It then analyzes the specific applications of data mining technology in the realm of business decision-making through various practical cases, such as market segmentation, product sales trend analysis, and customer relationship management. These cases not only demonstrate the practical utility of data mining technology but also highlight its diversity of applications in different business scenarios. The latter part of the article focuses on the evaluation framework for the effectiveness of data mining applications, showcasing the strengths and weaknesses of various data mining techniques through comparative experiments and providing a set of evaluation criteria. The article aims to combine theory and practice, offering valuable references and guidance for enterprises in the selection and optimization of data mining solutions.

Keywords: Data Mining, Decision Support Systems, Effectiveness Evaluation, Market Segmentation, Product Sales, Customer Relationship Management

1. Introduction

Data mining techniques play a crucial role in supporting intelligent decision-making for enterprises in the era of big data, with applications in various industries like retail, banking, and telecommunications [1]. These technologies extract valuable insights from massive data, such as analyzing retail consumer behavior, predicting credit risks, and studying customer usage habits. However, their success relies on considering business characteristics, data quality, algorithm selection, and human expertise. Establishing feedback mechanisms is also essential for optimization. This paper explores data mining's applications in enterprise decision support systems, offering practical insights to different industries aiming to gain a competitive edge in the age of big data.

BDEDM 2024, January 12-14, Ningbo, People's Republic of China

DOI 10.4108/eai.12-1-2024.2347185

2. Theoretical Foundation and Background of Data Mining Techniques

Data mining in enterprise decision support systems is built on a strong theoretical foundation, drawing primarily from statistics, machine learning, and databases[2]. Statistics provides the basis for discovering patterns in large datasets using methods like regression analysis, clustering, and association rules. Machine learning contributes by supporting predictive modeling with algorithms such as decision trees, neural networks, and support vector machines for tasks like classification and regression. Database technology plays a crucial role in collecting, storing, and managing massive data, with database management systems and data warehouses serving as essential platforms. This combination of statistics, machine learning, and databases forms the core of data mining technology, enabling enterprises to uncover data patterns and gain business intelligence for informed decision-making [3].

3. Data Mining Techniques in Decision Support

3.1. Data Preprocessing and Feature Engineering

Before applying data mining in enterprise decision support systems, it's crucial to preprocess raw data to enhance data quality. This preprocessing includes cleaning, outlier removal, deduplication, and other steps[4]. Simultaneously, feature engineering is essential, as it significantly affects mining results' quality. For example, when dealing with user data, one can extract statistical features like user age distribution, behavioral features like purchase frequency, and content features like product descriptions. In the context of an online shopping platform, features such as "user age," "browsing product categories," and "browsing duration" can be derived from user click and purchase data [5]. Feature engineering should consider business requirements and data characteristics and may involve feature selection, mapping, and combination to enhance model learning and predictive performance.

3.2. Application of Data Mining Algorithms

On the preprocessed dataset, various data mining algorithms can be applied to discover patterns and build models [6]. For example, classification algorithms can be used to categorize users into different types, and their formula can be represented as:

$$f(A) = B \tag{1}$$

Where A represents the input user features, and B represents the user's category. Clustering algorithms are employed to identify user groups, and their core formula can be:

$$\min \sum_{i=1}^{k} \sum_{a \in C_i} ||a - m_i||^2 \tag{2}$$

Where C_i is the ith cluster, and m_i is the center of that cluster. Frequent itemset mining algorithms like the Apriori algorithm have a core formula:

$$\sup(X) = \frac{X}{Z} \tag{3}$$

Where X is the count, and Z is the total number of transactions. Utilizing these techniques, an online store can train a purchase intent prediction model, i.e., given user features A, predict the probability P(B|A) of them buying a particular product. Decision tree models can be simplified into a series of conditional statements, for example, if (I>T) and (N=0), then (preference=H). Data mining technology enables the quantification of complex relationships and a deeper understanding of user behavior. However, the selection and optimization of algorithms require the expertise of data scientists. In a business environment, data mining can uncover patterns to enhance decision-making, but the ultimate business strategy still relies on the judgment of management.

3.3. Methods of Data Mining in Market Segmentation

Market segmentation involves identifying customer groups using data mining techniques,as shown in Figure 1. For example, analyzing customer data like age, occupation, and consumption behavior helps identify groups like "Fashion Enthusiasts" and "Budget-conscious Families." These insights enable customized marketing and product strategies for each group, improving overall business performance[7].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{4}$$

Where Y represents market response, X_i represents different customer features, β_i represents coefficients, and ϵ represents the error term.

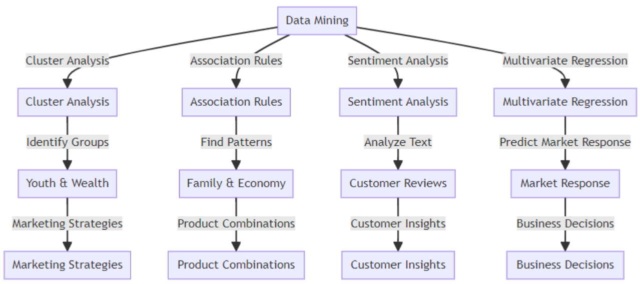

Fig. 1. Methods of Data Mining in Market Segmentation.

3.4. Methods of Data Mining in Product Sales and Promotion Decision-Making

Data mining enables businesses to increase product sales and optimize promotion decisions. For instance, analyzing user online behavior and transaction data may reveal that promotions have a weak impact on sustaining purchase intent, and aggressive promotions can actually reduce brand loyalty [8].

$$A = f(B, C) \tag{5}$$

Where A represents the influence of promotional activities on sales, B represents user behavioral data, and C represents transaction data.

Similarly, using user rating data and review data, a company may discover that functional products garner more trust from users based on word-of-mouth and reviews. Consequently, resources are allocated to improving the quality of these products rather than increasing promotional efforts.

$$D = g(E, F) \tag{6}$$

Where D represents the enhanced brand trust based on user ratings and reviews, E represents user rating data, and F represents user review data.

Additionally, conducting product bundle association analysis can reveal that bundling specific product combinations significantly enhances overall conversion rates.

$$G = h(I) \tag{7}$$

Where G represents the improved conversion rate through product bundling, I represents different product combinations. By employing these analytical methods in combination, businesses can formulate more effective sales strategies:

$$J = k(A, D, G) \tag{8}$$

Where J represents the overall sales strategy, k represents the integrated analytical method, combining metrics such as sales volume, brand trust, and conversion rate.

The functions f, g, h, and k in these equations represent various analytical methods and logical relationships, illustrating how data mining can be used to optimize product sales and promotion decision-making.

3.5. Methods of Data Mining in Customer Relationship Management

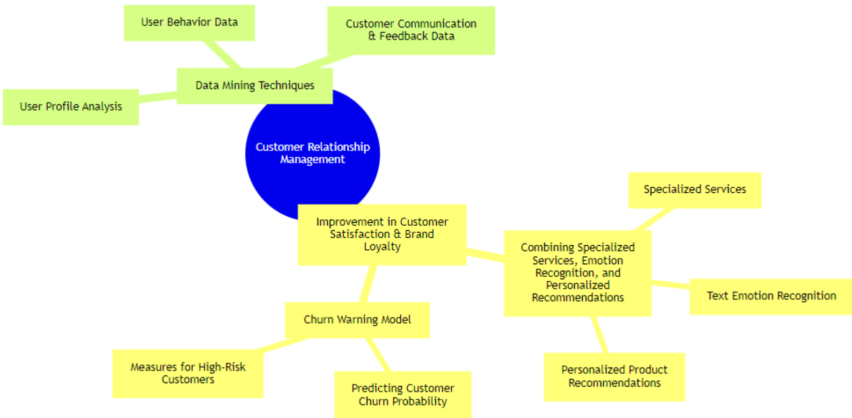

Fig. 2. Data Mining Methods in Customer Relationship Management.

Through data mining, businesses can gain a deeper understanding of their customers, provide personalized services, and establish long-term relationships, as depicted in Figure 2. For example, by utilizing user profiling analysis, they can offer exclusive service representatives and purchase discounts to high-value customers. By training text sentiment recognition models using customer communication and feedback data, they can proactively address customer sentiments. Furthermore, by using user behavioral data, they can recommend products of interest, enhancing personalized shopping experiences [9]. These initiatives can lead to increased customer satisfaction, improved brand perception, and enhanced loyalty.

$$G = i(A, C, E) \tag{9}$$

Where G represents the enhancement of customer satisfaction and brand loyalty, i represents the method that combines exclusive services, sentiment recognition, and personalized recommendations. Additionally, businesses can implement churn prediction models internally to forecast the likelihood of customer attrition and take retention measures for high-risk customers. By comprehensively utilizing data mining techniques to strengthen customer relationships, enterprises can significantly boost their core competitiveness.

4. Evaluation of the Application Effectiveness of Data Mining Techniques in Enterprise Decision Support

4.1. Presentation of Experimental Results in Data Mining Applications

After the application of data mining techniques, it is typically necessary to test the effectiveness of the models through online experiments or offline evaluations. Taking an example of an e-commerce recommendation system, this system needs to respond to users' queries in real-time and provide personalized recommendations [10]. In such a scenario, A/B experiments can be designed, where the control group uses a basic recommendation algorithm, while the experimental group employs a more complex deep learning recommendation algorithm. Online traffic is allocated to these two control groups in a certain proportion. During the experiment, user behavior data such as clicks, favorites, and additions to the cart are recorded. After a period of experimentation, the click-through rates for each group are compared, as shown in Table 1:

Table 1. Comparison of Click-Through Rates for User Groups.

User Group	Clicks A (Basic Recommendation)	Clicks B (Advanced Recommendation)
Big Customers	1.2 million	1.5 million
Regular Customers	800,000	1 million

Based on the experimental results, the deep learning algorithm improved click-through rates, leading the e-commerce platform to adopt it for its business value. This case underscores the significance of online experiments, enabling enterprises to assess model performance using real user data, leading to more informed decisions. This approach ensures a better user experience and greater business value in production.

4.2. In-Depth Analysis of Application Effectiveness and Method Comparison

During the evaluation phase, businesses need to compare the effectiveness of different data mining algorithms to determine the model that best suits their specific problem or task. For example, consider a bank conducting credit risk modeling, where they compared the classification performance of three models: logistic regression, neural networks, and gradient boosting decision trees (GBDT). Below are some metrics and the comparison results as shown in Figure 3:

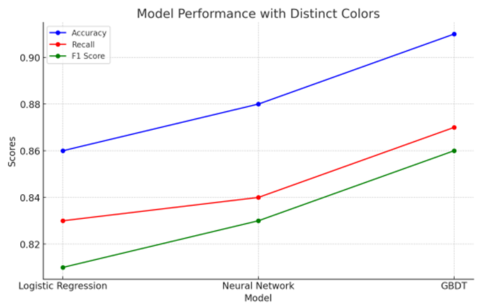

Fig. 3. Comparison of Classification Performance of Different Models.

Based on Figure 3 results, GBDT stands out as the best performer. Further research revealed that GBDT's ability to automatically combine weak classifiers makes it well-suited for complex tasks like credit risk assessment. Consequently, the bank chose to implement the GBDT model for credit decisions. This emphasizes the importance of a comprehensive algorithm selection process aligned with business needs. The evaluation phase of data mining and machine learning is critical for optimizing model performance and enhancing the accuracy of business decisions.

5 Conclusion

Applying data mining techniques in business, especially in areas like market segmentation, product sales, and customer relationship management, is vital for enterprise decision-making. However, effective utilization involves more than just applying algorithms. It entails analyzing business requirements, designing tailored feature engineering solutions, and experimenting with different methods. Furthermore, establishing a continuous evaluation mechanism is crucial for monitoring and optimizing data mining outcomes. To maximize its value in complex market environments, data mining technology must be closely integrated with specific business needs and data environments, along with continuous experimentation, evaluation, and optimization efforts.

References

[1] Huang Z L Y .Research of data mining and web technology in university discipline construction decision support system based on MVC model[J].Library hi tech, 2020, 38(3).

[2] Chao L .Application of Data Warehouse and Data Mining Technology in Decision Support System[J].Journal of Tianjin Vocational Institutes, 2019.

[3] Sulhi A .Data Mining Technology Used in an Internet of Things-Based Decision Support System for Information Processing Intelligent Manufacturing[J].IJIIS: International Journal of Informatics and Information Systems, 2021.

[4] Ohashi K , Mendori T , Kodama H .Investigation of principal factor decision support system using data mining methodology for surface grinding wheel[J].International Journal of Abrasive Technology, 2019, 9(4):303.

[5] Akhmedyanova G F .Application of OLAP-technology in decision support system for enterprise management[J].IOP Publishing, 2020.

[6] F. An, B. Zhao, B. Cui and Y. Chen, "Selective Virtual Synthetic Vector Embedding for Full-Range Current Harmonic Suppression of the DC Collector," in IEEE Transactions on Power Electronics.
[7] F. An, B. Zhao, B. Cui and Y. Ma, "Asymmetric Topology Design and Quasi-Zero-Loss Switching Composite Modulation for IGCT-Based High-Capacity DC Transformer," in IEEE Transactions on Power Electronics.
[8] F. An, B. Zhao, B. Cui and Y. Chen, " DC Cascaded Energy Storage System Based on DC Collector with Gradient Descent Method," in IEEE Transactions on Industrial Electronics.
[9] F. An, B. Zhao, B. Cui and R. Bai, "Multi-Functional DC Collector for Future All-DC Offshore Wind Power System: Concept, Scheme, and Implement," in IEEE Transactions on Industrial Electronics, 2022.
[10] F. An, W. Song, K. Yang, S. Yang and L. Ma, "A Simple Power Estimation with Triple Phase-Shift Control for the Output Parallel DAB DC-DC Converters in Power Electronic Traction Transformer for Railway Locomotive Application," in IEEE Transactions on Transportation Electrification, 2019.