e-ISJN: A4372-3114
p-ISJN: A4372-3115

ISSN: 2321-7782 (Online)
ISSN: 2347-1778 (Print)
Impact Factor: 7.529

Volume 11, Issue 4, April 2023

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Data Mining Applications in Business Intelligence: An Analysis in the Modern Globalization Era

Rina
Open Scholar,
Masters and UGC NET in Computer Science,
MDU, Rohtak, India.

Abstract: ***Data mining is a valuable asset for businesses, enabling them to increase profits and reduce marketing and manufacturing expenses. By analyzing historical data, companies can uncover hidden patterns and information, leading to better business decisions and strategies. Data mining is often used as a "Swiss army knife" for businesses to increase revenue, reduce expenses, retain existing customers, and identify prospective consumers. However, large-sized data requires special consideration, as data mining algorithms are typically designed for small-sized data only. Big data analytics can help address this issue by identifying essential features and reducing the complexity of algorithms when dealing with high-dimensional data. Data noise, which is incorrect and invalid, complicates model identification and decreases model precision. Data preprocessing techniques can be used to eradicate data disturbance, but most data mining techniques are designed for static data. Outlier data, which is significantly different from the rest of the data, degrades classification and clustering techniques' performance. To produce accurate results, algorithms must correctly manage outlier data. Identifying knowledge contained within outliers can be challenging, and real datasets often contain missing data. Pre-processing techniques can be used to replace missing values with appropriate values. Overall, data mining is a valuable tool for businesses to improve their operations and stay competitive in the market.***

Keywords: Data Mining, Business Decisions, Data, Business Intelligence, Computer Science, Computer Applications.

Content

For all manufacturers, enterprises, and industries, data is a novel asset. Utilising data mining for business intelligence is possible. Diverse business disciplines can use various data mining tools to increase profits and decrease marketing and manufacturing expenses. Until historical data stored in the company's database is converted into valuable information and knowledge that helps the company make better business decisions and strategies that give it a competitive advantage over its rivals, it is equivalent to waste. Tools and techniques for data mining enable us to analyse the data and uncover concealed patterns and information within the data. Data mining enables the business to conduct an effective market analysis, gain a deeper understanding of consumer feedback on similar products, evaluate the strategies and business policies of competitors, and retain loyal customers, all of which result in intelligent business decisions. Data mining is currently utilised as a "Swiss army knife" for business organisations. The majority of businesses use data mining to increase revenue, reduce expenses, retain existing customers, identify prospective consumers, etc.

The Extract, Transform, and Load phase requires the extraction of data from multiple sources. Following the application of various preprocessing techniques, the unprocessed data are converted into a suitable format. Transfer all transformed data to a centralised location. Implementing various data analysis techniques (e.g., OLAP and OLTP) and storing all data in a multidimensional data format (e.g., a data cube) are required to store and manage data. The business analyst organises the stored

data so that the end user can access it efficiently and swiftly according to their needs. Then The data is analysed according to the needs of the end consumers. Finally, results are presented in a visually appealing and user-friendly format (charts, graphs, diagrams, etc.) to facilitate comprehension of the extracted knowledge.

Knowledge discovery in databases (KDD) facilitates the extraction of relevant information from large databases. KDD involves multiple stages, such as data collection, preprocessing, data extraction, and post-processing. Data mining is therefore one of the primary processes that follow data preprocessing. In the step of data preprocessing, data selection, cleansing, preparation, and transformation are carried out. In the subsequent step of data mining, knowledge is extracted from the preprocessed data, followed by evaluations and the presentation of the results. To implement data mining techniques, multiple disciplines, such as statistics, algorithms, information science, machine learning, pattern recognition, database technology, soft computing, and visualisation, collaborate. Various statistical and artificial intelligence methods are used to derive knowledge from the database in order to implement various data mining duties. Various other fields, including computer science (data structure, algorithm, database management system), mathematics (optimisation techniques, linear algebra, set theory, calculus, etc.), information science, soft computing, and visualisation techniques (pie chart, bar plot, histogram, scatter diagram, etc.), are applied to support both of the aforementioned techniques.

Various algorithms are utilised in data mining to attain distinct duties. The algorithm analyses the data and determines a model that corresponds to the data's characteristics. The data model can be either descriptive or predictive. A predictive model is utilised to predict the values of unlabeled data without a class label. Classification (decision trees, regression, artificial neural network, SVM, k-NNN), regression (linear, logistic), and time series analysis are examples of predictive models. Using descriptive models, the data's structure and relationships are uncovered. Clustering k-Means, k-Medoidsids, k-Mode, DBSCAN, and hierarchical clustering are examples of descriptive data mining. Usclustering identifies ring groups based on information present in the data. Due to the fact that clustering does not require a class label, the unsupervised classification, aion, model is developed using training data with class labels (hence classification is a supervised technique). Next, we use the created model to classify any unknown data without a class label. A few examples of classification techniques include artificial neural networks, support vector machines, and decitrees. Trinds of dynamic data instances are identified by the evolution analysis.Using association mining, a relationship between the records is identified. The fundamental mission of association mining is rule generation g the cover and support. Outliers, or records that deviate significantly from the majority of records, can be identified through data mining techniques. Using outmining, various knowledge and patterns of interest can be identified.

Customer Relationship Management inherently prioritises repeat customers. Implementing customer-focused business policies and strategies is crucial for acquiring, retaining, and expanding customer base. A business must analyse data to obtain insight for this purpose. Here is where data mining comes into action. Using data mining techniques, manufacturers can identify the purchase behaviour, immediate likes and dislikes, and demands of consumers, and based on this information, they can modify or develop strategies to retain existing customers. Note that cross-selling and up-selling are two prominent CRM-based instruments for boosting profitability. Data mining is used to determine what to offer to whom and at what time. In addition, business analysts assist managers in identifying prospective customers who are oblivious of existing or new products and services by utilising data mining. Finally, managers are able to identify marketing strategies and offers that are suitable for such consumers. Using data mining techniques, it is possible to identify customer segmentation. The primary objective of segmentation is to supply products and services in accordance with client preferences. In addition, businesses can create personalised marketing messages for each consumer segment. Data mining employs clustering techniques to identify consumer behavioural segments. Data mining also aids in identifying segments of vulnerable consumers, allowing for their retention and satisfaction to be increased by catering to their needs.

Data mining techniques, such as clustering and classification, enable the identification of the latent factors that influence consumers' decisions regarding a specific brand or product. Using data mining, enhanced target marketing strategies can be

implemented to increase product sales and the company's profit. Through market basket analysis, we can identify groups of frequently purchased items. Using data mining techniques (e.g., association mining), it is possible to identify a group of items that are frequently purchased together by consumers. This information may aid the retailer in comprehending the requirements and purchasing patterns of customers and modifying the store's layout by positioning items accordingly. Using data mining algorithms, e-commerce websites can provide consumers with personalised recommendations based on their searches or purchases. A data mining algorithm analyses a customer's past purchases in order to recommend additional products.

Large-sized data require special consideration, as data mining algorithms are typically developed for small-sized data only. Algorithms for data mining that are not line-time intricate are inefficient at managing vast quantities of data. Notate that big data analytics is one of the potential solutions to this problem. (i) not all features are essential for data mining, and (ii) the complexity of the algorithms may increase when using high-dimensional data. As potential solutions, we may employ feature selection and feature transformation to ensure that only essential features are identified and utilised for data mining. In this manner, we simultaneously resolve both issues. Note that selecting the best features for the data mining is not a simple task, and special experience and knowledge are required. The data that is incorrect and invalid is referred to as noise data. The presence of noise in data complicates model identification. It decreases the precision of the models. Techniques for data preprocessing can be used to eradicate data disturbance. Most data mining techniques are designed to process only static data. Therefore, we must modify the existing algorithms or develop new ones in order to effectively manage dynamic data. The data that is significantly different from the remainder of the data is known as outlier data. Outliers degrade the performance of classification and clustering techniques; consequently, algorithms must correctly manage outlier data to produce accurate results. Note that special knowledge can sometimes conceal within outliers, and that identifying the knowledge contained within the outliers is a difficult endeavour. Real datasets typically contain missing data, i.e., missing values are not present in records; however, pre-processing techniques can be used to replace the missing value with an appropriate value.

Non-relevant features may degrade the performance of the data mining algorithm; therefore, we use feature selection methods to identify the algorithm's relevant features. Overfitting issues must be adequately addressed by data mining algorithms (especially in supervised techniques). The causes of overfitting are (a) a small number of representative data; (b) excessive training of the data; (d) the presence of outliers; and (e) a large number of polynomial terms. There are numerous techniques for addressing such overfitting issues, but discussing all of them is beyond the scope of this thesis. Multimedia data consists of various data categories, including video, text, animation, image, and sound, among others. Managing such diverse categories of data to produce a result is a challenging endeavour. Different patterns (bar plot, pie chart, scatter diagram) are used by data mining algorithms to represent the results, and an expert and experienced individual is required to interpret the results.

Clustering enables the identification of data's inherent classifications. It is a technique for dividing a given dataset into categories termed clusters. Data characteristics within a cluster have the highest degree of similarity, whereas data characteristics between clusters have the greatest degree of dissimilarity. In other words, a cluster consists of data that are extremely similar to one another and extremely dissimilar to data in other clusters (also known as automatic classification). Clustering is an unsupervised learning technique as there is no predefined class designation to categorise the data into clusters. Since no class identifiers are transmitted to create the groups, we can say that in clustering, groups are formed by observing the data itself or by comparing the data to one another. Clustering is known as data segmentation in the business world, where significant quantities of business-related data are classified into clusters or segments. To enhance the quality of the clusters, preprocessing techniques (the elimination of incomplete values, transformation, and standardisation) are utilised. Real-world datasets contain missing and noisy data the majority of the time; therefore, to implement clustering techniques, we eliminate missing and noisy data using various statistical and machine learning techniques. In addition, techniques such as feature selection, feature extraction, and feature transformation are employed as part of data preprocessing. Occasionally, post-processing procedures are employed to enhance the legibility and content of the clusters. Popular post-processing operations

include: (a) Outlier detection and removal technique: the elimination of groups with a limited number of data points. (b) merging the groups: minor clusters that are near to one another are merged to create a cluster of an appropriate size. Large clusters are subdivided into smaller, more manageable clusters.

References

1. A.K. Jain, M.N. Murty and P.J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, vol. 31, no. 3.
2. Berry, Michael J. A., Gordon S. Linoff, Data Mining Techniques for Marketing, Sales, and Customer Support,2/e, Wiley Publishing, Inc., Indianapolis, Indiana(2004).
3. Bishnu P. S., Bhattacherjee V.: A Modified K-Modes Clustering Algorithm. PReMI 2013.
4. Bishnu P. S., Prasad S., Bhattacherjee V.: A dimension reduction technique for KMeans clustering algorithm. RAIT 2012.
5. Bishnu P. S., Prasad S., Bhattacherjee V.: Volume-based clustering for arbitrary shaped clusters. Int. J. Comput. Vis. Robotics 3(3) (2013).
6. Capó, M., Pérez, A. & Lozano, J.A. “An efficient K-means clustering algorithm for tall data.” Data Min Knowl Disc 34, (2020).
7. Dunham, Margret H., Data Mining: Introductory and Advanced Topics, Pearson Education, Inc.,(2003).
8. Han, J., Kamber, M.: Data Mining: Concepts and Techniques.3/e, Morgan Kaufmann, New York (2012).
9. Jain, Anil K., Data clustering: 50 years beyond K-means, Pattern Recognition Letters 31 (2010).
10. Kim WC, Mauborgne R (2005) Blue ocean strategy: how to create uncontested market space and make the competition irrelevant. Harvard Business School Press, Boston ISBN: 978-1591396192.
11. Kotler P, Silva GD, Armstrong G, Haque E (2013) Principles of marketing: a south Asian perspective, 13th edn. Pearson, New Delhi.
12. Provost F. and Fawcett T., Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly Media, 2013.

