Predictive Analytics in Business Analytics: Decision Tree

Chee Sun Lee[1], Peck Yeng Sharon Cheang [1]*, Massoud Moslehpour[2,3]

[1]School of Management, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.

[2]Department of Business Administration, Asia University, Taichung, Taiwan

[3] Department of Management California State University, San Bernardino, California, USA

*Corresponding author Email: sharon@usm.my

Received: August 25, 2021; First Revision: September 10, 2021;

Last Revision: December 10, 2021; Accepted: January 02, 2022;

Published: January 13, 2022

Abstract

Purpose: Business Analytics was defined as one of the most important aspects of combinations of skills, technologies and practices which scrutinize a corporation's data and performance to transpire data-driven decision-making analytics for a corporation's future direction and investment plans. In this paper, much of the focus will be given to predictive analytics, which is a branch of business analytics that scrutinize the application of input data, statistical combinations and intelligence machine learning statistics on predicting the plausibility of a particular event happening, forecast future trends or outcomes utilizing on-hand data with the final objective of improving the performance of the corporation. While it has been around for decades, predictive analytics has gained much attention in the late 20th century. This technique includes data mining and big data analytics. Last but not least, the decision tree methodology, a supervised simple classification tool for predictive analytics, is fully scrutinized below for applying predictive business analytics and decision tree in business applications.

Design/ Methodology/Approach: A systematic literature review was conducted in predictive analytics and decision tree. The literature review explains various fields' latest predictive analytics and decision trees. All the research papers are obtained from two databases: Web of Science and Scopus, which are widely acknowledged by the scientific and research communities that contain top-quality peer-reviewed journals.

Findings: This study reviews the application of predictive analytics and decision tree in business decision-making across various fields.

Practical implications: This paper will strongly contribute to providing significant inputs to analysts or researchers in business analytics, predictive analytics and decision tree as it presents recent evidence of the applications of various fields. This review will be in the interest of academics and practitioners in business analytics, especially predictive analytics.

Keywords: Business Analytics (BA), Predictive Analytics (PA), Machine Learning (ML), Decision Tree (DT)

JEL classifications: D70, D79, D81

Introduction

Business analytics (BA) is one of the qualitative methodologies to derive valuable meanings based on data. Statistical methods to help boost business information and business analytics have been applied across many fields such as health care, stock markets, medicine, forecasting, and other areas for making informed business decisions. However, current trends focus on applying BA in predicting analytics (PA). PA is a branch of analytics that scrutinizes the application input data, statistical combinations, and intelligence machine learning statistics to predict a particular event's plausibility and forecast future trends. PA is much focused on forecasting markets and the manufacturing field. However, it would significantly impact the fields if the output were valuable.

Predictive analytics is also often used in the decision tree (DT) as it is deemed a user-friendly predictive tool where users can easily interpret the data. DT is a supervised easy learning algorithm focused on deducing the class or value of target variables according to the machine learning (ML) order trained by the training data (Vlahakis et al., 2020; Sarker, 2021). The approach is easy to use and interpret with simple mathematics without statistical knowledge or complex formulas. This approach is a user-friendly methodology as the data required is easily prepared without computing complex calculations. Moreover, when the variables have been built up, less intervention on data optimization is required.

In this paper, business analytics, predictive analytics, and decision tree will be explained, and their recent application will be provided.

Business Analytics

BA is a mixture of techniques, technologies and applications used to scrutinize a corporation's data and performance to transpire data-driven decision-making analytics for the corporation's future direction and investment plans (Bayrak, 2015; Kristoffersen et al., 2021). Data-driven corporations will manage their data as their corporate assets and actively look for ways to turn it into a competitive advantage against their competitors (Bawack and Ahmad, 2021). In this new era of big data, data-driven analytics is the way forward for major corporations in the manufacturing, information technology, marketing and logistics sectors. They are eager to define consumer spending and behavior to maximize profits (Bibri and Krogstie, 2021).

BA is made up of three types of analytics – descriptive analytics, prescriptive analytics and predictive analytics. Descriptive analytics interprets the historical data sets for a certain timeframe to identify valuable trends and patterns. The process includes drilling down into on-hand data to explore and understand details such as the occurrence of events, the value of operations, and the failure mode (Loeb et al., 2017; Kaur et al., 2018; Ondes, 2021). In general, descriptive analytics can be understood as the process that uses current data to provide insights that can help corporations manage and/or improve their business processes.

Descriptive analytics (DA) has been used to understand the impact of COVID-19, such as characterized workspace, changes in consumer behavior, and marketing, operations and e-supply networks and global value chains for future resilience (Sheng et al., 2021). Besides that, DA has also been used to understand household food insecurity in the African American community based on food conditions, characteristics, and perceptions of residents in this food environment. The data collected and analyzed through descriptive analytics shows that most of the African American community suffers from food insecurity regularly, especially those from lower-income households, those on food-assistance programs, and those without access to a motor vehicle (Jones et al., 2021).

DA was also used in the correlation of weather reports and electricity consumption of academic buildings in Melaka. The authors signify that the locations with higher rainfall show a lower consumption of electricity (Nasaruddin et al., 2021). Moreover, DA was also used to determine the smartphone purchase intention of consumers in Nepal, where price factor plays a significant influence on the purchase intention while brand personality and features do not play a significant role (Rai, 2021).

Prescriptive analytics use mathematically or computationally techniques to obtain an outcome that will give the optimum result in a given scenario to improve performance (Arismendy et al., 2021; Lana et al., 2021). Next, prescriptive analytics also examines opportunities within a decision, correlation in within decisions, influences that affect these decisions with the end goal of producing the finest solution in real-time (Arismendy et al., 2021).

Prescriptive analytics has been used to enhance planning-based sports by reducing human expert cognitive biases that can induce injury while training (Houtmeyers et al., 2021). Likewise, in stock market prediction, prescriptive analytics was used to study the flow pattern of stocks, which could help stockbrokers effectively invest in the stock platform with minimal risk (Meenakshi et al., 2021).

A prescriptive model was developed in the healthcare industry to reduce the risk of 30-day readmission, leading to an annual saving of at least $20 million in the United States by analyzing medical reports of 722,101 patients of general surgery (Bertsimas et al., 2020). Prescriptive analytics was also used to optimize healthcare inventory management by improving the quality of replenishment decisions and reducing the probability of emergency orders based on patients' number, type and length of stay in the ward (Galli et al., 2020).

Predictive analytics (PA) applies statistics to forecast future trends or outcomes with the current on-hand data to improve the corporation's performance. The following section will explain PA in more detail.

Predictive Analytics

Predictive analytics is a branch of analytics that uses input data, statistical combinations and ML statistics on predicting the probability of a particular event happening, forecast future trends or outcomes utilizing on-hand data with the final objective of improving the performance of the corporation (Kumar and Garg, 2018; Davenport et al., 2020; Espadinha-Cruz et al., 2021; Izagirre et al., 2021). It captures the relationship among factors to assess risk from a set of conditions by assigning scores, weightage or parameters to deduce the future trends or outcomes. By applying PA, the corporation will effectively interpret big data for its benefits (de Medeiros et al., 2020; Brynjolfsson et al., 2021).

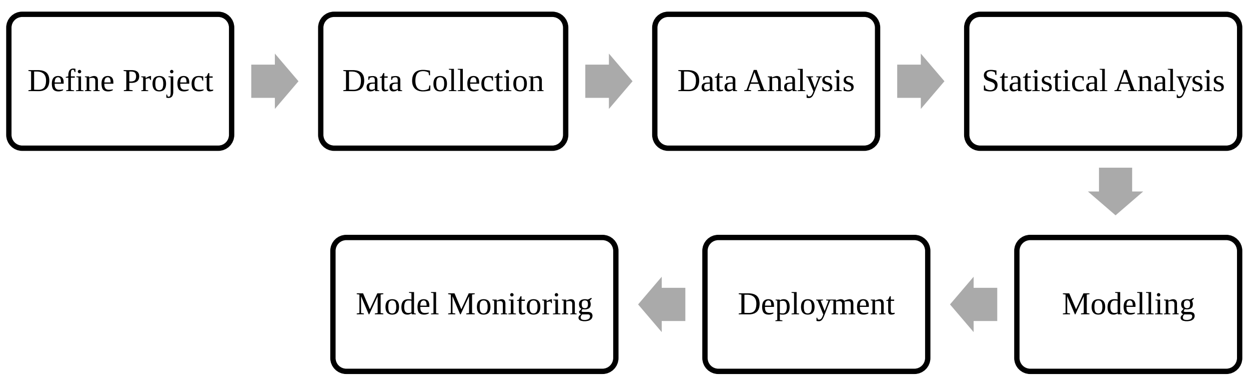
PA methodology allows corporations to be proactive, future-orientated, forecast outputs and behaviors based on data and not by assumptions without any supporting data or information. In addition, PA also suggests actionable instructions to benefit users from its predictions (Javaid et al., 2021; Lo et al., 2021). Moving forward, PA will be integrated into business applications and will no longer be a premium domain of mathematicians and statisticians (Dagnino, 2021; Saxena et al., 2021). Moreover, corporations will make use of PA due to the following reasons:

- Influx size and a class of data
- Utilizing current data to predict or generate valuable outputs and direction
- Higher speed, cost-efficient computers and supercomputers
- User-friendly software
- Harsh economic setting and a need to create a healthy competitive differentiation

Application to PA begins with identifying the project, deliverables, scope, business objectives and dataset for the prediction. The data collection phase is critical to the success of the analytics. Data is typically gathered from various data sources, which must be correlated to create a complete picture of the customers' interactions. Data preparation is then conducted to inspect, clean and transform the data before it undergoes statistical analytics to discover important information. Finally, statical analytics will be performed to validate the hypotheses, and the data will be tested using standard statistical models (Kumar and Garg, 2018; Biecek and Burzykowski, 2021).

After completing the prework for predictive analytics, the process will be continued with modeling, where the user will use the predictive modeling tools to generate accurate predictive models. When the models are in place, a process called deployment in the everyday decision-making process to get results, reports and output through automation (auto-mail/ auto message) based on modeling can be executed to obtain a predictive decision from the model build. Lastly, the model is monitored frequently to ensure the predicted model continuously gives correct predictions. Figure 1 illustrates the PA evaluation procedure (Kumar and Garg, 2018; Biecek and Burzykowski, 2021; Surucu-Balci et al., 2021).

Figure 1: Block Diagram of the Predictive Analytics evaluation procedure



Literature review for the PA is based on the latest published research papers in the area of customer relationship management (CRM), health care, collection analytics (budget planning of agencies or stakeholders), cross-sell (Analyze customer spending, usage, and other patterns), fraud detection, underwriting (predicting the chances of default, bankruptcy, and others), education and manufacturing. The literature review in Table 1 consists of 24 papers on applications of PA in various fields, the PA methodologies used, and the respective contributions.

Table 1. Past studies of predictive analytics (PA) in various fields and the tools used

Author and Year	Field	Predictive Tool Used	Key Takeaways
Sabbeh (2018)	Customer Relationship Management (CRM)	Regression Analytics, DT, Support Vector Machine (SVM), Ensemble Method (Random Forest, Boosting)	Customer retention can be predicted using PA to improve its business
Al-Zuabi et al. (2019)	CRM	Naive Bayes, Ensemble (Random Forest, Boosting), Logistic regression	PA helps in contributing to telecommunications corporation's end-to-end solution that approaches customer data (customer gender and age) from multiple aspects
Wassouf et al. (2020)	CRM	DT, Ensemble Method (Random Forest, Boosting)	PA helped increase customer loyalty and its overall revenue based on the application of classification algorithms onto the loyalty levels
Sneha and Gangil (2019)	Health Care	DT	PA helped in the prediction of the early stage of diabetes mellitus, which helps in preventing the complications of diabetes at the early stage

Kalyankar et al. (2020)	Health Care	Naive Bayes, DT,	PA used in predicting the probability of diabetic patients using Pima Indian diabetic and non-diabetic women
Sharma and Gupta (2021)	Health Care	Bayes Point Machines, Two class Logistic Regression	PA used COVID-19 India big data to assist and support the healthcare department by helping in early predictions
Tolba and Al-Makhadmeh (2021)	Health Care	Blockchain-based mobile edge computing framework (BMECF), Deep belief network (DBN)	PA used in predicting and improving the accuracy of medical data handling in smart grid healthcare systems
Grover et al. (2020)	Collection Analytics and Health Care	Little's missing completely at random (MCAR) Test	PA used to identify medical resource allocations implications for the COVID-19 pandemic patients
Kaufman et al. (2019)	Collection Analytics	DT	PA used to predict USA government court decisions and provide important factors impacting the court's decision

Xu (2021)	Collection Analytics	Gray Forecast	PA on the structure of port collection and distribution in China (Ningbo Zhoushan Port) which shows demand container sea-rail combined transportation demand will increase in the next five years
Purnamasari et al. (2020)	Cross Sell	Naive Bayes C4.5 DT	PA on the telecommunications customer potential as a new customer through classification of customer data
Punjabi et al. (2021)	Cross Sell	NSEpy, Quandl, NSEtools	Stock trend prediction and market sentiment prediction
Meire (2021)	Cross Sell	Random forest, Support vector machines (SVMs), Neural networks, Naive Bayes Stochastic gradient boosting (SGB)	PA on the customer comeback rate and defines the main criteria and challenges of customer comeback as customer comeback rate will directly impact the revenue of a corporations
Hussein et al. (2021)	Fraud Detection	Fuzzy rough nearest neighbor (FRNN), Sequential minimal optimization (SMO),	PA in credit card fraud detection and help identify the rate of detection rate and

		logistic regression (LR)	false alarm rate
Sawangarreerak and Al-Thanathamathee (2021)	Fraud Detection	Frequent Pattern (FP) Growth Algorithm, DT	PA in fraudulent patterns of Financial Statement for Open Innovation (2710 financial statements from Federation of Professional Accountings, Thailand and help users identify fraudulent relationships on financial statements
Seera et al. (2021)	Fraud Detection	Naive Bayes, DT, Artificial Neural Network (ANN) Logistic Regression	PA in payment card fraud detection from Statlog (German Credit), Statlog (Australian Credit), and Default Credit Card and help users identify fraudulent payment cards fraud detection
Misu and Madaleno (2020)	Underwriting	Models for Assessing Bankruptcy Risk Altman (1968, 2000), Conan and Holder (1979), Springate's 1978), Taffler (1982, 1983), Zmijewski (1984)	PA in bankruptcy Risk of Large Companies: European Corporations and improve the knowledge of bankruptcy prediction of European Corporations

Kou et al. (2021)	Underwriting	Two-stage multi-objective feature-selection	PA in bankruptcy prediction for SMEs in China -3,500,000+ SMEs, including enterprise type, industry, operational status (bankrupt or active) and improve the knowledge of bankruptcy prediction for SMEs in China
Matzavela and Alepis (2021)	Education	DT	PA in student academic performance which contributes to parameters enhancement of a student education performance
Ning et al. (2020)	Manufacturing	Analytical thermal modeling, Analytical distortion modeling	PA in part distortion in metal additive manufacturing which contributes to high and efficient thermal computational efficiency
Ayvaz and Alpay (2021)	Manufacturing	Correlation Analysis	PA in maintenance for manufacturing tools using IoT data which contributes to predicting the useful time remaining in a machine before

			failure
Huo and Chaudhry (2021)	Manufacturing	3D vision of mode network, Heat map, Hierarchical cluster analysis	PA in evaluating global expansion location decision in the Chinese manufacturing sector which contributes to predicting the impact of financial leverage on global expansion decisions in the Chinese manufacturing sector
Kumar et al. (2021)	Manufacturing	Keras/TensorFlow	PA in Energy Management in a smart factory which contributes to predicting energy consumption in smart manufacturing
Ruschel et al. (2021)	Manufacturing	Bayesian networks (BN) Time Series	PA, which contributes to predicting manufacturing cycle time

This review will explore the decision tree, a type of supervised classification tool that is easy to interpret (Sarker, 2021). DT is an established tool that can be used without statistical knowledge and does not need complex formulas (Kingsford and Salzberg, 2008; Sarker, 2021). In addition, DT is user-friendly, and the output can be easily interpreted as compared to other supervised machine learning tools that require statistical knowledge such as Naive Bayes (NB), Logistic regression (LR), Support vector machine (SVM) and Random Forest (RF) (Kingsford and Salzberg, 2008; Sarker, 2021).

Decision Tree (DT)

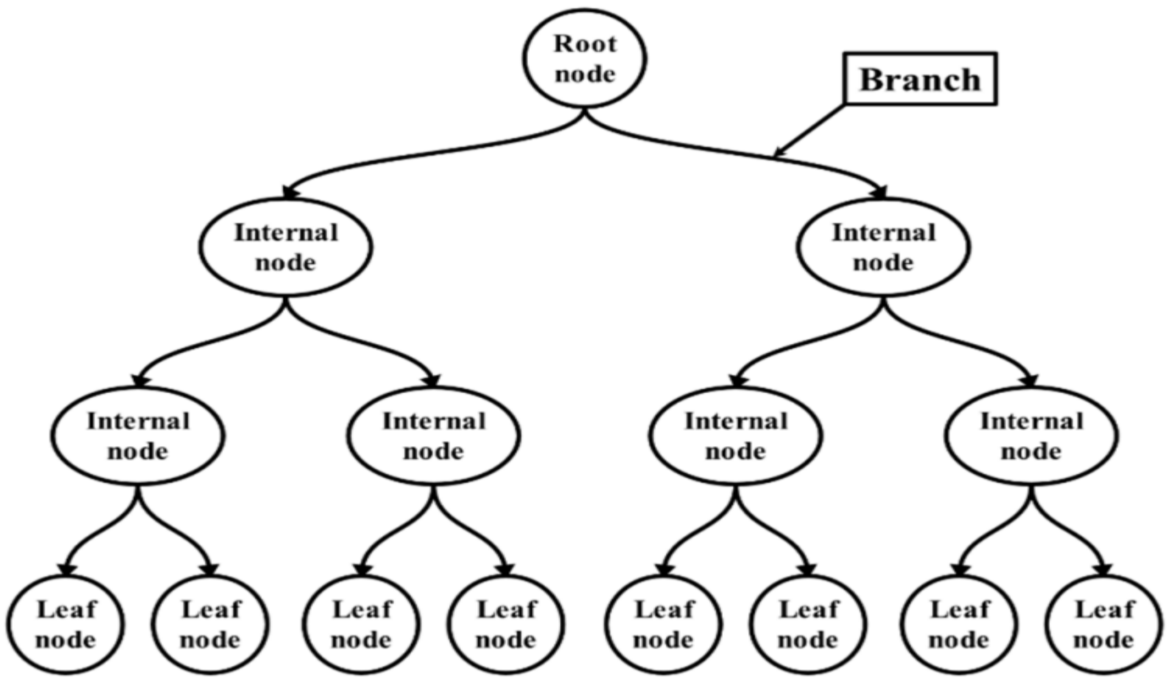
The decision tree is a supervised simple classification tool that can separate data records into designated categories by applying specific conditions in the decision-making process. It is an established tool, and one of the most powerful with relatively small learning curves for interpretability, and is regularly applied in numerous settings such as image processing, ML, data mining and identifications of patterns (Kingsford and Salzberg, 2008; Song and Lu, 2015; Sawant et al., 2021). Not only that, the decision tree was ranked the most more easily interpreted than other supervised machine learning algorithms such as Naive Bayes (NB), Logistic regression (LR), Support vector machine (SVM) and Random Forest (RF), thus justifies for the simple mathematics without even requiring statistical knowledge and no complex formulas (Kingsford and Salzberg, 2008; Sarker, 2021).

DT is a tree-based technique in which any path beginning from the root is described by data separating sequence up to Boolean outcome (either true or false) at the leaf node was achieved (Jijo and Abdulazeez, 2021). It follows a series of questions that provide separation at each level and split points derived from the questions that can be discrete values, a range, or a probability distribution (Hartman, 2021). In applying DT, users can explore the robustness of DT in handling various types of datasets with a mixture of categorical and/or numeric variables, and it can also handle missing data at a specific column (Song and Lu, 2015; Hartman 2021). Moreover, DT is also be used to identify significant variables for predicting an outcome, as DT can be applied to various types of input data (Manogna and Mishra, 2021).

DT input data uses row format, where the rows are known as records, and the columns are known as features. Thus, each row is allocated to a class label that correlates to its designated target (Kingsford and Salzberg, 2008; Hartman, 2021). DT structure is built using nodes and branches, with each node consisting of a specific count of values corresponding to the respective target class for all records within a similar node. The preliminary results of DT were that the target class with the greatest number of records present within the distribution will be displayed on the node (Hartman, 2021). In DT nodes classification and prefix, the parent node (root node) is the beginning node of the tree located at the peak of the DT and in the similar parent node, users will be able to view all the records present. The parent node has an extensive system extending from it, and it is connected through branches to the internal nodes (Kingsford and Salzberg, 2008; Hartman, 2021).

Internal nodes are nodes that branch out from the parent node. These internal nodes are easily identified as they are connected through branches (Abbas et al., 2021). Internal nodes will have branches connected to other internal nodes or leaf nodes. Leaf nodes have branches extended into them but with no branches extending out of them. It is sometimes defined as the end of the nodes and hence, represents the result of combinations of decisions or events (Song and Lu, 2015; Hartman, 2021). Branches in the node represent a split in the dataset. This split will often be associated with questions listed within the response to the description at the branch. DT split can be present in binary or range mode, with numerous answers taken from each of the trait inputs in the DT (Kingsford and Salzberg, 2008; Song and Lu, 2015; Hartman, 2021). The decision tree structure diagram is presented in Figure 2 (Do et al., 2019).

Figure 2. Decision Tree structure (Do et al., 2019)



DT is constructed using an algorithm that repeatedly sorts input data into smaller groups according to the class label. The algorithm is based on the measure of data impurity that determines the split of each node (Kingsford and Salzberg, 2008). There are various impurity types, including Gini impurity, entropy, information gain, and classification error, where nodes were split using a different mode of impurity. After completing the checking, impurities are premeditated for every child node, and its entire impurity for the split is the weighted average of the impurity in the child nodes (Jijo and Abdulazeez, 2021; Singh and Chhabra, 2021; Li et al., 2021). Then, the impurity of each test is compared, and the split with the lowest impurity is chosen. This split process is continued for each node in the tree so that child nodes are "purer" (i.e., homogeneous) in terms of the outcome variable (Korstanje, 2021; Li et al., 2021). After completing the split, the node is finally considered a leaf node (Korstanje, 2021; Li et al., 2021).

Split stopping prevents the DT from growing further to avoid overfitting, which will reduce the reliability of the DT. In overfitting, there will be many child nodes, but there will be a small number of leaf nodes which prevents the prediction capability of DT. This event is also denoted as poor generalizability (i.e., lack robustness) (Song and Lu, 2015; Li et al., 2021). Therefore, in DT, a stopping process needs to be in place to prevent overly complex models, and this includes the lowest number of records in a leaf and node before splitting and depth (i.e., number of steps) of any leaf from the root node (Maass, and Storey, 2021). Furthermore, split stopping must be aligned with the direction of the research. Based on the past findings, the target proportion of records in leaf nodes to be between 0.25 and 1.00% of the total training data set to avoid overfitting and underfitting (Berry and Linoff, 1999; Maass and Storey, 2021).

Last but not least, pruning is another strategy to prevent overfitting in DT. It is often applied as an alternative to prevent overfitting when the split stop is not conclusive (Al-Akhras et al., 2021). Initially, the DT is grown to a large tree and is being trimmed off by removing nodes that provide less additional information (Al-Akhras et al., 2021). The standard method to select optimized sub-tree from several DT is to select the list of history that consists of mistakes in its prediction, such as the predicted incident of the designated target were not predicted correctly. The next method includes choosing a validation dataset, such as sorting the sample size in half and trying out the model created on the training dataset. As for small-scale data sets, it can be performed through cross-validation, which translates into separating the sample into ten groups or 'folds,' and trying out the model generated from 9 folds onto the 10th fold, repeated for all ten combinations, and averaging the rates or erroneous predictions (Song and Lu, 2015; Akhras et al., 2021).

In the DT node trimming process, pruning can be divided into pre-pruning (forward pruning) and post-pruning (backward pruning). Pre-pruning utilizes the Chi-square test or other comparison adjustment methodology to stop the production of non-significant branches (Song and Lu, 2015; Biehler and Fleischer, 2021). After a complete decision tree is developed, post pruning is utilized to detach the branches to improve the precision of the final classification (Song and Lu, 2015; Biehler and Fleischer, 2021).

Systemic Review of Decision Tree (DT)

The decision tree is at the forefront of practical tools for classification in many different applications. Its importance has been noticed in the early 21st century and is growing.

The literature review for DT methodology, as shown in Table 3, is based on 20 research papers published across the years in various fields and applications, and the contribution is summarized.

Table 3. Systemic Review of DT in various fields

Author and Year	Sector/ Industry/ Domain	Algorithm & Tool	Key Takeaways
Emam et al. (2021)	Clinical Trial	Area Under the Receiver Operating Characteristic (AUROC)	DT on oncology clinical trial using sequential Decision Tree which contributes synthetic clinical trial data utility for a sequential synthesis method
Johnson et al. (2021)	Clinical Trial/ Paediatric clinical trials	Simcyp Population-Based Simulator (Version 16.1)	DT is used in pediatric clinical trials as part of either a pediatric investigation in the European Union or as a case study plan in the United States which contributes to predicting age-related changes in physiology and biochemistry
Van Pelt et al. (2021)	Medical/ Covid 19 PCR testing strategy on college students returning to campus	Second-order Monte Carlo simulation/ TreeAge Pro Version 2019	DT was used to predict RT-PCR testing on all students on return to campus if there are no resources

			available to perform RT- PCR test
Chee et al. (2021)	Medical/ Asian American Breast Cancer Survivors	Decision Tree/ SPSS 26.0	DT was used in understanding Asian American Breast Cancer Survivors based on background
Van Benthem and Herdman (2021)	Accidents	Boosted decision tree model	DT can be used to predict accident risk for older pilots, which can be used to build cognitive health screening procedures for older pilots
Cao et al. (2021)	Accidents	Decision Tree/ IBM SPSS 23/ IBM SPSS modeler 18.0	DT on Occupational health and safety (OHS) training of China Construction workers to reduce injury risk
Pappalardo et al. (2021)	Transportation	CART/ Decision Tree	DT on Performance of Lane Support Systems which will help in reducing frontal crash accident
Garcia Marquez et al. (2019)	Manufacturing	Logical Decision Tree	DT helps in cost reduction in wind turbine manufacturing
Merayo et al.	Manufacturing	Top-down induction	DT was used in

(2019)		of decision trees (TDIDT)	material selection applied to manufacturing in Industry 4.0 which will help the designer to optimize material selection
Zeng et al. (2019)	Manufacturing	C5.0 algorithm information gain/ Decision Tree	DT was used in coal manufacturing of Chinese cities, which helps in understanding regional patterns, and coal cities in different regions of China
Antosz et al. (2020)	Manufacturing	CART/ Decision Tree	DT was used in Lean Maintenance concept implementation in manufacturing enterprises which will reduce in corporations operating cost
Panjwani et al. (2021)	Manufacturing	Decision Tree	DT was used in predicting pathogen safety evaluation in biological manufacturing processes, which will help streamline the drug process development

Liou et al. (2021)	Manufacturing	Decision Tree	DT was used in Additive Manufacturing on fault detection capabilities which helps in detecting the defects in real-time
Yeboah-Ofori and Boachie (2019)	Supply Chain	Decision Tree	DT was used in predicting Malware attack in a Cyber Supply Chain
Garcia et al. (2019)	Supply Chain	J48 algorithm/ Decision Tree	DT was used in understanding the Brazilian cotton clothing supply chain, which will contribute to Brazilian cotton clothing sustainability
Zangaro et al. (2019)	Supply Chain	CART/ Decision Tree	DT was used in optimizing assembly line feeding mode for livestock
Chen et al. (2020)	Supply Chain	ID3/ Decision Tree	DT was used in understanding the environmental cost control system of manufacturing enterprises
Kaparthi and Bumblauskas (2020)	Supply Chain	Conditional inference Tree (CTree) technique	DT was used in designing predictive maintenance

			systems, which will contribute to maintenance decision making
de Magalhaes (2021)	Supply Chain	CHAID/ Decision Tree	DT was used in final decision-making for online grocery shopping using the CHAID algorithm, which will contribute to predicting customer purchase trend
Qian et al. (2021)	Supply Chain	CHAID/ Decision Tree	DT was used in predicting air pollutant emissions and supply chain in China, which will provide insights for governmental agencies to implement green supply chain management

Conclusion and future research directions of the study

In summary, business analytics and its application in predictive analytics is an established methodology to extract and predict valuable inputs to generate impactful insights. This review is essential as it provides a fundamental guideline for authors seeking to understand predictive analytics, especially decision tree users. The future work of DT and PA is to incorporate new fields and ideas such as supply chain, manufacturing, medical, and transportation and better incorporate the usage of DT into PA. DT has numerous potentials to become the most influential PA tool as it has a user-friendly methodology and can be used without any deep knowledge of statistics. This paper's significance and perspective highlight the usage of DT and PA. However, this work has some limitations regarding its scope in terms of limitations. The articles analyzed were mainly carried out from recent empirical studies up to 2021 and will require a new review on upcoming years' research to provide the newest studies on business analytics. The limitations of DT are that it would need to know its target (predicted data) and inputs data prior to performing any PA.

Funding

This discourse paper review is a nonprofit review that does not acquire any specific stake from a merchant, general public or nonprofit organization.

Acknowledgment

The third author would like to thank Asia University and California State University San Bernardino for their support.

Conflict of interest

The authors claim no conflict of interest.

References

Abbas, S., Hodhod, R., & El-Sheikh, M. (2021). Retrieval of behavior trees using map-and-reduce technique. *Egyptian Informatics Journal*, 1.

Al-Akhras, M., El Hindi, K., Habib, M., & Shawar, B. A. (2021). Instance reduction for avoiding overfitting in decision trees. *Journal of Intelligent Systems*, 30(1), 438-459.

Al-Zuabi, I. M., Jafar, A., & Aljoumaa, K. (2019). Predicting customer's gender and age depending on mobile phone data. *Journal of Big Data*, 6(1), 1-16.

Antosz, K., Pasko, L., & Gola, A. (2020). The use of artificial intelligence methods to assess the effectiveness of lean maintenance concept implementation in manufacturing enterprises. *Applied Sciences*, 10(21), 7922.

Arismendy, Luis, Carlos Cardenas, Diego Gomez, Aymer Maturana, Ricardo Mejía, & Christian G. Quintero M. (2021). "A Prescriptive Intelligent System for an Industrial Wastewater Treatment Process: Analyzing pH as a First Approach." *Sustainability* 13, no. 8, 4311.

Ayvaz, S., & Alpay, K. (2021). Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Systems with Applications*, 173, 114598.

Bawack, R. E., & Ahmad, M. O. (2021). Understanding business analytics continuance in agile information system development projects: an expectation-confirmation perspective. *Information Technology & People*, 1.

Bayrak, T. (2015). A review of business analytics: A business enabler or another passing fad. *Procedia-Social and Behavioral Sciences*, 195, 230-239.

Berry, M., & Linoff, G. (1999). *Mastering data mining: The art and science of customer relationship management*. John Wiley & Sons.

Bertsimas, D., Li, M. L., Paschalidis, I. C., & Wang, T. (2020). Prescriptive analytics for reducing 30-day hospital readmissions after general surgery. *PloS One*, 15(9), e0238118.

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43, S133-S142.

Bibri, S. E., & Krogstie, J. (2021). A novel model for data-driven smart sustainable cities of the future: A strategic roadmap to transformational change in the era of big data. *Future Cities and Environment*, 7(1).

Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: explore, explain, and examine predictive models*. CRC Press, Taylor and Francis.

Brynjolfsson, E., Jin, W., & McElheran, K. (2021). The Power of Prediction: Predictive Analytics, Workplace Complements, and Business Performance. *Workplace Complements, and Business Performance*, 1.

Cao, Z., Chen, T., & Cao, Y. (2021). Effect of Occupational Health and Safety Training for Chinese Construction Workers Based on the CHAID Decision Tree. *Frontiers in Public Health*, 9, 512.

Chee, W., Yi, J. S., & Im, E. O. (2021). Information Needs of Asian American Breast Cancer Survivors: A Decision Tree Analysis. *Journal of Cancer Education*, 1-10.

Chen, M., Liu, Q., Huang, S., & Dang, C. (2020). Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. *Enterprise Information Systems*, 1-20.

Dagnino, A. (2021). Industrial Analytics. *In Data Analytics in the Era of the Industrial Internet of Things*, 21-46. Springer, Cham.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24-42.

de Magalhaes, D. J. A. V. (2021). Analysis of critical factors affecting the final decision-making for online grocery shopping. *Research in Transportation Economics*, 101088.

de Medeiros, M. M., Hoppen, N., & Maçada, A. C. G. (2020). Data science for business: Benefits, challenges and opportunities. *The Bottom Line*, 33(2).

Do, M., Byun, W., Shin, D. K., & Jin, H. (2019). Factors influencing matching of ride-hailing service using machine learning method. *Sustainability*, 11(20), 5615.

Emam, K. E., Mosquera, L., & Zheng, C. (2021_. Optimizing the synthesis of clinical trial data using sequential trees. *Journal of the American Medical Informatics Association*, 28(1), 3-13.

Espadinha-Cruz, P., Godina, R., & Rodrigues, E. M. (2021). A review of data mining applications in semiconductor manufacturing. *Processes*, 9(2), 305.

Galli, L., Levato, T., Schoen, F., & Tigli, L. (2020). Prescriptive analytics for inventory management in health care. *Journal of the Operational Research Society*, 1, 1-14.

Garcia, S., Cordeiro, A., de Alencar Naas, I., & Neto, P. L. D. O. C. (2019). The sustainability awareness of Brazilian consumers of cotton clothing. *Journal of cleaner production*, 215, 1490-1502.

Garcia Marquez, F. P., Segovia Ramirez, I., & Pliego Marugan, A. (2019). Decision making using logical decision tree and binary decision diagrams: A real case study of wind turbine manufacturing. *Energies*, 12(9), 1753.

Grover, S., McClelland, A., & Furnham, A. (2020). Preferences for scarce medical resource allocation: Differences between experts and the general public and implications for the COVID-19 pandemic. *British Journal of health Psychology*, 25(4), 889-901.

Hartmann, J. (2021). Classification Using Decision Tree Ensembles. *In The Machine Age of Customer Insight*. 1.

Houtmeyers, K. C., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, 16(11), 1719-1723.

Huo, D., & Chaudhry, H. R. (2021). Using machine learning for evaluating global expansion location decisions: An analysis of Chinese manufacturing sector. *Technological forecasting and social change*, 163, 120436.

Hussein, A. S., Khairy, R. S., Najeeb, S. M. M., & ALRikabi, H. T. (2021). Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression. *International Journal of Interactive Mobile Technologies*, 15(5).

Izagirre, U., Andonegui, I., Eciolaza, L., & Zurutuza, U. (2021). Towards manufacturing robotics accuracy degradation assessment: A vision-based data-driven implementation. *Robotics and Computer-Integrated Manufacturing*, 67, 102029.

Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2021). Significance of Quality 4.0 towards comprehensive enhancement in manufacturing sector. *Sensors International*, 100109.

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.

Johnson, T. N., Abduljalil, K., Nicolas, J. M., Muglia, P., Chanteux, H., Nicolai, J., Gillent, E., Cornet, M., & Sciberras, D. (2021). Use of a physiologically based pharmacokinetic–pharmacodynamic model for initial dose prediction and escalation during a paediatric clinical trial. *British Journal of Clinical Pharmacology*, 87(3), 1378-1389.

Jones, R. E., Walton, T. N., Duluc-Silva, S., & Fly, J. M. (2021). Household Food Insecurity in an Urban Food Desert: A Descriptive Analysis of an African American Community. Journal of Hunger & Environmental Nutrition, (1), 1-19.

Kalyankar, G. D., Poojara, S. R., & Dharwadkar, N. V. (2017). Predictive analysis of diabetic patient data using machine learning and Hadoop. *In 2017 international conference on I-SMAC*, 619-624.

Kaparthi, S., & Bumblauskas, D. (2020). Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management*, 37(4), 659-675.

Kaufman, A. R., Kraft, P., & Sen, M. (2019). Improving supreme court forecasting using boosted decision trees. *Political Analysis*, 27(3), 381-387.

Kaur, P., Stoltzfus, J., & Yellapu, V. (2018). Descriptive statistics. *International Journal of Academic Medicine*, 4(1), 60.

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, 26(9), 1011-1013.

Korstanje, J. (2021). The Decision Tree Model. *In Advanced Forecasting with Python*, 159-168.

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multi-objective feature selection. *Decision Support Systems*, 140, 113429.

Kristoffersen, E., Mikalef, P., Blomsma, F., & Li, J. (2021). Towards a business analytics capability for the circular economy. *Technological Forecasting and Social Change*, 171, 120957.

Kumar, M., Shenbagaraman, V. M., Shaw, R. N., & Ghosh, A. (2021). Predictive data analysis for energy management of a smart factory leading to sustainability. *In Innovations in electrical and electronic engineering*, 765-773.

Kumar, V., & Garg, M. L. (2018). Predictive analytics: a review of trends and techniques. *International Journal of Computer Applications*, 182(1), 31-37.

Lana, I., Sanchez-Medina, J. J., Vlahogianni, E. I., & Del Ser, J. (2021). From data to actions in intelligent transportation systems: a prescription of functional requirements for model actionability. *Sensors*, 21(4), 1121.

Li, W., Ma, X., Chen, Y., Dai, B., Chen, R., Tang, C., Luo, Y. and Zhang, K. (2021). Random Fuzzy Granular Decision Tree. *Mathematical Problems in Engineering*, 5578682.

Liou, F., Spark, M. T., Flood, A., & Joshi, M. (2021). Applications of Supervised Machine Learning Algorithms in Additive Manufacturing: A Review. *Preprints*, 2021010588.

Lo, F. Y., Wong, W. K., & Geovani, J. (2021). Optimal combinations of factors influencing the sustainability of Taiwanese firms. *International Journal of Emerging Markets*, 1.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. *National Center for Education Evaluation and Regional Assistance*.

Maass, W., & Storey, V. C. (2021). Pairing conceptual modelling with machine learning. *Data & Knowledge Engineering*, 101909.

Manogna, R. L., & Mishra, A. K. (2021). Measuring financial performance of Indian manufacturing firms: application of decision tree algorithms. *Measuring Business Excellence*, 1.

Matzavela, V., & Alepis, E. (2021). Decision Tree Learning Through a Predictive Model for Student Academic Performance in Intelligent M-Learning Environments. *Computers and Education: Artificial Intelligence*, 2, 100035.

Meire, M. (2021). Customer comeback: Empirical insights into the drivers and value of returning customers. *Journal of Business Research*, 127, 193-205.

Meenakshi, N., Kumaresan, A., Nishanth, R., Kumar, R. K., & Jone, A. (2021). Stock market predictor using prescriptive analytics. *Materials Today: Proceedings*, 1.

Merayo, D., Rodriguez-Prieto, A., & Camacho, A. M. (2019). Comparative analysis of artificial intelligence techniques for material selection applied to manufacturing in Industry 4.0. *Procedia Manufacturing*, 41, 42-49.

Misu, N. B, & Madaleno, M. (2020). Assessment of bankruptcy risk of large companies: European countries evolution analysis. *Journal of Risk and Financial Management*, 13(3), 58.

Mosavi, N. S., & Santos, M. F. (2020). How prescriptive analytics influences decision making in precision medicine. *Procedia Computer Science*, 177, 528-533.

Nasaruddin, A. N., Tee, B. T., Tahir, M. M., & Jasman, M. E. S. M. (2021). Data Assessment on the relationship between typical weather data and electricity consumption of academic building in Melaka. *Data In Brief*, 35, 106797.

Ning, J., Praniewicz, M., Wang, W., Dobbs, J. R., & Liang, S. Y. (2020). Analytical modeling of part distortion in metal additive manufacturing. *The International Journal of Advanced Manufacturing Technology*, 107(1), 49-57.

Nwankwo, W., & Ukhurebor, K. E. (2020). Data Centres: A Prescriptive Model for Green and Eco-Friendly Environment In The Cement Industry In Nigeria. *International Journal of Scientific and Technology Research*, 9(5), 239-244.

Ondeş, R. N. (2021). Research trends in dynamic geometry software: A content analysis from 2005 to 2021. World Journal on Educational Technology: *Current Issues*, 13(2), 236-260.

Panjwani, S., Cui, I., Spetsieris, K., Mleczko, M., Wang, W., Zou, J.X., Anwaruzzaman, M., Liu, S., Canales, R. and Hesse, O. (2021). Application of machine learning methods to pathogen safety evaluation in biological manufacturing processes. *Biotechnology Progress*, e3135.

Pappalardo, G., Cafiso, S., Di Graziano, A., & Severino, A. (2021). Decision tree method to analyze the performance of lane support systems. *Sustainability*, 13(2), 846.

Punjabi, P., Vaswani, P., & Kubal, A. (2021). Modelling Stock Trading Platforms Leveraging Predictive Analysis Using Learning Algorithms, International *Journal of Research and Analytical Reviews*, 2348-1269.

Purnamasari, I., Handayanna, F., Arisawati, E., Dewi, L. S., & Sihombing, E. G. (2020). The Determination Analysis of Telecommunications Customers Potential Cross-Selling with Classification Naive Bayes and C4. 5. *In Journal of Physics: Conference Series*, 1641, 1, 012010.

Qian, Y., Li, Z., & Tan, R. (2021). Sustainability analysis of supply chain via particulate matter emissions prediction in China. *International Journal of Logistics Research and Applications*, 1-14.

Rai, B. (2021). Factors Affecting Smartphone Purchase Intention of Consumers in Nepal. *The Journal of Asian Finance, Economics, and Business*, 8(2), 465-473.

Ruschel, E., Loures, E. D. F. R., & Santos, E. A. P. (2021). Performance analysis and time prediction in manufacturing systems. *Computers & Industrial Engineering*, 151, 106972.

Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications*, 9(2).

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *Springer Nature Computer Science*, 2(3), 1-21.

Sawangarreerak, S., & Thanathamathee, P. (2021). Detecting and Analyzing Fraudulent Patterns of Financial Statement for Open Innovation Using Discretization and Association Rule Mining. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(2), 128.

Sawant, N. V., Panicker, V. V., & Anoop, K. P. (2021). Predictive Analytics in Food Grain Logistics: Supervised Machine Learning Approach. *Optimization Methods in Engineering*, 459-466.

Saxena, M., Bagga, T., & Gupta, S. (2021). Fearless path for human resource personnel through analytics: a study of recent tools and techniques of human resource analytics and its implication. *International Journal of Information Technology*, 1, 1-9.

Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2021). An intelligent payment card fraud detection system. *Annals of Operations Research*, 1-23.

Sharma, S., & Gupta, Y. K. (2021). Predictive analysis and survey of COVID-19 using machine learning and big data. *Journal of Interdisciplinary Mathematics*, 24(1), 175-195.

Sneha, N., & Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data*, 6(1), 1-19.

Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2021). COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. *British Journal of Management*, 32(4), 1164-1183.

Singh, M., & Chhabra, J. K. (2021). EGIA: A new node splitting method for decision tree generation: Special application in software fault prediction. *Materials Today: Proceedings*, 1.

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130.

Surucu-Balci, E., Balci, G., & Yuen, K. F. (2020). Social media engagement of stakeholders: A decision tree approach in container shipping. *Computers in Industry*, 115, 103152.

Tolba, A., & Al-Makhadmeh, Z. (2021). Predictive data analysis approach for securing medical data in smart grid healthcare systems. *Future Generation Computer Systems*, 117, 87-96.

Wassouf, W. N., Alkhatib, R., Salloum, K., & Balloul, S. (2020). Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study. *Journal of Big Data*, 7(1), 1-24.

Xu, D. (2021). Analysis on the structure of port collection and distribution in China. *In IOP Conference Series: Earth and Environmental Science*, 791, 1, 012080.

Van Benthem, K., & Herdman, C. M. (2021). A virtual reality cognitive health screening tool for aviation: Managing accident risk for older pilots. *International Journal of Industrial Ergonomics*, 85, 103169.

Van Pelt, A., Glick, H. A., Yang, W., Rubin, D., Feldman, M., & Kimmel, S. E. (2021). Evaluation of COVID-19 testing strategies for repopulating college and university campuses: a decision tree analysis. *Journal of Adolescent Health*, 68(1), 28-34.

Vlahakis, G., Kopanaki, E., & Apostolou, D. (2020). Proactive decision making in supply chain procurement. *Journal of Organizational Computing and Electronic Commerce*, 30(1), 28-50.

Yeboah-Ofori, A., & Boachie, C. (2019). Malware Attack Predictive Analytics in a Cyber Supply Chain Context Using Machine Learning. *In 2019 International Conference on Cyber Security and Internet of Things (ICSIoT)*, 66-73.

Zangaro, F., Minner, S., & Battini, D. (2020). A supervised machine learning approach for the optimisation of the assembly line feeding mode selection. *International Journal of Production Research*, 1-22.

Zeng, L., Guo, J., Wang, B., Lv, J., & Wang, Q. (2019). Analyzing sustainability of Chinese coal cities using a decision tree modeling approach. *Resources Policy*, 64, 101501.