The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business

Mohd Shahrizan Abd Rahman, Nor Azliana Akmal Jamaludin, Zuraini Zainol, Tengku Mohd Tengku Sembok
Department of Computer Science, Universiti Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia

***Abstract*—In today's rapidly advancing technological landscape and evolving business paradigms, the pursuit of insightful patterns and concealed knowledge beyond conventional big data becomes imperative. This pursuit serves a crucial role in aiding stakeholders, particularly in the realms of tactical decision-making and forecasting, with a particular focus on business strategy and risk management. Strategic and tactical decision-making holds the key to sustaining the longevity, profitability, and continuous enhancement of the oil and gas industry. Therefore, it is paramount to address this need by uncovering the most effective Decision Tree (DT) techniques for various challenges and identifying their practical applications in real-life scenarios. The integration of big data with Machine Learning (ML) stands as a pivotal approach to foster data-driven innovation within the oil and gas sector. This study aims to offer valuable insights and methodologies for efficient decision-making, catering to the diverse stakeholders within the oil and gas industry. It focuses on the exploration of optimal DT techniques for specific problems and their relevance in practical situations. By harnessing the potential of machine learning and collaborative efforts among research scientists, big data practitioners, data scientists, and analysts, the study strives to provide more precise and effective data. Furthermore, it is imperative to recognize that not all stakeholders are mathematicians. In project management, a holistic approach that considers humanistic perspectives, such as risk analysis, ethics, and empathy, is crucial. Ultimately, the output and findings of any system must be accessible, comprehensible, and interpretable by humans or human groups. The success of these insights lies not just in their mathematical precision but also in their ability to resonate with and guide human decision-makers. In this light, the study emphasizes the human element in data interpretation and decision-making, acknowledging that the system's output will require human interaction, analysis, and ethical considerations to be truly effective in driving positive outcomes in the industry.**

Keywords*—Decision-making strategies; decision tree family; business decisions; upstream; oil & gas; predictive analysis; project control; project planning; machine learning algorithms

I. Introduction

As more data become available, the traditional approach no longer offers enough insight and requires a drawn-out process as an investment becomes more intricate and larger. It is now time to look at alternative options. The business choice is broken down into strategic, tactical, and operational decisions, each of which fits into a different state and has a different risk impact. A strategic decision is to determine if a new investment opportunity is worthwhile or not, such as selecting the candidate project that would yield the highest Return of Investment (ROI) while staying within the authorised Work Program Budget (WPB) planning. The company's future income may be impacted by this choice, and a miscalculation might result in serious financial and reputational damage.

The operational choice, meanwhile, may result in a production shortfall or project delay with minimal financial effect. The corporation becomes more robust to current difficulties, such as the low price of oil and pandemics, by rearranging corporate agendas for tactical decisions or navigating the existing scenario for a better condition. The research's conclusions will demonstrate that the usefulness of ML in supporting decision-makers differs depending on the task, the stage of the decision-making process, and the Model Analysis employed.

Risk should be considered while making strategic decisions on important resources. Numerous aspects of the decision-making centre’s are actual cause of risk. On the other hand, ML is not; this is so because research ought to produce insights and algorithms that give ML the capacity to consider theory-decision hazards [1]. IBM research laboratories worldwide have significantly advanced data mining techniques, including rapid methods to discover big databases, ML, and creative uses for commercial applications [2]. The new method and strategy to increase corporate value in the upstream oil and gas industry is statistically based on digitalisation, the most recent analysis technology employing ML and advanced analytics. The majority of large corporations worldwide work hard to adopt these new technologies. Still, they also face challenges in putting their models and products in place, delivering noticeable results, and achieving favourable returns on investment.

Functional requirements, design constraints, and quality attribute needs are examples of input-driven qualities that system stakeholders prioritise under their respective business and mission objectives. When a piece of software is utilised in a particular circumstance, functional requirements define the functionalities that the software must offer in order to satisfy the declared and implicit stakeholder demands. Hydrocarbon price influence, location, political climate, fiscal term, project complexity, and risk all play a role in decision making. Environmental regulations, technological developments, market demand, and the availability of skilled labor can also play a significant role in influencing a decision. Project selection decisions should also take into account macroeconomic trends, political stability, and the energy industry's potential for future growth. The consistency and ongoing expansion of data are crucial to income. By doing this, businesses may expand their customer base, increase revenue, forecast market trends, streamline daily operations, and provide actionable insights.

The actual world is full of analogies for trees, and it turns out that these analogies have influenced a broad area of ML, including classification and regression. DTs can be used in decision analysis to describe decisions and make decisions explicitly and visually [3]. It makes use of a tree-like decision mechanism, as the name would imply. Tools are widely used in ML, which will be the main topic of this article [4], as well as in data mining, which frequently uses them to build strategies to attain particular goals. The linkages between the traits and their significance are obvious. Similar to DTs, regression trees also predict continuous values. Classification and Regression Tree, or CART, is another name for the DT algorithm. Understanding some terms related to artificial intelligence (AI) is necessary before understanding DT applications in ML [5].

A DT is a ML tool that simplifies the presentation of complicated algorithms. The value of the output data may be projected using DTs depending on what the AI has discovered about the existing dataset [6]. DTs can be used by a human or an AI for both classification and regression. Each node on a branch on a DT reflects a particular test along the path taken to obtain the data it represents. The general public may grasp DTs in ML rather well. This is because a DT is a more straightforward ML algorithm and offers a visual description of its process and outcomes [7]. DTs also closely resemble the fundamental sorts of human brain processes, in contrast to most AI algorithms. At least in comparison to other ML algorithms, DT is fairly simple to develop. Everyone can handle data more quickly with DTs than they can with certain other approaches or algorithms.

Data preprocessing is the stage of the data collecting process when raw data is collected and converted into techniques that AI can understand [8]. Businesses upstream locate and harvest raw resource reserves. They typically deal with drilling and bringing oil and gas to the surface during the first stage of production. Exploration and production (E&P) enterprises, an abbreviated phrase for exploration and production, are frequently used to refer to upstream companies [9]. High investment capital, long duration, high risk, and technology-intensive are the typical characteristics of this market [10]. Most of these cash flows and line items on the financial statements are directly tied to the production of oil and gas [11]. Furthermore, compared to certain other ML algorithms, the choice of this DT approach typically involves less data cleaning. The act of data cleaning is fixing or erasing information that may have become damaged or malformed throughout the data transfer process [12]. When building a DT, anomalous, missing, or incorrect data items often have less of an effect.

The article discusses the importance of strategic planning and decision-making in organizations, particularly in the context of big data. It highlights the need for a focus on intangibles, such as technology and skilled labor, to generate revenue and improve efficiency. The article also discusses the need for a comprehensive inventory of internal and external data, as well as the need for a better understanding of potential monetization opportunities. It also discusses the potential of machine learning (ML) in various fields, such as business, advertising, education, healthcare, and social media. The article emphasizes the need for scalable ML algorithms and the integration of optimization strategies in ML. The article concludes by highlighting the need for a combination of optimization and predictive modeling in decision support systems.

II. Background and Related Studies

A. *Decision Making towards Big Data*

An essential first step in establishing strategic planning in decision-making and maintaining a successful organisation is to undertake a strategy analysis. Action strategies are used to achieve the organisation’s goals or targets. For an organisation to advance, daily strategy planning is required. Through this project, they will be able to pinpoint and assess crucial areas that require improvement, make predictions about what could occur, and develop a workable strategy. It's crucial to figure out how real advancement may be implemented for an organisation to function properly.

Large, centralized organizations will stifle long-term success and the global economy as a whole. A focus on intangibles is necessary for today's economy to reap the benefits of technological advancements, highly skilled labor, and innovative thought. Monetization is widely acknowledged to be crucial. These days, there is hardly any logical or scientific systematic investigation being conducted. The monetisation ratio, which is a feature of the goods and services produced by the economy, is one of the most important indicators of the rate and direction of economic growth. Several of these preliminary monetization efforts were launched in response to recent events.

Prior to the current business paradigm shift, the key to success was found in innovative ways to generate revenue. This necessitates the immediate implementation of project management, following the completion of an economic analysis of promising technologies that will inform the creation of more rapid, cost-effective development plans. Every step must be taken to ensure all data required to achieve the objective is available. Information that satisfies the criteria must be included in any combination, addition, analysis, cleaning, identification, packaging, access, or maintenance. Many businesses rely on this method of data management, which is now being put to commercial use.

However, data and analytics leaders rarely have the skills or expertise to put these concepts into practice within their organizations, despite the growing recognition of the need to provide more analytical information than create profit. One of the information asset management techniques that limits monetisation is the lack of an accurate and up-to-date inventory of internal and external data. Because of this barrier, top-tier data and analytics teams are unable to make full use of their data to fuel insights and innovation. Companies can't find new sources of revenue or base decisions on data-driven insights if they don't know what information they already have. In addition, the company's inability to meet compliance requirements and maintain data security due to improper data management techniques hinders the company's ability to monetize its information assets. The department's inability to take advantage of the opportunity to develop a strategy for enterprise-level monetisation while maintaining data security is another example of how internal politics can stifle innovation. Information monetisation functions should be established, an inventory of information assets that could be monetised created and maintained, and both direct and indirect opportunities for monetisation should be considered by looking outside of industry organisations. Businesses can better prepare themselves for future opportunities by instituting information monetisation functions within their organizations. Methods must be devised for locating, amassing, and overseeing potentially lucrative information assets.

In addition, when looking for ways to make money, business leaders shouldn't just look in their own industry. Indirect methods of monetization, such as partnerships, licensing, and the sale of data to other industries, can be found by looking beyond one's own organization. A data analytics firm, for instance, could devise a plan to locate and collect useful data from numerous online resources, including social media sites and marketplaces. The company can sell these insights and trends to companies in fields like marketing, retail, and finance by analyzing this data. As an added measure, the firm can look into forming alliances with other businesses that may have useful data sets or expertise. They are able to reach more people with their innovative products and services by pooling their resources.

B. Machine Learning

Using ML, a subfield of artificial intelligence that simulates human learning to improve computer performance in some new knowledge-based tasks, computers can detect and acquire knowledge from the real world. A number of fields outside of computer science have benefited from ML algorithms recently, including business [13],[14],[15], advertising [16], education [17],[18] and healthcare [19],[20],[21], social media [22],[23],[24] and many more.

ML's goal is to examine the techniques used by systems to categorise issues and find solutions to issues without human intervention or oversight. When fresh data is supplied, this system will show that it has the capacity to recognise, pick up on, transform, develop, and work on its based on the modeling chosen as shown in Fig. 1.

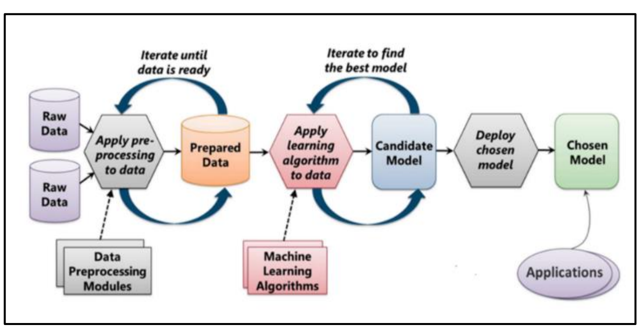

Fig. 1. Workflow for ML Project by David Chappel

ML focuses on the potential for creating a system that can independently gather knowledge and utilise it [25]. Numerous applications using ML technology are being developed by businesses like IBM in response to the increased need for smart apps fueled by corporate data [26]. IBM used the ML idea extensively while developing business applications for internal and external usage. Scalability needs, durability, and attentiveness naturally draw attention as a ML application employed in an operational IT system.

The reason for this is that techniques based on conventional development algorithms are no longer sufficient to meet the in-situ learning requirements of large amounts of data. Scalable ML algorithms must therefore be designed and implemented. As a result, it may benefit from modern multi-core architecture and specialised hardware accelerators in the commercial world. Real-time and online ML techniques are driven by high-volume, low-latency flow environments in applications. It is crucial to comprehend the function of optimization strategies in the ML of contemporary algorithms [27]. When applied to optimization, ML takes on an entirely new dimension. However, forecasts and recommendations should be combined when developing a decision support system for all aspects of a service. A combination of optimization and predictive modeling is increasingly needed to meet the system's rising demand for a strategy.

C. Machine Learning Methods

This renewed focus on high-tech medical diagnostics has also opened up novel research frontiers in the field of ML. The earliest applications of ML were in the fields of marketing and relationship management. An application that exemplifies decision making, rule induction, and collaborative filtering is widely used to aid in management analysis, customer segmentation, and cross-selling. ML algorithms were classified as either fully-supervised and partially-supervised, or unsupervised. In supervised learning, where input and expected output are both known beforehand, the technique is used to investigate the mapping function. Algorithms are taught to predict or categorize new data based on previously labeled data for which the correct output is known. Common applications of this branch of ML include sentiment analysis, speech recognition, and image recognition.

Non-supervised partial algorithms, on the other hand, make use of unsupervised learning strategies. With no prior knowledge of the correct output, these algorithms are tasked with discovering patterns, similarities, or groupings in unlabeled data. Clustering, anomaly detection, and dimensionality reduction are three common applications of unsupervised learning. Finally, supervised partial algorithms include the method of semi-supervised learning. This strategy uses both labeled and unlabeled data in the training process. It makes use of the limited amount of labeled data [28].

Classification and regression are the two main tasks in supervised learning. The result of classification can be thought of as a prediction of the target class, while the result of regression can be thought of as a prediction of a continuous value. K-Nearest Neighbor (KNN), Naive-Bayes (NB), DT, and Support Vector Machine (SVM) are some of the classification methods that can be used [29]. Several algorithms, such as NB, SVM, and neural networks, can accomplish this. Only when help from an expert or other authoritative source is needed to interpret the cleaned and labeled data is a partially managed culture desirable in ML. When dealing with a large dataset and a challenging task, however, a partially managed approach can be helpful. Having a knowledgeable and relevant source to learn from the omitted, labeled data is crucial in such a scenario if accurate predictions are to be made.

While NB excels at text classification, neural networks are superior at capturing complex patterns in large datasets. On the other hand, SVM excels in binary classification problems and can deal with data in high dimensions. Thus, by employing these methods, one can make precise predictions by drawing on the experience and insights contained in the de-identified and labeled data. Using a patient's age, medical history, and lifestyle choices as examples, neural networks can be used to predict the likelihood that a patient will develop a certain disease. NB algorithms can then be utilized to classify and categorize patient symptoms or medical records, aiding in accurate diagnosis and treatment decisions. Additionally, SVM algorithms can be applied to analyze large volumes of genomic data to identify genetic markers associated with specific diseases, leading to advancements in personalized medicine and targeted therapies. Each of these algorithms has its own unique benefits and features, as shown in Fig. 2

Algorithms	Decision Tree	Non-Linear SVM (based on libsvm)	Linear SVM (based on liblinear)
Types	Discriminant	Discriminant	Discriminant
Characteristics	Classification tree	Super-plane separation, kernel trick	Super-plane separation
Learning policy	Regularized maximum likelihood estimation	Minimizing regular hinge loss, soft margin maximization	Minimizing the loss of regular hinge. Soft margin
Learning algorithms	Feature selection, generation, prune	Sequential minimal optimization algorithm (SMO)	Sequential dual method
Classification strategy	IF-THEN rule according to tree spitting	Maximum class of test samples	The maximum weighted test sample

Fig. 2. Algorithms comparision.

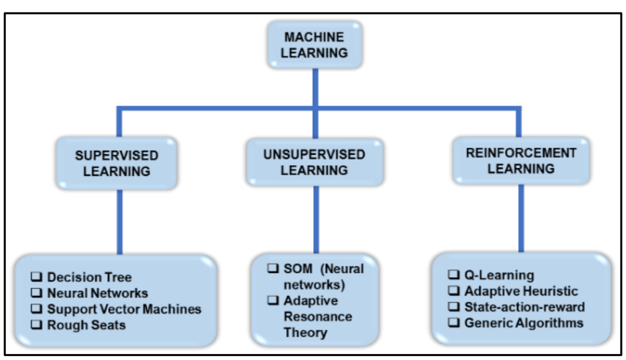

Fig. 3. ML category.

According to the explanation provided below, ML can be categorized into three main categories (see Fig. 3):

- Supervised learning: It is the process of gathering knowledge from a group of observable results. Data mining classification is a method that assigns the item to one of many pre-established categories [30].
- Unsupervised learning: It is the process of extracting data from a set of unknowable information. Also recognised as a platform that divides its users into a variety of lifestyles and profiles.
- Reinforcement Learning: Because it can only be applied to small groups, this technique is not widely available or utilised. Trial and error learning is another name for this process.

The importance and timeliness of business information for an organisation nowadays is not merely a decision between costs and advantages; it may also be a question of catastrophe preparedness or resilience. ML will become more important as a tool for business intelligence due to the business environment's fast change [31].

D. Decision Tree

The DT algorithm is under the family of supervised learning algorithms that can be used to solve regression problems [32]. Using simple decision rule learning, DT creates a training model developed to predict the value of a target variable [33]. Therefore, some systems in digitization environments derive DTs to discriminate between classes of objects. Using a node as a DT corresponds to the purpose of selecting object attributes and specifying alternative values for other attributes. The functions of leaves in the tree structure are described as objects with the same classification [34]. There are two types of Variable DTs either Categorical or Continuous. The difference between the two types is the target variable. According to Y.Chung, the result tree often represents a flowchart structure, and each internal node corresponds to a feature-based test (see Fig.4). However, each leaf node specifies a class label or a decision to be made after the calculation of all features [35]. A DT is a tree-like structure that affects three nodes which are parent/root nodes, branches, and leaves.

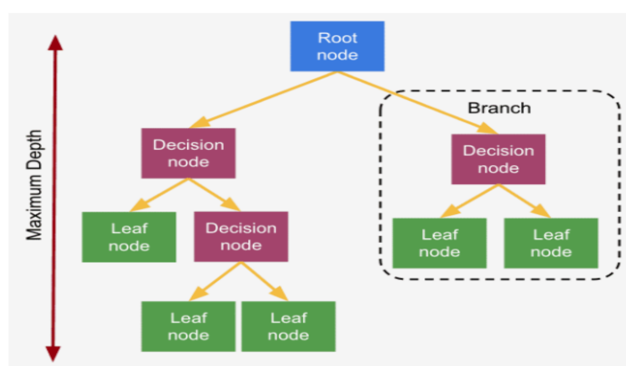

Fig. 4. Flowchart structure of node.

The advantage of using DT as a technique in ML data analysis is the ability to classify unknown records very quickly and resolve redundant attributes correctly and robustly in the presence of noise if a method like overfitting is provided [36]. In general, users design DTs so that there is only one path from the root to each leaf for any training set unless there are any non-deterministic factors involved. Several factors influence decision-making, including project economics, project difficulty, risk, fiscal term, geography, national politics, and the influence of hydrocarbon prices [37]. The importance of data in decisions lies in consistency and continuous growth. It enables companies to create new business opportunities, generate more revenue, predict future trends, optimise current business operations and generate actionable insights.

E. Decision Tree Regression

DT Regression is used for classification tasks, but it is possible to use it for regression tasks. For a training vector x ∈ , Rn (where n is some feature) and a training label y ∈ Rl (i = 1, 2,... l represents some label) the regression tree algorithm recursively divides the feature domain into smaller regions (class separately). Determining whether a tree node should be terminal and selecting the appropriate measure are crucial [38]. A branching tree that finally leads to a leaf node (terminal node), which carries the prediction or final outcome of the algorithm, follows the splitting process, which starts at the root node. Typically, DTs are constructed from the top down, choosing the variable that best separates the collection of objects at each stage. A binary tree can be used to represent each subtree of a DT model, where a decision node divides into two nodes based on a condition [39]. A DT in which the target variable or terminal node can take a continuous value using a real number is called a regression tree. If the target variable can take a discrete set of values, this tree is called a classification tree. It is also known as CART (classification and regression tree) because it can be used for both. It builds various models in a tree-structured form. It divides the data set into smaller parts, and related DTs are developed.

III. Result

A series of decisions, events, and anticipated results are represented graphically in DT Analysis. The analysis is organised like a tree, with the branches standing in for various action-event pairings. Each decision's conditional reward is determined by taking into account possible action combinations. When a decision-making process is multi-level, which happens when an event occurs in a series of levels, the DT Analysis approach is appropriate [40]. As a result, the DT Analysis approach is logically organised and appropriate for situations involving decision-making. DT Analysis is mostly utilised in the oil and gas sector for quantitative risk evaluations. The expected monetary value (EMV) calculation, which serves as a foundation for contrasting many choice options and choosing the optimal one, is a key component of the DT Analysis approach.

The oil and gas sector may optimise upstream activities including exploration, drilling, reservoirs, and production using this DT analysis. Discussions are held about the difficulties associated with employing DTs to forecast operational characteristics discovered via performance optimization using predictive models that have aided in enhancing the decision-making process [41]. DTs are classified in terms of the removal or decrease of uncertainty. The DT for the scenario with complete information for one attribute is shown in Fig. 5 along with three attribute groups. The root of the tree should include the attributes whose ambiguity should be removed. The information needed to build groups depends on that information. Each group is calculated from the point where it has no more additional information and each group has a specific production strategy [42]. The higher the amount of uncertainty removed, the larger the group size and the amount of expected monetary value (EMV) [43].

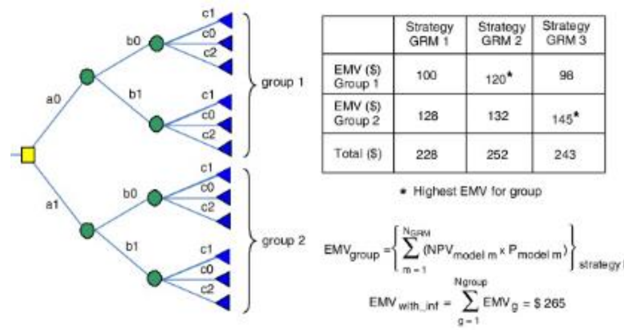

Fig. 5. An example of a DT and EMV calculation for a case with complete information for one attribute.

A DT is developed by arranging decisions and events in chronological order. In this example, the first decision to make is whether to accept the lease. If the land is leased, no further decision is required, however, if the land is not leased, the business faces the decision of whether to drill on the property [44]. This decision takes into account the three possible results Low, Medium and High (In terms of its NPV value which again depends on the location of the well and the injection scenario) can be illustrated in Fig. 6. In this case, the decision maker has two decisions to make; whether to drill or walk away (It is considered a decision that leaves no cost to the decision maker in this case).

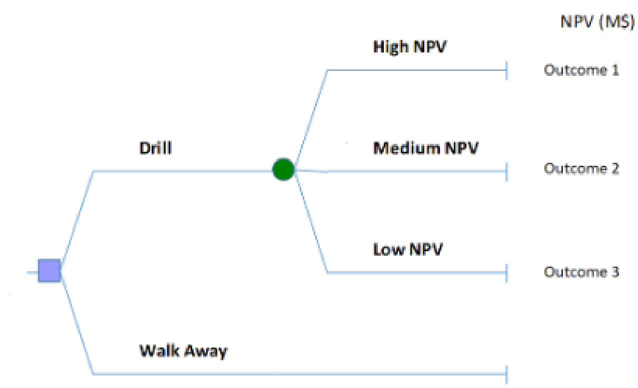

Fig. 6. A Simple DT for the case of decision making for drilling a 5-point pattern.

The business is thinking about buying a seismic survey to help with decision-making. According to corporate experts, there is a 0.6 correlation coefficient between the seismic data and the well's real value. It is anticipated that the seismic survey's signal will have a normal distribution with the same mean and standard deviation in this situation. However, the value of information (VOI) that can be added to the decision by the information gathering must be considered (see Fig. 7). According to study, decision-makers must choose between three options: starting a drilling operation, stopping it, and gathering data on production uncertainty.

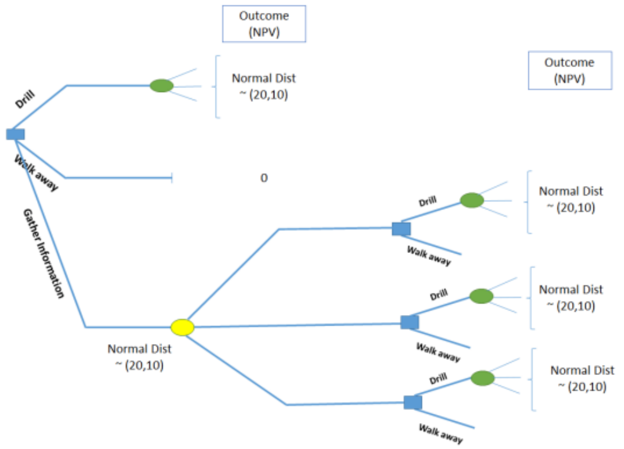

Fig. 7. DT Problem with information.

IV. Discussion

Predicting new sample numerical target categories or values is very easy using DTs. That is one of the main advantages of this type of algorithm [45]. Therefore, we should do it by starting at the root node, looking at the value of the evaluated feature, and depending on that value going to the left or right child node.

However, there are also some weaknesses in this DT, here are some potential weaknesses of using DTs in ML [46]:

- Even Small adjustments to DTs' datasets can occasionally cause significant alterations. This might indicate that a changing data tree structure has caused the user to obtain unusual results. This is why the DT approach is seen as unstable.
- For huge data sets, DTs could be less reliable in predicting the outcome. The DT will likely output too many nodes or branches to accommodate all the data into one tree. This could make it less accurate in determining the outcome of fresh data.
- The ideal model for predicting continuous variables might not be DTs. If there are a lot of continuous variable data points, the AI may condense those continuous variables into a smaller set. Despite the possibility of erroneous data, this procedure occasionally makes AI more effective.

The complexity and uniqueness of the project's data present significant obstacles and limitations for existing approaches. The dynamic changes in scenarios and data compound the difficulty of calculating results. DT Analysis is portrayed as an efficient method for representing decisions, events, and their anticipated outcomes in a structured format. Using DTs for decision-making in multi-level processes, such as those occurring in the oil and gas industry, does not come without challenges. The complexity and unreliability of the data involved is a significant limitation of DT in the oil and gas sector. Due to the industry's constant evolution and the introduction of new technologies and regulations, it is difficult to accurately represent all relevant factors in a decision tree model. Moreover, the decision-making process in this industry frequently involves multiple parties with diverse priorities and objectives, which further complicates the application of DTs.

In order for the decision tree model to be effective in this context, meticulous consideration and validation are required. Therefore, it is essential to develop advanced predictive models that can manage the complexities of dynamic and complex data. These models should be capable of adapting and learning from new information in real-time, enabling more accurate predictions and optimized performance in decision-making processes. Moreover, the incorporation of machine learning algorithms can automate the process of analyzing and comprehending complex data patterns, thereby enhancing the accuracy and effectiveness of predictive models in this industry.

V. Conclusion

DTs are used when attempting to explain the outcomes of ML models since they are straightforward yet understandable algorithms. Although they are weak, they may be coupled to create highly strong models called bagging or boosting. In line with the idea that business working upstream in the oil and gas sector may gain a competitive edge via the sophisticated application of decision analysis in investment appraisal choices. Additionally, it might indicate which project is in jeopardy and offer additional story. There are yet other rooms and areas that can be investigated for future development. Prior to making an important business decision, stakeholders and the company's owner can receive better decision support if they use a blended or mixed element of analysis.

The focus of "Industry 4.0" is on digitization, mechanization, and computer technology. Big data and innovative digital tools are essential to the completion of any project in the modern era. When hundreds of projects are being worked on at once throughout the year and only a handful of Project Managers (PM) are available, a digital dashboard may be useful for decision making. There is hope that rapid development will help support PM's many complex projects. Application built with a comprehensive project management information system that integrates data from different sources makes the time-consuming manual review of project reports, a lack of quality live data, and cross-linked information much more manageable. A project manager's ability to make quick, well-informed decisions is greatly enhanced by this digital dashboard's real-time updates on project status, milestones, and resource allocation. Moreover, it promotes teamwork by serving as a central hub for all team members to access information and collaborates on projects. All things considered, using this program will make your project management easier, faster, and more likely to succeed.

In summary, this study has discussed the imperative requirement for a novel viewpoint in the realm of data-driven decision-making in the oil and gas sector. Nevertheless, it is crucial to acknowledge that the difficulties pertaining to data management, knowledge integration, and process data usage are not limited just to this particular industry. Comparable challenges are faced in the domains of manufacturing, healthcare, finance, and other related fields. Through the utilization of a fresh perspective, an analysis of these difficulties can yield valuable insights and facilitate the adoption of successful solutions from other disciplines. Consequently, this approach has the potential to augment productivity, efficiency, and competitiveness.

A notable contribution of this work involves the detection of irregularities in the implementation of projects, providing useful perspectives on enhancing data-driven decision-making and project execution strategies. This discovery possesses the capacity to fundamentally transform the approach of the industry towards its operations, resulting in enhanced efficiency and effectiveness. Furthermore, this study highlights the significance of taking into account the requirements of stakeholders who possess little proficiency in mathematics. Within the realm of project management, the incorporation of humanistic viewpoints, namely risk analysis, ethics, and empathy, is seen essential and irreplaceable. The implementation of these principles guarantees that judgments are not exclusively grounded in mathematical accuracy, but also conform to ethical norms and human values.

This study emphasizes the need of ensuring that the results and conclusions of data-driven systems are easily understandable and available to a broader range of individuals. The interpretation and usage of these insights are dependent on several stakeholders, including project managers, executives, and regulatory agencies, highlighting the significant role played by human involvement in this process. Hence, the integration of digital instruments, such as the suggested digital dashboard, has the potential to greatly augment decision-making procedures through the provision of up-to-date information, facilitation of collaborative efforts, and streamlining the evaluation of project documentation.

Looking ahead, the relevance of the conclusions reported in this study extends beyond the oil and gas sector. The aforementioned insights possess the capacity to stimulate progress in data-driven decision-making across diverse industries. Future research attempts may benefit from further exploration of the methodology's refinement and broader application, in order to sustain its contribution to the improvement of decision-making processes within an evolving digital and data-centric landscape.

REFERENCES

[1] A Sani, "Machine Learning for Decision Making," in Universite de Lille 1,. English, 2015.

[2] C Apté, "The Role of Machine Learning in Business Optimization," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010.

[3] C. Dowell, "Machine Learning for Downstream Oil and Gas Refineries: Applications for Solvent Deasphalting," Submitted to the System Design and Management Program, Degree of Master of Science in Engineering and Management. Massachusetts Institute of Technology (MIT), B.S., University of California, 2021.

[4] Sousa, T. P. Ribeiro, S. Relvas, and A. P. B. Póvoa,"Using Machine Learning for Enhancing the Understanding of Bullwhip Effect in the Oil and Gas Industry,"in Mach. Learn. Knowl. Extra, 2019, pp. 994-1012.

[5] Johnson, Joseph G., and J. R. Busemeyer, "Decision making under risk and uncertainty," in Wiley Interdisciplinary Reviews: Cognitive Science 1.5, 2010, pp. 736-749.

[6] Regehr, Cheryl, et al, "Improving Professional Decision Making in Situations of Risk and Uncertainty: A Pilot Intervention," in The British Journal of Social Work 52.3, 2022, pp. 1341-1361.

[7] A. K. Abayomi, and S. C. Morrish, "Conceptualizing post-disaster entrepreneurial decision-making: Prediction and control under extreme environmental uncertainty," In 2022 International Journal of Disaster Risk Reduction 68, 2022: 102703.

[8] Bai, Wensong, et al, " Where business networks and institutions meet: Internationalization decision-making under uncertainty," in Journal of International Management 28.1, 2022: 100904.

[9] Petracou, V. Electra, V. Xepapadeas, and A. N. Yannacopoulos, "Decision Making Under Model Uncertainty: Fréchet–Wasserstein Mean Preferences." In Management Science 68.2, 2022, pp. 1195-1211.

[10] Nautiyal, Aditi, and A. K. Mishra, "Machine learning approach for intelligent prediction of petroleum upstream stuck pipe challenge in oil and gas industry," in Environment, Development and Sustainability, 2022, pp. 1-27.

[11] Wang, Xiaoyan, et al, "Application Research of Petroleum Basic Data Mining System Based on Intelligent Computing and Decision Tree Algorithm," in Wireless Communications and Mobile Computing 2022.

[12] A. Navada, A. N. Ansari, S. Patil, B. A. Sonkamble, "Overview of Use Decision Tree algorithms in Machine Learning," in IEEE Control and System Graduate Research Colloquium, 2011.

[13] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, "Machine learning: An artificial intelligence approach," in Springer Science & Business Media, 2013.

[14] M. S. Abd Rahman, N. A. A. Jamaludin, Z. Zainol, and T. M. T. Sembok, "Machine Learning Algorithm Model for Improving Business Decisions Making in Upstream Oil & Gas," in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021: IEEE, pp. 1-5

[15] A. S. H. Lee, Z. Yusoff, Z. Zainol, and V. Pillai, "Know your hotels well! -- An Online Review Analysis using Text Analytics," International Journal of Engineering & Technology, vol. 7, no. 4.31, 2018, pp. 341-347.

[16] Q. Cui, F. S. Bai, B. Gao, and T. Y. Liu, "Global Optimisation for Advertisement Selection," in Sponsored Search. Journal of Computer Science and Technology, Vol 30(2), 2015, pp. 295-310.

[17] P. N. E. Nohuddin, Z. Zainol, and M. H. A. Hijazi, "Study of B40 Schoolchildren Lifestyles and Academic Performance using Association Rule Mining," presented at the Global Research Conference 2020 (GRACE 2020), 2020.

[18] K. I. M. Fadilah, Z. Zainol, M. Ebrahim, and A. S. H. Lee, "Covid-19 Effect On Undergraduate Computing Students' Performance At Higher Education: Pilot Study," in 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2021, vol. 6: IEEE, pp. 1-6.

[19] I. Kononenko, "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. Artificial Intelligence," in medicine, Vol 23(1), 2001, pp. 89-109.

[20] A. Lee, N. Claudia, Z. Zainol, and K. Chan, "Decision Tree: Customer Churn Analysis for a Loyalty Program Using Data Mining Algorithm," in International Conference on Soft Computing in Data Science, 2019: Springer, pp. 14-27.

[21] N. K.-W. Chan, A. S.-H. Lee, and Z. Zainol, "Predicting employee health risks using classification ensemble model," in 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP), 2021: IEEE, pp. 52-58.

[22] M. J. Paul et al., "Social media mining for public health monitoring and surveillance," in Biocomputing 2016: Proceedings of the Pacific symposium, 2016: World Scientific, pp. 468-479.

[23] X. Liu, H. Shin, and A. C. Burns, "Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing," Journal of Business Research, vol. 125, pp. 815-826, 2021.

[24] M. Rouse, "Machine Learning Definition," 2011. (*http://whatis.techtarget.com/definition/machine-learning)*

[25] Z. Zainol, P. N. E. Nohuddin, A. S. H. Lee, N. F. Ibrahim, L. H. Yee, and K. A. Majid, "Analysing political candidates' popularity on social media using POPularity MONitoring (POPMON)," SEARCH Journal of Media and Communication Research, no. Special Issue: GRACE 2020 Conference, pp. 39-55, 2021. (*https://fslmjournals.taylors.edu.my/wp-content/uploads/SEARCH/SEARCH-2021-Special-Issue-GRACE2020/SEARCH-2021-Special-Issue-GRACE2020.pdf*)

[26] A. Simon et al, "An Overview of Machine Learning and its Applications," in International Journal of Electrical Sciences & Engineering (IJESE), Vol 1(1), 2015, pp. 22-24.

[27] T.M. Mitchell, "The discipline of machine learning," Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.

[28] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in IBM Journal of Research and Development, Vol 3(3), pp. 211–229.

[29] C. Apte, L. Morgenstern, and S. J. Hong, "AI," at IBM Research, IEEE Intelligent Systems, Vol 15(6), 2000, pp. 51-57.

[30] P. Tan, M. Steinbach, and V. Kumar "Cluster Analysis: Basic Concepts and Algorithms," in Introduction to Data Mining, Addison-Wesley, Boston, 2005.

[31] A. Azevedo, and M. F. Santos, "Business Intelligence: State of the Art, Trends, and Open Issues," 2009 in International Conference on Knowledge Management and Information Sharing, 2009, pp. 296-300

[32] B. Tahar and A. Mohsin, "Classification Based on Decision Tree Algorithm for Machine Learning," in 2021 Journal of Applied Science & Technology Trends, Vol 02(1), 2021, pp.20-28, 2021. *(ISSN: 2708-0757*)

[33] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, "General Issues in Machine Learning," in Machine Learning: An Artificial Intelligence Approach, Chapter 1, 1984. *(ISBN 978-3-662-12405-5 (e-Book), DOI 10.1007/978-3-662-12405-5)*

[34] P. L. Tu and J. Y. Chung, "A new Decision-tree Classification Algorithm for Machine learning," in Proceedings Fourth International Conference on Tools with Artificial Intelligence TAI '92, 1992, pp. 370-377. (*DOI: 10.1109/TAI.1992.246431*)

[35] M. Somvanshi, P. Chavan, S. Tambade, and S.V. Shinde, "A Review of Machine Learning Techniques using Decision Tree and Support Vector Machine," in International Conference on Computing Communication Control and Automation (ICCUBEA), 2016, pp. 1-7, (*Doi 10.1109/ICCUBEA. 2016.7860040*)

[36] A. Cook, P. Wu, K. Mengersen, "Machine Learning and Visual Analytics for Consulting Business Decision Support," Queensland University of Technology Brisbane, Australia, 2015.

[37] I. Baturynska and K. Martinsen, "Prediction of Geometry Deviations in Additive Manufactured Parts: Comparison of Linear Regression with Machine Learning Algorithms," in Journal of Intelligent Manufacturing, 2020. (*https://doi.org/10.1007/s10845-020-01567-0)*

[38] Bahaloo, Saeed, M. Mehrizadeh, and A. N. Marghmaleki. "Review of the application of artificial intelligence techniques in petroleum operations," Petroleum Research, 2022.

[39] Gryzlov, Anton, S. Safonov, and M. Arsalan. "Intelligent Production Monitoring with Continuous Deep Learning Models," in SPE Journal 27.02, 2022, pp. 1304-1320.

[40] Sieberer, Martin, and T. Clemens, "Hydrocarbon Field (Re-) Development as Markov Decision Process," in SPE Reservoir Evaluation & Engineering 25.02, 2022, pp. 273-286.

[41] L. Fan, et al, "A systematic method for the optimization of gas supply reliability in natural gas pipeline network based on Bayesian networks and deep reinforcement learning," in Reliability Engineering & System Safety 225, 2022: 108613.

[42] H.H Lou, et al. "A novel zone-based machine learning approach for the prediction of the performance of industrial flares." Computers & Chemical Engineering 162, 2022: 107795. (*DOI:10.1016/j.compchemeng.2022.107795*)

[43] L. Anderson, and L. Thompson, "The impact of uncertainty on expected monetary value in decision-making," in Organizational Behavior and Human Decision Processes, 139, 2017, pp. 23-35.

[44] E. B. Priyanka, et al, "Digital twin for oil pipeline risk estimation using prognostic and machine learning techniques," in Journal of industrial information Integration 26, 2022: 100272.

[45] M. D. Choudhry, et al. "Machine Learning Frameworks for Industrial Internet of Things (IIoT): A Comprehensive Analysis," in 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT). IEEE, 2022.

[46] Desai, J. Nitesh, S. Pandian, and R. V. Kumar, "Big data analytics in upstream oil and gas industries for sustainable exploration and development: A review," in Environmental Technology & Innovation 21, 2021: 101186.