Understanding Decision Trees

Vedant Marwadi - 23208466

Artificial Intelligence,
Master of Computer and Information Sciences,
Auckland University of Technology

Behind every breakthrough in predictive technology lies a simple question: how do we make the best choice from a sea of possibilities? The answer, in many cases, comes in the form of decision trees (Blockeel et al., 2023). This essay first explains the foundations of artificial intelligence (AI) and the concept of machine learning (ML) in order to understand how predictive models function. The main focus, however, is on exploring the evolution of decision trees, tracing their roots from the theoretical foundations to their modern applications. It also examines how decision trees process data, highlighting their strengths and addressing common challenges. Furthermore, the discussion extends to contemporary advancements that enhance decision tree performance. Finally, the essay draws attention to recent research from the International Joint Conference on Artificial Intelligence, which introduces novel optimizations in decision tree-based strategies, showcasing how these innovations are shaping AI's future.

Amid the rapid and relentless march of technological progress, AI has emerged as a revolutionary force, pushing the limits of human potential (Păvăloaia & Necula, 2023). Its roots can be traced back to the 1950s when visionary computer scientists like Alan Turing first speculated about the possibility of machines capable of thinking and learning (Turing, 1950). AI, at its core, is about engineering machines to think and act in ways that typically demand human intellect and insight (Hassani et al., 2020). However, the journey from Turing's early ideas to today's advanced AI was not immediate. The need for AI grew slowly over time, mainly because of the exponential increase in data and the growing difficulty of the problems we encountered as a result. It was only when big data and high-performance computers became available that AI could really begin to grow and show its true potential (van Assen et al., 2022).

A key driver behind this transformation is machine learning, one of the most crucial components of AI. While AI sets the broader goal of mimicking human intelligence, ML provides the mechanism through which this can be achieved (Manakitsa et al., 2024). It focuses on developing algorithms that improve their performance on specific tasks through experience, without the need for explicit programming (Mitchell, 2013). By teaching systems to self-improve through data, recognise complex trends, and operate with minimal human input, ML forms the ever-shifting engine behind modern AI solutions (Mukhamediev et al., 2022).

To understand how ML achieves this, it is important to examine its process, which is not a single leap into complexity, but rather a structured journey. It begins with collecting large amounts of data. But the data alone is not enough; first, it needs to be cleaned and organized to remove any errors that could confuse the model (Mahesh, 2019). Once the data is ready, it is split into two sets: the training set, which helps the model learn, and the test set, used to check how well it has learned. The model goes through a cycle of making predictions, learning from its mistakes, and adjusting its internal settings during training, much like how humans improve by practicing and correcting errors. After training, the model faces its final challenge: performing on new, unseen data from the test set. This step checks whether the model can handle real-world situations, not just the examples it has already seen. If it passes this test, the model can then be deployed to make predictions in real time based on new information (Bi et al., 2019).

Building on this structured process, the specific learning approach taken, whether supervised, unsupervised, or reinforcement, dictates how the model absorbs and adapts to the data. In supervised learning, the model relies on a dataset where every input is explicitly linked to a known output, allowing it to progressively map inputs to the correct outcomes by example (Mukhamediev et al., 2022). Unsupervised learning, on the other hand, removes this guidance. Instead of predefined labels, the model must shift through raw data to detect underlying patterns or groupings on its own, without direct supervision. Reinforcement learning stands apart by immersing the model in a dynamic environment where actions trigger responses which the model uses to adjust its behavior over time and maximize its cumulative success (Bhattacharjee & Bhattacharjee, 2020)

Among the various supervised learning algorithms, one of the most intuitive and effective is the decision tree (Quinlan, 1986). What makes decision trees unique is their origin in decision theory, a branch of economics and mathematics that deals with making choices under uncertainty (Fürnkranz, 2010). Long before their use in machine learning, decision trees were inspired by the work of early 20th-century thinkers John von Neumann and Oskar Morgenstern, who sought to model human decision-making in uncertain situations. They envisioned decisions as branching paths where each choice leads to new possibilities, breaking down problems into manageable, smaller decisions (Von Neumann & Morgenstern, 1947).

Building on this theoretical foundation, the early stages of decision tree development took the form of simple graphical tools. These early models were not computational, but they introduced essential ideas such as outcome probabilities, branching paths, and decision nodes, which were instrumental in transforming decision trees (Demirović & Stuckey, 2021). The real breakthrough, however, came when decision trees moved from theoretical frameworks and visual aids to fully-fledged computational tools. This transition began in the 1960s and 1970s, when researchers started developing algorithms capable of making decisions based on data. The most significant leap came forward in 1979, when Australian computer scientist Ross Quinlan introduced the Interactive Dichotomizer 3 (ID3) algorithm. ID3 was a landmark in the history of decision trees, as it was one of the first algorithms that enabled decision trees to be used effectively in machine learning tasks (Quinlan, 1986).

Quinlan's ID3 algorithm was just the beginning. In 1993, he developed the C4.5 algorithm, which improved upon ID3 by enabling decision trees to handle continuous attributes and by introducing pruning methods to prevent overfitting, making the models more reliable (Tsoi & Pearson, 1990). Around the same time, another important algorithm was developed: the CART (Classification and Regression Trees) algorithm, introduced by Leo Breiman and colleagues in 1984. While ID3 was primarily focused on classification tasks, CART expanded the versatility of decision trees by supporting both classification and regression, making it a more flexible tool in areas like finance and healthcare, where both types of tasks are commonly encountered (Salzberg, 1994). Both methods aimed to create decision trees that could accurately classify data, but they took different approaches to splitting and optimizing the tree structure. Together, these two algorithms formed the foundation for modern decision tree techniques.

With these advancements in place, the 1990s and early 2000s marked a period of rapid growth for machine learning, and decision trees took center stage as powerful tools for data mining and predictive modeling (Bock et al., 2019). The basic structure of a decision tree consists of nodes, branches, and leaves. The process begins at the root node, where the algorithm examines all features and selects the ones that best separates the data into meaningful groups. This decision is not random, it is guided by metrics like Gini impurity or information gain. Gini impurity helps estimate how mixed up the data in a node is. A node with high Gini impurity means the data is a messy mix of different categories. The goal of each split is to make the groups more organized by reducing this mix (Menze et al., 2009).

On the other hand, information gain measures how much “confusion” or uncertainty is reduced by making a split. A higher information gain means the split does a good job at separating different categories. This way, the algorithm picks the feature that best separates the data (Alhaj et al., 2016). The process continues until the tree reaches a stopping point—this could be a maximum tree depth, a minimum number of samples in a node, or when the nodes become “pure” enough (meaning they mostly contain data from one category). By stopping at the right time, the tree avoids being too complex (memorizing every detail) or too simple (missing important patterns) (Alpaydın, 2010).

This approach naturally leads to several key advantages, with transparency being the most prominent. Unlike "black-box" models such as neural networks, decision trees offer clear decision paths, making them straightforward for users to understand and interpret (Hassani et al., 2020). This interpretability is especially valuable in fields such as healthcare and finance, where decision-makers must fully comprehend the reasoning behind a model’s predictions. Boruah and his team demonstrated this transparency by developing a Transparent Expert System of Rules, which extracted decision rules from decision trees while enhancing comprehensibility through the reduction of redundant rules (Boruah et al., 2022). Similarly, Sprogar and colleagues emphasized the importance of decision trees’ transparency for critical applications, such as medical systems, where explaining decisions is vital (Sprogar et al., 2001). To further enhance both transparency and user comprehension, Zhang and the group introduced cascading decision trees, which shortened the explanation depth, making decision trees even more user-friendly (Zhang et al., 2020).

In addition to their transparency, decision trees excel in adaptability. They are capable of handling both categorical and numerical data, while also modeling complex relationships between input variables and target outputs. Dumitrescu and associates illustrated this versatility by combining decision trees with logistic regression to improve credit scoring models, highlighting their adaptability within the financial sector. Almuallim and peers extended this view, demonstrating that decision trees are also suitable for survival analysis (Almuallim et al., 2002). These experiments demonstrate the flexibility of decision tress across multiple domains.

Another key strength of decision trees is their non-parametric nature, which allows them to handle various data types without needing assumptions about the underlying data distribution (Costa & Pedreira, 2023). Mitrofanov and co-researchers highlighted this advantage, noting that decision trees can accommodate diverse datasets without requiring prior data transformations (Mitrofanov & Semenkin, 2021). Talekar further supported this by showing that decision trees perform well across a wide range of tasks, from regression to classification, without necessitating data normalization (Talekar, 2020). This flexibility makes decision trees highly applicable in a variety of real-world scenarios, making them an indispensable tool in machine learning.

Despite their many strengths, decision trees also have some notable limitations. One of the most significant challenges is their tendency to overfit the training data. As trees grow deeper, they create highly specific splits that can lead to overfitting, particularly when the data is noisy or contains irrelevant features (Quinlan, 1986). Moreover, they can be surprisingly unstable. Small changes in the dataset can lead to dramatically different tree structures, affecting the consistency of predictions (Piramuthu, 2008). Furthermore, decision trees often falter on imbalanced datasets, leaning heavily toward the majority class and neglecting the minority, which can result in misclassification in important cases (Leevy et al., 2018).

However, advancements in machine learning have offered solutions to many of these challenges. In particular, the development of ensemble learning techniques has transformed the performance and reliability of decision trees. By combining the strengths of multiple decision trees, ensemble methods such as random forest and gradient boosting help mitigate issues like overfitting and instability. For instance, random forest constructs multiple trees using various subsets of the data, averaging their predictions to improve accuracy and reduce overfitting (Breiman, 2001). Additionally, strategies like oversampling the minority class or undersampling the majority class during training have been found to enhance performance on imbalanced datasets (Yang et al., 2024). Recent developments have also integrated decision trees with other machine learning techniques to create hybrid models that harness the advantages of various algorithms. For example, decision trees have been combined with neural networks and support vector machines to improve their performance on complex tasks (Zhou, 2012).

Moreover, innovation in decision tree optimization continues to evolve, pushing the boundaries of their applicability beyond traditional tasks. For instance, research presented at the International Joint Conference on Artificial Intelligence (IJCAI) highlighted the use of Monte Carlo Tree Search (MCTS) to optimize decision tree-based strategies for Satisfiability Modulo Theories (SMT) solvers, which significantly outperformed state-of-the-art solvers by solving 42.7% more instances on challenging benchmarks (Lu et al., 2024). In parallel, advancements in eXplainable AI (XAI) have further explored how decision trees can be used to generate local, abductive, and contrastive explanations in AI systems. These explanations, particularly when incorporating domain-specific knowledge, improve interpretability while maintaining computational feasibility (Audemard et al., 2024). Together, these studies highlight how innovations in decision tree research are enhancing both the efficiency and transparency of AI, with wide-ranging applications from technical problem-solving to explainable decision-making.

In conclusion, it can be said that decision trees have evolved from simple theoretical constructs into highly sophisticated tools at the heart of modern artificial intelligence and machine learning. Their transparency, adaptability, and non-parametric nature have made them indispensable in various domains, from healthcare to finance. Despite challenges such as overfitting and instability, advancements like ensemble methods, hybrid models, and innovations in explainable AI have addressed these issues and expanded their applicability. Nonetheless, future research should focus on further refining decision trees to enhance their robustness, particularly in handling imbalanced datasets and improving stability under small data variations. Additionally, integrating decision trees more deeply with emerging AI techniques, such as deep learning and reinforcement learning, could unlock new possibilities in tackling more complex and dynamic problems. Moreover, research into explainable AI should continue to emphasize creating more intuitive and domain-specific explanations, ensuring decision trees remain both powerful and user-friendly. By pursuing these directions, decision trees can continue to play a pivotal role in shaping the future of AI.

References

Alhaj, T. A., Siraj, M. M., Zainal, A., Elshoush, H. T., & Elhaj, F. (2016). Feature selection using information gain for improved structural-based alert correlation. *PloS one*, 11(11), e0166017.

Almuallim, H., Kaneda, S., & Akiba, Y. (2002, January). 3 - Development and Applications of Decision Trees. In C. T. Leondes (Ed.), *Expert Systems* (pp. 53–77). Academic Press. https://doi.org/10.1016/B978-012443880-4/50047-8

Alpaydın, E. (2010). Machine learning : The new AI. *(No Title)*. Retrieved September 17, 2024, from https://cir.nii.ac.jp/crid/1130282272626148736

Audemard, G., Lagniez, J.-M., Marquis, P., & Szczepanski, N. (2024). Deriving Provably Correct Explanations for Decision Trees: The Impact of Domain Theories. *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*, 3688–3696. https://doi.org/10.24963/ijcai.2024/408

Bhattacharjee, J., & Bhattacharjee, J. (2020). Unsupervised and reinforcement learning. *Practical Machine Learning with Rust: Creating Intelligent Applications in Rust*, 107–139.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist [Publisher: Oxford University Press (OUP)]. *American Journal of Epidemiology*. https://doi.org/10.1093/aje/kwz189

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: From efficient prediction to responsible ai. *Frontiers in Artificial Intelligence*, 6, 1124553.

Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., & Klusemann, B. (2019). A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics [Publisher: Frontiers]. *Frontiers in Materials*, 6. https://doi.org/10.3389/fmats.2019.00110

Boruah, A. N., Biswas, S., & Bandyopadhyay, S. (2022). Rule extraction from decision tree: Transparent expert system of rules. *Concurrency and Computation: Practice and Experience*, 34. https://doi.org/10.1002/cpe.6935

Breiman. (2001). Classification and Regression Trees | Leo Breiman, Jerome Friedman, R. Retrieved September 17, 2024, from https://www.taylorfrancis.com/books/mono/10.1201/9781315139470/classification-regression-trees-leo-breiman-jerome-friedman-olshen-charles-stone

Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765–4800. https://doi.org/10.1007/s10462-022-10275-5

Demirović, E., & Stuckey, P. J. (2021). Optimal Decision Trees for Nonlinear Metrics. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 3733–3741. https://doi.org/10.1609/aaai.v35i5.16490

Fürnkranz, J. (2010). Decision Tree. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 263–267). Springer US. https://doi.org/10.1007/978-0-387-30164-8_204

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *AI*, 1(2), 143–155. https://doi.org/10.3390/ai1020008

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, 5(1), 42. https://doi.org/10.1186/s40537-018-0151-6

Lu, Z., Siemer, S., Jha, P., Day, J., Manea, F., & Ganesh, V. (2024, April). Layered and Staged Monte Carlo Tree Search for SMT Strategy Synthesis [arXiv:2401.17159 [cs]]. Retrieved September 17, 2024, from http://arxiv.org/abs/2401.17159

Mahesh, B. (2019, January). *Machine Learning Algorithms -A Review* (Vol. 9) [Journal Abbreviation: International Journal of Science and Research (IJSR) Publication Title: International Journal of Science and Research (IJSR)]. https://doi.org/10.21275/ART20203995

Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Technologies*, 12(2), 15. https://doi.org/10.3390/technologies12020015

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10, 1–16.

Mitchell, T. M. (2013). *Machine learning* (Nachdr.). McGraw-Hill.

Mitrofanov, S., & Semenkin, E. (2021). Tree retraining in the decision tree learning algorithm. *IOP Conference Series: Materials Science and Engineering*, 1047. https://doi.org/10.1088/1757-899X/1047/1/012082

Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., Muhamedijeva, E., & Yelis, M. (2022). Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics*, 10(15), 2552. https://doi.org/10.3390/math10152552

Păvăloaia, V.-D., & Necula, S.-C. (2023). Artificial Intelligence as a Disruptive Technology—A Systematic Literature Review [Number: 5 Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, 12(5), 1102. https://doi.org/10.3390/electronics12051102

Piramuthu, S. (2008). Input data for decision trees. *Expert Syst. Appl.*, 34, 1220–1226. https://doi.org/10.1016/j.eswa.2006.12.030

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Salzberg, S. L. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning*, 16(3), 235–240. https://doi.org/10.1007/BF00993309

Sprogar, M., Kokol, P., Zorman, M., Podgorelec, V., Lhotská, L., & Kléma, J. (2001). Evolving groups of basic decision trees. *Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001*, 183–188. https://doi.org/10.1109/CBMS.2001.941718

Talekar, B. (2020). A Detailed Review on Decision Tree and Random Forest. *Bioscience Biotechnology Research Communications*. https://doi.org/10.21786/BBRC/13.14/57

Tsoi, A. C., & Pearson, R. (1990). Comparison of three classification techniques: CART, C4.5 and Multi-Layer Perceptrons. *Advances in Neural Information Processing Systems*, 3. Retrieved September 18, 2024, from https://proceedings.neurips.cc/paper/1990/hash/1068c6e4c8051cfd4e9ea8072e3189e2-Abstract.html

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, New Series*, 59(236), 433–460. http://www.jstor.org/stable/2251299 van Assen, M., Muscogiuri, E., Tessarin, G., & De Cecco, C. N. (2022). Artificial Intelligence: A Century-Old Story. In C. N. De Cecco, M. van Assen & T. Leiner (Eds.), *Artificial Intelligence in Cardiothoracic Imaging* (pp. 3–13). Springer International Publishing. https://doi.org/10.1007/978-3-030-92087-6_1

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior, 2nd rev. ed* [Pages: xviii, 641]. Princeton University Press.

Yang, C., Fridgeirsson, E. A., Kors, J. A., Reps, J. M., & Rijnbeek, P. R. (2024). Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. *Journal of big data*, 11(1), 7.

Zhang, J., Santolucito, M., & Piskac, R. (2020). Succinct Explanations With Cascading Decision Trees. *ArXiv*, abs/2010.06631. Retrieved September 17, 2024, from https://consensus.app/papers/explanations-with-cascading-decision-trees-zhang/fdeba58d2d2f55eca3c94046dc0f9b11/

Zhou, Z.-H. (2012, June). *Ensemble Methods: Foundations and Algorithms* (0th ed.). Chapman; Hall/CRC. https://doi.org/10.1201/b12207