**Decision Trees**

**The CART Decision Tree used to classify numerical Wine Quality data**

In implementing decision trees (DTs), I selected a Wine Quality dataset for classification based on chemical measurements. This dataset was particularly compelling because it connected to a real-world commercial problem and offers rich analytical possibilities. With 1,599 instances and 11 chemical features, it provided an excellent framework for showing what DTs can do in meaningful situations. Predicting wine quality based on chemical properties is crucial for the wine industry, where small changes in chemical composition can greatly affect how much consumers enjoy the product and how well it sells. While the Weather dataset was useful for initially testing how DTs handle categorical data, the Wine Quality dataset presented a more interesting challenge by combining numerical measurements with quality ratings.

The structure of the Wine Quality dataset worked especially well for DT analysis. Its continuous features, like acidity, alcohol content, and sulphates, fit naturally with how CART DTs make binary splits, allowing the algorithm to find clear thresholds for predicting quality. These chemical properties are particularly useful because winemakers can adjust them during production to improve their product. Additionally, the quality ratings in the dataset range from 0 to 10 in order of increasing quality, making DTs particularly appropriate since they're good at handling this kind of stepped relationship between values. When the model makes mistakes, it's less problematic if it confuses similar quality levels (like 6 and 7) than if it confuses very different quality levels (like 3 and 8). DTs excel at understanding these kinds of ordered relationships, providing a clear framework for assessing quality.

Before building the DT, I developed several hypotheses based on established wine research and industry knowledge.

1. The main hypothesis was that alcohol content would be the strongest predictor of wine quality, since alcohol plays a fundamental role in how the wine feels and tastes.
2. Secondary hypotheses suggested that sugar levels would strongly influence quality ratings, particularly in determining how balanced and complex the wine tastes.
3. I also hypothesised that sulphur dioxide levels would have minimal impact on quality predictions, since these compounds are mainly used to preserve the wine rather than enhance its taste. These hypotheses gave a way to compare the DT findings against what wine experts typically assume.

The CART DT proved to be a better choice than ID3 for this analysis because it handles continuous data more effectively. While ID3 works well with categorical features, its approach of creating multiple splits at once would have resulted in too many branches when dealing with 11 continuous features, making the results harder to understand. CART's method of creating binary splits, combined with using Gini impurity to decide where to split, allowed for a more effective model. These features made the calculations more efficient and also provided more stable results when working with continuous chemical measurements. These advantages made CART the ideal choice for this dataset, ensuring the model could identify meaningful patterns without becoming overly complex.

The CART model achieved 67.73% accuracy when using a maximum depth of 7 and requiring at least 20 samples for each split. While this might not sound impressive at first, it represents strong predictive ability in a problem with 11 possible classes, far better than the 9% accuracy you'd get from random guessing. Most of the model's mistakes happened between neighbouring quality levels, showing that it understood the fundamental structure of wine quality. These results highlight how well the model could extract meaningful insights, even in a complex scenario with many possible outcomes. The settings were carefully chosen to balance understanding patterns while avoiding overfitting. A depth limit of 7 prevented the model from becoming too complex while keeping decision paths clear, and requiring 20 samples per split ensured statistical reliability while maintaining enough detail to catch important patterns.

When analysing which features mattered most, I found surprising results that challenged my initial hypotheses. Fixed acidity emerged as the most important predictor, appearing in 47 decision points, followed A graph of blue squares with white text

Description automatically generatedby volatile acidity with 27 points and total sulphur dioxide with 15 points. This directly contradicted my main hypothesis about alcohol content being the most important factor. Furthermore, the significant impact of sulphur dioxide challenged my assumption about its minimal role, suggesting that preservative compounds might affect perceived quality in more complex ways than previously thought. The relatively modest influence of residual sugar also went against my secondary hypothesis, indicating that sweetness plays a more subtle role in quality assessment than expected.

Beyond just accuracy numbers, the model provided practical value by connecting theoretical analysis with real-world applications. The thresholds it identified offer clear guidance for winemakers to improve their processes, and the ranking of important features suggests where to focus future research. For instance, the unexpected importance of acidity-related variables suggests that collecting more detailed acid measurements could be valuable.

Overall, the Wine Quality dataset proved to be a fairly strong choice for implementing DTs. The CART algorithm effectively handled the continuous features, captured the ordered relationships in the data, and produced results that were easy to interpret. The model challenged several common assumptions about what determines wine quality and also revealed unexpected patterns. While the model's accuracy suggests room for improvement, its insights about important features and specific chemical thresholds provide substantial practical value for winemaking. This analysis demonstrated both the strengths of DTs and their ability to uncover surprising relationships in complex, real-world datasets, pointing toward new directions for both research and practical applications in the wine industry.

**Recent Advances of Decision Trees**

In terms of approximation capabilities of DTs, several major developments have emerged. Traditional DT use univariate splits that evaluate a single attribute at a time, but recent research has explored multivariate trees that consider multiple attributes in each split. While these enhance approximation capacity, they bring challenges in interpretability and computational cost. Innovations like Sparse ADTrees and Fisher's ADTree have addressed these challenges by using Linear Discriminant Analysis to find optimal splits while maintaining efficiency. Another significant advancement has been in model trees, which employ predictive models in their leaves instead of simple labels or constant numbers. This advancement extends regression trees beyond piecewise constant functions and has been applied to classification tasks through predictive models such as Naive Bayes and logistic regression.

Generalisation has seen strong progress via several approaches. Traditional pre-pruning and post-pruning techniques continue to be essential elements of DT learning. A particularly noticeable development has been in optimal trees - algorithms that can find the mathematically best tree structure for a given dataset. While traditional greedy approaches like CART may produce sub-optimal solutions, new algorithms using mathematical programming and Boolean satisfiability have emerged. Notable examples include OCT (Optimal Classification Trees) and DL8.5, which can guarantee finding the optimal tree structure while remaining computationally feasible for moderately-sized datasets.

Interpretability has received increased attention in recent research. Tree size is the aspect most commonly associated with interpretability, with shallow trees regarded as more interpretable than deeper ones. However, the relationship between tree size and interpretability is complex and depends on various factors. Recent research has explored visual pruning techniques that allow direct visual interpretation without requiring a separate validation set. Additionally, methods have been developed to push broader and more accurate rules to the top of the tree, reducing the need to investigate deeper leaves.

A particularly innovative recent development has been the emergence of quantum decision trees (QDTs). These maintain the directed graph structure of classical trees but employ quantum kernels at split nodes to generate separating hyperplanes in higher-dimensional quantum feature spaces. The use of Nyström quantum kernel estimation helps prevent overfitting while maintaining the benefits of quantum feature spaces. QDTs have shown particular promise in handling multi-class problems without additional computational overhead, unlike quantum Support Vector Machines (SVMs) which require specific strategies for multiple classes. Both theoretical analysis and numerical experiments have shown QDTs perform better than quantum SVMs while needing fewer kernel estimations.

Gradient-based trees represent another significant advance, combining traditional DT structures with optimisation techniques. These approaches have shown improvements in accuracy while attempting to maintain interpretability through various techniques, including post-hoc interpretations and path labelling. Research has also advanced in handling specific data characteristics, including unlabelled data through semi-supervised learning approaches, cost-sensitive learning, and monotonic constraints.

These developments represent significant progress in DT research, enhancing their capabilities while addressing traditional limitations in both classical and quantum computing contexts.

**A Real-World application of a Decision Tree**

The Large Hadron Collider (LHC) demonstrates a particularly sophisticated implementation of Decision Trees through Boosted Decision Trees (BDTs). These ensemble machine learning models combine multiple shallow DTs into a more robust predictive framework, operating iteratively with subsequent trees focusing on previously misclassified data points. The final predictive outcome is determined through weighted voting mechanisms, enabling BDTs to accommodate complex data distributions while maintaining high accuracy, establishing them as a significant methodological choice in applications demanding both computational efficiency and statistical reliability.

In the context of the LHC, these BDTs serve a crucial function in processing the extraordinary volume of data generated by proton collisions occurring at 25-nanosecond intervals. These collisions produce data volumes in the order of tens of terabytes per second, presenting substantial data management challenges. To address this, the facility employs a sophisticated filtering system, designated as the trigger, wherein the Level-1 Trigger (L1T) must process approximately 40 million events per second while maintaining decision-making capabilities within a 10-microsecond window.

A particularly noteworthy implementation of BDTs can be observed in the Compact Muon Solenoid (CMS) experiment's muon detection system, where the implementation has demonstrated threefold enhancement in the filtering of low-energy muons compared to preceding methodologies. Through the implementation of hls4ml, these BDTs demonstrate remarkable computational efficiency, processing information within 12 clock cycles (equivalent to 60 nanoseconds at 200 MHz), rendering them particularly suitable for real-time trigger system applications. In particle jet classification tasks, BDTs have demonstrated comparable performance to Deep Neural Networks while requiring reduced computational resources, positioning them as an optimal solution for physics applications that necessitate both rapid processing capabilities and high accuracy requirements.

These advancements in BDT methodology have provided physicists with enhanced analytical capabilities, facilitating improved jet classification, more precise muon energy measurements, and enhanced detection of rare particle events. The successful implementation of BDTs in addressing these complex physics challenges demonstrates their significant utility in scientific applications requiring both precise analytical outcomes and exceptional processing efficiency. This application exemplifies how theoretical advances in decision tree methodology can be successfully translated into practical solutions for complex scientific challenges.

References

* Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F.D., Matos, T. and Reis, J., 2009. Using data mining for wine quality assessment. *Discovery Science, 12th International Conference, DS 2009*. Berlin: Springer. - <https://doi.org/10.1007/978-3-642-04747-3_8>
* Dataset Source: UCI Machine Learning Repository. (2009). Wine Quality Data Set. -<https://archive.ics.uci.edu/ml/datasets/wine+quality>
* GeeksforGeeks, n.d. CART (Classification And Regression Tree) in Machine Learning. *GeeksforGeeks*. -<https://www.geeksforgeeks.org/cart-classification-and-regression-tree-in-machine-learning/>
* Srikumar, M., Hill, C.D. and Hollenberg, L.C.L., 2024. A kernel-based quantum random forest for improved classification. *Quantum Machine Intelligence* - <https://doi.org/10.1007/s42484-023-00131-2>
* Costa, V.G. and Pedreira, C.E., 2023. Recent advances in decision trees: an updated survey. Artificial Intelligence Review - <https://doi.org/10.1007/s10462-022-10275-5>
* Summers, S., Di Guglielmo, G., Duarte, J.M., Harris, P., and others, 2020. Fast inference of Boosted Decision Trees in FPGAs for particle physics. - <https://doi.org/10.1088/1748-0221/15/05/P05026>
* Woodruff, K., 2017. Introduction to Boosted Decision Trees. Machine Learning Group Meeting, New Mexico State University -<https://indico.fnal.gov/event/15356/contributions/31377/attachments/19671/24560/DecisionTrees.pdf>