Predicting Beer Styles

Ava Weisheit
CSCI-B365
Indiana University
Bloomington, US
avaweish@iu.edu

***Abstract*- This project proposes a machine learning approach to classify beer styles based on their chemical and sensory attributes. Using the "Top Beer Information" dataset from Kaggle we will develop a decision tree classifier to predict beer styles. The goal is to explore how measurable characteristics influence the classification of beer to evaluate the effectiveness of a decision tree model in identifying beer styles.**

I . INTRODUCTION

Beer is one of the most widely consumed alcoholic beverages in the world, with thousands of unique styles and flavor profiles. As college students, many of us are no strangers to beer, but have you ever wondered what makes beers taste different? Beer styles are defined by different attributes such as alcohol content, bitterness, sweetness, and flavor profiles.

In this project, we aim to classify different beer styles based on their chemical and sensory characteristics. Our dataset sources from Kaggle includes thousands of individual beers with different attributes, each one classifying the beer into a specific “style”. The dataset includes attributes such as alcohol by volume (ABV), bitterness units (IBU), and flavor trait ratings. These features will help us to explore the correlation between taste and style to train a model that can predict a beers style by analyzing its attributes. The objective of this project is to develop a machine learning model using a **decision tree classifier** to accurately predict a beer’s style based on its measurable attributes.

II . DATA DESCRIPTION AND STATISTICS

The dataset used in this project was obtained from Kaggle: [Top Beer Information dataset by Stephen Polozoff]. It contains 5,558 entries and 21 features, which span both categorical and numerical attributes essential for beer classification. These features include Basic identifiers such as the beer's name, style, style key, and brewery, which help distinguish different types and brands of beer, descriptive attributes such as description, alcohol by volume (ABV), and average rating, which offer insights into the beer’s composition and popularity, and flavor descriptors which form the core of this dataset’s analytical value. These flavor descriptors include numerical ratings for Astringency, Body, Alcohol, Bitter, Sweet, Sour, Salty, Fruits, Hoppy, Spices, and

Malty. Each of these sensory descriptors is quantified, allowing us to construct a comprehensive flavor profile for every beer.

III . IMPLEMENTATION PLAN

To build a reliable beer style classification model, we followed a structured pipeline consisting of data preprocessing, feature selection, model development, and evaluation. First, we addressed missing values by replacing them with the mean of each respective column. We removed non-informative columns such as “Name,” “Key,” and “Brewery” to reduce noise and improve model performance. Since beer styles in the original dataset were highly imbalanced, we simplified the problem by selecting the 10 most frequent styles and randomly sampling 50 entries from each, resulting in a balanced dataset of 500 total samples.

We selected a set of meaningful features that reflect the chemical and sensory makeup of each beer. These flavor descriptors like Astringency, Body, Alcohol, Bitter, Sweet, Sour, Salty, Fruits, Hoppy, Spices, and Malty. These attributes were chosen for their direct relevance to beer classification based on flavor. Although we could have tested on more chemical properties, we wanted to see if the tree could distinguish sensory traits of beer styles.

For our classification model, we used a Decision Tree Classifier due to its interpretability, flexibility with unscaled data, and to make it visualizable. We split the data into 70% training and 30% testing sets to ensure both model learning and generalization. We experimented with several hyperparameters and found that limiting the tree’s maximum depth to 5 provided a good balance between performance and interpretability.

To evaluate the performance of our model, we used a variety of metrics including accuracy, precision, recall, F1-score, and a confusion matrix. These metrics allowed us to analyze not just the overall correctness of the model, but also how well it performed across each of the 10 selected beer styles.

IV . METHODOLOGY

A. Handling Missing Values:

Missing values will be handled by replacing them with the feature’s mean. Non-informative columns such as “Name,” “Key,” and “Brewery” will be removed to reduce noise. Because the variables already share a common scale, no further standardization is required. The decision tree classifier will be trained to predict beer style, which is an algorithm that naturally handles unscaled data. The features used to train the model will be: Bitter, Sweet, Sour, Salty, Fruits, Hoppy, Spices, and Malty. These features are chosen because they reflect the sensory makeup of each beer style. The target variable is the style of beer, encoded by the “Style Key,” which is based on flavor profiles and other attributes. There are 112 unique styles overall, making this a large multi-class problem. To keep the model manageable, we'll limit our scope to the 9 most frequent styles. No further imbalance correction is needed.

B. *Modeling Approach:*

We will use a Decision Tree Classifier because it is easy to visualize, handles continuous variables without scaling, and provides clear rules that users can inspect. We will split the data by 70% training and 30% test. To further tune and prevent overfitting within our tree we will restrict the max depth of the tree to 5 levels to prevent overly complex trees. To assess the performance of our tree we can calculate: the overall accuracy on the held-out test set, the precision, recall, and F1-score, and a confusion matrix.

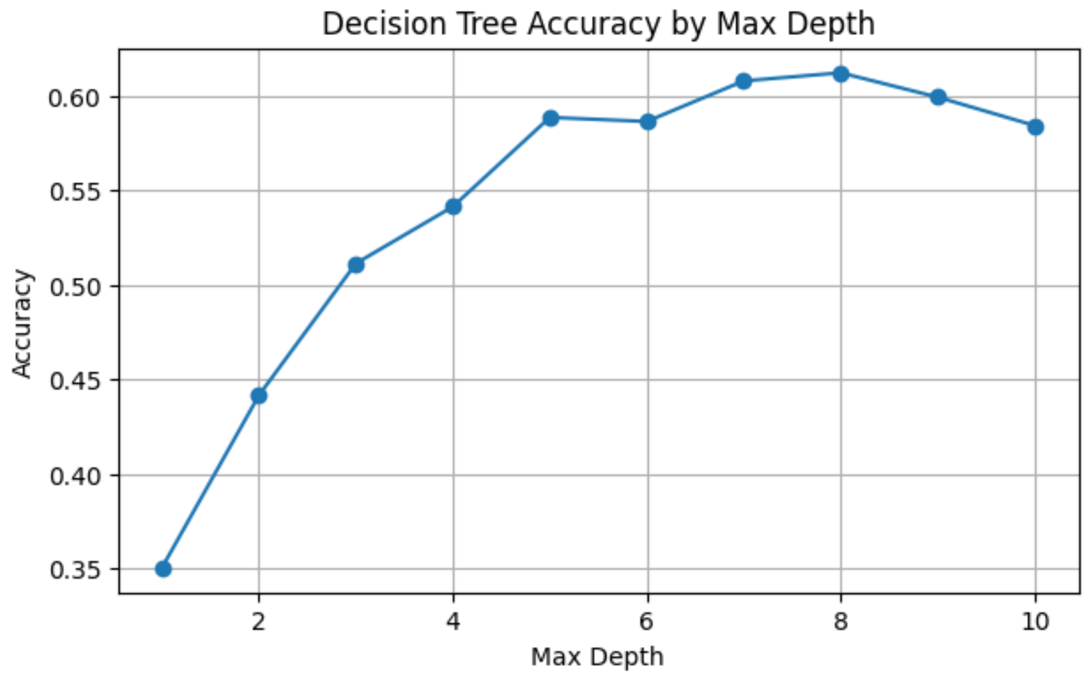
V . CHALLENGES AND TIMELINE

Date	**Phase**	**Deliverables**
April 20	Project Proposal	Defined project objective, selected preliminary dataset options, outlined methodology, and submitted formal project proposal.
April 23	Data Preparation & Initial Development	Data is imported, cleaned, preprocessed, and modeled. Report draft includes methods for handling the data and intro slides are prepared
April 25	Model Implementation & Analysis	Chosen data mining task is implemented, trained, and tested. Details of implementation and initial results are put into the report and put on slides.
April 28	Final Model Evaluation & Documentation	Code is completed, and outputs have been reviewed and verified by all team members. Results are interpreted and written in detail in the report and in slides. Code and slides are finished.
May 6	Final Review & Submission	Report has been completed to final draft and been reviewed and looked over by all team members. Code, report, and slides are turned in.

VI . MODEL TRAINING INSIGHTS

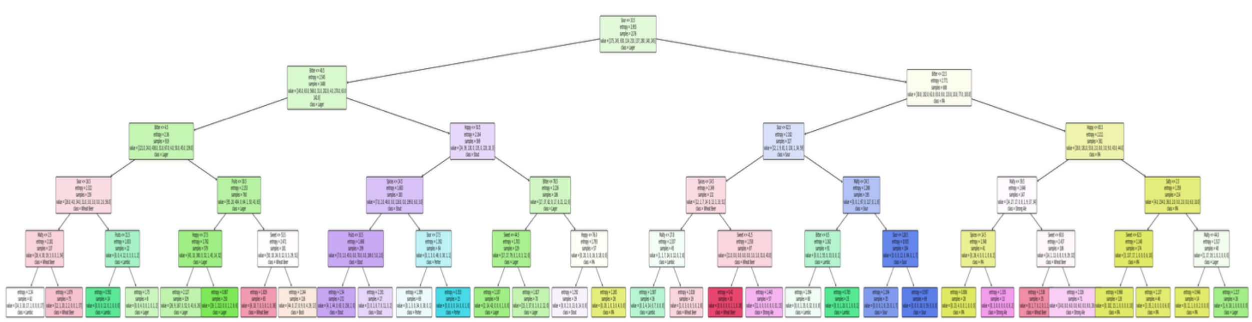
A. *Testing Decision Tree Accuracy by max depth*

After testing and plotting accuracy across multiple depths, we observed that the depth with the highest accuracy is 8. However, this depth appears to overfit the model, so we chose to use a max depth of 5 for better generalization.



B. *Analyzing the Decision Tree*

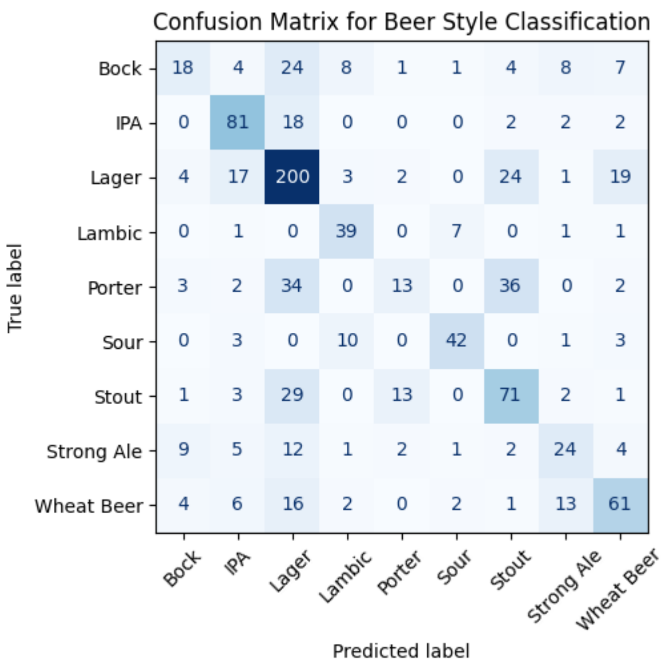
Training group: 70%
Testing group: 30%
Max depth: 5



The model was trained on 70% of the dataset and tested on the remaining 30%. In visualizing the tree, darker node colors correspond to higher prediction accuracy, providing an intuitive way to assess how well the model classifies each beer type. Notable colors include yellow, blue, purple, and green, each corresponding to a different beer category. Yellow represents IPAs, which achieved a precision of 0.66. The tree consistently identifies IPAs after the third decision node—Sour > 33.5, Bitter > 22.5, and Hoppy > 65.5—resulting in darker yellow nodes clustered on the right side of the tree. Blue corresponds to Sour beers, which have the highest precision at 0.79. All blue nodes are concentrated within a single branch, reflecting the fact that Sour beers are easily distinguished due to their high sour content. In contrast, green represents Lambics, which appear across multiple branches of the tree, indicating that this beer type cannot be easily classified by any single sensory attribute. This visualization highlights how the decision tree captures the varying degrees of predictability among beer types and provides a clear, color-coded representation of classification accuracy. The darkness and distribution of each color effectively convey the model's performance, demonstrating which beer types are well-separated by the sensory features and which are more ambiguous, reinforcing the interpretability of the model.

VII . KEY FINDINGS

A. Overall Accuracy with Confusion Matrix



In the confusion matrix, each row corresponds to the actual beer style, while each column corresponds to the predicted style. The diagonal boxes represent instances where the model correctly predicts the beer style, whereas off-diagonal boxes indicate misclassifications. Ideally, the diagonal boxes would appear darker, reflecting higher accuracy and fewer false positives. In this matrix, however, several off-diagonal boxes are noticeable, suggesting that the model occasionally predicts the wrong beer style. This pattern highlights that, while the decision tree captures some relationships between sensory features and beer types, it may not be the most reliable classifier for achieving consistently accurate predictions across all beer styles. The matrix visually reinforces the limitations of the model and emphasizes areas where misclassification is more common, providing insight into which beer styles are more difficult for the tree to distinguish based on the available sensory predictors.

B. Classification Report

Classification Report (per beer style):

	precision	recall	f1-score
Bock	0.46	0.24	0.32
IPA	0.66	0.77	0.71
Lager	0.60	0.74	0.66
Lambic	0.62	0.80	0.70
Porter	0.42	0.14	0.21
Sour	0.79	0.71	0.75
Stout	0.51	0.59	0.55
Strong Ale	0.46	0.40	0.43
Wheat Beer	0.61	0.58	0.60
accuracy	0.59	0.59	0.59
macro avg	0.57	0.55	0.55
weighted avg	0.57	0.59	0.57

The classification report provides a breakdown of model performance for each beer style using precision, recall, and F1-score metrics.

- **Precision** measures the proportion of correctly predicted instances among all instances predicted as a given style. For example, IPA has a precision of 0.66, meaning 66% of beers predicted as IPA were actually IPA. Sour has the highest precision at 0.79, indicating that predictions for this style are the most reliable, whereas Porter has the lowest precision at 0.42, meaning it is frequently misclassified as another style.
- **Recall** measures the proportion of actual instances correctly identified. Lambic has the highest recall at 0.80, showing that 80% of all Lambics in the dataset were correctly identified. In contrast, Porter and Bock have very low recall (0.14 and 0.24, respectively), indicating the model struggles to correctly identify these styles.
- **F1-score** combines precision and recall to provide a balanced metric. The highest F1-scores are for Sour (0.75), IPA (0.71), and Lambic (0.70), suggesting these styles are most reliably classified. Lower F1-scores for Porter (0.21), Bock (0.32), and Strong Ale (0.43) highlight that the model performs poorly on these styles.

Overall accuracy is 0.59, meaning the model correctly classifies 59% of all beer instances. Macro averages (unweighted) of 0.57 precision, 0.55 recall, and 0.55 F1-score suggest moderate performance across all classes without weighting for class size. Weighted averages (considering class frequency) are slightly higher, reflecting that common styles like IPA and Lambic contribute more heavily to overall performance.

This report indicates that the decision tree classifier performs best for beers with distinct sensory profiles (Sour, IPA, Lambic), while styles with less distinctive features (Porter, Bock, Strong Ale) are frequently misclassified. It highlights both the strengths and limitations of using this model with the given sensory predictors.

REFERENCES

[1] Polozoff, S. (n.d.). *Top Beer Information*. Kaggle. https://www.kaggle.com/datasets/stephenpolozoff/top-beer-information