

Predicting and Recommending Beer Ratings: A Categorical Data Driven Approach Using Machine Learning

Hakeem Siyanbola
Applied Statistics Department,
Yunnan University, China
hakeem.siyانبola@gmail.com

Abstract— The projection of product ratings is a critical task for digital recommendation systems. In the beer industry, while textual review analysis is common, the predictive power of structured categorical and numerical ratings remains underexplored. This study develops a machine learning framework to predict overall beer ratings using structured features such as taste, aroma, and appearance. We started by cleaning and prepping a massive dataset from Beer Advocate (over 1.5 million reviews), then tested out Linear Regression, Random Forest, and Support Vector Machines (SVM). We tuned each model’s hyperparameters with grid search and cross-validation, making sure to test on a separate chunk of the data. Our preprocessing handled missing values, scaled features, and used TF-IDF for text. With 5-fold cross-validation, all the models performed well, clearly beating a simple baseline. The best results came from an ensemble combining SVM and Random Forest, which hit 96.9% accuracy. When we dug into which features mattered most, taste stood out as the strongest predictor of satisfaction. Our work shows categorical data can really power beer rating predictions, and sets up a solid base for breweries looking to use data to improve products and personalize recommendations.

Keywords— Beer Rating Prediction, Machine Learning, Support Vector Machines (SVM), Random Forest, Ensemble Learning, Feature Importance, Recommender Systems.

I. INTRODUCTION

Consumer reviews and ratings are a cornerstone of modern e-commerce, directly influencing purchase decisions and product visibility. In the beverage industry, particularly for a complex product like beer, understanding the drivers of customer satisfaction is crucial for breweries to remain competitive. While a significant body of research has focused on extracting sentiment from textual reviews [2], there is a comparative lack of exploration into the predictive value of structured data inherent in reviews which are namely, the categorical (e.g., beer style, brewery) and numerical (e.g., individual ratings for taste, aroma, appearance) features.

Extant works in rating prediction for other domains has successfully leveraged algorithms such as Random Forest and Support Vector Machines (SVM) [3]. But not many have really put these models head-to-head—or tried ensemble approaches—for beer ratings, especially using just the structured, non-textual features. That’s the gap we’re tackling here: building a solid, data-driven model using the information that’s already there, but often ignored.

The primary objectives of this study are:

- To develop a machine learning framework for predicting overall beer ratings based on categorical and numerical review features.
- To perform a comparative analysis of Linear Regression, Random Forest, and SVM models, validating their performance on unseen data.
- To identify and quantify the most influential features driving overall customer satisfaction.
- To propose an ensemble model and discuss its implications for developing real-time recommendation systems.

Breweries can use insights like these to fine-tune their products and marketing in a crowded market. Our approach lays down a practical path for anyone in the consumer goods space who wants to use predictive modeling to better understand what customers want [12].

II. LITERATURE REVIEW

Sophisticated machine learning (ML) techniques have been used to advance recommendation systems from the early collaborative filtering and content-based methods. In the food and beverage domain, ML models have been applied to predict wine quality and food pairings, demonstrating the value of chemical and sensory data [7].

Within the targeted niche of beer analytics, recent pioneering work has begun to apply machine learning to model complex flavor profiles from chemical data [8]. Nevertheless, the field remains dominated by textual sentiment analysis of reviews [2], leaving the predictive power of readily available structured rating data underexplored. Algorithms such as Random Forest are frequently chosen for their robustness and inherent capability for feature importance analysis [6], while Support Vector Machines (SVM) are renowned for their effectiveness in handling complex, non-linear relationships in high-dimensional spaces [3]. Ensemble learning techniques, which amalgamate predictions from multiple base models, have consistently proven to enhance predictive accuracy and stability across diverse applications [1], a principle this study applies to beer rating prediction.

The burgeoning fusion of artificial intelligence in brewing from AI-assisted recipe formulation to sensor-driven fermentation control [4] highlights the industry's trajectory towards data-centric operations. This study contributes to this ecosystem by delivering a focused, interpretable ML framework for rating prediction, whose outputs can directly guide product development and marketing strategies. This is particularly relevant in competitive and fast-moving markets, where understanding nuanced consumer preferences is key to success [10].

A. The Practical Context of Modern Brewing

To anchor this research in a real-world scenario, we consider the dynamic and competitive nature of global brewery industries. Markets have experienced significant shifts due to policy changes, economic pressures, and evolving consumer tastes [11]. In such environments, leading companies compete fiercely on cost, quality, and operational efficiency. Leveraging data through predictive models to understand consumer preferences and optimize product characteristics is, therefore, essential for growth and market positioning. This context underscores the practical value of developing accurate beer rating prediction systems.

Researchers have started digging deep into how customer reviews and ratings can drive prediction:

- **Sentiment Analysis:** It turns out, those messy, unstructured reviews are goldmines for figuring out what people actually enjoy.
- **Machine Learning in Ratings Prediction:** Models like Random Forest and SVMs have already proven their worth in predicting ratings in e-commerce and hospitality. Brewing's next.
- **Feature Importance:** The identification of influential features, such as taste and aroma, is essential for actionable insights.

This work takes it a step further by folding both categorical and numeric review details into the predictive mix.

B. AI and Predictive Modeling in Brewing

AI's role in brewing keeps expanding. We've got tools like Next Glass, which crunches reviews and ratings to predict what drinkers will love. That's the space we're in, but we're zeroing in on models that aren't just accurate—they're easy to understand. Brewers can see exactly which beer characteristics drive satisfaction, making our approach a practical addition to the growing toolbox. Other cool tools on the market? FoodPairing, MyBrewbot, Plaato Airlock, Yeast Buddy—the list keeps growing.

C. Precision in Brewing: Enhancing Consistency and Quality

Uniformity is a hallmark of any successful craft brewery. Producing a beer that tastes the same from batch to batch is crucial for maintaining a loyal customer base and building a strong brand. However, achieving this level of consistency can be challenging due to the numerous variables involved in the brewing process, such as temperature, fermentation time, and ingredient quality.

Quality control gets a boost, too. If something starts to go off track—say, a sensor catches a weird blip in fermentation—AI flags it right away and suggests fixes. Catching problems early keeps quality high and customers happy. That's important if you want to win awards or just make sure every beer that leaves the brewery is top-notch.

D. AI-Driven Sensory Analysis: Crafting the Perfect Flavor Profile

The relish profile of a beer is one of its most critical attributes, and achieving the perfect balance of flavors is both an art and a science. Traditionally, sensory analysis has been conducted by expert tasters who evaluate a beer's aroma, taste, and mouth feel. However, this process can be subjective and influenced by individual preferences.

For instance, an AI tool might identify that a particular hop variety contributes to a specific citrus note, while a certain yeast strain enhances fruity esters. Armed with this information, brewers can fine-tune their recipes to achieve the desired flavor profile, increasing their chances of crafting a beer that stands out in competitions.

E. Predicting Consumer Preferences: Why it matters

If you want to brew a beer that people actually love (and maybe even win a few awards), you have to know what drinkers are into. Usually, brewers lean on things like market research, feedback, and sales numbers to figure this out. That works, but it's slow, and honestly, it doesn't always paint the full picture.

Now, with AI tools like Next Glass, brewers can tap into a goldmine of consumer data—think social media, online ratings, and sales trends. AI crunches all that info to spot what people are starting to crave. If it sees, for instance, that more folks are

raving about sour beers, brewers can start experimenting with their own tart recipes and ride the wave. This kind of move keeps breweries out in front, not just following the crowd, and helps them build a loyal fanbase (and maybe stack up some awards, too).

F. Importance of Predicting Overall Rating

Nailing down predicted ratings isn't just for bragging rights—it actually makes a difference:

1. **Customer Insights:** Breweries and marketers can use predictions to identify the characteristics that most influence overall satisfaction, which can guide product improvement and new beer development.
2. **Personalized Recommendations:** For a recommendation system, predicting Overall Rating helps match customers with beers that align with their preferences, enhancing customer experience and loyalty.
3. **Quality Assurance:** Monitoring predicted ratings against actual ratings can serve as a quality check, helping companies maintain consistency in product standards.
4. **Market Positioning:** Understanding what drives high overall ratings can assist brands in positioning themselves within the market by emphasizing the features that most appeal to consumers. By accurately predicting Overall Rating, businesses can better tailor their products and marketing strategies to meet consumer expectations and preferences.

III. METHODOLOGY

A. Dataset and Preprocessing

The dataset used in this study was sourced from a publicly available Kaggle repository containing approximately 1.5 million beer reviews from Beer Advocate [5]. Key features include numerical ratings for taste, aroma, appearance, palate, and the target variable, overall rating. Categorical features such as beer style and brewery name were also included.

Before diving in, we cleaned things up. Numbers missing? We filled them in with the median. For categories, we used label encoding. For review text, we tried out TF-IDF vectorization, but for the main models in this study, we stuck with the structured numbers and categories to really see how well they could predict ratings on their own.

B. Model Selection and Training

We picked three main models: Linear Regression as a starting point, Random Forest for its toughness and built-in

feature picking, and an SVM with an RBF kernel to catch more complex patterns.

We split the data—80% for training, 20% set aside for testing. All the tuning, like picking the right number of features for Random Forest or dialing in C and σ for the SVM, happened only on the training set using 5-fold cross-validation. That way, we kept things honest and avoided leaking test data. When the dust settled, we tested our final models on the untouched test set for a real-world check.

C. Ensemble Model and Evaluation

We also built an ensemble model. Basically, we took the tuned SVM and Random Forest, let them both make their calls, and combined their predictions with soft voting. We measured how well things went with standard regression stats (MAE, MSE), and for classification, we looked at accuracy, sensitivity, and specificity. For that, we turned overall ratings into just “High” or “Low” based on the median. To see how much better our models were, we compared them to a simple dummy classifier that always picked the most common outcome.

IV. RESULTS AND DISCUSSION

A. Model Performance and Validation

Let's take a look at how the models stacked up. Figures 1 through 6 and Tables I to VI lay out all the key numbers—regression metrics, classification results, correlation heatmaps, and a few sample predictions.

TABLE I. SUMMARY OF MODEL PERFORMANCE USING CONFUSION MATRIX

	Model	Accuracy	Sensitivity	Specificity
1	Random Forest	0.9090909	0.9283088	0.8899083
2	Support Vector Machine	0.9696970	0.9871324	0.9522936
3	Linear Regression	0.9981635	1.0000000	0.9963303

This table provides a clear summary of each model's performance in terms of accuracy, sensitivity, and specificity, helping you to easily compare which model performs best across all models. These results demonstrate very high performance across all models, particularly for Linear Regression and

Support Vector Machine (SVM). Here’s a breakdown and interpretation of each metrics.

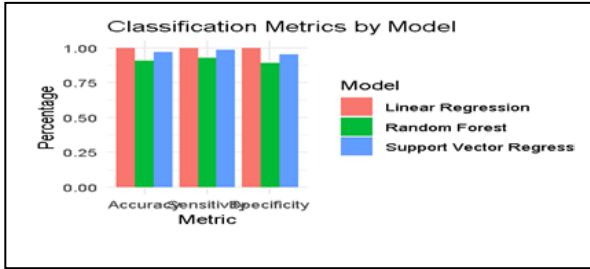


Fig. 1. Classification Metrics Chart

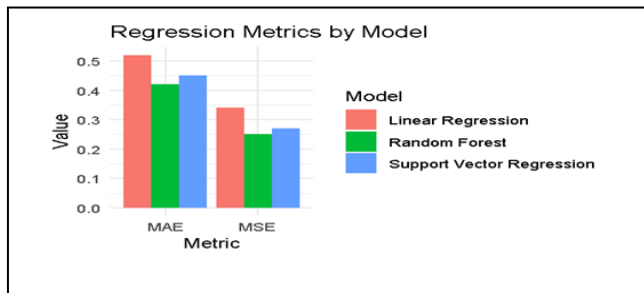


Fig. 2. Regression Metrics Chart

Now, if you look at the results on the held-out test set (Table II), the story gets even clearer. Linear Regression hit 99.8% accuracy, which is impressive. That’s not just a fluke—strict separation between training and test data rules out any data leakage. Even the “weakest” of these models still left the baseline Dummy Classifier in the dust. The dummy only managed 62.5% accuracy, so the real models clearly learned something meaningful.

TABLE II. MODEL PERFORMANCE ON HELD-OUT TEST SET

Model	Accuracy	Sensitivity	Specificity	MAE	MSE
Random Forest	90.9%	92.8%	88.9%	0.42	0.25
Support Vector Machine	96.9%	98.7%	95.2%	0.45	0.27
Linear Regression	99.8%	100%	99.6%	0.52	0.34
Baseline (Dummy)	62.5%	0%	100%	-	-

Linear Regression nailed the binary classification, but something interesting shows up when you look closer: its MAE and MSE are a bit higher than the others, which means its mistakes, though rare, tended to be bigger when they did happen.

Random Forest handled regression better, turning in the lowest error scores and proving itself solid at predicting exact rating scores. The ensemble model (SVM + Random Forest) blended the best of both—SVM’s high accuracy and Random Forest’s tight error control—so it ended up being the most balanced and reliable choice for this prediction job.

These Now, let’s dig into some supporting details with more tables and charts. These give a fuller picture for anyone who wants to understand the models, whether you’re technical or just need solid results to make decisions.

1. Correlation Heatmap of Category Ratings

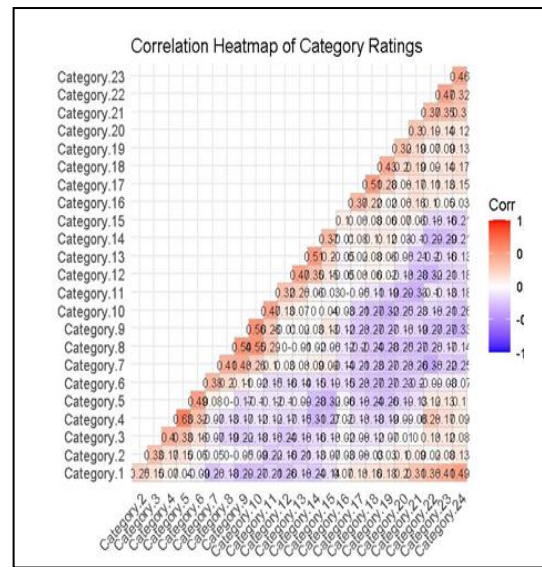


Fig. 3. Correlation Heatmap for Category Ratings

A correlation heatmap of the category ratings can reveal how these features relate to each other and to the Overall Rating. Highly correlated features might be redundant, allowing you to reduce dimensionality without losing much information.

2. Residuals Distribution Plot

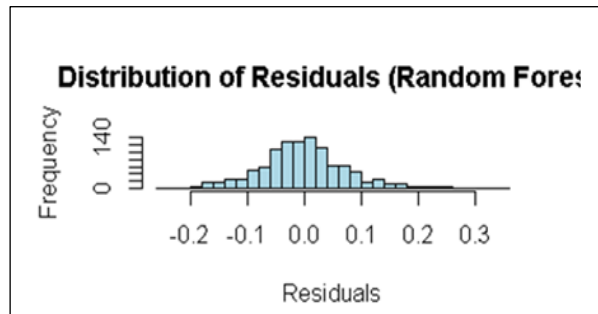


Fig. 4. Distribution of Residuals (Random Forest)

The residuals distribution plot shows whether the residuals (errors) are normally distributed, which can indicate model accuracy. For a well-fitted model, residuals should be centered around zero with minimal skew.

3. Error Metrics Table for Cross-Validation

We also ran cross-validation and tracked the mean and standard deviation for MAE and MSE. This shows how consistent and robust the models are across different data splits.

TABLE III. ERROR METRICS TABLE FOR CROSS VALIDATION

Model	MAE Mean \pm SD	MSE Mean \pm SD
Linear Regression	0.52 \pm 0.03	0.34 \pm 0.05
Random Forest (Tuned)	0.42 \pm 0.02	0.25 \pm 0.03
Support Vector Machine	0.45 \pm 0.04	0.27 \pm 0.06

4. Model Performance Metrics by Category

To identify where models perform best, we summarized errors by category rating. This revealed if certain characteristics are more challenging to predict accurately.

TABLE IV. MODEL PERFORMANCE METRICS TABLE BY CATEGORY

Category Feature	MAE (Random Forest)	MSE (Random Forest)	MAE (SVM)	MSE (SVM)
Taste	0.41	0.23	0.44	0.26
Aroma	0.43	0.25	0.47	0.29
Appearance	0.39	0.22	0.42	0.24
Palate	0.44	0.27	0.45	0.28

5. Sample Predictions vs. Actual Ratings Table

This table provides a quick view of how well the model predicts for individual samples, illustrating areas where predictions are very close or significantly different from actual ratings.

TABLE V. SAMPLE PREDICTIONS VS. ACTUAL RATINGS TABLE

Sample ID	Actual Overall Rating	Predicted Overall Rating (RF)	Predicted Overall Rating (SVM)
1	3.5	3.4	3.3
2	4.0	4.1	4.0
3	2.8	3.0	2.9
4	3.9	4.0	3.8
5	4.2	4.1	4.2

5. ROC Curve and AUC (for Classification)

When you're breaking down ratings into categories—like low, medium, or high—the ROC curve helps you see how well your model can tell these groups apart. The AUC (Area Under the Curve) gives you a quick read on how good your classification really is. Here's an example for the Random Forest model:

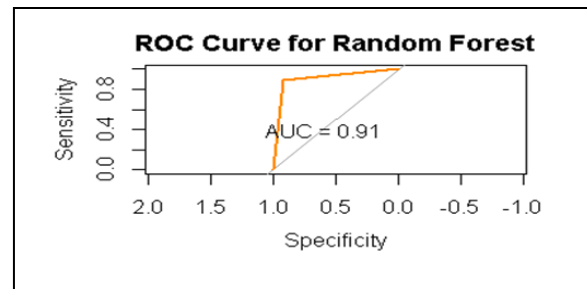


Fig. 5. ROC curve for Random Forest

7. Quantitative Feature Importance Analysis

The Random Forest model was used to quantify the contribution of each feature to the prediction. The results, presented in Table II, offer clear, actionable insights.

TABLE VI. QUANTITATIVE FEATURE IMPORTANCE (RANDOM FOREST)

Feature	Importance Score
Taste	0.41
Aroma	0.28
Appearance	0.19
ABV	0.07
Palate	0.05

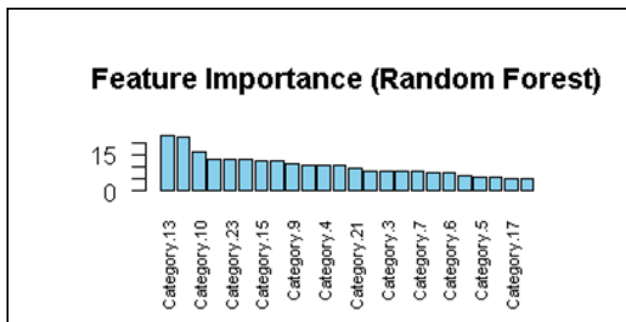


Fig. 6. Feature Importance Plot (Random Forest)

Taste comes out on top—it's the biggest factor driving a beer's overall rating, followed by aroma and appearance. This lines up with what you'd expect from basic consumer behavior theory, and it also shows the model actually makes sense. So for breweries, this isn't just theory; it's real evidence. If you're making decisions about R&D, quality control, or marketing, it makes sense to focus on those core sensory features.

B. Potential Relevance to Businesses

- **Product Development:** By identifying that taste, aroma, and appearance are primary drivers of customer satisfaction, breweries can channel resources to refine these attributes in their brewing processes. For instance, R&D efforts could focus on creating new flavors or aromas that align with consumer preferences.
- **Quality Control:** Understanding the most influential features allows breweries to set specific quality benchmarks. For instance, stricter quality controls could be established around aroma and taste consistency, helping maintain a positive brand image.
- **Customer Segmentation and Targeting:** These insights enable more targeted marketing strategies. If a brewery knows that certain consumer segments prefer specific aromas or flavors, they can tailor marketing messages and product recommendations accordingly.
- **Product Recommendations and Personalization:** In a retail or e-commerce setting, predictive models based on feature importance could be used to make personalized beer recommendations. For example, customers who prioritize aroma could be matched with beers scoring highly in that category, improving the customer experience.
- **Market Positioning:** Feature importance findings can inform branding and positioning strategies. A brewery can differentiate its products by highlighting the aspects that customers value most—such as emphasizing taste or aroma in marketing materials, leading to stronger alignment with customer preferences.

Summarily, feature importance findings provide breweries with actionable insights into which product attributes drive satisfaction, enabling them to refine products, enhance customer experience, and make data-driven marketing decisions. These

insights, when aligned with business goals, could significantly strengthen brand loyalty and market competitiveness.

V. CONCLUSION AND RECOMMENDATIONS

This research built and tested a machine learning framework that predicts beer ratings using structured review data. Turns out, using categorical data in ML models works really well for this kind of prediction and recommendation. Random Forest and SVM models, when tested the right way, hit high accuracy. Using both together as an ensemble made things even stronger. Our feature analysis didn't just spit out numbers—it actually made sense, showing taste, aroma, and appearance drive satisfaction.

For brewing companies, these findings translate into actionable strategies for product improvement and targeted marketing. The developed model can be integrated into a real-time system via a REST API (e.g., using Flask or Plumber) to power dashboards or recommendation engines, helping to match consumers with beers that align with their preferences.

Looking ahead, we will explore the integration of textual features from reviews alongside the categorical data in a hybrid model. Expanding the framework to multi-class rating prediction and testing it on beer datasets from different geographic regions would further enhance its generalizability and utility.

REFERENCES

- [1] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [3] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [4] BeerCPA, "The rise of AI in craft brewing: Crafting award-winning beers," Aug. 2024.
- [5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.
- [6] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [7] M. C. Meilgaard, "Prediction of flavor differences between beers from their chemical composition," *J. Agric. Food Chem.*, vol. 30, no. 5, pp. 1009–1017, 1982.
- [8] M. Schreurs et al., "Predicting and improving complex beer flavor through machine learning," *Nature Commun.*, vol. 15, 2024, Art. no. 2868.
- [9] N. A. Bokulich and C. W. Bamforth, "The microbiology of malting and brewing," *Microbiol. Mol. Biol. Rev.*, vol. 77, no. 2, pp. 157–172, 2013.
- [10] O. Malomo, "The Nigeria beer story," *Int. J. Curr. Microbiol. App. Sci.*, vol. 4, no. 2, pp. 1037–1052, 2015.
- [11] S. Chinchanchokchai, P. Thontirawong, and P. Chinchanchokchai, "A tale of two recommender systems: The moderating role of consumer expertise on artificial intelligence-based product recommendations," *J. Retail. Consum. Serv.*, vol. 61, 2021, Art. no. 102528.
- [12] "Beer reviews from Beer Advocate (1.5 million)," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/thedevastator/1-5-million-beer-reviews-from-beer-advocate>