BeautyBot:
Product Ratings Prediction

by: Kasia J. Windett

Non-Technical Overview

Problem Statement

More times than often, brands in the beauty industry have a hard time predicting whether or not their consumers will like a product they're selling. Brands, retailers, product development and marketing teams often have trouble with predicting product success and customer satisfaction, identifying optimal pricing and stock strategies, and understanding customer preferences.

Proposed Solution

A solution that is ideal for this type of problem is building a predictive model using data science. To be more specific, there will be multiple models built and developed to find the best one, such as linear regression. This ideal model will predict customer ratings, that way businesses can analyze their goals and customers more effectively.

Impact

This solution will add value to businesses by having product improvement, inventory optimization, higher customer loyalty, and reduced return rates. Beauty companies will gain insight from the predictive feedback to tailor their formulas, packaging, or marketing strategy. Beauty retailers, such as Sephora, can predict which products will be rated high to stock more inventory or reduce their overstock on low-demand products. Customers will have fewer returns to make because the products will meet their expectations, if not, surpass them.

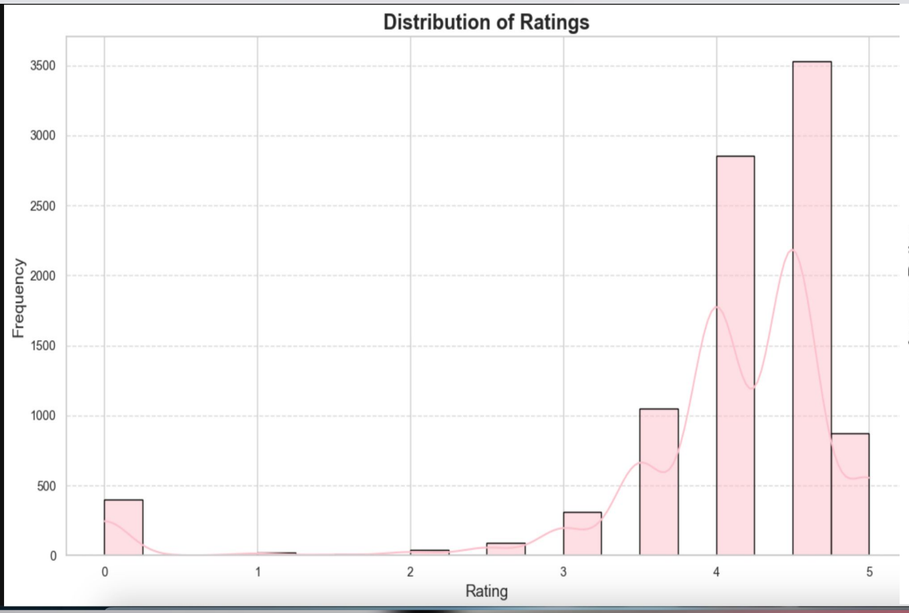
Data & Preprocessing

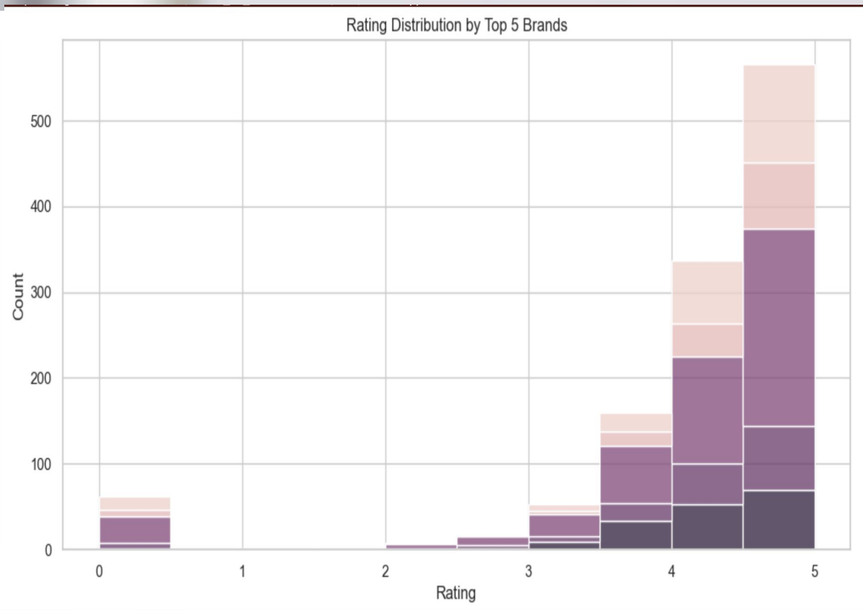
The dataset used in this project is from Kaggle, called Sephora Website. It has data that was scraped from Sephora's website and includes over 9,000+ beauty products. To preprocess the data, all columns were analyzed to find their description, data type, and to see if they had any missing values. Irrelevant columns, like URL, were deleted, duplicates were also deleted, and new columns were added using other columns.

Variable	Description	Type
id	The product ID on Sephora's website	Int
brand	Brand of the product	Object
category	Category of the product	Object
name	The name of the product	Object
size	The size of the product	Object
rating	The rating of the product	Float
number_of_reviews	The number of reviews of the product	Int
love	The number of people loving the product	Int
price	The price of the product	Float
value _price	The value of discounted products	Float

URL	The URL link of the product	Object
MarketingFlags	Whether the product was exclusive/online-only or not	Bool
options	The options available on the website for the product like colors and sizes	Object
details	Details of the product	Object
how_to_use	The instructions of the product	Object
ingredients	The ingredients of the product if available	Object
online_only	If the product is sold online	Int
exclusive	If the product is sold exclusively on the website	Int
limited_edition	If the product is limited edition	Int
limited_time_offer	If the product has a limited time offer	Int

Key Insights





	MSE	R2
Linear Regression	0.578931	0.022863
Decision Tree	0.506623	0.144907

Model Results and Interpretations:

Chosen Model: Random Forest Regression with hyperparameter tuning.
It has the best overall key performance indicators compared to the other models show.

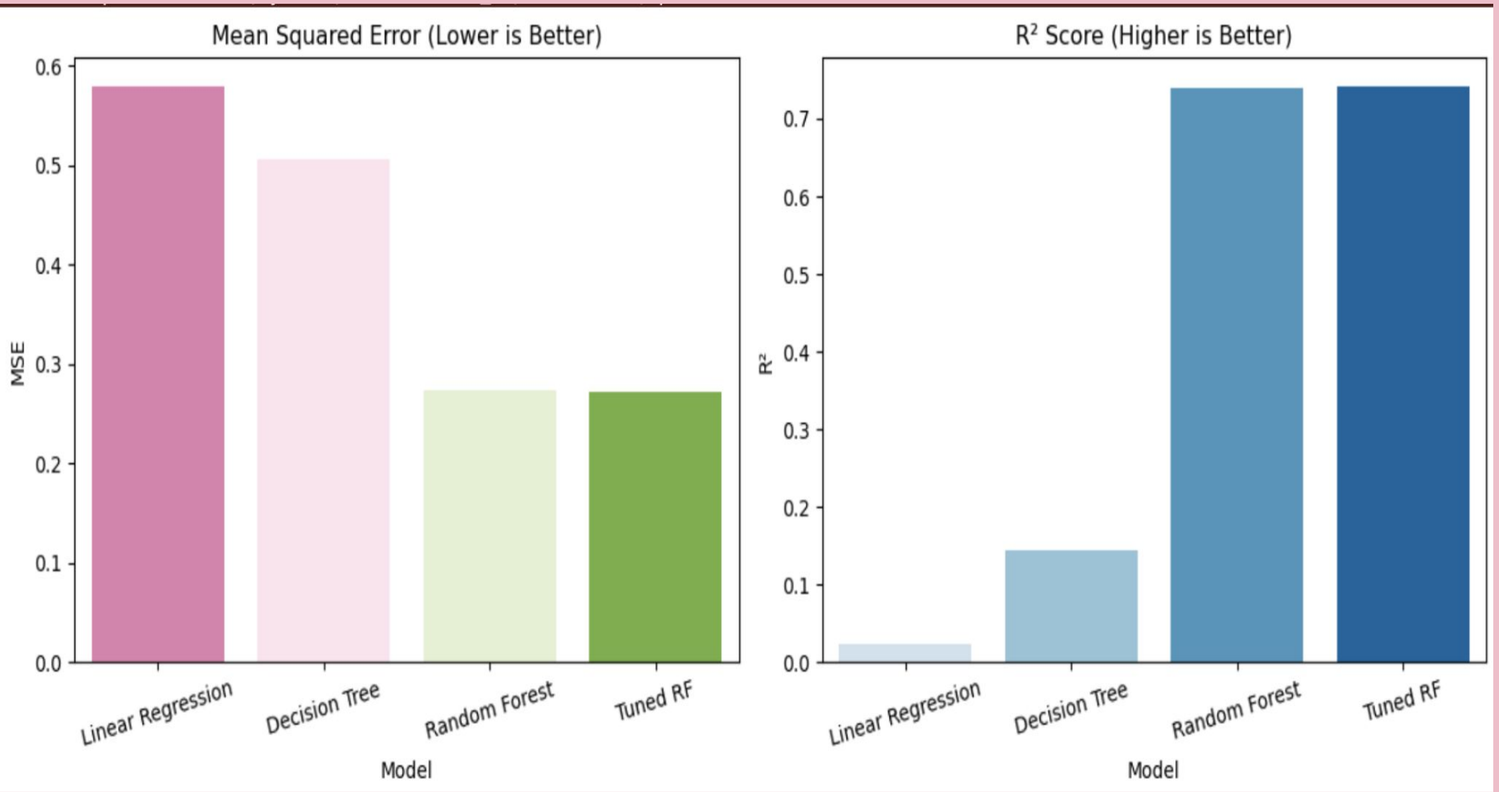
Random Forest Results:

```
Random Forest MSE: 0.2732
Random Forest R²: 0.7398
```

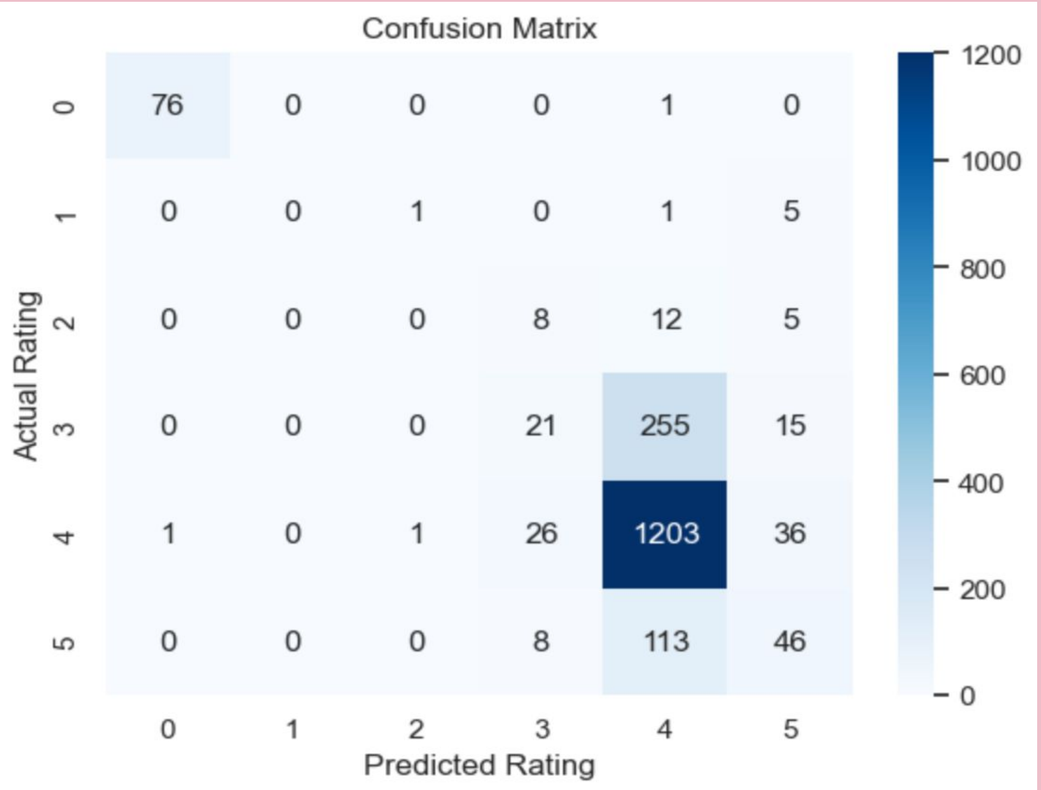
Random Forest (Tuned):

```
Best R² Score on Validation Sets: 0.7257978163200326
Test Set MSE (Best Model): 0.2716
Test Set R² (Best Model): 0.7413
```

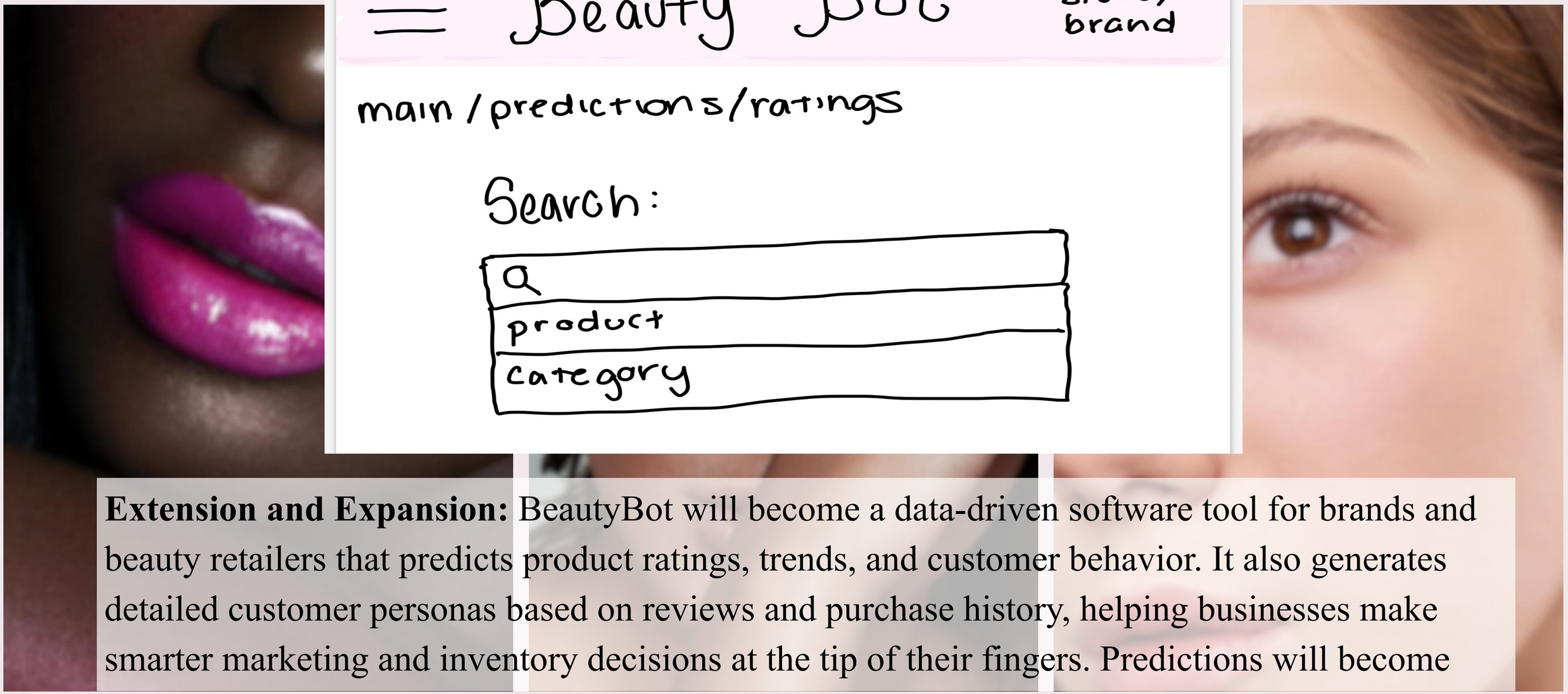
Model Comparisons:



Random Forest Classification:

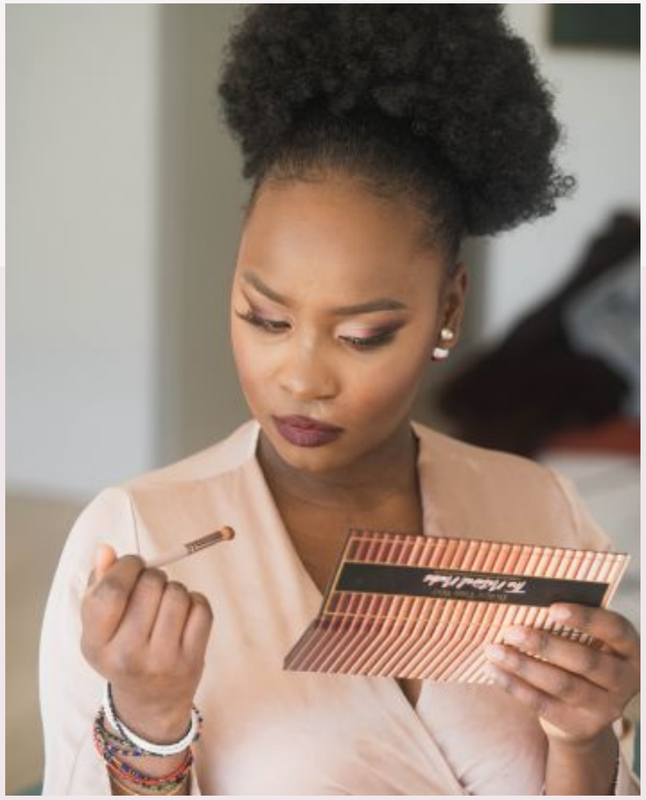


Next Steps



Extension and Expansion: BeautyBot will become a data-driven software tool for brands and beauty retailers that predicts product ratings, trends, and customer behavior. It also generates detailed customer personas based on reviews and purchase history, helping businesses make smarter marketing and inventory decisions at the tip of their fingers. Predictions will become stronger with using models such as XGBoost, yielding higher accuracy and lower bias.

thank you



Where beauty meets brains — BeautyBot makes smarter strategies look effortless.