Predicting Coffee Ratings from Expert Reviews

Kate Meredith

Agenda

- The Problem
- Background
- Data
- Cleaning & Pre-Processing
- Exploratory Data Analysis (EDA)
- Modeling & Results
- Key Takeaways
- Future Work

The Problem

How might we better understand the relationship between review and rating, so that coffee roasters can achieve top scores and distinguish themselves in a highly competitive market?

Background

- Global coffee market: $100 billion USD and growing
- Highly competitive, dominated by large companies
- Growing interest in socially and environmentally responsible coffee
- Ratings and reviews influence consumer behavior
- Smaller roasters can distinguish themselves through expert reviews (“cupping”)

The Data

- Scraped for this project from CoffeeReview.com
- Reviews span 1997 to 2022
- Reviews conducted by a small, specially trained team
- Data includes:
  - Descriptive info (e.g. year of review)
  - Evaluative numeric data (e.g. aroma score)
  - Descriptive text (e.g. where to buy it)
  - Evaluative text (e.g. blind assessment)

Data Wireframe

Features

Month	Year	Bean Agtron	Ground Agtron	Aroma	Acidity	Body
1-12	1997-2022	1-100	1-100	1-10	1-10	1-10

Flavor	Aftertaste	Roaster Location	Roaster Longitude	Origin Latitude	Origin Longitude	The Blind Assessment
1-10	1-10	Lat + Lon	Lat + Lon	Lat + Lon	Lat + Lon	Text

Target

Overall Score
50-100

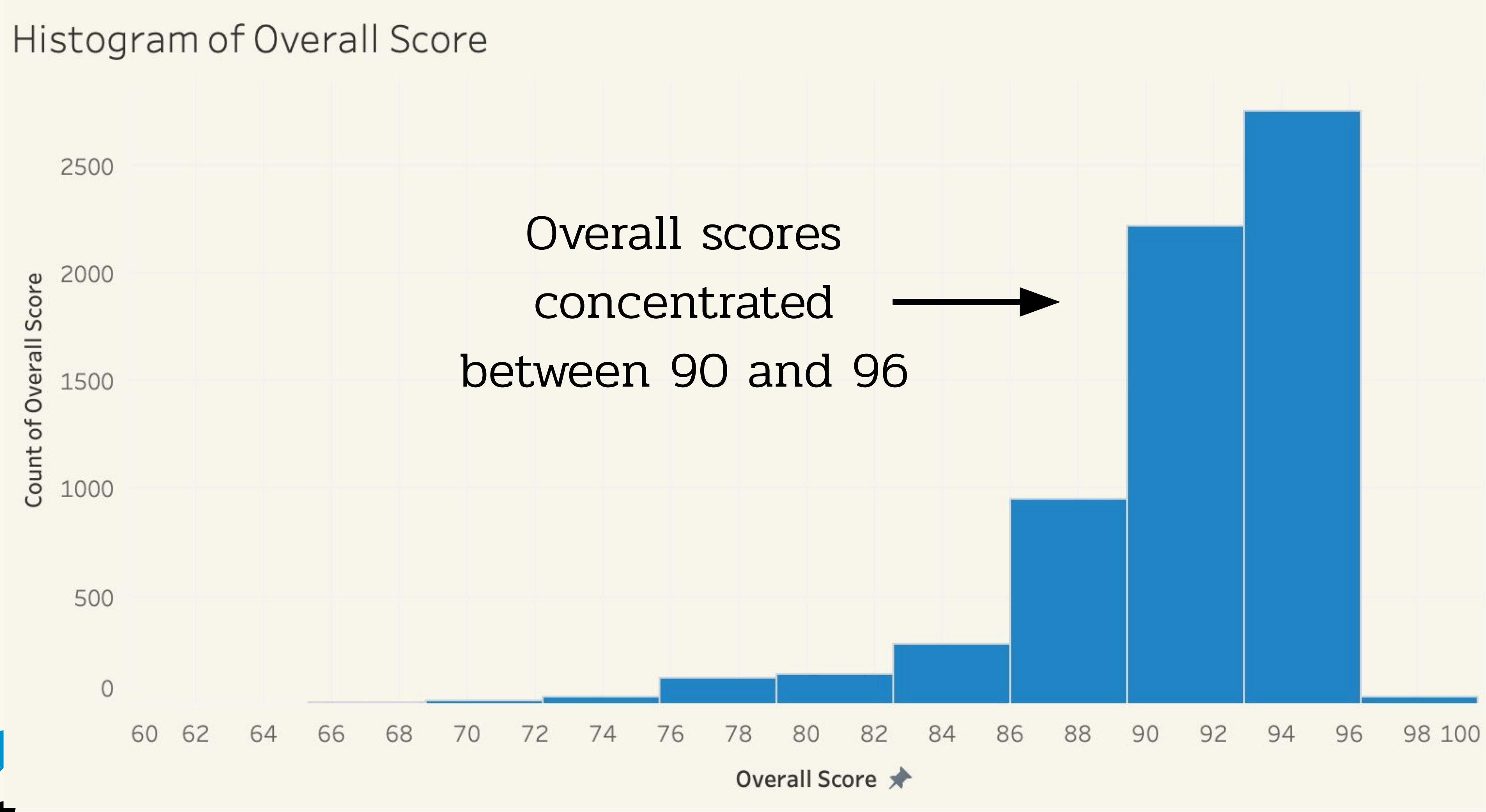
Cleaning & Preprocessing

- Sorting scraped data to match values with the correct feature
- Missing data filled in
- Values transformed to more useable format (location, roast level)
- Dropped: ‘with milk’, ‘estimated price’ and descriptive paragraphs
- Natural language processing on text

Exploratory Data Analysis (EDA)

- Key insights:
  - Subscores: aroma, acidity, flavor and aftertaste:
    - Highly correlated with target (overall score)
    - Scores tend to be high (7-9)

Exploratory Data Analysis (EDA)



SLIDESMANIA.COM

Modeling & Results

- 3 Stages:
  - Baseline models on numeric data
  - Baseline model on transformed text data
  - Optimized models on combined dataset
- Looking for R^2 close to 1 and low Mean Absolute Errors (MAE)

Modeling & Results

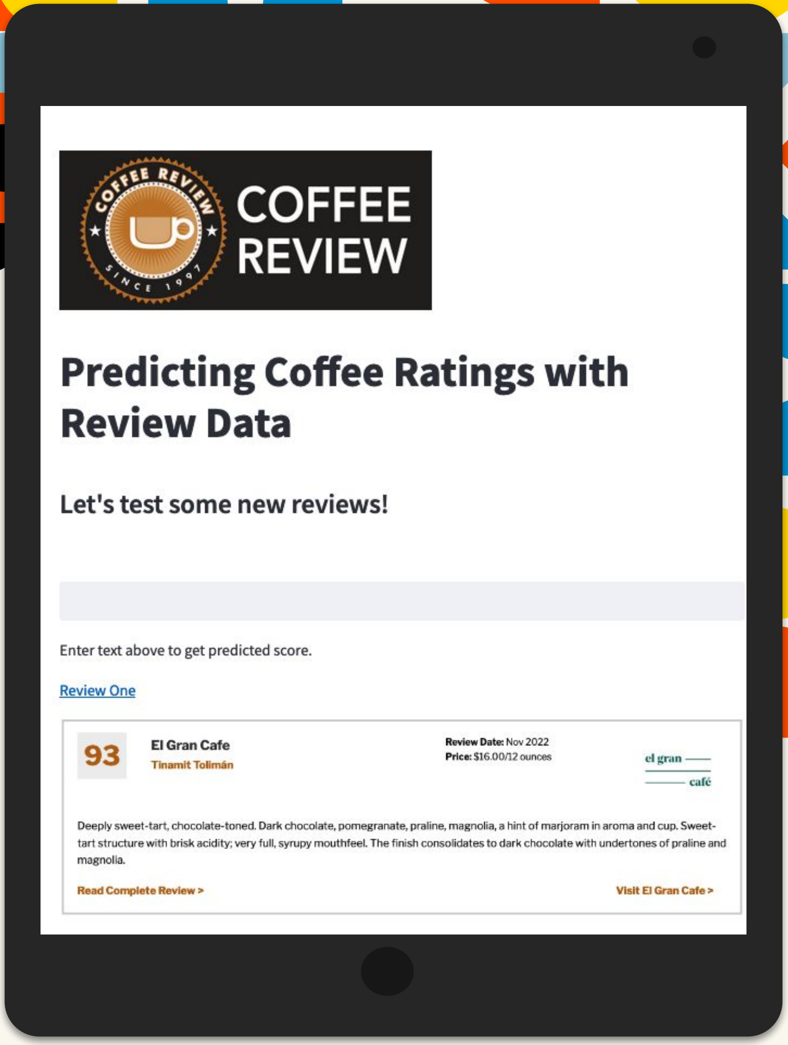
- Baseline models on non-text data (in order of performance
  - XG Boost Regressor (R^2: 0.918)
  - Random Forest (R^2: 0.916)
  - Linear Regression (R^2: 0.898)
  - Neural Network (R^2: 0.894)
  - KNN (R^2: 0.843)
  - Support Vector Regression (R^2: 0.892)

Modeling & Results

- 4 transformation methods tried:
  - Count Vectorizer
  - TFIDF Vectorizer (stemmed)
  - TFIDF Vectorizer (not stemmed)
  - Word Embedding
- Selected method: TFIDF Vectorizer (not stemmed): R^2: 0.744

Text Model & Results

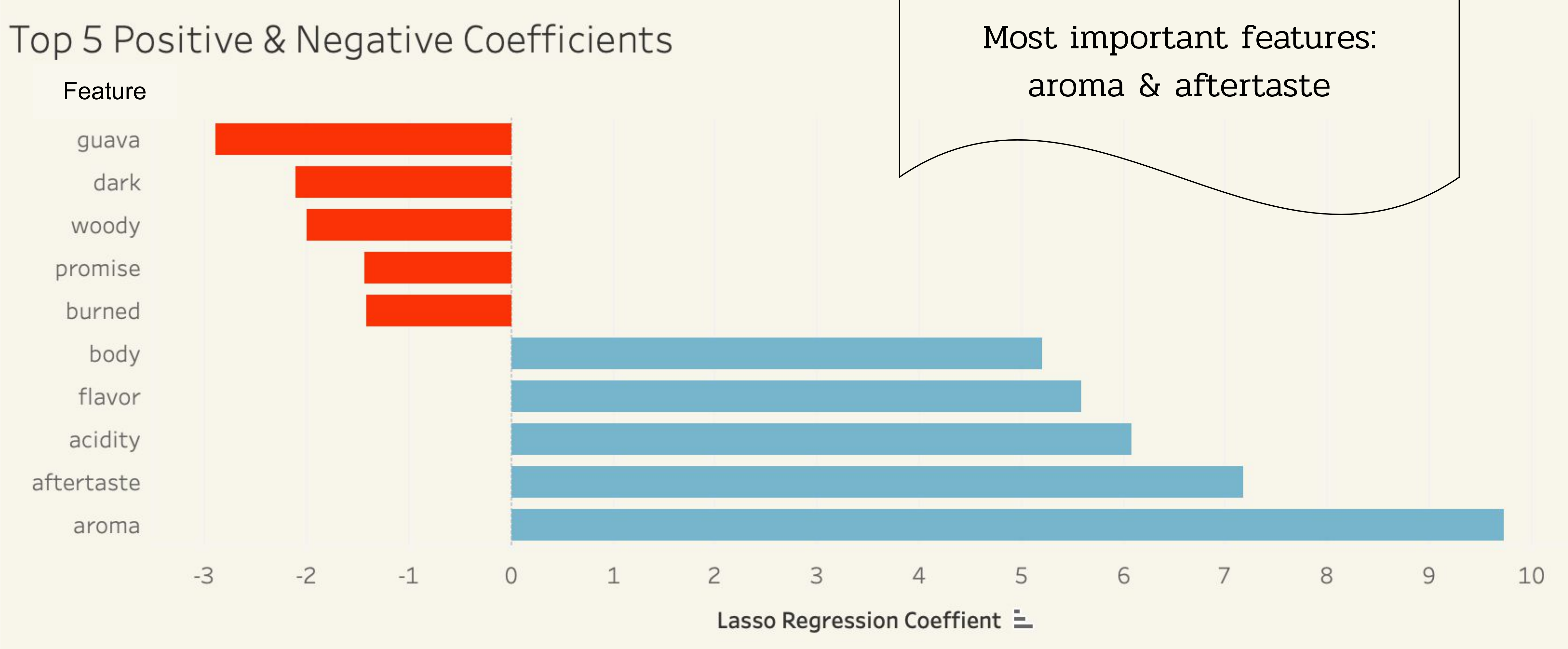
- Best text-only model R^2: 0.744
- Let's see how well the text reviews predict results.



Combined Model & Results

- Optimized models on combined dataset
  - XG Boost Regressor (Validation R^2: 0.906 & MAE: 0.593)
  - Lasso Regression (Validation: R^2: 0.904 & MAE: 0.674)
- **Final Model: Lasso Regression**
  - Test R^2: 0.910
  - Test MAE: 0.691
  - Conclusion: We can expect the model to predict scores within 1 point of accuracy

Key Takeaways



Most important features:
aroma & aftertaste

Future Steps

- Including additional text in the final model
- Trying other methods for dealing with missing values
- Optimizing additional models on the combined dataset
- Explore topic modeling

Thank You!

Let's Connect.

linkedin.com/in/kate-m-meredith/

Github

github.com/KMere21

Credits.

Presentation Template: SlidesMania

Images: Unsplash

Pattern design made with shapes based on Paaatterns

Please keep this slide or mention us and the other resources used in the footer of a slide.