Edinburgh Airbnb Star Rating Outcomes Prediction through Review Texts Analysis

airbnb

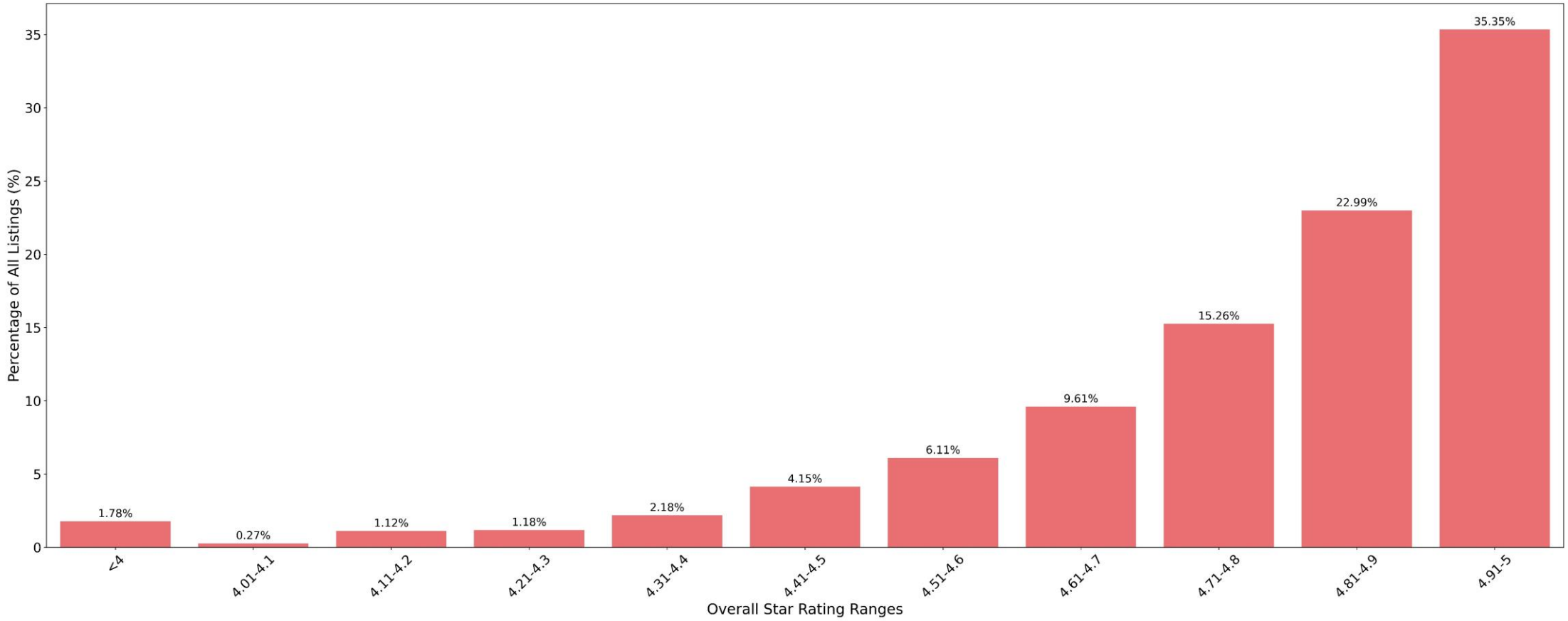
Sprint 2 Progress

Problem Statement	Predict Airbnb Listing Star Ratings and Perform Text Feature Extraction
Methodology	NLP, Matching Learning, Deep Learning
Potential Outcomes	Well Trained Model, Interpretable Results, and Actionable Insights
Current Results	First GridSearch Completed, Key Features Extracted.

airbnb

Distribution of Airbnb Listings by Star Rating

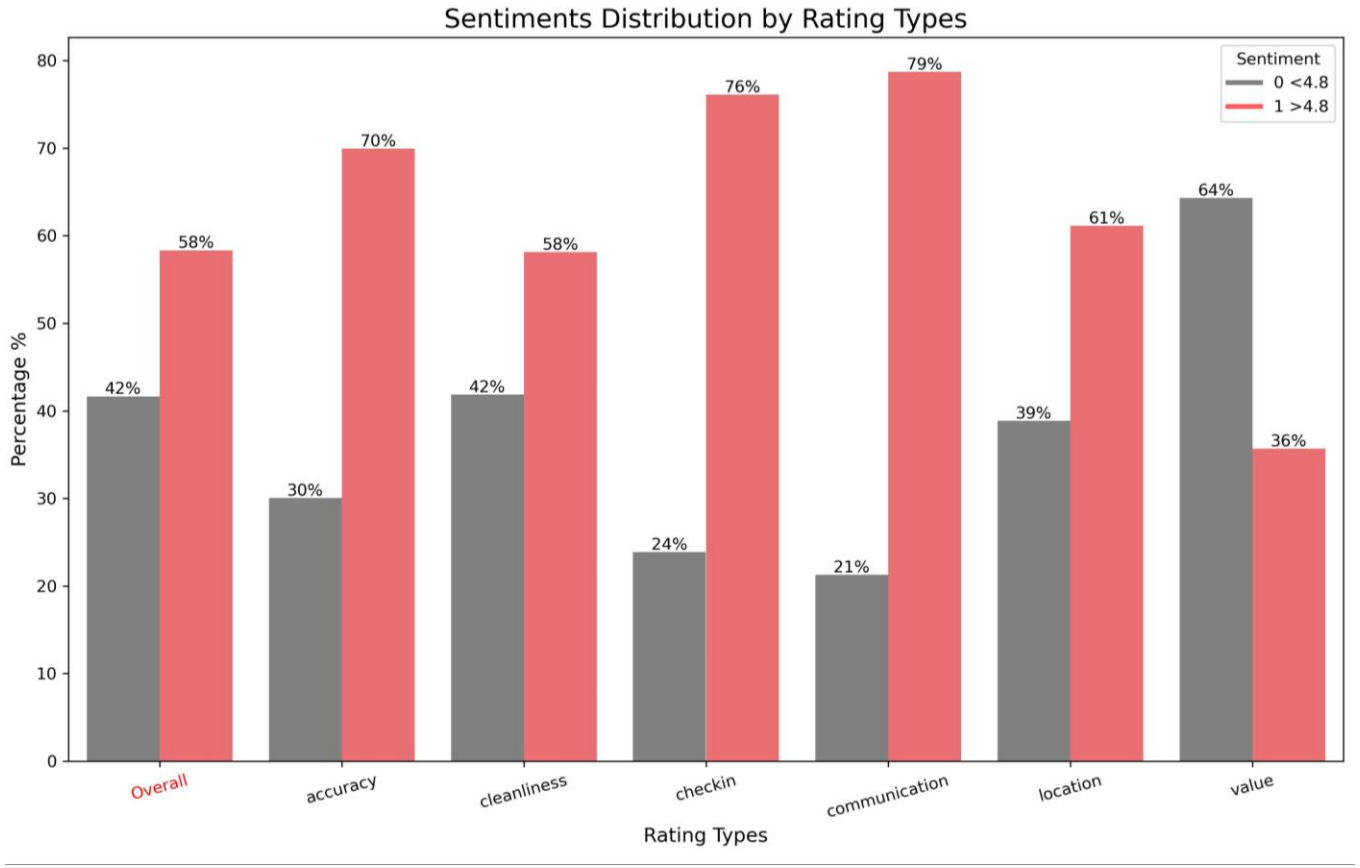
- 74% of the listings have above 4.7 overall rating.
- Only less than 2% of the listings have overall rating less than 4.



airbnb

Distribution of Transformed Sentiment Scores

- Overall Sentiment Scores are about balanced.
- Only Value Sentiment Scores Class 0 exceeds Class 1.



Aggregated Review Data

Listing Details → **Reviews**

Listing 1	Rating 1	Review A
Listing 1	Rating 1	Review B
Listing 1	Rating 1	Review C
Listing 2	Rating 2	Review D
Listing 2	Rating 2	Review E

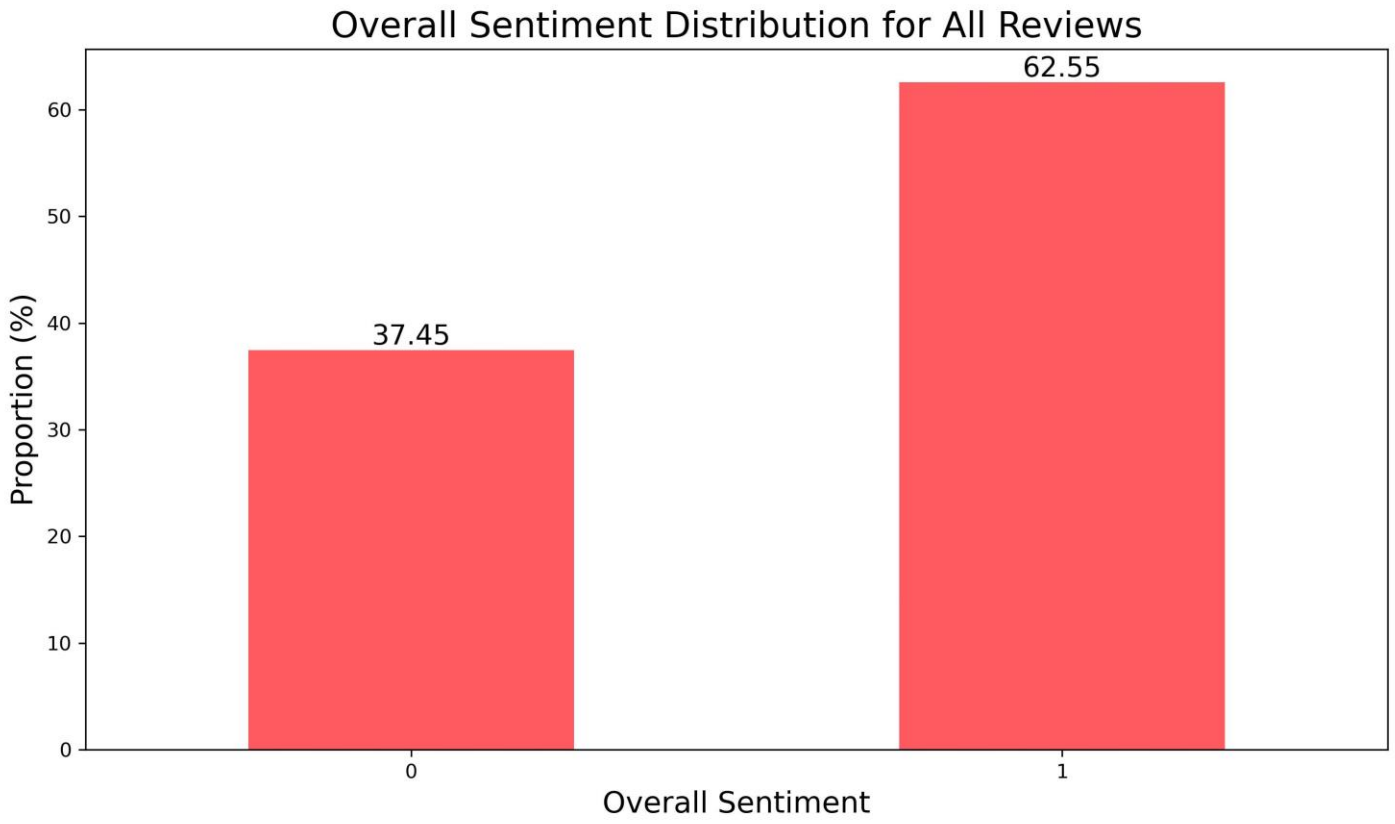
Reviews Collapsed → **Listings**

Listing 1	Rating 1	Review 1 + Review 2 + Review 3
Listing 2	Rating 2	Review 4 + Review 5

airbnb

Overall Sentiment Distribution for Reviews Not Collapsed

- Overall Sentiment Scores are imbalanced after aggregation.
- Down sampled training data with care to overcome this issue.

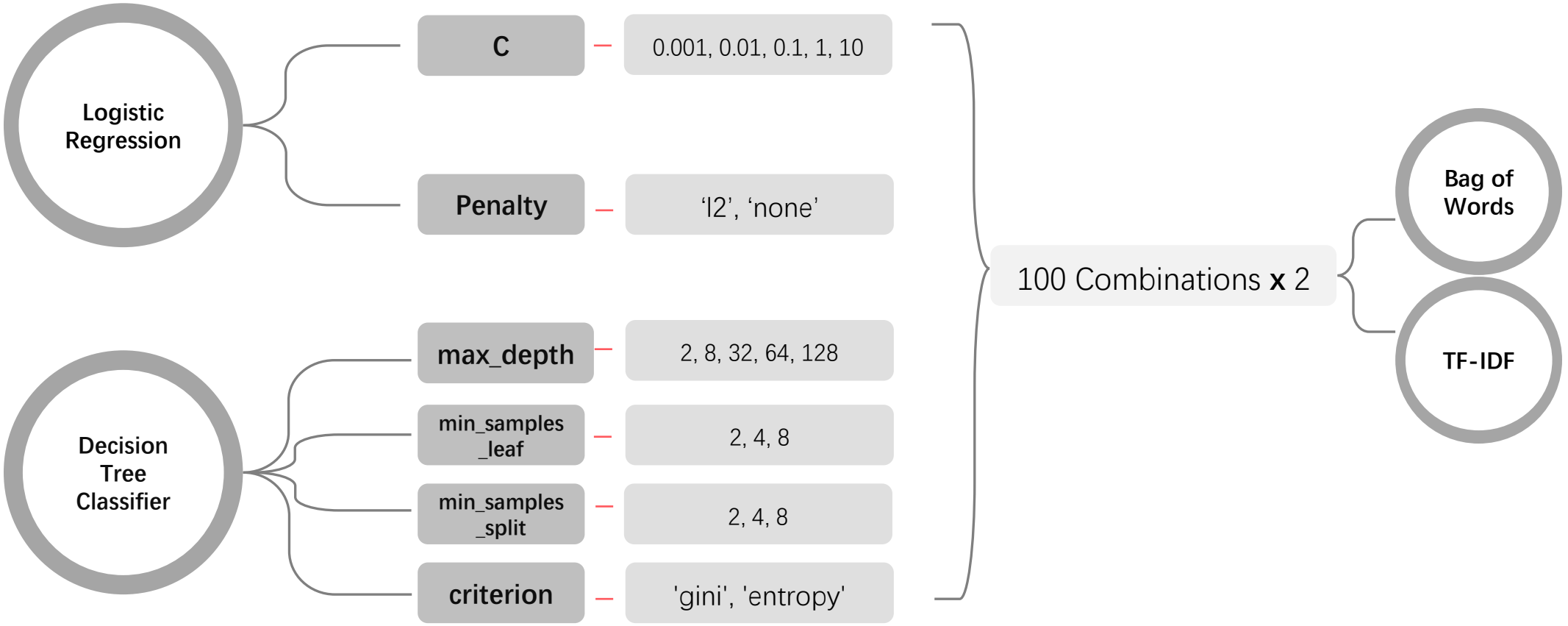


Most Common Words in All Reviews

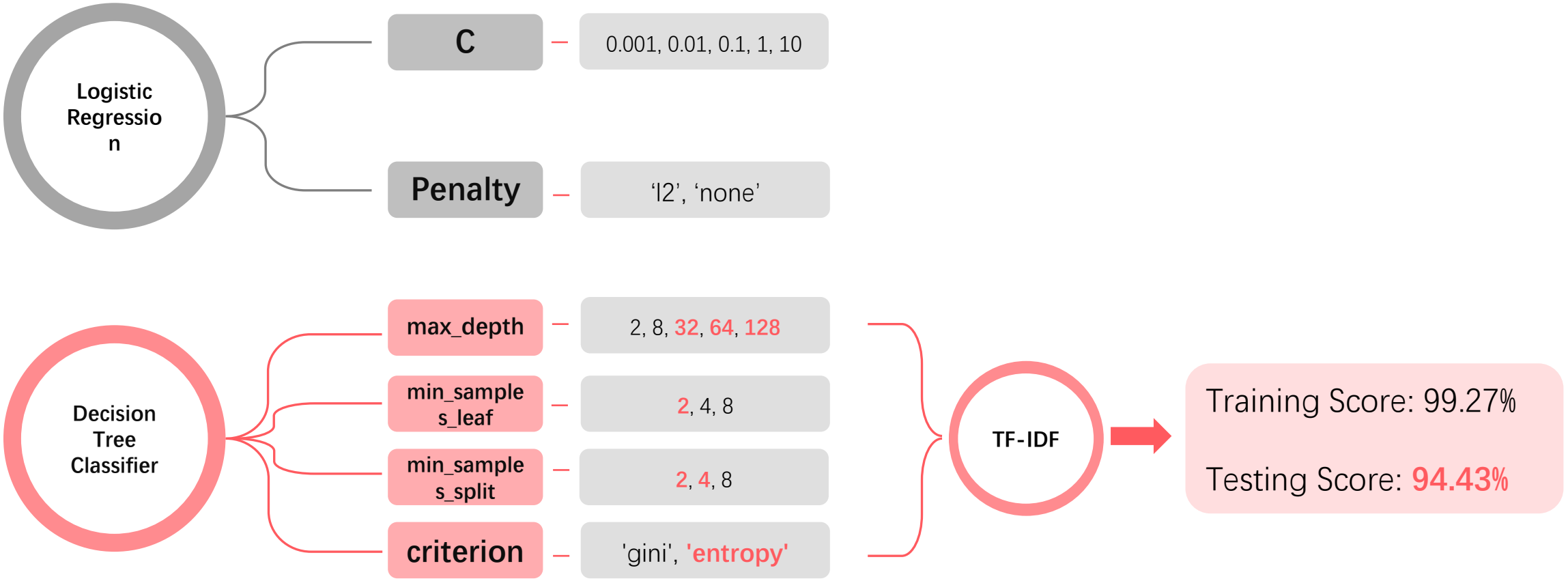
- Most common words just by counting the frequency are ‘Edinburgh’, ‘flat’, ‘apartment’.

airbnb

First GridSearchCV Models and Hyperparameters

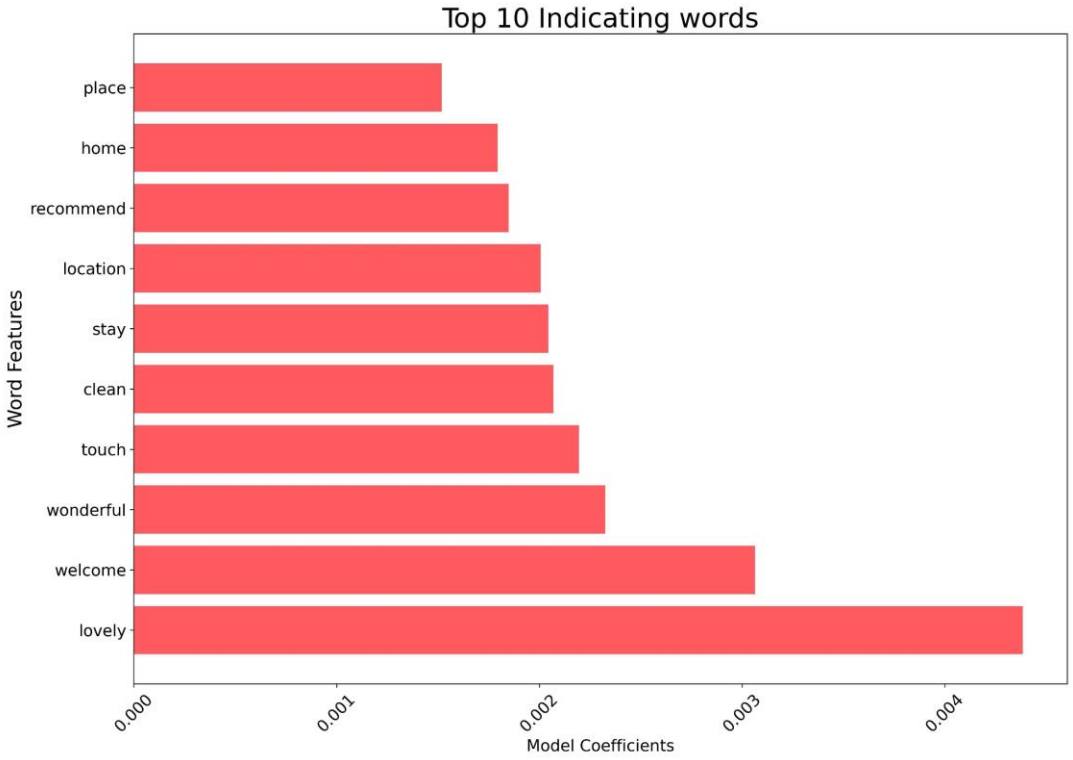


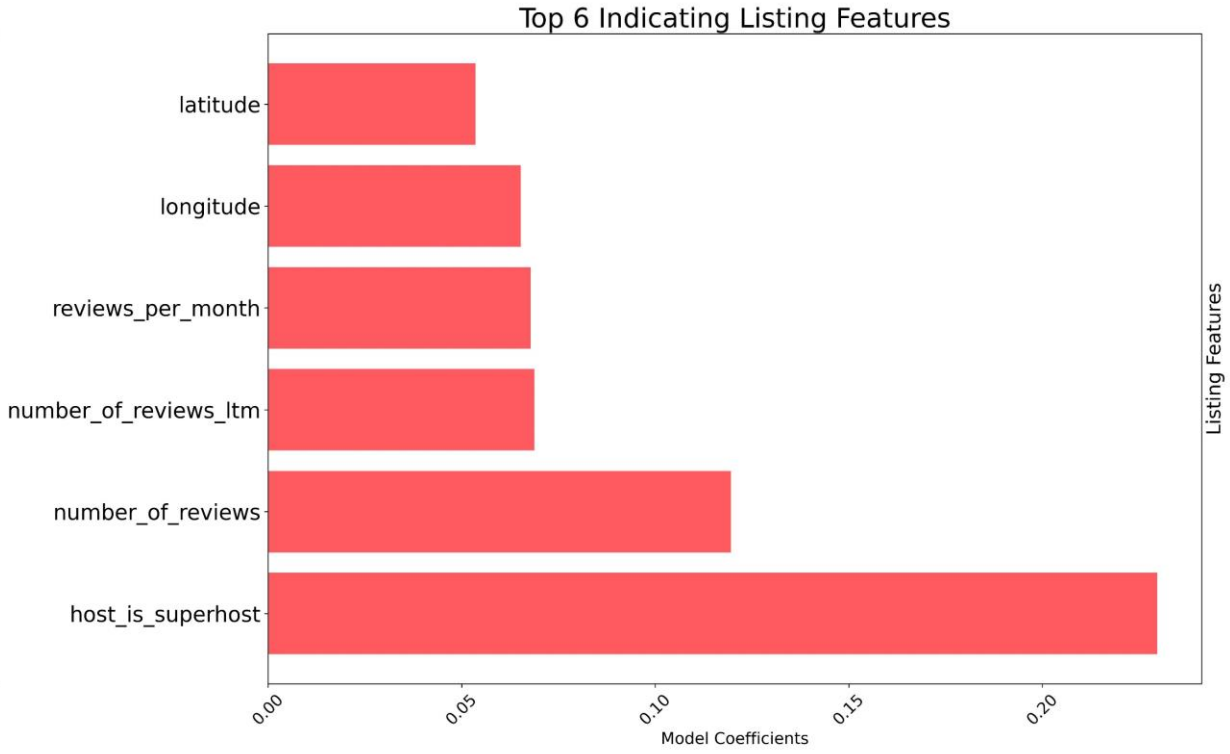
First GridSearchCV Models and Hyperparameters



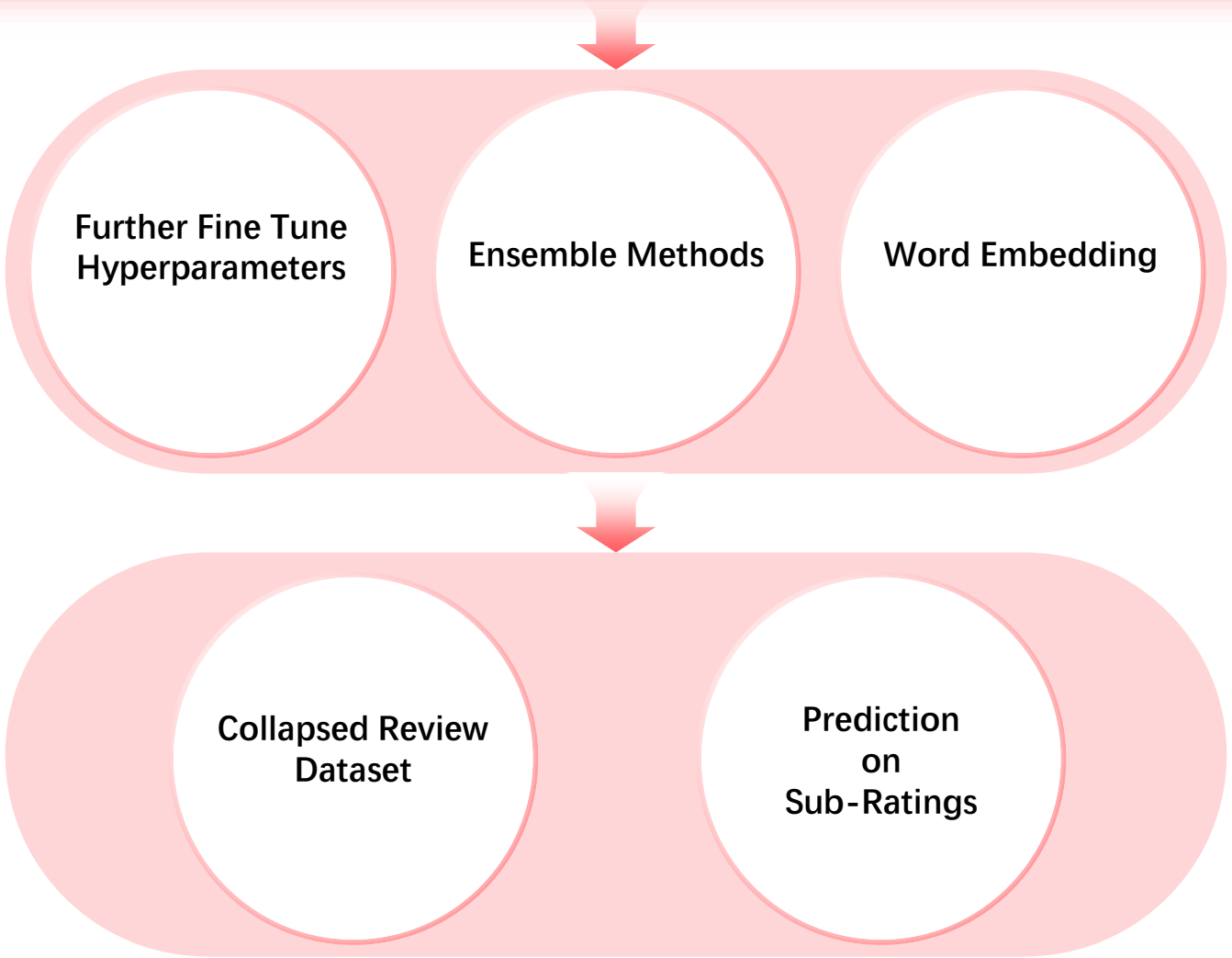
Best Model Feature Extraction

- Words with highest model coefficients are ‘lovely’, ‘welcome’, etc.
- Listing Feature with highest model coefficient is ‘host_is_superhost’.





Next Steps



airbnb