Edinburgh Airbnb Star Rating Outcomes Prediction Through Review Texts Analysis

airbnb

Problem Statement

How can Airbnb hosts improve their rentals?

What are the key factors that determine an outstanding Airbnb listing?

How can I answer this question through the analysis of guest reviews?

airbnb

Key Issue About Datasets

- No ratings scores for individual reviews

Listing Dataset

Listing ID	Listing Features	Rating Scores
1	Features_1	4.3
2	Features_2	4.6

Reviews Dataset

Listing ID	Reviews
1	A
1	B
1	C
2	D

Aggregate Listing and Review Data

- Same listing details on different reviews if they comment on the same Airbnb
- Half a million datapoints
- Assumed all reviews of same listing share same score

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

airbnb

Aggregate Listing and Review Data

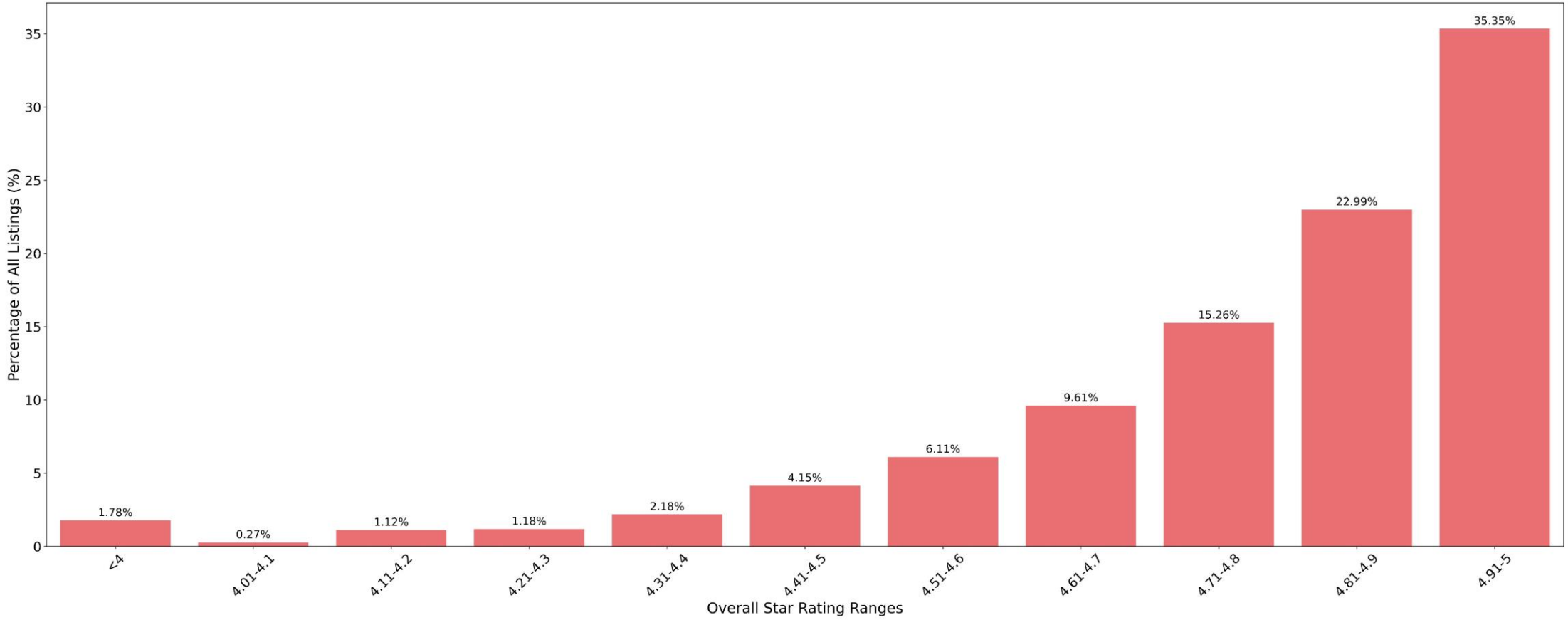
- Reviews condensed to same listings
- Only about 6000 datapoints
- Better interpretability as overall rating score corresponds to overall reviews

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

airbnb

Listings Distributions by Overall Star Ratings

- 58% of the listings have above 4.8 overall rating.
- Only less than 2% of the listings have overall rating less than 4.
- Therefore, a threshold of 4.8 is selected to classify listings



airbnb

Modelling Workflow

Uncondensed Review Dataset

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

GridSearch (General Sweep)

GridSearch (N-grams)

Condensed Review Dataset

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

GridSearch (General Sweep)

GridSearch (N-grams)

airbnb

Modelling Approach

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

GridSearch (General Sweep)

Purpose: Baseline Analysis

Limitation: Difficult to interpret

GridSearch (N-grams)

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

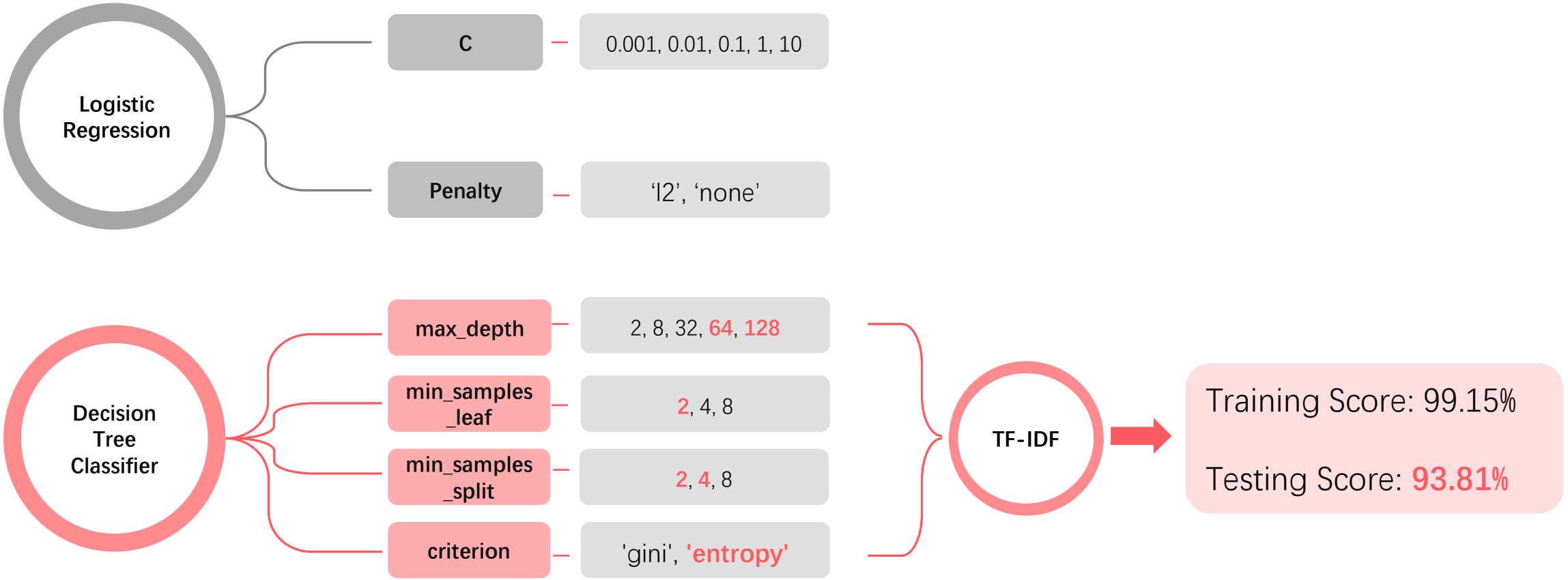
GridSearch (General Sweep)

GridSearch (N-grams)

airbnb

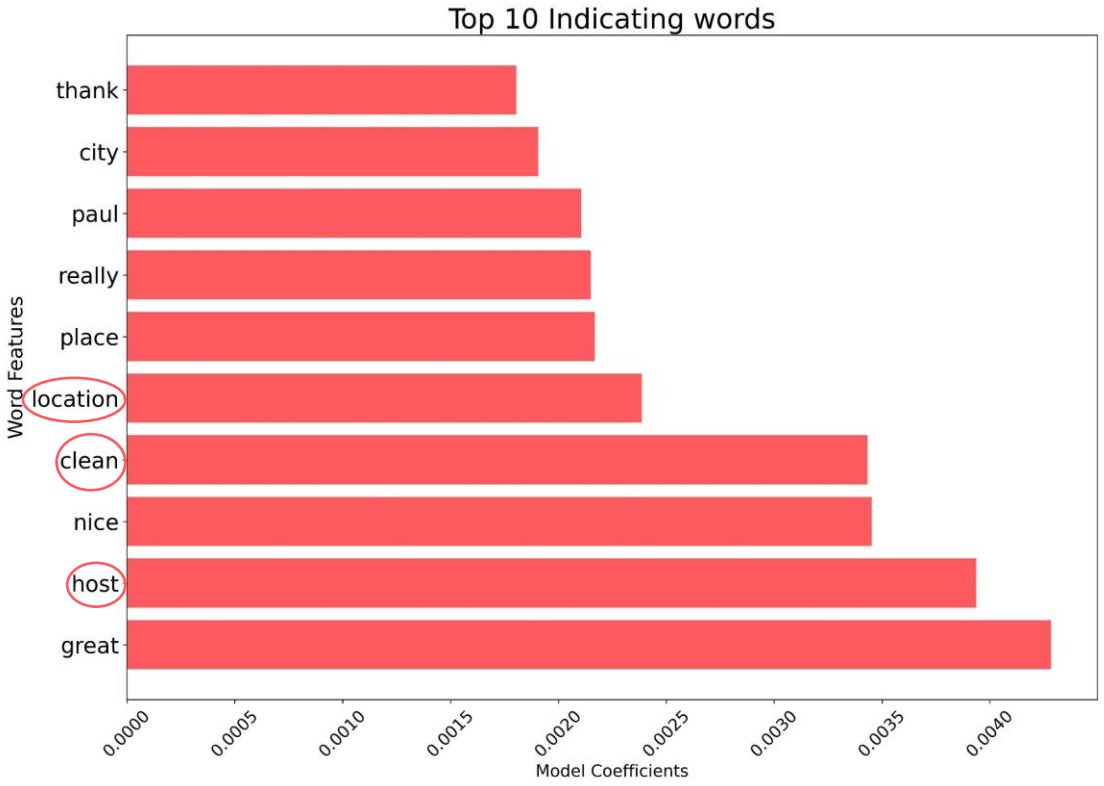
GridSearch 1: General Sweep

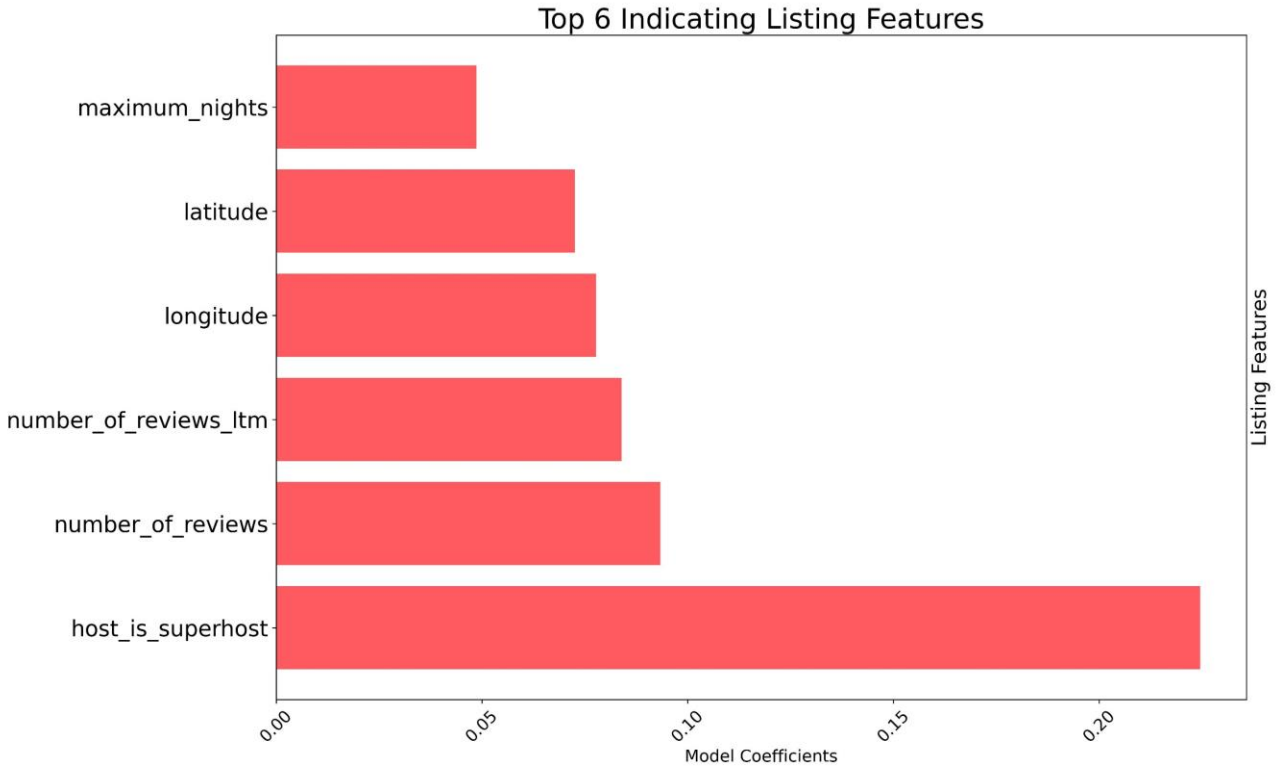
- Dataset: Same listing details on different reviews



GridSearch 1: General Sweep Extracted Features

- Dataset: Same listing details on different reviews





Next Approach

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

GridSearch (General Sweep)

- Purpose: Baseline GridSearch
- Limitation: Difficult to interpret

GridSearch (N-grams)

- Purpose: Better Interpretability
- Limitation: Under aggregation assumptions

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

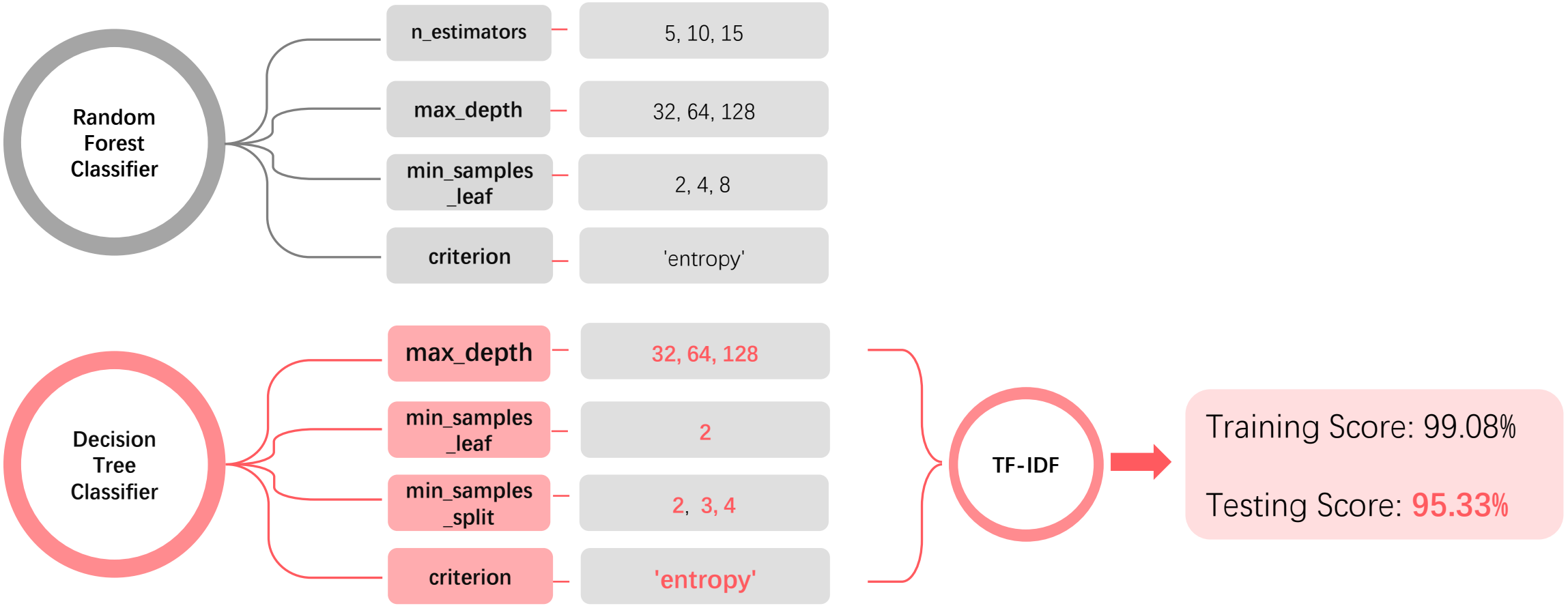
GridSearch (General Sweep)

GridSearch (N-grams)

airbnb

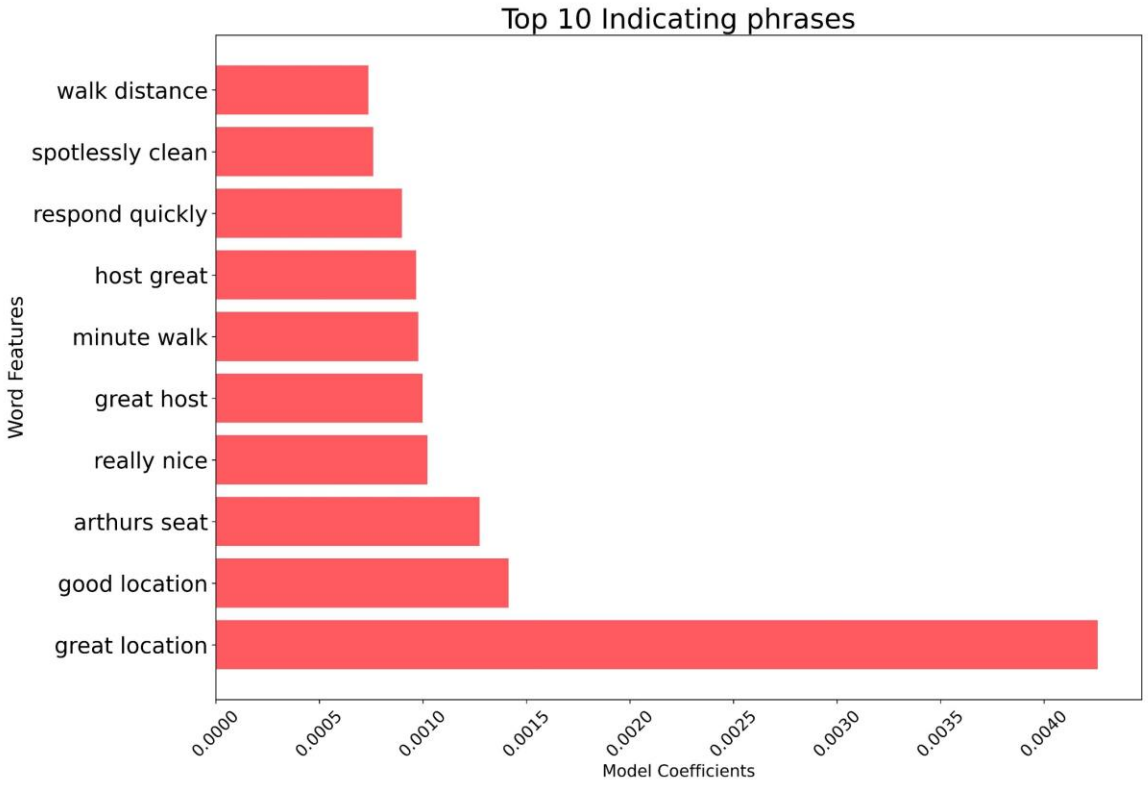
GridSearch 2: N-grams

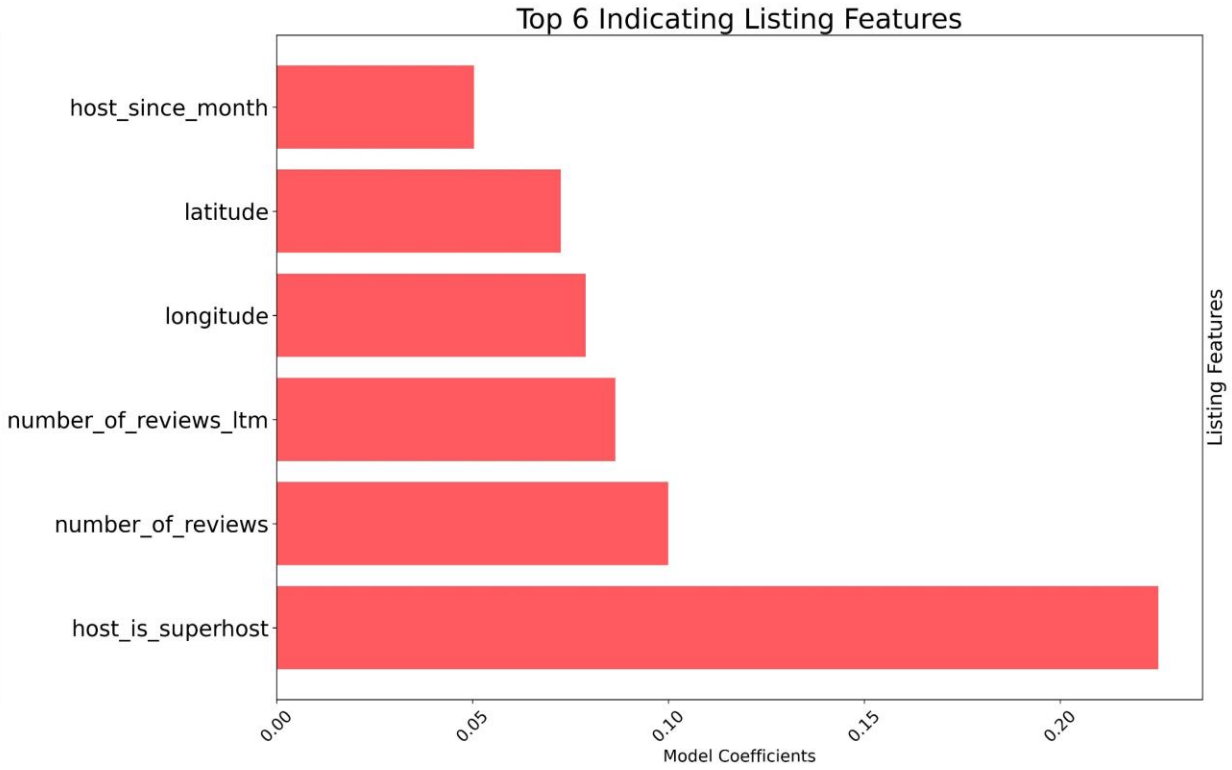
- Dataset: Same listing details on different reviews



GridSearch 2: N-grams Extracted Features

- Dataset: Same listing details on different reviews





Next approach

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

- GridSearch (General Sweep)
  - Purpose: Baseline GridSearch
  - Limitation: Difficult to interpret
- GridSearch (N-grams)
  - Purpose: Better Interpretability
  - Limitation: Under aggregation assumptions

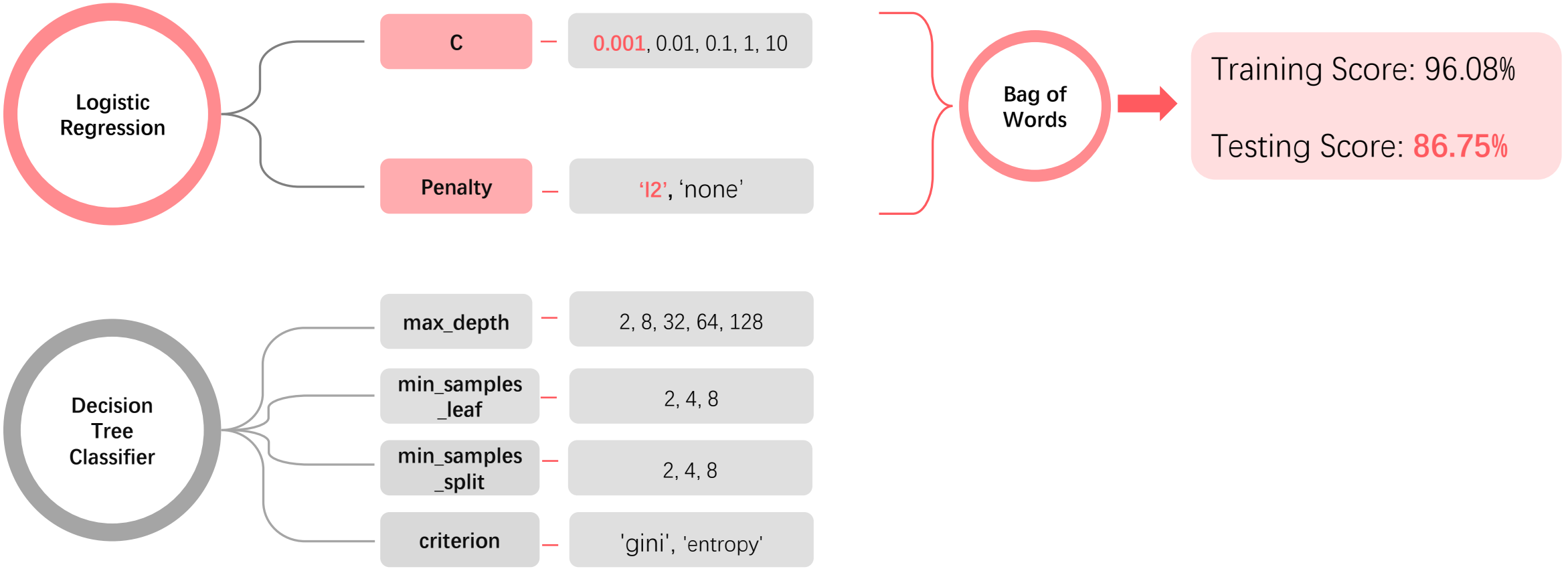
Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

- GridSearch (General Sweep)
  - Purpose: Better generalization
  - Limitation: No actionable findings
- GridSearch (N-grams)

airbnb

GridSearch 3: General Sweep

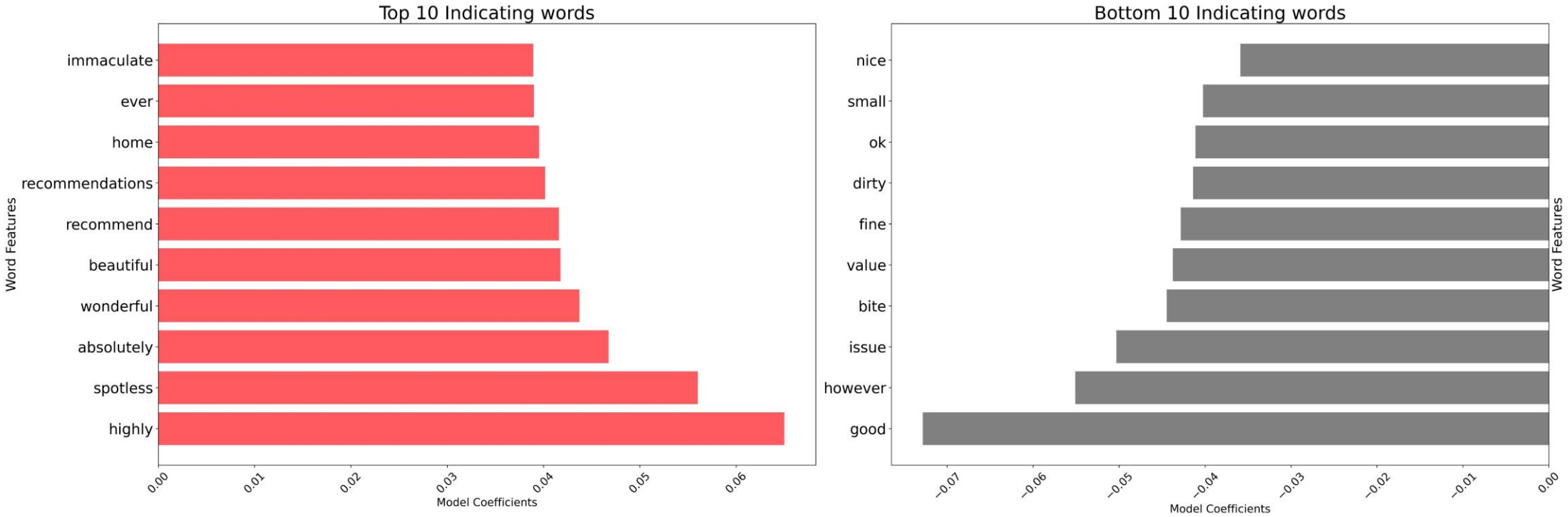
- Dataset: Reviews condensed to same listings



airbnb

GridSearch 3: General Sweep Extracted Features

- Dataset: Reviews condensed to same listings



airbnb

Modelling Workflow

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A
1	Features_1	4.3	B
1	Features_1	4.3	C
2	Features_2	4.6	D

GridSearch (General Sweep)

- Purpose: Baseline GridSearch
- Limitation: Difficult to interpret

GridSearch (N-grams)

- Purpose: Better Interpretability
- Limitation: Under aggregation assumptions

Listing ID	Listing Features	Rating Scores	Reviews
1	Features_1	4.3	A + B + C
2	Features_2	4.6	D

GridSearch (General Sweep)

- Purpose: Better generalization
- Limitation: No actionable findings

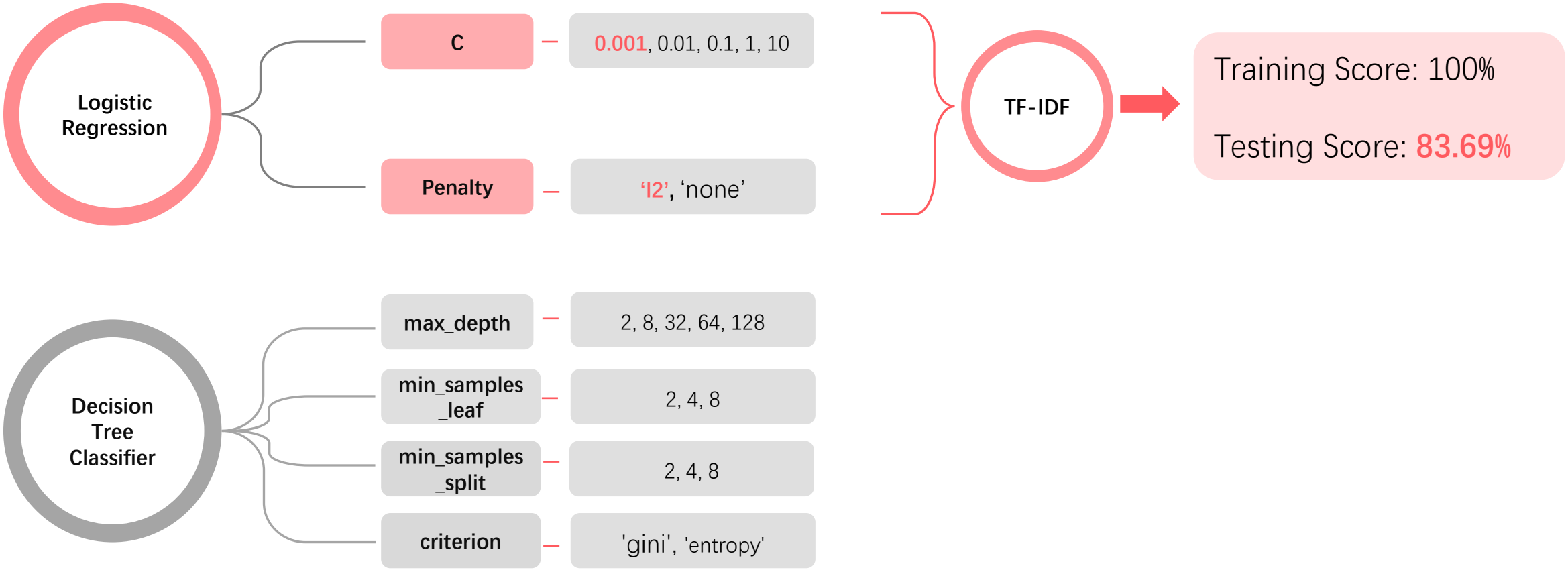
GridSearch (N-grams)

- Purpose: Better generalization & Interpretability
- Limitation: Negative indicators

airbnb

GridSearch 4: N-grams

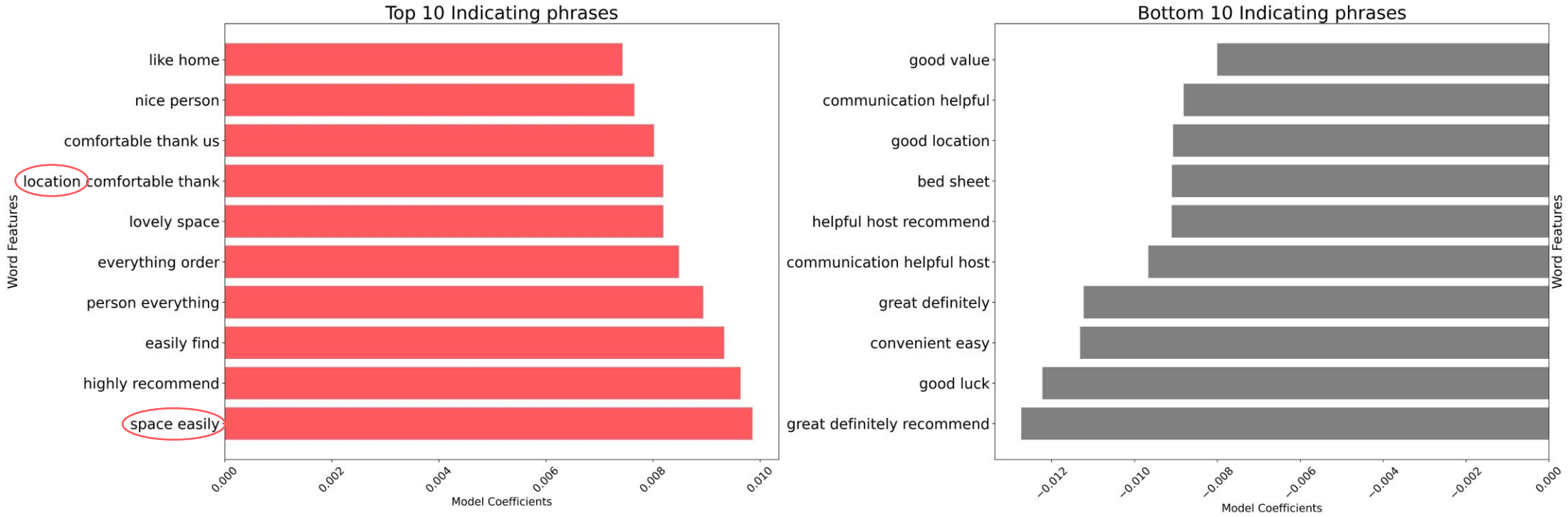
- Dataset: Reviews condensed to same listings



airbnb

GridSearch 4: N-grams Extracted Features

- Dataset: Reviews condensed to same listings



airbnb

Is it possible to extract sub-rating specific key features from guests reviews?

airbnb

Modelling on predicting sub-rating outcomes

Targets

- Location
- Cleanliness
- Communication
- Check-in

Data

- Uncondensed Review Dataset

Threshold

- Rating score of 4.8

Vectorizer

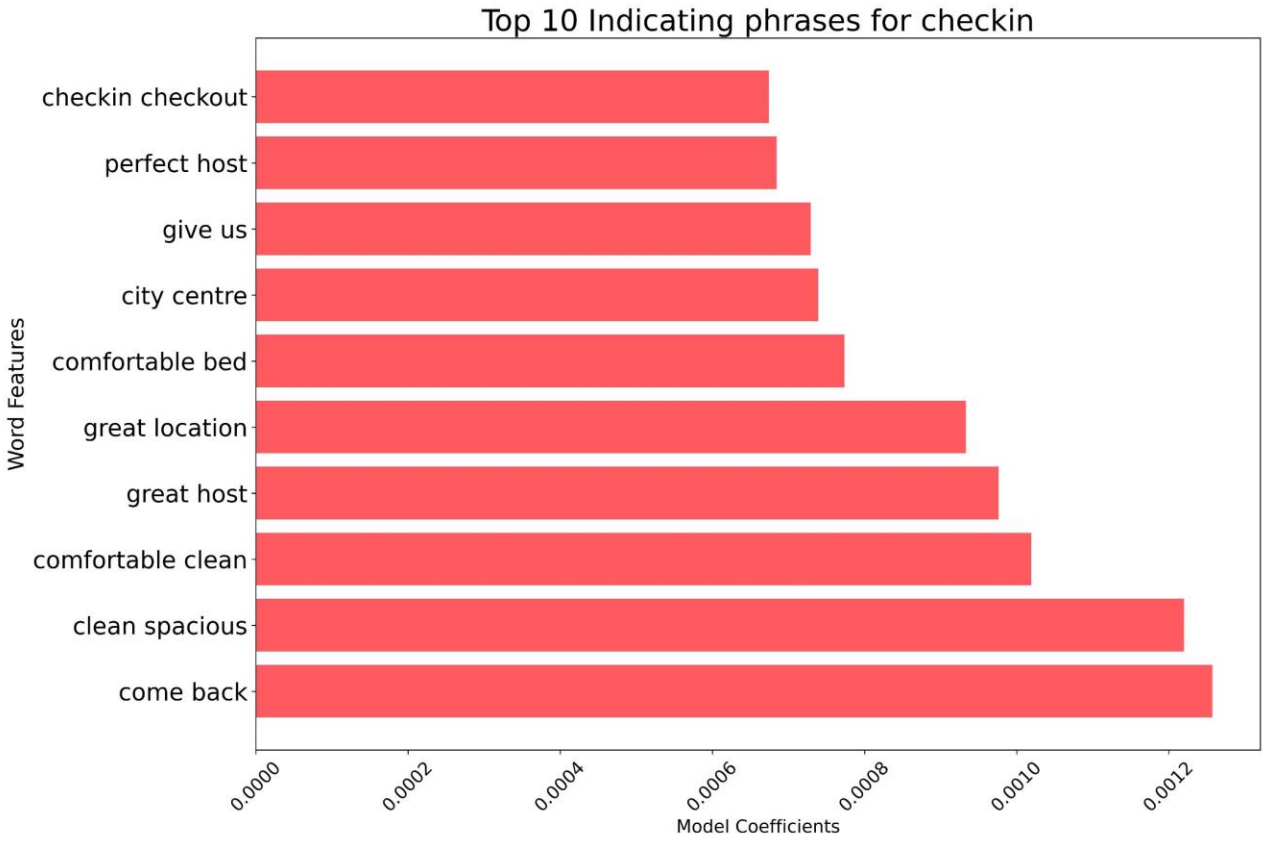
- Bag of Words with N-grams
- TF-IDF with N-grams

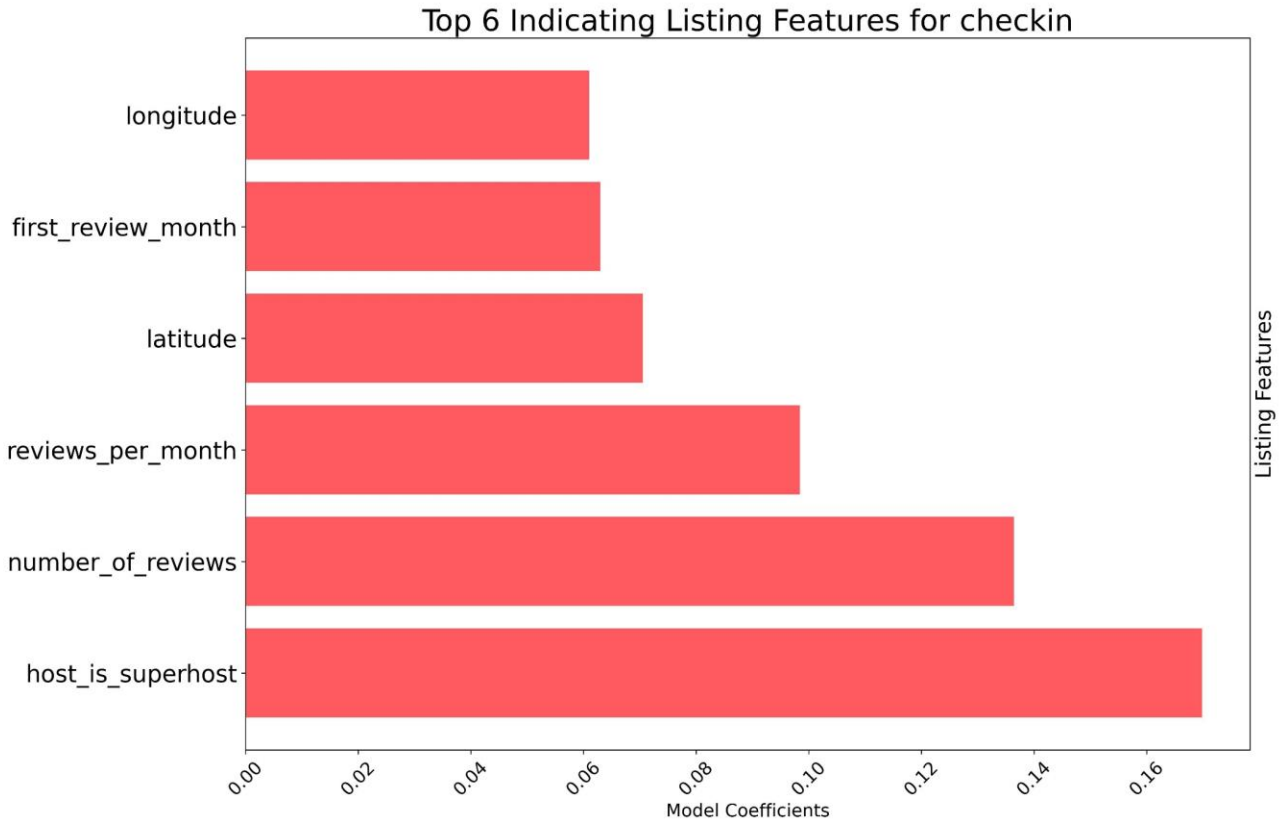
Model

- GridSearch on Decision Tree Classifier
- Small ranges of hyperparameters

airbnb

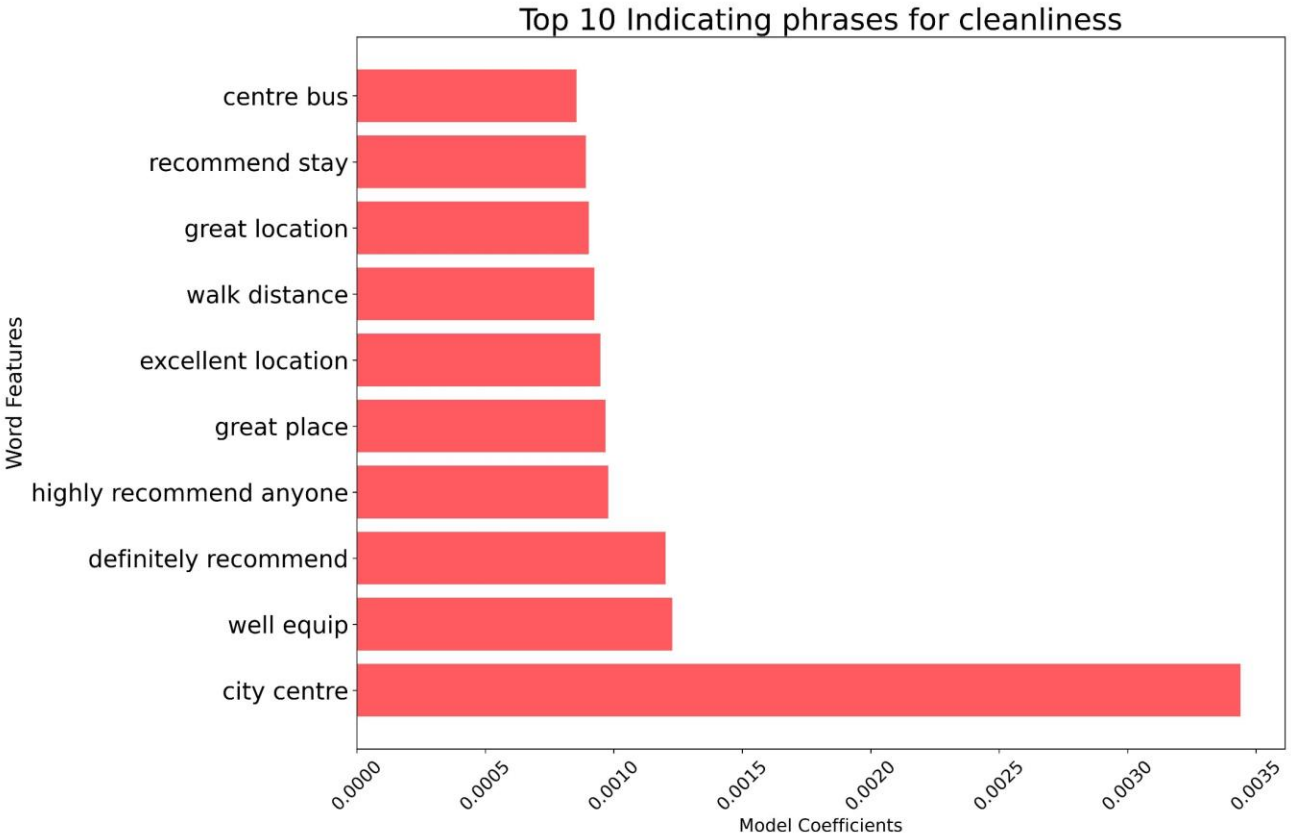
Sub-rating Check-in Predictions Results

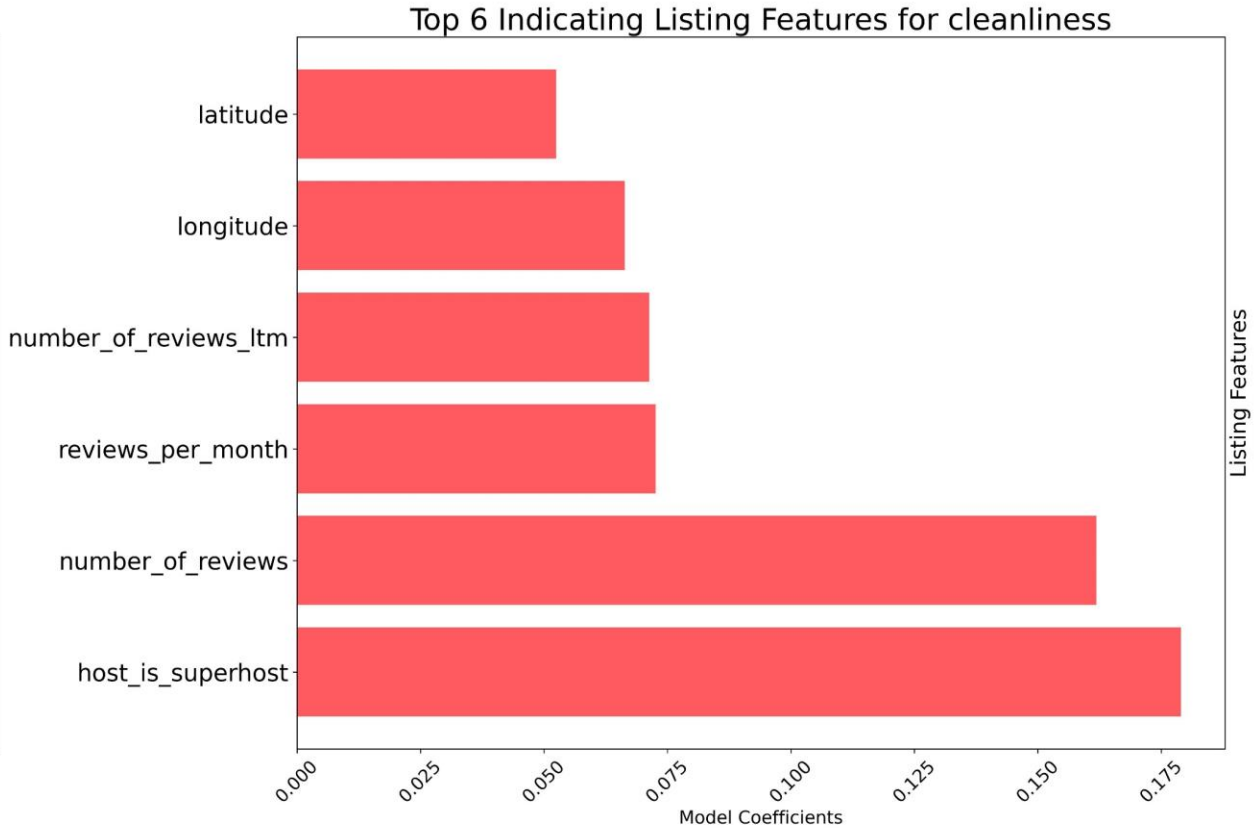




airbnb

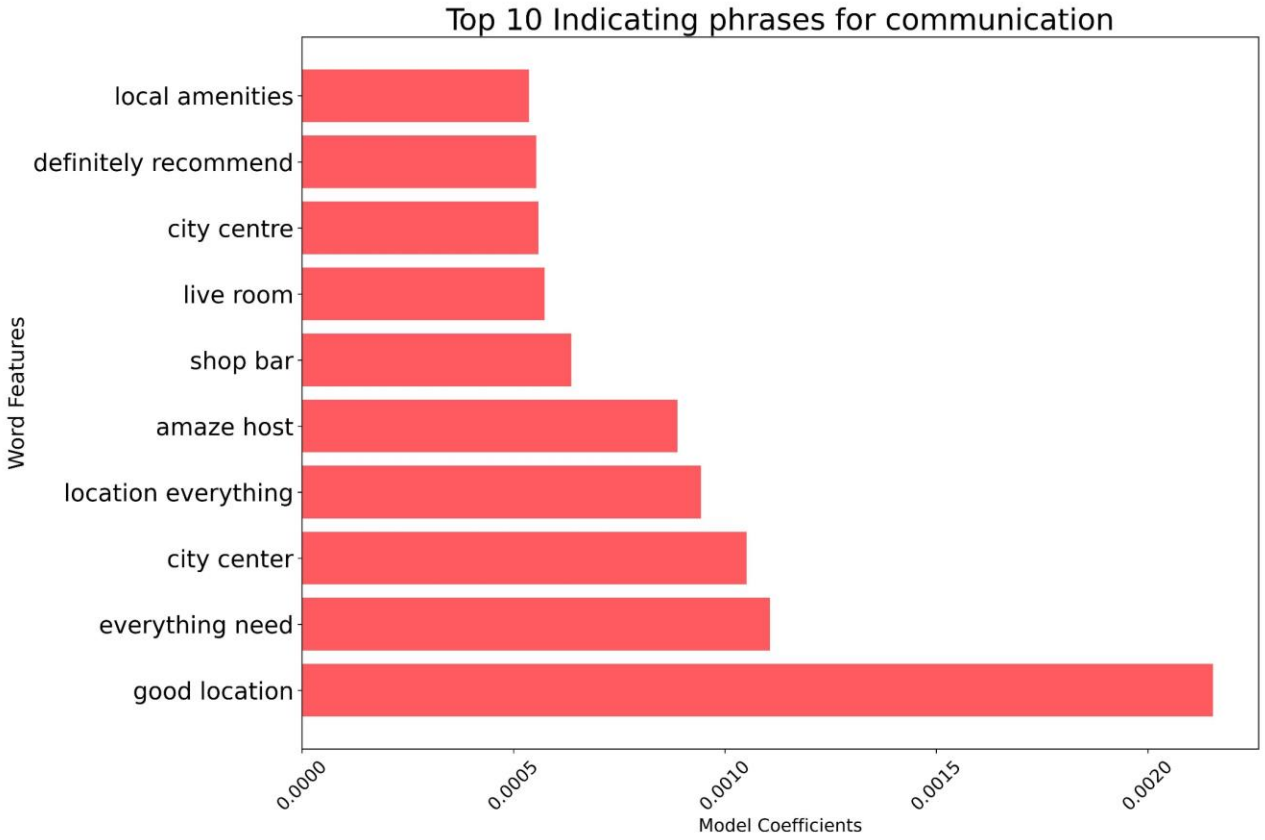
Sub-rating Cleanliness Predictions Results

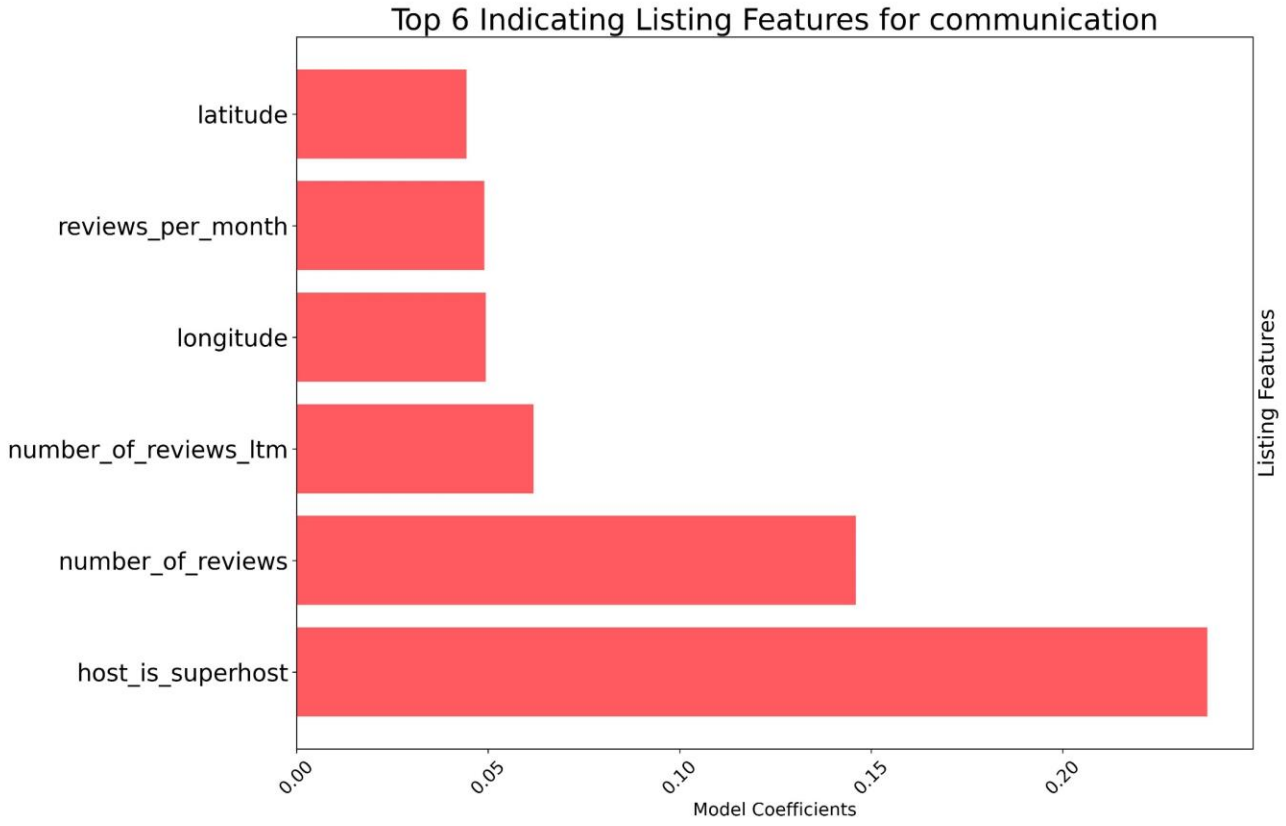




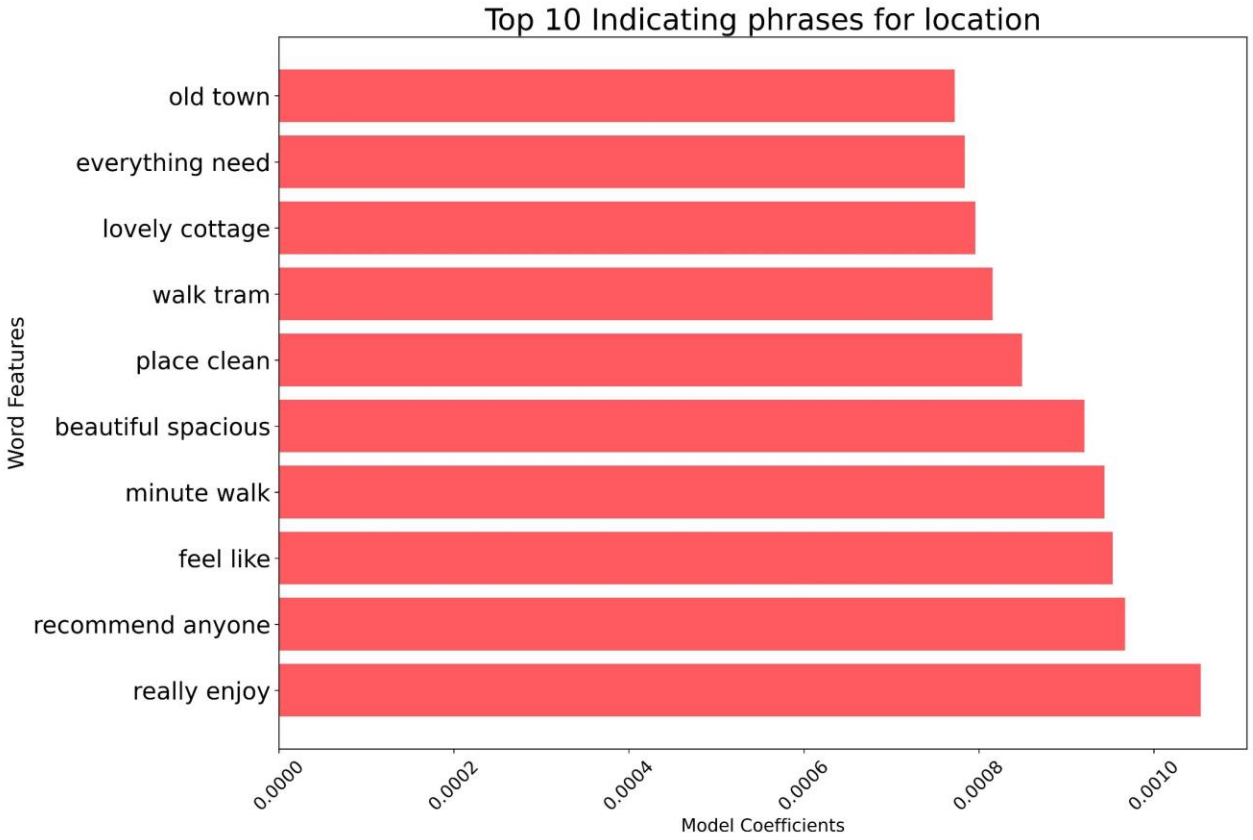
airbnb

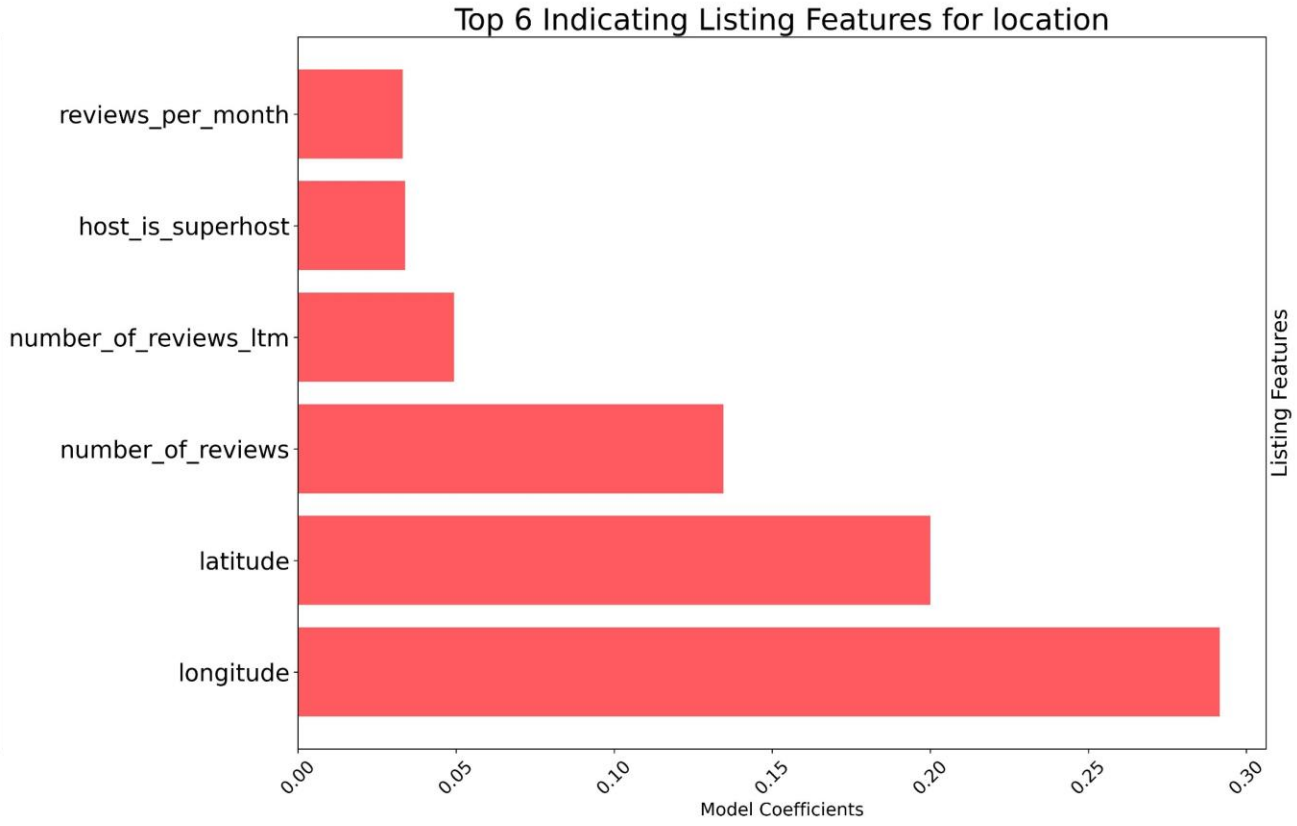
Sub-rating Communication Predictions Results



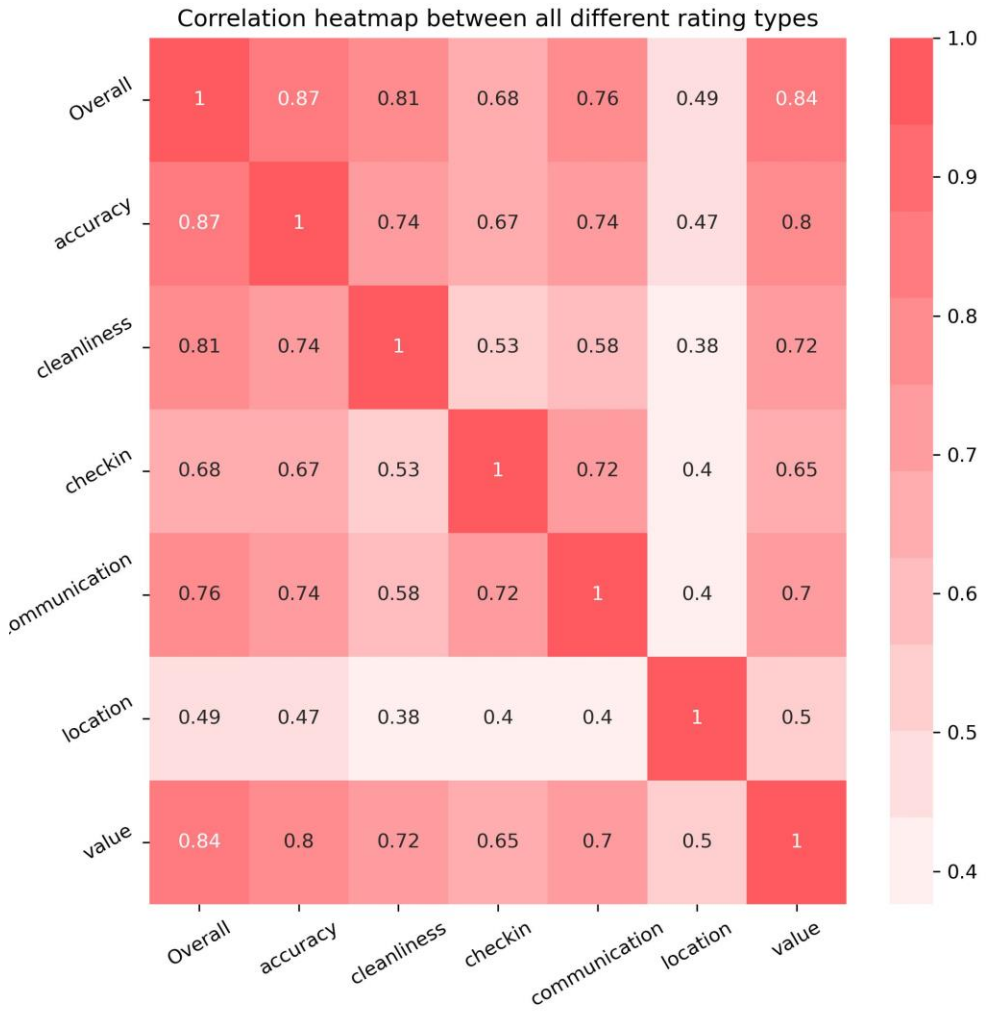


Sub-rating Location Predictions Results





Rating Scores Correlations



Next Steps

- Obtain corresponding rating scores for individual reviews
- Feature Engineering for sub-ratings
- Include more in-depth models to better interpret results

airbnb