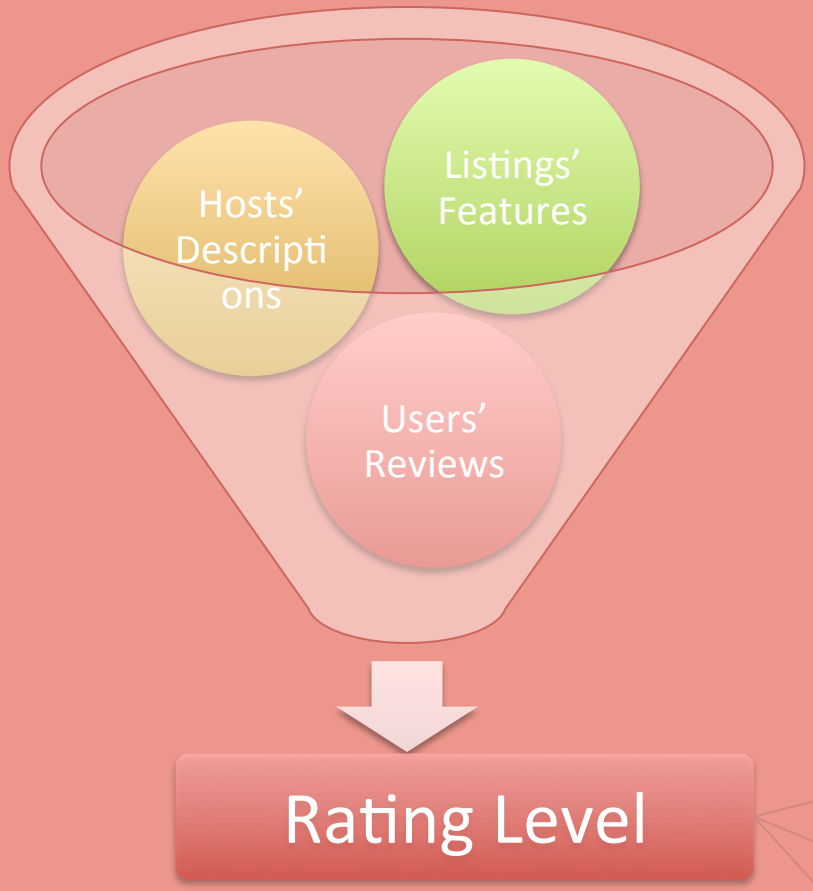
Airbnb Rating Level Classification

Presenter | Yiwen Xi
Email | yxx183@case.edu
Github| Florence17
Linkedin| yiwenxi

Project Overview



Goal

- Build a classifier to identify the rating level of an Airbnb listing

Motivation

- Help users avoid the real bad places quickly and pick up the only good ones
- Help Airbnb know better about their listings in the eyes of guests
- Improve classification accuracy by understanding guests' reviews

Best | 5 stars

Good | 4 stars

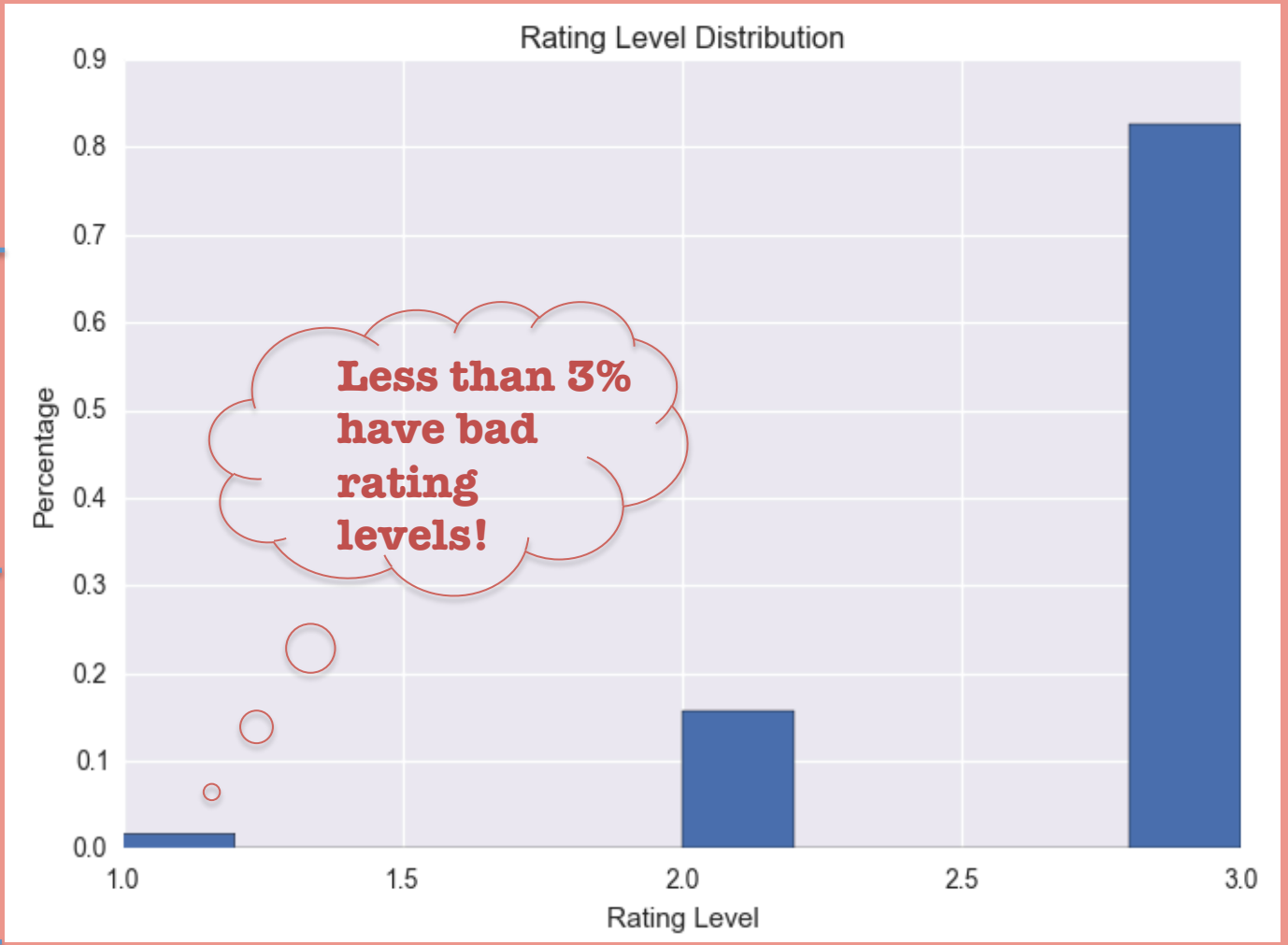
Bad | 3 stars or less

Imbalanced Data

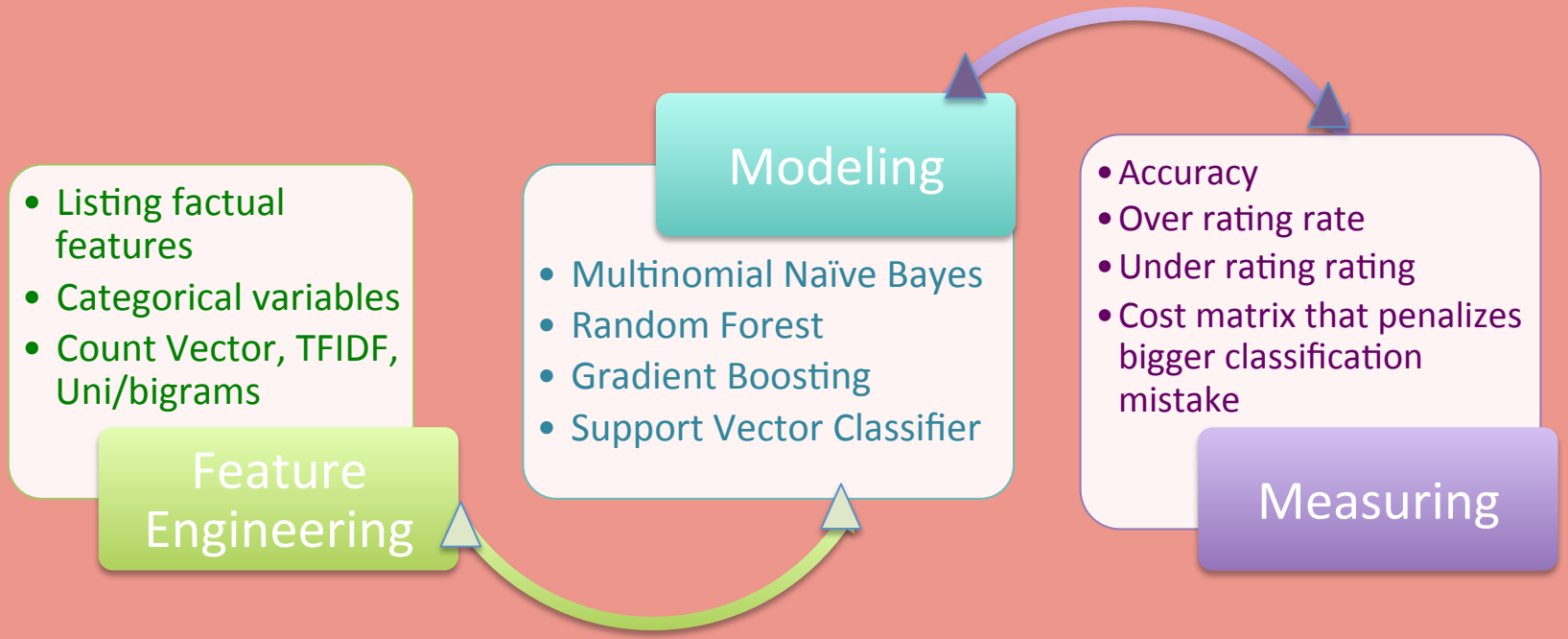
Not Enough data to tell the classifier what makes a bad rating

Different **Sampling Methods**

Under Sampling works best in detecting the bad rating level

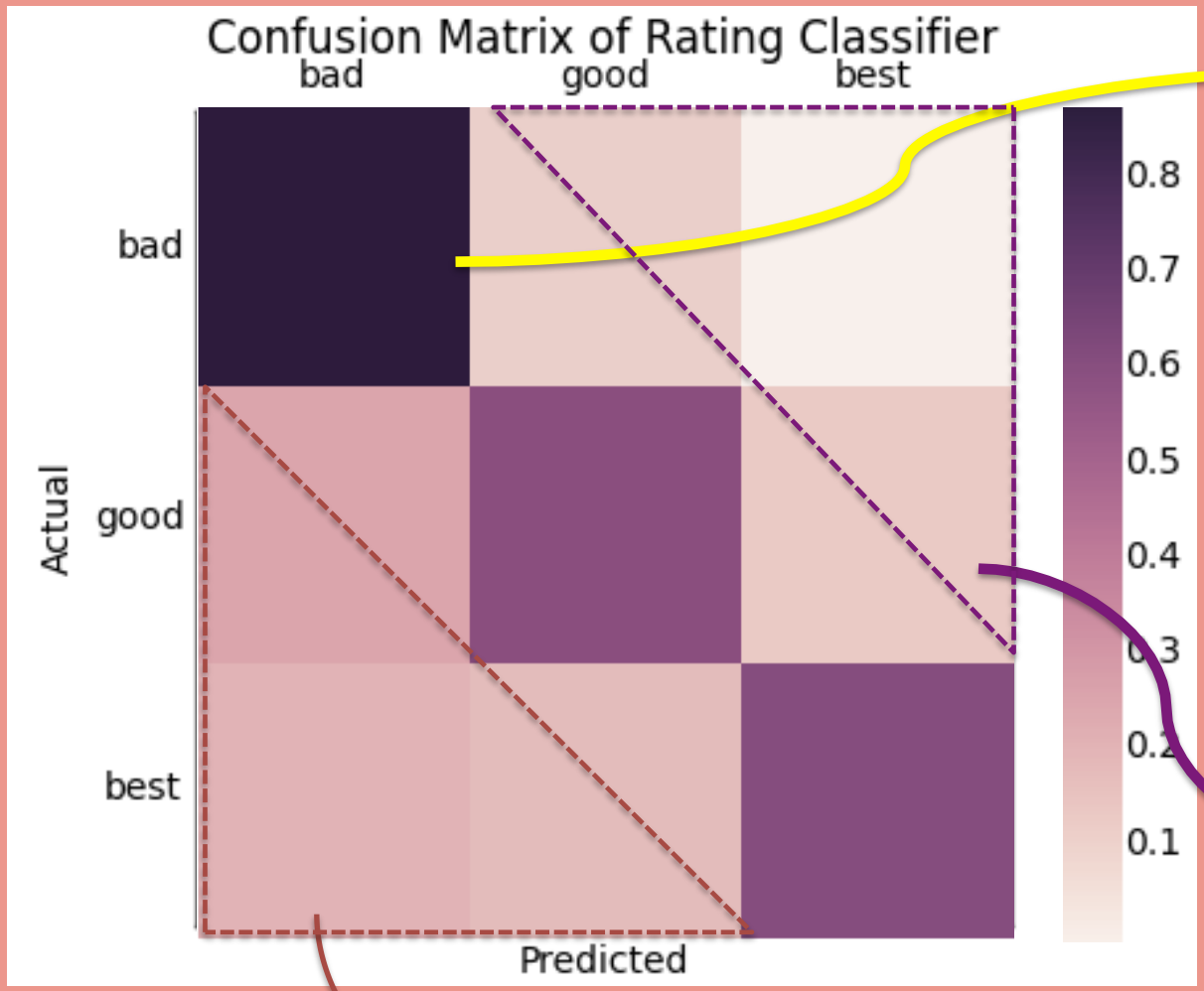


Model Building and Selection



	Model	Cross Validation Accuracy	Test Accuracy	Over Rating	Under Rating
Baseline	Random Forest with listing features	53%	48% Overfit	27%	25%
Final	Random Forest with Review TF-IDF Matrix	69%	69%	10%	21%

Random Forest with Review TFIDF Matrix

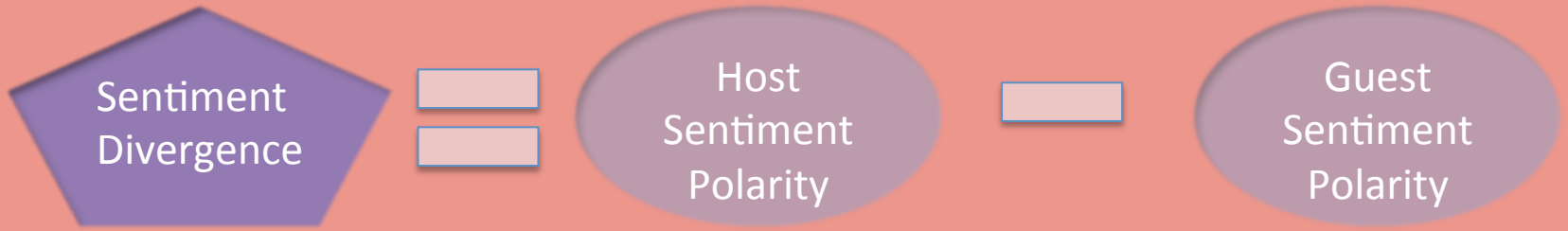


Good at detecting the bad ratings☺

Over Rating is not so common ☺

Under rate a little bit ☹

Sentiment Analysis



1.9

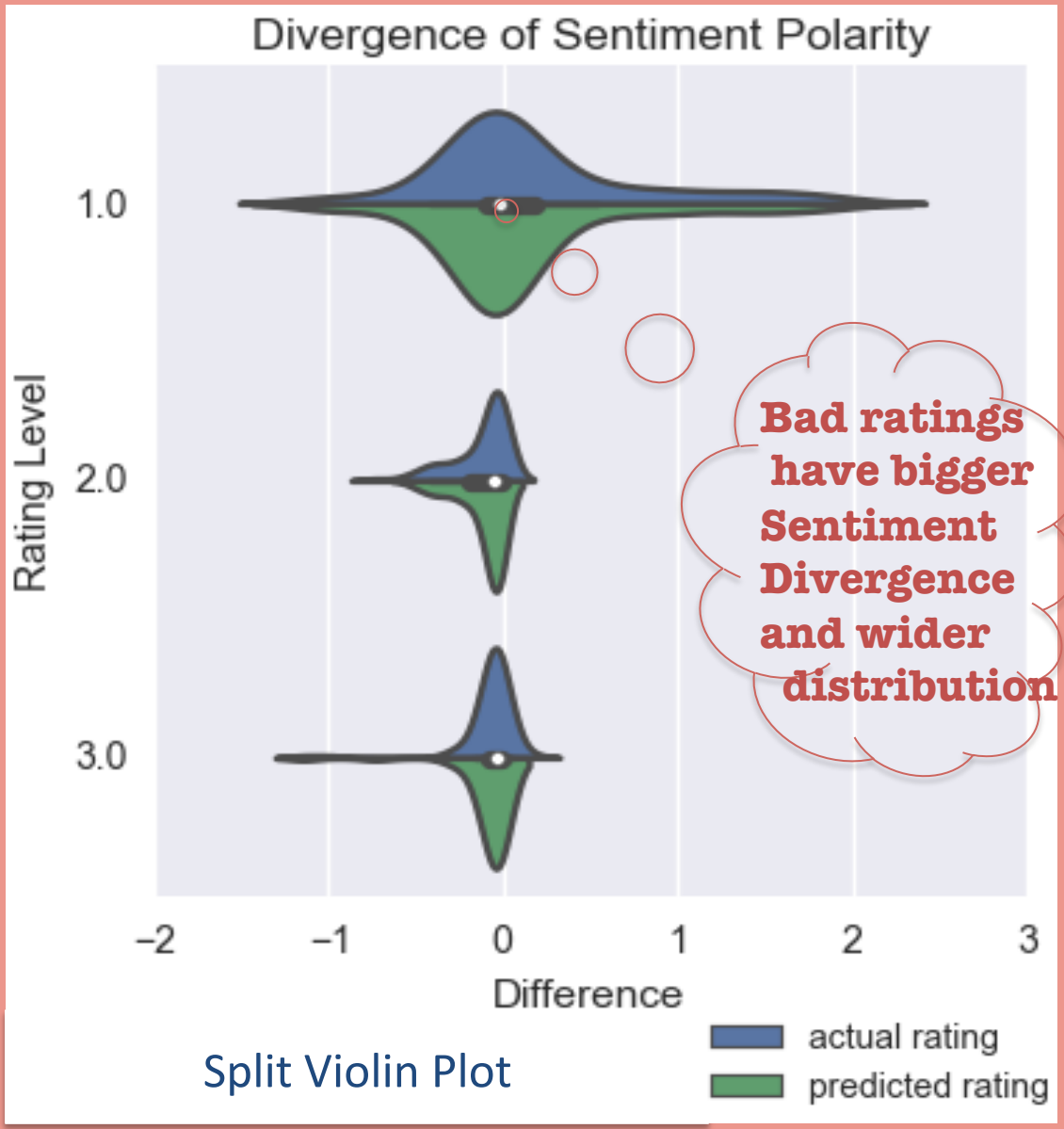
Great older 60's style ranch home in one of the best areas in Beverly Hills.
Great backyard for entaining and having events.
Nice big kitchen and open areas closet to Sunset Strip and Hollywood.
4 bedrooms 3 baths fire place.

0.9

Being charge an extra fee for a carpet that was already dirty with gum stuck on it and a so called broken door in a house that had broken shelves everywhere is not fair to me and my group. Completely not happy with this stay or how airbnb is handling this. We spent over 3k. We shouldn't be forced to spend more.

-1.0

Distribution of Sentiment Divergence



Split Violin Plot

Bad ratings have bigger Sentiment Divergence and wider distribution

How good the classifier matches up with the actual rating

Symmetricity

Wider the distribution, More divergent of the sentiment from guests to hosts

Width

Next Step

- Further improve model's accuracy by introducing sentiment features, listings picture info, etc
- Cluster listings into different groups and build a classifier for each
- Topic Modeling on reviews to find latent listing features guests care about
- Webapp to help user quickly target the best listings

Thank you!

- Questions?