Sentiment Analysis Yelp

Springboard Capstone Project 2
Serik Omarov

Background

- Yelp
- Online Social network services
- Consumer provides reviews and ratings
- 1 to 5 star- rating
- Customer feedback on services, quality and location etc

Business Problem

- Will the customer reviews help in the indication of the provided rating?
- How Can restaurants access their success and faults based on reviews?
- What aspects of the business are correlated between positive and negative sentiments?

Outline

- ❑ **Prepare Data for Supervised ML**
- ❑ **Descriptive Analytics**
- ❑ **Build Model**
- ❑ **Extract Results**

Project Approach

- Latent Sematic Analysis and Singular Value Decomposition
  - ✓ Topic Modeling
  - ✓ Dimensionality Reduction
  - ✓ Relationship between documents and Terms

- Logistic Regression
  - ✓ Predictive Binary Classification
  - ✓ Distinguish between two classes
  - ✓ BOW and TF-IDF to extract terms and coefficients for feature importance

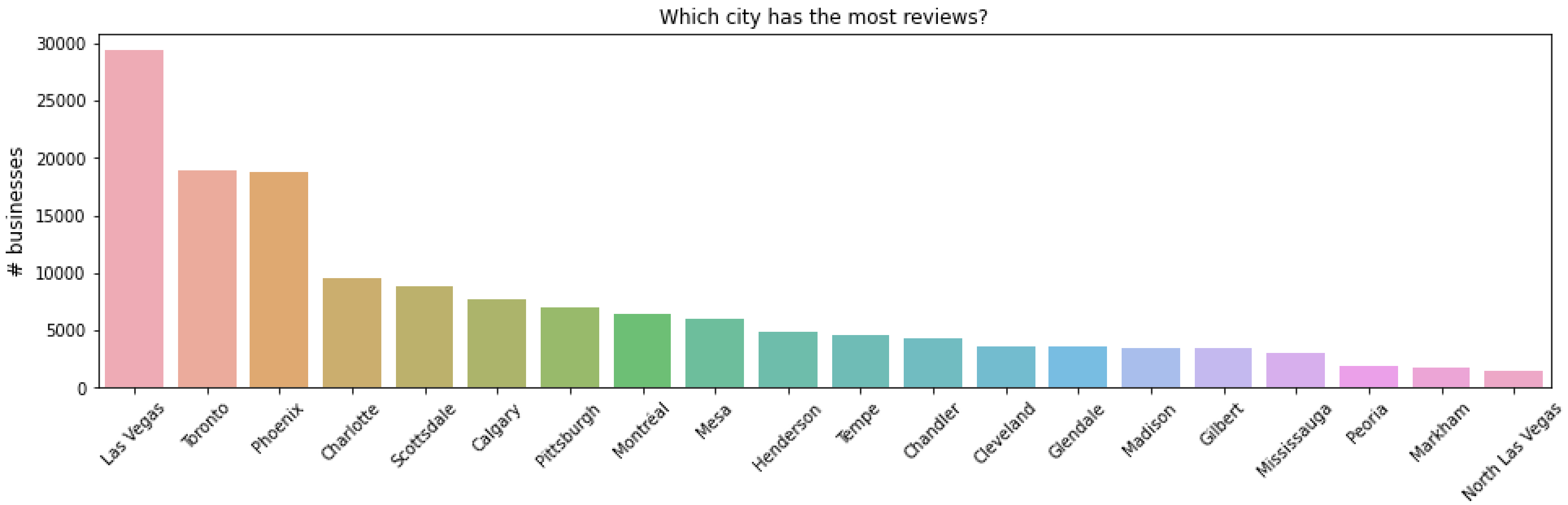
Data Sources

Project Approach

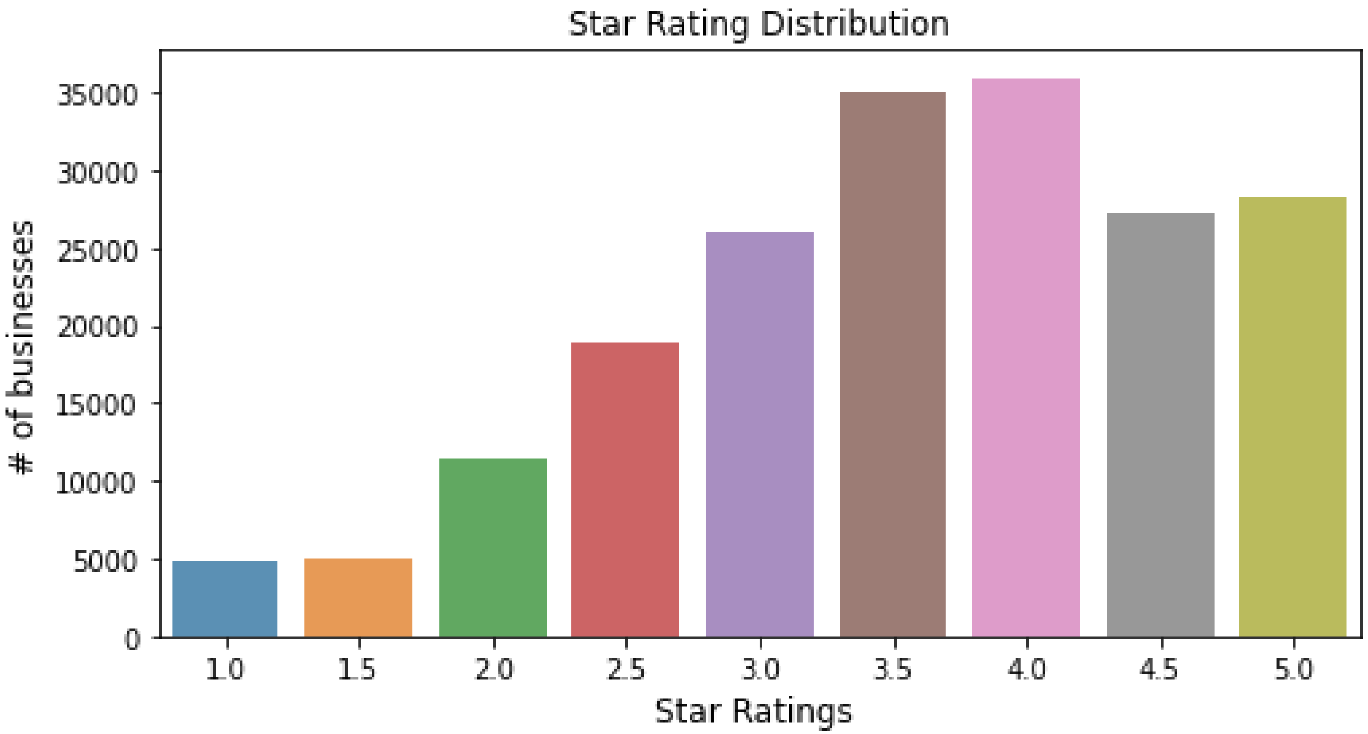
- Latent Sematic Analysis and Singular Value Decomposition
  - ✓ Topic Modeling
  - ✓ Dimensionality Reduction
  - ✓ Relationship between documents and Terms

- Logistic Regression
  - ✓ Predictive Binary Classification
  - ✓ Distinguish between two classes
  - ✓ BOW and TF-IDF to extract terms and coefficients for feature importance

Exploratory Data Analysis



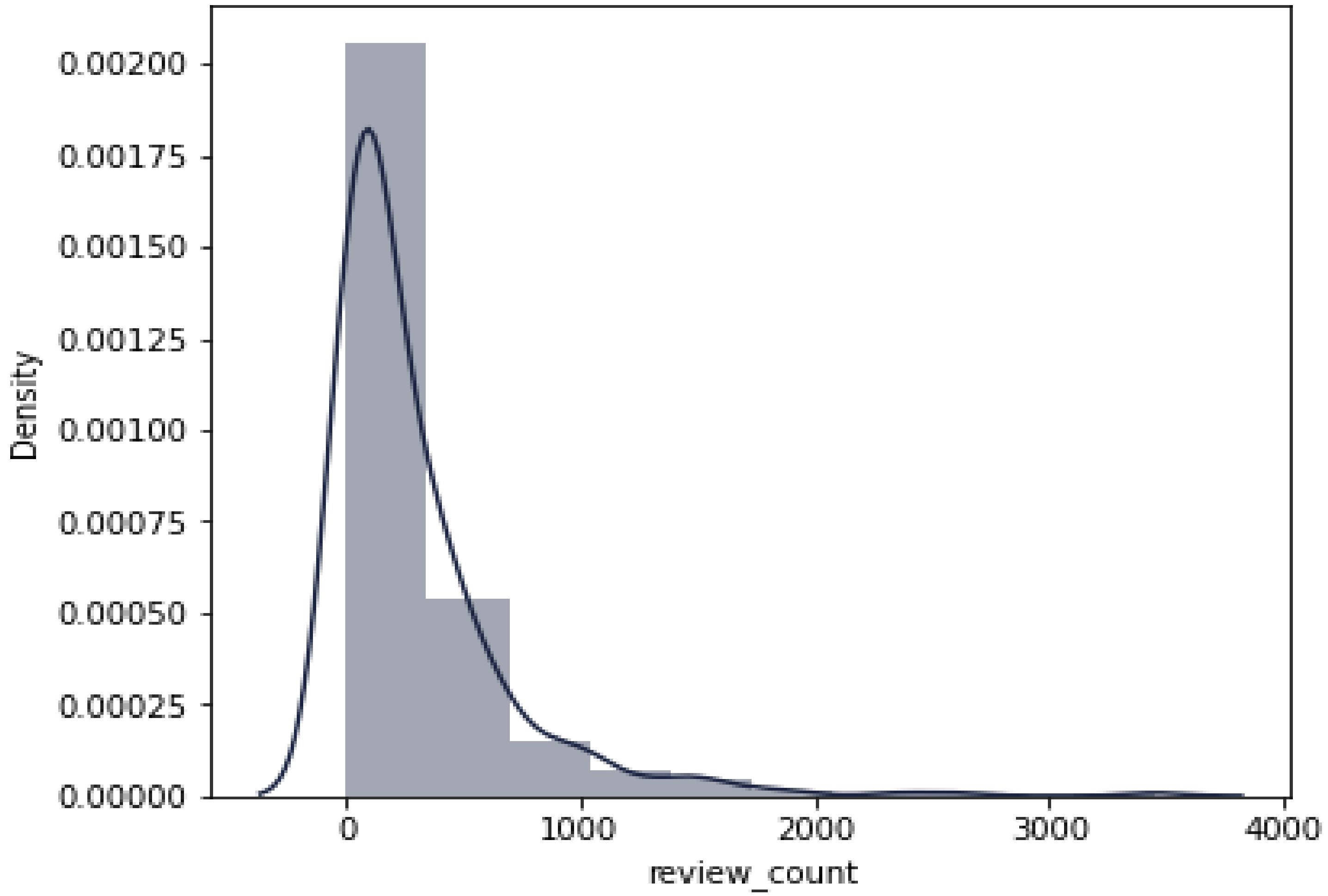
Exploratory Data Analysis



Exploratory Data Analysis: Las Vegas

Exploratory Data Analysis: Las Vegas

Distribution of Count of Reviews



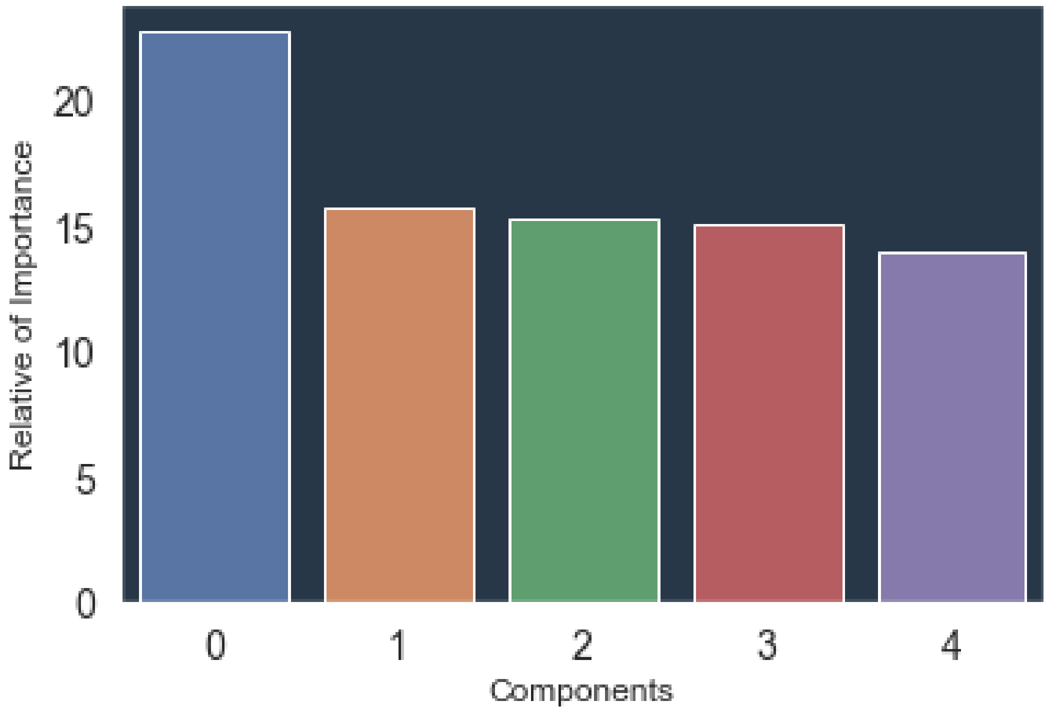
Exploratory Data Analysis

Negative Sentiment

Positive Sentiment

Latent Semantic Analysis (LCA)

- ❑ 5 concept of the top 10 words
- ❑ Understand relationship between document and terms
- ❑ Average conceptual idea of consumer's reviews and experiences

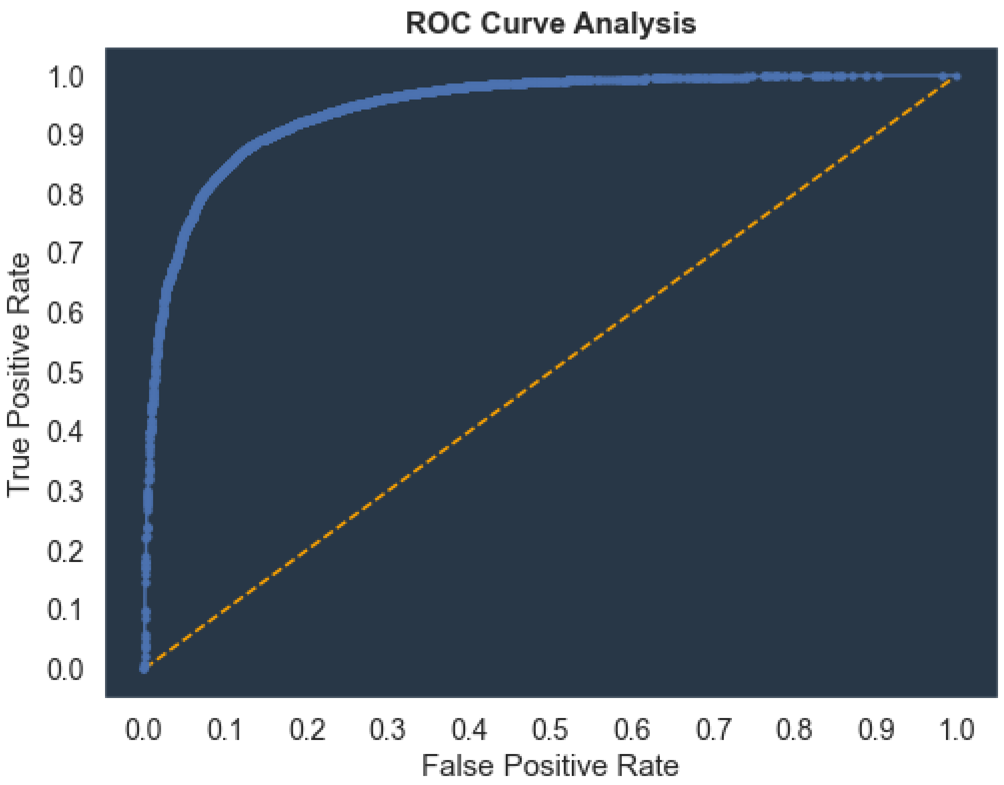



```
{'Concept 0': [('come back', 0.2431157699795417),
  ('sushi place', 0.1799710778081639б),
  ('happi hour', 0.1780473263387264 4),
  ('la vega', 0.1680606516389793 2),
  ('great servic', 0.1601527456538924),
  ('great food', 0.1536805335141426),
  ('servic great', 0.1356425593111377),
  ('food great', 0.1315095547684829 2),
  ('first time', 0.1296723260082838 5),
  ('highli recommend', 0.12603940331979532)],
 'Concept 1': [('great food', 0.5218811933526182),
  ('great servic', 0.4926298680860966 6),
  ('food great', 0.4280575951443961),
  ('servic great', 0.2475597987304875 6),
  ('great price', 0.08176026154274607),
  ('place great', 0.05714059964624758),
  ('love place', 0.05361362620297335),
  ('food servic', 0.05290096731960285),
  ('definit come', 0.05038352348897965),
  ('great atmospher', 0.04896926979287501)],
 'Concept 2': [('happi hour', 0.6312600412663628),
  ('hour menu', 0.1125805100234496 5),
  ('sushi place', 0.1124378037487351),
  ('great servic', 0.0967818749023581 1),
  ('la vega', 0.0952437599803730 3),
  ('best sushi', 0.0924332602548735 8),
  ('great food', 0.0859152098812962 5),
  ('food great', 0.0638524163425380 5),
  ('great happi', 0.0606336073345181 6),
  ('favorit sushi', 0.0583404106288695 8)],
 'Concept 3': [('happi hour', 0.6548203854641816),
  ('come back', 0.3810942669475722 6),
  ('definit come', 0.2359821426704506 2),
  ('hour menu', 0.1187132177995879 6),
  ('would definit', 0.0610645089374114 4),
  ('great happi', 0.0594124989567593 5),
  ('hour price', 0.0518912039382881 1),
  ('late night', 0.0510609240053496 2),
  ('back tri', 0.0449520715613066 8),
  ('first time', 0.0437589900449026 4)],
 'Concept 4': [('come back', 0.3800730436786854 4),
  ('sushi place', 0.3465569870321958 4),
  ('best sushi', 0.3108785488245836 6),
  ('definit come', 0.2601014523884045),
  ('la vega', 0.2152988359892752),
  ('happi hour', 0.1397904583508747),
  ('favorit sushi', 0.1395310684048973 2),
  ('one best', 0.0817862459749592 5),
  ('sushi restaur', 0.0720665027759246),
  ('ayc sushi', 0.0683681270613398 1)]}
```

Logistic Regression Implementation

- ✓ **N-Grams (n_gram_grange) : Uni, Bi, Tri**
- ✓ **Feature Selection(max_feature)**
- ✓ **Word Frequency Exclusion (min/max_df)**

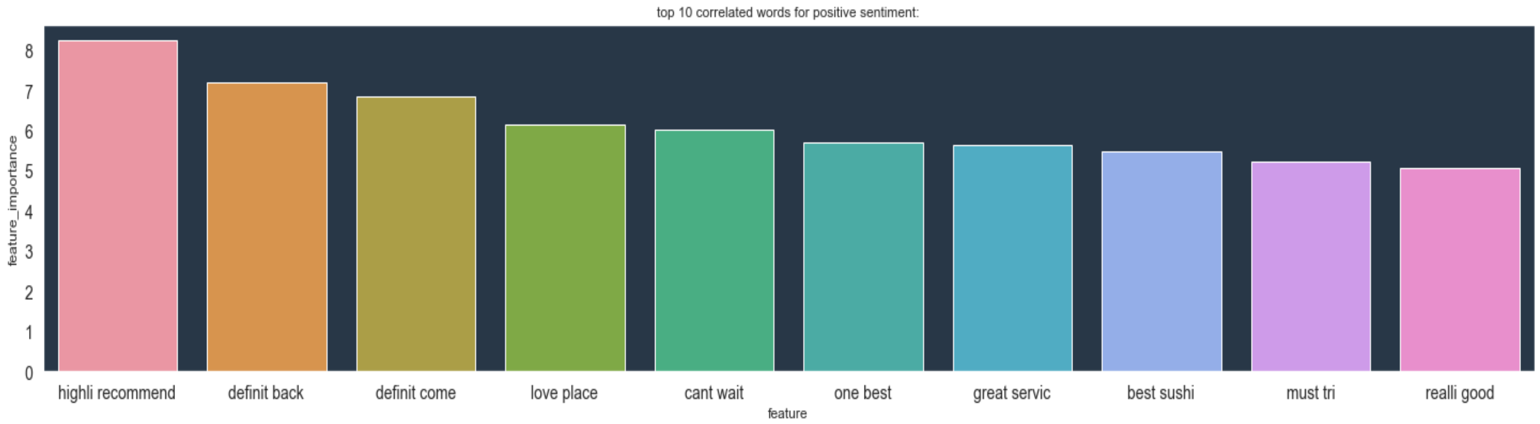
Logistic Regression

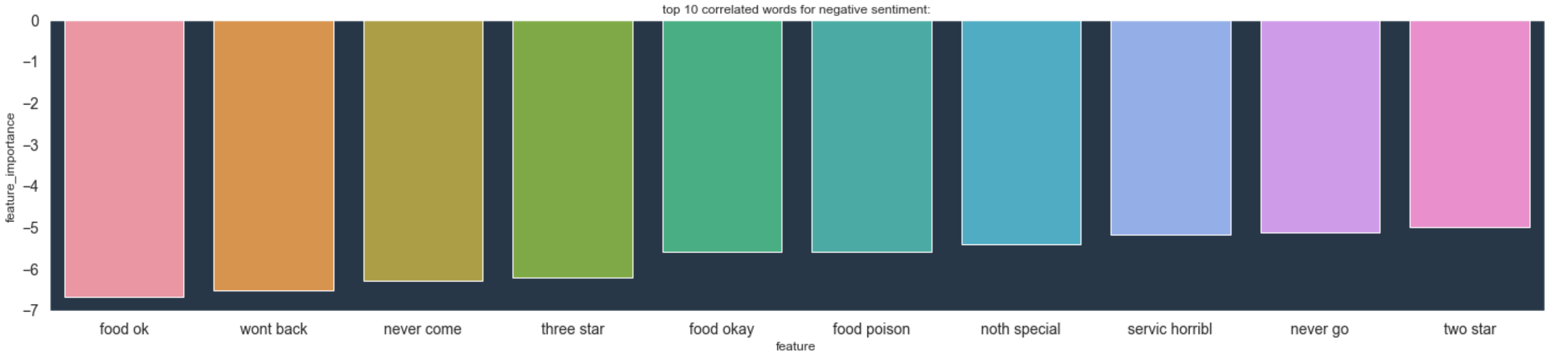


Evaluation Metrics

- **Accuracy:89.25%**
- **Recall:96.55 %**
- **F-1: 92.99 %**
- **Precision: 89.69 %**

Feature Importance





Conclusion

- **Throughout this analysis I found that sentiment words were more positive than negative of Yelp users ‘experiences at the designated restaurants. In relation to the positive sentiments of users, there is a positive correlation between that of positive reviews with high ratings, and negative reviews with lower ratings. By segmenting the area to Las Vegas and categorizing Japanese restaurants, i was able to gain insight on how they operate.**

- **Each of the individual users provide their opinions throughout their reviews; as the positive outweighs the negative, Japanese restaurants are providing great dining services and food to their customers, which increases the positivity disclosed in their review, as well as an input of a higher rating. Although, through this project, Japanese restaurant owners can also view aspects that drive more negative sentiments which they can take initiative on and remedy over time.**