

# ***Predicting Success on***

Ramin Ostad and Will Sundstrom



September 24,

# Objective

Our aim is to develop a success prediction for new restaurants

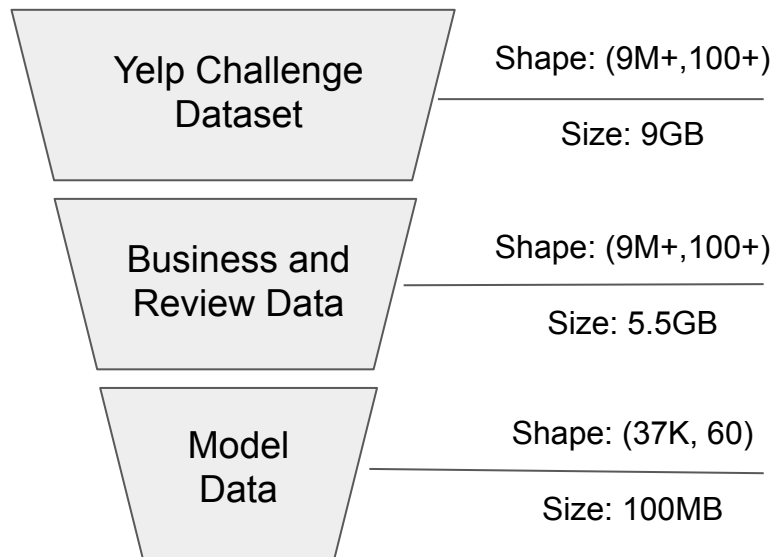
Success is defined as 4+ stars

The predictor can help new restaurants evaluate whether they are 'on track' to becoming successful restaurants



# Data overview

The dataset comes from Yelp and was honed down in Cloud SQL



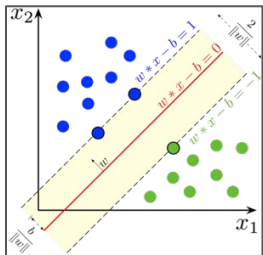
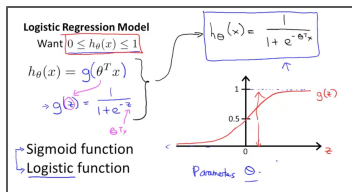
## Data Overview:

- Data provided through Yelp 2019 Challenge
- Used Google Cloud SQL to store data and run initial cuts
- Types of data:
  - Business information
  - Reviews and tips
  - User information
- Data ultimately used in model:
  - Business information
  - Aggregate characteristics of reviews for each restaurant

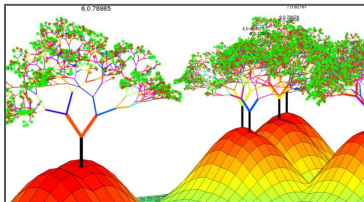
# Approach

## Logistic Regression, SVM and Random Forest

Logistic  
Regression



Random  
Forest



## Success Metrics

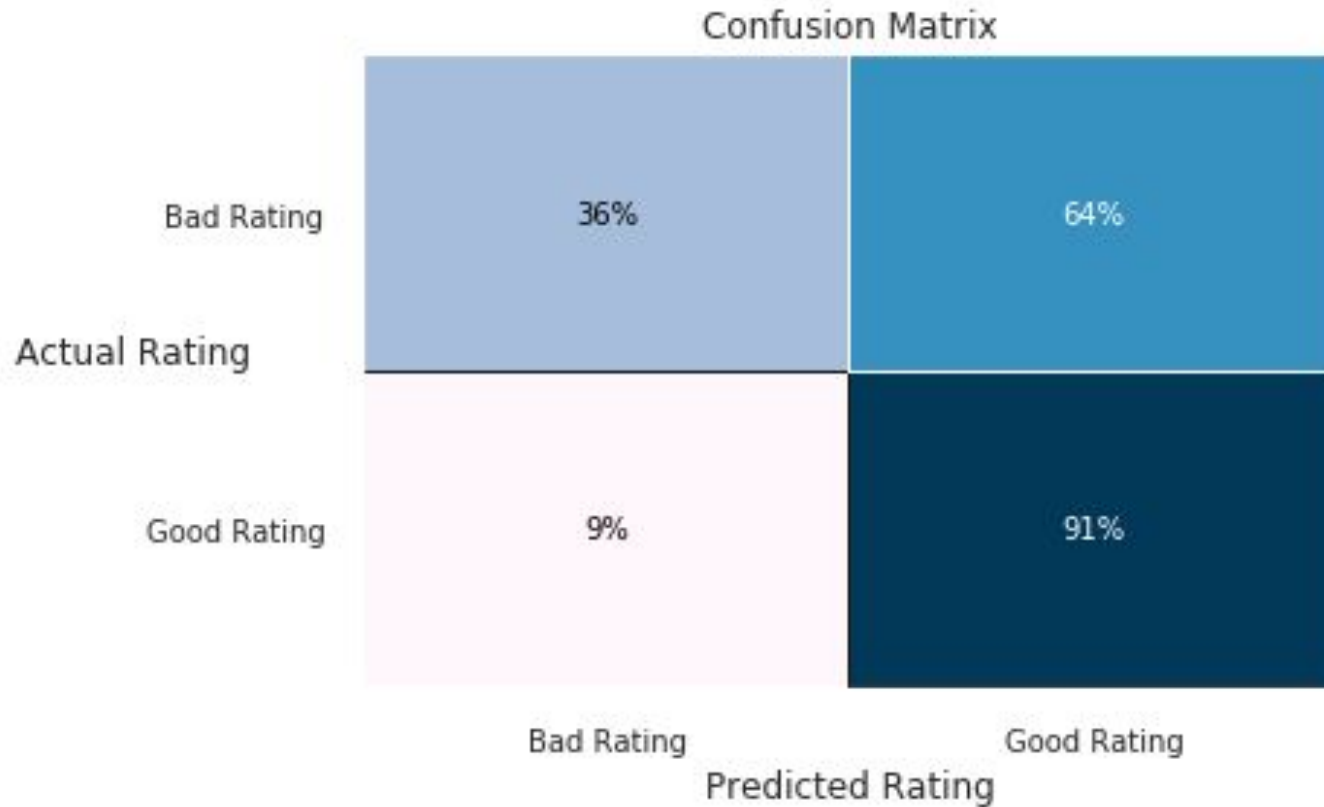
- Accuracy
  - False positive and false negatives are equally bad outcomes
- Time Taken
- Model Interpretability

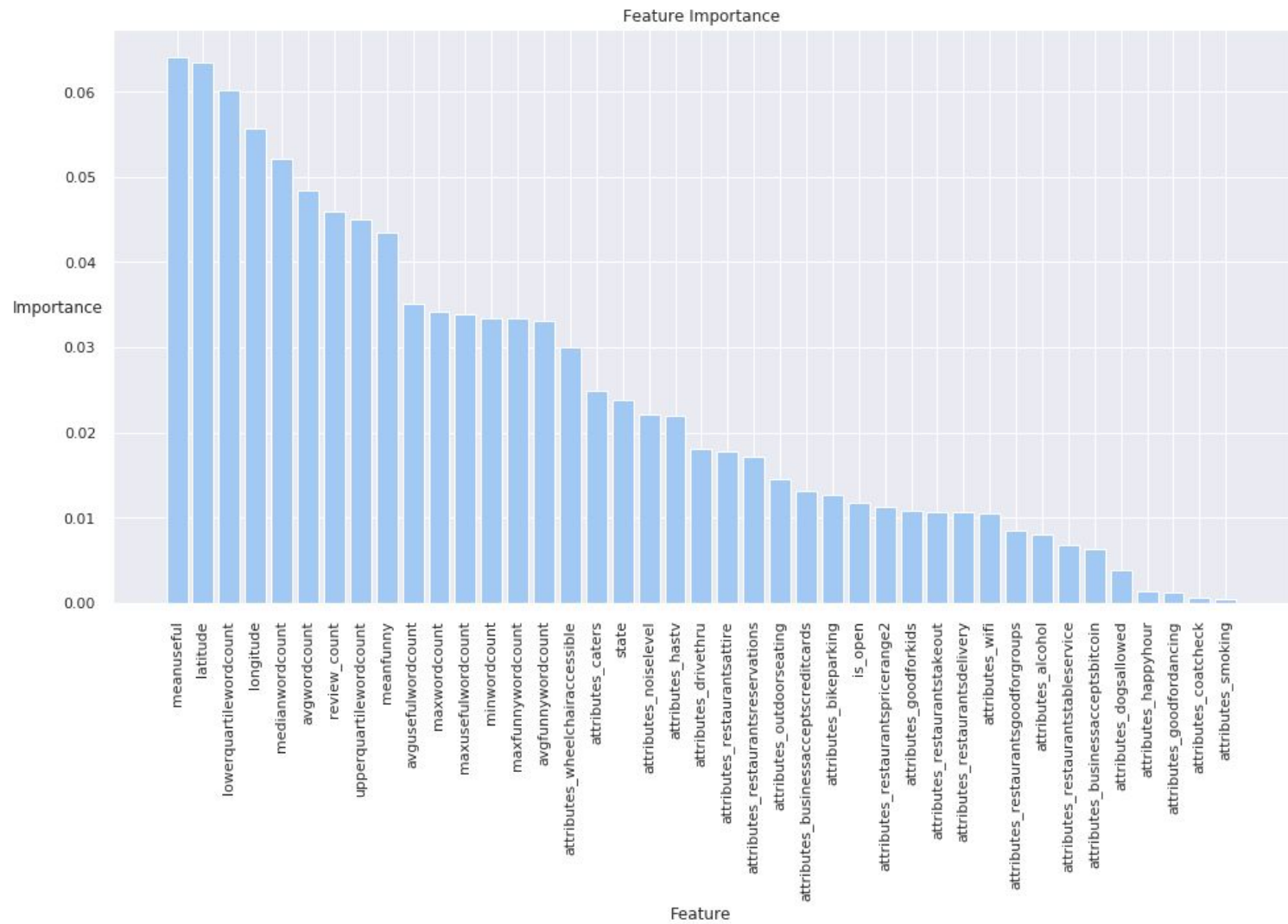
# Model Selection

## Comparing Average Time Taken and Accuracy

Classifier	Accuracy	Time Taken	Interpretability
Random Forest	71.7585%	6.97 seconds	Feature Importance
SVM Nystroem	71.3564%	54.92 seconds	
SVM Monte Carlo	70.4559%	33.68 seconds	
Logistic Regression	70.9743%	4.60 seconds	Odds Ratios

# Model Performance





# Conclusion

## Takeaways

- Random Forest is effective and efficient
- Review length and engagement is predictive
- Location, location, location

## Next Steps

- NLP on review text
- Review time series
- Look at check-in data
- Weight reviews based on user influence metrics



***Thank You!***