

Amazon Product Sentiment Analysis

Srinivasan Ramakrishnan



Overview



01

**Business
Understanding**

02

**Data Understanding/
Preparation**

03

Modeling

04

Evaluation

05

Recommendations

06

Next Steps



Business Understanding

- Companies, including Amazon, need a sentiment analysis to improve their products
- The sentiment analysis -> increased customer satisfaction

Data Understanding

- Data is from AWS
- 19 columns with around 9 million reviews
- Predicting star rating (from 1 star to 5 stars)
- Focusing on star rating and review body features



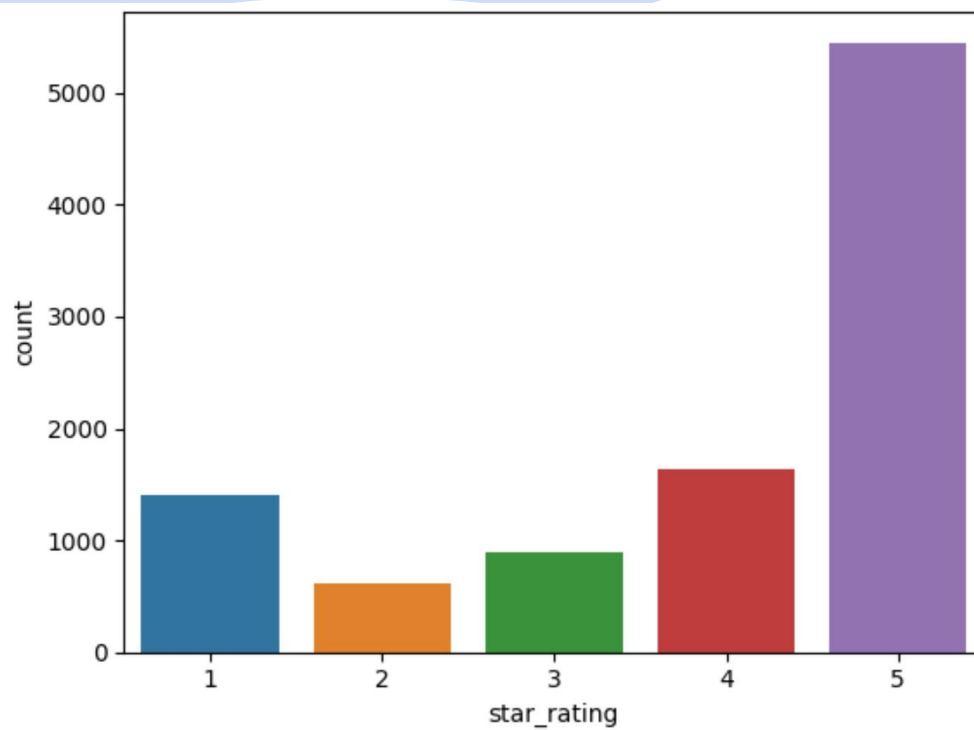
Preparation



**Fix class
imbalance**

**Dropped
null/unnecessary
values**

**Randomly
sampled 10,000
reviews**



Preparation



1.

Tokenization

- Word Tokenizer

2.

Stop Words

- Removed stopwords

3.

Normalization

- Snowball Stemmer

4.

Vectorization

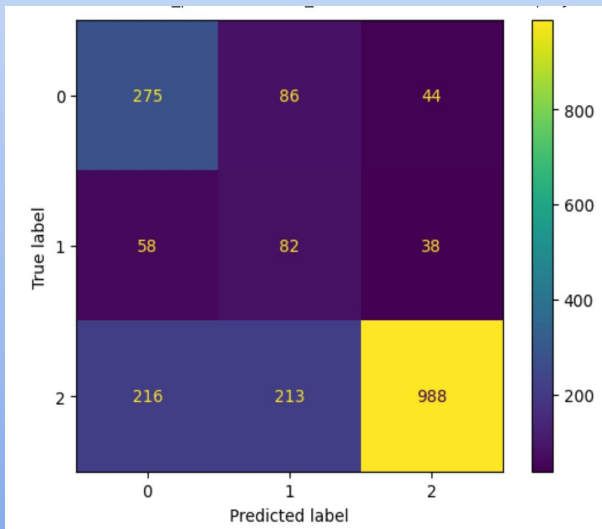
- TFIDF Vectorizer



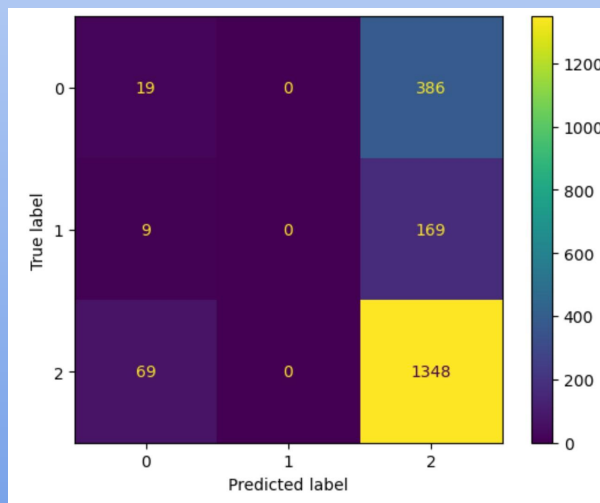
Modeling



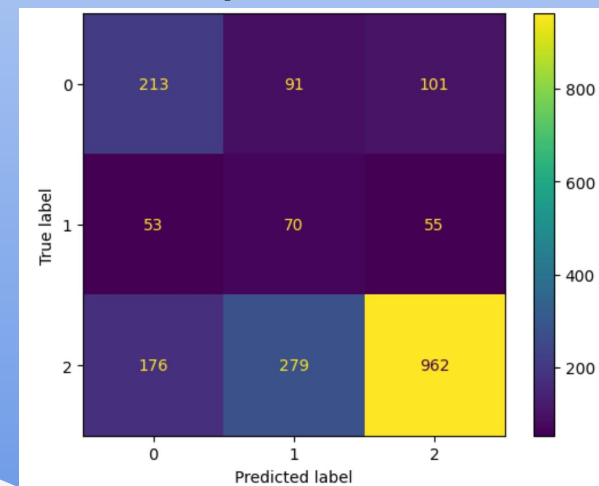
XGB Classifier -> **77% precision**



KNeighbors Classifier -> **54% precision**



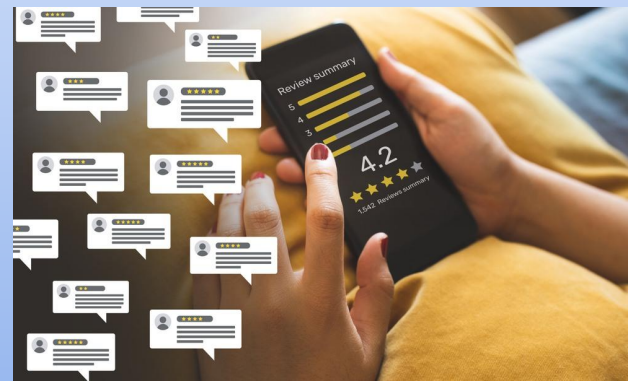
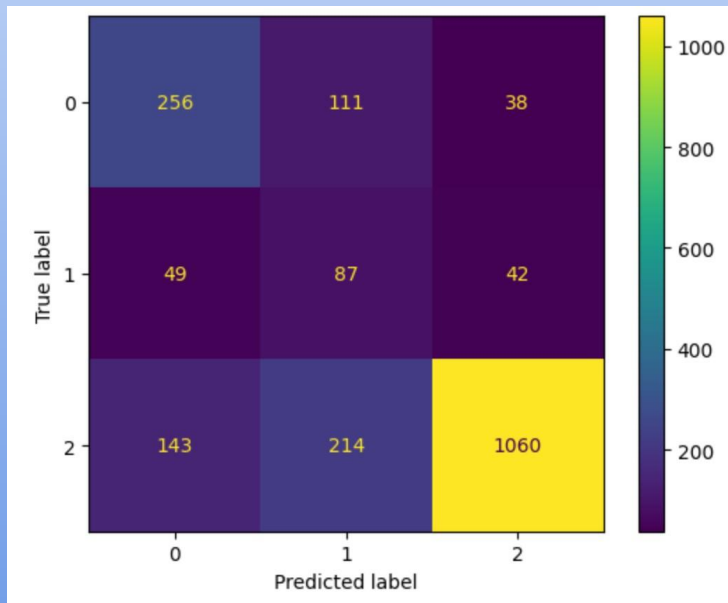
Random Forest Classifier -> **72% precision**



Evaluation



Base Model: Logistic Regression -> **79% precision**

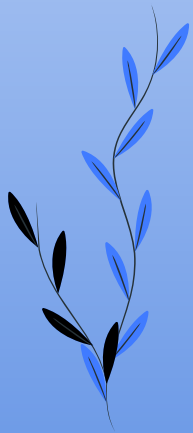


Recommendations



**Use the
Logistic
Regression
Model**

**Model not best
with negative/
neutral
reviews**





- Try using lemmatization instead of stemmer
- Sample a larger set of the data to see how accurate the model is
- Improve the precision in the negative and neutral class

Next Steps