PREDICTING PRODUCT RECOMMENDATIONS FROM AMAZON REVIEWS

QMSS GR5067 NLP FOR THE SOCIAL SCIENCES
GERALD LEE (GL2668)

OVERALL ARCHITECTURE

DATA SOURCING

Repository of Product reviews webscrapped from Amazon by Jianmo Ni (2014 - 2018)

DATA CLEANING

Text Cleaning and Preprocessing

DATA LABELLING

Labelling Recommended Products based on 5-star ratings

SENTIMENT ANALYSIS

Use AFINN lexicon to determine if sentiment of reviews

MACHINE LEARNING

Built model to classify reviews into Recommended or Not Recommended

EVALUATION

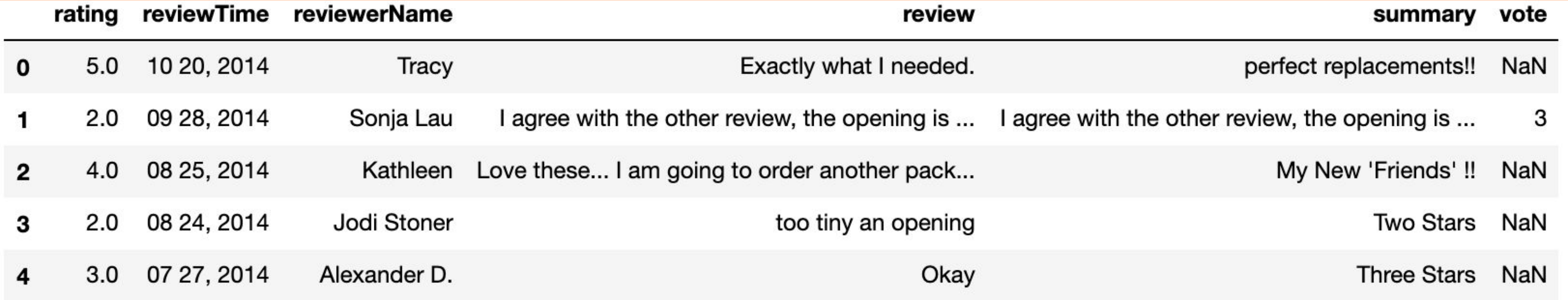
Evaluate Model and Visualize Outputs

OVERVIEW OF DATA

- Amazon Product Reviews from 2014 - 2018
- Webscrapped by Jianmo Ni
- 883,636 rows x 6 columns

product rating 1 - 5 stars

plain-text review

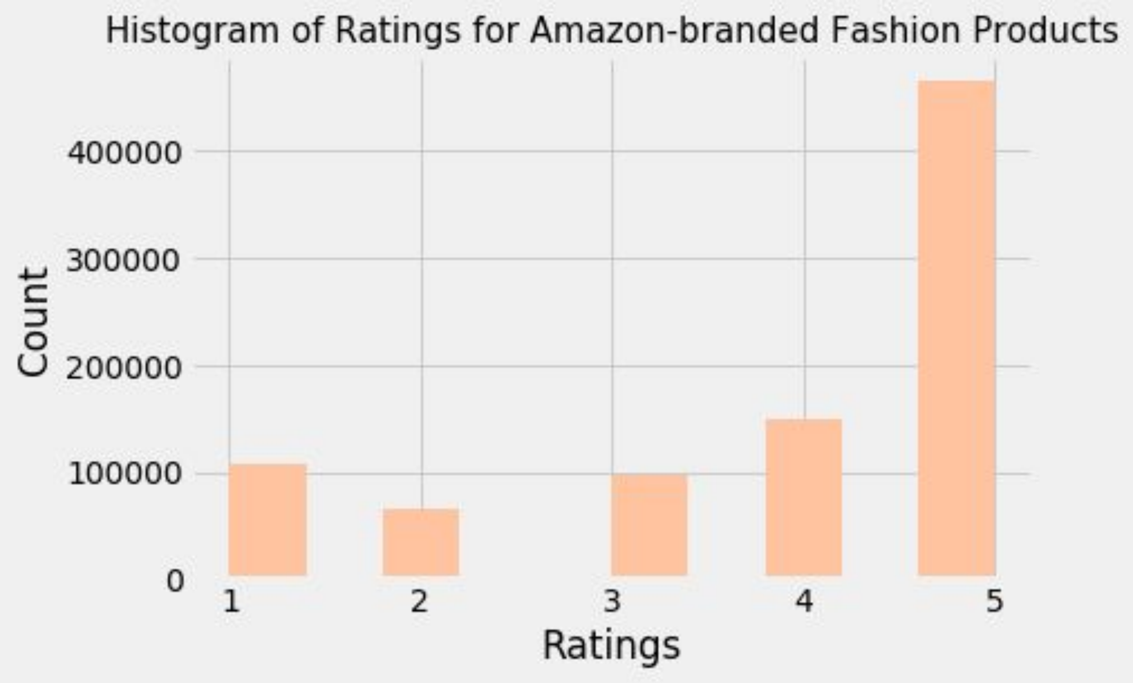


	rating	reviewTime	reviewerName	review	summary	vote
0	5.0	10 20, 2014	Tracy	Exactly what I needed.	perfect replacements!!	NaN
1	2.0	09 28, 2014	Sonja Lau	I agree with the other review, the opening is ...	I agree with the other review, the opening is ...	3
2	4.0	08 25, 2014	Kathleen	Love these... I am going to order another pack...	My New 'Friends' !!	NaN
3	2.0	08 24, 2014	Jodi Stoner	too tiny an opening	Two Stars	NaN
4	3.0	07 27, 2014	Alexander D.	Okay	Three Stars	NaN

time of rating / review

number of votes for whether review was helpful or not

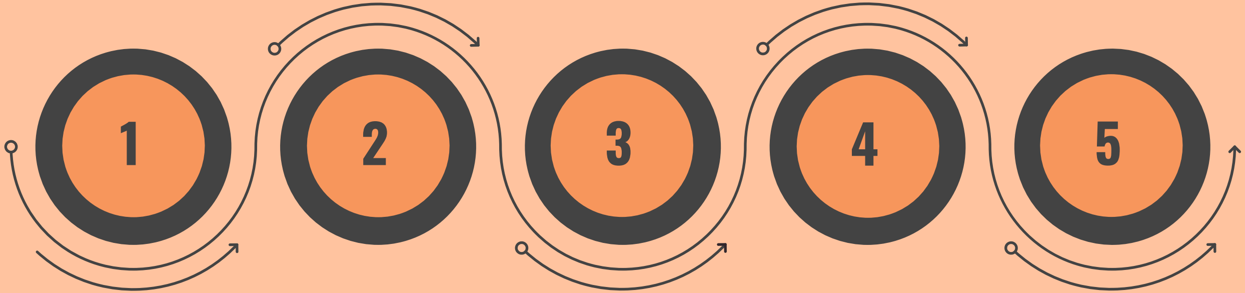
INITIAL INSIGHTS FROM EXPLORATORY DATA ANALYSIS



Rating Inflation

What does 5-star mean?

METHODOLOGY



Remove Punctuations, non-ascii characters and numbers

STEP 1

STEP 2

Change all text to lowercase, remove stop words and tokenize text

Use WordNetLemmatizer to lemmatize the text

STEP 3

STEP 4

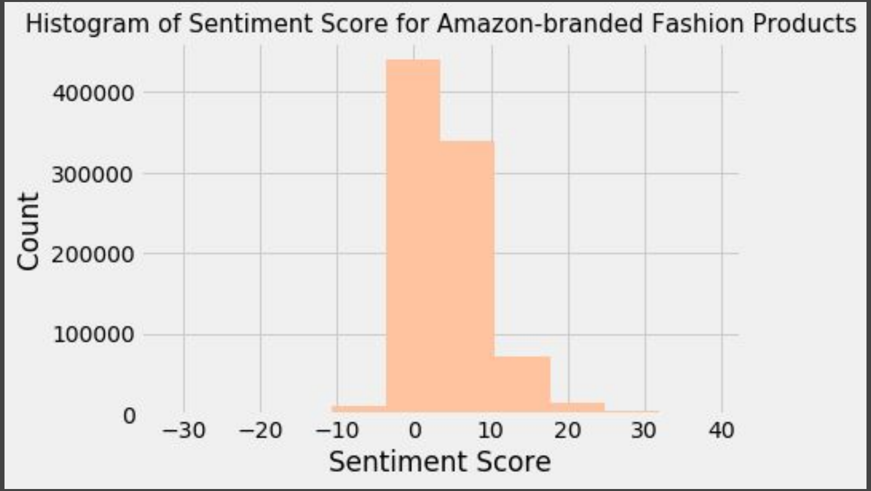
Carry out sentiment analysis with the AFINN lexicon for each review

Compare the Sentiment Score produced from SA process

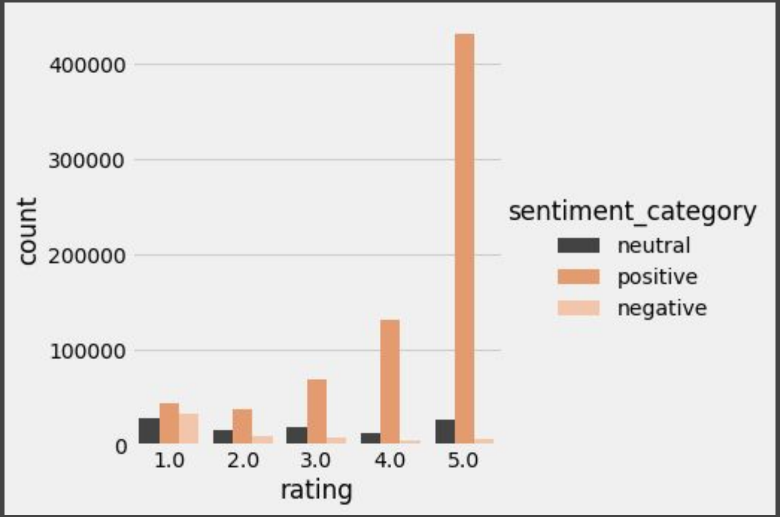
STEP 5

PERFORMANCE MEASURES: SENTIMENT SCORES

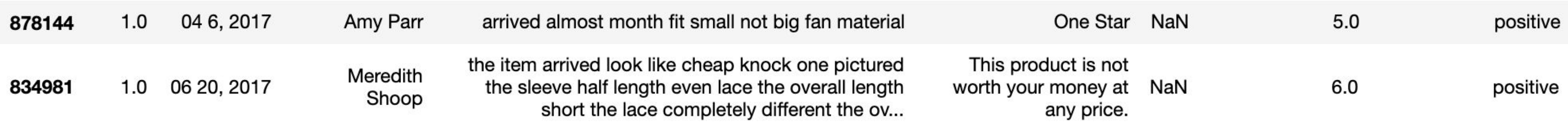
LESS INFLATED / NORMAL DISTRIBUTION



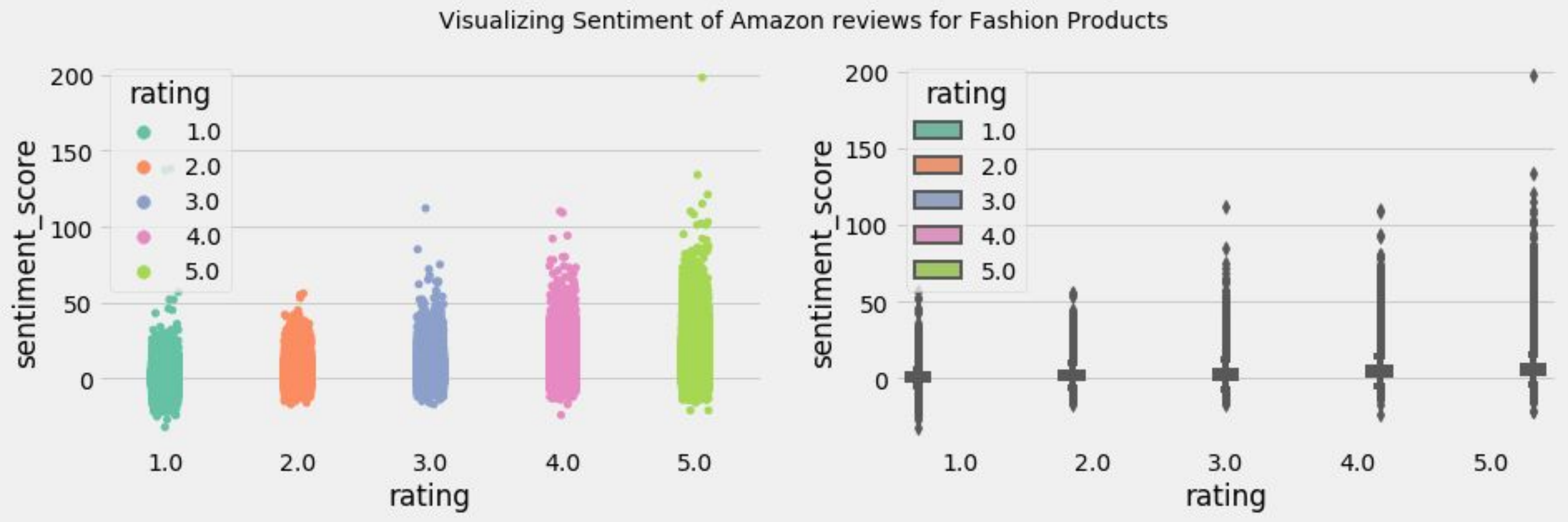
BUT NOT ALWAYS ACCURATE



PERFORMANCE MEASURES: SENTIMENT SCORES



878144	1.0	04 6, 2017	Amy Parr	arrived almost month fit small not big fan material	One Star	NaN	5.0	positive
834981	1.0	06 20, 2017	Meredith Shoop	the item arrived look like cheap knock one pictured the sleeve half length even lace the overall length short the lace completely different the ov...	This product is not worth your money at any price.	NaN	6.0	positive



PERFORMANCE MEASURES: ML CLASSIFICATION

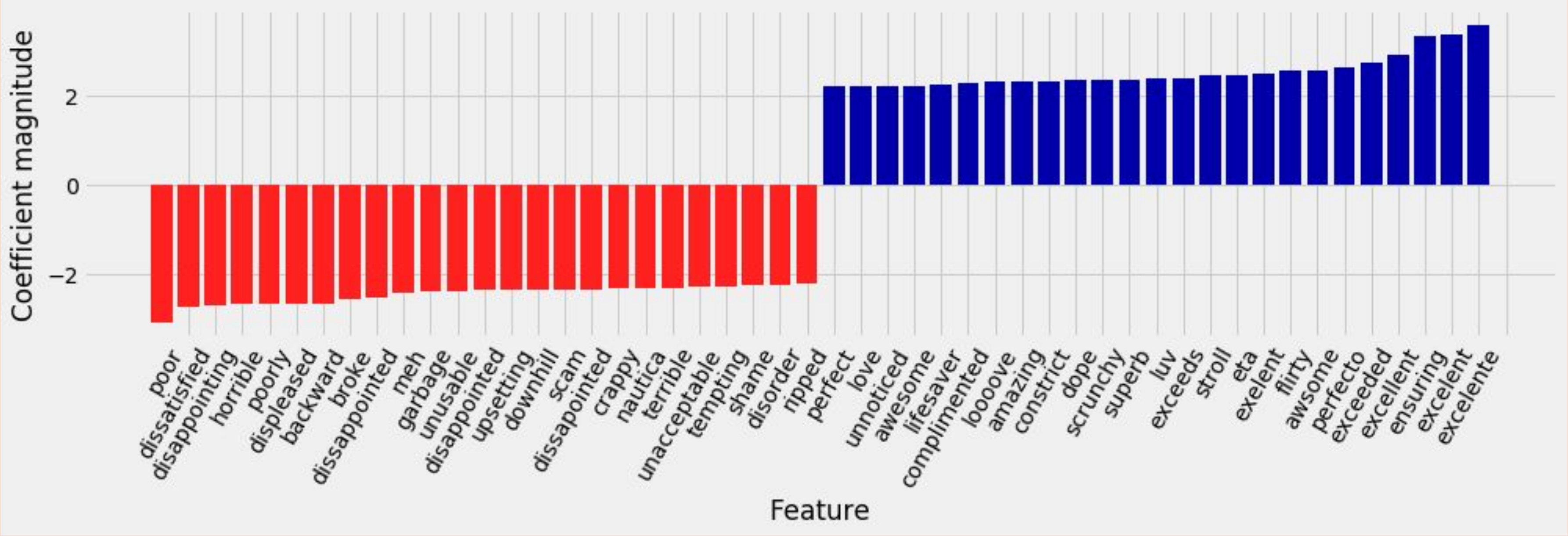
LOGISTIC REGRESSION MODEL TO PREDICT IF A PRODUCT IS RECOMMENDED OR NOT

- ❑ Logistic Regression Model (C = 1)
- ❑ Mean Cross-Validation accuracy: 0.88

	Precision	Recall	F1 Score
Not Recommended	0.84	0.75	0.79
Recommended	0.90	0.94	0.92
Accuracy			0.88

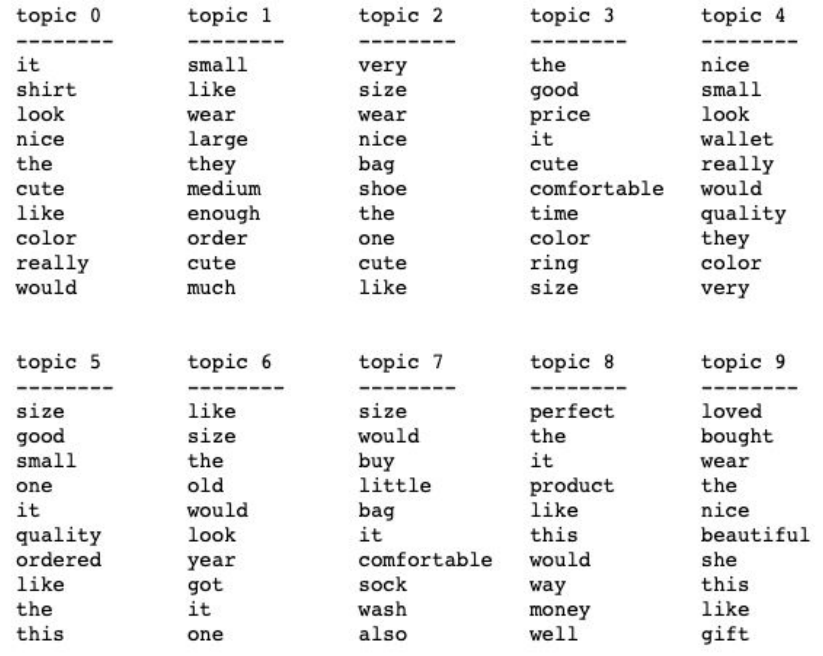
PERFORMANCE MEASURES: ML CLASSIFICATION

COEFFICIENTS OF FEATURES (UNIGRAMS AND BIGRAMS)



PERFORMANCE MEASURES: LDA

TOPIC MODELLING WITH LATENT DIRICHLET ALLOCATION



```
topic 0       topic 1       topic 2       topic 3       topic 4
--------      --------      --------      --------      --------
it            small         very          the           nice
shirt         like          size          good          small
look          wear          wear          price         look
nice          large         nice          it            wallet
the           they          bag           cute          really
cute          medium        shoe          comfortable   would
like          enough        the           time          quality
color         order         one           color         they
really        cute          cute          ring          color
would         much          like          size          very


topic 5       topic 6       topic 7       topic 8       topic 9
--------      --------      --------      --------      --------
size          like          size          perfect       loved
good          size          would         the           bought
small         the           buy           it            wear
one           old           little        product       the
it            would         bag           like          nice
quality       look          it            this          beautiful
ordered       year          comfortable   would         she
like          got           sock          way           this
the           it            wash          money         like
this          one           also          well          gift
```

- 10 Components
- Did not yield very distinctive groups
- Might have to try with more components

BUSINESS CASE

A plug-in analytical tool for E-Commerce platforms to vet products based on reviews and ratings from customers