

Topic Modelling and Reviews Mining

DSI-14 Capstone Project

TABLE OF CONTENTS

01 INTRODUCTION

02 EXPLORATORY DATA ANALYSIS

03 MODELLING

04 DEPLOYMENT

05 CONCLUSION

01

INTRODUCTION

EVERY MINUTE...



510,000

comments



16,000,000

text messages



156,000,000

emails



HIGHLY UNSTRUCTURED

AMBIGUOUS

COMPLEX

A central white vertical line is positioned on the left side of the slide, extending from the top to the bottom. A horizontal white line extends from the bottom of this vertical line to the right edge of the slide.

PROBLEM STATEMENT

1. Automatically segregates reviews by topics
2. Extracting product features and opinions from reviews

BUSINESS STAKEHOLDERS

E-commerce companies

It can also be fine-tuned and used by any company that receives a huge volume of text data

SCOPE

Topic Modelling and Reviews
Mining on Amazon Cell
Phones Reviews

USE CASE

Improve user interface and platform experience by:

- segregate reviews into topics
- summarise long reviews into just the main points/features and their respective sentiments.

SCOPE

APPROACH

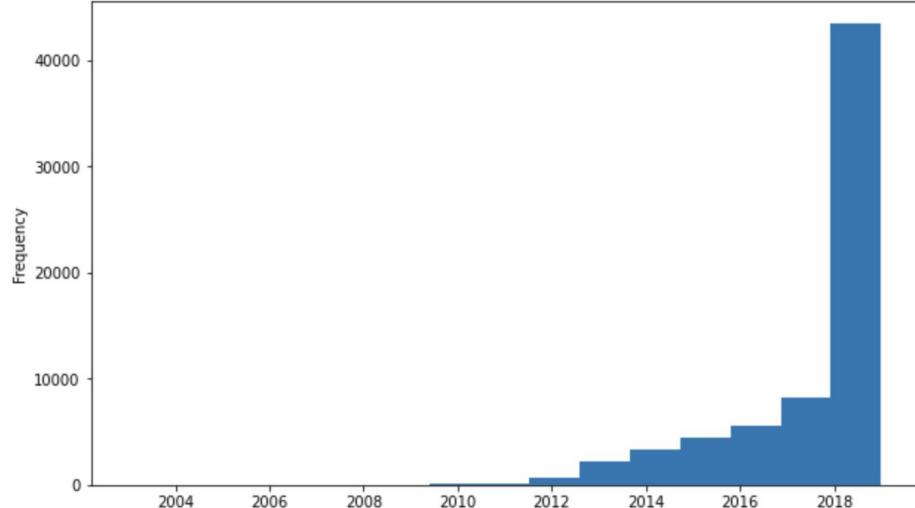
- Models trained based on 22,000 Amazon Cell Phone Reviews
- Unsupervised Learning using **Latent Dirichlet allocation(LDA) to do topic modelling**
- Features Extractions and Sentiments Predictions using **Logistic Regression Classifier, Valence Aware Dictionary and SEntiment Reasoner (VADER) and Bidirectional Encoder Representations from Transformers (BERT)**

02

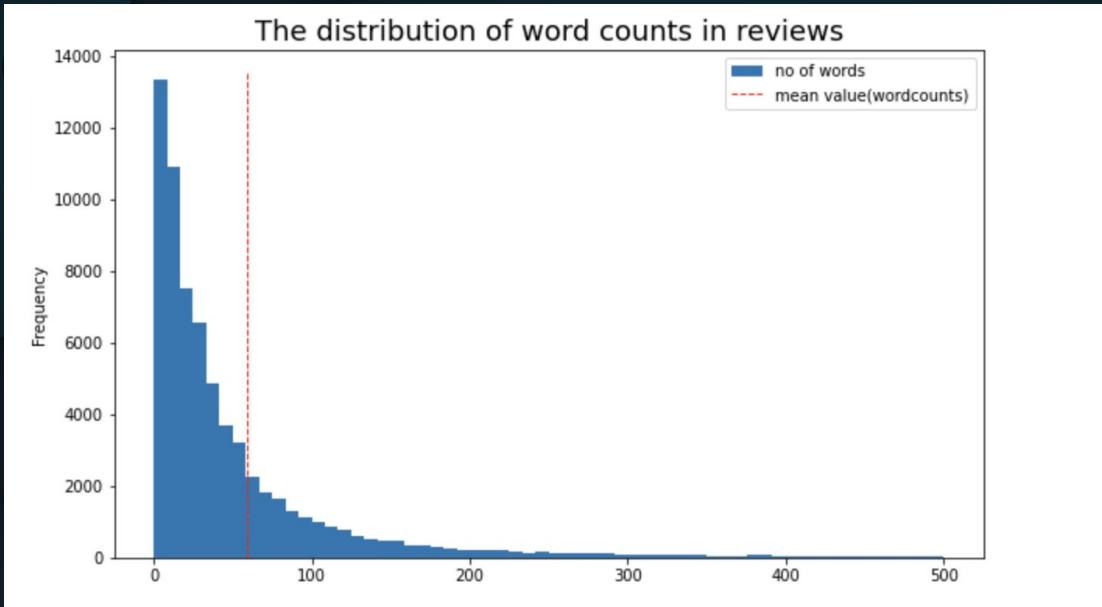
EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

Distribution of reviews by year



EXPLORATORY DATA ANALYSIS



03

MODELLING

Unsupervised Topic Modelling

SIM CARD

need wifi
carrier
card
sim
set issue
network datum **work**

SELLER

refurbish thank
work
condition brand
well purchase
seller price
arrive

SCREEN AND CAMERA

time
take **camera**
video app
thing feel case
screen
display

PRICE AND QUALITY

money
quality
camera value
feature **price**
love
picture
note better

PROBLEMS

screen
issue warranty time
work month
problem stop week
update

BATTERY

screen
battery
charge port hour
day charger scratch
life power

03

MODELLING

Features Extractions and
Sentiment Predictions



ACCURACY

	VADER	LOGISTIC REGRESSION	BERT
ACCURACY SCORE	70.7%	71.4%	86.3%
DETECT NEGATION?	SOME	NO	YES
DETECT CONTEXT?	NO	NO	YES

NOTE:

Logistic Regression was trained using overall rating

BERT was fine-tuned using overall rating

Overall rating was converted to 3 classes: 1-2: negative, 3: neutral, 4-5: positive

LET'S COMPARE THE PREDICTIONS OF THE 3 METHODS

Example:

I don't like the screen. It is different from the description. Please don't buy it.

VADER rating: [(5, 'screen')]

Logistic Regression rating: [(5, 'screen')]

BERT rating: [(1.0, 'screen')]

Amazon example:

Eh wouldn't buy again One - It comes in a weird box Two it had more scuffs and scratches than I'd like for the price 3 - I had to return it because of how over all lame it came and how it showed up/ battery was pretty warn/scratches/ and I didn't get the awesome feeling of unboxing it..

VADER rating: [(5, 'battery')]

Logistic Regression rating: [(5, 'battery')]

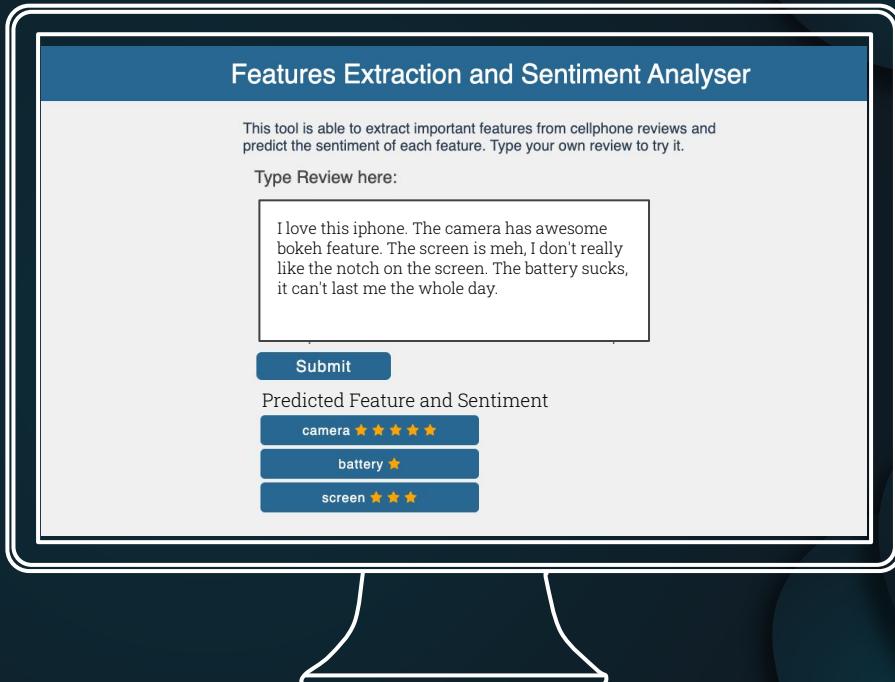
BERT rating: [(1.0, 'battery')]

04

DEPLOYMENT

DEPLOYED TO SHOWCASE HOW THE MODEL WORKS

Click here for live [DEMO!](#)



PRACTICAL USE CASES

AUTOMATICALLY GENERATES AGGREGATED OVERALL RATING FOR EACH FEATURE WITHIN A LISTING



SEGREGATES REVIEWS BY TOPICS

Read reviews that talk about..

problems

seller

battery

problems

charger and battery

price

problems

SUMMARISE SUPER LONG REVIEWS
INTO IMPORTANT PRODUCT
FEATURES AND RATINGS FOR EVERY
FEATURE

One of the few phones worth it's crazy price SMARTPHONE INFO I've been buying the latest and greatest smartphones since the early HTC days and have owned just about every significant Samsung smartphone release since then. Every year I will stray over to other manufacturers to see what the big and small players have to offer. Most recently I've been getting back to the fold, right back to my Samsung phone as the daily driver. Which is usually a Note. Now that Samsung has released another Note series for an increasingly high price, let's see how it fares. In particular, since the Note 9 is the Note 8's replacement. Price: Excellent build and design. The high-resolution OLED display somehow looks better than before + Superfast performance; zero lag + Much improved speakers that you can "feel" + The bezels continue to disappear. Camera: Excellent wide-angle lens. The high-resolution camera still work great + Lots of internal storage + expandable storage + Generous battery that truly lasts all-day + Fast wired and wireless charging for those who don't mind the heat. CON: The screen is still a bit of a pain to evolve + 3 Camera lenses for any scenario + Great picture & video quality with some useful features CONS: - Hole punch camera is distracting - The in-screen fingerprint reader takes a few steps back from the physical sensor - No headphone jack - Price keeps going up OEMOS still costs more than the Galaxy Note 9, the Galaxy Note 10+ feels much more substantial than I

Submit

Predicted Feature and Sentiment:

battery ★★★★☆

screen ★★★★☆

charger ★★★★☆

05

CONCLUSION

SENTIMENT PREDICTIONS

Inaccuracy due to the way it was fine-tuned.
Example, the word "okay" or "fine" is often predicted as 5 stars

SENTIMENT PREDICTIONS

Unable to detect sarcasm

FEATURE EXTRactions

Unable to identify synonyms and typos

LIMITATIONS

FUTURE STEPS

SENTIMENT PREDICTIONS

Inaccuracy due to the way it was fine-tuned (3 classes). Example, the word "okay" or "fine" is often predicted as 5 stars

Fine-tune it with 5 classes labels instead

Train the model with more neutral comments

SENTIMENT PREDICTIONS

Unable to detect sarcasm

Needs further research on this

FEATURE EXTRACTIONS

Unable to identify synonyms and typos

Generate more synonyms

Employ module that is able to detect typo eg. Textblob

LIMITATIONS

THANKS!

Elise Nerissa
DSI-14 Capstone Project