

## ABOUT OUR DATASET



aggle



9560 review



7 column

# Snapchat Reviews

To  
wi  
cu  
  
Ap  
dif  
an

## **TABLE OF CONTENTS**

- Introduction
- Data preprocessing
- EDA
- Topic Modeling
- Models
- Conclusion



## Introduction

Topic modeling with reviews  
will help Snapchat to improve  
customer experience.

Applying sentiment analysis to  
differentiate between negative,  
and positive reviews.

ing

## ABOUT OUR DATASET



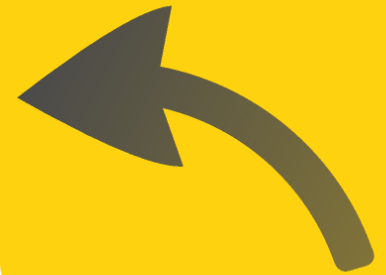
from Kaggle



9560 review



7 column



# Data Preprocessing

● Data Cleaning

● REMOVE NUMBERS, CAPITAL LETTERS,  
PUNCTUATION, AND SPELL CHECKING

● REMOVING STOP WORDS

● TOKENIZING AND LEMMATIZING

IDF	NB2-TFIDF
-----	-----------

746	0.693
-----	-------

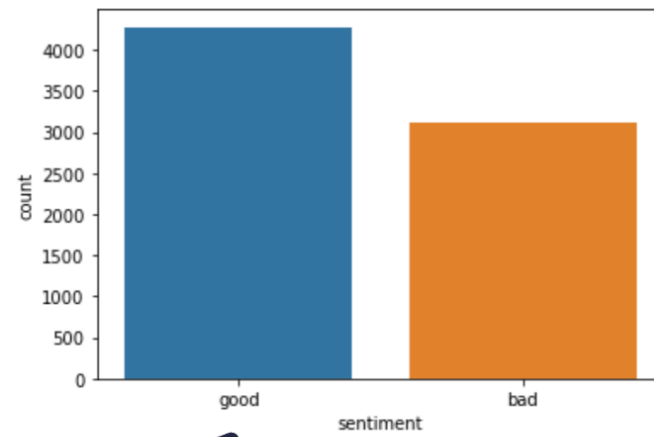
722	0.658
-----	-------

.917	0.990
------	-------

808	0.791
-----	-------

# EDA

Reviews







## NLP negative reviews



# Topic Modiling

## NMF Topicing

### Positive Reviews

Topic 0 ( **Service** )

like, good, see, make, change

Topic 1 ( **Filter** )

friend, love, filter, great, fun

Topic 2 ( **Story** )

story, post, video, private, posted

### Negative Reviews

Topic 0 ( **Locked Account** )

account, locked, get, back, email, support

Topic 1 ( **Update App** )

update, back, chat, map, change, look

Topic 2 ( **Camera Quality** )

camera, picture, quality, photo

## Corecx Topicing

(Problem) :fix,tried,working,wifi,try,problem,bug

(App): dark,mode,friend,fun,family,communicate

(Feature):feature,like,change,add, love,better,idea



## Topic Modiling

NMF Topicing

# Models

	LogReg1	LogReg2	NB1	NB2	LR1-TFIDF	LR2-TFIDF	NB1-TFIDF	NB2-TFIDF
Accuracy	0.820	0.828	0.768	0.693	0.811	0.803	0.746	0.693
Precision	0.837	0.835	0.809	0.658	0.831	0.824	0.722	0.658
Recall	0.860	0.880	0.789	0.990	0.849	0.843	0.917	0.990
F1 Score	0.848	0.857	0.799	0.791	0.840	0.833	0.808	0.791

## **Conclusion**

Overall, the second Logistic Regression model (using n\_bigram and counts) seems to best classify positive and negative reviews.

