KICKSTARTER Project Analytics

A NLP Approach to Increase the Chance of Building a Successful Crowdfunding Project

Dao Yong | Hanyu | Jing Jie | Jocille | Raymond | Soo Yeon

Date: 6 April 2021

PROJECT OVERVIEW – BACKGROUND

About Kickstarter

- Global **crowdfunding** platform
- Fund **All-or-Nothing** model
- Collect **5%** of the successful funds raised.

Business Problem

- **38.61%** projects fail, according to Kickstarter
- **$229,775,700 USD** opportunity cost

Stakeholders

- New aspiring **project starters** -> successfully raise funds
- **Kickstarter management** -> collect fees from successful projects

BUSINESS USE CASES

Fundraisers will be provided with the **sentiments of backers** towards project categories of their choice.

Fundraisers will be **recommended** with related and similar projects.

OUR GOAL

To use text mining to help future Kickstarter campaigns to increase their success rate

DATASET AND TOOLS

Octoparse pandas python jupyter

Web Robots Kickstarter Datasets

- Data is crawled and published once a month
- Includes short project description, location, amount raised, goal and project status

Crawled Data From Kickstarter

- Using Octoparse to crawl project comments from Kickstarter

Our Final Datasets

- Combined through project names
- 605 Projects in final datasets
- 25158 comments

SOLUTION OVERVIEW (1)

1. Topic Modelling and Classification

- Gather description & title
- To identify new key trending topics
- Optimally recommend categories for fundraisers

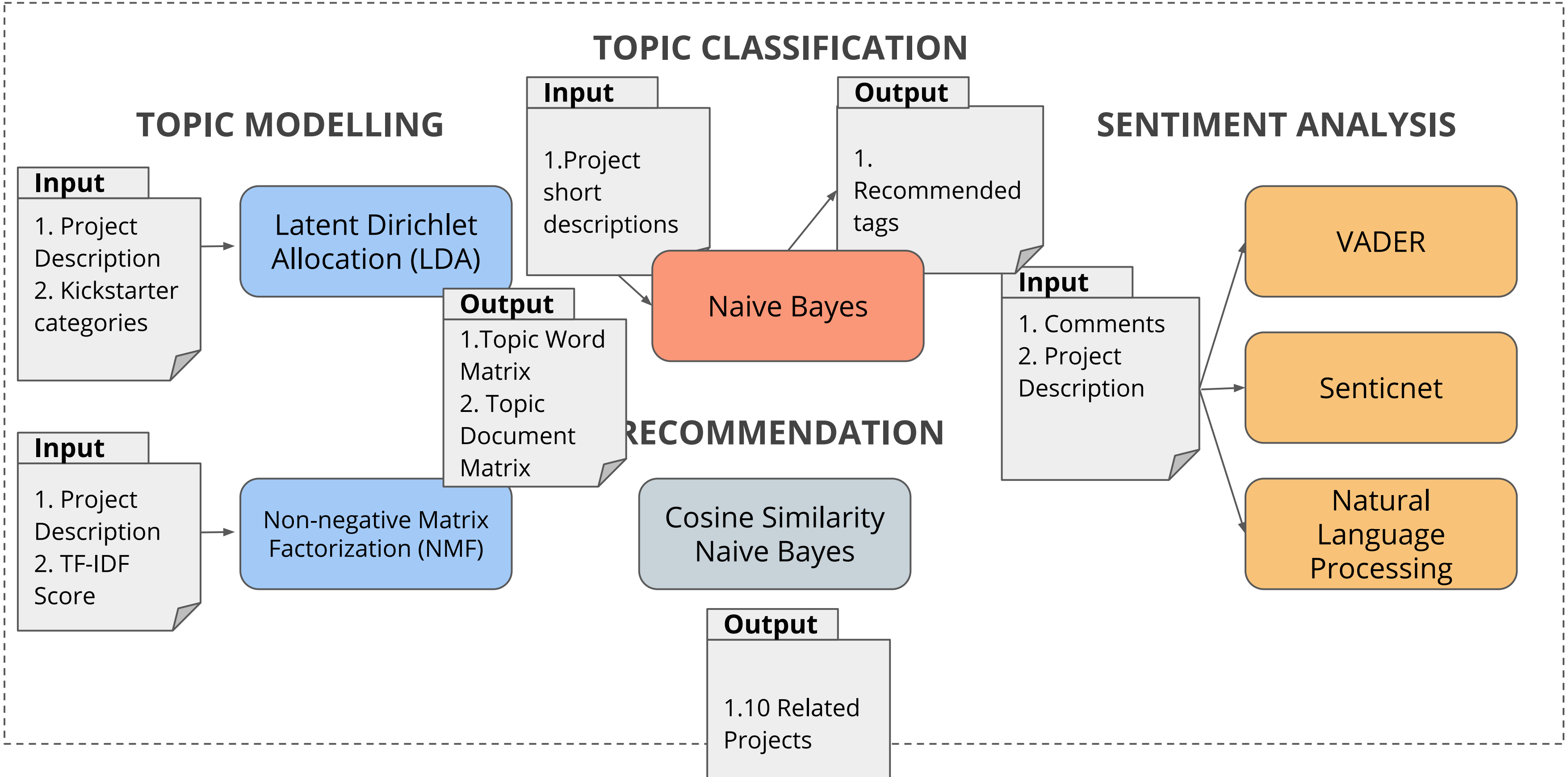
2. Sentiment Analysis

- To find out general attitude towards a campaign from comments
- To draw business insights from the different project categories in a top down manner

3. Content Based Recommender

- To **recommend** similar and related projects to fundraisers assist their research

SOLUTION OVERVIEW (2)



TASK 1 - TOPIC MODELLING

PURPOSE: To identify trending topics and keywords

APPROACH: LDA

- Unsupervised classification of documents

RESULTS:

- Coherence Score: 0.38
  Word intrusion: 0.27

APPROACH: NMF

- Application of TF-IDF vector to document and model the topics

RESULTS:

- Coherence Score: 0.42
- Word intrusion: 0.40

CHALLENGES:
Choosing the number of topics and finding the right hyperparameters

CHALLENGES:
Choosing the right number of topics and setting TF-IDF to work with NMF

TASK 2 - PROJECT TOPIC CLASSIFICATION

PURPOSE: Optimally recommend categories for fundraisers

APPROACH:

Naive Bayes Classification Technique

Train: Trained the model with 100 documents for each category

Validate: Validated with 20% of the volume of the trained model (Unique data)

Test: Tested with 10% of the volume of the trained model (Unique data)

RESULT:

99.29% accuracy with validation

1.42% error with testing (Prediction)

CHALLENGES:

- Non-english descriptions
- Wrong outputs in some cases

TASK 3 - SENTIMENT ANALYSIS

PURPOSE: To find out the general attitude towards different groups of projects.

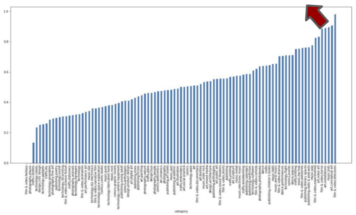
APPROACH:

1. **Comparison** of sentiment results on comments and projects descriptions from VADER and Senticnet.
   a. Word level - from model
   b. Sentence level - from model or average score
   c. Project level - average score
2. Natural Language Processing technique to find out the **most commonly used words** in different projects across locations, categories, status etc.
   a. Ngram

RESULT:

- **Visualisation** & **Insights**!

film & video,television



Sentiment Scores by Category

Unigram for All Comments

- **VADER** performs better!

CHALLENGES:

- Non-english comments and misspelling.
- Limitation on VADER.

BONUS TASK - CONTENT BASED RECOMMENDER

PURPOSE: To find similar & related projects

APPROACH:

- Base on topic modeling, group the projects into different groups
- Use cosine similarity to find related projects
- Comparison between normal cosine similarity and topic based cosine similarity

```
Precision@k = (# of recommended items @k that are relevant) / (# of recommended items @k)
```

RESULT:

Precision @ K	LDA + Cosine	0.636
	NMF + Cosine	0.472
	Cosine	0.916

CHALLENGES:

- Using the short description/category might not be enough
- Cosine similarity only matches the exact words
- Lexical ambiguity

DEMO

	Findings	Insights	Analysis
1	**NMF:** Topic 9: film and horror appear frequently	• Most kickstarter project ask for funds in areas of **films** and **horror**. • Trending topic for people asking for fundings	Allows Kickstarters to **gauge** funding based on the topics
2	**Sentiment Analysis:** • Technology-related projects receive most comments. • Art/illustration raise the most amount.	• Imbalance of public attention and who gets the fundings. • **Technology-related projects** have the potentials to grow in the future.	Many people are looking forward to advanced technology, but taken aback by the chance of failure.
3	**Sentiment Analysis:** Successful projects in Hong Kong and Germany do not receive as much comments	• People from Hong Kong and Germany are likely to be non-native English speaker • Project descriptions will not be translated.	**Auto-translation** of the project description can be implemented to increase the project exposure to non-english speaking countries.
4	**Sentiment Analysis:** • Most successful project category is Art/Illustration. • Card games such as Tarot and fictional characters are popular.	Audience and backers tend to look for products that are not easily found on mainstream markets on Kickstarters.	Fundraisers and Kickstarters should pay more attention to projects that are original, and in the field of **horror/weird fiction** and **fortune-telling**.

WHAT WORKS WELL AND WHAT DOES NOT

GOOD

- **NMF** for Topic Modelling
- **Cosine Similarity** for Recommender
- **VADER** is able to find the most positive and negative comments

BAD

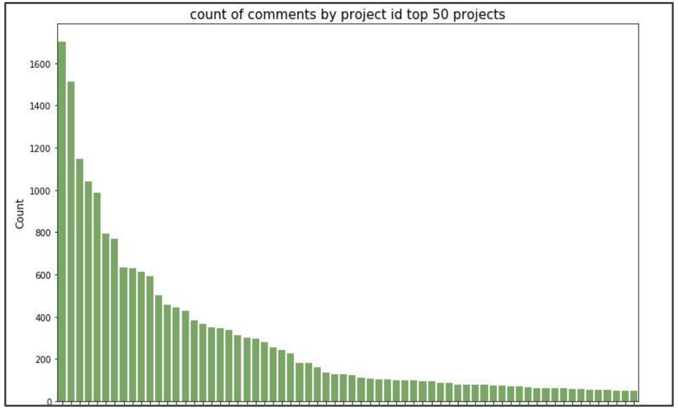
- Recommender with **topic modelling** is still not optimal
- **Senticnet** did not perform well for our sentiment analysis

LIMITATIONS

Datasets

- Our dataset is a percentage of all the Kickstart projects
- Our approach is **scalable** once full dataset is obtained.

Unbalanced number of comments



Non-english and multilingual comments; Misspelling

- Removed all the non-English comments using Langid model.
- Some sentiments are not considered.

FUTURE WORKS

- Work on non-english comments for sentiment analysis
- Further tuning of hyperparameters
- Lexical ambiguity and dictionary for unknown words

Beyond the class?

- **NMF**
  - Model used for document clustering
  - Non negative elements needed
  - Multiplication of 2 matrix (W & H) to derive the combine matrix (V)
- **Sentiment Analysis using VADER and Senticnet**
  - VADER - Valence Aware Dictionary for Sentiment Reasoning. It uses a combination of **sentiment lexicon** (phrases, emoticons, acronyms) and **grammatical rules**.
  - Senticnet - Concept-level sentiment analysis. It focuses on **semantic analysis** through the use of **semantic networks**.

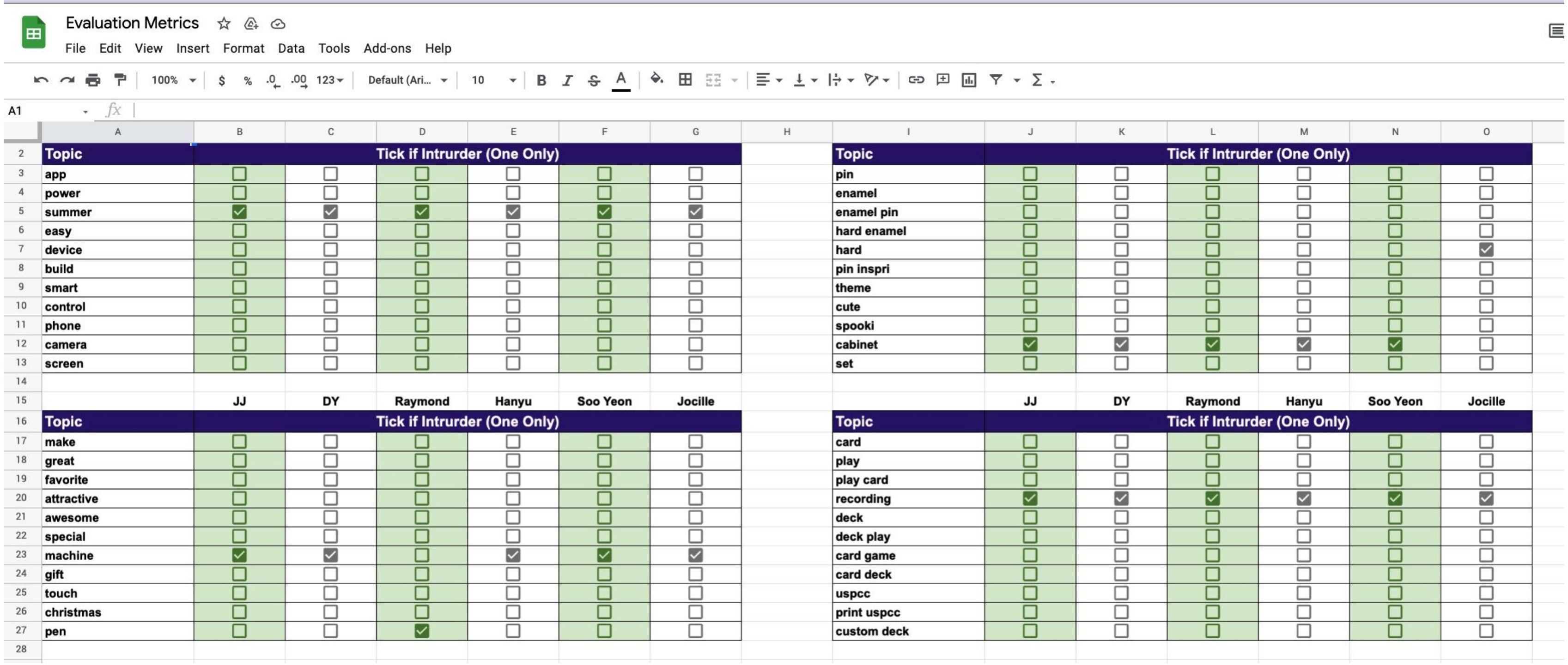
Thank You!

QnA

Why VADER Performs Better?

Model	Most Positive Comments	Most Negative Comments
VADER	Congrats Sam, it is a great idea and I look forward to watching and if given the chance to participate I'll be there. I'm a truckin biker of the old school I make more time to ride these days as I have been a successful owner-operator in the trucking business. As a cold war vet from '69 to '71 I relish experiencing our great countries history and meeting the folks who hold secrets and stories not readily available to most. Best wishes and best of luck.	Fuck it, why not. But you better hit someone with that stick.
Senticnet	Nicely presented Kickstarter - Good Luck! @ jerice50: Thank you very much.	Congratulations on funding!

How We Calculate Word Intrusion Score



Topic	Tick if Intrurder (One Only)					
app	☐	☐	☐	☐	☐	☐
power	☐	☐	☐	☐	☐	☐
summer	☑	☑	☑	☑	☑	☑
easy	☐	☐	☐	☐	☐	☐
device	☐	☐	☐	☐	☐	☐
build	☐	☐	☐	☐	☐	☐
smart	☐	☐	☐	☐	☐	☐
control	☐	☐	☐	☐	☐	☐
phone	☐	☐	☐	☐	☐	☐
camera	☐	☐	☐	☐	☐	☐
screen	☐	☐	☐	☐	☐	☐

Topic	Tick if Intrurder (One Only)					
pin	☐	☐	☐	☐	☐	☐
enamel	☐	☐	☐	☐	☐	☐
enamel pin	☐	☐	☐	☐	☐	☐
hard enamel	☐	☐	☐	☐	☐	☐
hard	☐	☐	☐	☐	☐	☑
pin inspri	☐	☐	☐	☐	☐	☐
theme	☐	☐	☐	☐	☐	☐
cute	☐	☐	☐	☐	☐	☐
spooki	☐	☐	☐	☐	☐	☐
cabinet	☑	☑	☑	☑	☑	☐
set	☐	☐	☐	☐	☐	☐

	JJ	DY	Raymond	Hanyu	Soo Yeon	Jocille
Topic	Tick if Intrurder (One Only)					
make	☐	☐	☐	☐	☐	☐
great	☐	☐	☐	☐	☐	☐
favorite	☐	☐	☐	☐	☐	☐
attractive	☐	☐	☐	☐	☐	☐
awesome	☐	☐	☐	☐	☐	☐
special	☐	☐	☐	☐	☐	☐
machine	☑	☑	☐	☑	☑	☑
gift	☐	☐	☐	☐	☐	☐
touch	☐	☐	☐	☐	☐	☐
christmas	☐	☐	☐	☐	☐	☐
pen	☐	☐	☑	☐	☐	☐

	JJ	DY	Raymond	Hanyu	Soo Yeon	Jocille
Topic	Tick if Intrurder (One Only)					
card	☐	☐	☐	☐	☐	☐
play	☐	☐	☐	☐	☐	☐
play card	☐	☐	☐	☐	☐	☐
recording	☑	☑	☑	☑	☑	☑
deck	☐	☐	☐	☐	☐	☐
deck play	☐	☐	☐	☐	☐	☐
card game	☐	☐	☐	☐	☐	☐
card deck	☐	☐	☐	☐	☐	☐
uspcc	☐	☐	☐	☐	☐	☐
print uspcc	☐	☐	☐	☐	☐	☐
custom deck	☐	☐	☐	☐	☐	☐

Full SOLUTION OVERVIEW