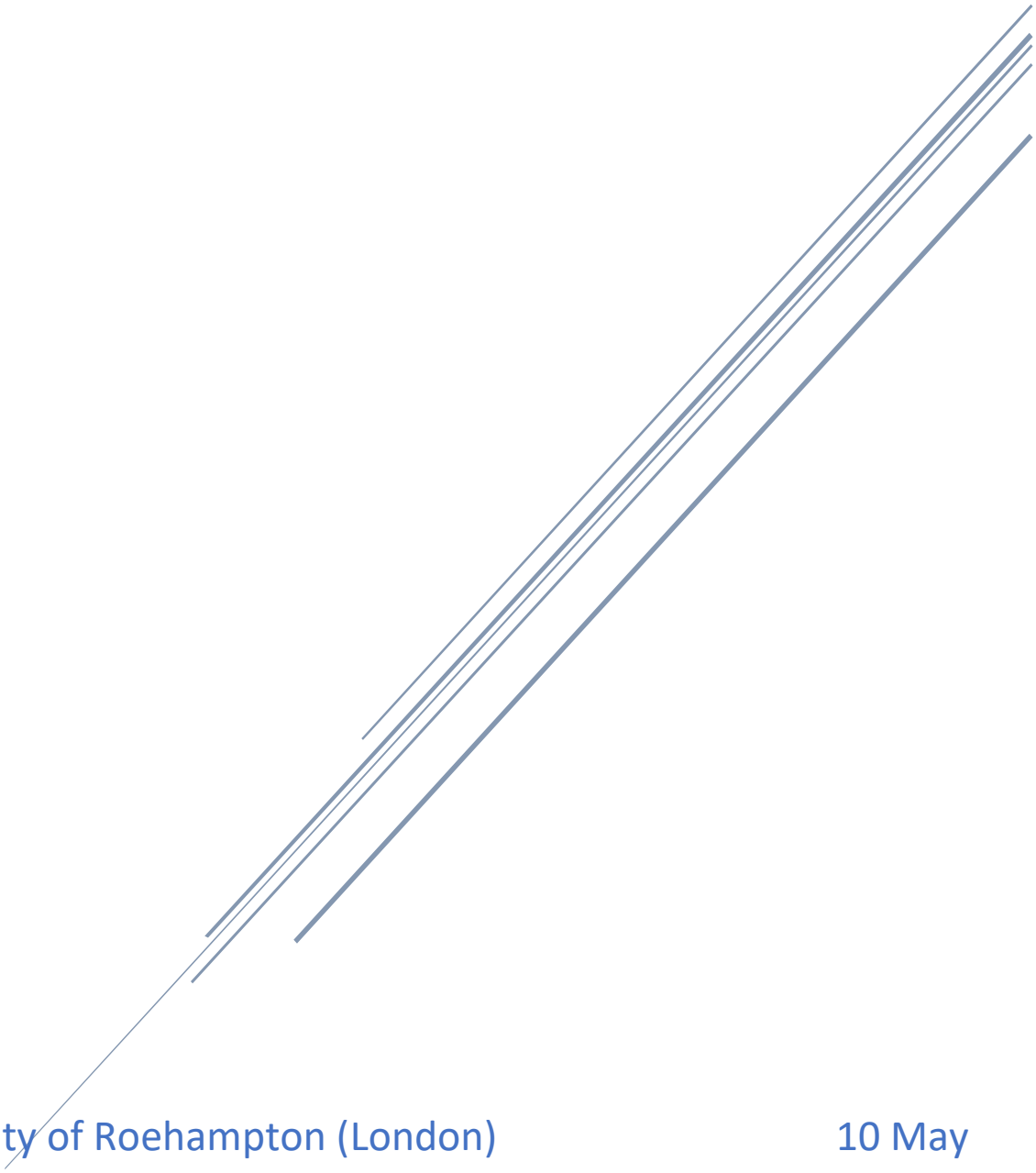
BAYES THEOREM

Artificial Intelligence



University of Roehampton (London)

10 May 2022

Tony, YIT19488399

Tony, YIT19488399

Artificial Intelligence with Dr Gu & Dr Arturo

07 May 2022

INTRODUCTION

From academic journals and magazines, one interesting Artificial Intelligence (AI) application that I found is about calculating the sentiment from reviews from people who watch videos and comment directly on the video while watching it. This is known as Danmaku videos and is popular in China and Japan. Symbols, Chinese letters, English letters and emojis are used in this case.

Naïve Bayes with Sentiment Dictionary is the best algorithm to analyse sentiment, but other algorithms were used like Support Vector Machine, Maximum Entropy, TextCNN, etc.

To clarify, this publication received an economical grant from China and the authors appreciate it.

BODY

Naïve Bayes theorem is applied to each word from reviews and classifies each review text into 7 human sentiments: like, happy, anger, sad, fear, disgust and surprise. Like and happy are defined as positive sentiment analysis while the other 5 are defined as negative sentiment analysis. Naïve Bayes classifies each text/sentence by predicting the probability to belong to a certain sentiment category based on the highest probability.

One of the challenges that experience the authors of this paper is collecting all the reviews because anyone can comment while watching the video. This can be a problem because of different ages, social statuses, levels of education, believes or ideas. This diversity makes this study not focused on or targeted to any specific group. Another challenge from the study is when there are too many comments in a short period. Collect all the comments in a short period of time can be challenging to collect all the information because of the hardware and phisical limitations. Polarity detection from the sentiment analysis plus popularity monitoring based on the number of reviews and the number of different users' reviews.

One of the ethical issues that I found, and the authors did not comment on in the study is how reviews are collected. I suppose that anyone who wants to write a review should be registered on the platform. This data is valuable and can be sold or reported to authorities with precise information about the author of the review. Another ethical issue that the authors commented on in this study is using the sentiment analysis for economical or business purposes to make precise marketing campaigns for example.

If I am successful in finding a good team and the economical grant is huge enough, I will do the same sentiment analysis but in real-time. And it could be done to any video in any language. Because this study shows high accuracy using the Naïve Bayes theory, I will not try to find a better algorithm. But could be a great idea to experiment with other algorithms. Authors use N-GramNB, N-GramSVM and TextCNN only. Another improvement will be observing the people that watch the video and applying a machine learning algorithm like image recognition to classify the sentiment of the whole video. Face expressions, movement of the eyes, and maybe the movements of the hands should be collected, analysed and finally classified. This study is more precise but a very different approach is

used even no Naïve Bayes theory. This study-investigation-research project will require signing a contract with users and is less respectful of the privacy of my users-volunteers.

Interesting screenshots from the paper can be found here:

FIGURE 1. Example of danmaku on a network video platform.

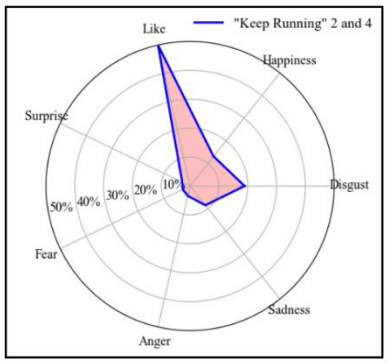


FIGURE 2. Radar map of danmaku sentiment categories.

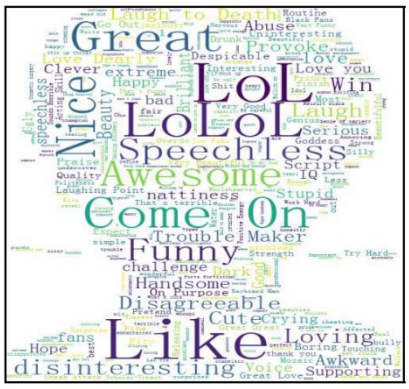
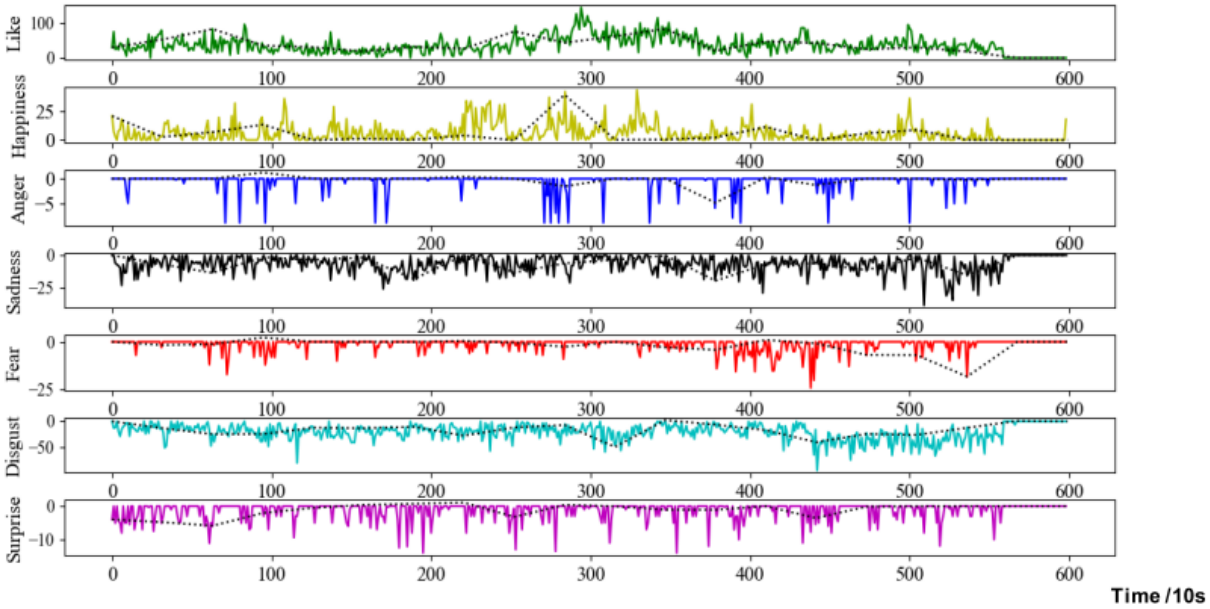
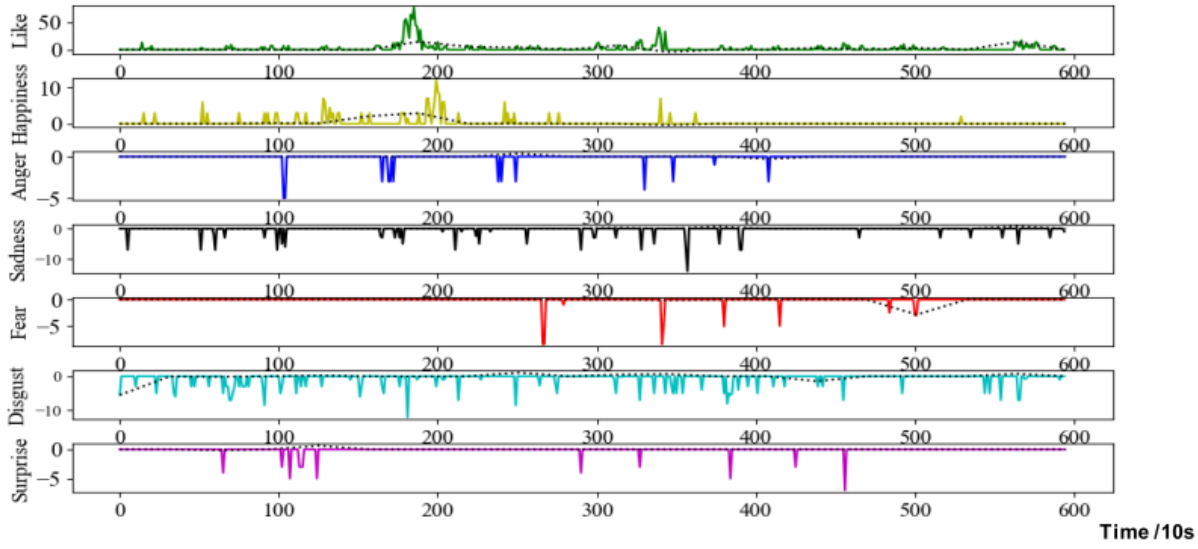


FIGURE 3. Word cloud of danmaku sentiment.



(a) “Keep Running” Season 2 Episode 10



(b) “Keep Running” Season 4 Episode 8

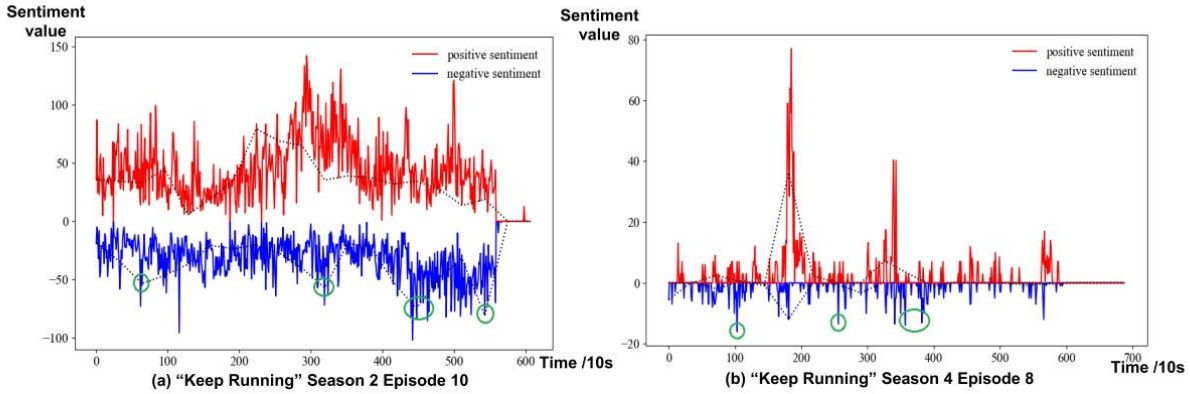


FIGURE 6. Time distribution of the two sentiment polarities of danmaku comments.

TABLE 2. Precision, recall, f1 and accuracy scores of sentiment classification algorithms.

Sentiment classification algorithms	Positive			Negative			
	Precision	Recall	F1	Precision	Recall	F1	Accuracy
SD-NB	0.867	0.783	0.823	0.889	0.936	0.912	0.882
N-gram-NB	0.872	0.359	0.509	0.722	0.979	0.831	0.750
SVM	0.687	0.518	0.591	0.753	0.862	0.804	0.735
TextCNN	0.661	0.770	0.712	0.850	0.767	0.806	0.768

CONCLUSION

The paper is difficult to understand and is large. In total 12 pages. I required 4 times to read it carefully. Many concepts and ideas on the same academic paper. And I made my notes-summaries. Therefore, I decided to do another essay about the same topic, but it is a founded economically conference. Only 2 times it was necessary to read it.

END essay 1

Tony, YIT19488399

Artificial Intelligence with Dr Gu & Dr Arturo

09 May 2022

INTRODUCTION

Indonesia has 260 million people and the National Health Insurance Administrator (BPJS Kesehatan) from 2014 has more than 80% of the population covered, launching and managing a Mobile JKN application. This academic paper analyse the sentiment of the reviews from Indonesian users on the Android application from Google Play Store. Positive and negative sentiments analysis allow to discover and understand user perception against service standards of Mobile JKN. Accuracy for the Naïve Bayes Classifier is 93% and shows more negative than positive reviews for the Mobile JKN application. The result is for future improvement and optimization of the application.

This paper is economically granted by Universitas Padjadjaran Indonesia and is classified as a conference.

BODY

Opinion reviews from the users of the JKN Mobile application are classified with Naïve Bayes Machine Learning Technology. This sentiment Classification type requires little

training data, it is fast and makes real-time predictions. Naïve Bayes Classifier discovers user perception against service quality of Mobile JKN.

The Naïve Bayes algorithm uses more than 20 thousand reviews. Exactly 13417 reviews are for training (6337 labelled as positive reviews and 7080 reviews labelled as negative sentiment). And 7139 reviews are for analysing the model after trained. It follows the classic 70%-30% ratio. Data is cleaned, uniformised and pre-processed in the Indonesian language. Changes to the original text are done. Case folding, filtering, tokenization, and stemming process have been carried out. The summary of the Naïve Bayes Classifier can be observed in the confusion matrix (see image screenshot from the original paper). The trained model created can predict and classify the sentiment of the Mobile JKN application reviews with very high accuracy, 93%. Only two predictive labels (-1 and +1 for negative and positive reviews respectively).

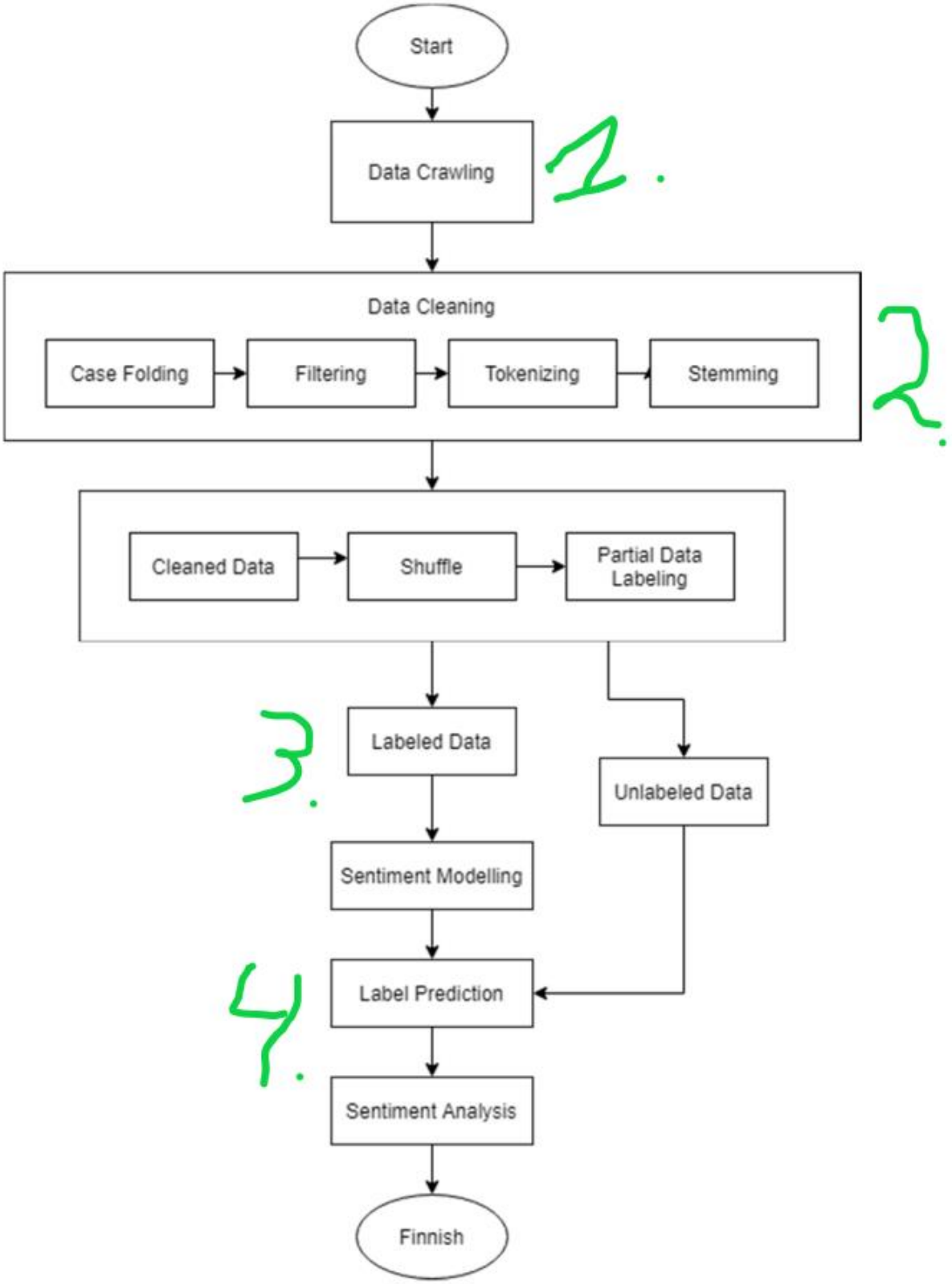
The prediction result using this model is 35% for positive reviews and 65% for negative reviews. Overall, users complain about features like updates frequency, and problems while logging and sing in whit phone number verification. While there is no precisse information about positive reviews.

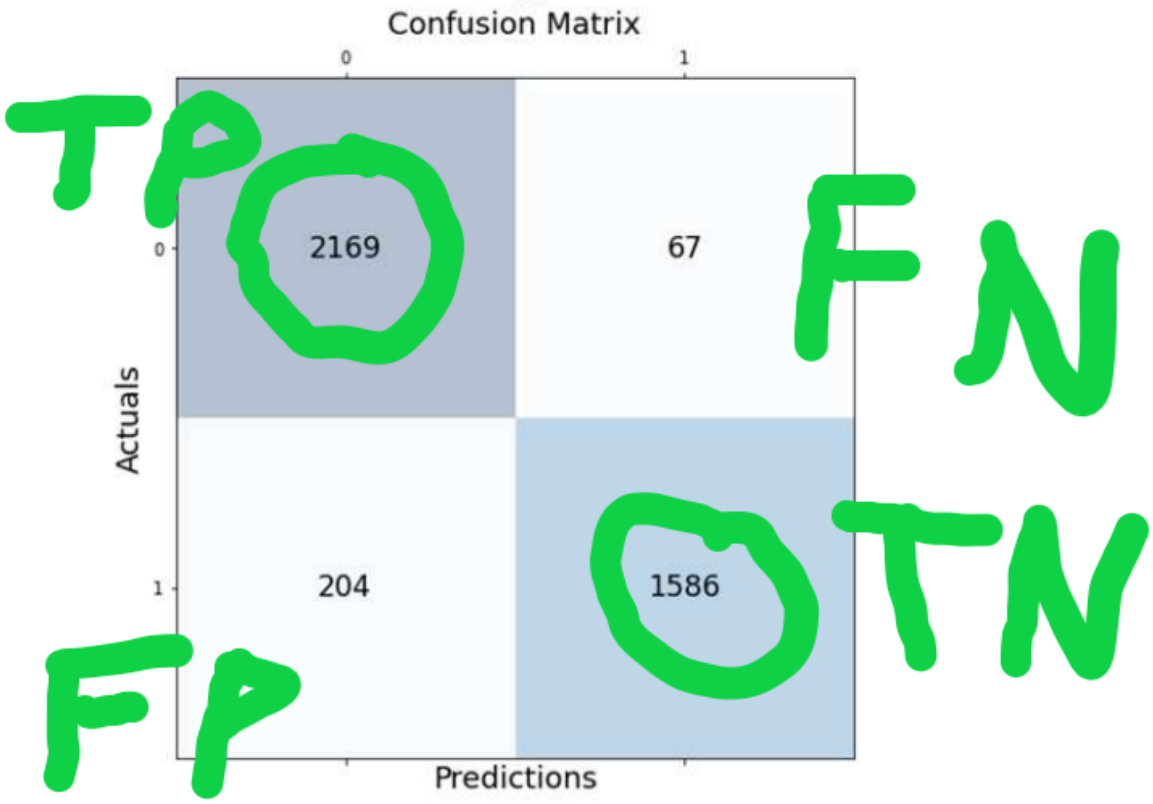
CONCLUSION

In summary, the whole study is accessible and understandable for an undergraduate student. Many similar projects can be found online about movie reviews, Amazon customer reviews or Twitter sentiment data. This is a supervised machine learning technique because data are labelled (- and + sentiment classification). Naïve Bayes algorithm applied

to text sentiment classification analyse the frequency of each word (token) in each label and predicts/classifies future text based on these words from the model created and trained.

Mathematics and statistics are necessary to better understand how this algorithm works.


Start
Data Crawling
Data Cleaning
Case Folding
Filtering
Tokenizing
Stemming
Cleaned Data
Shuffle
Partial Data Labeling
Labeled Data
Unlabeled Data
Sentiment Modelling
Label Prediction
Sentiment Analysis
Finnish



Confusion Matrix of classification model

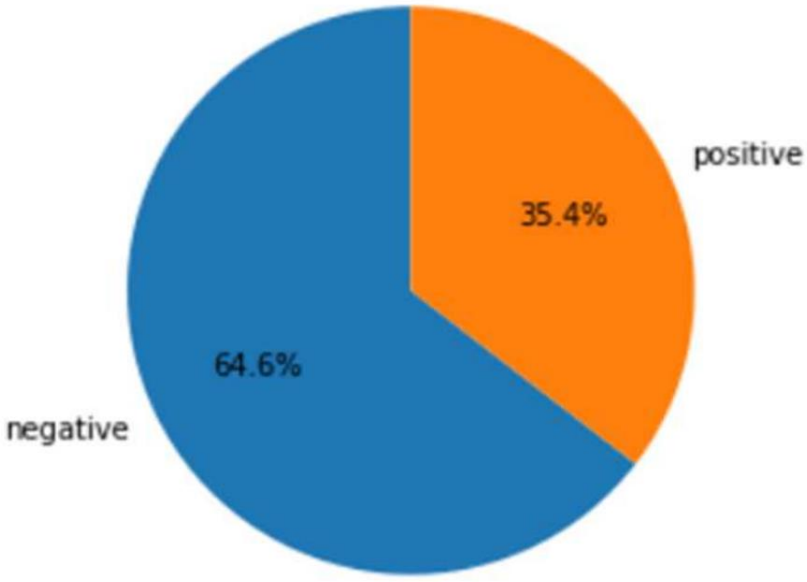
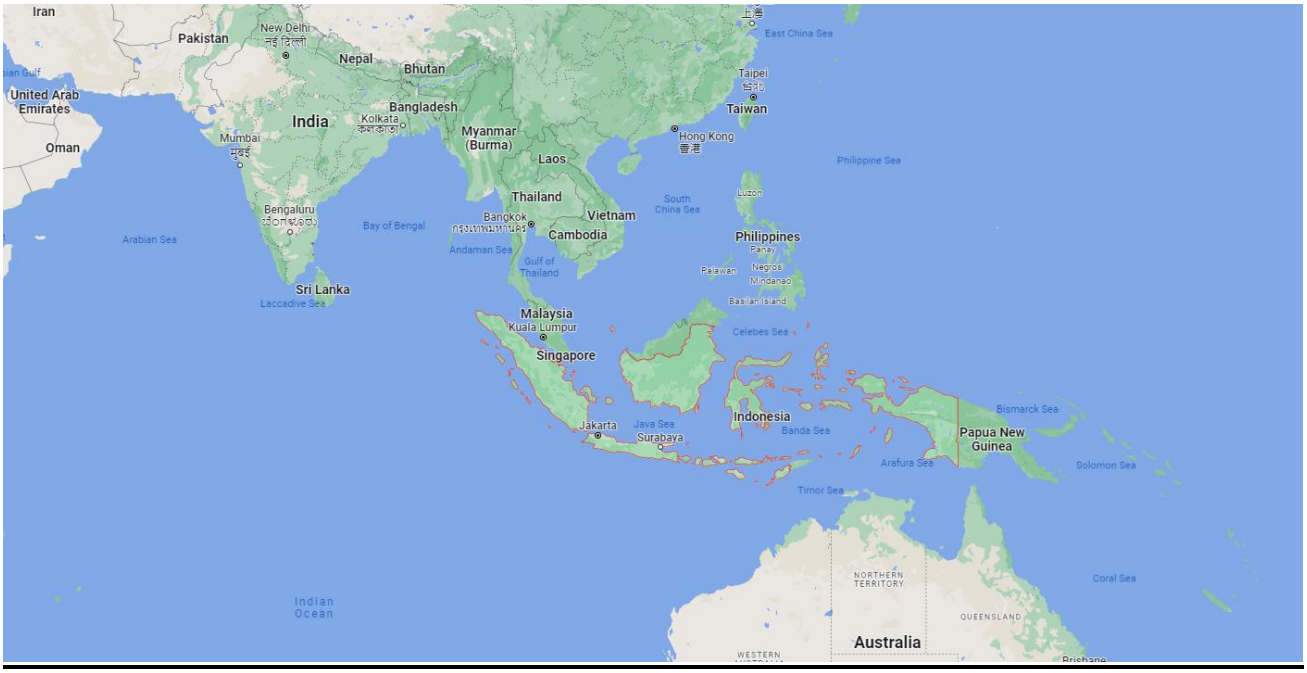


Fig. 3. Comparison of Sentiment Reviews for Mobile JKN Application Version 3 and above



END essay 2

REFERENCES

https://ieeexplore.ieee.org/document/9060892

https://ieeexplore.ieee.org/document/9689726/authors#authors