UNIVERSITE ABDELMALEK ESSAIDI
ECOLE NATIONAL DES SCIENCES APPLIQUES
D'AL HOCEIMA

Department: Mathematics and Computer Science

ANALYSIS OF POLITICAL SENTIMENTS USING YOUTUBE DATA: A CASE STUDY OF THE PALESTINIAN-ISRAELI CONFLICT

Made by:

ELGHARBAOUI ABDELGHAFOR

Supervised by:

Pr. AZIZ KHAMJANE

Academic Year: 2023-2024

ANALYSIS OF POLITICAL SENTIMENTS USING YOUTUBE DATA: A CASE STUDY OF THE PALESTINIAN-ISRAELI CONFLICT

Abdelghafor Elgharbaoui

Abstract

Due to people high activity on social media, a huge amount of data is generated each day, a big part of it is related to people discussions and interactions with events and trends, not only general information, but also political opinions and attitudes. Many research efforts used that potential of social media to analyse and predict the public opinion towards political events. Since people reaction around the globe about the current war that started on the 7th of October shows an unseen sympathy toward Palestinians, this work aim to measure the political attitude of people from different nations about the Palestinian-Israeli conflict after that event, using comments on the news shared by many famous YouTube news channels.

Keywords

Public opinion, YouTube comments, Political attitudes, Palestinian-Israeli conflict.

1. Introduction

Social media platforms have grown explosively over the past decade. People from all over the world have been using them extensively to express their views and discuss topics of interest. The large number of users, the variety of discussed topics and the massive volumes of posted content have made social media a rich source to understand and predict the population attitudes. Mining social media for political opinions may provide a faster and less expensive alternative to traditional polls.

In this paper, we propose an analytical study that uses a sample of YouTube comments data in different languages to measure and analyse the political public opinion in several countries around the world towards the Palestinian-Israeli conflict.

2. Related Works

Sentiment analysis is the major area of interest in the field in NLP. It is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities. And due to the huge amount of data related to people opinions and attitudes toward events such as political events and conflicts in social media, many researches focused on studying, analyzing and predicting the public opinion about conflicts between entities, and which pole is making political victories, since public opinion is the power that politicians obey, or to predict future actions, and take the appropriate decision.

For example, [1] this work presents an analytical study for measuring the political public opinion towards the Palestinian-Israeli conflict by using Twitter data. The study uses a novel data analysis model that leverages two levels of analysis: country-level analysis and individual-level analysis. The country-level analysis aims to explore the country's overall attitude towards Palestine by: 1) Identifying counties that generated the most topic-focused tweets, 2) Measuring the friendliness of each country towards Palestine. 3) Analysing the change of sentiment over time. The individual-level analysis aims to analyse data based on the activity and background of individuals. The attitudes of opinion leaders and ethnic groups were analyzed and discussed in light of countries' attitudes.

Another work, detect the bias of Palestine/Israel conflict in western media, [2] this research paper report a text mining experimental study, that’s have conducted on western media analysis to identify patterns in the press orientation and further in the media bias towards side to another. They have followed the text mining techniques and machine learning in an effort to detect the bias in news agencies. They have crawled news articles form seven major outlets in the western media. Then they have made preprocessing to convert them into useful structured form, building sentiment classifiers that be able to predict articles bias.

In general, we conclude that social media and media platforms proved to be an effective way to measure and study the public reaction associated with important events.

3. Overview of The Proposed Approach

The adopted approach consists of the following steps: data collection, data pre-processing, political sentiment analysis where we select the appropriate classifier for this task.

Afterwards, a case study that utilizes the proposed model to analyse the international public opinion towards the Palestinian-Israeli conflict is presented in detail. Data collection and pre-processing steps are described. The paper reports on the experiments conducted to compare several sentiment classifiers and to train and evaluate our own sentiment classifier.

4. Data Collection

Apify is a cloud platform for web scraping, browser automation, and data for AI. It includes many ready-made tools, we used an actor named “YouTube Comments Scraper” [3] to scrap comments of the videos shared by YouTube news channels, it provides 5$ for free, and get regenerated each month, the price of scraping 1000 comment is 2$. There is an official YouTube API, but it has a limitation of 100 comment for each call.

In total, 2606 comments were collected. These comments were posted by 1981 users during the period from Oct. 07 2023 to Dec. 24 2023, and the war still in continuation. We think that comments posted at this period is enough to measure the political attitudes of people around the word toward the conflict.

General information	Total number of comments	2606
	Number of users	1981
	English comments	42,44%
Channel information	Number of channels	7
	Top Sources of comments	NBC News, BBC News, France24, CNN Brazil, ZDF
	Min comments by channel	100
	Max comments by channel	600

5. Data Pre-processing

Comments often have special characteristics that make their pre-processing different from that of ordinary texts, and may contain special texts, such as hashtags, URLs, emoticons and usernames. For the pre-processing of comments, we followed the next steps:

- Striping and lower casing: removing withespace at the beginning or end of the comment and lower casing the letters.
- Normalization of emoji and emoticons: transform each emoji to its meaning in text, and replace emoticons (special ordering of characters) with text.
- Cleaning: removing usernames, numbers, HTML tags, URLs, and any left punctuation.
- Spell checking: detect the language of the comment in order to perform spell correction. As an example of the output of this step, a comment like "I looove palestin. Happi to visit it" will be corrected to "I love Palestine. Happy to visit it".
- Translation: translate non-English comments to English.
- Removing stop words: keep only words that have I direct influence on the comment class.
- Lemmatization: break the words down to its root meaning based on the language dictionary.

6. Political Attitude Analysis

It is assumed the comment is an opinion and therefore we need to know its polarity classification, which is positive, negative or neutral. To achieve this, we used a supervised approach for sentiment analysis. First we train our model on a training dataset, the we can use it on the collected dataset to extract insights.

We trained different models and a manually labeled dataset of 300 rows, and the results is shown in the next two tables.

Table 1 shows the comparison between classification algorithms, where we notice that Multinomial Naïve Bayes have the best accuracy.

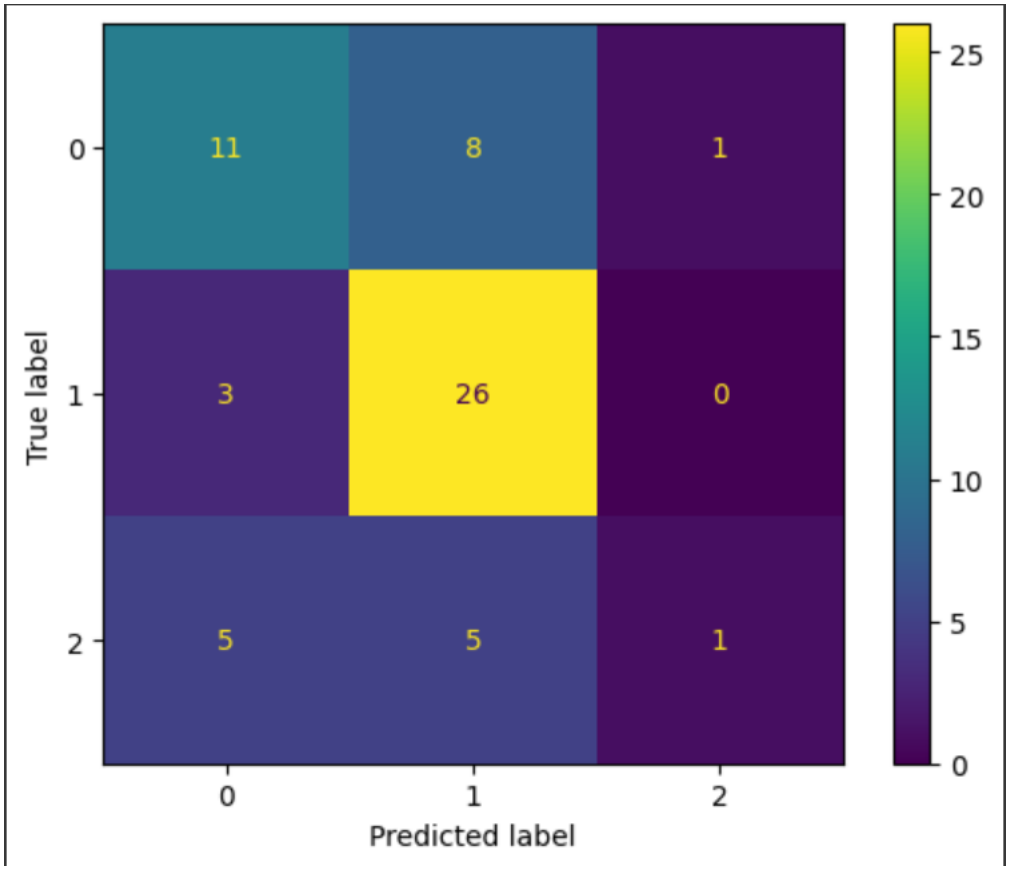
Table 1. Comparison between sentiment classifiers

Model	Accuracy	Precision	Recall	F1-Score
MultinomialNB	**63%**	67%	90%	76%
SVM	**57%**	48%	88%	62%
Xgboost	**47%**	51%	66%	58%
Bert base	**50%**	48%	67%	56%

We remark that our classifiers cannot reach the desired accuracies, and that's totally normal due to the difficulty of the task since comments depend directly on the video content, not like tweets that are somehow independent, and contain enough information to classify it correctly.

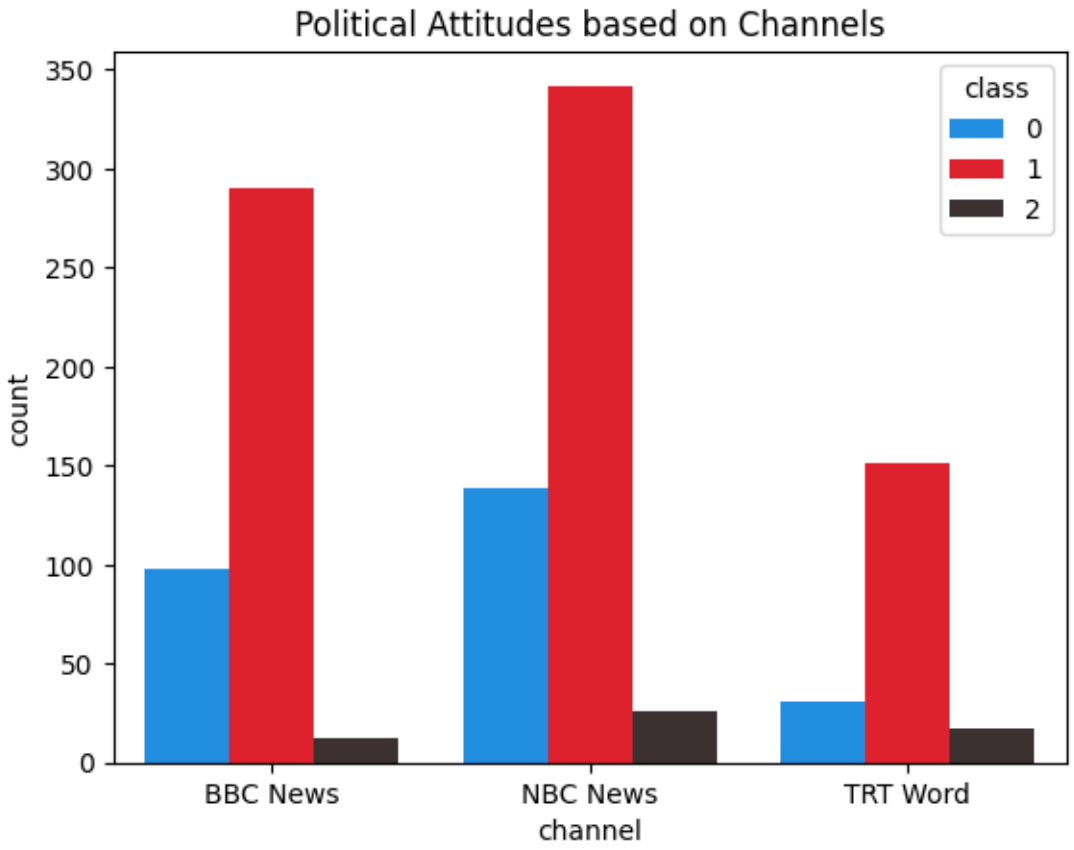
Table 2 shows the testing results of the trained classifier, which is Multinomial Naïve Bayes, a good known algorithm for text classification. Precision and recall values for each class were calculated by creating the confusion matrix.

Table 2. The confusion matrix of the trained naïve bayes model



7. Data Analysis and Results

The sentiment classifier built in the previous section was used to measure the sentiments of all tweets in the dataset, excluding those used to build and test the classifier. In total, only English comments (1106 comments) we measured due some technical issues.



We notice that the number of pro-Palestinian comments on NBC News, BBC News and TRT Word is big, and far from Neutral and pro-Israeli comments, which prove the western media is finally changed, and now the public opinion serve Palestine.

8. Conclusion

Based on the previous results, we can easily conclude that the 7th Oct war is turning point of the word public opinion toward this historical conflict. Specially in the western media, where the previous bias to Israel due to sharing only the Israeli narrative decreased a lot, and people start knowing the real history of the conflict, and start understanding which side deserve the support.

That means in the next days, by the will of ALLAH, we will see consecutive victories, not only geographical, but also political, the majority of word countries now is in Palestine side, which will enforce a high pressure on Israel abuses and irregularities against Palestinians.

9. Future Work

Using YouTube comments on news, find out to be a complex task, this is because the comment is simply a reaction to the video content, the correct classification of the comment don't rely only on the comment text, but also what the video present. When videos have different contexts and we treat all the same, now the joy or sadness, the blame or compliment might not be enough to correctly classify the comment if it lakes to additional topic keywords. That triggers new approaches to effectively perform this task, such as clustering videos based on the context in order to treat each category differently, since when we the topic is fixed, the comment/reaction can easily be classified. Another approach is to classify the video first based on the title or the summary of the content, then start classifying the comments based on information we have about the video.

REFRENCES

[1] I. Al-Agha and O. Abu-Dahrooj, "MULTI-LEVEL ANALYSIS OF POLITICAL SENTIMENTS USING TWITTER DATA: A CASE STUDY OF THE PALESTINIAN-ISRAELI CONFLICT"

[2] W. F. Al-Sarraj and H. M. Lubbad, "Bias Detection of Palestinian/Israeli Conflict in Western Media"

[3] https://console.apify.com/actors/p7UMdpQnjKmmpR21D/console