

Sentiment Analysis of Israel Political Tweets

Rami Fisher 204249759

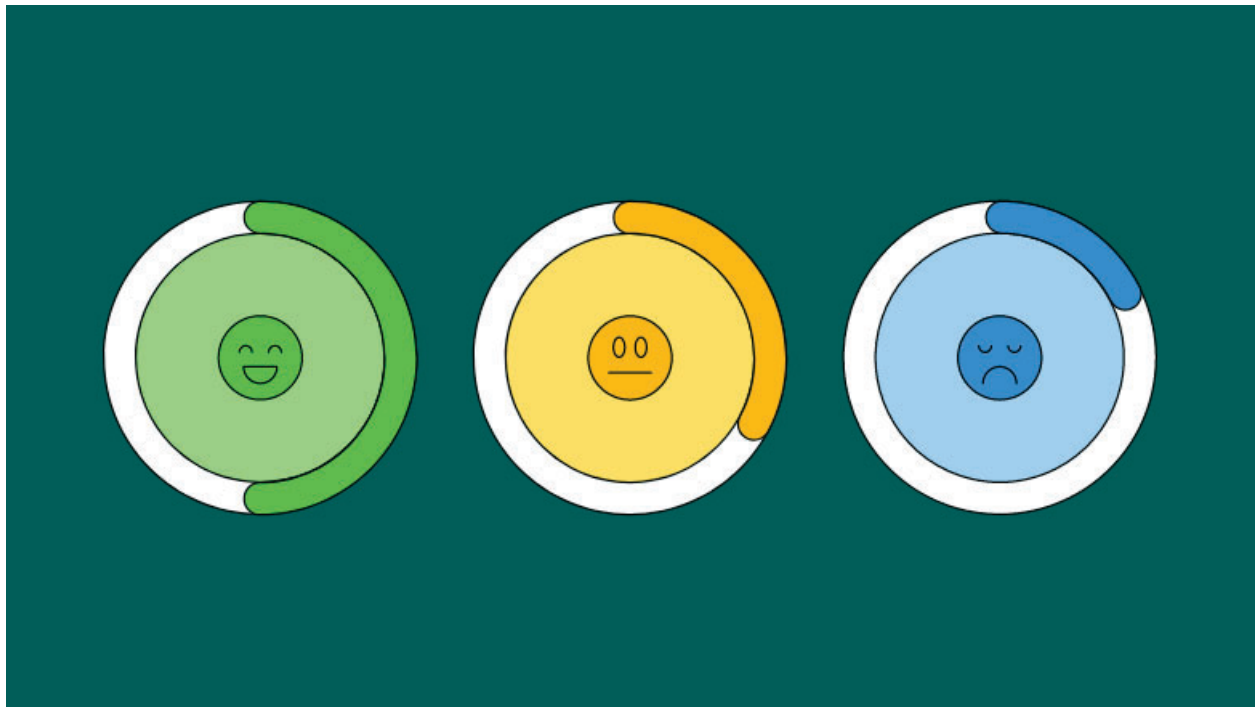


Table Of Content

General Project Definitions.....	3
Research Question.....	3
Story Background.....	4
Data Challenge.....	4
Data Challenge #1 -Bias.....	4
Data Challenge #2- How do we handle ambiguity?.....	5
Data Challenge #3 - Data Collection.....	5
Data Process and Block Diagram.....	6
1. Data Collection - Description and tools.....	7
2. Data Preprocess.....	9
Stop and Get Some Insights.....	10
3. Sentiment Identification.....	11
4. Feature Exploration.....	14
5. Sentiment Visualization.....	16
Conclusions.....	19
Related Academic Articles.....	19

General Project Definitions

Our project will focus on the field of **text processing**, specifically on **sentiment analysis**. The project will be focused on the **research** aspect of the field.

NLP sentiment analysis refers to the use of natural language processing techniques which analyze the subjective information present in text data, such as **opinions and emotions**. It involves the use of advanced algorithms and models to automatically extract sentiment-related information from textual data and classify it into positive, negative, or neutral sentiments.

Twitter is a social media platform that enables users to share their **opinions, thoughts, and feelings** about various topics in short, text-based messages called tweets.

NLP Twitter sentiment analysis aims **to extract valuable insights** from this massive amount of textual data by identifying the sentiment of tweets, which can be **positive, negative, or neutral**.

Research Question

Can we Identify the change in the military-political situation in Israel by sentiment based on tweets in 2023?

We seek to understand how constitutional crises are reflected in the content and thousands of tweets, and how these patterns can be detected and analyzed using NLP methods. Ultimately, the goal of this research would be to develop new insights into the dynamics of constitutional crises and the ways in which they are communicated through social media.

Story Background

Lately, in Israel, a constitutional crisis has given rise to two distinct groups among the citizens, each with differing perspectives on proposed reforms and their potential impact on democracy.

One group advocates in favor of the **reforms (רפורמה)**, asserting that they are necessary **for preserving and strengthening democracy**. This group emphasizes the need for change to address concerns related to concentration of power, corruption, and accountability within the government. Their supporters argue that these reforms would lead to a more inclusive political system, ensuring transparency, and restoring public trust.

In contrast, the **other group** opposes the proposed reforms, contending that they represent a **revolutionary shift (מהפכה)** in the government that could potentially **undermine Israeli democracy**. They argue that the reforms may disrupt the balance of power and weaken democratic checks and balances that have historically been in place.

Data Challenge

Data Challenge #1 - Bias

The data will be based on **Tweets** that are related to **political people\issues**. We are aiming to extract sentiment scores to interpret public opinion towards the Israel political people\issues. To **ensure bias towards both groups involved**, it is crucial to select objective words that are relevant to the concerns expressed by each group.

For instance, terms such as 'רפורמה' (reform) or 'מהפכה' (revolution) would introduce subjectivity and potential bias. As a neutral approach, we have decided to collect tweets containing the term 'דמוקרטיה' (democracy), which represents a common denominator for both groups. By utilizing this term, we aim to encompass a broad range of opinions and perspectives while maintaining unbiasedness.

The collection and analysis of tweets using the keyword 'דמוקרטיה' will allow us to assess the sentiment and discourse surrounding the democratic functioning of the government.

Data Challenge #2- How do we handle ambiguity?

In order to smooth off the potential error of ambiguity in individual twitter. We would use the average score of multiple Twitter texts which can help mitigate the impact of ambiguity. When dealing with sentiment analysis, aggregating the sentiment scores of multiple texts can provide a more robust and reliable measure of sentiment.

When different texts have varying degrees of ambiguity, averaging their sentiment scores can help balance out the effects of uncertainty.

In addition we are using a pretrained sentiment model named BERT which considers the surrounding context of each word within a sentence. By taking into account the entire sentence or text, BERT can capture the context in which ambiguous words or phrases appear. This contextual understanding helps BERT disambiguate the sentiment expressed in the presence of ambiguous terms.

Data Challenge #3 - Data Collection

Collecting Twitter data can present several challenges due to the platform's nature and restrictions. Here are three main challenges and potential ways to address them:

API Limitations: Twitter imposes limitations on API access, including rate limits on the number of requests and the amount of data that can be retrieved. This can hinder the collection of a large volume of tweets within a short timeframe.

Data Bias and Representativeness: Data may suffer from biases and limitations in representativeness, as it represents a subset of users who are active on the platform.

Noise and Irrelevant Content: Twitter data often contains noise, spam, and irrelevant content, which can impact the quality of the collected dataset.

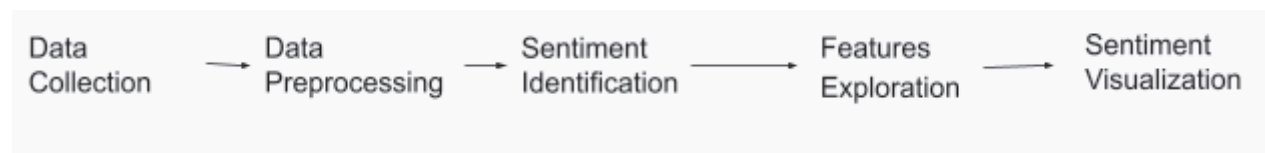
We had 3 optional ways for collecting the data for handling the challenges described above.

1. Twitter API: The Twitter API provides developers with programmatic access to Twitter's data.
2. Third-Party Tools and Libraries: Several third-party tools and libraries simplify the process of collecting Twitter data. Examples of such tools include Twython, Tweepy.
3. Web Scraping: Web scraping involves extracting data directly from the Twitter website by parsing the HTML or using browser automation tools.

In our project, we carefully evaluated and recognized that each tool possesses advantages and limitations. Given our specific requirements, **we have made a deliberate decision to utilize Web Scraping as our preferred method.** This choice was because of its capability to facilitate the collection of an unlimited volume of Twitter data. Moreover, Web Scraping empowers us with extensive filtering functionalities, including parameters such as #Likes, #StartDate, #EndDate, and others.

Data Process and Block Diagram

The graph diagram below presents the process we have implemented during our project.



The process is built from :

1. *Data Collection* - First , we will collect a dataset of tweets according to keywords,hashtags.The tweets will be collected using Twitter API tool or another python package which performs tweets export.
2. *Data Preprocessing* - include data cleaning, filtering, and transforming the raw tweet data into a usable format for analysis. This includes removing irrelevant data, such as URLs and mentions, correcting misspellings, and converting text to lowercase. This step may also include tokenization, stemming, and stop-word removal.

3. *Sentiment Identification* - using a pretrained model of sentiment tweets , in order to label our dataset to 3 main categories Positive , Negative and Neutral. In this stage we may use 2 different pretrained models in order to check the differences.
4. *Feature Exploration*- include selecting relevant features from the preprocessed tweets that are likely to be indicative of sentiment. Common features include bag-of-words, n-grams, word embeddings, and sentiment lexicons.
5. *Sentiment Visualization* - using data visualization tools and techniques to create visual representations of sentiment analysis results.

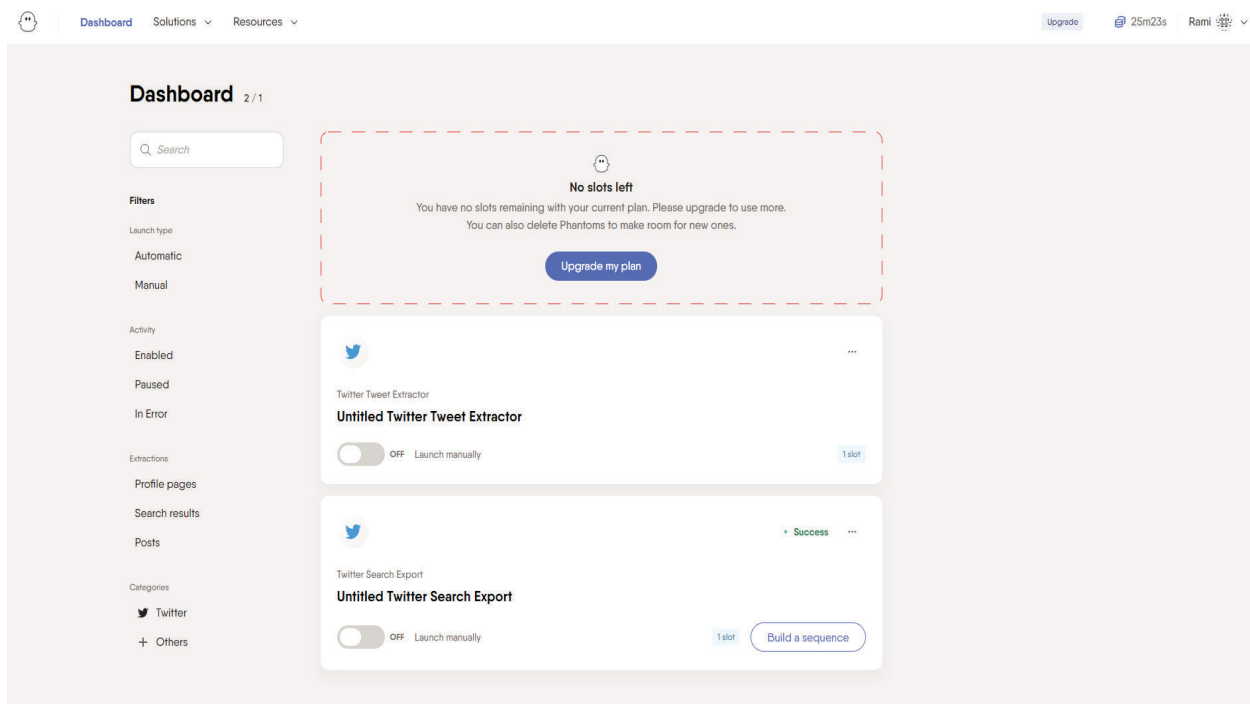
Finally , we would like to identify the evolving of the constitutional crisis through the sentiment score of related tweets.

1. Data Collection - Description and tools

As we have described before our data collection method is Web Scrapping.

We have used a tool named **Phantombuster** which is a versatile and comprehensive automation tool designed for web scraping and data extraction tasks.

In the figure below we can see the GUI of Phantombuster, in which we could have sent a URL link with the requested twitter and get a response with all the twitters.



Data Parameters Description

HashTag : "דמוקרטיה"

Start Time : 1.10.2022

End Time : 20.5.2023

Minimum Likes Per Twitter : 100 Likes

According to the requested parameters we have collected raw data of 2252 twitters.

In the Figure Below we can see examples of partial raw data twitters. The dataset contains the following features :

1. TweetDate - Published time of
2. TwitterId - Unique Id of the the twitter
3. Handle - The used name of the twitter's writerer
4. Text - Content of the twitter
5. ProfileUrl - Link to the profile of the user.
6. Name - Full name of the use
7. TweetLink - twitte link.
8. Timestamp - Download timestamp.

sentiment_raw_data = pd.read_excel("TwittesCombine.xlsx")								
sentiment_raw_data.head(10)								
	tweetDate	twitterId	handle	text	profileUri	name	tweetLink	timestamp
0	Fri Dec 09 10:15:40+0000 2022	1.421636e+18	mosaud08	אשכרה ראש הממשלה שלכם מסגין נגד תחזיות הבחירות...	https://twitter.com/mosaud08	محمد سعود Mohammed Saud	https://twitter.com/mosaud08/status/1601158631...	2023-05-21 17:08:07.270
1	Tue Dec 06 18:13:36+0000 2022	1.113447e+18	AnatiRina	מדובר 9/ בשאלות רטריות לחלוטין. חוק יסוד כבוד...	https://twitter.com/AnatiRina	Rina Anati	https://twitter.com/AnatiRina/status/160019174...	2023-05-21 17:08:07.270
2	Wed Dec 07 06:57:05+0000 2022	3.182426e+09	IsraelGaley	עו"ד ציון אמיר @sharongal על דברי יו"ר לשכת ע...	https://twitter.com/IsraelGaley	גלי ישראל	https://twitter.com/IsraelGaley/status/1600383...	2023-05-21 17:08:07.270
3	Mon Dec 05 10:35:48+0000 2022	9.981974e+17	OmerArvili	תודה לכל עובדי עיריית ירושלים (כולל זאת ששפחה מ...	https://twitter.com/OmerArvili	Omer Arvili	https://twitter.com/OmerArvili/status/15997141...	2023-05-21 17:08:07.270
4	Mon Dec 05 19:55:40+0000 2022	8.764084e+17	YairNetanyahu	שמים סוף לסרור המשפטי נגד אזרחי הימין אשר ישראל	https://twitter.com/YairNetanyahu	Yair Netanyahu	https://twitter.com/YairNetanyahu/status/15998...	2023-05-21 17:08:07.270
5	Wed Dec 07 19:23:14+0000 2022	1.057413e+18	BokerIris	זה מדוע אותי שהילדים שלנו לא יזכו לחופש ולזכו...	https://twitter.com/BokerIris	Iris Boker	https://twitter.com/BokerIris/status/160057165...	2023-05-21 17:08:07.270
6	Fri Dec 09 09:24:18+0000 2022	9.093824e+17	roznaay	בעקבות יוזח נוקב בסוויטר, מלך ראיון עם עמית ...	https://twitter.com/roznaay	יניב רוזנאי Yaniv Roznai	https://twitter.com/roznaay/status/16011457054...	2023-05-21 17:08:07.270
7	Wed Dec 07 10:21:22+0000 2022	1.169205e+18	berale_crombie	בואו נסכם שלפני שאתם מגלים מסיפים לזעזוע "למדיד לזע...	https://twitter.com/berale_crombie	ברלה קרומבי Il	https://twitter.com/berale_crombie/status/1600...	2023-05-21 17:08:07.270
8	Mon Dec 05 06:16:19+0000 2022	3.902784e+07	moshefeiglin	דמויות הקונקס הנגלים שזועקים עכשיו מורה, לא ממ...	https://twitter.com/moshefeiglin	משה סייגלין	https://twitter.com/moshefeiglin/status/159964...	2023-05-21 17:08:07.270
9	Tue Dec 06 06:01:47+0000 2022	3.574964e+07	mk1956	דמוקרטיה 1/ זה קודם כל קבלת גזר דין נכון הבחור...	https://twitter.com/mk1956	Moshik Kovarsky - מושיק	https://twitter.com/mk1956/status/160000757592...	2023-05-21 17:08:07.270

2. Data Preprocess

When dealing with Twitter data, it is crucial to acknowledge the distinctive conventions that set it apart from other online social media platforms. The pre-processing procedure entails several key steps:

1. Filtering Non-Standard Lexical Tokens: Following the tokenization process, non-standard lexical tokens such as mentions, hashtags, emoticons, and unconventional punctuation are removed. This ensures that the resulting tokens maintain a standardized format.
2. Eliminating Duplicate Tweets and Retweets: To preserve the uniqueness of each tweet, duplicate tweets and retweets are eliminated from the dataset. This step helps prevent bias and redundancy in the analysis.
3. Removing Standard Stop Words: Standard stop words, which are commonly occurring words with little semantic significance, are removed from the text. This step helps reduce noise and focuses on more meaningful content.

In our project we have implemented all 3 steps on the raw data. We have used the NLTK package in order to collect all the HEBREW stopwords, in addition we have added several words that weren't included in their library.

In the figure below we can see the result of applying the 'preprocess_data_text' function which includes all the steps we have described.

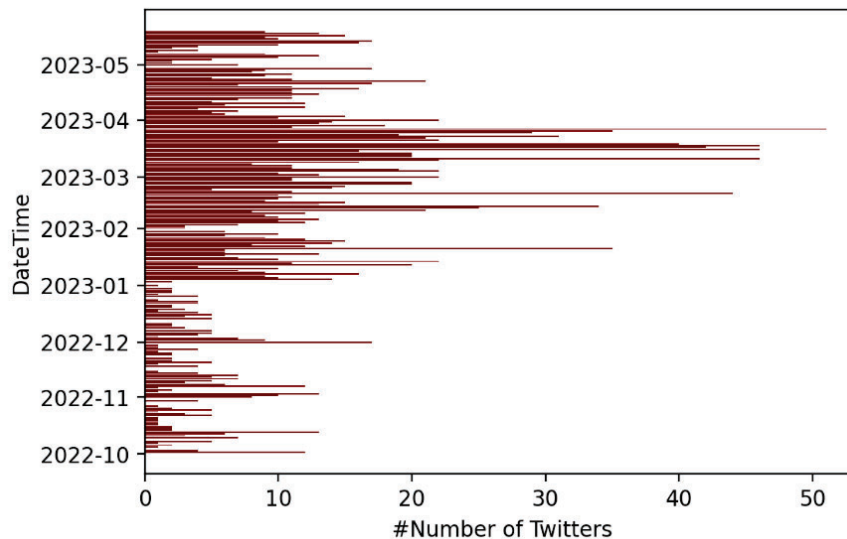
```
pd.set_option('display.max_colwidth', None)
sentiment_raw_data['Preprocessed_Text'] = sentiment_raw_data['text'].apply(preprocess_data_text)
sentiment_raw_data[['Preprocessed_Text', 'text']].head(5)
```

	Preprocessed_Text	text
0	אשכרה ראש הממשלה מפגין נגד תוצאות הבחירות והרפורמות שנתניהו מתכנן, ממש כמו הקפיטול. ואז יש כאן ואז ליצינים שמטיפים דמוקרטיה, אצלכם הבחירות המלצה	אשכרה ראש הממשלה שלכם מפגין נגד תוצאות הבחירות והרפורמות שנתניהו מתכנן, ממש כמו הקפיטול. ואז יש כאן ואז ליצינים שמטיפים דמוקרטיה, אצלכם הבחירות זה המלצה https://t.co/DvyJz3kxOW
1	מדובר בשאלות רטוריות לחלוטין. חוק יסוד כבוד האדם וחירותו הוא חלק מהחוקה מדינת ישראל כמדינה דמוקרטית, ובניגוד למה שטאוב טוען דמוקרטיה איננה רק "ביטוי של רצון האזרחים" אלא היא הגנה על זכויות אדם מפני עריצות השלטון. היא הגנה מפני עריצות הרוב ואיננה יכולה להתקיים בלעדיה	מדובר בשאלות רטוריות לחלוטין. חוק יסוד כבוד האדם וחירותו הוא חלק מהחוקה של מדינת ישראל כמדינה 9/ דמוקרטית, ובניגוד למה שטאוב טוען דמוקרטיה איננה רק "ביטוי של רצון האזרחים" אלא היא הגנה על זכויות אדם מפני עריצות השלטון. היא הגנה מפני עריצות הרוב ואיננה יכולה להתקיים בלעדיה
2	עוד ציון אמיר ל דברי יור לשכת עורכי הדין שקרא לצאת לרחובות מדברים דמוקרטיה כאילו הכרעה חודש ימים בציבור הישראלי שהלך להצביע בדיוק הדברים האלה של רפורמות במערכת הפעילים המשפט מסע חמפה בלתי נגמר שרון גל אבי חימי בחר שלא להתראיין	על דברי יו"ר לשכת עורכי הדין שקרא לצאת לרחובות: "מדברים על דמוקרטיה כאילו @sharongal עו"ד ציון אמיר ל לא היתה הכרעה לפני חודש ימים בציבור הישראלי, שהלך להצביע בדיוק על הדברים האלה של רפורמות במערכת "שרון גל: "אבי חימי בחר שלא להתראיין לחלוטין. המשפט. זה מסע חמפה בלתי נגמר
3	תודה לכל עובדי עיריית ירושלים כולל ששפכה מים שיצאו היום מהמשרד כדי לקלל הפעילים שהפגינו לבית המשפט בו התקיים דיון בנוגע לאדמות שמנסים לגזול מתושבי סילוואן הסיוע בלהראות לעולם ישראל דמוקרטיה	תודה לכל עובדי עיריית ירושלים (כולל זאת ששפכה מים) שיצאו היום מהמשרד כדי לקלל את הפעילים שהפגינו מחוץ לבית המשפט בו התקיים דיון בנוגע לאדמות שמנסים לגזול מתושבי סילוואן. אין על הסיוע שלכם בלהראות לעולם למה ישראל היא לא דמוקרטיה
4	שמים סוף לטרור המשפטי אזרחי הימין ישראל בתקופה האחרונה התפתחה תופעה הזויה אצל השמאל בישראל, כאשר הם משתיקים, מפחידים ומרוששים אזרחים ימנים, באמצעות תביעות השתקה הזויות וחסרות כל התופעה הזאת היא סכנה חמורה ביותר לחופש הביטוי בישראל, שבלעדיו אין דמוקרטיה לחלוטין. https://t.co/g4EYBidOsg	שמים סוף לטרור המשפטי נגד אזרחי הימין של ישראל! בתקופה האחרונה התפתחה תופעה הזויה אצל השמאל בישראל, כאשר הם משתיקים, מפחידים ומרוששים אזרחים ימנים, באמצעות תביעות השתקה הזויות וחסרות כל התופעה הזאת היא סכנה חמורה ביותר לחופש הביטוי בישראל, שבלעדיו אין דמוקרטיה לחלוטין. https://t.co/g4EYBidOsg

Stop and Get Some Insights

Before performing any sentiment model , we would like to get some insights from the preprocessed data.

In the figure below we can see the distribution of the tweets over the time from October 2022 until May 2023. It is clear to see that the number of tweets is increasing to local maximum at the end of January 2023 and again to global maximum in March 2023. This visual insights are making sense with the timeline of the announcement of the legal reform.



In the figure below (on the left) we can see the word cloud of the preprocessed data. As we could have expected the most dominant word is “דמוקרטיה” which is the filtered word which we have been looking for. In addition we can see the word “נתניהו” which is the prime minister of israel.



Most common words:

דמוקרטיה: 2176
ישראל: 412
הדמוקרטיה: 217
שלא: 183
הממשלה: 147
עוד: 145
אחד: 144
נתניהו: 143
ולא: 142
עכשיו: 140

On the figure above (on the right side) we can see the most common words in top 10 least. We can confirm that 'דמוקרטיה' is the most common word with **2176 times**. The second most common word is 'ישראל' with **412 times** which also makes sense.

On the figure below we can see the list of most common 2-grams. By using the **N-GRAM algorithm** we can understand what is the most common topic that is discussed in the tweets. We can see that the both 'מערכת המשפט' and 'ההפיכה המשטרית' are in the **top 10 topics** which confirms that our corpus includes the topics that we are aiming to analyze.

Most common 2-grams:

56: ישראל דמוקרטיה:
 48: מדינת ישראל:
 46: ההפיכה המשטרית:
 45: דמוקרטיה ליברלית:
 40: שלטון הרוב:
 39: דמוקרטיה אמיתית:
 36: דמוקרטיה שלטון:
 36: מערכת המשפט:
 34: המילה דמוקרטיה:
 33: דמוקרטיה מרד:

3. Sentiment Identification

At this point we have **2187 preprocessed tweets** which are related to our research topic.

We would like to classify each tweet as Negative / Positive / Neutral , for that we will use a pretrained sentiment model - **twitter-xlm-roberta-base-sentiment(HuggingFace)**. The model output is sentiment score to each one of the 3 main labels : "Positive", "Natural", "Negative".

The Final output of the model will be:

1. DateTime
2. TweetText
3. NegativeScore
4. NaturalScore
5. PositiveScore

	Date	Text	Negative_Score	Natural_Score	Positive_Score
2	2022-10-01 02:10:13	יודע נזכרתי בזה פתאום זוכרים שבמדינה דמוקרטית תפקידה התקשורת לבקר השלטון מבין נכון עכשיו השלטון מבקש לסגור ערוץ התקשורת מבקר השלטון אחלה דמוקרטיה	0.491700	0.294873	0.213427
3	2022-10-01 07:05:56	בתיאום מופלא הליכוד החליט להישאר בבית להרוס מבפנים דמוקרטיה למופת אכן ולבסוף כפרס בגידתו בבנט לקבל שריון בליכוד חשבון שרצים ומדענים ופועלים במפלגה שנים ערבי ולא חרדי והוא נעים לעין ולאוזן החוק שיחכה בסבלנות לפעם הבאה	0.600573	0.191813	0.207615
4	2022-10-01 17:41:03	אזהרה אמריקאית בדיחה ברק רביד התועמלן נשלח אתר הפח וואלה לאיים נתניהו התערבות בבחירות כנראה שרביד ואתר וואלה שמריץ לפיד מבינים דמוקרטיה	0.891799	0.087368	0.020833
5	2022-10-02 02:58:11	שניתן להבין מהטוויטר הימים האחרונים אם גוש השחיתות ימנע בבחירות סבבה מבחינתו הבחירות זוייפו אם הנאשם יזוכה במשפטו סבבה לאמערקת המשפט מושחתת וכתב האישום והמשפט היו תפורים דמוקרטיה הנאשם רהמ לאזז דיקטטורה השמאל	0.507878	0.357376	0.134746
6	2022-10-02 05:26:56	איש התקשורת שבמקרה מצא כסא ראש הממשלה אוהב ציטוטים בואו מעזור להיזכר באחד פעם שראש הממשלה תוקף התקשורת קצת פחות דמוקרטיה י לפיד	0.302301	0.362424	0.335275

In the figure above , we can see a sample from the output of the sentiment model. In order to understand and confirm the model reliability, we would again use the word cloud and n-gram algorithm.

In the figure below we have performed 1-gram and word cloud on the **top 10 positive sentiment scored tweets**. The words in the top 10 includes 'אדירה' 'טוב' 'ערכי' most of the words have **positive connotations** , this confirms that the model is performing well.



Most common words:

דמוקרטיה: 9
חוגגים: 5
טוב: 3
קפלן: 2
אדירה: 2
בקפלן: 2
המדינה: 2
ערכי: 2
לכל: 2
סרסון: 2

In the figure below we have performed 1-gram and word cloud on the **top 10 negative sentiment scored tweets**. The words in the top 10 include 'נורא' 'דם' 'רעים' 'נורא' most of the words have **negative connotations** this confirms that the model is performing well.



Most common words:

דמוקרטיה: 8
נורא: 3
אנשים: 3
זה: 2
ישראל: 2
בישראל: 2
רעים: 2
בה: 2
רוץ: 2
דם: 2

In the figure below we have performed 1-gram and word cloud on the **top 10 neutral sentiment scored tweets**. The words in the top 10 include 'הראל' 'דן' 'גלנט' 'עמיקם' most of the words are **names of people or cities** , this confirms that the model is performing well.



Comic Relief - which tweet is the most positive in the dataset?

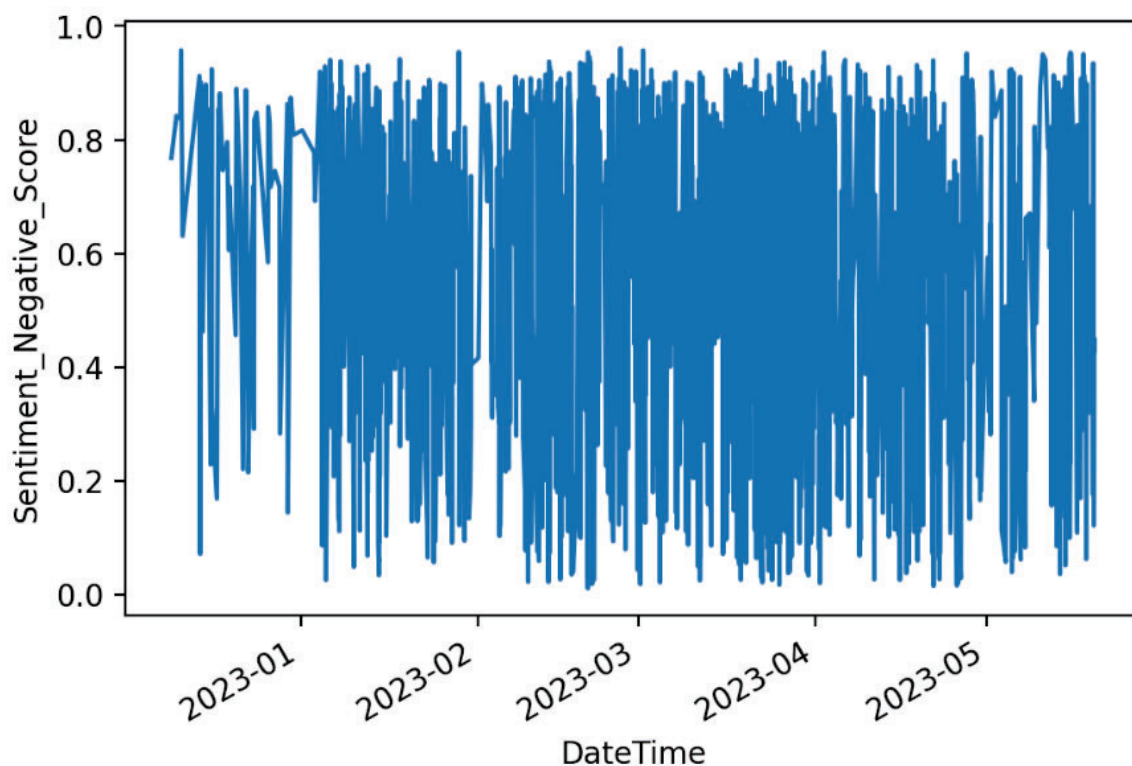
	Date	Text	Negative_Score	Natural_Score	Positive_Score
1738	2023-04-01 18:25:42	ולסיכום לצעוד אורדי הפועל מבלומפילד ניצחון קפלן להפגנה אחת החוויות הכי כיפיות משמחות מרגשות ומעולות שיש חד משמעית	0.019780	0.065801	0.914418

In the figure above we have printed the most positive tweet , which is a fan of a football club who's really happy because their team won the game and now he is planning to join a protest with that same happy feeling.

4. Feature Exploration

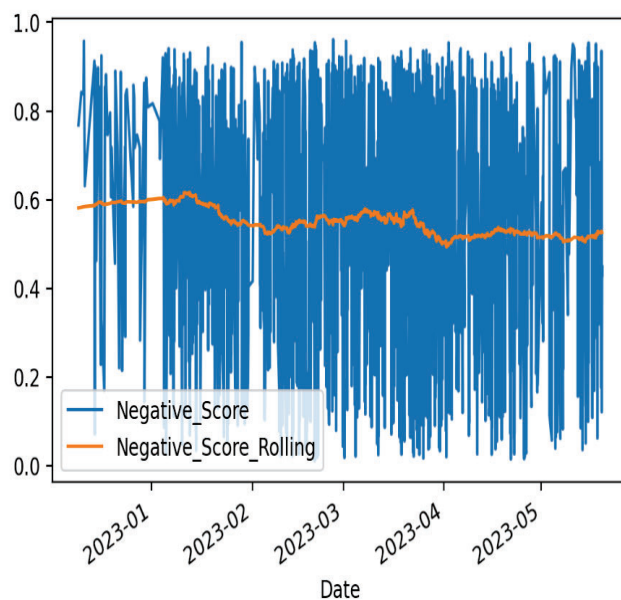
At this point we have the score for each one of the tweets in our dataset. In order to keep it simple we will focus on the **NEGATIVE scores values** (The results for the Positive scores would be dual).

Our research question is aiming at **identifying critical \ change moments in the sentiment of the citizens in israel**. As a starting point we will use **temporal analysis** in order to visualize the change in the sentiment over time(Using the raw data with no engineering).



At the figure above we can see the negative sentiment value of each tweet over time from Oct 2022 till May 2023. The graph is **very dynamic moving from 0 to 1**, so it will be hard to get any **insight** from the values of a single tweet. In addition we would like to **smooth out errors that are caused by ambiguity or bias**.

Rolling average, also known as **moving average**, is a statistical calculation that provides a **smoothed representation** of a dataset by taking the average of a subset of values over a specified window or period. By incorporating rolling average values as **new features**, we can **capture trends and patterns in the data over time**. This can be particularly useful in time series analysis or when dealing with sequential data.



	Negative_Score	Negative_Score_Rolling
Date		
2022-12-09 10:15:40	0.768007	0.581288
2022-12-10 07:12:39	0.842283	0.582690
2022-12-10 23:28:13	0.837626	0.583638
2022-12-11 02:45:44	0.957627	0.583902
2022-12-11 09:01:48	0.630902	0.584394
2022-12-14 08:02:57	0.912568	0.586835
2022-12-14 10:48:26	0.070979	0.586543
2022-12-14 11:02:14	0.393278	0.586447
2022-12-14 13:11:41	0.902229	0.588564
2022-12-14 21:50:24	0.462468	0.589419

At the figure above we can see how using rolling average can smooth the graph , now it would be easier to identify changes in the sentiment using the new explore feature.

To identify periods of **significant change**, we need to **set thresholds for High, Medium, and Low** sentiment scores. These thresholds will help us categorize sentiment scores and pinpoint when there has been a big shift in sentiment over time.

We have define the threshold according to average and standard deviation :

Upper threshold of 0.56 - Red Area

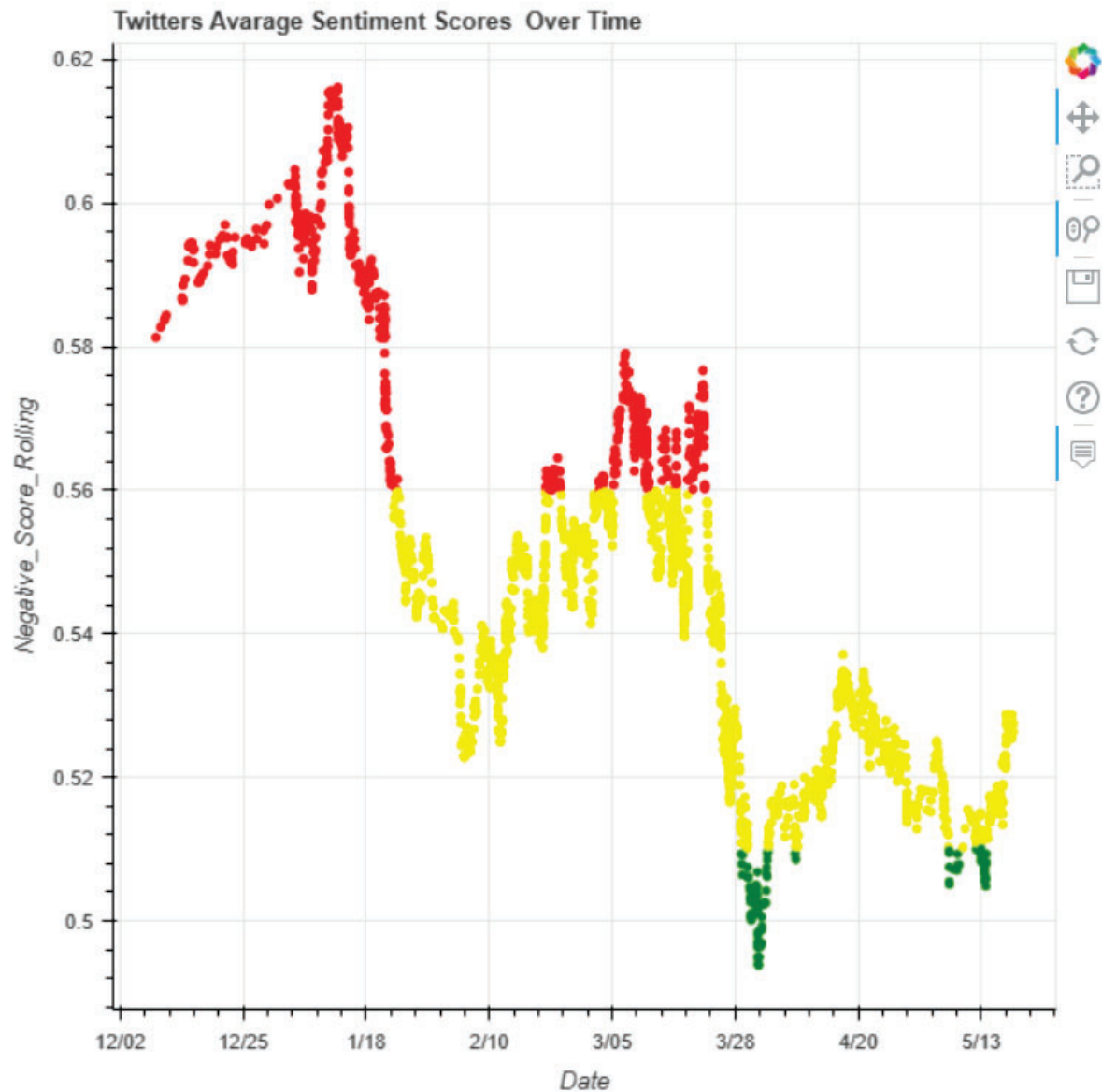
Between 0.56 and 0.51 - Yellow Area

Lower threshold of 0.51- Green Area

Those areas can be seen in the dataset below under the field **'AreaSentiment'**

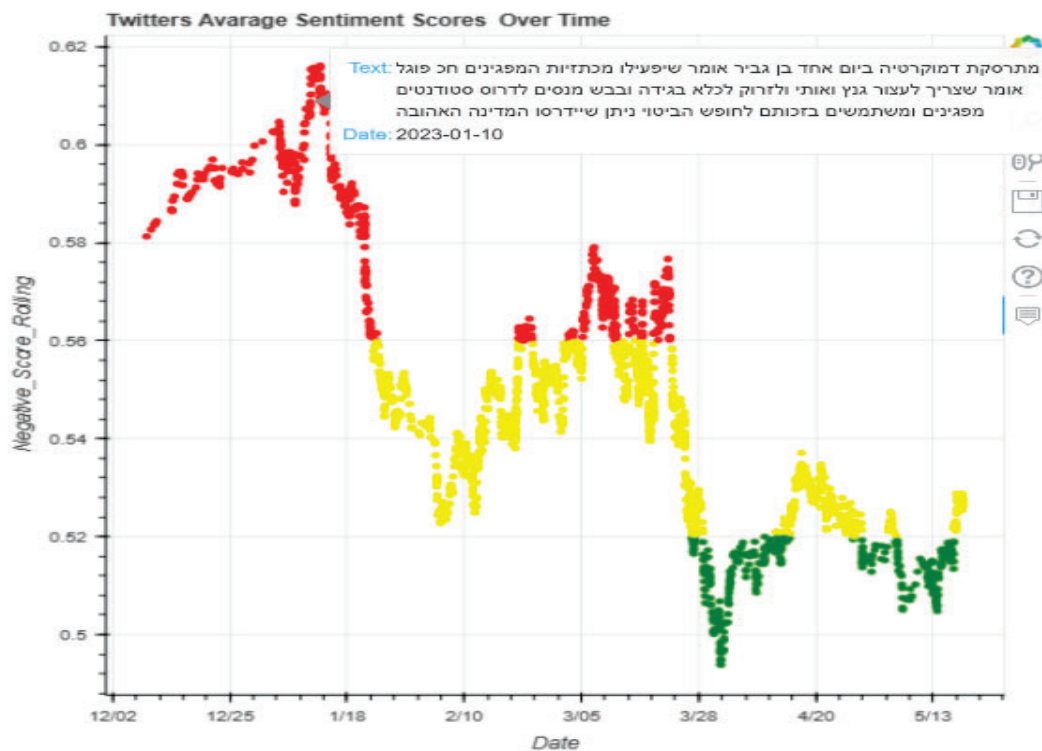
	Text	Negative_Score	Negative_Score_Rolling	AreaSentiment
500	יודע כמה מכם זוכרים השם המקורי קניון עזריאלי תל אביב כשתוכנן בשנות התשעים המאה הקודמת קניון השלום היום קניוני עזריאלי סוענים שהמילה דמוקרטיה משהו שאסור לכתוב שלטי הפרסום אצלם חתיכת מטאפורה שקרה בעשרים וחמש השנים האחרונות	0.474535	0.585198	red
501	עזריאלי מוחק המילה דמוקרטיה משלט פרסום מוחק האפשרות להיכנס אליו לחנוני פשוט לאללה	0.449802	0.583926	red
502	הסרטון שנאסר להקרנה ע"י קבוצת עזריאלי המילה דמוקרטיה	0.167736	0.581227	red
503	רשימה קניוני עזריאלי לשימושם שתדעו לעשות הקניות דמוקרטיה איננה עמדה פוליטית נקודה	0.305817	0.579100	red
504	עזריאלי כמקרה בוחן עבור המחנה חפץ החופש במדינה באו נראה לעצמנו שאנחנו לפעול כיחידה אחת ולגרום להצטרף היום בו החליטו שהמילה דמוקרטיה להופיע שלטי הפרסום נפילת מכירות ברשת עזריאלי יעשו העבודה	0.241824	0.576237	red
...
2184	שאני בטוח בו שיש התעוררות פעם בדור באמת אפילו קלייזא ושאני זורם מרכיביה כשיש טעויות כשהמסר מספיק טוב עבורי תמיד מסכים הדרך עושים נלחמים דמוקרטיה מנצחים	0.176512	0.526015	yellow
2185	ימים רעים עוברים אחריו המדינה הדמוקרטית והליברלים מאז כמה הממשלה הקייסנית והמסוכנת הזו במקום לדאוג לעתיד המדינה להשקיע בהייטק ברווחה ובחיטך הוססו מיליארד שקל לטובת ביטור השלטון הקמת מליציה לאומנית ליצירת דור ללא לימודי ליבה ללא לימודי דמוקרטיה וללא עתיד ההפיכה	0.934845	0.528742	yellow
2186	כמה שיעשו לינץ לגלית גוטמן המסר שהיא ניסתה להעביר יקלט מסר אמיתי ד' לביצה	0.121300	0.527727	yellow

5. Sentiment Visualization



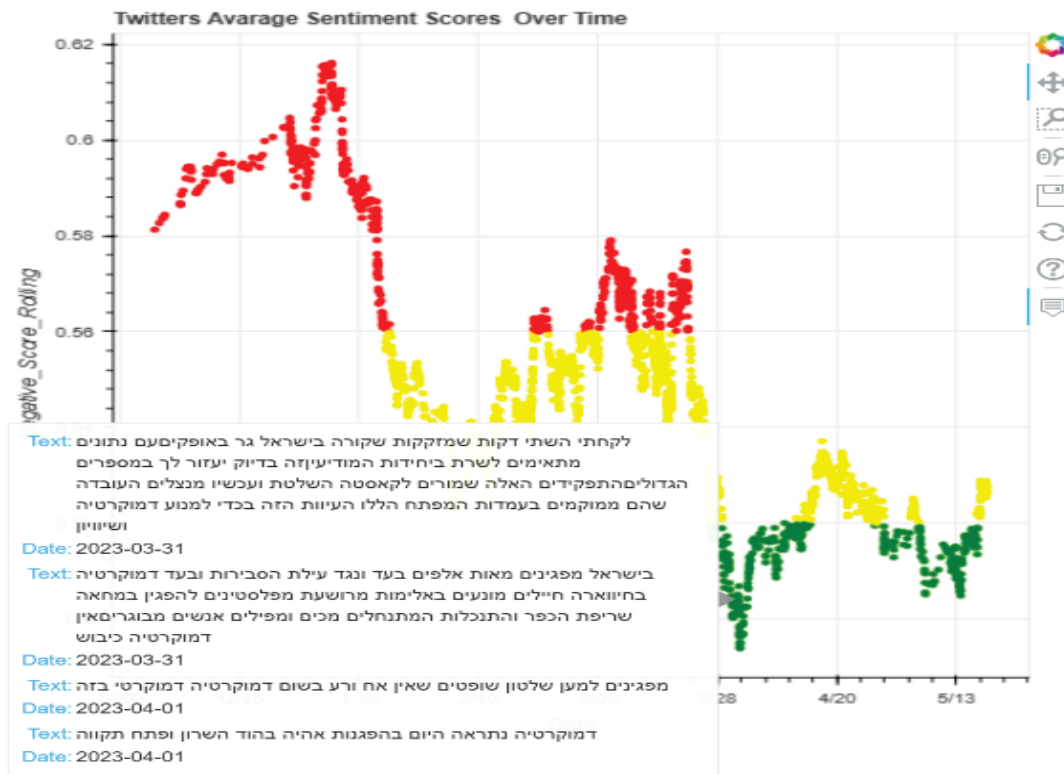
The attached figure shows an **interactive tool** created using **BOKEH**. It helps us visualize the rolling average sentiment over time. The graph uses different colors like **Green, Yellow, and Red** to represent **sentiment levels** based on the thresholds we set earlier. By selecting any point on the graph, we can see the corresponding tweet that was posted at that time. This tool allows us to easily track and explore how sentiment changes over time and understand the tweets associated with specific points on the graph.

According to the visualization tool we would like to focus on 2 main points which are the maximum and minimum of the sentiment graph.



At the attached figure above we have selected one of the tweets in the global maximum of the all sentiment graph. The maximum is occurring on **10.1.2023** which is the reform announcement week. The announcement was happening on **4.1.2023(הכרזה על הרפורמה)**, the delay in the sentiment of the graph can be explained by rolling average.

Surprisingly we can see a deep decline in the graph after a few weeks, Till the beginning of March approx at **9.3.23(יום ההתנגדות של המחאה)** where we can see a local maximum.



After the date **9.3.23**, the sentiment starts to decrease and continues to do so until it reaches the lowest point on the graph. This **lowest point occurs on 1.4.23**, just a few days after the Ayalon protest (מחאת איילון). Surprisingly, despite expectations, the **Ayalon protest is reflected in the social network as a moment of hope**, with the lowest negative sentiment recorded in the graph.

Another interesting point can be seen on **9-14.5.23** where we can see local minimums, this makes sense with the fact that military operations (**מבצע מגן וחוץ**) were executed at those days. It seems that israel citizens were not focused on the political situations on those days.

Conclusions

1. Getting a reliable and significant dataset of tweets is a very challenging task , caused by API limitations.(Web scraping)
2. Cleaning the dataset is an essential for getting reasonable results.(StopWords , Hashtags , Url etc.)
3. Relaying on a single tweet may cause misleading results,thus feature engineering was necessary for getting insights from the tweets.(Rolling Averaging)
4. Using our interactive tool we have identified areas of sentiment change in the military-political situation in israel. (הכרזת הרפורמה,יום ההתנגדות,מחאת איילון)
5. RoBERTa is performing well on identifying positive/negative/neutral words in HEBREW. As we could have seen in the top K tweets for each category.
6. We have generated an html (IsraelSentimentTemporalTool) file for testing the sentiment graph results.
- 7.Future work - getting better results we may have used weighted tweets score according to their #Likes.

Related Academic Articles

1. ***Sentiment Analysis of Political Tweets for Israel using Machine Learning*** - Amisha Gangway, Tanvi Mehta
Link - <https://arxiv.org/abs/2204.06515>
Related to our project: Similar to our project, the article also uses the collection of tweets on the Political issues , but unlike us, the focus is on the Israeli-Palestinian conflict. We could use the article to understand the work process (Data Collection,Data Preprocessing,Data Cleaning) and the relevant architecture for the practical execution of our process.
2. ***Analysis of Political Sentiment Orientations on Twitter*** - Mohd Zeeshan Ansaria, M. B. Aziza,*, M. O. Siddiqui, H. Mehraa, K. P. Singha
Link - <https://www.sciencedirect.com/science/article/pii/S1877050920306669>
Related to our project: Similar to our project, the article also uses the collection of tweets on Political issues , but unlike us, the focus is on the Elections in India. We could use the article to understand how to perform text vectorization.