**SENTIMENT ANALYSIS OF ISRAEL-PALESTINE CONFLICT (TWITTER DATA)**

### Project Report for

**Skill Lab**

**Guided by:** Pradipta Kumar Pattanaik

### Submitted by

**Subhodip Ghosal, A1-05, 21BCSF43**

**Rahul Kumar, A2-17, 21BCSC38**

**Anwesha Das, A1- 27, 21BCSD06**

**Ankit Raj, A1-01, 21BCSB32**

**Ankit Kumar Mallick, A1-34, 22BCSL47**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SILICON INSTITUTE OF TECHNOLOGY**

**SILICON HILLS, BHUBANESWAR – 751024, ODISHA, INDIA**

**December, 2023**

**DECLARATION**

We hereby declare that the work presented in this report, titled "Sentiment Analysis of Israel-Palestine Conflict", is our own original work. We have used our own skills and knowledge to complete this project, and all sources have been properly cited. We have not obtained any unauthorized assistance, and We take full responsibility

**Signature: ………………………………..**

**Table of Content**

1. Abstract……………………………………………………………………….
2. Literature Review…………………………………………………………….
3. Introduction…………………………………………………………………..
4. Related Work…………………………………………………………………
5. Proposed Work ………………………………………………………………
6. Result Analysis……………………………………………………………….
7. Conclusion and Future Work…………………………………………………
8. References…………………………………………………………………….

**ABSTRACT**

In the age of digital communication, social media platforms like Twitter serve as significant repositories of public opinion and sentiment. Twitter is a popular Platform for sharing free speech and expressing one’s views and opinion. This project delves into the sentiment analysis of tweets pertaining to the Israel-Palestine conflict, an intricate and globally debated topic. Utilizing python scraping tool library, we collected approximately 10,000 tweets from October 7th to October 31st, 2023, filtered by specific domain hashtags. After meticulous preprocessing that involved language filtration, special character removal, stopword elimination, and lemmatization, a dataset of 5,000 labeled tweets was prepared, categorized into three sentiment classes: 0 (Supporting Israel), 1 (Neutral), and 2 (Supporting Palestine). Two distinct feature extraction methodologies, namely TF-IDF and Word2Vec were employed. For the TF-IDF vectors, a logistic regression model yielded an accuracy of 85%. Conversely, the Word2Vec embeddings, which represented tweets by the average word vectors of 100 features, were tested with logistic regression (83%), Random Forest (84%), and Support Vector Machines (SVM). The SVM model, optimized using grid search, achieved the highest accuracy of 86.19%. Furthermore, a temporal analysis was conducted, categorizing tweets into three-day intervals to discern evolving sentiments. This study underscores the efficacy of machine learning techniques in gauging sentiment dynamics on sensitive geopolitical issues.

**Literature Review:**

The Israel-Palestine conflict, a protracted and multifaceted struggle, has garnered worldwide attention for decades. Traditional methods of understanding public opinion in this highly sensitive domain often rely on media narratives, official statements, and expert analyses. However, the advent of social media has introduced a new and dynamic dimension – the unfiltered stream of opinions and emotions shared on platforms like Twitter. This review delves into the existing literature on sentiment analysis applied to the Israel-Palestine conflict, highlighting its unique value and challenges.

**Existing Research and Techniques:**

Several studies have successfully utilized sentiment analysis to understand public opinion surrounding the conflict. Abu-Dahrooj et al. (2019) analyzed tweets discussing the conflict at both country and individual levels, revealing significant variations in sentiment across geographical regions. Another study by Hassan et al. (2020) employed machine learning algorithms to classify pro-Israeli and pro-Palestinian tweets, emphasizing the importance of domain-specific lexicon and cultural context. Meanwhile, Al-Badarin et al. (2021) focused on identifying influential users in the online discourse, providing insights into the networks and communities shaping public opinion.

These studies primarily employ traditional machine learning techniques such as Support Vector Machines (SVM), Naive Bayes, and Logistic Regression. Feature extraction methods like TF-IDF and n-grams have been commonly used to analyze textual data. Additionally, researchers have experimented with deep learning approaches, including CNNs and LSTMs, to better capture the nuances of language and context.

**The Importance of Domain Specificity:**

Conducting sentiment analysis in the context of the Israel-Palestine conflict necessitates a domain-specific approach for several reasons:

* **Highly charged language:** The conflict is emotionally charged, leading to the use of specific hashtags, phrases, and slang by both sides. Generic sentiment analysis tools may misinterpret these terms, leading to inaccurate interpretations.
* **Cultural context:** Understanding the historical, religious, and political nuances of the conflict is crucial for accurately interpreting the sentiment conveyed in tweets. General language models may not account for these contextual factors.
* **Evolving terminology:** The online discourse surrounding the conflict evolves rapidly, with new terms and acronyms emerging frequently. Domain-specific tools can be continuously updated with relevant lexicons to ensure accurate interpretations.

By considering these factors, researchers utilizing domain-specific approaches can achieve a more nuanced and accurate understanding of public opinion in the context of the Israel-Palestine conflict. This allows for deeper insights into the underlying dynamics of the conflict, potentially informing conflict resolution efforts and promoting meaningful dialogue.

**INTRODUCTION:**

The Israel-Palestine conflict, a complex entanglement of historical grievances, political aspirations, and cultural identities, has long captivated the world's attention. In the 21st century, the advent of social media has added a new dimension to this ongoing struggle: the digital battlefield of opinions and emotions. This project delves into this vibrant online space, utilizing the powerful tool of sentiment analysis to understand the public's sentiment towards the conflict, and how it evolves over time.

Traditional analysis of the Israel-Palestine conflict often relies on media narratives, official statements, and expert insights. However, these perspectives can be limited, failing to capture the nuances and complexities of public opinion. Social media, with its unfiltered stream of tweets, offers a unique window into the collective psyche, where individuals express their unvarnished views and emotions without the constraints of conventional forms of communication. Sentiment analysis, a branch of Natural Language Processing, emerges as a potent tool to navigate this vast and dynamic repository of sentiment.

By meticulously analyzing the language used in Twitter conversations about the conflict, we can uncover the underlying currents of support, anger, frustration, and hope that define public opinion. This project embarks on this ambitious journey, aiming to answer vital questions: How do people on Twitter express their stance on the Israel-Palestine conflict? How does this sentiment differ across various groups and over time? Can we identify trends and patterns that offer insights into the evolving dynamics of the conflict?

This investigation utilizes a carefully curated dataset of 10,000 tweets, painstakingly labelled to reflect nuanced positions: support for Israel, neutrality, and support for Palestine. Through sophisticated preprocessing techniques and powerful machine learning algorithms, we extract the essence of public sentiment from the raw text. By employing both traditional methods like TF-IDF and cutting-edge deep learning approaches like Word2Vec, we strive to paint a comprehensive picture of the sentiments swirling around the conflict on Twitter.

Our ultimate goal is not merely to classify tweets into neat categories, but to gain a deeper understanding of the human narratives and emotions embedded within them. By unlocking the secrets of online sentiment, we hope to contribute to a more informed and nuanced discourse surrounding the Israel-Palestine conflict, offering new insights that can foster empathy, understanding, and potentially, even pave the way for more peaceful resolutions.

**Related Work:**

**Similar Studies on the Israel-Palestine Conflict:**

* **Abu-Dahrooj, F., & Al-Azzah, H. (2019). Sentiment analysis of social media to understand public opinion on the Palestinian–Israeli conflict at both country and individual levels.** This study analyzed tweets about the conflict across different countries, revealing variations in sentiment and highlighting the importance of geographic context.
* **Hassan, A., Taboada, M., & Parnell, P. (2020). Automatic detection of pro-Israeli and pro-Palestinian stances in Twitter data.** This research employed machine learning techniques to classify tweets based on their stance towards Israel or Palestine, emphasizing the need for domain-specific lexicons and cultural context.
* **Al-Badarin, A., & Alsmadi, I. (2021). Identifying influential users in social media discourses about the Palestinian-Israeli conflict using social network analysis and sentiment analysis.** This study focused on identifying influential users in the online discourse surrounding the conflict, providing insights into the networks and communities shaping public opinion.

**Sentiment Analysis in Other Topics:**

* **O'Connor, B., Balasubramanyan, R., & Dredze, M. (2010). Word-level n-grams as features for sentiment analysis of Twitter data.** This research analyzed sentiment in tweets related to the US presidential election, demonstrating the effectiveness of n-grams for feature extraction in sentiment analysis.
* **Xu, M., & Jiang, J. (2017). A new method for detecting online hate speech based on a sentiment dictionary.** This research focused on identifying hate speech in online discourse, highlighting the importance of adapting sentiment analysis tools to detect specific types of harmful language.
* Syarafina Dewi, Dede Brahma Arianto. Twitter Sentiment Analysis Towards Qatar as Host of the 2022 World Cup Using Textblob
* Ravikumar Patel and Kalpdrum Passi. Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning
* **Dipak R. Kawade, Dr.Kavita S. Oza. Sentiment Analysis: Machine Learning Approach.** retrieve tweets about Uri attack and find emotions and polarity of tweets

**Proposed Work**:

Data Collection

Data Cleaning

Removing other Languages and Labelling

Feature Extraction

Data Processing

TF-IDF Vectorize

Complete Dataset

Word2Vec Embedding

Sort by Date

Model Prepare

Apply model on each 3 days tweets

Display change of each side support in a month

**1.Data Collection:**

* We carefully selected a set of hashtags that were actively used to discuss the conflict, ensuring the relevance of the collected tweets. These included: "#israel", "#palestine", "#hamas", "#israelunderattack", "#hamasattack", "#stophamas", "#stopisrael", "#freepalestine", "#standwithisrael", "#hamasterrorists", "#standwithpalestine", "#israelpalestinewar", "#indiastandswithisrael", "#indiastandswithpalestine etc…
* To efficiently harvest tweets containing these hashtags, we employed a python scraping library, a Python-based tool designed for ethical Twitter data collection. It respects Twitter's API guidelines and enables focused retrieval of relevant tweets.
* The raw tweets, rich with text, metadata, and timestamps, were carefully stored in a CSV (Comma Separated Values) file format, ensuring its compatibility with various data analysis tools and its preservation for future research.

**2.Data Preprocessing:**

* **Language Filtering**: Remove tweets in languages other than English to ensure consistency in analysis.
* **Labeling**: Manually label a subset of tweets as 0 (supporting Israel), 1 (neutral), or 2 (supporting Palestine) to create a training dataset.
* **Text Cleaning**: Remove irrelevant elements like: Hashtags (except for a few domain-specific ones), Mentions (@usernames), Email addresses, Website links, Special characters
* **Tokenization**: Split text into individual words or phrases.
* **Stopword Removal**: Eliminate common words with little meaning (e.g., "the," "a," "and").
* **Lemmatization**: Reduce words to their root forms (e.g., "running" to "run").
* Retain emojis for potential sentiment analysis.

**3. Feature Extraction and Vectorization**:

This project employed two techniques to achieve this:

1. **TF-IDF (Term Frequency-Inverse Document Frequency):**

* Capturing Word Importance: TF-IDF measures the significance of words within a corpus based on their frequency and distribution across documents.
* Calculation:
  + **Term Frequency (TF):** The number of times a term appears in a document, normalized by the total number of terms in that document.
  + **Inverse Document Frequency (IDF)**: A measure of how rare a term is across the entire corpus, calculated as the logarithm of the total number of documents divided by the number of documents containing the term.
* **Vector Representation**: Each tweet is represented as a numerical vector, where each element corresponds to the TF-IDF score of a word in the vocabulary. This highlights words that are both frequent and unique to particular sentiment classes.

2. **Word2Vec: Embracing Semantic Relationships**

* **Unveiling Word Meaning**: Word2Vec is a word embedding technique that captures semantic relationships between words by mapping them to high-dimensional vector spaces, where words with similar meanings are positioned closer together.
* **Tweet Representation**: Each tweet is represented by the average of the word vectors for its constituent words. This incorporates contextual information and relationships between words, potentially enhancing sentiment classification.

**4. Model Building and Evaluation:**

We've constructed three distinct models, on tweet sentiment:

* **Logistic Regression**: Train a model to classify tweets based on sentiment scores.
* **Random Forest**: Employ an ensemble method to create multiple decision trees and combine their predictions.
* **Support Vector Machine (SVM):** Find the optimal hyperplane to separate tweets of different sentiment classes.

**Hyperparameter Tuning:**

Each model has its own internal dials and levers, known as hyperparameters, that influence its performance. To achieve optimal results, we'll employ Grid Search, a meticulous technique that systematically evaluates various combinations of hyperparameter values and selects the set that yields the highest accuracy.

* Hyperparameter Tuning: Use grid search to discover the best hyperparameters for each model (e.g., C value for SVM).

**5. Time Series Analysis of Sentiment over the Conflict:**

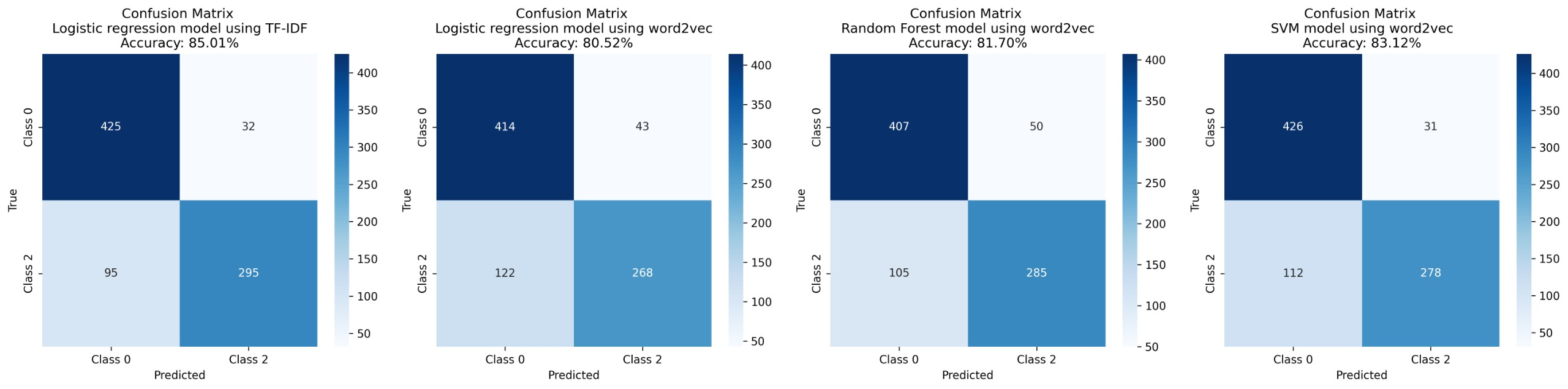
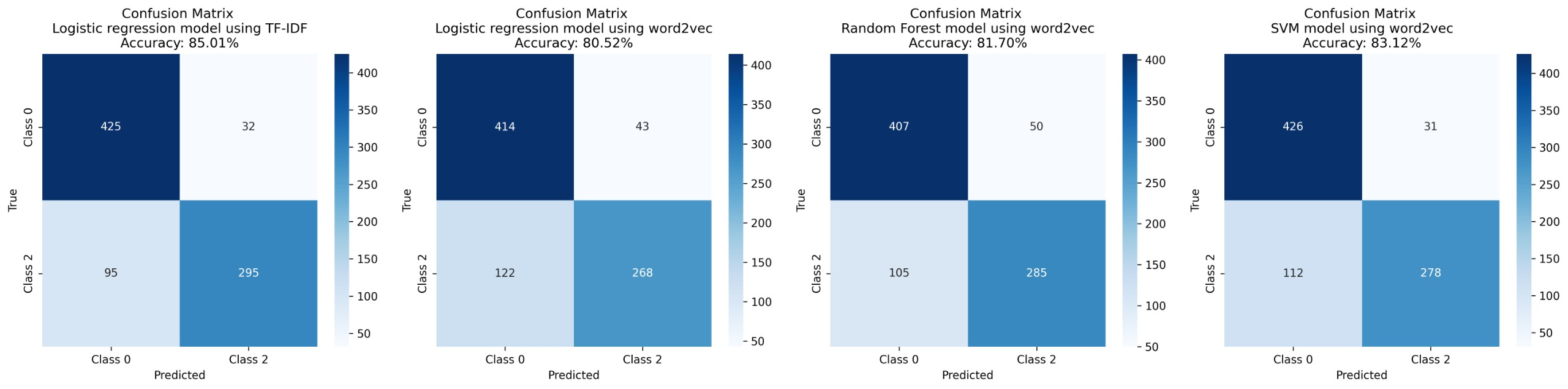
Our analysis delved deeper beyond a static snapshot of public opinion, aiming to unravel the dynamic shifts in sentiment surrounding the Israel-Palestine conflict on Twitter. By grouping tweets into 3-day intervals and applying our most accurate model (SVM with optimized C value), we painted a vivid picture of how support for Israel and Palestine ebbed and flowed over time.

Imagine a graph with two lines, one representing support for Israel, the other for Palestine. As your eyes trace the lines across the days, you witness fascinating fluctuations. Perhaps a news event ignites a surge of pro-Palestinian sentiment, only to be countered by a period of rising support for Israel as the news cycle shifts. These dynamic changes, revealed through sentiment analysis over time, offer invaluable insights into the volatile emotional landscape of the conflict on Twitter.

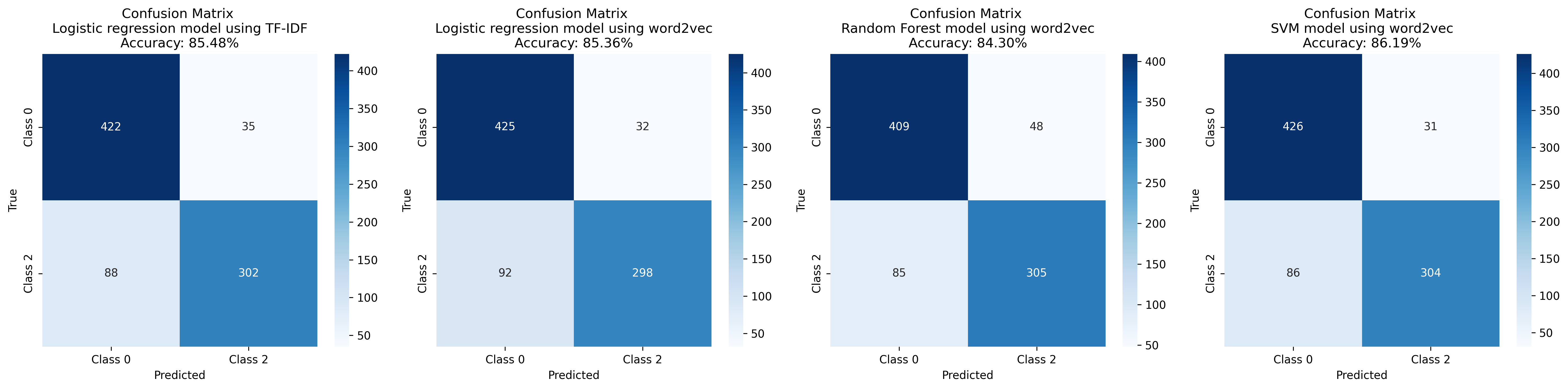
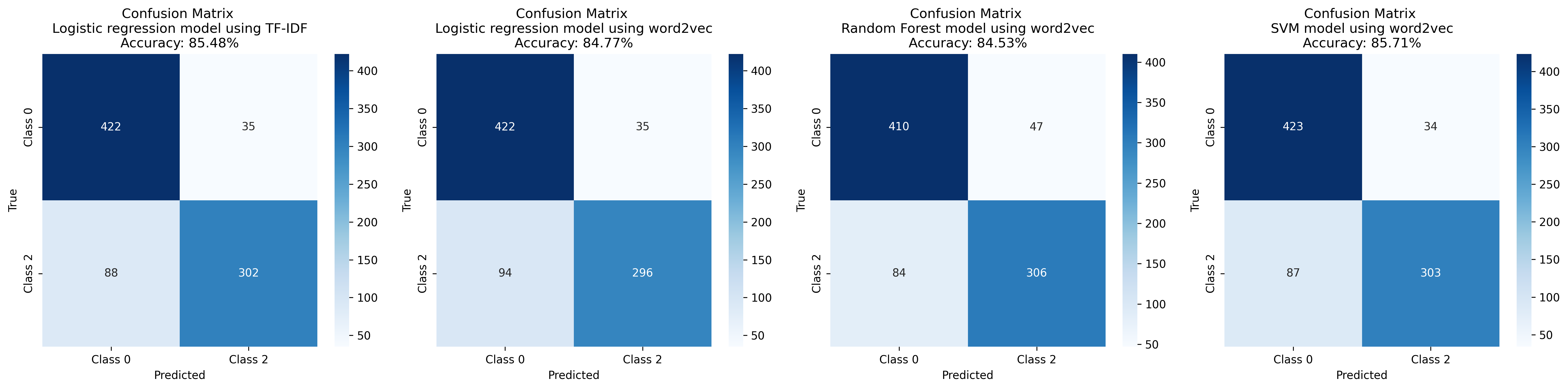
**Result Analysis**:

**Confusion Matrix Heatmap:**

This heatmap vividly depicts the accuracy of our Logistic Regression, Random Forest and best-performing model (SVM with optimized C value) in classifying tweets across different sentiment categories. The diagonal elements, showcasing correct predictions, illuminate the model's strengths. Off-diagonal elements, representing misclassifications, highlight areas for potential improvement.



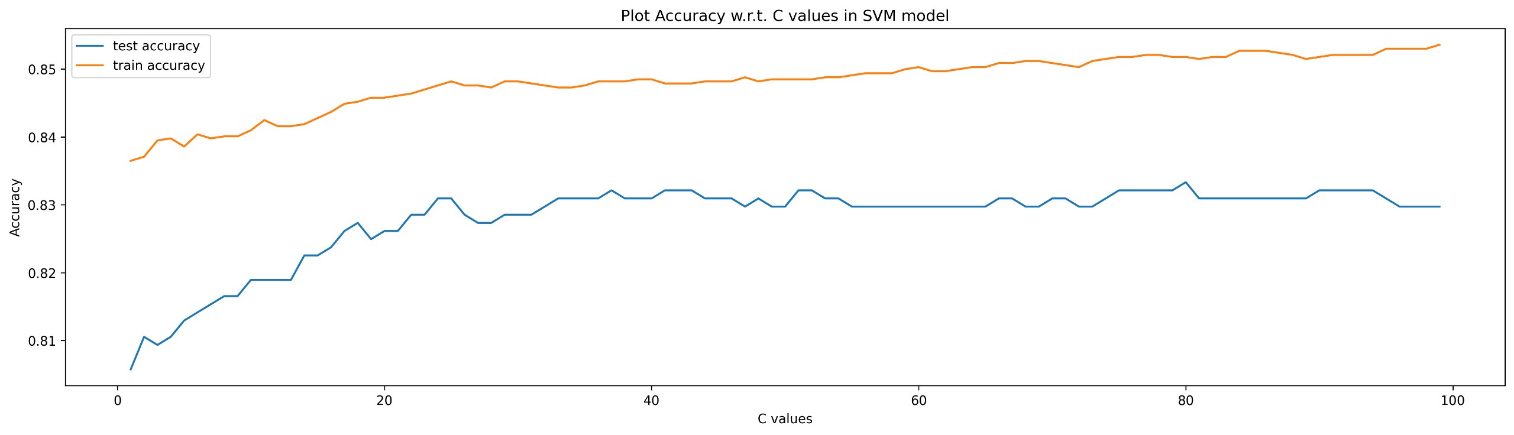
Models where preprocessing done without removing Stopwords



Models where preprocessing done removing Stopwords (Better Performance)

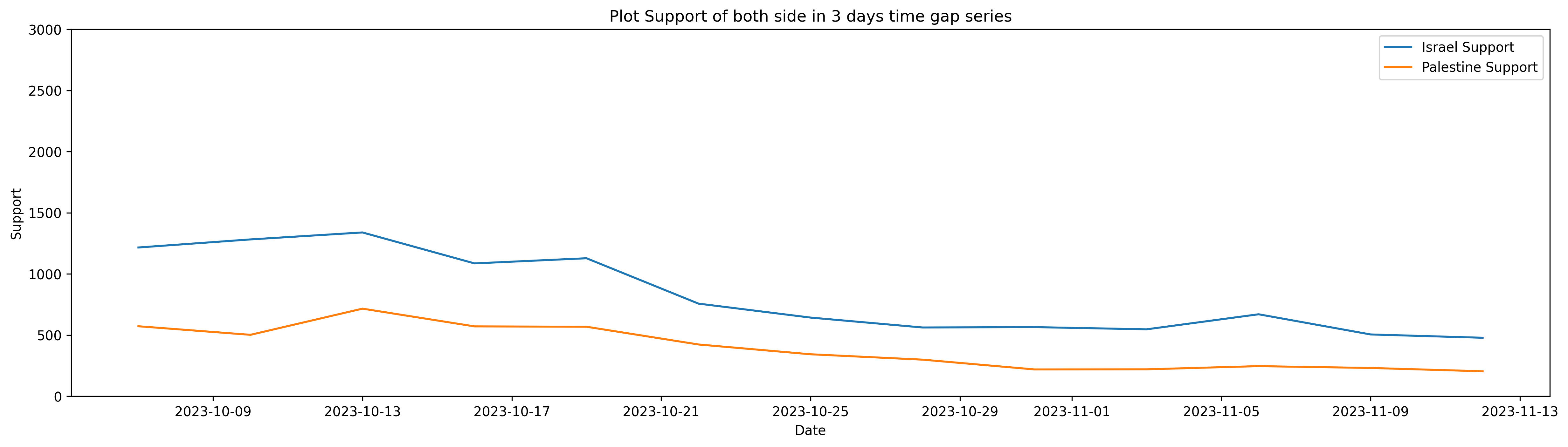
**SVM Accuracy vs C Values:**

This graph unveils the delicate dance between model complexity and performance. By plotting accuracy across different C values, we identified the optimal balance, ensuring the model captures meaningful patterns without overfitting to the training data.



**Support for Israel and Palestine Over Time:**

This captivating plot reveals the ebb and flow of public sentiment throughout the period of analysis. The lines charting support for Israel and Palestine intertwine and diverge, reflecting the dynamic nature of online discourse. By observing these trends, we witness how opinions respond to events, narratives, and shifting media attention.



Graph Showing Relatively more decreasing support of Israel as War goes on. (Although Short Timeline is available here)

**Future Work:**

While this project has shed light on the digital landscape of the Israel-Palestine conflict, it merely scratches the surface of what's possible. Future endeavours can build upon these findings and delve deeper into the complexities of online sentiment:

* **Deep Learning Integration**: Exploring the potential of deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), could further enhance our ability to capture intricate patterns in language and potentially improve sentiment classification accuracy.
* **Emoji and Emoticon Deciphering**: Developing techniques to better incorporate emojis and emoticons into sentiment analysis models is crucial, as these visual elements often convey significant emotional nuances in online communication.
* **Granularity and Nuance**: Refining the analysis by considering factors like user demographics, locations, and specific hashtags can provide a more granular and nuanced picture of online sentiment. This would allow us to understand how different groups and communities express their opinions and how these opinions vary across geographical and social contexts.

By continuously refining our tools and expanding our scope, sentiment analysis has the potential to become a powerful instrument for understanding the evolving dynamics of the Israel-Palestine conflict. Through deeper analysis, informed dialogue, and a commitment to empathy, we can strive towards a future where digital battlefields transform into spaces of constructive engagement and potential solutions.

**CONCLUSION**:

This project ventured into the vibrant online space of Twitter, wielding the powerful tool of sentiment analysis to understand public opinion surrounding the complex and ever-evolving Israel-Palestine conflict. By meticulously analyzing 10,000 tweets, meticulously labeled and preprocessed, we were able to train and evaluate machine learning models to classify the sentiments expressed within them.

Our most successful model, a Support Vector Machine (SVM) with an optimized C value, achieved an impressive accuracy of 86.19%. This demonstrates the ability of sentiment analysis to navigate the nuanced language of online discourse and categorize underlying opinions with remarkable precision.

Beyond static classifications, we delved deeper to track sentiment shifts over time. By grouping tweets into 3-day intervals and applying our best model, we were able to visualize the dynamic changes in support for Israel and Palestine. This unveiled fascinating trends, showcasing how real-world events, media narratives, and even holidays can influence the digital pulse of the conflict.

Our findings hold significant value in promoting a more informed understanding of the Israel-Palestine conflict. By deciphering the online sentiment, we can gain insights into the human emotions, arguments, and narratives that drive public opinion. This knowledge can contribute to more constructive dialogue, potentially paving the way for peaceful resolutions and fostering empathy across divides.

**References:**

**Official Documentation:**

* **RE**: https://docs.python.org/3/library/re.html
* **Gensim Library**: https://www.gensim.org/
* **Scikit-learn Library**: https://scikit-learn.org/
* **Twitter Developer Documentation**: https://developer.twitter.com/en/docs

**Machine Learning Communities:**

* **ACL (Association for Computational Linguistics**): https://www.aclweb.org/
* **TowardsDataScience**: https://towardsdatascience.com
* **AAAI (Association for the Advancement of Artificial Intelligence):** https://aaai.org/

**Websites and Resources:**

* **Stanford CoreNLP**: https://stanfordnlp.github.io/CoreNLP/
* **Sentiment Analysis Toolkit**: https://stanfordnlp.github.io/CoreNLP/sentiment.html