**Group 5**

**Sachini Jayawardena, Rumesh Mohan, Sajith Sanjeewa, Vibhath Algama**

**Determination of political biases of news reporting websites in Sri Lanka using Text Analysis**

**Abstract**

Identifying whether the news we read is biased or not is a difficult task especially in this advanced technological era. This project aims to understand and predict political biases of selected news websites using NLP tools and techniques with a blend of visualization for better and easy interpretation. We used methods of text analysis which gave us deep insights and knowledge to proceed on this research.

**Introduction**

News portals play a vital role in society in many ways. They keep people informed, bring essential topics into public discussions, forums and they gradually change the attitudes of the communities. Media is able to draw attention to particular events while ignoring other important news too. Also, the selection of what to report about a specific event or entity produces biasness [2].

Are sources of news biased or not? This is a difficult question to answer. There is no definitive ‘truth’ to measure against in the reporting of news, only different interpretations or presentations of the facts. As news articles can be written to present either a favorable or unfavorable view of an event, we decided to use text analysis to try to judge whether the news outlets offer a biased view in its reporting of stories, especially stories about political involvement or influence.

Biases in publishing can take two forms, the selection of stories to show and how such stories are phrased. A site opposing the Prevention of Terrorsm Act (PTA), may choose to report stories alleging a link between fundamental rights, and may write those stories with words like “breach”, “insult civil society” etc. A site in favor of the PTA might choose to ignore such claims and may describe using words like “reconciliation”, “strengthen democratic governance”, “ independent oversight”, etc.

Journalists select events, sources and from these sources the information they want to publish in a news article. The selection of news stories, how they interpret, in case of translation and the choice of words in reporting affect the reader’s perception of a particular story if the writer uses words with a positive or a negative connotation to refer to an entity, or by varying the credibility ascribed to the source [3].

**Literature Review**

Media plays a vital role in the life of each individual in a country. In most cases the media is the entity that determines the fate of a country in terms of the government that should rule the country which aligns mostly with the public sentiment. Hence this powerful tool called media can be utilized with various agendas for political and personal gains of the people who are part of it. News of a certain topic can be reported in different perspectives by different news reporting media via their broadcasting medium like websites etc. Bias is commonly understood as inclination or prejudice towards a certain point of view. A discourse or text that has a bias may have a certain agenda or promote a certain ideology.

There can be situations where political biases are involved with the reporting approach of these websites as it can be perceived by the public in different ways based on the news reporting websites. For example, these news reports will be subjected to biases made by the news channel thus changing the perspective of the news article in favor of a political party, political affiliations or a respective group of people. This project’s main approach is to collect news content from various different online news websites and then consolidate them for the determination of news biases. According to the News Literacy Project, news bias is usually “incidental and debatable rather than intentional and overt”.

**Table 1: Common types of bias in news reporting.**

| **Types of Bias** | **Definition** |
| --- | --- |
| **Partisan** | News reporting unfairly favors one political party, group or viewpoint. |
| **Demographic** | News organizations and journalists affect how they present or write an article, when the newsroom staff is not diversified. |
| **Corporate** | When the parent company or owner affects the presentation of news stories. |
| **Neutrality** | “False Balance”, when something is demonstrably known but the news articles attempt a neutral stance. |
| **Big story** | Focus too much on a specific topic. |
| **Ideological** | Promote a specific opinion on a topic. |
| **Spin** | Attempts to create a memorable story. |
| **Coverage** | Concern with the visibility of topics or entities such as person or country. |
| **Gatekeeping/Selection/Agenda** | Which stories are selected or rejected for reporting. |
| **Statement/ Presentation** | How articles choose to report on concepts. |

**Methodology**

* **Data Collection (Planning)**

We collected news text as data from two news websites which are famous in Sri Lanka and have a frequent to significant amount of viewers/readers everyday. The popularity of these news websites were interpreted and filtered out based on their reputation, years since publication established, amount of followers in social media, newspaper sales and TRP (if channel exists), etc. As websites carry a large amount of news on various topics it is practically impossible to scrape and process each and every news for studying their biases. Hence we thought of selecting three topics which have ample amount of news from these websites that are not too old as well as latest but on events that were held within a three years period within Sri Lanka considered as politically important, socially trending and critically acclaimed which had significant effects on the day to day life of a Sri Lankan. Thus we have decided to select the following three topics and two websites as mentioned below for this study.

**Selected Topics**

1. **Anti-Terrorism Act/Prevention of Terrorism Act (PTA)**
   * The government of Sri Lanka has gazetted a new Anti-Terrorism bill on 22nd of March 2023. This bill is claimed to amend the anti-terrorism bill that already exists but from the public perspective it is considered to be a sweeping attack on the basic democratic rights of the working class, youth and the rural poor.
2. **X-Press Pearl Ship (XPS) Cargo Vessel Sinking**
   * This is one of the worst maritime disasters that occurred in the world. The incident took place along the coastal region of Colombo in July 2021. Even though this occurred two years ago its effects and damages caused are still prevalent and the reason why it is still trending is the bribing of top officials in order to prevent the compensation that has to be paid by the ship owners for the damage caused to the maritime environment and its stakeholders.
3. **Local Government Election (LGE)**
   * The local government elections of Sri Lanka were planned to be held on March 9th 2023. The president of Sri Lanka who is also the finance minister has postponed it indefinitely due to the economic crisis and declaring the priority for economic recovery over the expenditure for elections.

All three topics selected have been under the limelight of the media over the past few weeks which are critically acclaimed and involve various stakeholders in Sri Lanka.

**Table 2: Description of websites selected for predicting the biases.**

| **News Website** | **Publication (Year)** | **Language/**  **Release Type** | | **Social Media Followers** |
| --- | --- | --- | --- | --- |
| **Ada Derana** | **Power House Ltd (2007)** | **English/Daily** | | **2M+ Followers** |
| **News First** | **Capital Maharaja Organization Ltd (2003)** | **English/Daily** | | **2M+ Followers** |

* **Data Collection (Execution)**

We crawled through the selected two websites for various other web pages that are linked to it using the Requests library and Beautiful Soup library. Then we scraped the relevant and required information from these websites using the Newspaper library for this research purpose. The Newspaper internally uses NLTK libraries and their modules. The Newspaper library uses article function to perform the scraping by downloading, parsing and processing each topic of the selected news website. Later these data are passed to a data frame for preprocessing them for other analytical purposes. The data frame was divided into following six columns to categorize the scraped data as **Article Title, Text, Summary, Keywords, Channel and Topic**. We followed this procedure for all 2 websites and corresponding 3 topics respectively.

**Table 3: Summary of article selection from each website.**

| **News Portal** | **Local Government Election (LGE)** | | **Prevention of Terrorism Act (PTA)** | | **X-Press Pearl Ship (XPS)** | |
| --- | --- | --- | --- | --- | --- | --- |
| **Number of Articles** | **Selected Period** | **Number of Articles** | **Selected Period** | **Number of Articles** | **Selected Period** |
| **Ada Derana** | **122** | **04-Sep-2021 To**  **14-May-2023** | **159** | **03-Jan-2020 To 03-May-2023** | **85** | **21-May-2021 To 12-May-2023** |
| **News First** | **148** | **18-Mar-2021 To 09-May-2023** | **47** | **14-May-2021 To 24-April-2023** | **211** | **21-May-2021 To 16-May-2023** |

* **Data Preprocessing**

Text Preprocessing is a crucial step in NLP where we clean the text data in order to convert it into a presentable form that is analyzable and predictable. Text preprocessing techniques are utilized for training the machine learning and AI models for better understanding of the data we are closely working with. The following are the procedures we followed during the preprocessing to enhance the collected data to be used for further analysis.

**Converting text into lowercase ➜ Removing HTML tags ➜ Removing special/accented character ➜ Converting numbers into words/removing numbers ➜ Expanding contractions ➜ Removing white spaces ➜ Word Tokenization ➜ Correcting words/repeating characters ➜ Lemmatization ➜ Removing stop words, sparse terms, and particular words.**

* **Tools Utilized**

**Table 4: Description of main tools used throughout the project**.

| **Tool Name** | **Usage** |
| --- | --- |
| **Beautiful Soup Library** | To pull and save content from the website, clean and parse them as required |
| **Matplotlib Library** | To create interactive, static and animated visualization in python |
| **Newspaper Library** | Used to extract and process data by curating and scraping news articles and online news websites allowing to retrieve data based on date, time, author and article title |
| **NLTK Library** | Used for working with NLP in python by providing various text processing libraries |
| **Requests Library** | Execute HTTP operations against a specific web server specified by its URL |
| **Scikit-Learn Machine Learn Library** | Efficient tool for predictive data analysis that allows implementing various machine learning models like classification, regression, clustering and statistical tools for analyzing them. |
| **Seaborn Library** | Data visualization for ML and other data science tasks |
| **Text Normalizer Library** | Normalizes the text to standardize it, reducing the amount of information the computer has to work with hence increasing the efficiency of the analysis |

We need to clean the scraped data to minimize the issues in the clarity and quality of text data and unusable features in the text.

**Unusable features in Text**

**Stop Words**

* Stop words make topical and grammatical sense when used in a sentence, but add no impact on determining the overall sentiment behind it. This includes words such as ‘the’, ‘is’, ‘and’, ‘I’, ‘you’, ‘at’, ‘be’, ‘by’, ‘for’, ‘from’ etc. The goal of cleaning the data should be to make the training data as meaningful as possible. Stop words do not contribute to the underlying sentiment of a sentence and are classified as noise for sentiment classification. Each data instance should have a majority of words that can indeed contribute to determine sentiment.

**Special Characters and Numbers**

* Special characters and numbers are unreliable indicators of sentiment in text based data.

**Negations**

* This includes words such as ‘not’, ‘un-’, ‘non-’. It is important to appropriately account for the impact of negations on polarity.

**Vagueness**

* Not all text contains the presence of highly polarized statements. This can add unnecessary features to a classifier in the form of vectors.

The following steps have been taken to clean the data to make it appropriate.

**Text Lowercasing**

* Common approach to lowercasing everything for the sake of simplicity. It maintains the consistency flow during the NLP tasks.

**Remove Stop Words**

* We have used the ‘nltk stopwords’ library which provides a list of stop words that are irrelevant to sentiment analysis. Each sentence of the input text is processed and the words matching in the ‘nltk stopwords’ library are removed.

**HTML Stripping**

* News scripts in the portals are in the form of HTML documents. We do not want any embedded javascript or CSS and filter tags and texts.

**Contraction Expansion**

* Contractions are words or combinations of words which are shortened by dropping letters and replaced by an apostrophe. This will expand the word as two separate words. Ex: let’s -> let us, we’ve -> we have.

**Accented Character Removal**

* Remove extra characters (Ex: Â, â) which represent accents. Accent characters do not affect the word’s meaning. If we don’t remove these symbols, the corpus will see words with and without an accent symbol as the different words while it is the same word.

**Remove Numbers/Numeric Characters**

* Numbers in the text have no significant meaning hence they are removed.

**Stemming and Lemmatization**

Figure 1: Summary of Stemming and Lemmatization.

* We have used lemmatization in this text normalization process. Lemmatization performs normalization using vocabulary and morphological analysis of words. Lemmatization aims to remove inflectional endings only and return the base or dictionary form of words. Lemmatizing makes much more sense than stemming. 

**Special Character Removal**

* In python strings are immutable which cannot change its content. This creates a new string with characters from the original string. String has been modified and unwanted characters (Ex: €,œ,™) have been deleted.

**Solution Approach**

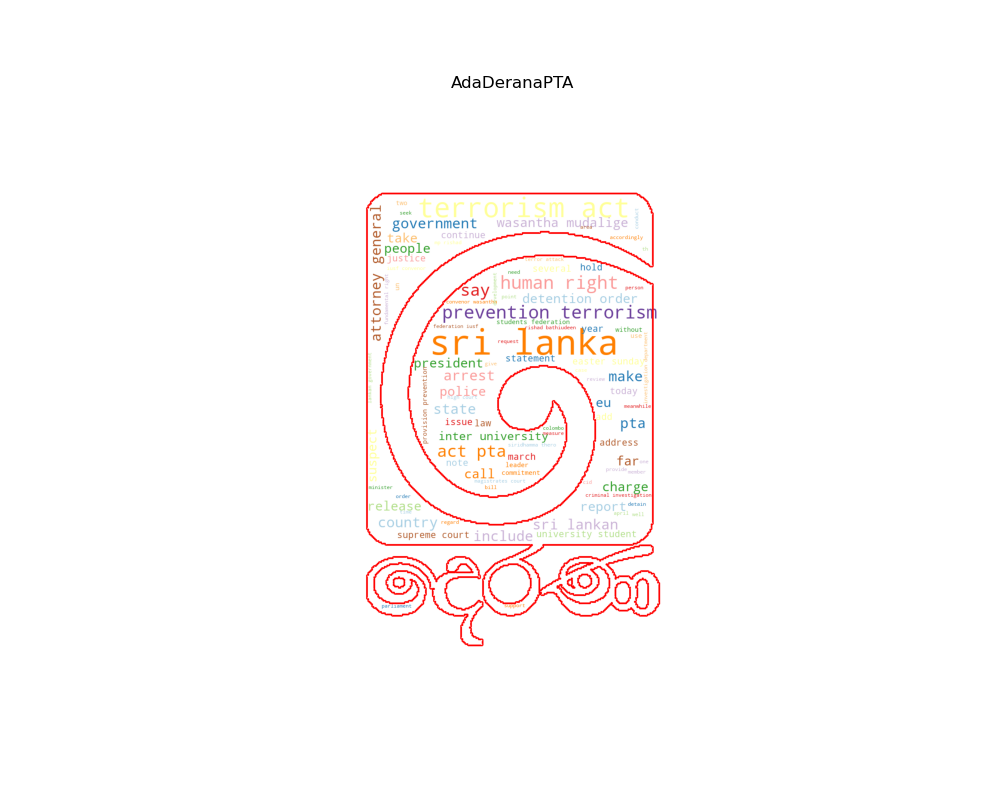
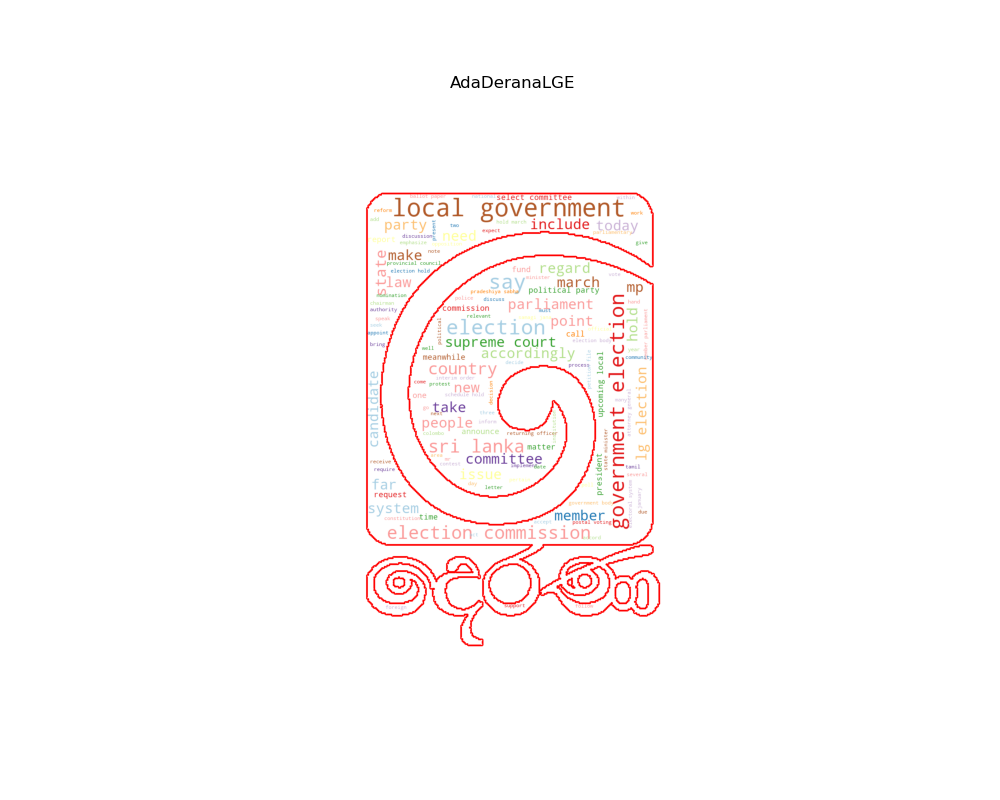
* **Bias Recognition** **:** Recognize the biased words or phrases from the news article.

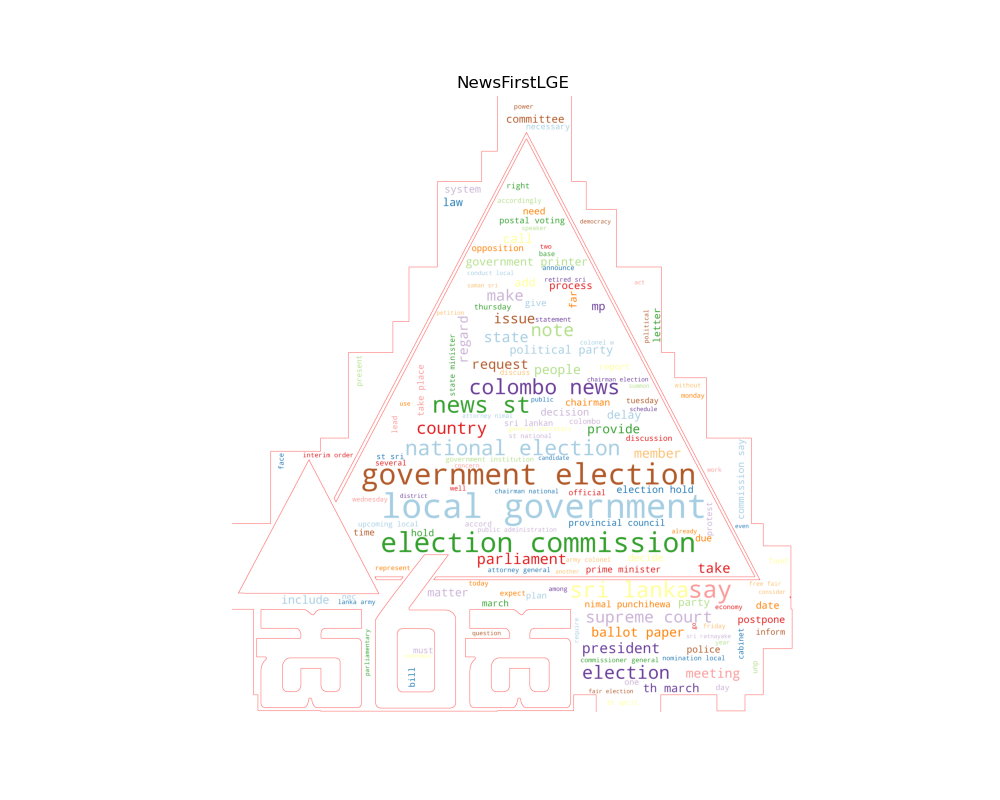
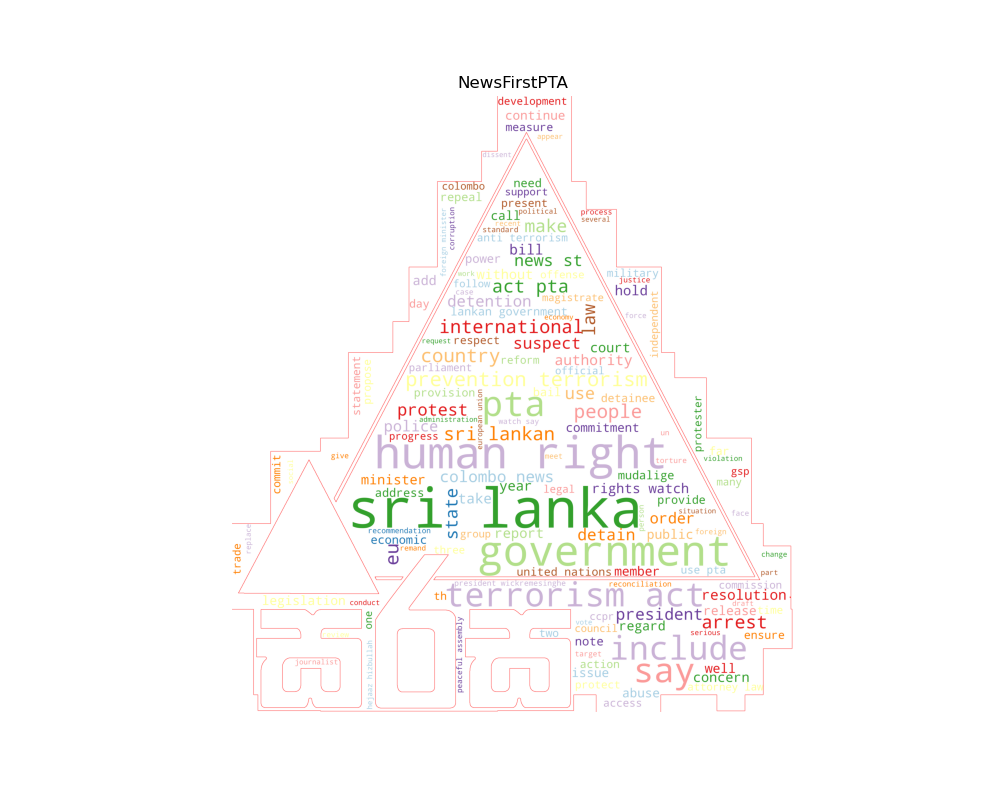
Bias recognition is to annotate the biased words or phrases in the news articles, indicating that each sentence is towards pro-government (1) or anti-government (0). The news article **“SJB’s P. Harrison pledges to support President Ranil’s re-election”,** for example, has been classified as a pro-government article (1). This news proposes instead of holding the Local Government Election, Presidential Elections can be held which favors the government opinion.

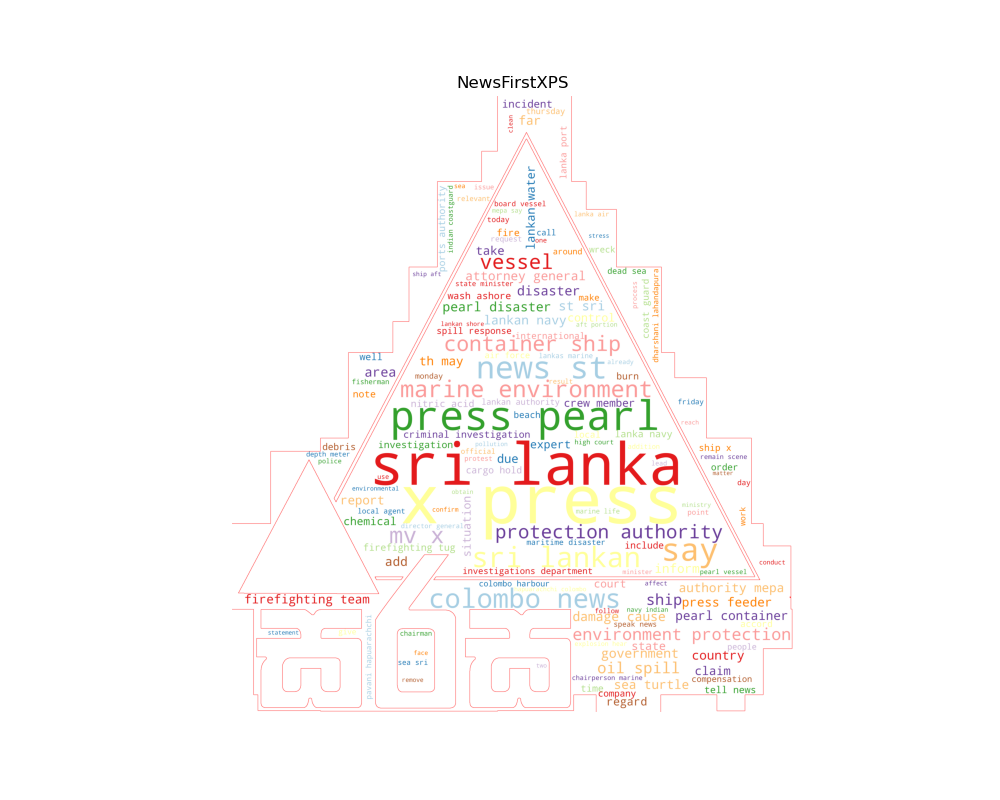
* **Bias Detection :** Detects whether a news article is biased or not.

Bias detection is to predict when reporting the selected news topics are pro-government or anti-government by analyzing each sentence.

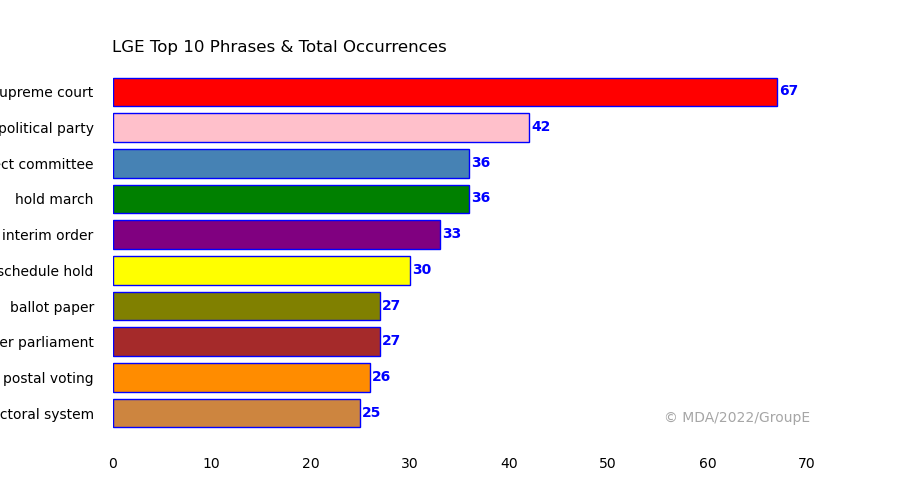
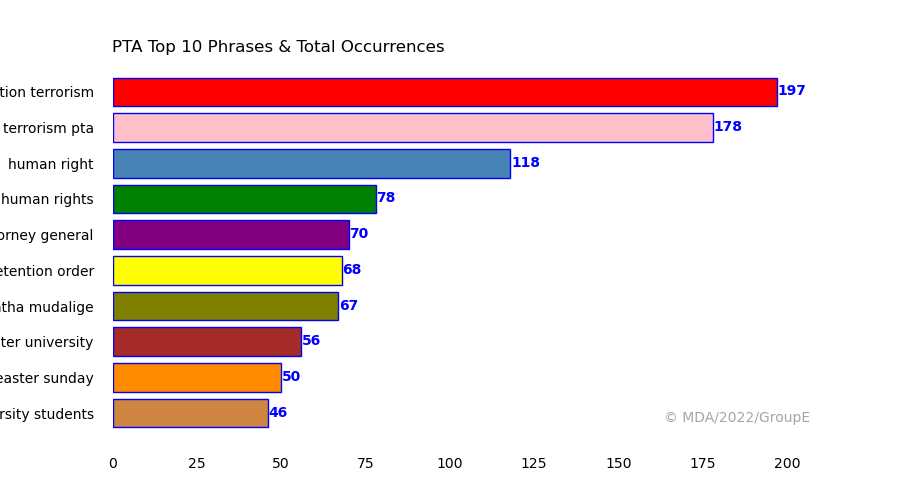
**Visualization**

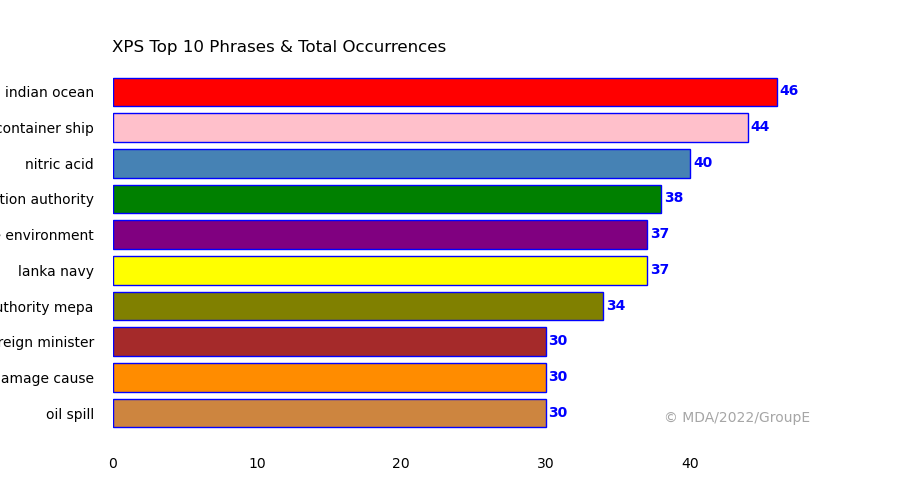
**Figure 2: Word Cloud(s) of text from the news articles**



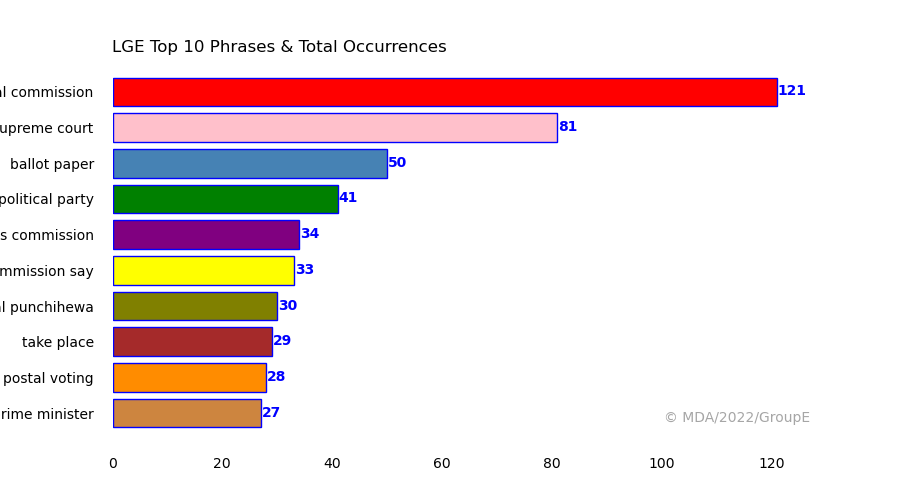
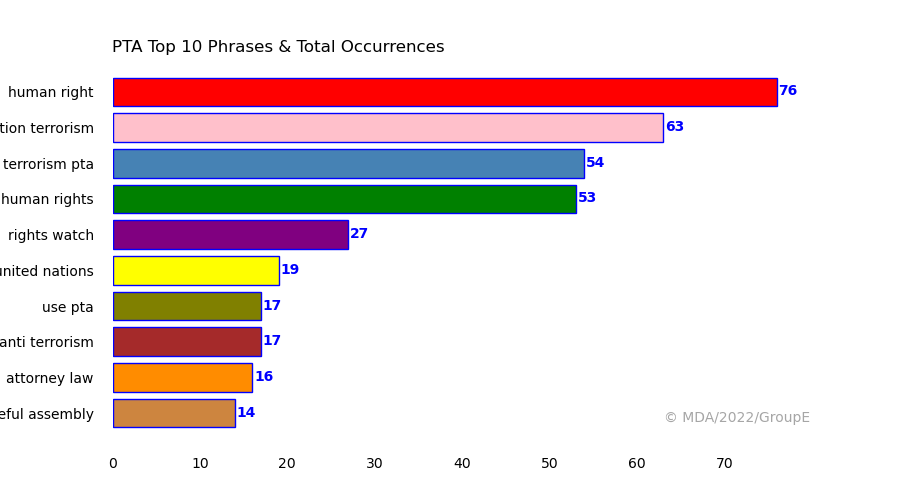


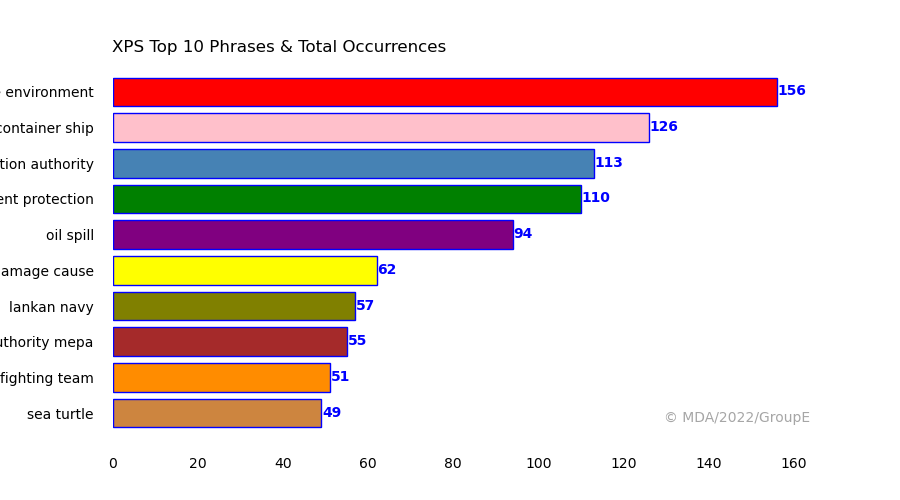
**Figure 3: Bar charts for top 10 Bigram frequencies for news text (AdaDerana)**





**Figure 4: Bar charts for top 10 Bigram frequencies for news text (NewsFirst)**





We have approached three main processes to fulfill this project.

1. **Sentiment Analysis using VADER**

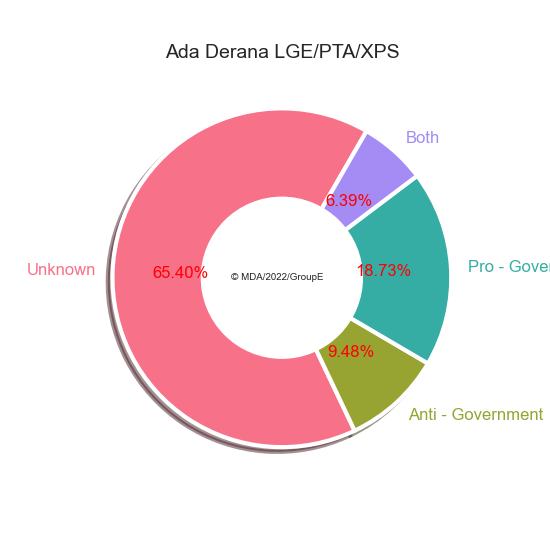
Sentiment analysis is basically defined as the process of identifying and categorizing opinions from a piece of text, thereby determining whether the opinion is positive, negative or neutral. This is also known as opinion mining. Most of the previous studies on finding news bias have used the polarity based on sentiment analysis. Those studies discussed the biases based on this polarity of the text data which is positive, negative or neutral.

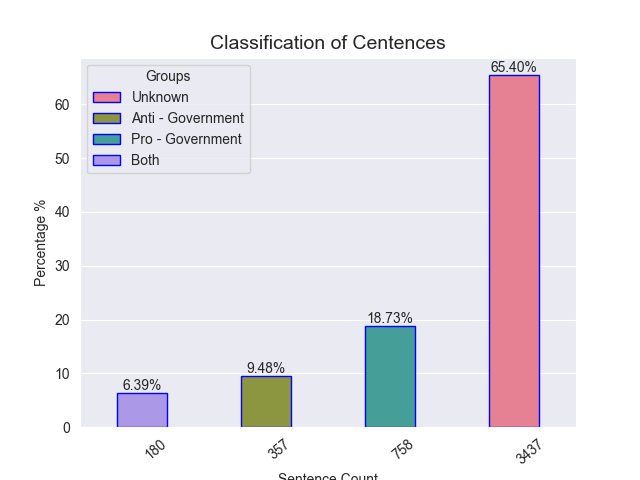
**Table 5: Polarity for the news corpus**

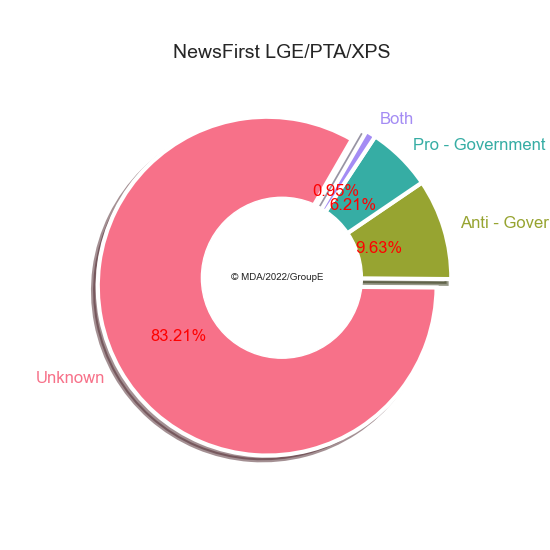
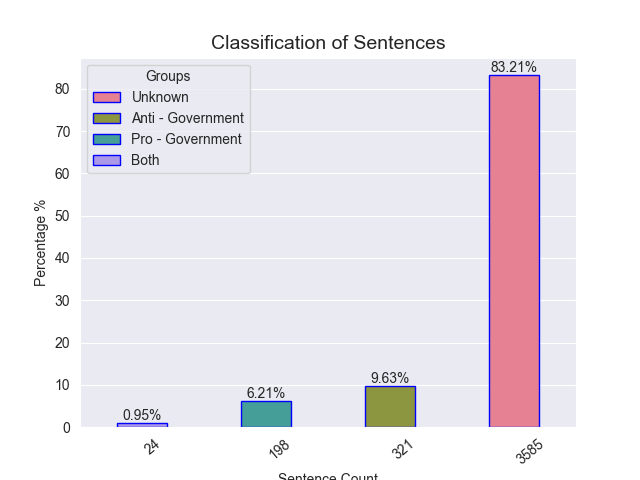
| **%** | **Ada Derana** | | | **News First** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **LGE** | **PTA** | **XPS** | **LGE** | **PTA** | **XPS** |
| **Sentence was rated as Positive** | **15.2** | **11.5** | **12.0** | **12.7** | **13.8** | **7.8** |
| **Sentence was rated as Negative** | **7.2** | **18.2** | **13.1** | **7.3** | **20.5** | **12.9** |
| **Sentence was rated as Neutral** | **77.6** | **70.3** | **74.9** | **80.0** | **65.8** | **79.2** |
| **Sentence Overall Rate** | **Positive** | **Negative** | **Negative** | **Positive** | **Negative** | **Negative** |

1. **Define a set of Pro-government and Anti-government key words and identify the biases.**

This examines the number of words from sentences that appear in the defined Pro-government list and in the Anti-government list. If a sentence described with Pro-government vocabulary, it is towards Pro-government and if a sentence described with Anti-government vocabulary, it is towards Anti-government. If a sentence has both vocabulary it will be both. If there is no such vocabulary, it is unknown.

**Figure 5: Four states of sentences (AdaDerana)**



**Figure 6: Four states of sentences (NewsFirst)**

1. **Designing Sentiment Analysis Model using Classification Machine Learning Models.**

Machine learning approach that looks at previously labeled data in order to determine the sentiment of never seen text data. It involves training a model using previously processed text to predict or classify the sentiment of the new input. In this approach we combined all three news topic corpus. Then, pre-labeled all the cleaned news items as either Pro-government (1) or Anti-government (0). Later we vectorized the cleaned text using **TfIdfVectorizer** to convert text into numbers a computer could interpret. After that we built six different models using machine learning algorithms. Each one was fed a list of extracted features, the words, and each label which is the process of training the algorithm. In order to test the algorithms, we used default data sizes for training and testing.

**Steps:**

1. **Get labeled data:** The labeled data consists of cleaned text and labels.
2. **Convert labeled data to encoded data:** Convert data to numeric form by encoding for computation.
3. **Create feature set:** Features are extracted using the TfidfVectorizer module. Features are the characteristics of the text whereas labels are the class of the text with those features (Pro-Government or Anti-Pro-Government).
4. **Split the data for train and test:** The data set is split for training and testing (test\_size=0.25,random\_state=1).
5. **Train classifiers:** Train different classifiers with the training data.
6. **Test classifiers:** Test the classifiers with the testing data.
7. **Evaluate:** Evaluating the classifiers with confusion matrix and its evaluation metrics, accuracy, precision, recall and f1 value.

**The Dataset**

The dataset is composed of the text of 772 articles taken in the period of 2020-2023. Each article is labeled among one of the two classes:

1. **1 => Pro-government:** These articles are from two portals which show some biases towards the government's opinion.
2. **0 => Anti-government:** These articles are from two portals which show some biases towards opposing the government’s opinion.

The main goal is to classify the text of an article as one of these classes.

**Machine Learning Models - Classification**

Normalized data are used for training classification machine learning models to predict the biases of the news articles. We trained and evaluated six different algorithms and compared the results. Each model we defined random state value into 0 and trained data set splitting random state value into 1 for achieving the same accuracy for every iteration. These models are used in previous similar studies.

1. **Logistic Regression**

We use Logistic Regression to learn a model to predict the probability of each document belonging to a given class. This model learns a set of weights for each feature and uses these weights to make predictions based on the input features. We use 1x10^9 as C value, a high value of C tells the model to give more weight that trust this training data a lot. This also says data may be fully representative of the real world data. Default ‘lbfgs’ (limited-memory BFGS) solver used which supports multinomial loss and works great in most situations.

1. **Random Forest Classifier**

This model works by generating several decision trees for classifying the inputs. We used 50 trees in the forest.

1. **Linear Support Vector Classification**

Vectors are lists of numbers which represent a set of coordinates in some space. So, provided we can find vector representations which encode as much information from the news text as possible.

1. **Multinomial Naive Bias Classifier**

One of the common models used in machine learning algorithms for text classification, Natural Language Processing. It assumes that all the features are uncorrelated from each other and uses the Bayes theorem principles to calculate the results.

1. **K-Nearest Neighbors**

This used to be classified by finding the K nearest matches in training data and then using the label of closest match to predict. Value of K is usually specified by us.

1. **Gradient Boost**

Gradient boosting algorithms are more effective at classifying complex data sets. Features of the data set are the variables used to solve the equation and another part of the equation is the label. We used the default 100 number of boosting stages to perform Gradient Boosting. Larger numbers usually result in better performance. The maximum depth limits the number of nodes in the tree.

**Results**

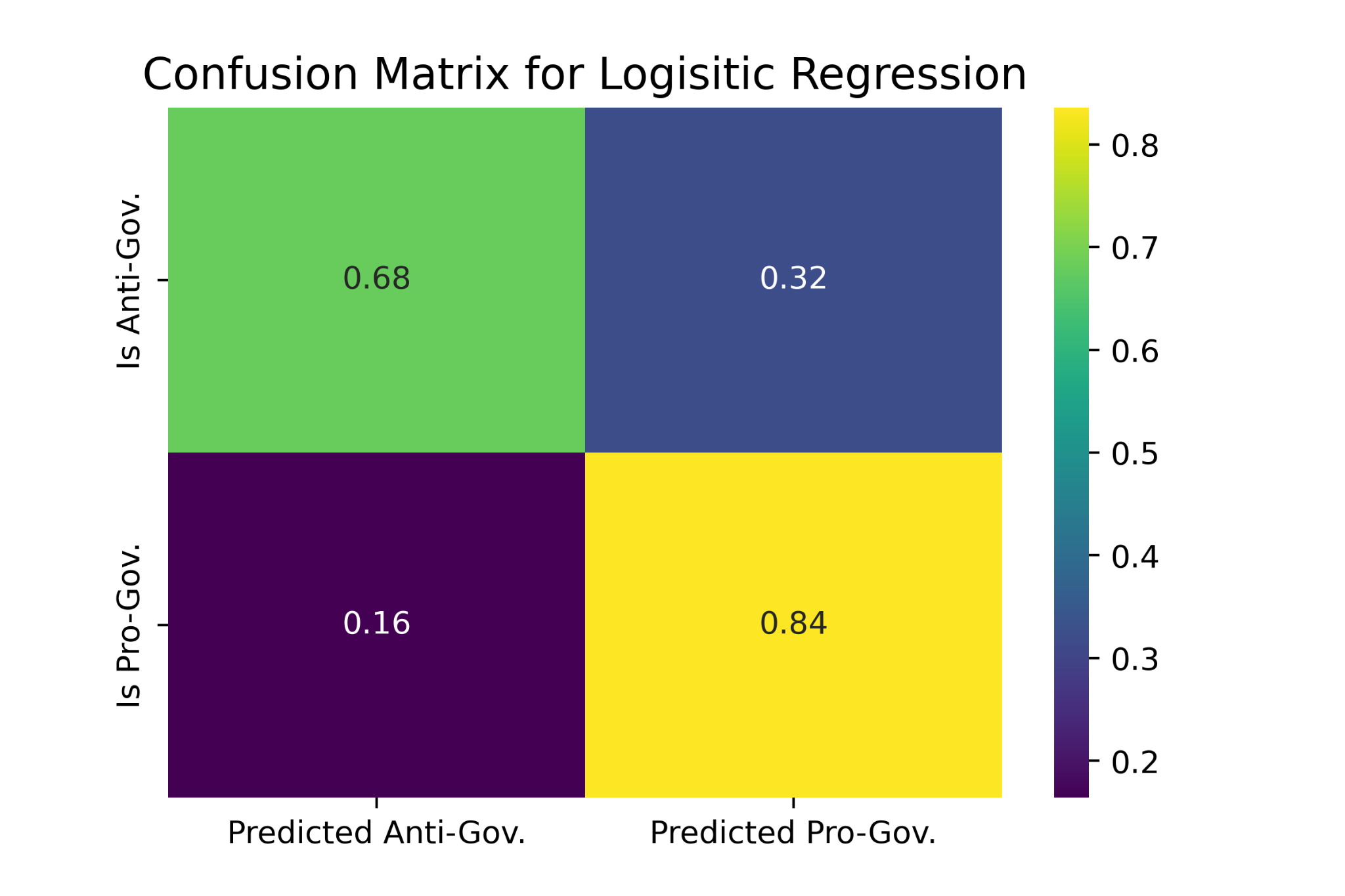
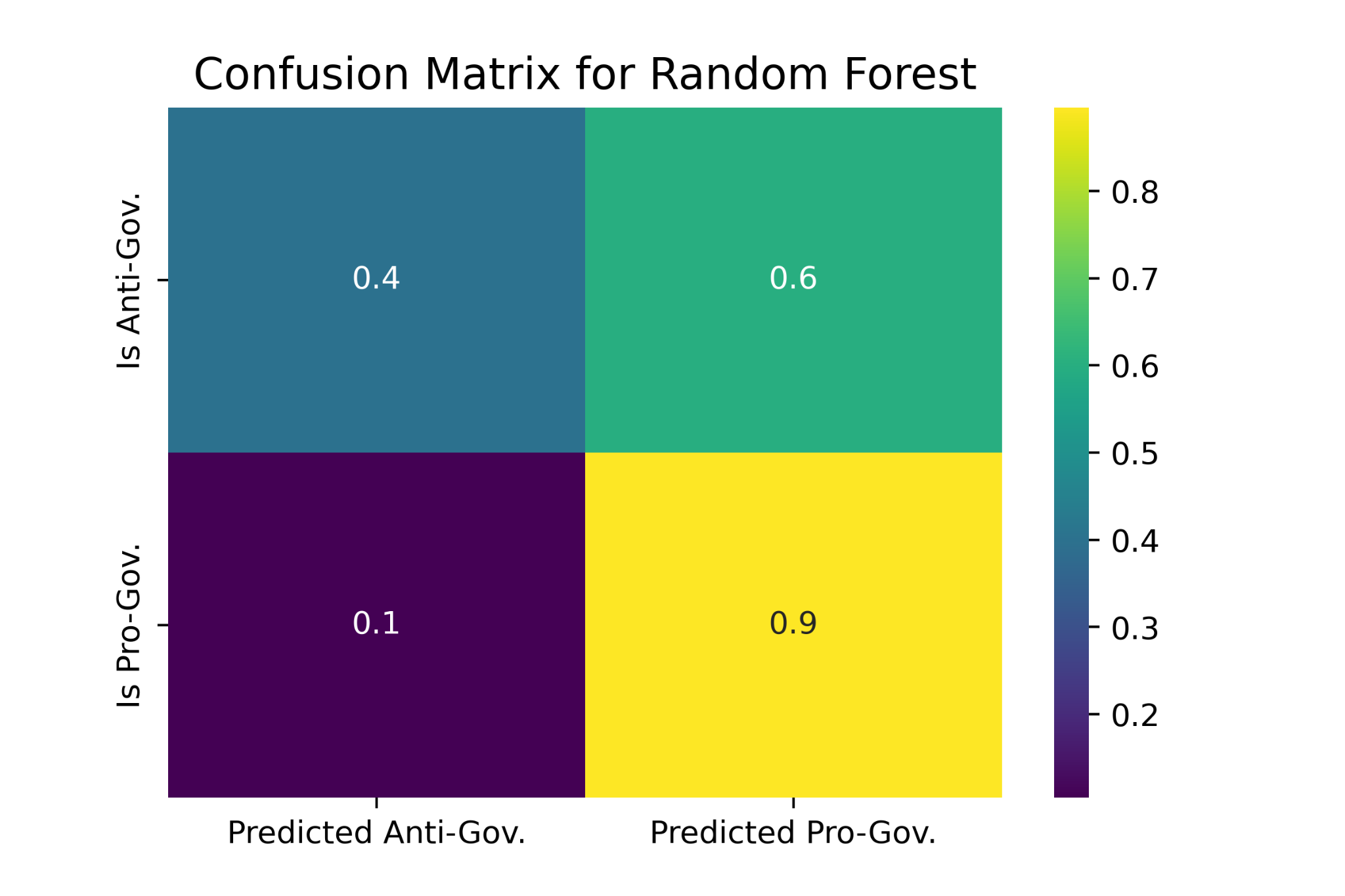
**Table 6: Classification results for AdaDerana portal and selected three news topics.**

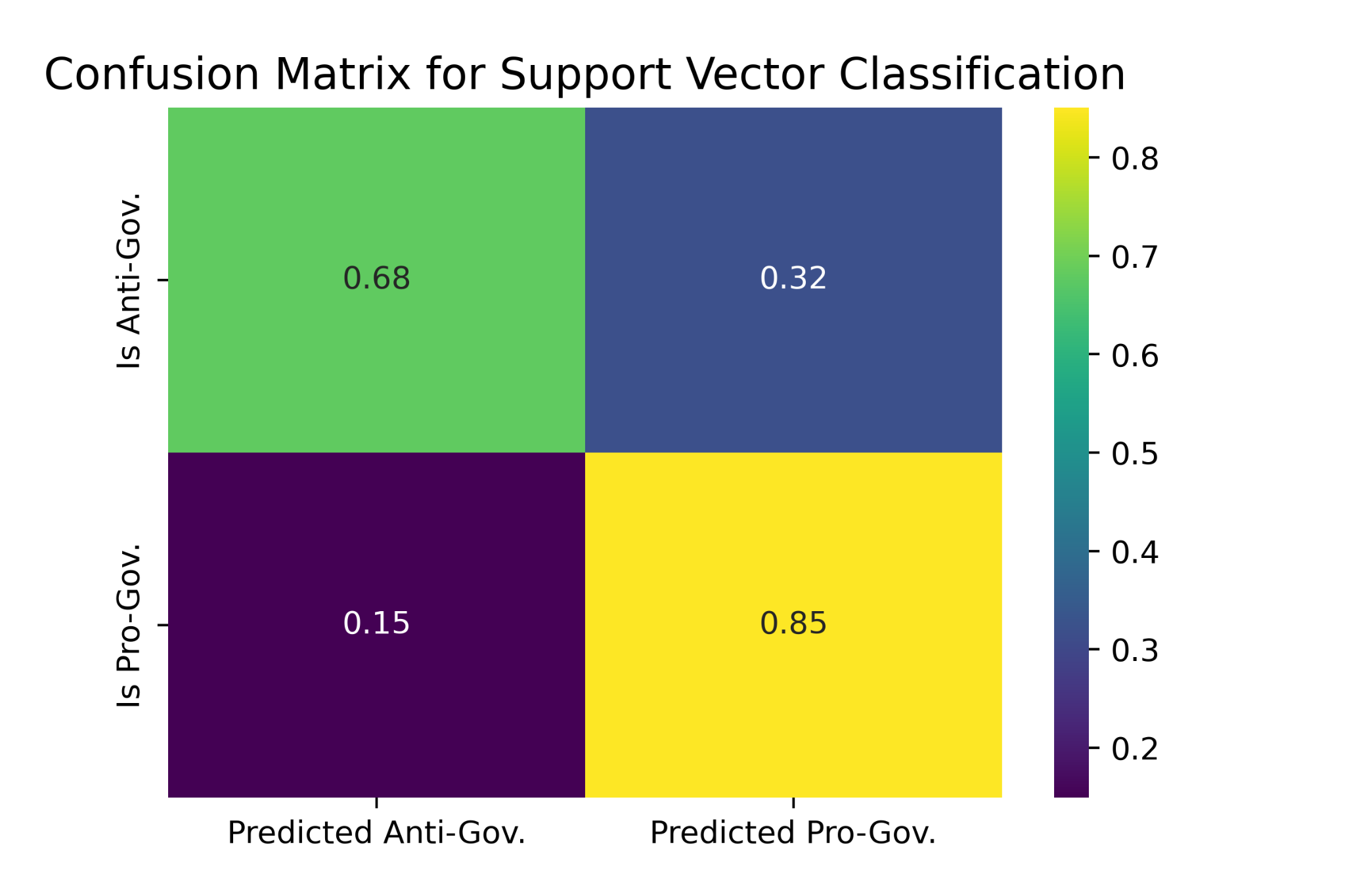
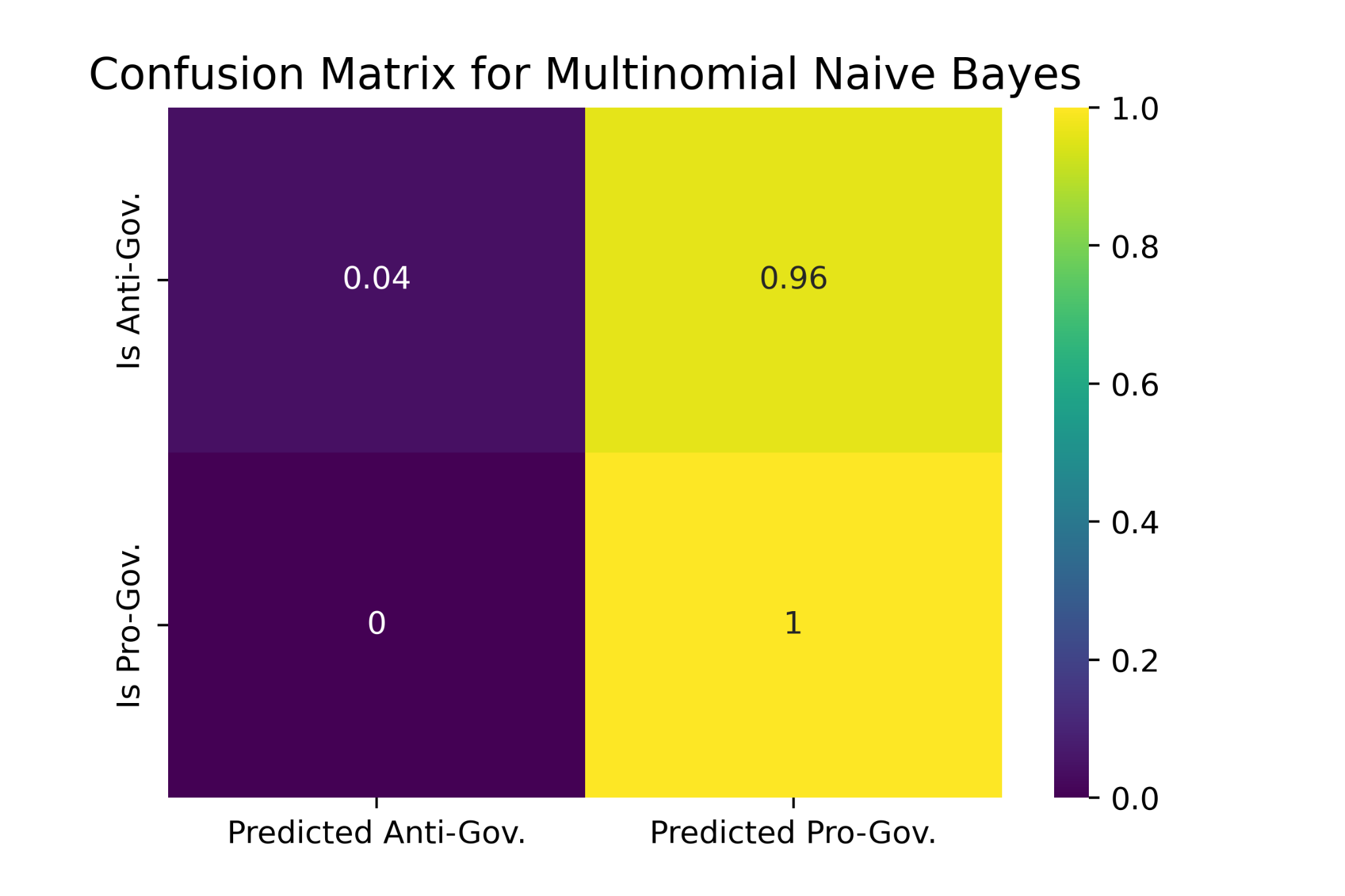
| **AdaDerana** | | **Classification Models** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Logistic Regression** | **Random Forest** | **Support Vector Classification**  **(CVS)** | **Multinomial Naive Bayes** | **K-Nearest Neighbour** | **Gradient Boosting** |
| **Classifier Accuracy (%)** | | 83.0 | 83.0 | 84.0 | 82.0 | 82.0 | 82.0 |
| **Anti-**  **Government** | **Precision** | 0.56 | 0.61 | 0.6 | 0.75 | 0.54 | 0.57 |
| **Recall** | 0.68 | 0.42 | 0.63 | 0.16 | 0.63 | 0.42 |
| **F1- Score** | 0.62 | 0.5 | 0.61 | 0.26 | 0.58 | 0.48 |
| **Pro-**  **Government** | **Precision** | 0.91 | 0.86 | 0.90 | 0.82 | 0.9 | 0.86 |
| **Recall** | 0.86 | 0.93 | 0.89 | 0.99 | 0.86 | 0.92 |
| **F1- Score** | 0.88 | 0.89 | 0.90 | 0.89 | 0.88 | 0.89 |

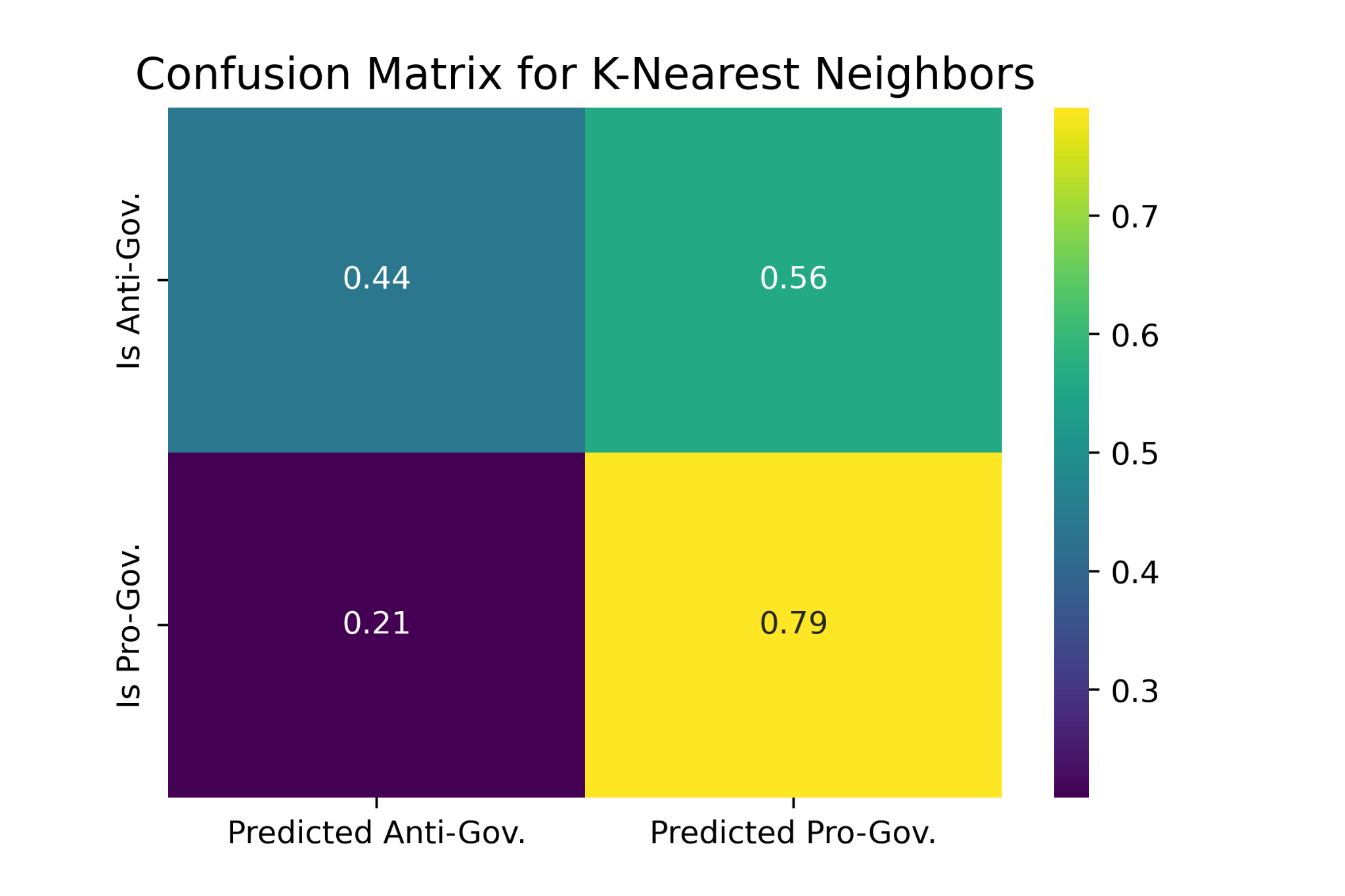
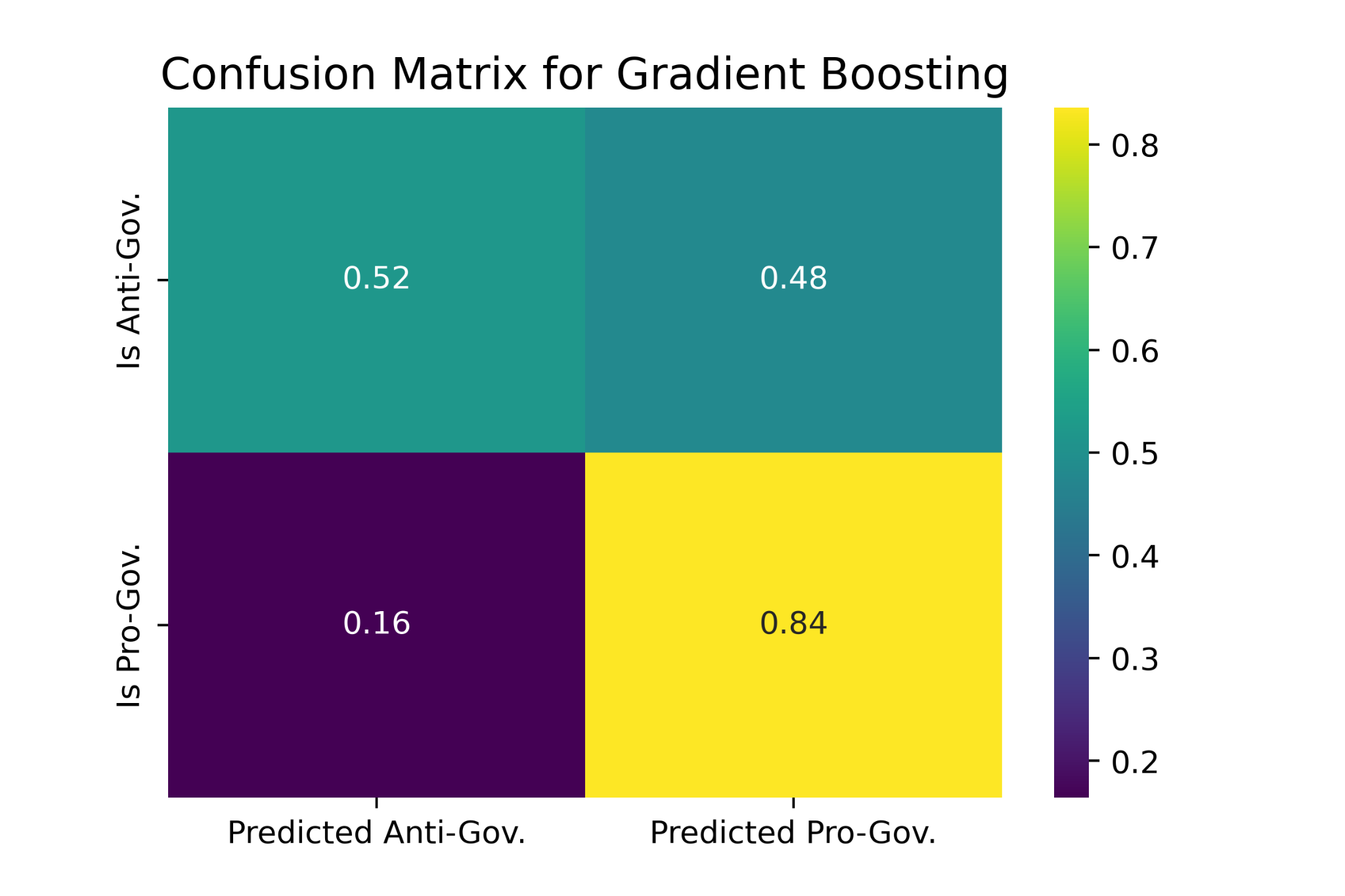
**Table 7: Classification results for NewsFirst portal and selected three news topics.**

| **NewsFirst** | | **Classification Models** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Logistic Regression** | **Random Forest** | **Support Vector Classification**  **(CVS)** | **Multinomial Naive Bayes** | **K-Nearest Neighbour** | **Gradient Boosting** |
| **Classifier Accuracy (%)** | | 70.0 | 70.0 | 73.0 | 66.0 | 70.0 | 74.0 |
| **Anti-**  **Government** | **Precision** | 0.74 | 0.69 | 0.74 | 0.65 | 0.74 | 0.75 |
| **Recall** | 0.8 | 0.94 | 0.88 | 1.0 | 0.82 | 0.88 |
| **F1- Score** | 0.77 | 0.8 | 0.80 | 0.79 | 0.77 | 0.81 |
| **Pro-**  **Government** | **Precision** | 0.59 | 0.71 | 0.68 | 1.0 | 0.6 | 0.69 |
| **Recall** | 0.51 | 0.27 | 0.46 | 0.05 | 0.49 | 0.49 |
| **F1- Score** | 0.55 | 0.39 | 0.55 | 0.10 | 0.54 | 0.57 |

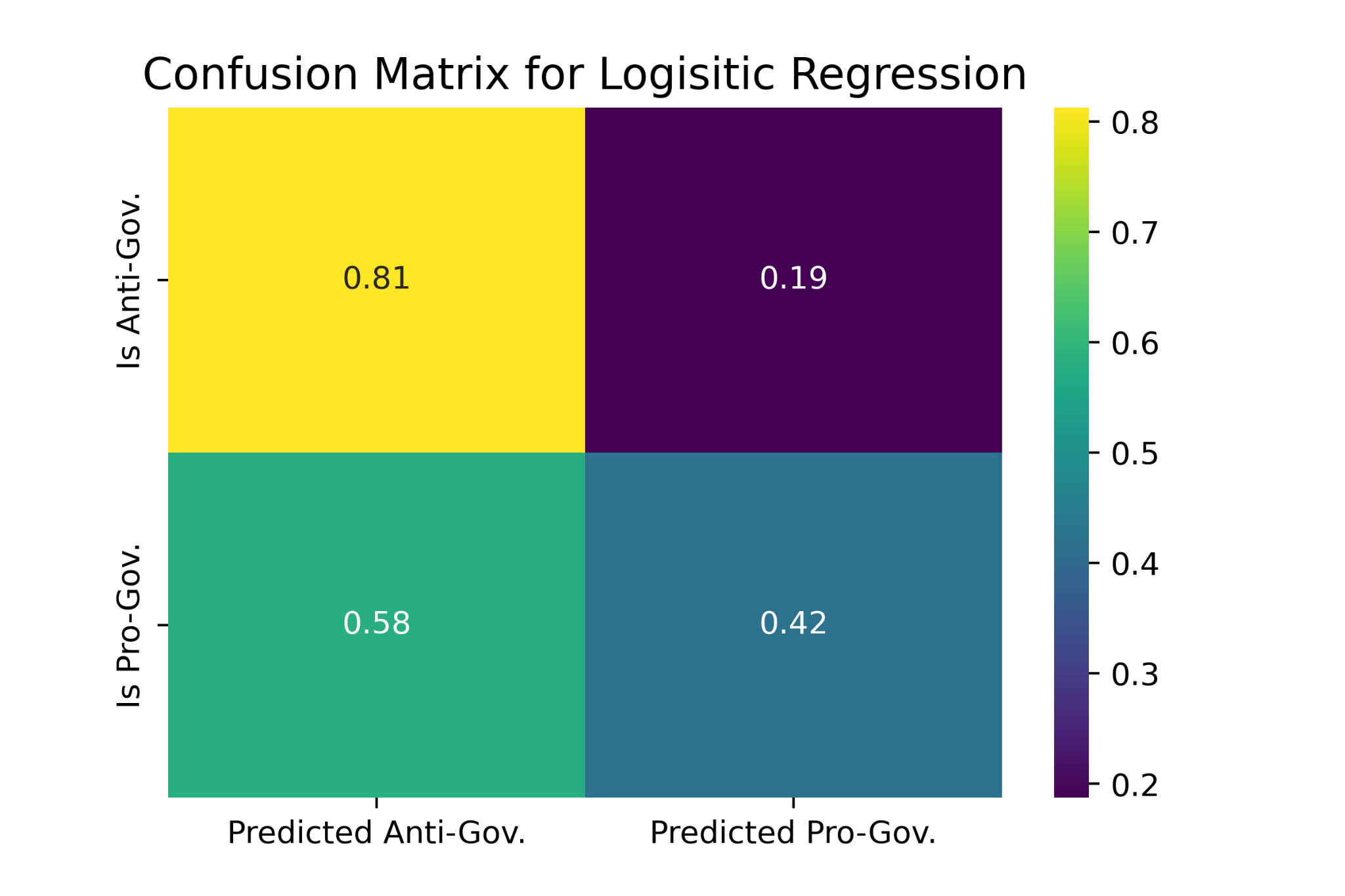
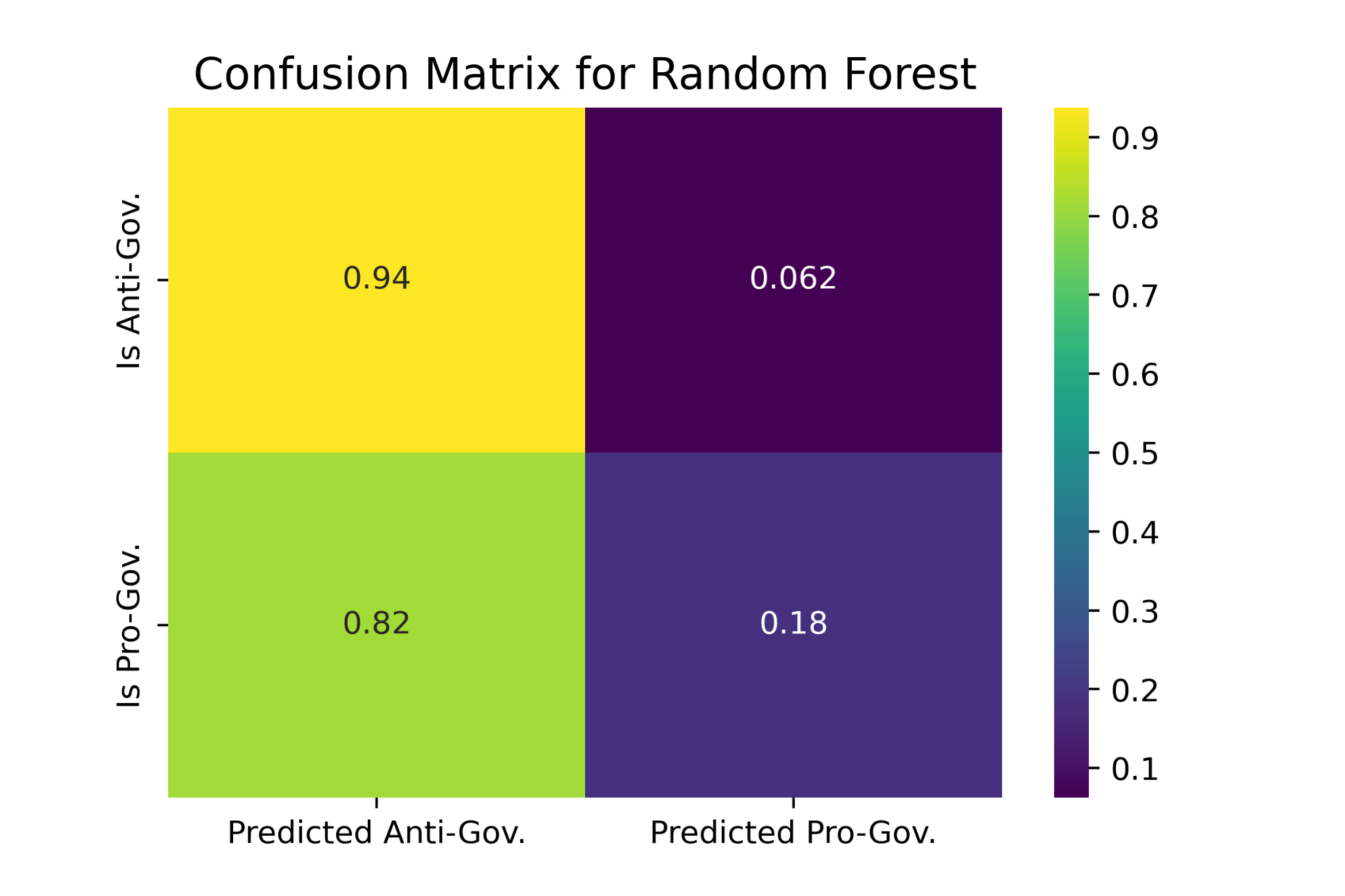
**Figure 7: Confusion Matrices for AdaDerana portal and selected three topics.**

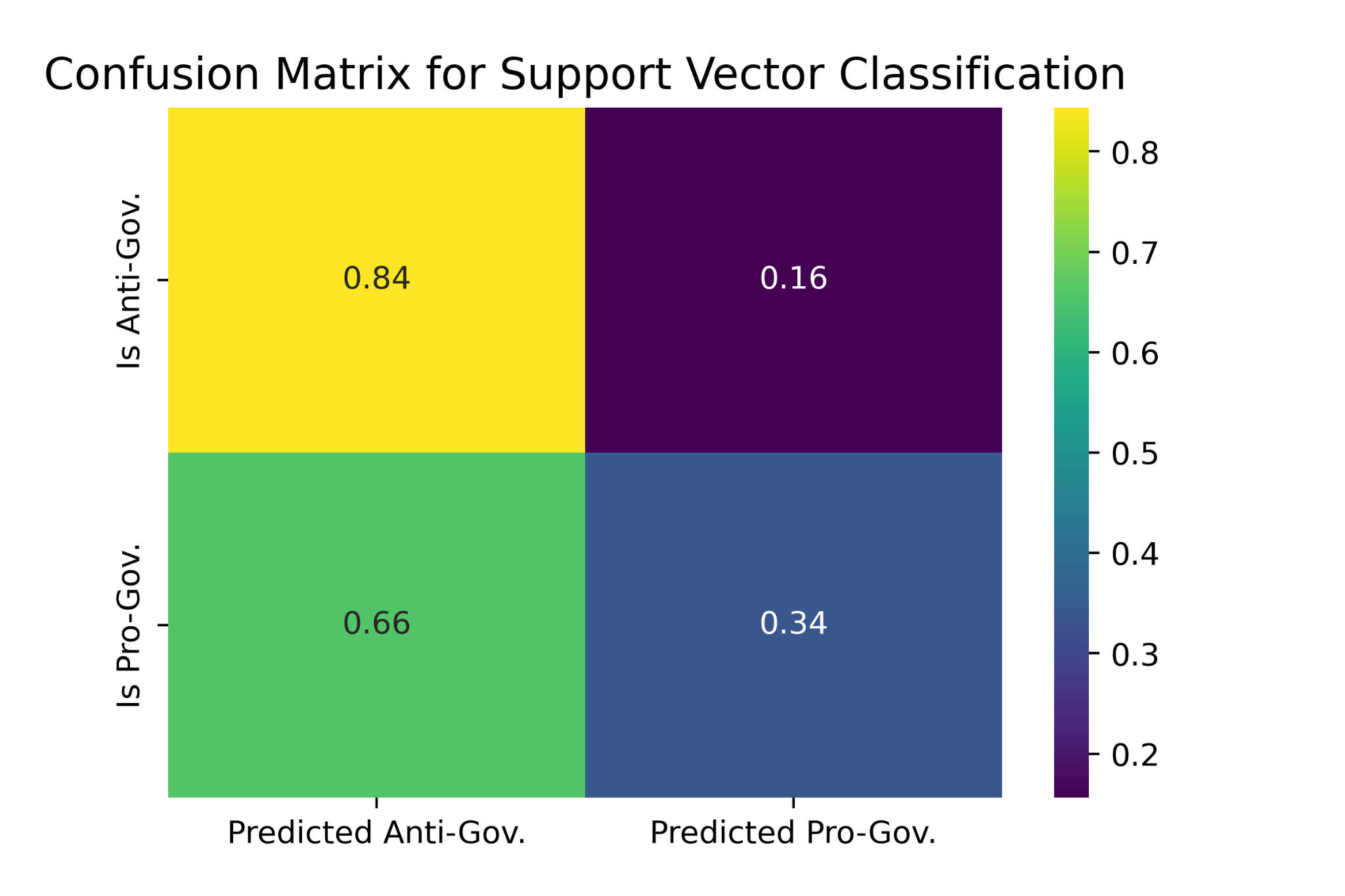
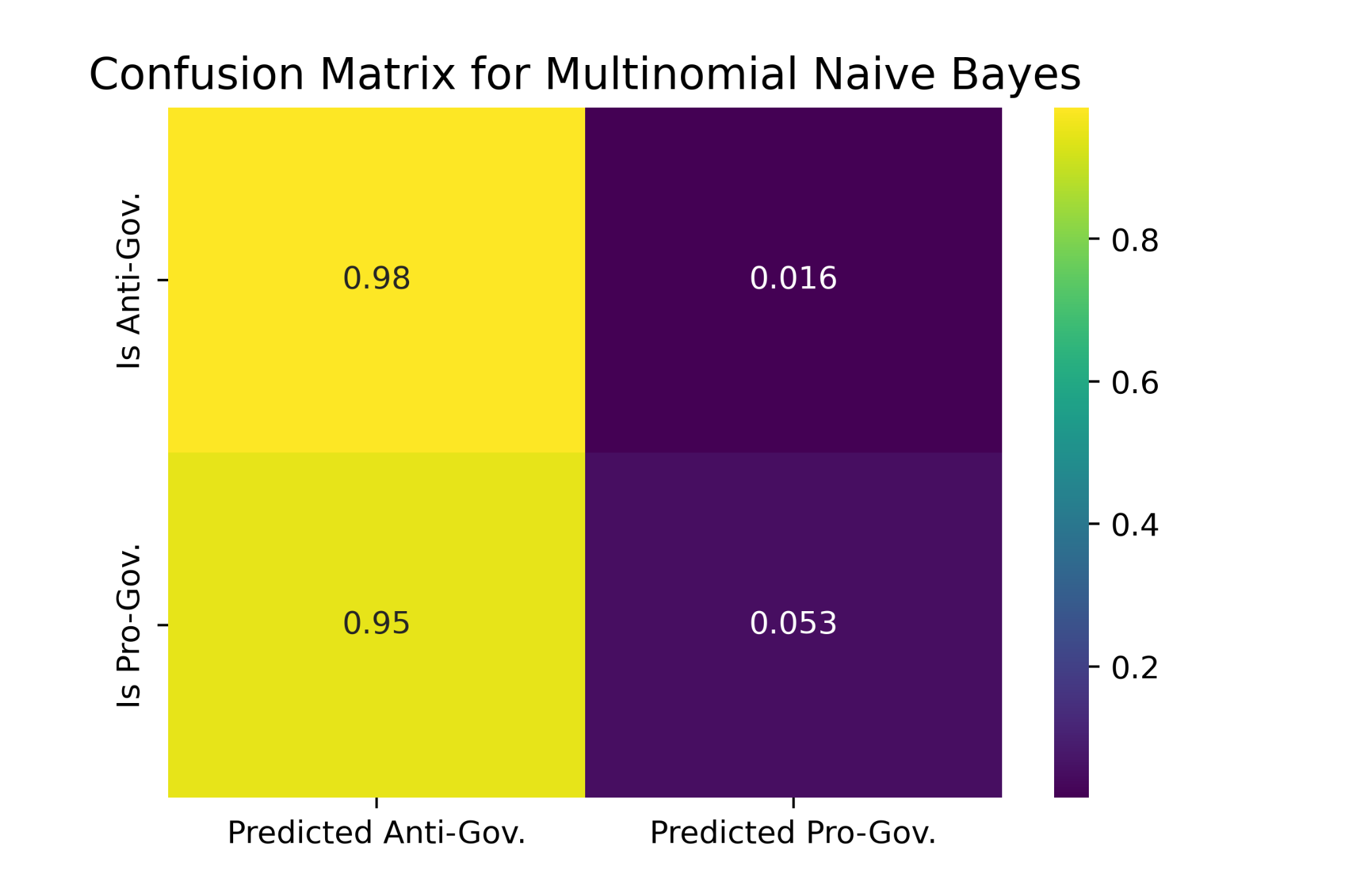
****

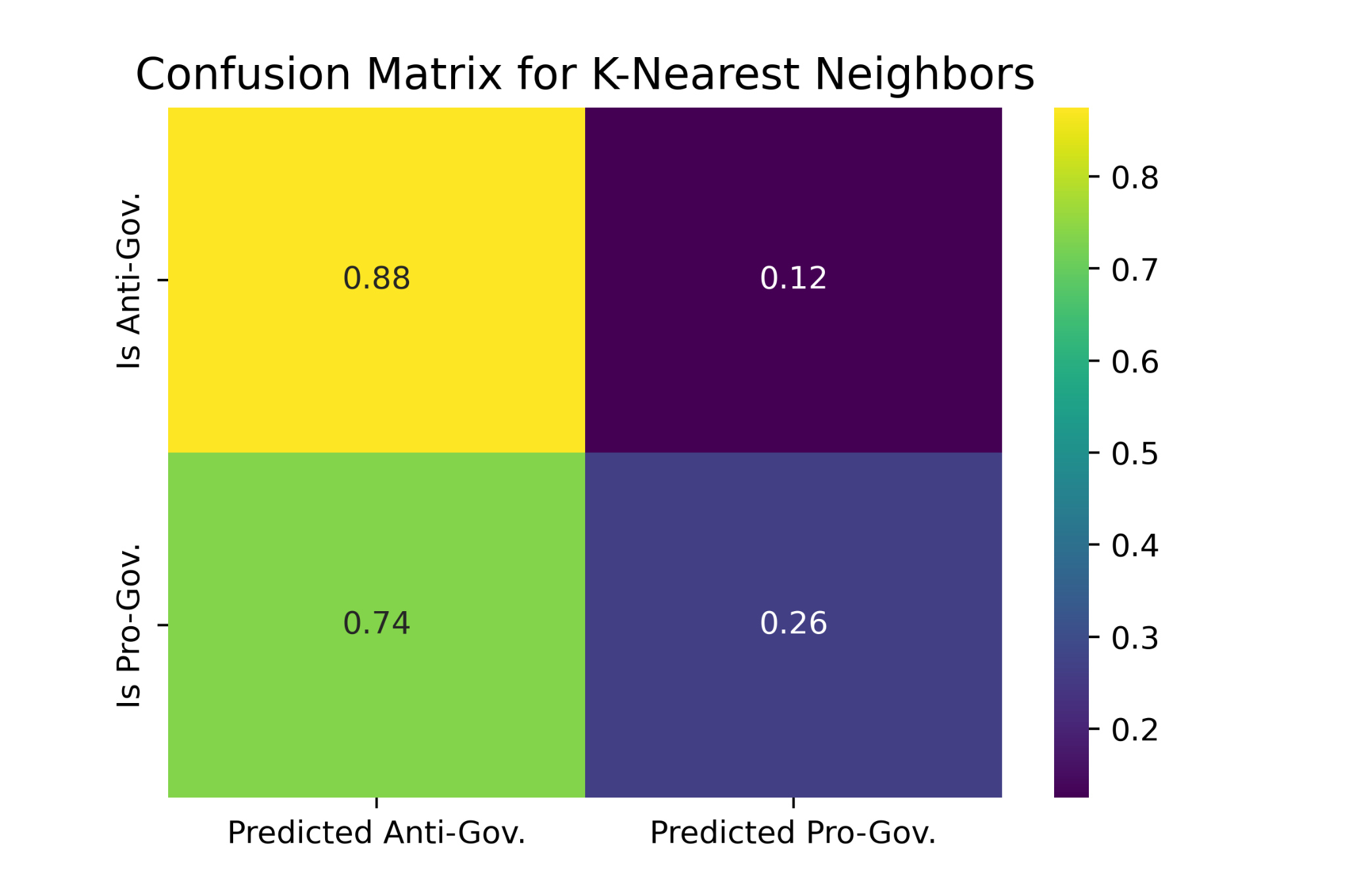
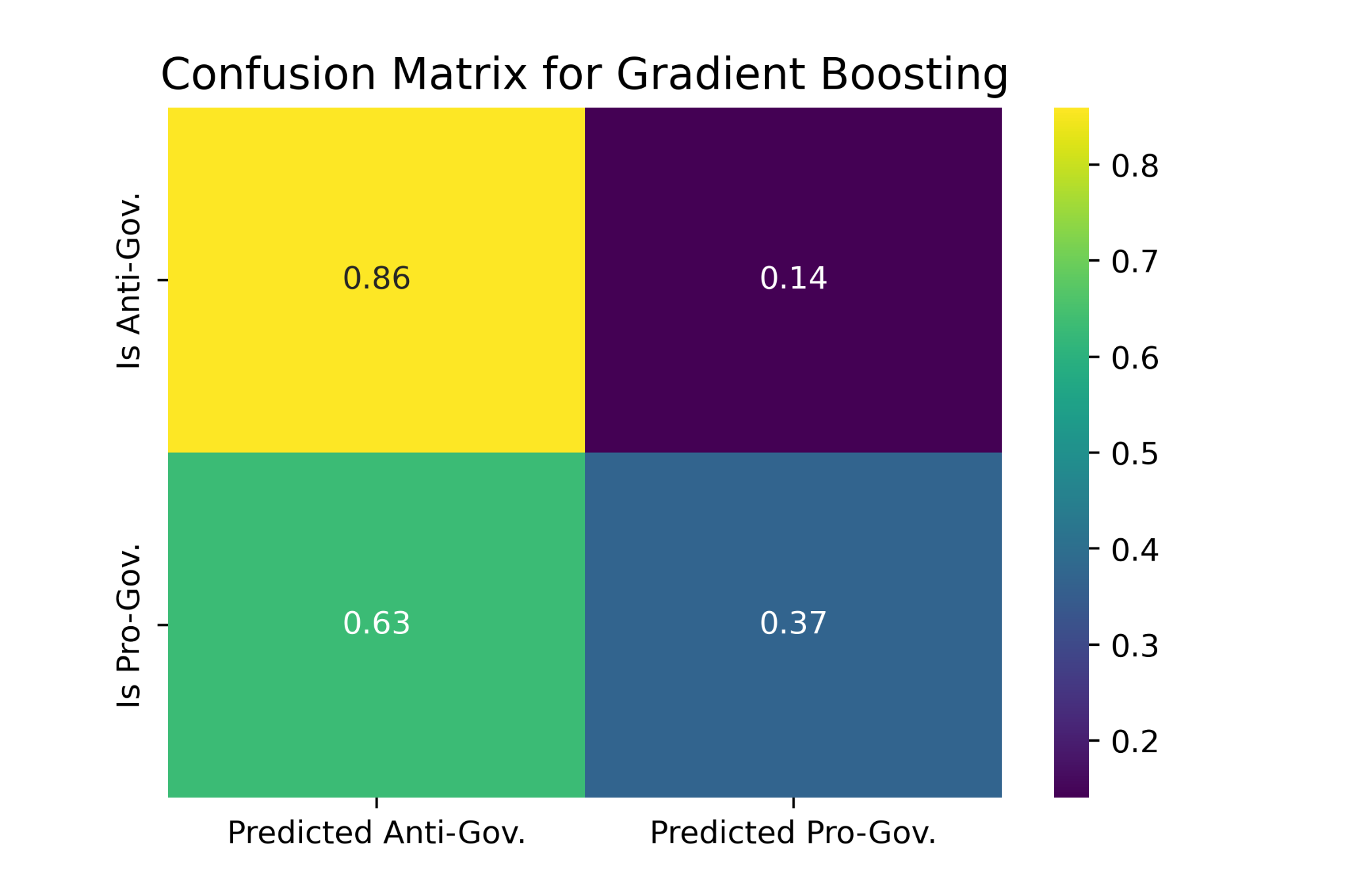
**** 

****

**Figure 8: Confusion Matrices for NewsFirst portal and selected three topics.**







**Conclusion**

In this study different natural language processing methods have been applied to analyze the news articles of AdaDerana and NewsFirst websites for the selected topics of Anti-Terrorism Act/Prevention of Terrorism Act (PTA), X-Press Pearl Ship (XPS) Cargo Vessel Sinking and Local government election (LGE).

As the **first approach** a sentiment analysis was conducted using VADER to find the polarity of the published news contents. The results of sentiment analysis suggest that the news content of both sites have a positive sentiment on LGE topic and negative sentiment on XPS and PTA topics.

In the **second approach** the existence of pro-government and anti-government key words in the news content was used to identify political biases of news sites. Based on the analysis, after considering all three topics, AdaDerana news site’s percentage of pro-government is high compared to NewsFirst.

In the **third approach** we tried to explore how natural language processing paired with classification machine learning models, can be used for detecting bias in news articles.

All of the models for AdaDerana dataset (306 sentences) scored high (above 80%) on the classification accuracy matrices. F1 scores for the Pro-government classification is remarkably higher than the F1 scores for the Anti- government classification. Also the actual and predicted percentage for all six models (True Negatives, i.e Actual => Pro. Predicted => Pro.) in the confusion matrices have higher values while type I (False Positive) and type II (False Negative) errors have less percentages. So we can conclude that AdaDarna news articles are more towards Pro-government as we labeled.

All of the models for NewsFirst except Multinomial Naive Bayes for NewsFirst dataset (406 sentences) scored relatively high (70-74%) on the classification accuracy matrices. F1 scores for the Anti-government classification is higher than the F1 scores for the Pro-government. The actual and predicted percentage for all six models (True Positive, i.e. Actual=> Anti. Predicted=> Anti.) in the confusion matrices have higher values while type I and type II errors have less percentages. This can conclude that NewsFirst articles are towards Anti-government as we labeled.

Some more tunings in the parameters and training iterations of the learning algorithms can probably score higher accuracy. These models can be improved with more data from different sources. This would further boost the generalization of the algorithms and produce more accurate results. For example one news portal might report the same sentence or use words in a common order in multiple articles. The model will learn these occurrences and classify new articles that contain similar wordings and ordings as biased, regardless of the real bias.

* **It is important to note that labeling news articles as biased is a relatively controversial topic. Bias could be changed as the opinion of a particular person, bias for some are not for others, vice-versa. These results are not an accurate representation of the truth, but gives a rough idea of the bias.**

**References**

[1] Paranyushkin, D. (2018, October 4). Measuring Discourse Bias Using Text Network Analysis | by Dmitry Paranyushkin. Towards Data Science. Retrieved May 15, 2023, from <https://towardsdatascience.com/measuring-discourse-bias-using-text-network-analysis-9f251be5f6f3>

[2] H. W. A.-K. S. Wei-Fan Chen, "Learning to Flip the Bias of News Headlines," pp. 79-88, 2018.

[3] Hamborg, F., Donnay, K. & Gipp, B. Automated identification of media bias in news articles: an interdisciplinary literature review. Int J Digit Libr 20, 391–415 (2019). <https://doi.org/10.1007/s00799-018-0261-y>.

[4] Hamborg, F. (2023). Media Bias Analysis. In: Revealing Media Bias in News Articles. Springer, Cham. <https://doi.org/10.1007/978-3-031-17693-7_2>

[5] Sri Lankan breaking news, Hot News - Adaderana -Truth first. Sri Lanka News. (n.d.). https://www.adaderana.lk/

[6] First, N. (n.d.). Newsfirst.lk - Sri Lanka: Latest breaking news and top stories: Sirasa News. sample image. https://english.newsfirst.lk/

[7] What is text mining, text analytics and Natural Language Processing?. What is Text Mining, Text Analytics and Natural Language Processing? Linguamatics. (n.d.). https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing

[8] Dwivedi, R. (n.d.). Introduction to text analytics and models in Natural Language Processing. Analytics Steps. <https://www.analyticssteps.com/blogs/text-analytics-and-models-natural-language-processing>

[9] S. R. Muhammad Umar Nadeem, "Detecting Bias in News Articles using NLP Models," at Stanford University.

[10] A. G. Shri Bharathi, "Determination of news biases using content sentiment analysis algorithm," Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, no. 2, pp. 882-889, 2019.

[11] Bellows, Martha, "EXPLORATION OF CLASSIFYING SENTENCE BIAS IN NEWS ARTICLES WITH MACHINE LEARNING MODELS" (2018). Open Access Master's Theses. Paper 1309.

<https://digitalcommons.uri.edu/theses/1309>

[12] S. Perera, I. Perera and S. Ahangama, "Exploring Twitter Messages during the COVID-19 Pandemic in Sri Lanka: Topic Modelling and Emotion Analysis," 2022 2nd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2022, pp. 148-153, doi: 10.1109/ICARC54489.2022.9753866.

[13] S. Perera, I. Perera and S. Ahangama, "The effects of Social Media User Behavioural Factors on Hate Speech Detection," 2022 2nd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2022, pp. 131-136, doi: 10.1109/ICARC54489.2022.9753942.

[14] Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020. [What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context](https://aclanthology.org/2020.acl-main.308). In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3364–3374, Online. Association for Computational Linguistics.

[15] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>

[16] <https://scikit-learn.org/stable/supervised_learning.html#supervised-learning> (accessed June, 2023)

[17] T. Thomsen, "Do Media Companies Drive Bias? Using Sentiment Analysis to Measure Media Bias in Newspaper Tweets," in Johns Hopkins University, Maryland, 2018.

[18] <https://investigate.ai/investigating-sentiment-analysis/designing-your-own-sentiment-analysis-tool/>

(accessed June, 2023)