Detecting Propagandistic Poster Title: A Machine Learning Approach

Md Intiajul Alam Shah, Riaz Mahmood, Hasan Abdullah, Tasnimul Hassan, Taskin Mohammad Mubassir

Department of Computer Science & Engineering, BRAC University

Supervisor: Md. Golam Rabiul Alam

Abstract

Propagandistic poster title detection is the procedure of identifying the poster titles which leads towards misinformation, lies, fake or invalid news and which doesn't have proper and valid grounds. Our research aims to detect the poster titles which have propaganda either implicitly or explicitly in it. A dataset consists of more than 3000 propagandistic and non-propagandistic poster titles that were formed manually for this research.

Research Objectives

The objectives of this research are:

- To understand the machine learning (ML) models and how it works.
- To detect propagandistic information from poster texts and also to select non pro pagandistic information.
- To develop a model for a propaganda detection system based on machine learning.
- To evaluate the model.
- To offer recommendations on improving the model.

Data Preprocessing

Here, to process data we do data cleaning, lower case, tokenization, removing special characters, removing stop words and punctuation and stemming.

	Propaganda text	Type	num_characters	num_words	num_sentences	transformed_text
0	The Company You Keep	0	20	4	1	compani keep
1	Rampage	0	7	1	1	rampag
2	Before Sunset	1	13	2	1	sunset
3	Sword Art Online: Progressive - Aria Of A Star...	0	56	11	1	sword art onlin progress aria starless night
4	The Personal History Of David Copperfield	0	41	6	1	person histori david copperfield

Figure 1: Transformed texts

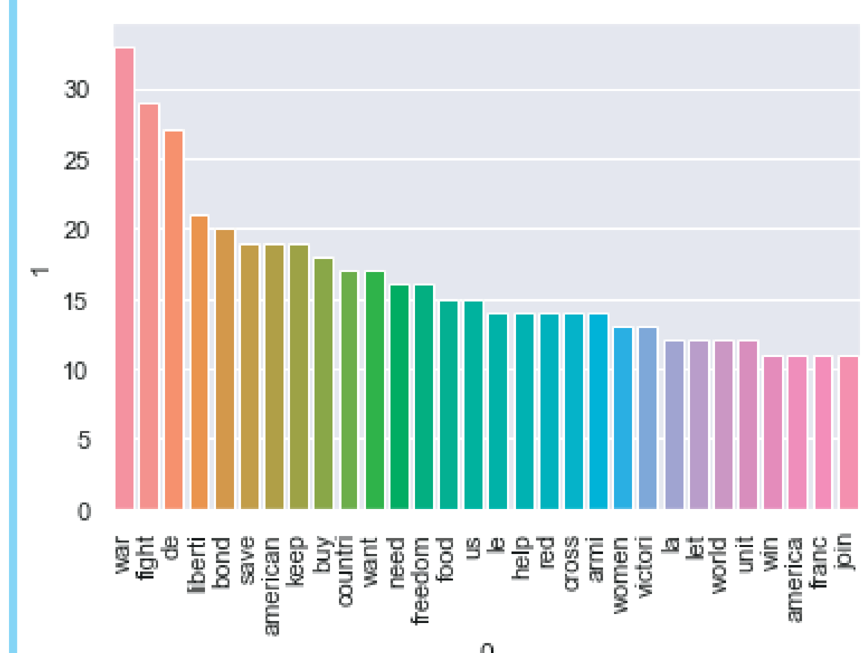

Figure 2: Propaganda words

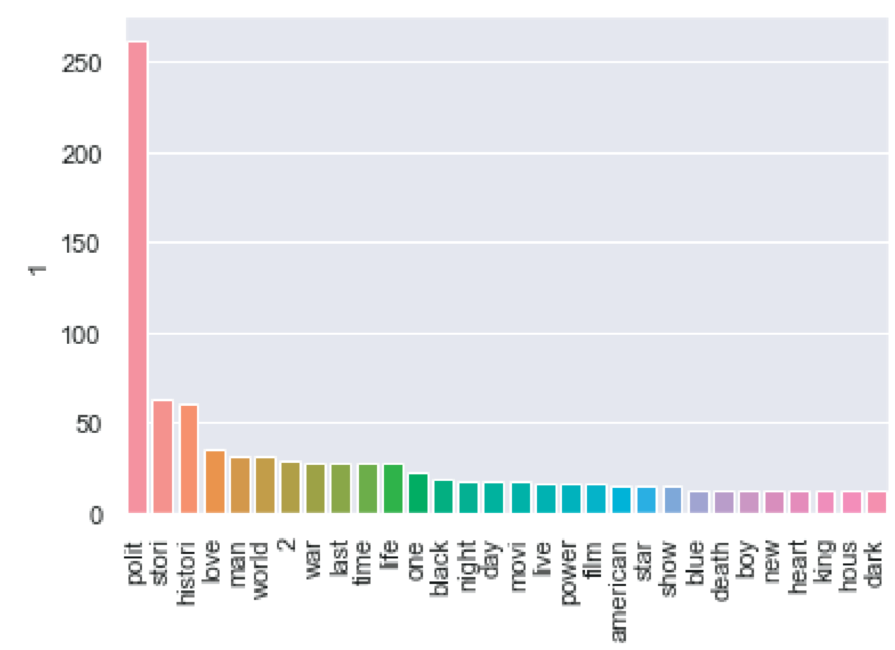

Figure 3: Non-propaganda words

Workflow

Our major goal is to construct a system utilizing machine learning to recognize propaganda in poster titles and The steps of our system are given below in the flow chart:

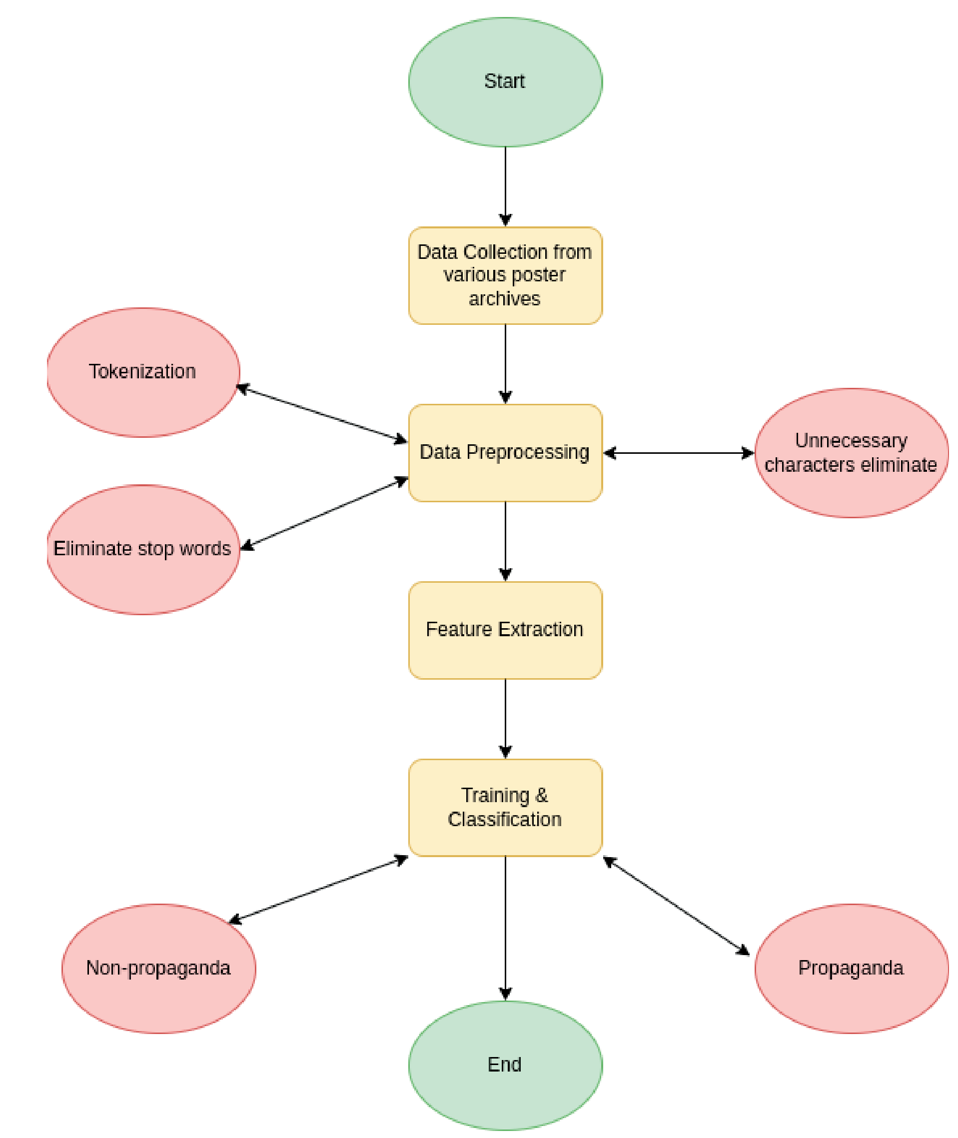

Figure 4: Workflow

Literature Review

A paper discusses a study that focuses on detecting propaganda at the sentence level in news articles. It introduces a strong binary classifier using transfer learning and a unified neural network model called BERT-BiLSTM-Capsule [1]. Another paper introduces a way to detect propaganda in news articles by examining smaller sections of text instead of labeling entire documents or news sources. They create a collection of articles with manual annotations, specifically focusing on propaganda techniques at a more detailed level also they achieve 63.20% accuracy by using the BERT-based model. [2]

Result Analysis

	Model	Accuracy Score
0	Logistic Regression	83.45%
1	KNN	83.80%
2	Random Forest	87.11%

Random Forest

	precision
0	0.88
1	0.82
accuracy	
macro avg	0.85

KNN Accuracy Score: 83.45%

	precision
0	0.83
1	0.84
accuracy	
macro avg	0.84

Logistic Regression:

	precision
0	0.83
1	0.92
accuracy	
macro avg	0.88

Figure 5: Accuracy and precision table

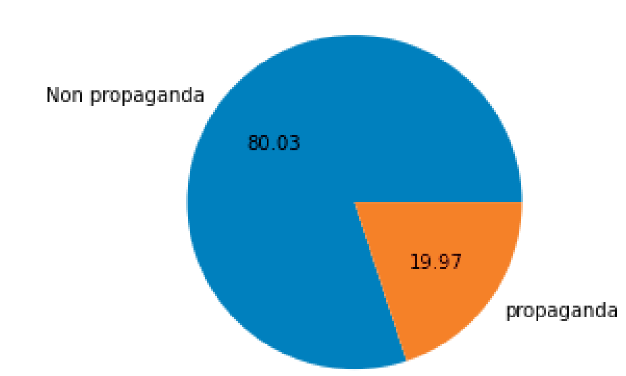

Figure 6: Pie-chart of the data

The Random Forest model has the highest accuracy score of 87.11%, followed by KNN with an accuracy score of 83.80% and Logistic Regression with an accuracy score of 83.45%. It sums up that the Random forest model came up as the best model to identify propagandistic titles of the poster. But our data is imbalanced. So, here we focus on precision.

References

1. Vlad, G., Tanase, M., Onose, C., & Cercel, D. (2019b). Sentence-Level Propaganda Detection in News Articles with Transfer Learning and BERT-BiLSTM-Capsule Model.

2. Da San Martino, G., Yu, S., Barrón-Cedeño, A., Petrov, R. A., & Nakov, P. (2019b). Fine-Grained Analysis of Propaganda in News Article.