

Project Initiation Document

Fake News Classifier

Prepared by

Simon Lowry, D20129695

Table of Contents

<u>1</u>	<u>PROJECT OVERVIEW</u>	<u>2</u>
1.1	PROJECT OBJECTIVE	2
1.1.1	SUCCESS CRITERIA	2
1.2	PROJECT ASSUMPTIONS	2
1.3	RISKS	2
<u>2</u>	<u>SOLUTION DESIGN</u>	<u>3</u>
2.1	PROPOSED ANALYTICS SOLUTION	3
2.1.1	SOLUTION USES	3
2.2	DATA REQUIREMENTS	3
2.2.1	DATA SOURCE DESCRIPTION	3
<u>3</u>	<u>PLAN OF ACTION</u>	<u>4</u>

1 Project Overview

1.1 Project Objective

Over the past decade or more, there has been an increased prevalence of what has been deemed fake news. Fake news is crafted information that is used to try and create a certain intended impression which is usually a deceitful one. In some countries it's been used to sow distrust, to create false narratives, to spread lies, damage the reputation of people. Its impact can't be underestimated as it's leading to a lot of polarization in our society, a distrust in core pillars of society (like the use of science in government), and in some countries playing a role in undermining their whole political systems of democracy. In certain countries there are actual factories of people whose main job is to create fake news and propaganda to further certain agendas. Ultimately it is a big problem for our current society and not one that will disappear any time soon. In an era where social media platforms have become a central hub for a lot of people's time and attention tools which can help in the fight against blight are becoming increasingly important. Creating tools with AI solutions can help to combat the proliferation of fake news. Especially since in a number of sites, they are used by literally billions of people across the world. The ability of humans to comb through millions of posts every day is not feasible so having tools that can run at scale and help in this process can be a huge asset and is of increasing importance. The aim of this project is to create a solution that can help in this process of classifying fake news.

1.1.1 Success Criteria

Ascertain which of the applied models performs best not just on the training data but also generalizes well to other data as well. This evaluation will be based on metrics such as precision, recall and f1-score.

1.2 Project Assumptions

There is an implicit trust by using this data source from kaggle that the labelling for the fake news data set has been correctly performed and as a result will be a good data source for attempting to create a tool that can identify fake news. Further analysis of the data source will be carried out as part of the data analysis section of the project.

1.3 Risks

The introduction of a tool for judging fake news is best served as an aid to people to be used in conjunction with a human-in-the-loop to validate the various classifications performed by the tool. This is important because there will be a number of scenarios whereby it won't be completely clear even for a human whether a given piece of text is legitimate or fake news. The human observer would be able to take advantage of gaining more context by, for example, visiting the user's profile and using other indicators from their profile and other previous posts they've made. They could also perform some research around the topic as well to assess its validity. This could help to battle against any false-positive scenarios that could potentially frustrate users and damage the reputation of the site if done repeatedly at scale. The classifier could be used as a helpful tool that could potentially flag posts which are considered to have a good chance of being fake news and then the human-in-the-loop could make the final call on whether it is fake news or not.

2 Solution Design

2.1 Proposed Analytics Solution

The proposed idea is to try out a number of different classifiers models and determine which approach produces the best results. The models that will be applied are Naive Bayes, Support Vector Machines, Random Forest algorithms to ascertain the most effective model for determining fake news amongst these.

2.1.1 Solution Uses

The data source for the Fake news tool i'm looking to create is from kaggle:

<https://www.kaggle.com/c/fake-news/data>

2.2 Data Requirements

The approach will be to implement a classifier which can apply natural language processing techniques and tools to textual information, for this reason, it's important that the data is fit for purpose. This data set adequately meets the initial requirements of being in textual format for the variables where the model will be applied however the data will still require a deeper dive into assessing its data quality which will happen in the data analysis and data cleaning stages. This will help preprocess the data in order to get it into a format and shape that's best suited for the models we're looking to apply and to gain the best results from those models. Data preparation and its impact is becoming an increasingly and widely spoken about area of AI projects as it plays a crucial role in the performance of the model. The results of this data analysis and data cleaning will be included in the final solution report.

2.2.1 Data Source Description

There is a total of 26680 entries that have been compiled into this dataset. The data has already been split into two different csv files. It's already been partitioned into training and testing data which takes away from needing to perform that step. They are divided as follows:

Training Set total entries: 20800

Test Set total entries: 5880

This makes for a split of 22% for the test data and 78% for the training data, which is close to the standard 80/20 split that can be used commonly in machine learning projects.

train.csv - This is the full training dataset with these given attributes:

- id - unique id for a news article
- title - the title of the news article
- author - author of the news article
- text - the text of the article, could be incomplete
- label - a label that marks the article as potentially unreliable
 - 1 - unreliable
 - 0 - reliable

test.csv: A testing training dataset with the same attributes of train.csv without the label.

3 Plan of Action

Data analysis

During this stage I'll be looking to get a sense of the dataset and its properties with performing some data analysis. This will help to summarize some of the main characteristics of the dataset, uncover some details about it and gain a better understanding of the dataset overall. This will be done using a combination of using libraries such as pandas and matplotlib or tableau for producing some visualizations.

Data Cleaning and Preprocessing

After gaining more insight into the data set there will then be a data cleaning and preprocessing stage. During this stage, the data will undergo a number of procedures in order to change into a format that allows for the models to obtain greater performance. This will involve applying methods to deal with missing data, removing any unnecessary special character, as well as punctuation marks, removing urls, removing double spaces & line breaks. There will also be the removal of stopwords, which are commonly occurring words within a given language which don't actually provide any additional meaning such as 'a' or 'the'. The data will then be changed into lower case as well.

Model Selection

As mentioned previously, the models that will be applied are Naive Bayes, Support Vector Machines, Random Forest algorithms. Each of these will be applied with a few to finding the most effective model for determining fake news. If time allows there may also apply some deep learning techniques applied and included in the model selection.

To assess the effectiveness of the models, the following evaluation metrics will be used: precision, recall, and f1 score. These are commonly used metrics that can be used to assess the performance of binary classification problems.

Precision gives the number of correct predictions made.

$$\text{Precision} = \text{TruePositives} / (\text{TruePositives} + \text{FalsePositives})$$

Recall gives the number of positives class predictions from all of the positive examples in the dataset.

$$\text{Recall} = \text{TruePositives} / (\text{TruePositives} + \text{FalseNegatives})$$

F1-score is the combination of both precision and recall into a single calculation which enables both of these to be captured.

$$\text{F1-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Results and Discussion

The results of the model performance evaluation will lead then to some discussion and visualizations in the final solutions report. This will illustrate how well our models have performed on the data set and which of the models has been shown to be the most effective for the given problem.