CLIMATE CHANGE SENTIMENT ANALYSIS

1. Introduction

Climate change is the long-term changes in temperature and weather patterns. Although these changes are natural, since the 1800s, human activities have been the primary driver of climate change owing to the burning of fossil fuels (such as coal, oil, and gas), which produces greenhouse gases. Climate change will impact the lives of future generations. Hence, Climate change is one of the most important concepts in the current scenarios that people need to be aware of. But some people think that climate change is just a hoax or a conspiracy theory while some people think that climate change is a severe threat to life on earth.

In this project, we propose to analyze the opinion of people by performing sentiment analysis on tweets by Twitter users on this topic. We plan to do an exploratory data analysis and train models based on various traditional machine learning algorithms and a Deep Learning Model using Recurrent Neural Network approach using Text in Tweets as features.

By performing this analysis, we get a rough idea of how aware and how supportive people are towards the cause of the prevention of climate change and plans can be made to bring awareness to people.

2. Dataset

We are using **Twitter Climate Change Sentiment Dataset**[1] taken from the Kaggle website containing more than 43 thousand tweets based on the topic of climate change in the period between April 27th, 2015, and February 21st, 2018.

These tweets are labeled by three reviewers independently into 4 classes:

- 2(News): the tweet links to factual news about climate change
- 1(Pro): the tweet supports the belief of man-made climate change
- 0(Neutral): the tweet neither supports nor refutes the belief in man-made climate change
- -1(Anti): the tweet does not believe in man-made climate change

3. Implementation Methods

Method 1: Traditional Machine Learning Approaches

As part of this approach, we perform Exploratory Data Analysis on the data to get some statistics about the statistics of sentiments of tweets.

Then we perform Multi-Class Classification on the tweets by using Document classification techniques by converting the text into vectors or other numeric formats to give the tweet text as input features for Machine Learning algorithms. We train models to train systems using supervised learning to detect the sentiment of tweets and check how accurate these approaches are.We plan to train models using algorithms like Logistic Regression, Support Vector Machines, Random Forests, etc.,

Method 2: Traditional Machine Learning Approaches

We plan to perform a Token-based Text Analysis using Natural Language Processing and find the keywords in tweets affecting the sentiment of tweets and perform statistical analysis on the frequency of these buzzwords on each sentiment.

Using these keywords and cleaned tweet texts, we try to use the Long Term Short Memory approach to train a recursive neural network. We plan to use predefined models in the Keras library for training the network.This approach is a semi-supervised learning approach as we are using keywords defined in the text analysis as features for training the network.

4. Evaluation Method

For Testing and Evaluating the models trained, we split the dataset into two sets: one set we use for training the system and the other for testing and evaluating the models.

For Machine Learning approaches we use Metrics like Precision, Recall, F-1 Score, Support, Train, and Test Error percentages for each sentiment class by using methods defined in the Scikit-learn library. And for the Deep Learning approach, in addition to the above metrics, we use evaluation metrics defined in the Sequential Model of Keras library.

5. References

[1] Twitter Climate Change Sentiment Dataset