Aisha B., Allen S., Ritika S.

CS301-102

Group 6

Professor: Dr. Pantelis, TA: Yashwee Kothari

Due Date: April 24, 2022

**Group Six Project:**

Partly Sunny With a Chance of Hashtag

**Competition Website**

<https://www.kaggle.com/competitions/crowdflower-weather-twitter/overview>

**Abstract**

As the years go by, social media platforms, such as Facebook, Instagram, and even Twitter, are being used more frequently. We are at a point where we can use these platforms to predict the weather and when it has happened. In this project, our team was provided a data set with tweets from Kaggle, which was related to the weather. Using those tweets, we have to come up with a python program that will not only predict what the weather is but will also predict when it happened (past, current, or future). In this project report, the reader will understand what this project is all about and its process for it. They will know what related works group six has looked into, the data that has been collected, methods that have been used, and how did the experiment has went, and also what a whole team has concluded at the end of this project.

**Introduction**

In the era of big data, social media applies many information methods. In this project, we will be provided a data set with tweets related to the weather. With these provided tweets, our group will create a program that will read these tweets and analyze them to see what sort of weather the tweet references and will figure out when the weather occurred (past, present, or future). This project is to create a software that will analyze a set of tweets which will be given to the software itself and from there it will determine if the weather is one of the three sentiments (positive, neutral, or negative) and then tell if the referenced weather in the tweet occurred in the past, present, and or future. The main objective of this project is to be able to assign certain labels to tweets based on the weather information that we can learn from them. This is clearly a classification problem so approaches like decision trees are being taken into consideration since they fit our requirements. While the competition’s performance metric is based on how many labels are accurately assigned, we will be focusing on getting the outputs right.

**Related Work**

Many companies have done similar work towards this type. Space Weather Prediction Center Space has Weather Scales that were introduced to communicate to the public the current and future space weather conditions. They analyze different readings and also developed a computer model to predict the weather and its warnings better than before. The Space Weather Prediction Center Space has produced forecasts for different types of space weather phenomena.

Weather is one of the biggest influences in our everyday lives. However, as the concept of data analytics continues to expand, the role of social media in the future of weather tracking is being more notable. In a recent research completed by the Warwick University (Britain) found that certain words on social media can be used to predict weather conditions before they actually occur. Nataliya Tkachenko of the Warwick Institute had once said “our analysis demonstrates that metadata in social media image postings enables them to be used as social sensors, which can serve as a valuable supplement to instrument-based systems for predicting and monitoring floods, and other kinds of natural hazards”. So other than the Space Weather Prediction Center Space, many institutes have done similar research close to ours.

**Data**

For this report, we will be working with data (tweets) that are from Kaggle. The set of tweets consists of words such as “hurricane”, “hot”, “cold” “rainy”, etc. It consists of four files, and a total of 57 columns. The four files given are “sampleSubmission.csv”, “test.csv”, “train.csv”, and “variableNames.txt”. Each of these contains important information which is why when we download it we use:

*kaggle competitions download -c crowdflower-weather-twitter*

We will be using many functions that we have used throughout that we have learned this semester that will be mentioned in the next section (the Methods Section).

**Methods**

Since this challenge is mainly dealt with keywords and phrases, we had to use certain methods. To start off we imported many packages such as NumPy, Pandas, Matplotlib, and also JAX. Grid search was also used to search for the best matching parameters. The fit regression model is also needed to organize the details. Although NumPy and JAX go hand in hand, we believed that both functions should be used to make the project more readable, run faster, and would be easier to debug whenever there is an issue with compiling the program when it finishes.

The biggest problem for us was to download the data and have the python program analyze the data line by line. After that problem was fixed, the next obstacle that came in our was was to have the program decide what should its output be on its own.

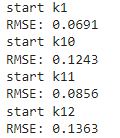
**Experiments**

For this experiment, we used a fit regression model to predict the tweets’ weather and their time.

Part of The Given Data:



The Expected Output Would Be:



A storm usually consists of rain, clouds, and a lot of wind. This given output is the mean squared error for each feature. This means that for K1, there is clouds. Zero would mean we were 1000% determining if there were clouds or not.

**Conclusion**

In conclusion, this project applied pattern and logistic recognition. We learned that with NumPy, Pandas, Matplotlib, and JAX, we can create a python program that could read a set of tweets and predict the weather reading those tweets, and can tell when has/will it happen. After doing this project, we strongly feel that we are heading towards an era where we can predict much more than just weather with the use of social media.

**References**

* NOAA US Department of Commerce. (n.d.). National Weather Service. Retrieved April 2022, from https://www.weather.gov/
* The Weather Company: The internet of weather things. Social Media for Business Performance. (n.d.). Retrieved April 2022, from https://smbp.uwaterloo.ca/2017/03/the-weather-company-the-internet-of-weather-things/