**Weather Analysis: To Trust or Not to Trust?**

**Have you ever found yourself thinking why is the weatherman always wrong? Oftentimes, especially in a climatologically-schizophrenic state such as the Great State of Georgia, a ‘meteorologist’ will predict a day to be 65 degrees and sunny before reality turns out to be a much more a disappointingly dreary 45 degrees with overcast.**

**Problem Statement: The conventional wisdom is that weather can be fickle and difficult to forecast with acute accuracy, however there is an entire industry built around the idea that weather is predictable based on meteorological models. How accurate are these models when compared with the raw data?**

**Hypothesis: Meteorologists are generally wrong about their predictions for weather forecasts.**

**Null Hypothesis: Meteorologists are generally correct about their weather forecasts.**

**Expected Outcome: We believe that we will fail to reject the null hypothesis, which would show that meteorologists are indeed generally incorrect in their weather forecasts.**

**For the purposes of this project, at least two data sources will be needed: an hourly forecast from a reputable weather service for over the course of 24 hours with the chosen day of 3/20/2019. We then gathered raw minute-by-minute weather data independently collected by our robot R.A.M.B.O. for the same day of 03/20/2019.**

**Due the seriousness of the topic at hand, the founding members of the United Nations have took over the project and will handle all testing.**

**Slavoljub Petkovic, VP of Data Science: data gathering, project planning**

**Hardeep Chahal, VP of Finance: data cleaning, analysis, project planning**

**Andrew Smith, VP of Operations: analysis, technical report, project planning**

**Technologies to be used: Raspberry pi 3 B+, SQLite, Excel, Jupyter Notebook, Dark Sky API, Python [dependencies: SQLAlchemy, datetime, time, board, BUSIO, adafruit\_bme680, sqlite3, pandas, JSON].**

**ETL: Describes the process of extraction, transformation, and loading that is required of data before manipulation and analysis can be done.**

**R.A.M.B.O.: mostly just a fun name for the Raspberry Pi 3 B+ used to collective the real-time weather data which stands for Raspberry pi Automated Meteorological BME680 Operative**

**BME680: Bosch Sensortec’s integrated environmental sensor used in this case as a mobile weather station.**

**API: Application Programming Interface - a set of subroutine definitions, communication protocols, and tools for building software.**

**FTP: File transfer protocol -**

**PW: Predicted weather - the weather forecast as gathered from the darksky.net api.**

**AW: Actual weather - the actual weather measurements gathered from R.A.M.B.O.**

**To begin the project, Slav will source code to extract the forecasted weather for 03/20/2019 Dark Sky. While beginning the external data extraction, the company enlisted R.A.M.B.O to gather data as part of our internal testing. Both tests will be extracted at the same zip code, 30092 – Norcross, GA for a full 24 hours.**

**The following illustrates what will be used and parameters set:**

**PW Data Set**

* **URL:** [**https://darksky.net**](https://darksky.net)
* **Extraction Protocols: Python API calls**
* **API:** [**https://api.darksky.net/forecast/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/37.8267,-122.4233**](https://api.darksky.net/forecast/*******************/37.8267,-122.4233) **(API has been excluded)**
* **Available Parameters: Apparent temperature, atmospheric pressure, cloud cover, dew point, humidity, liquid precipitation rate, moon phase, nearest storm distance, nearest storm direction, ozone, precipitation type, snowfall, sun rise/set, temperature, text summaries, UV index, wind gust, wind speed, wind direction**
* **Loaded Parameters: Time, Temperature, Pressure, Humidity**
* **Access: requires an email address to obtain a trial api key and usage of a python script provided by the website**

**AW Data Set:**

* **Technology: Raspberry Pi with a BME680 integrated environmental sensor**
* **Location: Zip 90032**
* **Available Parameters: Pressure, humidity, temperature, gas, altitude**
* **Loaded Parameters: Pressure, humidity, temperature, gas, altitude**
* **Access: Access to the listed technology and usage of a python script provided by Bosch SensorTech, manufacturer of the BME680**

**As described above, we gathered two datasets to compare against one another: (1) the PW and (2) the AW. To gather these two datasets, we used very different resources.**

**For the PW, a standard API call was made to darksky.net using a python3 script provided by Dark Sky to gather the hourly forecasts for the zip code 30092 from 12:00 am to 11:59 pm on 3/20/2019. Dark Sky API parameters allows users pull 1000 calls per day for free, after which, the price per call is $0.0001.**

* **The Dark Sky Company specializes in weather forecasting and visualization. Dating back to 2011, the company had the crazy idea that robots could predict the weather with down-to-the-minute precision, and thanks to a handful of generous strangers, they were able to give it a shot. Since then, those robots have become "scarily accurate," powering their very own award-winning weather app, Dark Sky, in addition to thousands of other businesses, apps, and crazy ideas.**
  + **Based on the Dark Sky’s reputation, accuracy and API, we decided this was a qualified resource for data.**
  + **The API call was based on utilizing Python 3 with which we extracted:**
    - **Time (24-hour time scale)**
    - **Temperature (Fahrenheit)**
    - **Pressure (sea-level air pressure in milibars)**
    - **Humidity**
* **Our API called upon hourly forecasts measuring each of the variables previously mentioned which was stored into a SQL database.**

**For the AW, we had to gather our own dataset using R.A.M.B.O. To do this, our data-gatherer would place R.A.M.B.O. outdoors at zip code 30092 where, using a weather-data-gathering library provided by Bosch Sensortec gathering information based on our python application. R.A.M.B.O. measured the weather at one-minute intervals from 12:00 am on 3/20/2019 to 12:00 am on 3/21/2019 for a total of 1440 data points.**

* **We used R.A.M.B.O. to accurately source and pull real-time data points from the environment, independent of a third-party’s service such as Dark Sky.**
* **R.A.M.B.O. utilizes the Python 3 language in concert with the technology found in the BME680 to yield:**
  + - **Time (Linux timestamp)**
    - **Temperature (Celsius)**
    - **Gas**
    - **Humidity**
    - **Pressure**
    - **Altitude (feet)**

**In order to provide any sensible analysis, we had to bring the data into same format first. We imported, cleaned, modified, and exported our data from Jupyter Notebook back to a SQL Database. Issues between the databases included imperial units from Dark Sky versus metric units from the Bosch Sensortec and the time being displayed as Linux Timestamps. To correct these issues, we converted the Linux timestamp into datetime as well as correcting the imperial units (Fahrenheit) into metric units (Celsius). Since R.A.M.B.O recorded data on minute level we can explore min and max and mean of temperature for every hour and compare them against the forecasting. This could be useful to identify which of the values are correlated most closely to forecasted values.**

**As noted earlier, Dark Sky is a highly reputable company for weather forecasts, based on their API versus other websites, we found them to be very helpful even if we were trying to disprove them. As for R.A.M.B.O., the newly minted BME680 sensor is manufactured by BOSCH, a well-known, trusted, global brand that has been in existence for over 120 years.**

**As for data security, we adhered to Dark Sky’s provisions while creating our API and pulls. The API key is the only security parameter we faced, which is kept away from the source code, instead being imported as api\_key from config.py for secrecy.**

**After conducting the data extraction, cleaning and display, we as a team were pleased with the process. Major issues during this process was non-existent, any minor hiccups were quickly handled the internal staff. As for our clientele needs, we have opened SQLite visuals from them to see the data sets, but effectively we just wanted to give them their answer. However, screenshots below show our results from Dark Sky and R.A.M.B.O.**



