**Memorandum**

**DATE:** 01/12/2019

**TO: XX**

**FROM:** TIMOTHY MADEGWA

**SUBJECT:** PREDICTING AIR QUALITY IN DAKAR SENEGAL

Here is the report you requested on 27/11/2019. It relates to the prediction of air quality in Dakar Senegal using machine learning methods.

**PREDICTING AIR QUALITY IN DAKAR SENEGAL**

Presented to:

XX

Presented by:

Timothy Madegwa

01/12/2019

**EXECUTIVE SUMMARY**

**Purpose and method of this report**

The goal of this project was to build a predictive regression model from historical PM2.5 pollution data from January 2010 – March 2019 and historical weather data from August 2011 to November 2019 using Python programming language.

**Findings and conclusions**

There was a seasonal trend in the PM2.5 pollution levels, higher temperatures and higher levels of relative humidity resulted in lower levels of PM2.5 pollution and on average, the months of December, January, February and March experienced higher levels of PM2.5 pollution.

**Recommendations**

Recommendations to build better predictive models include:

* One point of collection for PM2.5 pollution data and weather data
* Consistent data collection to avoid missing data

Table of Contents

[Introduction 5](#_Toc26141213)

[Challenges faced 5](#_Toc26141214)

[Scope of the study 6](#_Toc26141215)

[Findings and Conclusions 6](#_Toc26141216)

[Recommendations 8](#_Toc26141217)

[References 8](#_Toc26141218)

# **Introduction**

The goal of this project was to predict air quality in Dakar, Senegal based on ground sensor data that records PM2.5 pollution and historical weather data. The predictive model regression model was built from historical data of PM2.5 pollution from January 2010 – March 2019 and historical weather data from August 2011 to November 2019 using Python programming language with its related libraries (NumPy, Pandas, Matplotlib and Sci-kit learn).

To do this, we used three machine learning algorithms:

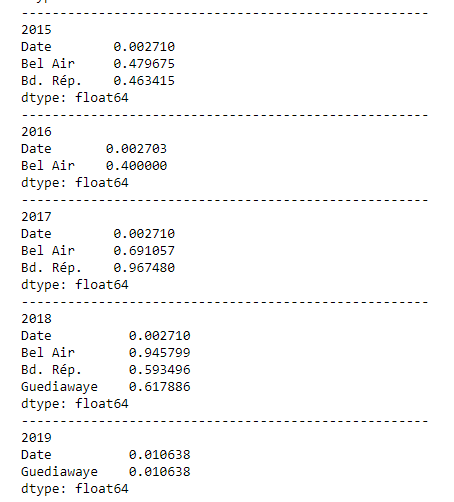
* Random Forest Regressor
* XGB Regressor
* Linear regression

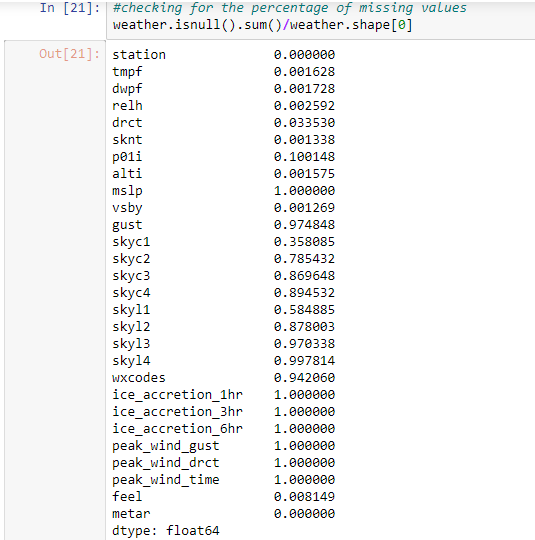
With the Random forest regressor came out on top with an error of 8.798.

# **Challenges faced**

In conducting this project, we encountered the following challenges:

* Lack of readily available historical weather data for Dakar, Senegal.
* Lack of a single dataset containing both weather data and PM2.5 pollution data
* Inconsistencies in the readings of PM2.5 pollution in the different stations
* Missing data from the historical PM2.5 pollution data
* Missing data from the historical weather dataset





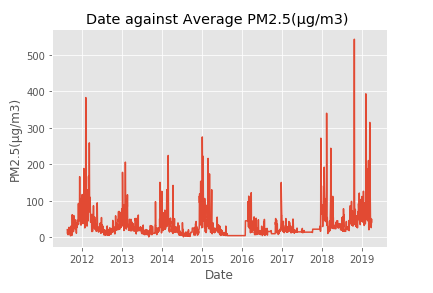
# **Scope of the study**

* Predicting the levels of PM2.5 pollution
* The effects of weather conditions on the PM2.5 pollution

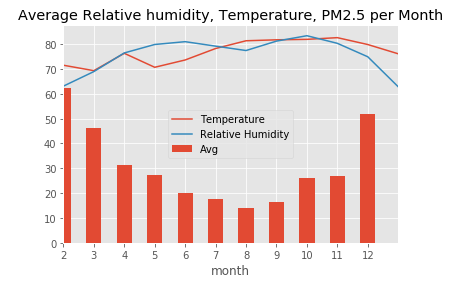
# **Findings and Conclusions**

Based on the analysis of the Dakar PM2.5 pollution levels and the weather data, we came up with the following conclusions:

* There was a seasonal trend in the PM2.5 levels.



* Higher temperatures resulted in lower levels of PM2.5 pollution
* Higher levels of relative humidity resulted in lower PM2.5 pollution levels
* On average, the months of December, January, February and March experienced higher levels of PM2.5 pollution.



# **Recommendations**

After considering the findings and conclusions of this analysis, we offer the following recommendations in an effort to build better predictive models:

* One point of collection for PM2.5 pollution data and weather data
* Consistent data collection to avoid missing data

# **References**

Historical weather data for Dakar, Senegal <https://mesonet.agron.iastate.edu/request/download.phtml?network=SN__ASOS>