Research Proposal

0. **Student Information**

Alan Uthuppan - alan.uthuppan@temple.edu

1. **Real-Time Air Quality Prediction Using Machine Learning**

2. **Objective & Significance**

The primary objective of this project is to create a Long Short-Term Memory recurrent neural network in Philadelphia to forecast PM2.5 and PM10 values. Given the health impacts of air pollution, especially in densely populated urban areas, this system is critical for public health warnings and environmental policy decisions. By predicting air quality fluctuations, the model can help take preventive measures against the harmful effects of air pollution on health, ecosystems and overall quality of life.

3. **Background**

Air pollution is a growing problem as it is directly linked to chronic health conditions, environmental degradation and economic impacts. The complexity of air quality data, which is affected by a variety of atmospheric conditions and human activities, makes its prediction difficult but crucial. Real-time monitoring and predictive analytics of air quality can play a key role in alerting people to harmful pollution levels and helping policymakers develop targeted emissions reduction and air quality control strategies.

4. **Proposed Approach**

The project will utilize extensive data from the OpenAQ data platform (https://openaq.org/about/initiatives/openaq-data-platform/). The platform aggregates real-time and historical data on air quality, specifically pollutants such as PM2.5 and PM10.

I will explore an LSTM (Long Short-Term Memory Network) time series forecasting model in order to. These models will help predict future air quality levels based on historical data trends and seasonal changes.

Data preprocessing will include missing value handling, normalization, and feature engineering to extract meaningful air quality predictors.

The performance of the models will be evaluated through rigorous cross-validation techniques. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared will be used to assess the accuracy and reliability of the predictions.

5. References

OpenAQ. (2023). OpenAQ Data Platform. Retrieved from https://openaq.org/about/initiatives/openaq-data-platform/

Shi, X., & Zhu, X. (2018). Machine Learning for Spatial Environmental Data: Theory, Applications, and Software. Environmental Modelling & Software, 51, 150-156.

Zheng, Y., Liu, F., & Hsieh, H. P. (2013). U-Air: When Urban Air Quality Inference Meets Big Data. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13), 1436-1444.