

GOKUL BALAGOPAL

Tel: +1(832)919-4205 **Address:** Richardson, TX, 75080

Email: gokulb3194@gmail.com

GitHub: <https://github.com/gokulbalagopal>

EDUCATION

PhD in Electrical Engineering

Aug 2019 – Present

University of Texas at Dallas, Texas (GPA - 3.53/4.00)

Master's in Electrical Engineering

Jan 2017 – May 2018

University of Houston, Texas (GPA - 3.70/4.00)

Bachelor of Technology in Electronics Engineering

Sep 2011– May 2015

Kerala University, India (CGPA - 7.04/10.00)

TECHNICAL SKILLS

Statistical Analysis, Machine Learning, Data Analysis

Programming language: Julia, Python, R, C++, MATLAB, Shell Script, Perl, JS, SQL

Libraries: MLJ, pandas, numpy, keras, scikit-learn, dplyr, ggplot,

Soft skills: Time Management, Communication, Team Player, Critical Thinker, Problem Solver

WORK EXPERIENCE

PhD Researcher at University of Texas at Dallas

Jan 2020 – Present

Characterizing the Temporal Variation of Airborne Particulate Matter using Variograms (Julia, R and Python)

- The primary objective of this study is to ascertain the optimal frequency of PM level measurements in a neighborhood, given that the Environmental Protection Agency (EPA) aggregates data on an hourly basis prior to its inclusion in their database.
- Currently there is no established methodology for quantifying the sudden variation in particulate matter (PM) concentration that transpired within the given hourly interval.
- The investigation is done in one of the most polluted areas of DFW, known as Joppa, and the frequency of measurements are calculated using a geostatistical method called variogram which is used for finding the location of gold nuggets in a mine.
- Measurement time ranged from 1 to 2 minutes for the actively polluted areas indicating how frequently the monitor must be active.
- An air parcel trajectory model called HYSPLIT was used to determine the course of polluted air, hence facilitating the identification of its origins.

Calibration of Airborne Particulate Sensors using Machine Learning (Julia and Python)

- The idea behind this study was to provide affordable and high accuracy air quality monitoring systems for low-income neighborhoods in DFW.
- Employed research-grade sensors, like Palas as reference and calibrated the cheaper monitors with them using machine learning. This doubled their accuracy and reduced the installation cost to less than one percent.
- Various regression models were used in our study and among them, neural network model yielded the greatest correlation coefficient of 0.99. A total of 17 measurements obtained from the cheaper nodes were utilized as the independent variables, while 6 measurements of PM levels were used as the target variable.
- Live Dashboards to visualize the time series data of the cheaper monitors were created using python's plotly package. This helps in detecting outages, anomalies in sensor measurement and successfully handling them.

Software Engineer – QBurst, Trivandrum, India

Jun 2015 – Jun 2016

- Coordinated with a team in developing an IOT based platform using Angular JS for client.
- Analyzed the social media data to understand the current trends and sentiments of the public towards beverages such as Pepsi, Coca-Cola using Facebook and Twitter API's.

ACADEMIC PROJECTS

Classification of Deepsat-4 dataset from Kaggle (Python)

- Processed data with over 4 million data points as batches of 1000s. Reshaped the data to 28 x 28 x 4 format, with Red, Green, Blue and Near InfraRed (NIR) channels.
- Neglected the NIR channels and normalized the data to obtain the color images to obtain clarity on how the satellite images differ. Created gray scale images after reshaping the data into 4,000,000 x 28 x 28 x 1 matrix.
- Singular Value Decomposition (SVD) was used to find that only 10 components were needed to explain more than 90% of the data. These 10 components were taken from the matrix obtained after doing SVD. The count 10 was obtained from the knee point in the SVD plot.
- Red, Blue, Green mean values for each of the images were also used as additional features.
- Used various machine learning classifiers like Logistic regression, Support Vector, Linear Discriminant Analyzer, Decision Tree, and Random Forest. Best performance of 96% was obtained using Random Forest.