LINGDUO LUO

+1(213) 359-9081 Los Angeles, CA, 90007

lingduoluo109@gmail.com ⋄ lingduo@usc.edu ⋄ linkedin.com/in/lingduo

ABOUT ME

- Self-driven, action-oriented data analyst with SQL/Python skills, adept at multitasking in data science and GIS.

EDUCATION

Master of Spatial Data Science, University of Southern California   August 2021 - May 2023
Relevant Coursework: Machine Learning for DS, Foundations of Data Management

Bachelor of Geographic Information Science, Sun Yat-Sen University   August 2017 - June 2021

Summer Session Visitor, University of California, Berkeley   July 2019 - August 2019

SKILLS

- **Programming Languages**: Python (Advanced), SQL (Advanced), R (Intermediate), Go (in progress)
- **Data Analysis**: Data Modeling, Big Data Analysis, Machine Learning
- **Tools and Technologies**: Scikit-learn, Github, Tableau, Spark, Hadoop, MATLAB, GeoPandas, SPSS, AWS, Firebase, DynamoDB, MongoDB, MySQL, NoSQL, Regex, ArcGIS Pro/online/desktop/API, Google Earth Engine

EXPERIENCE

Department of RS and GIS, Guangzhou Institute of Geography [Demo]   March 2020 - December 2020
Research Assistant   *Guangzhou, China*

- Utilized **Python**, **SQL**, and **JavaScript** to analyze urban land use changes and surface temperature, resulting in a **10x** improvement in computation speed and valuable insights into underlying patterns.
- Cleaned and preprocessed Guangzhou hourly weather API data using **Python** and **SQL**, ensuring high-quality data.
- Utilized **JavaScript** to operate **Google Earth Engine**, obtained sub-pixel landuse classification by linear mixture model.
- Operated **Google Earth Engine** using **JavaScript** to obtain sub-pixel landuse classification by linear mixture model, improving land use pattern analysis detailed accuracy by **13%**.
- Processed Landsat8 Surface Reflectance data, applied masks and calculated indices (NDVI and Land Surface Temperature).

PROJECTS

National Parks & Areas Travel Planner Web Application [Website] [Git]   February 2023 - May 2023

- Developed and deployed a **user-interactive web application** on PythonAnywhere, enhancing travel planning for U.S. National Parks Systems, simplifies data integration and exploration, delivering location-based insights to users.
- Implemented data collection processes via **web scraping** techniques using **Python** libraries (BeautifulSoup & requests) to gather National Park System data from Wikipedia, increasing data accuracy by **10%**.
- Leveraged **APIs** (OpenWeatherMap API & Google Maps API) to collect weather forecasts & driving directions respectively, utilized **Python** for data cleaning, preprocessing, handling diverse datasets to improve quality and consistency.
- Constructed website interface employing **Flask**, **HTML**, and **Python**, delivering interactive features such as route search, weather forecasts, and information about the destination park, bringing a **3X** reduction in time on travel planning.

Power Plant Energy Output Prediction and Analysis [Git]   February 2022

- Constructed a **multiple linear regression model** using 9600 records, accurately predicting the hourly electrical energy output of the plant, which **enhanced** the model's predictive capabilities.
- Optimized a sklearn **KNN regression model**, achieving a **20% reduction** in testing error by eliminating non-significant variables, thereby improving the model's predictive accuracy and reliability.

Medical Data Binary Classification [Git]   Jan 2022 - Feb 2022

- Conducted **comprehensive data analysis** on the Vertebral dataset, including data pre-processing, exploratory data analysis, and performance metrics calculation, leading to a more robust and reliable data foundation for model building.
- Implemented a **k-nearest neighbors algorithm** for binary classification, optimizing the model by exploring different distance metrics and weighting, which resulted in a low testing error of **8%**.