Xinbo Lu

Xinbo.lu@outlook.com | (206) 399-9918 | Seattle, WA
linkedin.com/in/xinbo-lu/ | github.com/shinbolu

EDUCATION

UNIVERSITY OF WASHINGTON, Seattle, WA
Master's in Information Management - Data Science and Business Intelligence track,   *06/2023*
- *Course Work: Machine learning, natural language processing, web development, Product Management, Business Intelligence application development, SQL and NoSQL database development*

UNIVERSITY OF WASHINGTON, Seattle, WA
Bachelor's in Geography - Geographical Information System; Minor in Informatics,   *06/2019*
- *Course work: Object Oriented Java and Python programming, database design, data structures, spatial statistics, system design, data science in R and Python*

PROFESSIONAL EXPERIENCE

BAW BAB TECHNOLOGIES, Seattle, Washington
Software Engineer - Data Services, September 2023 – Present; Software Engineer Intern, October 2022 – June 2023
- Developed to-C facing progressive web apps with features such as user onboarding, geolocation, geocoding, plus code calculation using Python Django, Google API, and OpenStreetMap to resolve the missing address difficulty in last-mile delivery in Kenya
- Designed and developed a full stack B-to-B web application for e-commerce and logistics companies to handle missing shipping information and route planning by utilizing Typescript, Prisma, NextJS, and PostgreSQL. The application is able to streamline user experience in the onboarding process, package delivery, route planning, and task scheduling
- Constructed 11 CI/CD data pipelines using Azure Data Factory and DevOps to consolidate user information and geospatial information, including user profiling, geolocation, geocoding, plus code generation. The consolidated data schemas helped the company to unify and standardize diverse user address inputs and improved logistics accuracy and efficiency by 34%
- Implemented 3 real-time data pipelines to monitor and process shipping progress and live position tracking of delivery by integrating 12 data sources using Kafka and pySpark
- Batch standardized and verified government-provided digital infrastructure files using pySpark. The new verified data reduced backend server and third-party API utilization rate by 80% month over month when observed using Azure Monitor
- Deployed spatial and computer vision machine learning models to classify buildings and other ground objects in Kenya to label uninformed communities by using remote sensing images, TensorFlow, and Mask-RCNN model

ESRI, Redlands, California
Advanced Analytics Intern, June 2022 – September 2022
- Collaborated with ArcGIS clients in the energy and manufacturing industry and translated their business requirements in operational efficiency and market shares to market and spatial analysis solutions by integrating customer segmentation, location-allocation, network analysis, market penetration, and site suitability analysis to their ArcGIS Online portal
- Implemented data engineering toolsets on top of ArcGIS Pro Toolbox and ArcGIS API for Python, which emphasizes validations for diverse data formats, geospatial attributes, and data outliers. The toolset evaluated and improved ArcGIS clients' data accuracy and quality by detecting null values, inconsistent data format with ArcGIS field definition, and invalid geospatial attributes including point, lines, and area shapes
- Constructed an iOS application during the hackathon event to automatically remind users to apply sun protection products based on the product description, real-time weather conditions, and user preferences. The development consists of designing UIUX prototypes, implementing a full stack application with Swift, which interacts with Weather API and ArcGIS SDK for iOS, and performing functional and performance tests

LONGFOR GROUP, Beijing, China
Software Engineer – Data Platform, July 2019 – August 2021
- Maintained, governed, and developed 40+ data pipelines in Apache Hive, Kylin, and Spark using SQL and Python by integrating source systems to assist interdisciplinary teams. Increased annual labor efficiency rate by 90+%
- Compiled 4 interactive Tableau dashboards and 17 reports for the rental and e-commerce division with topics including actuals, forecasting, risks, and revenue
- Developed Python machine learning models to analyze and predict house price, revenue, and location selection by constructing, training, and tuning statistics models such as multivariate regression, spatial temporal, and XGboost. All models had an 80+% accuracy and helped 600+ marketing and operation professionals increase 24% in net profits
- Built and improved operational and financial metrics backend system (data mart) using Python Django by adding 120+ indicators to help analysts monitor business health. The updated metrics system increased reports development by 70+%
- Containerized 11 analytics services covering customer profiling and segmentation, location-allocation, and corporate finance using Docker. The dockerized applications enabled more efficient use of system resources and reduced operational expenses by 15%, and improved Agile development life cycle efficiency by 28%
- Constructed and distributed 13 RESTful APIs and 6 Kafka MQ to cross-functional teams using AliCloud's OpenAPI Gateway and MQ for Kafka service. Improved real-time and near-real-time data processing efficiency by 50+%