Serah Varghese

Berkeley, CA 94704, USA
serahverg@gmail.com | 510-973-9246 | LinkedIn | Github:Serahverg-8

Availability: ***September 2023***

EDUCATION

University of California Berkeley - Berkeley,USA *August 2022 - **August 2023***
Master of Analytics *from Department of Industrial Engineering and Operations Research | GPA: 3.72/4.00*

Christ University - Bangalore,India *July 2016 - August 2020*
Bachelor of Technology in Computer Science Engineering | GPA: 3.88/4.00

SKILLS

Languages: Python, SQL, C++, Java, Shell Script, JavaScript, AMPL
Web/Cloud : HTML, CSS, Django, React.js, Node.js, AWS , Databricks, Snowflake, Langchain
Software: PowerBI, Excel, Git
Domain: Machine Learning, Data Engineering, Deep Learning, Optimization, Analytics, Prompt Engineering

WORK EXPERIENCE

Kone *May 2023 - August 2023*
Data Scientist

- Created a new ETL pipeline in Databricks to clean and transform the buisness tender data so that it can be consumed by the MLOps pipeline to identify improvements in winning tenders.
- Implemented feature engineering and selection techniques (filter methods, statistical methods & embedded methods) on the tender data that reduced the number of features by 65% while maintaining the roc-auc metric at 81.04% (Best Model XGBoost).
- Automated Sales Scorecard Reports by connecting various datasources, formatting the data using DataBricks, performing data warehousing techniques and analysing it using PowerBI which reduced manual effort by 82%.

Oracle Cerner *November 2021 - July 2022*
Data Analyst II

- Developed an ETL pipeline to collect and transform 600,000 ticket data in Python for storage in Snowflake.
- Identified 30 potential microservices to automate as part of the new IaaS platform using collected and mapped data from interacting with clients.
- Presented analysis to stakeholders using PowerBI reports to automate redundant manual processes and reduced man-hours by 36%.

Oracle Cerner *June 2020 - October 2021*
Software Engineer

- Managed server-side systems and support of 600+ Business Applications hosted on Windows and Linux servers using DevOps tools to manage infrastructure.
- Reduced the downtime of applications by 20% by analyzing server monitoring data using Python to pin down recurring alerts with large resolution.
- Developed a Django application to reflect real-time usage statistics of applications for prioritization of tasks through data pulled from F5 and stored in PostgreSQL.
- Developed a Python application to reduce the time taken for bulk JIRA creation by 80% for work and project tracking.
- Created an automated data pipeline to Azure data lake for reporting of productivity post 'Covid' from monitoring and logging tools using Jenkins, Python, and Selenium.

Oracle Cerner *December 2019 - May 2020*
Software Intern

- Developed a single dashboard to reflect monitoring data from 3 tools using APIs, SQL and Django for quicker access and data visualization.
- Increased efficiency by 15% through the development of an organization-wide application(Opencards) to streamline the hiring process using Microsoft PowerPlatform and Power Automate.

OTHER PROFESSIONAL EXPERIENCE AND CERTIFICATIONS

Vice President of Berkeley Analytics Consulting Organization (Client relations) *March 2023 - Present*
VP of BACO, a student-led analytics consulting organization at the University of California Berkeley that partners with Berkeley IEOR Analytics

Chartered Institute of Management Accountants certification, (CIMA), UK *July 2017 - April 2021*
Cleared Strategic Level, Modules: Strategic Management, Risk Management, Management Accounting,Financial accounting and Ethics.

RELATED PROJECTS

Transportation Optimization using Uber/Lyft Data | *Time Series Forecasting,PricePrediction,Optimization,Clustering* *February 2023*

- Predicted future demand using timeseries models such as SARIMA,XGBOOST to predict future demand. RMSE:23 with XGboost
- Predicted price using several models using gridsearch and cross validation.Accuracy over 90% with XGboost
- Clustering using Kmeans and optimization using gurobi to identify potential locations to open hubs based on demand.
- Identification of optimal routes through the nearest hubs for logistic usecases

Predicting Continuity of Yelp Business Listings | *Pandas, Matplotlib, Cross-validation, Feature Engineering* *November 2022*

- Predicted if a business listed on Yelp will close or continue operations using Python.
- Identified boosting algorithm to be the most accurate (85%) among six different models evaluated using bootstrapping.
- Applied NLP on a subset of the dataset to categorize reviews as positive or negative sentiments. Incorporated sentiments into the model to improve the accuracy by 1.25%.