

Anuj Saboo

778-862-8169 | saboo.anuj17@gmail.com | linkedin.com/anujsaboo | anuj-saboo.github.io

SUMMARY

Meticulous professional with 4+ years of experience winning 8 performance awards. Proficient in data engineering, statistical analysis, machine learning and natural language processing to generate insights and deliver value to decision makers displaying exemplary technical and leadership skills.

TECHNICAL SKILLS

Languages: Python, SQL, R, C++, HTML, CSS

Data Science Frameworks: Pandas, Scikit-learn, Spark, Hadoop, Keras, Spacy, NLTK, Plotly Dash, Power BI

Miscellaneous: Azure, AWS, Docker, Elasticsearch, MS SQL, SSIS, AWS DynamoDB

Horizontals: Data Engineering, Statistical Analysis, A/B Testing, Hypothesis Testing, Machine Learning, NLP

EXPERIENCE

Data Scientist

May 2020 – Present

Transport Canada

Ottawa, Canada

- Developed ETL pipelines, performed cleaning, feature engineering and statistical analysis using Azure Databricks and Data Lake storage to detect patterns and report KPI's to evaluate risk and safety for multimodal transport
- Produced business intelligence dashboards explaining features influencing risk and accidents using regression and tree based models, applying dimensionality reduction and training clustering models
- Led and designed an end-to-end NLP workflow to process transport incident text notifications, identify and translate language. Developed a word embedding based text similarity model to recommend similar incidents
- Devised a full stack web-app on Azure to perform Social Media Analytics in real time that helps quantify regulatory benefits using Sentiment Analysis and analyze topics using Graph Networks improving response time by 90%

Database Analyst

Aug 2014 – Mar 2017

Societe Generale

Bengaluru, India

- Database platform Subject Matter Expert leading the project transition at customer site delivering diverse and complex analytical and automation solutions using Python, Powershell and MS SQL Server & BI stack
- Engineered a pipeline for data collection, processing and modelling to monitor, cluster, measure and report server performance metrics through dashboards providing descriptive and predictive analytics for infrastructure planning
- Developed a text processing solution to process database and server logs in order to quantify root causes, analyse performance bottlenecks and automate service handling saving \$100K per year
- Programmed an in-house end to end data orchestration solution for data replication, reporting and high availability for 180 databases resulting in savings of \$200K per year for the organization

EDUCATION

Simon Fraser University - 4.18/4.33

Burnaby, BC

Master of Science in Computer Science, Big Data Specialization

Sep 2019 – Apr 2021

Manipal Institute of Technology - 8.09/10

Manipal, India

Bachelor of Engineering in Information Technology

July 2010 – May 2014

CERTIFICATIONS

Microsoft Azure | AZ 900: Azure Fundamentals, AI 900: Azure AI Fundamentals

May 2021

deeplearning.ai | Natural Language Processing Specialization

December 2020

Learning Tree | Advanced SQL Queries

December 2020

PROJECTS

London Commute Analysis | Python, Spark, Kafka, PostgreSQL, TextBlob, Dash

Oct 2019 – Dec 2019

- Developed ETL pipelines and built a UI to analyse performance and safety of London transportation network
- Utilized geospatial data to create cluster and heatmaps for Geo Analytics of accident hotspots and infrastructure

Flight Delay Prediction | R, Classification Algorithms, Paramter Hypertuning

Oct 2019 – Dec 2019

- Performed data cleaning, data analysis, developed and tuned a classification model to predict flight delays using algorithms such as Logistic Regression, SVM and Random Forest reporting causal factors for delays