Anuj Saboo

778-862-8169 | asaboo@sfu.ca | linkedin.com/anujsaboo | anuj-saboo.github.io

SUMMARY

Meticulous professional with 4+ years of industry experience winning 8 performance awards designing data focused solutions in banking, retail and transportation domains using engineering pipelines, analytics, 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, Airflow, Spacy, NLTK, Keras, Power BI, Docker, Azure
Expertise: Big Data | Data Engineering | Statistical Analytics | Machine Learning | NLP | Deep Learning

EXPERIENCE

Data Scientist
Transport Canada

May 2020 – Present
Ottawa, Canada

- Led and designed an end-to-end NLP workflow to process transport incident emails, identify language and recommend similar incidents to detect emerging risks, patterns and inform predictions
- Developed ETL pipelines, performed cleaning, feature engineering and statistical analysis on transportation accidents to segment data and detect patterns by applying dimensionality reduction, training clustering and anomaly detection algorithms
- Leveraged visualization and communication skills to explain risk indicators, analyze performance and cluster properties to answer challenging business problems to executives and stakeholders through interactive dashboards
- Developed and deployed a portal on Azure cloud to perform Social Media Analytics on twitter data in real time that helps understand public opinion using Sentiment Analysis and analyze topics using Graph Networks improving response time by 90%

Senior Developer
Mu Sigma

Mar 2017 – July 2017
Bengaluru, India

- Developed ETL workloads and designed dashboards for reporting customer activity of a digital marketing company
- Worked as a single point of contact for database hosting and performance tuning achieving high level of customer satisfaction

Database Administrator & Solution Designer
Societe Generale

Aug 2014 – Mar 2017
Bengaluru, India

- Spearheaded the design and deployment of analytical solutions using large scale data infrastructure using SQL and NoSQL data sources being the Subject Matter Expert leading the project transition from customer site at Paris
- Engineered a pipeline for data collection, preprocessing and regression modelling to monitor, measure and report infrastructure resource allocation through analytical dashboards
- Developed a text processing solution to process database and server logs leading to quantification of root causes, analysis of performance bottlenecks and automated service handling
- 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 - 4.14/4.33
Master of Science in Computer Science, Big Data Specialization

Burnaby, BC
Sep 2019 – Apr 2021

Manipal Institute of Technology - 8.09/10
Bachelor of Engineering in Information Technology

Manipal, India
July 2010 – May 2014

CERTIFICATIONS

- Natural Language Processing Specialization | deeplearning.ai | December 2020
- Advanced SQL Queries | Learning Tree | December 2020
- Analyzing Data with PowerBI | Learning Tree | June 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
- Applied sentiment analysis on twitter feeds to generate insights on customer feedback

Flight Delay Prediction | *R, Classification Algorithms, Paramter Hypertuning*   Oct 2019 – Dec 2019
- Performed data cleaning, exploratory data analysis and developed a classification model to predict flight delays
- Evaluated and tuned machine learning models using algorithms such as Logistic Regression, Random Forest, SVM and XGBoost reporting causal factors for delays

Strategic Asset Manager | *NLP, Stock Prediction, Dashboard, Chatbot*   Feb 2020 – Apr 2020
- Parsed edgar reports and applied sentiment analysis to be used as feature in the stock price prediction model
- Developed the web-based portal with chatbot integration to recommend investment decisions
- Used the BERT model to develop a system to answer questions about the company using its edgar reports

Find Me Love | *Recommendation Systems, Correlation, Clustering*   Feb 2020 – Apr 2020
- Built a Recommendation System on Libemseti dataset to suggest dating profiles using cold and hot start strategies
- Applied Content based and Collaborative filtering methodologies such as Correlation, Nearest Neighbor etc.

Real Time Speech Activated Assistant | *Object detection, OpenCV, Speech Processing*   Oct 2019 – Dec 2019
- Trained an image classification model and developed an application for object detection through speech processing and translation to classify everyday objects in a live video stream
- Worked with Google Open Images V4 dataset for training the model and explored creation of a custom dataset using web scraping and bounded box labelling

Extra Curricular

- Subject Matter Expert at Societe Generale for Automotive Leasing Division
- Director at Manipal Network, a media channel at Manipal University
- Co-Founder at DesiKapde (Apparel brand)