**EDUCATION:**

**Stevens Institute of Technology, Hoboken, NJ |** Master of Science in Information Systems (data science) | **Aug 2019 – May 2021**

**Uttar Pradesh Technical University, India |** Bachelor of Technology in Electronics and computer Engineering| **Sep 2009 – May 2013**

**SOFTWARE SKILLS**

**Languages**: Python, R, SAS, SQL, JavaScript, Scala, C, MATLAB

**Databases & Tools:** MongoDB, NoSQL, MySQL, PostgreSQL, Hadoop, Presto, Snowflake, Airflow, Spark, Hive, Salesforce, Git, Github

**Data Visualization:** Tableau, PowerBI, Quicksight, Qliksense, Matplotlib

**Cloud Technologies:** AWS (S3, Redshift, IAM, EC2, DynamoDB), GCP (BigQuery, Composer, Kubernets),Azure(databricks,datalake)

**Data Science:** Regression, Classification, Clustering, Feature Engineering, Hypothesis tesing (A/B test), ANOVA, t- test, chi- square, Deep Learning (CNN, RNN, LSTM), Natural Language Processing (BERT), Decision Tree, Naïve Bayes, KNN, LLM, Causal inference

**PROFESSIONAL EXPERIENCE:**

**Intuit Inc., California | Role: Data Scientist May 2023 – CURRENT**

* Build a predictive model to customer churn based on past behavior and engagement. Used logistic regression, decision trees, or random forest algorithms using python. Increased customer retention by 20%.
* Built a text classification model utilizing NLP techniques for mail classification, sentiment analysis and customer segmentation.
* Built a Data pipeline involving data extraction, data cleansing, data mining and modelling using python scripts, Redshift SQL queries.
* Performed hypothesis A/B testing and correlation analysis to study impact of promotions across merchant channels like salesforce.
* Conducted causal impact analysis of a marketing campaign using Difference-in-Differences and Synthetic Control methods
* Deployed, tested, monitored ML model on GCP using MLOps frameworks Docker, Kubernetes, Airflow, Kuberflow, Argo Workflows .
* Performed ANOVA, t-tests, and chi-square analyses using statistical packages (R, SAS) to uncover key insights on marketing strategy
* Developed, published, maintained Tableau and Quicksight dashboards to provide recommendation to customer success, sales team.
* Developed ETL, ingesting ~5Tb/day from iceberg, hive using PySpark, performed Spark optimizations to reduce runtime by 66%.

**FACEBOOK (contract via EPITEC), California | Role: Business Data Analyst 2 March 2022 - March 2023**

* Developed Performance matrix using NoSQL, Python, Tableau to track KPIs, maintained SLA, Performed UATs, for recommendation.
* Developed Python scripts for data validation, using pandas, optimized SQL queries and built ad-hoc Tableau dashboards and reports.
* Built Prediction model using python (Logistic regression, SVM) to forecast customer behavior and provided stakeholders insights using Tableau, Increased customer conversion rate by 21% by recommending improvement to supply chain enterprise team.
* Analyzed performance of new hardware using Python & Presto SQL to identify & bucket requests into categories, for feature launch.
* Collaborated with cross functional stakeholders to gather business requirements, define KPIs, write use cases to prioritize backlog.

**BRIGHT MIND ENRICHMENT, New York |** **Role: Data Science ML Engineer**  **September 2021 - March 2022**

* Leveraged Geospatial with Random Forest, Logistic Regression and K mean clustering to converting maltreatment and disease risk predictions into actionable intelligence that stakeholders can use to efficiently allocate limited child welfare resources.
* Performed Logistic regression, correlation and feature engineering (PCA) utilizing health care data to predict risk disease in patients.
* Developed predictive time series models using SAS, R, and SPSS to forecast healthcare cost trends and utilization patterns.
* Classified sentiment using Natural Language Processing (word to vector, tf-id, word embedding) on reviews, built ad-hoc SQL report.
* Built NLP Text classification model (named entity recognition), utilized LLM BERT, to find key words in legal & health documents.
* Deployed, monitored ML applications (CI/CD) on AWS utilizing PyTorch, TensorFlow and MLOps deep learning frameworks (Caffe).

**GRAVITE EDUVENTURE, India |** **Role: Decision scientist**  **February 2018 - June 2019**

* Performed K means clustering on customer data to provide customer segmentation according to customer purchasing habits.
* Developed Marketing Mix Models for ecommerce clients using Bayesian algorithm to study impact of promotions across channels.
* Implemented statistical modelling, A/B testing, built reports (Excel) for stakeholders to identify pain points, product improvements.

**HCL TECHNOLOGIES, India| Role: Analyst July 2016 - January 2018**

* Optimized Advanced SQL query by indexing and normalization to improve data retrieval, reduced processing time by 25%.
* Processed large data sets (1 TB) using python scripts to see patterns, reported using Tableau to give insights on customers
* Performed data wrangling, mining, modelling on big data using python, Advanced SQL and monitor KPI to provide business insights.
* Performed Exploratory Data Analysis (EDA), ANOVA using Excel, (index, pivot tables) and VBA, generated reports using Tableau.
* Implemented a time series forecasting model with ARIMA in SAS to forecast sales. Conducted fraud detection analysis using anomaly detection algorithms in R, identifying fraudulent transactions with 90% precision.

**J K T CONSULTING Pvt Ltd, India|** **Role: Analyst June 2015 - July 2016**

* Executed SQL stored procedure, designed data profiling packages in SSIS to uncover data quality issues, Developed ETL test scripts.