**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, Java, Scala, C, XML, Html5.

**Python Libraries:** Keras, TensorFlow, Numpy, Pandas, NTLK, SciPy, PyTorch, Pyspark, OpenCV, Matplotlib, Geopandas, libpysal

**Frameworks, Libs, Version Control:** nodejs, GraphQL, Presto Query Engine, git, github, bitbucket

**Databases & Tools:** MongoDB, PostgreSQL, Tableau, Looker, Power BI, Hadoop, Airflow, Spark, Hive, Databricks, Snowflake

**Cloud Technologies:**), GCP (Pubsub, GKE, Compute Engine, Spanner, BigQuery, Composer), MS Azure, AWS (S3, IAM, EC2)

**Machine Learning:** Regression, Classification, Clustering, Dimensionality Reduction, Ensemble Methods (Random Forest), Neural Nets, Deep Learning (CNN, RNN, LSTM), Natural Language Processing(BERT), Decision Tree, Naïve Bayes, LLM, GENAI

**PROFESSIONAL EXPERIENCE:**

**VISYS CLOUD TECHNOLOGY, California | Role:Sr Data Engineer May 2023 - CURRENT**

* Develope Python scripts for ETL processes, using Pandas, Pyspark to automate data ingestion, cleansing, transform, loading tasks.
* Implement, maintain Pyspark applications to extract raw payload from external systems through REST and SOAP APIs, applying transformations to ensure high-quality input for dimensional, fact models.
* Developed ETL, ingesting ~5Tb/day from iceberg, hive using PySpark, performed Spark optimizations to reduce runtime by 66%.
* Implemented REST API's using Python to manage data center resources, build CI/CD pipeline to automate steps in software development process. Utilized Google Kubernetes Engine, Pubsub, BigQuery for data warehousing to deliver business insights
* Collaborate with stakeholders to gather requirements, design, develop, maintain user-friendly & scalable systems, ML applications.
* Deployed, tested, monitored ML models on GCP using MLOps frameworks Docker, Kubernetes, Airflow, Kuberflow, Argo Workflows.

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

* Thrashed the landing-time of several pipelines by 30% by tuning, optimizing SQL queries and building ad-hoc data pipelines.
* Developed Performance matrix using SQL, Python, Tableau to monitor KPIs, maintained SLA, Performed UATs for recommendation.
* Designed Star, Snowflake Schema Data Models for Data Warehouse, created data dictionaries and maintained documents.
* Built Prediction model using python (regression, SVM) to forecast customer behavior, provided stakeholders insights using Tableau.
* Analyzed performance of new hardware using Python & Presto SQL to identify & bucket requests into categories, for feature launch.

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

* Developed Python scripts to automate data validation, cleaning process using pandas, improved process by 70%.
* Designed, developed scalable data pipelines using Databricks, Airflow, AWS to integrate, process data from systems (Workday, Jira).
* Built sentiment classification model using NLP techniques on reviews such as word embeddings (word2vec, TF-IDF), vector search.

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

* Implemented a time-series forecasting model using Prophet and BigQuery ML to predict demand trends, 12% reduction in stockouts.
* Built a recommendation system using XGBoost in Python, improved accuracy by 25% through advanced feature engineering (PCA).
* Developed a customer segmentation model using clustering algorithms (K-means, DBSCAN) to identify customer behavior groups.
* Implemented statistical modelling (SAS, R), performed hypothesis A/B testing, built reports for stakeholders to identify pain points.

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

* Developed a customer churn model (logistic regression, SVM, Naïve bayes) using python, performed PCA with .84 accuracy.
* Optimized Advanced SQL query by indexing and normalization to improve data retrieval, reduced processing time by 25%.
* Designed Star, Snowflake Schema Data Models for Data Warehouse, created data dictionaries and maintained documents. Developed ETL (Extract, Transform, Load) pipeline to scrape data using Pyspark REST APIs in JSON format, loaded them in MongoDB.
* Built and scheduled the dataflow architecture for refreshing Tableau dashboards using Python, Shell scripts and Spark SQL.

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

* Performed data wrangling & modelling on large data sets using python, SQL and monitor KPI to provide business insights.
* Automated 40% manual steps using Advanced SQL queries, analyzed CX data using SQL, python(pandas) for 150 Product categories.
* Designed ETL workflows in SSIS and Talend to integrate data from different data sources, MySQL, SQL Server, PostgreSQL.

**XOVIAN, India|** **Role: Software Engineer January 2014 - May 2015**

* Executed SQL stored procedure, designed data profiling packages in SSIS to uncover data quality issues, Developed ETL test scripts.
* Scheduled tests in Selenium using JUnit for regression testing of Actimize Watch List Filtering (WLF) in all environments.