

---

**EDUCATION****Missouri University of Science and Technology**

August 2022 – May 2024

Master of Science in Computer Science

- Coursework: Big Data, Business Analytics, Database Management Systems, Project Management, Algorithms

**Gandhi Institute of technology and Management**

August 2018 – May 2022

Bachelor of Technology in Computer Science Engineering

- Coursework: Data Mining, Operations and Supply Chain Management, Machine Learning, Computer Networks

---

**SKILLS****Programming:** Python (pandas, NumPy, SciPy, matplotlib), SQL (SQL Server, MySQL, PostgreSQL), R, PySpark**Analytics Tools:** Excel, Tableau, Power BI, Alteryx, Looker**Databases & Cloud Services:** AWS (Lambda, Glue Studio, Athena, Redshift, QuickSight), Snowflake, Salesforce**Machine Learning:** TensorFlow, OpenCV, Scikit-learn, A/B Testing, Regression, Clustering, Classification**Certifications:** AWS Solution Architect Associate

---

**WORK EXPERIENCE****Suhan Infotech Pvt Ltd, India**

Jan 2021 – June 2022

**Data Analyst**

- Produced real-time data pipelines using SQL & Python, enhancing the speed and reliability of financial transactions by 27%.
- Executed a comprehensive end-to-end data analytics system, incorporating custom visualization tools using **R**, **Tableau**, and **Power BI**; planned data accessibility for 15 cross-functional teams, accelerating decision-making processes.
- Leveraged **Excel**'s advanced functions & **VBA** macros to streamline financial reporting processes, reducing 30% report generation time.
- Crafted complex **SQL** queries to aggregate vast retail datasets, enhancing sales report accuracy by 35%.
- Maintained a relational database in **Salesforce** to store processed data using **SQL**, increasing data security by 50%.
- Unified a custom machine learning algorithm with **Python Scikit-Learn**, analyzing over 500,000 data points; eliminated data insights and decreased analysis time by 25 hours each month for the analytics team.
- Tracked real-time sales using **Power BI** dashboards to spot emerging trends, advancing peak season sales by 25%.
- Scheduled Python-based tools to automate customer segmentation, enhancing the precision of targeted marketing campaigns and optimizing effectiveness by 20%.

**Technocolabs Softwares, India**

April 2020 – September 2020

**Data Scientist**

- Led a smart contract vulnerability analysis project, utilizing advanced data labeling techniques, **Python (pandas, NumPy)** and **Unsupervised Machine Learning** to analyze over 500 contracts.
- Improved risk profiling accuracy by 30% by streamlining detection workflows and analyzing risk tags using **frequency** and **correlation** analysis and **SQL** to uncover prevalent vulnerabilities.
- Conducted validation of peer-generated data labels, achieving 95% accuracy rate and ensuring consistency across datasets of more than 10,000 entries with attention to detail, revamping data reliability.
- Built interactive dashboards in **Power BI** to visualize vulnerabilities and risk profiles, leveraging Python (**matplotlib**, **seaborn**) for advanced statistical visualizations, with optimized reporting efficiency by 40% for over 50 stakeholders.

---

**PROJECT EXPERIENCE****YouTube Data Pipeline**

- Deployed an ETL pipeline using **AWS** services to extract, transform, and load YouTube video statistics, converting JSON and CSV data into Parquet format, enabling efficient data querying with **Athena**.
- Engineered data ingestion and transformation workflows with **AWS Glue** and **Lambda**, minimizing manual intervention by 70% and ensuring real-time readiness of cleansed data.
- Designed an **AWS QuickSight** dashboard to visualize regional YouTube video performance, enabling real-time insights and improving decision-making by 60%.

**Network Intrusion Detection System**

- Implemented **Random Forest Classification** on the UNSW-NB15 dataset using **Python** to detect anomalies with 96% accuracy by processing inputs from network traffic data packets, serving as a firewall.
- Executed **SQL queries** to retrieve, preprocess, and validate network traffic data, Utilized **TensorFlow** and **Multimodal Classification techniques** to compute various intrusion probabilities, cutting back decision-making time by 30%.
- Applied **matplotlib**, **seaborn** data visualization, creating heat maps and charts for intuitive interpretation and designed a user-friendly interface for inputting network parameters and obtaining intrusion probability.

**HR Analytics Dashboard**

- Created an HR Analytics dashboard in **Power BI**, visualizing key employee metrics, enabling leadership to make data-driven decisions that refined workforce management efficiency by 30%.
- Automated employee data pipelines using Power BI and **Data Analysis Expressions (DAX)** to analyze datasets with 1,470 records, increasing data accuracy by 25% with faster, data-driven HR decisions.