Sarthak Bora

585 6259140 sb8708@rit.edu https://www.linkedin.com/in/sarthakbora https://github.com/SarthakBora12

Education

Rochester Institute of Technology, Rochester, NY, USA **Jan 2023 – May 2025**

Master of Science, Information Technology and Analytics

Technical Skills

Languages and Databases: Python, R, SQL, MySQL, MongoDB, PostgreSQL, SQL Server, pgAdmin, Neo4j.
Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Keras, Transformers, Spark (PySpark), LLMs(OpenAI, HuggingFace) Selenium, MLflow, RegEx, dplyr, Time Series Forecasting.
Statistics & Modeling: Predictive Modeling, Regression, Classification, PU Modeling, Hypothesis Testing, SAS, Stata
Visualization & BI Tools: Tableau, Power BI, Qlik, Alteryx, QGIS, ggplot, Looker, MS Excel, Matplotlib, Seaborn.
Tools & Platforms: GitHub, Docker, Kubernetes, Apache Airflow, Kafka, Data Warehousing, CRM, JavaScript.

Experience

Rochester Institute of Technology – Biomedical Engineering Department **Aug 2024 – May 2025**

Data and Administrative Assistant *Rochester, NY*

- Analyzed student performance, attendance, and lab engagement data using Python and SQL, identifying at-risk students and enabling faculty to improve academic support; contributed to a 10–15% improvement in engagement tracking.
- Built Power BI dashboards to monitor departmental expenses, TA hours, and supply usage, reducing monthly budget by 12%; also conducted NLP-driven research on AI in biomedical engineering to support faculty proposal development.

SeisoTech Pvt. Ltd. **Jan 2022 – Jan 2023**

Analytics Engineer *Pune,India*

- Automated and orchestrated 10+ ETL pipelines using Apache Airflow, deploying workflows on AWS EC2 to streamline data integration across systems; reduced data refresh lag by over 22% and enhanced pipeline scalability for real-time.
- Built and operationalized predictive churn models using Logistic Regression and XGBoost, achieving a 25% increase in retention forecasting accuracy; leveraged MLflow for experiment tracking and managed model versions on AWS S3.
- Architected robust data transformation workflows using Python and SQL, integrating CI/CD pipelines for automated testing and deployment, and optimized data flows into AWS Redshift for scalable analytics across teams.
- Created and maintained interactive dashboards in Tableau, Power BI, and Looker, connected to Snowflake; enabled self-service BI for business users and reduced manual report generation time by over 50%.
- Executed A/B testing and campaign analytics to evaluate the performance of marketing initiatives and product features; delivered actionable insights that informed content strategy, user segmentation, and experimentation decisions.
- Collaborated cross-functionally with stakeholders in product, sales, and engineering to define key metrics, validate data models, and deliver insights; documented reusable analytics queries and reporting templates to support team scalability.

StandardWings Technologies Pvt. Ltd. **May 2021 – Dec 2021**

Data Analytics Intern *Nashik,India*

- Streamed real-time GPS and IoT data using Apache Kafka and processed vehicle telemetry with PySpark, leveraging AWS S3 for centralized storage and achieving a 23% reduction in response time.
- Constructed a containerized admin dashboard using Flask and Docker, deployed on AWS EC2, enabling real-time monitoring and enhancing accessibility for operations teams.
- Pioneered interactive dashboards in Tableau and Power BI connected to AWS Redshift; applied time-series and clustering analysis, improving fuel efficiency and increasing scheduling accuracy by 18%.

Projects

Restaurant Recommendation System Using Yelp Data | *Python, BERT, SVD, Streamlit, NLP*

- Developed a hybrid restaurant recommendation system by combining BERT-based sentiment analysis with SVD-based collaborative filtering on 2.5M+ Yelp reviews, and launched an interactive Streamlit web app to deliver personalized, emotion-aware dining suggestions in real time.

Detection of Dyslexia Based on Eye Movements | *Python, SVM, Random Forest, Decision Tree*

- Engineered machine learning classifiers (SVM, Random Forest) to detect dyslexia from eye movement patterns, performed in-depth feature engineering on fixation and saccade data, and achieved 91% model accuracy to support early cognitive diagnosis and personalized educational interventions.

Credit Risk Prediction Model | *Python, SQL, Scikit-learn, AWS, Power BI*

- Created a credit risk prediction model using Logistic Regression and XGBoost on 500K+ financial records, engineered key features such as credit utilization and delinquency history, and visualized risk tiers in Power BI to support early identification of high-risk applicants and improve credit decision strategies.