VIJAY SIVAKUMAR

US Citizen | ☎+17325585267 | ✉ vijaysivakumarprof@gmail.com | 🔗 Linkedin | Portfolio | Github

EDUCATION

Rutgers University || New Brunswick, United States   August 2023 – December 2024
Master of Science in Data Science (GPA: 3.7/4.0)

Ramaiah Institute of Technology || Bangalore, India   August 2017 – May 2021
Bachelor of Engineering in Information Science

SKILLS

- **Programming and Tools**: Python, R, SQL, Java, Snowflake, AWS, Git, SAS, SPSS, MS Excel, Jira, Selenium, Flask, Power BI, Tableau, SAP ERP, SAP BODS, SAP S4 Hana.
- **Gen AI and ML**: Langchain, Hugging Face, Prompt Tuning, Fine Tuning, RAG, NLP, Deep Learning.
- **Libraries/ Frameworks**: Pandas, NumPy, Scikit-learn, PyTorch, Keras, OpenCV, Matplotlib, Seaborn.
- **Data Techniques**: Web Scraping, Neural Networks, Supervised and unsupervised Machine Learning, Exploratory Data Analysis, Data Visualization, A/B Testing, Time Series Analysis, ETL, Forecasting, Data Wrangling, Statistical Modeling, Feature Engineering, Data Cleaning.

WORK EXPERIENCE

Rutgers Health, New Jersey: *Research Assistant – Data Science*   08/2024 – 1/2025
- Analyzed clinical trial data from the IRON PREP study at IHHCPAR using Python and SAS; performed EDA and logistic regression to identify key treatment predictors and support healthcare decisions.
- Built preprocessing pipelines to clean, normalize, and encode health outcome data across 30+ features; improved model performance by 22% after feature selection.
- Analyzed over 10K+ records for the CAPHE project using regression and survival analysis, revealing a 5-year lifespan gap for Asian Americans under environmental stressors.

Mercedes Benz Research and Development, Bangalore: *Software Engineer*   08/2021 – 08/2023
- Collaborated on AI POC projects using ~200K diagnostics and telemetry readings; applied statistical thresholds and time-series analysis to detect anomalies for predictive maintenance initiatives.
- Automated 300+ end-to-end test cases across TIS, XRP, and XW apps using Selenium WebDriver and Java, reducing QA cycles by 50% and catching critical edge-case bugs pre-release.
- Developed a Python-based Jira REST API script to automate sprint board updates, streamlining task tracking and saving ~10 hours/week for the development and testing teams.
- Contributed to developing an after-sales recommendation system by analyzing ~50K service records; visualized KPIs using Matplotlib and Seaborn to support model validation and reporting.

Kimberly Clark, Bangalore: *Data and Analytics Intern*   01/2021 – 07/2021
- Built a live Supplier Data Quality dashboard in Power BI, integrating SAP MDG/ECC sources via BODS pipelines; flagged master data mismatches with >94% accuracy and generated $23.5K in annual savings.
- Engineered backend ETL workflows using SAP S/4 HANA and optimized SQL queries to reduce dashboard refresh time by 40%; boosted performance KPIs for business users.
- Conducted A/B testing on dashboard prototypes, improving load time by 40% and user engagement by 25%; implemented visual optimization based on feedback from supply chain analytics stakeholders.

PROJECTS

Customer Risk Intelligence System   March 2024
Developed a unified ML framework to predict loan default, fraud, and churn using XGBoost, LightGBM, Random Forest, and Logistic Regression (ROC-AUC up to 81%). Delivered SHAP-driven model interpretability and integrated Power BI dashboards for real-time customer risk profiling and portfolio-level risk analytics.

Examplify: LLM-Powered Exam Prep Tool   November 2024
Engineered an LLM-based exam assistant using fine-tuned LLaMA 3.2-1B and RAG architecture. Generated context-aware questions and flashcards from academic PDFs via FAISS and LangChain, achieving 74% accuracy and 0.83 F1 score. Optimized fine-tuning with QLoRA and 4-bit quantization for efficient, low-resource deployment.

Twitter Data Search Application   May 2024
Built a triune database architecture using Snowflake, MySQL, and MongoDB to manage 1M+ tweets. Built a Streamlit-based search app with sentiment analysis, topic modeling, and emotion detection. Improved query performance 70× through indexing and custom caching strategies.