SAHANA GIRISH

(217) 979-5487 | sgirish2@illinois.edu | **LinkedIn** | **Website**

EDUCATION

University of Illinois Urbana-Champaign   **August 2023 - May 2025**
Master of Science in Information Management (MSIM)   4.00/4.00 GPA

Visvesvaraya Technological University   **August 2017 - July 2021**
Bachelor of Engineering in Electronics and Communication Engineering (ECE)   3.64/4.00 GPA

SKILLS

Languages: Python, SQL, R, MATLAB
Libraries: Flask, Pandas, TensorFlow, Scikit-Learn, PyTorch, Keras, OpenCV, Request, NumPy, Matplotlib, SciPy, Seaborn
Technologies: AWS (SageMaker, Lambda, Glue, Batch, ECR), Azure (Data Factory, DataBricks, Stream Analytics), MLflow
Others: GitLab, Docker, Tableau, PowerBI, Jupyter, Spark, Talend, PyCharm, Linux, Statistics, Agile, Jira, MS Suite

---

WORK EXPERIENCE

Data Science Intern, **Werfen – R&D**   **August 2024 – Present**
- Applied statistical curve fitting techniques to digitize the instrument time series data to aid new feature development
- Formulated a robust assay flagging strategy to handle the reported false positives thereby reducing it by 96%
- Delivered customized datasets, extracted new features and optimized the existing pipeline through complex SQL queries leading to 62.85% faster runtime while adhering to GMP for quality, traceability, and regulatory compliance

Data Science Intern, **OSF Healthcare – Advanced Analytics**   **May 2024 – August 2024**
- Implemented LSTM and CNN time series forecasting models on ECG signals to predict short-term cardiac abnormalities
- Derived temporal, morphological and frequency domain features resulting in 95% average precision and 96% recall
- Designed insightful dashboards and assessed operational KPIs using statistical process control to support the mergers & acquisitions team in due diligence and market strategy analysis thereby enabling real-time monitoring of process variation
- Employed A/B testing and control group analysis to identify areas of synergy, operational strategies and potential risks

Data Scientist, **Comviva Technologies**   **September 2021 – July 2023**
- Led the design, development, containerization, and AWS SaaSification of an automated AI/ML product - Mobilytix AIx
- Expanded Mobilytix AIx by incorporating REST API, model performance visualization tools and deep learning modules
- Streamlined the cleaning and processing of large datasets to construct demand forecasting machine learning models, yielding an 8% to 10% reduction in customer churn
- Optimized interactive Power BI dashboards through query folding and data modeling for presenting performance metrics
- Supervised interns to meticulously extract, transform and load data by using Apache NiFi and Spark

Machine Learning Intern, **Compsoft Technologies**   **July 2021 - September 2021**
- Performed market analytics and sentiment analysis on social media comments using NBSVM, achieving 84% accuracy
- Increased the accuracy to 86.7% through NLP enhancements such as text cleaning, stemming, and tf-idf techniques
- Designed and orchestrated data pipelines with AWS, S3, and Glue, optimizing python scripts and MYSQL to enable efficient ETL process and data storage

---

PROJECTS

Business consultancy for a Korean-based motion capture company   **April 2024**
- Provided a data-informed B2B strategy to enter the US retail market with in-store analytics, BI and warehouse management system use cases by carrying out detailed market research, product and data analysis, and visualizations
- Performed cohort analysis to examine customer lifetime value and shopping behaviors, providing actionable recommendations for targeted marketing efforts, supply chain optimization and customer retention plans

Diabetes Prediction using Machine Learning Classifiers: Random Forest, Naïve Bayes and XGBoost   **January 2023**
- Processed the Pima dataset using 5-fold cross validation and maximized the metrics through hyper parameter tuning
- XGBoost outperformed Naïve Bayes with a sensitivity, specificity and AUC of 81.2%, 94.5% and 2.02% respectively

Customer Segmentation and Next Best Offer Recommendation on Bank Marketing Datase   **December 2023**
- Segmented high-value customers by K-Means clustering and RFM analysis for marketing strategies to improve retention
- Built predictive models to determine the propensity score of customers accepting offers and achieved a recall of 86.1%