SAHANA GIRISH

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

WORK EXPERIENCE

***Data Science Intern* - Werfen (R&D)** **August 2024 - December 2024**

- Integrated statistical curve fitting techniques and SQL to digitize 3B+ records of time series data for new feature development
- Formulated a scalable flagging strategy and threshold analysis, reducing false positives by 86%, thereby improving reliability
- Partnered with product managers in analyzing sales trends and UX, providing actionable insights for better product experience
- Delivered customized datasets and enhanced data pipelines through complex SQL queries, resulting in 62.85% faster runtime

***Data Science Intern* - OSF Healthcare (Advanced Analytics)** **May 2024 - August 2024**

- Forecasted cardiac abnormalities with 88% sensitivity by implementing dual-detector LSTM and CNN time-series ML models
- Built automated testing, validation and feedback loops using BERT and RLHF GenAI systems to monitor chatbot functionality
- Measured a 20% improvement in the response performance of an Angular website by conducting A/B testing and control group analysis to validate positive reinforcement
- Collaborated with cross-functional teams in sprint planning workshops to develop clear and aligned project roadmap flowcharts

Senior *Data Scientist* – Comviva **September 2021 - July 2023**

- Led the full-stack development, Docker containerization and AWS SaaS deployment of an automated AI/ML product
- Boosted customer acquisition efforts by integrating REST APIs, deep learning modules such as ANN and CNN, and Tableau for model performance visualization
- Streamlined the cleaning and processing of large datasets to construct demand forecasting and upselling/cross-selling machine learning models, yielding an 8% to 10% reduction in customer churn
- Deployed Qlik Cloud and Monitor, obtaining a 30% run-time reduction in the dashboard generation process
- Accelerated SDLC by implementing sprint iterations and CRM skills for SaaS, increasing Product Life Cycle velocity by 60%
- Supervised a team of 5 in executing big data ETL workflows utilizing Apache NiFi and PySpark, to optimize CI/CD pipelines

***Machine Learning Intern* - Compsoft Technologies** **July 2021 - September 2021**

- Performed market and sentiment analysis on social media and website reviews using NBSVM, achieving an 84% accuracy rate
- Increased model accuracy to 86% through NLP enhancements such as cleaning, lemmatization, stemming, and tf-idf techniques
- Refined interactive Power BI reports with data modeling and query folding, effectively presenting SEO performance metrics

EDUCATION

University of Illinois Urbana-Champaign **August 2023 - May 2025**
Master of Science in Information Management (MSIM) 4.00/4.00 GPA
Courses: Machine Learning on Cloud, Methods of Data Science, Data Warehousing and BI, Data Mining, Data Storytelling

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

SKILLS

Languages: Python, SQL, R, MATLAB
Libraries: TensorFlow, PyTorch, Keras, OpenCV, NLTK, Transformers, LlamaIndex, LangChain, OpenAI, Scikit-learn, Flask
Technologies: AWS (SageMaker, Lambda, Data Lake, ECR, S3, Glue), Azure (ML, Databricks, Data Factory), GCP, MLflow
Others: Git, Docker, Tableau, Power BI, Dataiku, PySpark, PyCharm, Qlik, Hadoop, Airflow, JavaScript, Angular, CSS

PROJECTS

Mental Health Chatbot Using Llama on Reddit Data **January 2025**

- Extracted mental health-related Reddit posts from subreddits using the PRAW API and applied text preprocessing techniques
- Fine-tuned a Llama model using GCP's Vertex AI and GPUs for multi-class emotion classification attaining a 0.89 F1 score

Business consultancy for a motion capture company to enter the US retail market **April 2024**

- Provided a data-driven B2B strategy with in-store analytics, BI, and warehouse management system use cases by conducting detailed market research, and cohort analysis to examine customer lifetime value and shopping behaviors
- Recommended a framework for targeted marketing, supply chain optimization, and client acquisition to drive lead generation

Customer Segmentation and Next Best Offer Recommendation for e-commerce **December 2023**

- Segmented high-value customers using K-Means clustering and RFM analysis for marketing strategies to improve retention
- Built sales prediction models to determine the propensity score of customers accepting offers, attaining a recall of 86.1%