

# SAIKUMARREDDY POCHIREDDYGARI

+1 315-742-4440 • spochire@syrr.edu • linkedin.com/in/saikumarreddyp • portfolio

## WORK EXPERIENCE

### Machine Learning Engineer, TATA Consultancy Services

June 2018 – Aug 2022

- Enhanced customer satisfaction to 95.44% through seamless migration of models to AWS, utilizing the AWS tech stack
- Pioneered a hybrid end-to-end MLOps CI/CD Architecture framework combining AWS IaaS, Databricks Spark Jobs, MLFlow, Snowflake, and Domino Model monitoring to optimize data processing and production model performance
- Developed and deployed a time series forecasting model to predict customer churn, resulting in a 15% reduction in churn rate. Utilized Prophet for forecasting and implemented an automated pipeline with airflow & Jenkins for model training, deployment, and monitoring.
- Designed custom XGBoost images and developed 3 innovative data pipelines workflows and 15 Airflow routines, reducing infrastructure billable hours by 10% & enabled training in distributed environments with low latency inference
- Evaluated the feasibility of building models, AI/ML/DL models, model development and A/B tests to predict user adoption for retail client. Explained impact of features on user adoption, informing optimized retention strategies and targeted A/B testing.
- Crafted Experiments and tested POC prototypes for finance clients, significantly speeding up business deal closures

### Research Assistant, CCDS, SU

Nov 2022 – Present

#### Data Science Duties

- Developed 25 innovative features for enhancing ML reproducibility using manual coding with MS Excel Spreadsheet
- Enhanced Geo-data Wiki development platform by streamlining navigation using Python and Docker resulting in good experience
- Executed large-scale data scraping and analysis of climate-related discussions from 1M Reddit posts and 5M tweets using Python
- Created a dynamic Image data Visualization Dashboard utilizing Django, Python, CSS, and JavaScript for glitch viewing

#### Data Engineer Duties

- Led thorough data models requirement analysis for the MongoDB to Azure data warehousing migration project
- Engineered efficient ETL data pipelines with Azure Data Factory, Databricks handling over 2TB of MongoDB data
- Implemented an incremental data update strategy, ensuring a smooth 2TB data migration with minimal downtime
- Developed a DAG for unstructured data extraction from application APIs resulting in time spent for data collection efforts by 70%

## TECHNICAL SKILLS

**Programming Languages:** Python, R, SQL, Pyspark, Scala

**Certifications:** Google Tensorflow Certified, Azure Certified Data Scientist, Azure Certified Data Engineer, IBM Data Science & AI

**Frameworks:** MLFlow, Flask, Django, DVC, GIT, XML, Kubernetes, Agile, TensorFlow, API, Apache Spark, Hadoop, Kubeflow, Seaborn, Matplotlib, PCA, UMAP, tSNE, FASTAPI, Pandas, Numpy, Json, SKlearn, Hive, Jenkins, Streamlit, Keras, Pytorch, BASH

**Database Management/ETL:** SQL, Mongo DB, Cassandra DB, NoSQL, MS SQL Server, Spark, Snowflake, SQL Database

**Machine Learning:** Linear/Logistic Regression, Support Vector Machines, XGBoost, Random Forest models, Decision Trees, k-means clustering, PyTorch, Keras, Tensorflow, Pandas, CNNs, RNN, LSTMs

**Cloud:** Azure ML Studio, AWS Sagemaker, Azure Databricks, Azure Devops, Synapse, Azure Machine Learning, Azure Data Factory, AWS Glue, AWS Redshift, AWS Codebuild, AWS Codepipeline, AWS Cloud Formation, AWS Athena, AWS Lambda, AWS StepFunctions, AWS Managed Kafka, AWS Devops, AWS Athena, AWS Glue, AWS Redshift, AWS S3, AWS SNS, AWS EKS

**Software Tools:** Github, Docker, MS Excel, MS PowerPoint, MS Word, JIRA, Tableau, Gitlab, Neo4j, Linux

## EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY

May 2024

**M.S. Applied Data Science**

*Relevant Coursework:* Natural Language Processing, Machine Learning, Business Analytics, Deep Learning, Statistics, Mathematics

## PROJECTS

**Fraud Transaction Anomaly Detection** – Tableau Dashboard, Data Engineering, MLOPS, AWS ([Link](#), [Docs](#))

Nov 2023 – Dec 2023

- Heightened transaction security by predicting fraudulent activities through a statistical modeling approach
- Innovated by engineering 17 pivotal features, improving the detection of fraudulent transaction activities
- Realized a 17% improvement in the F1 Score for fraud detection and deployed an optimized model to AWS with a development of Tableau dashboard for continuous performance monitoring

**Bad Server Activity Detection Challenge** - Python, Validation, Statistical Modelling, Kubernetes, AWS ([Link](#))

Nov 2023 – Dec 2023

- Engineered and operationalized a machine learning model on AWS, utilizing Amazon Kinesis for real-time data streaming and Amazon EKS for scalable model deployment
- Innovated in the realm of feature engineering by developing 18 distinctive features registered them in Hive meta store and Redshift, pivotal in elevating the model's predictive accuracy
- Accomplished a 58% enhancement in the F1 Score for detecting server anomalies, significantly bolstering system dependability, and reducing downtime

**Facebook Friend Case Study** - Python, Machine Learning Algorithms, EDA, Visualization, Statistics, Graph ([Link](#))

Mar 2022 – Apr 2022

- Assessed potential link between two users to enable better user experience by using machine learning and large datasets analysis
- Performed feature engineering to create 52 features for each user in a graph and modelled parameters that best fit the algorithm
- Improved F1 Score by 7% in identifying link between two user nodes in a graph network using machine learning