Megha Gulati

Data Science Portfolio GitHub meghagulati03@gmail.com +91-9818946463

Technical Skills

Machine Learning: Classification (Logistic Regression, Random Forest, Decision Trees, XGBoost), Model Evaluation (ROC-AUC, F1, Precision/Recall), Time Series Forecasting, Feature Engineering, Risk Scoring, Imbalanced Learning (SMOTE)

Deep Learning: Artificial Neural Networks (ANN), Convolutional Neural Network (CNN),Recurrent Neural Networks (RNN), TensorFlow, Multi-Agent Systems

Programming Languages: Python, SQL, Java

Libraries/Packages: Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, Matplotlib, Seaborn, PyMC, TensorFlow

Cloud Platforms: Google Cloud Platform (GCP), Azure ML, Georgia Tech PACE HPC Cluster

Tools Frameworks: Jupyter Notebook, VS Code, Git, Tableau, Selenium, Serenity, YOLOv5

Professional Experience

Publicis Sapient — Client: Goldman Sachs **2018 – Present**

Data Scientist

- Developed machine learning models for **Credit Default Risk Prediction** using classification algorithms such as **Logistic Regression, Random Forest, and XGBoost**, optimizing for ROC-AUC and F1-score to improve recovery prioritization and reduce false outreach.

- Integrated **multi-bureau credit scores** (FICO, TransUnion, VantageScore) and modeled repayment behavior using delinquency history, charge-off risk, and hardship program enrollment patterns.

- Designed an interpretable **risk grading system (A to D)** based on model probabilities and customer payment data, used to drive collections comms strategies.

- Aligned model outputs with downstream business actions, including **hardship and settlement program recommendations**, supporting both **regulatory compliance** and effective collections.

Automation Quality Engineer

- Led end-to-end QA strategy for financial recovery systems, identifying critical defects pre-launch and improving product reliability across high-stakes banking workflows.

- Enhanced data validation and test automation frameworks, ensuring that new data models and predictions maintained high accuracy and reliability across diverse banking workflows..

- Built and maintained automation frameworks using Java, Selenium, Rest Assured, Serenity, reducing regression testing time by 70% and accelerating CI/CD releases.

Projects

Bayesian Air Quality Forecasting | *Python, PyMC, Jupyter, Time-Series, Matplotlib*
Developed a Bayesian Inference model using PyMC to forecast PM2.5 air pollution levels in India.

Applied **Gibbs Sampling** and Bayesian ARIMA models to capture temporal dependencies in air quality data.

Face Mask Detection Using YOLOv5 | *Python, Deep Learning, Computer Vision, YOLOv5, Georgia Tech PACE HPC Cluster*
Trained a YOLOv5 model on a Kaggle dataset to detect face mask usage with mAP@0.5 = 74%.

Utilized Georgia Tech's PACE HPC Cluster for scalable model training, enabling faster experimentation with large image datasets.

Performed data augmentation and hyperparameter tuning to improve model performance.

Consumer Debt Analysis | *Python, Jupyter, Tableau, Time-Series, Matplotlib, Plotly*
Analyzed U.S. consumer debt trends and predicted repayment likelihood using ML models.

Education

Georgia Institute of Technology **Aug 2023 – Present**
Master of Science in Computer Science (OMSCS) — Specialization: Machine Learning *Online*

Guru Gobind Singh Indraprastha University **Aug 2013 – Aug 2017**
Bachelor of Technology in Information Technology — Graduated with 83.5% *New Delhi, India*