Manoj Kumar Ashok

(312)-284-9898 | mashok@depaul.edu | LinkedIn | GitHub | Portfolio | Chicago, IL

Education

DePaul University Chicago, IL
Master of Data Science, Concentrated on computational methods *Jan 2024 - Nov 2025*

Coursework: Data analysis and Regression, **Mining Big Data**, Advanced Machine Learning, Fundamentals of Data Science, Advanced Data analysis, Database processing for large scale analytics, **Neural Networks and Deep Learning**, **Natural Language Processing**.

Bharathiar University India
Bachelor's in Computer Applications - AI *July 2020 - Nov 2023*

Coursework: Python programming, Data Structures, Intro to AI and ML, statistics, NLP, Computer vision

Technical Skills

Languages: Python, SQL, R
Data Engineering: Apache Spark, Hadoop, Airflow, ETL
Databases: MySQL, Hive, MongoDB
Cloud & Pipelines: Azure Data Factory, Google Dataflow
Visualization: Tableau, Matplotlib, Seaborn
Version Control: Git
Also Familiar With: HTML, CSS, JavaScript, Oracle, ggplot2

Professional Experience

Research Assistant – Machine Learning (GANs) Jan 2025 – Present
DePaul University (Prof. David Ramsay) *Chicago, IL*

- Spearheaded research on Conditional GANs, integrating IcGAN and RoCGAN architectures to develop advanced AI-driven image enhancement models.
- Boosted model realism by **30%** (evaluated via FID and IS) and reduced training time by **40%** through hyperparameter tuning and pruning.
- Collaborated across departments to design scalable ML pipelines and validate models for real-time deployment in imaging applications.

Software Developer – Data Engineering Team Jan 2023 – July 2023
Zoho Corporation *Chennai, India*

- Developed ETL pipelines using Python and SQL to process CRM/Books datasets, increasing data throughput by **20%**.
- Automated Tableau dashboards using advanced SQL queries, reducing reporting lag by **24%**.
- Built KPI-tracking models to identify and resolve customer support inefficiencies, accelerating resolution speed by **10%**.
- Implemented and monitored Airflow DAGs for reliable data orchestration and reduced batch latency by **15%**.

Projects

Heart Disease Prediction Using ML | Logistic Regression, Gradient Boosting, MLP, Random Forest Mar 2025

- Achieved **88.6% accuracy** and **93.9% ROC-AUC** on clinical data using ensemble classifiers.
- Boosted recall by **21%** via stratified sampling and Random Forest feature selection.
- Outlined integration for real-time deployment in wearable devices and triage systems.

Predictive Analysis for Credit Limit | Python, Scikit-Learn, TensorFlow, SQL July 2024

- Improved credit limit prediction accuracy by **20%** using Ridge and Lasso regression models.
- Reduced anomalies by **25%** through outlier filtering and data normalization.
- Streamlined batch job automation via SQL scripting to boost data throughput by **50%**.

Bankruptcy Prediction Using Ensemble ML | XGBoost, LightGBM, Random Forest, SMOTE Mar 2025

- Trained ensemble models on 6.8k-record financial dataset, reaching **98.5% accuracy** and **97.6% F1-score**.
- Utilized SMOTE for minority oversampling, increasing recall by **42%** while maintaining precision.
- Designed full ML pipeline from data prep to tuning, deploying a production-ready ensemble model.