HARTHIK MANICHANDRA VANUMU

+91-7676452975 | Gmail | Linkedin | Github | Portfolio Website

Education

Manipal Institute of Technology (MIT) Bengaluru 07/2023 – 07/2027
B.Tech in Computer Science and Engineering (Artificial Intelligence)

Projects

End-to-End Predictive Analytics Pipeline & Optimization Framework | *Python, ETL, Optuna, SQL*

- Engineered an automated **ETL pipeline** to extract, clean, and merge disparate datasets (JSON/YAML) into a unified schema, utilizing median imputation and Yeo-Johnson transformations to ensure data quality.
- Defined and modeled **Key Performance Indicators (KPIs)** for season-long player valuation, transforming raw ball-by-ball records into actionable metrics for strategic scouting and decision support.
- Developed a **Joint Optimization Framework** using Optuna to simultaneously tune data augmentation (SMOGN) and model parameters, solving data scarcity issues where traditional sequential pipelines failed.
- Achieved a Test R^2 of **0.9075** and reduced forecast error margins (RMSE $\approx$ 19.5 runs), providing reliable quantifiable benchmarks for performance analysis.
- Proven **cross-domain scalability** by validating the pipeline on a medical insurance dataset (1,300 rows), demonstrating the system's robustness for diverse business regression tasks.

RBI NEFT Financial Data Analytics Platform | *Python, Flask, ETL, MySQL, Pandas*

- Engineered an automated **ETL pipeline** (Selenium, BeautifulSoup) to scrape and ingest unstructured financial datasets (Excel) from RBI repositories, transforming 7+ years of historical records into a structured **MySQL data warehouse**.
- Designed and implemented a **Business Intelligence dashboard** using Flask and Jinja2, enabling self-service querying and visualization of critical banking KPIs such as transaction volume and value volatility.
- Optimized data retrieval workflows by architecting efficient SQL schemas, facilitating complex trend analysis across thousands of banking nodes and reducing manual reporting latency.
- Developed dynamic visualization modules (Matplotlib) to benchmark **market share performance** (Top 10 Banks) and identify seasonal transaction patterns for strategic financial assessment.

AWS Cloud-Based ML Pipeline for Cost-Sensitive Classification | *Python, AWS (EC2/S3), Fuzzy Logic*

- Architected a **scalable inference pipeline** on **AWS** (EC2/S3) for medical diagnostic classification, ensuring high availability and data durability for model training artifacts.
- Built a **risk-management engine** using Fuzzy Logic that reduced safety-critical **False Negative rates by 50%** (0.105 to 0.053) by automatically deferring low-confidence predictions.
- Implemented a **cost-optimization module** that reduced overall **selective risk by 39%** (0.488 to 0.297), optimizing rejection thresholds to minimize financial and safety penalties in high-stakes scenarios.
- Designed a **robust validation workflow** (Three-Way Split) to audit model performance, ensuring reproducible deployment metrics across production-like environments.

Technical Skills

Languages: Python, SQL, C
Cloud & Infrastructure: AWS (EC2, S3), MySQL, Relational Databases, Git/GitHub
Machine Learning & AI: Regression & Classification, Transformers (Hugging Face), Tree-Based Models (XGBoost/CatBoost), Cost-Sensitive Learning, Fuzzy Logic, Optimization (Optuna)
Data Engineering: ETL Pipelines, Web Scraping (Selenium, BeautifulSoup), Data Cleaning & Imputation, Pandas, NumPy
Visualization & BI: KPI Definition, Dashboarding (Flask/Jinja2), Statistical Analysis, Matplotlib, Seaborn
Tools & OS: Jupyter Notebooks, VS Code, Linux/Unix, Excel

Publications

Performance Degradation of Deep and Traditional Machine Learning Models Under Market Volatility: An Inter-Day Stock Trend Comparison

- Accepted for Presentation at **IEEE INDICON 2025** (The flagship IEEE India Council conference).

A Machine Learning Framework for Data-Scarce Regression using SMOGN with Joint Hyperparameter Optimization: A Case Study with Cricket Performance Prediction

- Accepted for Presentation at **IEEE TENCON 2025** (The flagship conference of IEEE Region 10, Asia Pacific).

An Explainable and Resource-Efficient Transformer Pipeline for CPU-Based Document Summarisation and Question Answering

- Accepted for Presentation at **AICCoNS 2026**.

Key Achievements

Recipient, IEEE Student Travel Grant (2025)

- Awarded a competitive travel grant of **INR 25,000** from the IEEE Bangalore Section to present research findings at a IEEE TENCON 2025 (The flagship conference of IEEE Region 10, Asia Pacific).

1st Place, RoboRun Competition & TechTatva Selection

- Won MAHE Bengaluru's university-wide 'RoboRun' line follower robotics competition (1st Place).
- Subsequently, selected for the official, institutionally-funded MIT-Bengaluru team at TechTatva (MIT-Manipal's technical fest).