MANOGNA TUMMANEPALLY

Tampa, Florida | +1 (813) 869-1456 | manogna30@gmail.com | LinkedIn | GitHub | Portfolio

Education

University of South Florida, Tampa FL — **Aug 2023 - May 2025**
Master of Science, Artificial Intelligence & Business Analytics — *Tampa, Florida*

Jawaharlal Nehru Technological University, Hyderabad — **Jul 2018 - Aug 2022**
Bachelor of Technology, Information Technology — *Hyderabad, India*

Work Experience

DXC Technology | *Jr Data Scientist I* — **Oct 2021 - Jul 2023**

- Created ETL workflows using Integrate.io as per business requirements to collect, clean, and transform approximately 1.2 million records from customer and healthcare data sources, streamlining data readiness.
- Performed exploratory data analysis (EDA) on structured datasets covering 15+ variables, uncovering trends in customer behavior and service usage that supported budgeting and operational recommendations.
- Designed predictive models to estimate outcomes such as sales performance, customer churn, and healthcare utilization, improved model performance by 12–15% through targeted feature engineering and parameter tuning.
- Conducted A/B testing to compare marketing strategies for a retail client, contributing to a 35% boost in engagement, deployed models to production using AWS Lambda and S3, maintaining over 99% uptime for real-time predictions.

Verzeo Edutech Private Limited | *ML Intern with Python* — **Jun 2020 - Aug 2020**

- Developed a Python-based weather application that pulled real-time climate data from OpenWeatherAPI for over 15+ global locations; implemented ETL logic to extract, parse, and organize JSON responses for features like temperature, humidity, and wind speed.
- Conducted EDA on retrieved weather data to uncover geographic and seasonal patterns; built a temperature forecasting model using linear regression, achieving an average prediction error under 2.5°C, providing enhanced user-level insights.

Project Experience

Synthetic Voice & Health Data Generation — **Apr 2025 - May 2025**

- Generated a privacy-preserving synthetic dataset from real voice and health survey data (1,109 features, 442 rows) using generative models like CTGAN and TVAE, ensuring no sensitive information was copied.
- Engineered preprocessing pipelines to handle 96% missing data, applying imputation and dimensionality reduction to stabilize inputs for model training.
- Achieved 95%+ statistical similarity with the real dataset using KS-Statistic and Correlation Similarity, enabling safe sharing of high-fidelity synthetic data for ML applications.

Hate Speech Detection in Memes

- Built a multimodal hate speech detection system using text and image inputs from the Facebook Hate Meme Dataset, integrating TF-IDF, pre-trained CNNs, and dense neural networks to classify memes with a peak accuracy of 69.4% and F1 score of 51.4% across models.
- Designed three independent pipelines: CNN-based image classifier, a text-based classifier using TF-IDF + SVD, and a combined model for feature fusion, optimized preprocessing with lemmatization, normalization, and augmentation, improving recall in multimodal setups by 13% over unimodal baselines.

BAT: Deep Learning Methods for IDS in Wireless Networks

- Developed a deep learning-based intrusion detection system using the NSL-KDD dataset, achieving 84.25% classification accuracy, outperforming traditional CNN and RNN baselines by 4.12% and 2.96%, respectively.
- Designed the BAT-MC architecture, combining multiple 1D convolutional layers, BLSTM networks, and an attention mechanism to automatically extract and focus on high-impact features from raw network traffic.
- Preprocessed and normalized over 125,000 network traffic records, applying one-hot encoding and scaling, and transformed byte-level packet data into structured tensors for model training without manual feature engineering.

Technical Skills

- **Programming Languages:** Python, C++/C, Java, Pyspark, SQL, Java
- **Libraries:** PyTorch, TensorFlow, OpenCV, Scikit-learn, Pandas, Matplotlib, NumPy, Seaborn, Theano, BeautifulSoup
- **BigData Ecosystem & ETL:** Apache Spark, Airflow, Hadoop, Hive, Kafka, Azure Synapse, BigQuery
- **Tools:** Git, Docker, Kafka, GCP, AWS, Azure, CMake, FastAPI, Snowflake, Databricks, PowerBI, Tableau
- **Statistical Tools:** Descriptive and Inferential Statistics, Hypothesis Testing, A/B Testing, Time Series, ANOVA, Survival Analysis, Gradient Boost, ADA Boost, XGBoost
- **DataBase:** PostgreSQL, MongoDB, MySQL, OracleDB