Syam Kumar Mandapati

Hyderabad | msk041999@gmail.com | 9010390210 | Linkedin | GitHub | Portfolio

Professional Background

Data Scientist with **3.5 years** of experience in **Machine Learning (ML)**, **Natural Language Processing (NLP)**, **Deep Learning**, **Time Series**, and **Generative AI**. Expertise in **predictive modeling**, **time series forecasting**, and **statistical analysis** using **Python**, **Scikit-Learn**, **TensorFlow**, and **PyTorch**. Proficient in **end-to-end AI development** from **EDA**, **preprocessing**, **feature engineering**, **model building**, **model evaluation** to deployment using **Flask API** and **Microsoft Azure cloud services**.Currently deep diving into **Agentic AI architectures**, **RAG implementations**, **vector databases**, **embeddings**, and exploring **LangChain** framework.

Primary Skills

Technical Skills

- Machine Learning (ML)
- Deep Learning (DL)
- Natural Language Processing (NLP)
- Time Series Forecasting
- Exploratory Data Analysis (EDA), Data Preprocessing, Feature Engineering, Model building and Model evaluation
- Clustering
- Generative Artificial Intelligence (Gen AI)
- Prompt Engineering
- Multi Agent Orchestration (Agentic AI) creating end to end workflows using agents

Programming Languages, Frameworks, and Libraries

- Python, TensorFlow, Scikit-Learn, nltk, spacy, Numpy, Pandas, Open AI, PyTorch, PyCaret , Flask, Rest API's, DB's (SQL, MongoDB and FAISS vector DB - Basic understanding)

Development Tools and Environments

- GitLab, Postman, Linux, Putty, WinSCP, Visual Studio Code (VS Code), Spyder, Eclipse, Jupyter Notebook, Anaconda

Additional Skills

- Celery, Redis (Microservice Architecture)
- Automation and Scripting (e.g. Python Scripting)
- Microsoft Azure
- Agile Methodologies
- Soft Skills: Problem-Solving, Critical Thinking, and Collaboration

Professional Summary

- Explored and analyzed **20+ datasets** over 3 years, gaining expertise in **data distributions** (Gaussian, Poisson, Uniform), descriptive statistics and **hypothesis testing** (Z-score, T-score, null and alternative hypothesis).
- Conducted **Exploratory Data Analysis (EDA)**, and visualization using **Plotly, Matplotlib, and Seaborn**.
- Proficient in handling **missing values** using imputational methods and identifying/replacing **outliers**, plus **feature engineering** (encoding, scaling techniques: **Standard Scaler, Min-Max**; transformation techniques: **log10, Box-Cox, square, square root**) and **feature selection** (**recursive feature elimination, Lasso(L1), forward/backward selection**).
- Handled **class imbalance** using **SMOTE, SMOTETomek, and random over/under sampling**, and performed **cross-validation** (**K-Fold, Stratified K-Fold**) to ensure robust model performance.
- Developed and deployed **end-to-end ML solutions** including **Boeing customer tool price prediction**, **Cargill**

Major Incident prediction, **Tata Steel ticket assignment**, and **supply chain demand forecasting** (SKU/category level).

- Expertise in **supervised and unsupervised learning algorithms** such as **Random Forest, Logistic Regression, XGBoost, LightGBM, KNN, Naive Bayes, SVC, K-Means**, and **DBSCAN**.
- Improved model performance by **15% accuracy** and reduced runtime by **20 minutes** through **GridSearchCV** and **RandomSearchCV** for hyperparameter tuning and parallel processing with **n_jobs**.
- Built and optimized **NLP pipelines** using **text cleaning** (regular expressions, stop words removal, tokenization, lemmatization), **vectorization** (TF-IDF, Bag of Words, Word2Vec, Sentence Transformers - all-MiniLM-L6-V2), tokenizer and padding, along with **NER, parser, and tagger**, achieving **87% accuracy for sentiment analysis, 85% accuracy for ticket classification, and 99% for Major Incident prediction**.
- Implemented **clustering solutions** using **K-Means, DBSCAN, and Hierarchical Clustering**, leveraging **Sentence Transformers** for semantic understanding and **Elbow Method and Silhouette Score** for optimal cluster selection.
- Integrated **Generative AI solutions** using **Azure OpenAI**, reducing costs by combining **K-Means clustering with prompt engineering** for concise cluster naming.
- Utilized **pretrained models** like **XLM-RoBERTa from Hugging Face** for sentiment analysis and **Azure Speech-to-Text** for multilingual audio transcription.
- Proficient in traditional models (**ARIMA, SARIMA, SARIMAX, Prophet, Exponential Smoothing**) by ensuring stationarity (**ADF/KPSS tests, differencing**) and identifying parameters (**ACF/PACF**).
- Skilled in advanced deep learning models (**N-BEATS, TFT, NHITS, LSTM, GRU**) via Darts library to handle complex stochastic trends, seasonality, and noise created covariates to handle multiple seasonalities in data.
- **Deep Learning Expertise**: Proficient in tuning **batch size, epochs, and learning rates** for optimal performance, implementing **optimizers** (ADAM, SGD) and **activation functions** (Sigmoid, TanH, ReLU), and using **dropout and pooling layers** to prevent overfitting and reduce computational complexity.
- **Deep Learning Architectures**: Hands-on experience with **RNNs, LSTMs, and GANs** for tasks like **sequential data analysis, predictive modelling and synthetic data generation**, including addressing **vanishing gradients** and implementing **forward/backward propagation**.
- Evaluated models using **classification metrics** (**accuracy, precision, recall, AUC score, specificity, sensitivity**) and **regression metrics** (**Rsquare, Adjusted Rsquare, MAPE, MAE, MSE, RMSE, SMAPE**), addressing **high variance and high bias** scenarios.
- Deployed **GPT** and **embedding models** in **Azure AI Foundry**, configuring API connections for **multi-agent systems** and autonomous workflows.
- Developed a workflow using multiple agents on the **WisdomNext** platform to orchestrate the SDLC, where **prompts generate products** and automatically push code to **Git** with proper orchestration.
- Started Developing **Autonomous E2E Inventory Management System powered by Agentic AI** on **TCS WisdomNext** combining **Rapid AI forecasting** with **GenAI reasoning** for end-to-end automation.
- Created **Reasoning AI agent workflow** that interpret **RapidAI predictions** and generate explanation about why model has predicted this value(provided each feature addtional information and feature importance score of ML model).

Experience

Project - 1: TCS Cognix™ Rapid AI (Current)

November 2022 – Present

Client: Tata Consultancy Services
Role: Data Scientist

Roles and Responsibilities:

- **Developed and maintained Jupyter notebooks** for various datasets, streamlining **data preprocessing, model development, and evaluation.**
- **Designed and deployed AI solutions** tailored to business needs, ensuring automation and data-driven decision-making.
- **Optimized AI workflows** by implementing efficient **data handling, feature engineering, and model tuning** techniques.
- **Integrated multiple ML and DL algorithms** into the platform, improving prediction accuracy and clustering

capabilities.

- **Developed E2E time series forecasting solutions** to business data.
- **Worked on speech-to-text transcription** and **audio data analysis**, enabling **multilingual processing.**
- **Conducted rigorous testing and Validated models** using appropriate **evaluation metrics** to ensure reliability and scalability.
- **Documented processes and findings**, ensuring reproducibility and knowledge sharing.
- **Collaborated with cross-functional teams** to translate business problems into AI-driven solutions.

Project - 2: Machine First Delivery Model™ (MFDM™).AI

March 2022 – October 2022

Client: Tata Consultancy Services
Role: AIML Developer

Roles and Responsibilities:

- **Developed multiple Jupyter notebooks** for various use cases, including **Topic Modeling and Sentiment Analysis.**
- **Built predictive models** for **app installation through ads** using **machine learning algorithms.**
- **Developed AI solutions** for **employee attrition prediction** and **bank customer churn analysis.**
- **Enhanced pre-existing notebooks**, making them ready for **seamless integration with applications for business users.**
- **Deployed machine learning models** as **Flask APIs**, enabling real-time predictions and accessibility.

Education

JNTUK afflitiated VSM College Of Engineering And Technology, B.Tech in Mechanical Engineering — July 2016 – September 2020

- CGPA: 7.82

Achievements

- Promoted to **Innovator** role (Ninja → Digital → Innovator) at TCS by clearing **Elevate Wings** exams with distinction in all tracks: **Articulation**, **Business Skills**, **Process**, **Core Full Stack (Data Science)**, and **Core Programming (Python)**, and being selected for **HiTalent Associate** program.
- **Achieved A Band rating** and recognized as **Star Performer** in the team during last financial year
- **Honored with the Best Mentor Award, Badge, and Special Lanyard** for **mentoring associates** to complete the **AI Practitioner certification.**
- **Awarded by GEMS** for multiple achievements, including the **Xcelerate Warrior Certificate** and **MFDM AI Practitioner Certification.**

Certifications

- Earned **Microsoft Certified: Agentic AI Project Ready** professional badge
- Completed Udemy certifications in **Deep Learning**, **Machine Learning**, **Time Series Analysis**, and **Probability and Statistics** all are availble in my linkedin page.
- Achieved **MFDM AI Practitioner Intermediate** certification from TCS

Declaration

I hereby declare that the information provided above is true and accurate to the best of my knowledge.