Sai Teja Settipalli

Washington DC | (571)786-0557 | ssaiteja168@gmail.com | [LinkedIn](#) |[GitHub](#)

EDUCATION

ARIZONA STATE UNIVERSITY   Tempe, AZ
Master of Science – Data Science Analytics and Engineering, GPA: 3.90/4.00   Aug2023 – May 2025
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, GUWAHATI   Guwahati, India
Bachelor of Technology – Computer Science, GPA: 3.46/4.00   Aug 2019 – May 2023

PROFESSIONAL EXPERIENCE

Exeevo Inc   Bangalore, India
Data Engineering Intern   Jan 2023 – Jul 2023
- Developed & refined data pipelines that streamlined data extraction, transformation, and loading for client corporate events and sales data migration to cloud environments; optimized overall processing time by 30% while ensuring data integrity.
- Engaged with diverse departments to ensure quality data delivery while collaborating on requirement-gathering sessions that led directly to the design of 3 bespoke database solutions optimizing workflows across projects.
- Created and deployed automated solutions for cross-functional data mapping workflows, resulting in an impressive 70% decrease in processing times, boosting team efficiency by allowing resources to focus on more strategic initiatives.
- Conducted targeted investigations into dataset characteristics using data analysis techniques resulting in refined interpretations shared with stakeholders focused on fixing 3 major pain points identified during analysis sessions.

GRADUATE PROJECTS

AI Generated Text Detection using Deep Learning

- Developed an end-to-end data analytics pipeline using the Brazilian Olist E-commerce dataset to tackle challenges in customer retention, delivery accuracy, and demand forecasting. Conducted EDA and hypothesis testing to uncover key pain points, and built time-series models (Prophet, SARIMAX, XGBoost), optimizing performance with an ensemble model achieving 5.74% MAPE.
- Implemented a probabilistic CLTV model using Probabilistic models and Regression, followed by K-Means clustering into five tiers, enabling targeted retention strategies and optimized inventory planning for improved efficiency and satisfaction.

AI Generated Text Detection using Deep Learning

- Engineered and implemented a machine learning pipeline leveraging advanced Natural Language Processing techniques and advanced models (LSTMs, BERT, DeBERTa) to classify human-written versus AI-generated essays with 94% accuracy.
- Leveraged NLP tools to extract linguistic, stylistic, and semantic features, employing advanced vectorization methods to improve classification accuracy. Mitigated data imbalance challenges by incorporating external datasets, enhancing model robustness across varied text samples there by creating a scalable model.

Disaster Tweet Detection using NLP
- Executed comprehensive Natural Language Processing techniques to clean and preprocess a dataset exceeding 10,000 tweets; constructed custom embeddings that streamlined text classification accuracy by over 15% during training.
- Executed models based on GloVe with LSTM and BERT variations to understand the importance of context in classification and text processing finally obtaining an accuracy of 84%.

Multi Class Plant disease prediction
- Created an advanced deep learning framework for diagnosing simultaneous plant diseases using leaf imagery by building and testing diverse convolutional neural networks and enhanced detection accuracy through rigorous training.
- Assembled a comprehensive multi-label dataset comprised of over 3,000 images depicting infected leaves across six distinct categories to advance deep learning model and obtained an accuracy of 93%.

LEADERSHIP

Yuvaan - College Festival   Guwahati, India
Event Organization Team Lead   Jan 2020 – May 2022
- Formulated strategic communication plans that facilitated successful negotiations with potential sponsors; secured four long-term sponsorships for future events while fostering strong relationships in the community.
- Spearheaded planning and execution of diverse events by uniting efforts among 25 members to streamline processes; improved overall efficiency which resulted in a notable increase in attendee engagement levels.

SKILLS

Programming Languages: Python, Java, C, C++, JavaScript, SQL, R
AI & Machine Learning: Supervised & Unsupervised Learning, Reinforcement Learning, Time Series Analysis (ARIMA, SARIMA, Prophet), Forecasting, XGBoost, NLP, NLTK, RAG models, Transformers, Prompt Engineering, LLMs, Scikit-learn, TensorFlow, Keras, PyTorch, TensorFlow, Feature Engineering, Hyperparameter Tuning
Data Analysis & Visualization: Exploratory Data Analysis, Hypothesis Testing, Financial Analysis (Ratio Analysis, Forecasting, Scenario Modeling), Data Wrangling (Pandas, NumPy), Matplotlib, Seaborn, Power BI, Tableau, MS Excel
Big Data & Data Engineering: ETL Pipelines, Apache Airflow, SSIS, Hadoop, PySpark, Databricks, Data Warehousing, Snowflake, SQL Optimization, Data Quality Assurance, Kafka, Hadoop Hive, DBT
Cloud & DevOps: AWS (S3, EC2), Microsoft Azure, Docker, Git, CI/CD Pipelines, Linux, Shell Scripting
Software Development: Agile/Scrum, Spring, Flask, Django, GraphQL, SDLC, REST APIs, OOP