Sri Naga Kollepara

911 Bernard St, Denton, TX +1 (940) 222 7049 snmk.kollepara@gmail.com Github Linkedin

Summary

Computer Science graduate student with a background in Data, Machine Learning and a strong foundation in Python and SQL. Proficient in building scalable data processing pipelines, developing predictive models, and deploying machine learning workflows . Actively exploring opportunities in data science, machine learning, or software engineering to contribute to impactful, data-driven projects.

Education

University of North Texas **Aug 2023 – May 2025**

Master of Science, Computer Science *Denton, TX*

Work Experience

Accenture **Sept 2021 – Aug 2023**

Data Engineer - Big Data & ML *Bangalore, India*

Python, PySpark, Hadoop, TensorFlow, Scikit-learn, SQL, AWS, Tableau

- Built scalable data pipelines using PySpark and Hadoop to process large volumes of data across CSM Ingredients' financial and supply chain datasets, reducing ETL runtime by 35%.
- Developed and deployed ML models using TensorFlow and Scikit-learn on AWS SageMaker to predict client financial trends, improving forecast accuracy by 18%.
- Automated reporting workflows using SQL and Tableau dashboards, enabling business teams to access real-time KPIs and drive data-backed decisions.
- Collaborated with cross-functional teams in Agile sprints to integrate ML services into production via REST APIs, improving analytics delivery by 20%.
- Leveraged exploratory data analysis (EDA) and feature engineering on complex financial datasets to establish standardized data schemas, facilitating smoother integration with key partner platform and supporting the 18% improvement in financial trend forecast accuracy.

Projects

Predictive Modelling for Earthquake Magnitude and Location **Jan 2024 – Apr 2024**

Python, PyTorch, Deep Learning, CUDA, AWS S3, SageMaker, Hadoop, Pyspark, MapReduce

- Leveraged Apache Hadoop for distributed data storage and utilized Hadoop MapReduce and Apache Spark for comprehensive data processing of large seismic datasets, including cleaning, validation, and feature engineering.
- Implemented, trained, and evaluated predictive models, including Random Forest with AdaBoost Regressor for magnitude prediction and a Multi-Layer Perceptron Regressor for location prediction, demonstrating capabilities in algorithmic research and model optimization.

Accommodation Connect (GitHub) **Sept 2023 – Dec 2023**

Python, Flask, HTML/CSS, MongoDB, Scikit-learn, Hugging Face Transformers, Fast API

- Developed a full-stack web platform to assist students in discovering and posting accommodation listings, incorporating user authentication and roommate matching features.
- Implemented a machine learning model using NLP techniques (Hugging Face Transformers) to analyze user preferences and bio descriptions, recommending compatible roommates to optimize user satisfaction.
- Integrated a price prediction module using regression models to estimate fair rental prices based on location, amenities, and room features, showcasing skills in building practical analytical tools.

Technical Skills

Languages: Python, Java, SQL
Data & Machine Learning: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Hugging Face, Supervised Learning, Statistical Modeling, Algorithm Development, Experiment Design
Big Data & Data Engineering: PySpark, Hadoop, Spark, MapReduce, ETL Pipelines, Clustered Data Processing
Visualization: Tableau, Power BI, Business Intelligence Tools
Cloud/DevOps: AWS (S3, EC2, SageMaker, Athena, Lambda), GCP (BigQuery, Dataproc, Dataflow), Git
Databases: MySQL, PostgreSQL, MongoDB
Concepts: Data Structures, System Design, Agile Methodology, Object-Oriented Programming, Data Extraction, Quantitative Analysis