Rashmika Reddy Vookanti

+1 (682) 966-2434 |rashmikareddy777@gmail.com| LinkedIn |Github

SKILLS

- Programming : C++, C#, Java, Python, R
- Application Development : Spring MVC, .Net MVC, JavaScript(ES6), TypeScript, React, HTML5/CSS
- Databases : SQL DB, SQL Data Warehouse, MongoDB, Spark, Databricks, Hadoop, Extract Transform Load (ETL, ELT)
- AWS: EC2, Cloud Formation, EMR, SageMaker, S3, EMR Studio
- Azure - Azure Machine Learning, Synapse, Data Factory, Data Lake
- Data Science : Regression, Classification, Applied Statistics, Experimental design, Hypothesis testing, Data visualization
- Machine Learning : Time Series Analysis, Recommendation Systems, Natural Language processing(LSTM, Transformers, LLM)

PROFESSIONAL EXPERIENCE

AMAZON **SEATTLE, US**

Data Science Intern **May 2023 - Aug 2023**

End-to-End Machine Learning-Driven Insights System for Amazon Influencers

- Engineered a Spark-driven data processing system using EMR studio and S3 that can handle **10TB** of data.
- This system utilizes statistical and machine learning techniques to generate actionable insights for Amazon influencers, to understand customer demographics, revenue trends, product analytics.
- Developed a recommendation engine using item-item collaborative filtering techniques, to suggest products to Amazon influencers for optimal sales. Fine-tuning the **T5** model using SageMaker on a data-to-text translation dataset led to significant improvements in the quality of the text insights, making them more accurate and fluent.
- Architected a scalable backend with REST APIs (Spring MVC) within Amazon's mobile app, boosting user engagement **15%** with React/JavaScript for ML-driven insights on Amazon Influencers page.

MICROSOFT **HYDERABAD, INDIA**

Software Engineer(ML/Data) **July 2019 - July 2022**

- Processed, filtered **12TB** telemetry data of Azure Supply Chain using spark, developed a Machine Learning-based imputation for addressing the telemetry file unavailability issues using time series forecasting, which improved the accuracy by **~20%.**
- Developed an automated Data Quality monitoring solution using data pipelines and SQL, saving **~40 hours** of manual work per month and helping improve forecasting metrics of Azure VMs. Enhanced it by adding a power bi reporting layer aiding leaders in identifying telemetry issues.
- Developed an advanced analytics solution using Snowflake and python for a US-based healthcare provider, forecasting the number of COVID-19 cases with **12%** MAPE using ARIMA time series forecasting and exponential smoothing.
- Developed a web application using .NET MVC and React, featuring a hybrid recommendation system to manage staffing, optimize project assignments, and suggest projects based on employee skills and interests reducing manual effort by **40%.**
- Built a custom Data Quality Framework using a SQL rule engine for a machine manufacturer, resulting in an **18%** increase in reliability of machine failure predictions.
- Designed and implemented azure ELT-based data integration platform for a pharmaceutical manufacturer's supply chain. Seamlessly ingested data from **4** different sources into a SQL data warehouse, leveraging Spark for transformation, delivering near **real-time** insights for informed decision-making.
- Architected an end-to-end ETL pipeline for a government entity to implement visualization of smart city components like healthcare, environment and transportation analytics to view KPIs.

PROJECTS

- Implemented Named Entity Recognition of eBay's listings as part of its Machine Learning competition. Computed and contrasted models built using Bi-directional LSTM, BERT to tag Brand, product type, model etc. Evaluated performance of both the models and BERT achieved an F1 score of **86%**.
- Conducted a comparative analysis of network pruning and LLM-int8 quantization techniques on a pre-trained RoBERTa Language Model for sentiment analysis classification task, we achieved a significant reduction in inference time by **1.5X**.
- Enhanced Quora question similarity using TF-IDF, word2vec, Logistic Regression, and Support Vector Machine, achieving 50% lower log-loss compared to the baseline model Random Forest.

EDUCATION

University of Washington, Seattle **Sept 2022 - Mar 2024**

Master of Science in Data Science, GPA: 3.9/4

Courses – Machine Learning, Deep Learning, Statistics, Data Visualization, Applied Statistics and Experimental Design, Explainable AI

VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad **June 2015 – May 2019**

B. Tech in Computer Science and Engineering, GPA: 3.87/4

Courses - Big Data Analytics, Predictive Analytics, Artificial Intelligence & Neural Networks.