Eashan Joshi

585-202-4547 | eashan.joshi.rit@gmail.com | LinkedIn | GitHub | New York City, NY (Willing to relocate)

Data Scientist with 3 years of experience in Python, machine learning, software development and quantitative analysis

Education

Rochester Institute of Technology
Master of Science in Data Science — Dec. 2024

- Relevant Courses: Machine Learning, Software Construction, Statistics, Business Intelligence, Data Warehousing, Data Visualization
- Honors: RIT Academic Merit Scholarship - For top 10% students in the class

University of Pune
Bachelor of Engineering in Computer Science — May 2020

Experience

Generative AI Research Assistant — Aug. 2023 – Present
Rochester Institute of Technology

- Devised an analysis engine for the open source DevGPT dataset, mining 5M+ ChatGPT mentions with GPU tools (HPC) to identify trends
- Conducted quantitative analysis to uncover significant patterns, enhancing predictive models and strategic insights
- Led the development of a dynamic prompt engineering prototype using Python and JavaScript to enhance debugging interactions, improving AI response relevance and data quality by 30%

Software Engineer — Sep. 2020 – Sep. 2021
Yardi

- Improved fiduciary management efficiency by developing 80+ components in the Trust Accounting Application, leading to a $200,000 reduction in operational costs
- Optimized SQL queries, improving client data retrieval by 10% and utilized Power BI on client data, enhancing data analytics and visualization
- Designed data pipelines & scalable business solutions, reducing financial data error rate by 15% in insurance management
- Delivered on-call production support, resolving over 90 support tickets to ensure system performance & reliability

Projects

Duplicated Bug Report Detection | *Python, TensorFlow, NLP*

- Harnessed TF-IDF & Cosine Similarity to process Eclipse Bug Report Dataset, achieving 93% accuracy
- Employed BM25 on similar and dissimilar datasets, achieving recall rates up to 2.24% and 6.25% respectively and Implemented Shapiro-Wilk test on BM25 recall rates across datasets, revealing p-value of 0.17 indicating normality
- Evaluated a Siamese CNN model on textually similar & dissimilar datasets, obtaining AUC scores of 0.54 & 0.58

Risk Management of Listed Companies (Industry Project) | *Python, ML*

- Modeled financial data from FinSharpe, optimizing strategies for 362 high-risk stocks and boosting stakeholder satisfaction
- Applied Value at Risk (VaR) strategy to trading systems to capitalize on black swan events, securing an $850,000 investment

RevenueVision Spark Suite | *Apache Spark, AWS, Prophet*

- Developed a predictive model with Spark and Prophet, achieving 95% revenue forecast accuracy over 12 months
- Implemented debt-to-equity prediction, enhancing control and cutting processing time by 30%

Technical Skills and Certifications

Programming Languages:Python, R, SQL, Java, Scala, C/C++, .NET
Frameworks and Libraries: Scikit-Learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, NLTK, Spacy, Transformers (Hugging Face), Gensim
Data Tools and Technologies: Pandas, Numpy, Dplyr, Tidyverse, Matplotlib, Seaborn, ggplot2, Tableau, Power BI, Plotly, Dash, Apache Spark, Hadoop, ETL Tools, Airflow, PostgreSQL, MySQL, MongoDB, Cassandra
Certifications: AWS Certified Cloud Practitioner (Amazon Web Services), Google Project Management (Google)