Rajesh Marudhachalam

Toronto, ON
rajesh.marudhachalam@gmail.com | +1-[redacted]
https://linkedin.com/in/rajesh1804
https://github.com/rajesh1804
https://rajesh1804.github.io

Professional Summary

Data Scientist with 4+ years of experience delivering data-driven solutions that drive measurable business impact at scale. Specialized in applied machine learning, time series forecasting, anomaly detection, and NLP using LLMs and RAG pipelines. Proven ability to lead end-to-end model development—from experimentation to deployment—using Python, PySpark, SQL, and AWS.

Professional Experience

Bluecat Networks, Toronto, Canada

Data Science Developer2 July 2024 – Present

- Developed and deployed a Retrieval-Augmented Generation (RAG)-based LLM chatbot that accelerated network diagnostics, reducing incident response times by **40%**.
- Designed and operationalized time series forecasting models (Exponential Smoothing, ARIMA) to improve infrastructure capacity planning by **30%**, boosting network uptime and reliability.

Software Developer January 2024 – July 2024

- Built a real-time anomaly detection system for network metrics (e.g., packet drops, TCP anomalies) with **92% precision**, cutting false positives by **40%**.
- Partnered with infrastructure teams to productionize ML pipelines and scale monitoring across **500+ enterprise client devices**.

Machine Learning Researcher Intern May 2023 – December 2023

- Prototyped a weakly-supervised learning model using pairwise relation prediction to detect DNS tunnels, **achieving 67% detection accuracy** on unseen threats—**23% above baseline**.

JP Morgan Chase & Co, Bengaluru, India

***Software Engineer 2**, Wealth Management* January 2022 – August 2022

- **Accolades:** Recognised under '***Execution Excellence***' category for Q1 2022 for outstanding delivery in cross-functional initiatives.
- Led the design of a centralized HDFS-based data lake, enabling 10+ teams to access **15+ TB** of analytical data for decision-making.
- Re-engineered SQL workloads into Hive/Impala, reducing query latency by **60%** and streamlining reporting pipelines.
- Spearheaded the cloud migration of custom ETL frameworks to **AWS EKS** and **Snowflake**, improving throughput and reducing infra costs by **20%**.

***Software Engineer**, Wealth Management* August 2020 – January 2022

- **Accolades:** Recognised under '***Execution Excellence***' category for Q4 2020 for outstanding delivery in cross-functional initiatives.
- Built a modular PySpark ETL framework that automated ingestion of **10M+ records daily** and handled **5+ TB** of structured/unstructured data. Introduced schema validation and alerting, reducing pipeline failure rates by **80%**.
- Migrated legacy ETL pipelines (Informatica/Pentaho) to the new PySpark framework, achieving **11x speedup** (from 45 to 4 minutes).

***Software Engineer Intern**, Asset Management* January 2020 – July 2020

- **Accolades:** Recognised as one of the '***Top 6 performers***' among ~300 Asia-Pacific interns.
- Automated Excel-based reporting by building a React.js dashboard, saving analysts **~2 hours/day** and improving real-time business insights; adopted as the team's primary analytics interface.

Heptagon Technologies Pvt Ltd, Bengaluru, India

Data Science Summer Intern April 2019 – May 2019

- Built a sentiment classifier for political tweets using an SVM + TF-IDF pipeline; achieved **85% classification accuracy**. Applied NLP techniques (lemmatization, co-occurrence embedding) to improve model generalization across varying tweet structures.

Education

MSc in Applied Computing (MScAC) September 2022 – December 2023

University of Toronto, Canada

Coursework: Introduction to Machine Learning, Cloud based Data Analytics, Neural Networks and Deep Learning, Advanced Data Systems

B.Tech in Computer Science and Engineering August 2016 – May 2020

Vellore Institute of Technology, India

Coursework: Statistics, Calculus, Linear Algebra, Discrete Mathematics & Graph theory, Data Mining, Natural Language Processing

Accolades: *Merit Scholarship for Academic Excellence*

Projects

Quantifying Uncertainty in Ensemble Learning October 2022 – December 2022

- Analyzed predictive uncertainty in ensemble models (XGBoost, Random Forest, AdaBoost) using entropy and variance metrics, informing robustness tradeoffs for model deployment under real-world noise and variance.

Active Learning Strategies for NLP Classification October 2022 – December 2022

- Benchmarked transformer-based active learning techniques (e.g., entropy sampling, least confidence) on the 20 Newsgroups dataset, achieving a **20–25% reduction in labeling cost** versus baseline, with direct implications for scalable annotation workflows.

Publications

- Selvakumar K, **Rajesh M**, Eshwar S, Shraveen B.S, 'YouTube Video Ranking: An NLP based system', *IJRTE, Vol-8 Issue-4*. **(SCOPUS)**

Technical Skills

Programming Languages: Python, SQL, PySpark, Shell, JavaScript
Machine Learning & NLP: Scikit-learn, TensorFlow, Keras, NLTK, XGBoost, Random Forest, CNNs, RNNs, LSTMs
Data Visualization: Matplotlib, Seaborn, Plotly
Data Engineering & Big Data Tools: Airflow, Apache NiFi, Apache Spark, Hadoop, Hive, Impala, Kafka, Sqoop
Databases: Snowflake, Trino, Redshift, MySQL, PostgreSQL, Oracle, SQL Server
Cloud Platforms: AWS (EKS, S3, Lambda, EC2), Docker, Git

Certifications

AWS Certified Developer Associate [view credential], ***Tensorflow Developer*** [view credential], ***Deep Learning Specialization*** [view credential]