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 building scalable ML models, deploying production-ready pipelines, and solving high-impact business problems. Proven track record of reducing operational costs and improving decision-making through advanced analytics, NLP, and time series forecasting. Skilled in Python, PySpark, TensorFlow, and cloud platforms like AWS.

Professional Experience

Bluecat Networks, Toronto, Canada

Data Science Developer2 July 2024 – Present

- Engineered and deployed a Retrieval-Augmented Generation (RAG)-based chatbot that accelerated network diagnostics and cut incident response times by 40%.
- Created and operationalized time series forecasting models (Exponential Smoothing, ARIMA) to improve infrastructure resource planning by 30%, impacting network reliability and uptime.

Software Developer January 2024 – July 2024

- Built a high-precision anomaly detection framework for real-time network monitoring, achieving 92% precision and reducing false positive alerts by 40% and enhancing system resilience.
- Collaborated cross-functionally with infrastructure teams to scale ML-driven monitoring tools across 500+ client devices.

Machine Learning Researcher Intern May 2023 – December 2023

- Prototyped a novel weakly-supervised learning model using pairwise relation prediction to detect DNS tunnels, achieving 67% accuracy on zero-day threats—outperforming baselines by 23%.

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.
- Collaborated with stakeholders and designed a scalable HDFS-based data lake serving 10+ teams, unlocking access to 15TB+ of data and enabling data-driven decision-making.
- Re-engineered SQL-based reporting queries into Hive/Impala, slashing run times by 60% and improving analyst productivity.
- Led cloud migration of ETL pipelines to AWS EKS and Snowflake, boosting pipeline throughput and reducing infrastructure cost 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.
- Created a robust PySpark ETL framework that automated daily ingestion of 10M+ records and 5TB+ of data, incorporating real-time schema checks and alerting for data quality reducing pipeline failures by 80%.
- Migrated legacy pipelines (Informatica/Pentaho) to PySpark framework, Reduced pipeline runtime from 45 to 4 minutes (11x improvement) through Spark-based optimization and modular pipeline design.

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

- **Accolades:** Recognised as one of the '***Top 6 performers***' among ~300 Asia-Pacific interns.
- Replaced manual Excel-based reporting with an interactive React.js dashboard, saving analysts 2 hours/day and improving real-time business visibility; 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 Twitter data using TF-IDF + SVM pipeline; achieved 85% accuracy in classifying political tweets across 3 classes. 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

Investigating Uncertainty in Ensemble Methods October 2022 – December 2022

- Quantified uncertainty in ensemble models (Bagging vs Boosting) using entropy and variance metrics; analyzed performance on UCI datasets using XGBoost, Random Forest, and AdaBoost.

Investigating Query Strategies in Active Learning for NLP Tasks October 2022 – December 2022

- Benchmarked active learning strategies (e.g., Least Confidence, Entropy Sampling) for text classification on 20 Newsgroups dataset using transformer-based embeddings; achieved 20–25% label efficiency gains vs random sampling.

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]