Rajesh Marudhachalam

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

Dynamic self-starter with a track record of collaborating with cross-functional teams. Possesses a strong foundation in machine learning, data analysis, and problem-solving. Eager to leverage data-driven insights to drive business growth as part of your world-class team.

Education

MSc in Applied Computing (MScAC) September 2022 – December 2023
University of Toronto, Canada **CGPA**: 3.9/4.0
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 **CGPA**: 9.27/10.0
Coursework: Statistics, Calculus, Linear Algebra, Discrete Mathematics & Graph theory, Data Mining, Natural Language Processing
Accolades: *Merit Scholarship for Academic Excellence*

Work Experience

Bluecat Networks, Toronto, Canada
Data Science Developer January 2024 – Present

- Utilized a range of time series forecasting methodologies to accurately predict CPU and memory usage patterns in network devices, effectively mitigating network downtime for customer devices.

Machine Learning Researcher Intern May 2023 – December 2023

- Researched the viability of using deep learning to address the problem of DNS Tunnels, and prototyped a weakly-supervised learning method using pairwise relation prediction technique to detect both previously seen and unseen DNS tunnels with a 67% hit rate.

JP Morgan Chase & Co, Bengaluru, India
Software Engineer 2*, Wealth Management* January 2022 – August 2022

- **Accolades:** Recognised under '***Execution Excellence***' category for Q1 2022.
- Analysed the needs of business and collaborated with product managers, stakeholders, data scientists & analysts to design and build a data-lake that acts as one-stop shop serving data to solve the analytical business needs. [HDFS on YARN clusters, Hive & Impala].
- Led the effort to redesign on-demand data insights RDBMS SQL queries to HIVE/Impala queries, reducing run time by 60%.
- Performed a POC to port custom ETL framework to run on Amazon EKS with S3 & Snowflake, to migrate from on-prem servers.

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

- **Accolades:** Recognised under '***Execution Excellence***' category for Q4 2020.
- Conceptualized and implemented custom reusable ETL framework on PySpark with automated deployments for structured and unstructured data that processed more than 10M records per day, resulting in storage of 5+ terabytes of data. Built a recon module into the framework to check for inconsistencies (schema changes) & data gaps, and alert target users via email.
- Migrated existing Informatica/Pentaho pipelines to use the aforementioned PySpark ETL framework, this cut the run time 5x.

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

- **Accolades:** Recognised as one of the '***Top 6 performers***' among ~300 Asia-Pacific interns.
- Developed a multipage React.js dashboard to replace the existing MS-Excel reports and act as one-stop shop for visualization need of business, and eliminated the need to spend ~2 hours daily on report generation.

Heptagon Technologies Pvt Ltd, Bengaluru, India
Data Science Summer Intern April 2019 – May 2019

- Developed a ML model to determine the sentiment polarity of the data fetched from Twitter, this involved tokenization, lemmatization, vectorization to create higher dimensional word embeddings and capture fine grained relationships between tokens.

Research Projects

Investigating Uncertainty in Ensemble Methods October 2022 – December 2022

- This project investigates and tries to quantify the uncertainty of predictions of using Boosting and Bagging techniques in an ML.

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

- A comparative study on query strategies for AL for labelling unsupervised data for NLP tasks using state-of-the-art methods.

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

ML Libraries/Frameworks: Scikit-Learn, Keras, Tensorflow, Matplotlib, Seaborn, Numpy, Pandas, Scipy, Nltk,
Big-Data Technologies/Frameworks: Hadoop (Sentry & Ranger), Spark, Hive, Impala, Kafka, Sqoop, JSON, Parquet
AWS services: Lambda, EKS, Redshift, Sagemaker, EC2
Databases: MySQL, PostgreSQL, SQL Server, Oracle (LDAP & Kerberos), MariaDB, Snowflake, Trino

Certifications

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