Rajesh Marudhachalam

Toronto, ON

✉ rajesh1804@cs.toronto.edu | 📞 +1-437-973-6572

🔗 https://linkedin.com/in/rajesh1804
🔗 https://github.com/rajesh1804
🌐 https://rajesh1804.github.io

Experienced graduate student with passion for data science and significant analytics experience to bring to the table. With a team-oriented attitude, eager to apply my skill-set in data science in a research-centric role and contribute to enhance your firm's business efficiency, strategic goals, and profit.

🎓 Education

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

University of Toronto, Canada

Coursework (ongoing): Introduction to Machine Learning, Advanced Data Systems, Cloud based Data Analytics, Neural Networks and Deep Learning, Natural Language Computation

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

Vellore Institute of Technology, India   **CGPA**: 9.27/10.0

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

Accolades: Merit Scholarship for Academic Excellence

💼 Work Experience

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 a data-lake that acts as one-stop shop serving data to solve all analytical business problems. [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 at least 3 folds.
- Performed a POC to port previously build custom ETL framework to run on Amazon EKS using S3 and Snowflake.

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 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 former 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 the 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 email daily reports as Excel files.

Heptagon Technologies Pvt Ltd, Bengaluru, India

Data Science Summer Intern   April, 2019 – May, 2019

- Developed a model to determine the sentiment polarity of the data fetched from Twitter API, 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 – Present

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

Active Learning + Weak Supervision for NLP Tasks   October, 2022 – Present

- A comparative study between AL and Weak supervision 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)** 🌐
- **Rajesh M**, 'Prediction of Survivors in the Titanic Cruise', IJRTE, Vol-8 Issue-3. **(SCOPUS)** 🌐

⚒ Skills

ML Libraries/Frameworks: Numpy, Pandas, Scipy, Scikit-Learn, Matplotlib, Seaborn, Keras, Tensorflow, Nltk

Big-Data Technologies/Frameworks: Hadoop (Sentry & Ranger), Spark, Hive, Impala, Kafka, Sqoop, JSON, Avro, Parquet, ORC

AWS services: Lambda, EKS, Redshift, Sagemaker, EC2, Cloudformation

Databases: MySQL, PostgreSQL, SQL Server, Oracle (LDAP & Kerberos), MariaDB, Snowflake

Others: Git, Unix/Linux, Agile, QlikSense, MS-Excel, React.js, HTML5, CSS3

🎓 Certifications

AWS Certified Developer Associate, AWS [view credential]   January 2022

Tensorflow Developer, Deeplearning.ai [view credential]   September 2021

Deep Learning Specialization, Deeplearning.ai [view credential]   June 2021