

ARPIT RATHORE

www.github.com/rarpit1994 | www.linkedin.com/in/arpit-rathore | Ottawa, Ontario | +1-343-987-3262 |
rarpit1994@gmail.com

PROFESSIONAL EXPERIENCE

Senior Data Scientist

Aug 2021- Present

Environment and Climate Change Canada (ECCC)

Ottawa, Canada

- Utilized **Python** to develop non-compliance indicators using **Ensemble Learning algorithms** for Regression Analysis on the multi-year transactional database to increase targeted enforcement.
- Wrangled **PostgreSQL**-based database to utilize **Dash** framework to develop the **Text Analytics** driven dashboard to upscale the policymakers' decision-making and provide KPIs towards mandate.
- Utilized **Tensorflow** framework, **SQL**, and parallel computing to perform fraud detection to develop the Text Classifier on 100 M+ records.
- Utilizing **OpenCV**, and **Tesseract** on 10K+ images to perform text detection and extraction to implement digital solutions for clients.
- Developing a graph database **Apache AGE** on existing **PostgreSQL** database to perform network analysis etc. using Cypher Query Format.
- Responsible for development, administration, and managing of the Linux Workstation and **Microsoft Azure** workspace for the team.
- Presented results to team leads and wrote requested summary detailing value proposition and strategy to present to end clients and senior leadership.

Data Scientist

Jan 2021- Aug 2021

Department of National Defense (DND)

Ottawa, Canada

- Utilized **ARIMA**, **Prophet** and **LSTM** to develop **Time Series** forecasting to support decision-making on billion-dollar policies.
- Developed the **Microsoft PowerBI** dashboard to support weekly meetings of senior leadership to map the branch's vision and mandate.
- Reviewed the methodologies of **synthetic data generation** with the **GAN** framework to produce proof of concepts on various policies which were lacking data requirements for the analytics.
- Responsible for developing, planning, and managing the Master Project Tracker for the branch and stakeholders.
- Interface with stakeholders and clients within the organization.

Software Developer Assistant

Jan 2020- Mar 2020

Canadian Centre for Studies and Research in Bilingualism and Language Planning, University of Ottawa, Ottawa, Canada

- Re-developed a bi-lingual mobile application using **MEAN** stack, **ionic** Framework that increased application download to **18%**.
- Implemented the infrastructure for the application on **AWS** which increased the **25%** traffic on the application.
- Utilized the **Dash** Framework to develop the dashboard on user engagement data collected using application to perform targeted features development for upcoming versions.
- Gathered requirements from Professor and prepared wireframes to showcase the application flow.

RELEVANT SKILLS

Programming Languages: Python, SQL, Java

Big Data & Machine Learning Tools: Spark, Python (e.g., Scikit-Learn, NumPy, Pandas, Dask, PySpark, Matplotlib, Plotly)

Databases: PostgreSQL, Graph Database (Apache AGE), Oracle

Visualization Tools: Tableau, Microsoft PowerBI

Deep Learning Frameworks: Tensorflow, Keras, PyTorch

Data Science & Miscellaneous Technologies: A/B testing, ETL, Data Science Pipeline (cleaning, wrangling, visualization, modeling, interpretation), Statistics, Time Series, Experimental Design, Hypothesis Testing, OOP, APIs, Excel, Git, Docker

Cloud Computing Tools: Microsoft Azure, Databricks, Google Colab

EDUCATION

University of Ottawa, Ottawa, Canada

CGPA: 3.8/4.0

Sep 2019 – April 2021

Masters in Applied Data Science (MDTI)

Vellore Institute of Technology, Vellore, India

CGPA: 9.26/10.0

June 2017 – May 2019

Master of Computer Applications (MCA)

INTERNSHIP EXPERIENCE

Junior Analyst

May 2020 – Dec 2020

Employment and Social Development Canada (ESDC)

Ottawa, Canada

- Reviewed the project development and iteration tracking using **Microsoft Azure DevOps** and **Power BI**.
- Utilized the **Microsoft Azure DevOps** queries for data extraction for analytical dashboards using **Microsoft Power BI**.
- Responsible for planning and managing the sprint cycles (Agile) using Microsoft Azure DevOps
- Analyzes, highlights and presented the common elements between processes and rules for senior management.

Research & Development Intern

Dec 2018- Aug 2019

Axis India Machine Learning (A Machine Learning Research Lab)

Jaipur, India

- Designed data pipelines for preprocessing, scaling, and transforming raw data using **Scikit-Learn** and **Tensorflow**.
- Developed data generator for highly imbalanced data using the **Google cloud data prep**.
- Implemented **PyTorch** framework of **CNNs** with data augmentation to improve model robustness and accuracy on testing data.
- Utilized TensorFlow's **Keras** API to develop **RNNs** for stock prediction and sentimental analysis on quantitative & qualitative datasets.

MASTER RESEARCH PROJECTS

Intelligent Prediction of the Stock Market using BERT Word Embeddings, LSTM, Dense Networks and, XGBoost Techniques on News Headlines

- News Headline and market Data collection using Intrinio API and Text Preprocessing using NLTK and Regex.
- Labelling of Data using MACD Market Indicator and Word Embeddings using BERT model.
- Data Reduction using PCA and Data normalization using Min-Max Scaler Techniques.
- The Text Classification using LSTM, Dense Networks and XGBoost. LSTM performed best among these techniques.

Kernel Support Vector Machines for Multiclass Image Classification

- Classification of the MNIST Images using one vs all and all pairs' techniques in the SVM Classifier.
- The Support Vector Machine classifier attained 95.33 % accuracy, 95.7% recall and 95.67 % Precision.

PUBLICATIONS

1. Machine Learning Based Classification of Cervical Cancer Using K-Nearest Neighbor, Random Forest, and MLP Algorithms

- Journal of Computational and Theoretical Neuroscience, Volume 16, Numbers 5-6, May 2019, pp. 2523-2527(5)
- DOI: <https://doi.org/10.1166/jctn.2019.7925>

2. Stock Prediction using Machine-Learning Algorithms

- International Journal of Engineering and Advanced Technology (IJEAT), 2249-8958, Volume: 8, Issue-2S
- <https://www.ijeat.org/wp-content/uploads/papers/v8i2s/B10831282S18.pdf>