

Rupsa Chakraborty

✉ rupsachakraborty2622@gmail.com | ☎ +1 (848)437-1600 | in Rupsa-Chakraborty | 🌐 Rupsa25 |

Education

Rutgers University, New Brunswick, NJ, USA

Sep 2021 - May 2023

– **Master of Science, Computer Science**

– Relevant Coursework: Artificial Intelligence, Linear Algebra and Optimization, Data Structures and Algorithms, Machine Learning, Database Management, Data Interaction and Visual Analytics, Probability and Statistics

Vellore Institute of technology, Chennai, India

Aug 2015 - Aug 2019

– **Bachelor of Technology, Computer Science and Engineering**

– Relevant Coursework: Data structures and Algorithms, Database Management Systems, Image Processing

Technical Skills

– **Programming Languages:** Python, Javascript, Scala, C++, HTML, CSS, Shell Scripting, R

– **Libraries and Frameworks:** Express, NodeJS, NumPy, Pandas, Scikit-learn, TensorFlow, Pytorch, Matplotlib

– **Database and Tools:** SQL Databases, MongoDB, Airflow, Spark, Kafka, Tableau, Git

– **Cloud Services and Version Control:** AWS Redshift, S3, EC2, Azure, Github

Experience

Rutgers University, Research Assistant

Jun 2022 - Aug 2022

– Coded baseline models for Active Learning for object detection pipeline, by training DETR and other models for custom dataset and utilized the MakeSense.ai platform to annotate about 3000 images.

– Trained on PASCAL VOC dataset, which led to MAP of 62.3720. at 90000 iterations on the final baseline model which was comparable to the original paper which had used the COCO dataset. **Stack: Python, Pytorch.**

Phenom People Private Limited, Data Engineer

Feb 2021 - Jul 2021

– Built data pipelines by developing ETL scripts using python and SQL, to source, transform the raw data, automate data validations and report generations, reducing man hours required for data validations by 20%.

– Ensured adherence pep8 standards and proper coding practices, limiting technical debt to 5%.

– Reduced load time by 10 times by migrating to Amazon Redshift from PostgreSQL, leveraging the Massively Parallel Processing Architecture. **Stack: Python, SQL, Airflow, AWS.**

Academic Projects

Formula 1 Project

Jan 2023-Mar2023

– Analyzed Formula 1 racing data, by data ingestion, transformation using **Azure Databricks** and leveraged **Delta Lakehouse** features to perform operations on data using **PySpark** and **SQL**.

– Used **Azure Data Factory** to schedule the pipeline, incorporated an incremental load architecture for data ingestion, created dashboards for data visualization, taking advantage of IAM roles for permissions for integrating **Azure Data Lake Storage with Databricks.**, using the former for data warehousing.

Massive Data Mining

Jan 2023-Mar2023

– Employed **PySpark** to give friendship on recommendations using data consisting 49995 records.

– Performed Market basket analysis to find frequent itemsets on given browsing data consisting 31101 records.

What's cooking Kaggle

Aug 2018 - Apr 2019

– Developed an ML model for prediction of the type of cuisine from the list of ingredients given using **SVC**.

– Concepts of text mining such as use of **lemmatization**, **Tf-idf** on the text data and was able to achieve a final test score accuracy of **81.043%** with the highest ranked on leaderboard score being **83.216%**.

Publications

Image Processing based Edibility Analysis of Spinach Leaves

Oct 2018 - Feb 2019

– “**International Journal of Recent Technology and Engineering (IJRTE)**” in March 2019. Volume 7, Issue 6. ISSN No. 2277-3878.