

JUDHAJIT ROY

jroy13n96@gmail.com | (872)-218-5124 | <http://www.linkedin.com/in/judhajit-roy> | <https://github.com/Judhajit-Roy>

SUMMARY

Software Developer with 3+ years' experience in Big Data development for driving business solutions and in Machine learning and Deep Learning for creating models based on numerical data, images and text primarily using Python, SQL and Spark.

EDUCATION

Master of Science - Computer Science (GPA – 4.00)

Aug 2021 - Present

University of Illinois at Chicago, USA

Relevant Courses: Machine Learning, Visualization, Data Structures & Algorithms, Databases, Big Data, Deep Learning, Cloud Computing, Natural Language Processing, Causal Inference and Deep Generative Models

Bachelor in Electronics & Telecommunications Engineering

Aug 2014 - June 2018

University of Mumbai, INDIA

TECHNICAL SKILLS

Languages and Tools: SQL, Python, Java, Unix, HTML, AWS, EC2, Lambda, S3, Spark, Hive, Kafka, Presto, Hadoop, MongoDB, Docker

Machine Learning: Scikit-learn, NLTK, NumPy, Pandas, Seaborn, NLP, CNN, PyTorch, Tensorflow, LSTM, BERT, VAEs, Tableau, D3

WORK EXPERIENCE

Machine Learning Intern – Encyclopedia Britannica | Chicago, IL, USA

May 2022 – Aug 2022

- Experimented and developed several techniques for named entity recognition and key-phrase extraction from natural language.
- Created a pipeline to automate parsing and cleaning of XML data into natural language text and extract important key-phrase from top sentences using supervised and unsupervised algorithms.

Data Science Intern – Trippers | Monaco (Remote)

March 2021 - Aug 2021

A start-up working on perfecting an app to generate travel itineraries and connect travellers.

- Developed logical methods on python for the core algorithm to generate itineraries with the shortest and most apt route catering to the user preferences.
- Implemented a recommender system using the principles of text similarity that reduced manual efforts in searching cities to visit in Europe by 100%.

Data Engineer – Reliance Jio Platforms Ltd | Mumbai, India

Aug 2018 - May 2021

India's Largest LTE Network Provider with Annual Revenue of USD 1.7 Bn.

- Built ETL pipelines on Apache Spark to migrate Big Data of TB and PB memory size from one storage system to another.
- Developed and Optimized ETL pipelines for data aggregation to cut down the run time from hours to under 15 minutes.
- Scheduled the jobs using Cron expressions for batch processing and monitored the operation of over 100 ETL jobs.
- Performed feature generation using Advanced SQL and PySpark operations to curate data and created visualizations on Tableau.
- Implemented and maintained operation of real-time streaming of data gathered from OTT Services through Kafka brokers.

PROJECTS

Sentiment Analysis on Hotel Reviews:

- Pre-processed and Classified 20k+ TripAdvisor Hotel reviews into positive, negative and neutral sentiments.
- Applied cross-validation technique on the best models to improve the performance of the algorithms and compared the results.

Multi-label classification of Research Articles:

- Classified 20k+ research articles into multiple classes using various Machine Learning and Deep Learning algorithms.
- Compared the results and efficacy of the algorithms, obtained Macro-F1 of 0.72 in the best model (Bi-LSTM).

Language Translation using Seq2Seq Model:

- Built a Seq2Seq Model using the encoder-decoder architecture to translate English language text to Marathi language.
- The model translates English phrases into precise Marathi phrases and the same architecture can be reused for other language translations as well.

Log Statistics generation using Map-Reduce Framework:

- Built end to end pipeline that reads Log Data is processed using the map-reduce tasks to generate Log Statistics.
- Deployed pipeline on AWS where log data is in S3, and the Map-Reduce tasks run on EMR cluster to store results in S3 bucket.