Anshita Saxena

anshita333saxena@gmail.com +1 (514) 569-5243 Anshita Saxena

Education

Sep 2022 – Present

University: University of Montreal, Canada
(MILA- Montreal Institute of Learning Algorithms)
*D*egree: Masters in Computer Science (Machine Learning Specialization), **Grade**: A
*C*ourses: Machine Learning, Data Science, Deep Learning, Natural Language Processing, Geometric Data Analysis
Internships: *Deep Learning Researcher Intern at Hydro-Québec Research Institute*
Evaluated as an ***Excellent Intern*** *(Score: 96/100), May-Dec 2023;*
Machine Learning Intern at Ericsson GAIA, Jan-Apr 2024

Aug 2013 – Jul 2017

University: Dr. A.P.J. Abdul Kalam Technical University, AKTU India
College: Meerut Institute of Technology, MIET Group, Meerut.
Degree: B.Tech in Computer Science and Engineering (**Percentage**: 82.26/100 [Honours])

Selected Honors And Awards

2022

Diversity Award (Tuition Fees Scholarship/Exemption)
Organization: University of Montreal

2019, 2020, 2021, 2022

Awards: IBM Eminence and Excellence Award, IBM Impact on the Business and Significant Achievement, IBM Service Excellence Award, IBM Growth Award, ScyllaDB Innovator Award
Organizations: IBM India Pvt. Ltd, ScyllaDB (NoSQL Product Company)

2014, 2015, 2016

Awards: Academic Excellence Award, Codezilla Coding Award
Organizations: Meerut Institute of Technology, Meerut Institute of Engineering & Technology, MIET Group

Skills: Soft and Technical Skills, Certifications, and Courses - Online Learning

Soft Skills

Leadership, Teamwork, Adaptability, Positivity, Interpersonal Skills, Creative thinking

Technical Skills

Interests- LLMs, Generative Modelling (Generative AI), Natural Language Processing (NLP), MLOps
Databases- ScyllaDB, MySQL, PostgreSQL, Elasticsearch. Programming Languages- Python, PySpark.
Tech- VS Code, Google Colab, Azure, AWS SageMaker, GCP, IBM Cloud, IBM CP4D, Git, Docker, Kubernetes, Flask.
Libraries - PyTorch, pandas, numpy, sklearn, streamlit, HuggingFace, CometML, Optuna, Ray, LangChain

Certifications and Skill Courses

Microsoft: Azure Fundamentals: 2021, IBM Certified Big Data Engineer: 2020, Enterprise Design Thinking Practitioner, IBM Watson Knowledge Catalog Essentials, Cognitive Practitioner (IBM), Python for Data Science (IBM), Deep learning specialization (Coursera)

Work Experience

Jan 2024 – Present | Ericsson Montreal GAIA (Global AI Accelerator) Lab as Machine Learning Intern

May 2023 – Dec 2023 | Hydro-Québec Research Institute (IREQ) Montreal as Deep Learning Research Intern

Apr 2018 – Aug 2022 | IBM Corporation as Data Scientist and Senior Big Data Engineer

Jul 2017 – Apr 2018 | Astrea IT Services as Software Developer

Publications

2019 **Optimal Partition Search**
A. Saxena and A. Saxena, "Optimal Partition Search," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1-8, doi: 10.1109/ICECCT.2019.8869459. Link: https://ieeexplore.ieee.org/abstract/document/8869459

2017 **DeepCoder: An Approach to Write Programs**
A. Saxena, A. Saxena, J. Patel, "DeepCoder: An Approach to Write Programs," 2017 International Conference on Advanced Research and Innovation in Engineering (ICARIE), 2017 International Journal of Engineering and Manufacturing Science (IJEMS), 2017, pp. 9-13, Vol. 7, No. 1, Research India Publications. Link: https://www.ripublication.com/ijems_spl/ijemsv7n1_02.pdf

Projects

Hydroquebec, Montreal - Generation of High-frequency data from Low-frequency data

- Implemented the WaveNet model approach worked successfully to address a comparatively complex problem statement characterized by diverse sources of information operating at various levels. Implemented MAML-based sinusoidal regression approach on measurement data to accomodate minute and second level measurement modality. Implemented MLP as a baseline model to predict the frequencies at second modality based on minute modality.
- Libraries such as ydata-synthetic to apply TimeGAN and run on multi-node GPU cluster. Improvements done on BiLSTM.
- All models were implemented in PyTorch, JAX, and Keras.

Question Answering

- Used Huggingface, Pytorch, Haystack for model building and evaluation
- Compared BM25 and Dense Passage Retriever for Reader on the basis of Recall Evaluation Metrics for top-3 reader and reader.
- Evaluation and Exact Match/F1 score for Retriever Evaluation.
- Fine Tune on SQUAD, SubjQA, SQUAD+SUBJQA datasets to compare the performance.

Hockey NHL Project

- Downloaded the live hosted data using Python scripts and API. Cleaned the data according to the project requirements. Created the visualizations using contours, the intention is to showcase the shot generation intensity.
- Applied logistic regression, XGBoost, Neural Networks, Ensemble methods (Decision Trees, LGBM, Random Forests), Log models in Comet ML. Used Shapley and Lime to analyze the feature importance of the features.
- Developed flask API for prediction and download model from the registry. Created first docker containers for passing the data to generate prediction service (using Flask API) and second docker container for WEB UI through Streamlit. Establish the docker communication network.

Youtube Content GPT:- (Deployed App)
Used open-source google/flan-t5-xxl LLM hosted in Huggingface. Integrated Wikipedia tool with Huggingface LLM using LangChain and user interaction through the Streamlit app.

IBM - Broadridge
Implementation of end-to-end pipeline and a loan prediction machine learning model for AI Data Platform practice. Used Random Forest and XG-Boost Model development. Integrated various platforms to build a digital twin that will predict the feature mapping information and take the recommendations from experts for continuous improvement of the model. Developed and presented various prototypes to clients and the IBM Global leadership team.

IBM - BAT (British American Tobacco)

- Data Migration from one ScyllaDB Cluster to another ScyllaDB Cluster. Data Transformation using PySpark and Data Extraction using Apache Spark (Scala) using datastax spark-cassandra-connector framework.
- Integration of several platforms (IBM COS, ScyllaDB, PostgreSQL, Kafka, Slack) by Python.
- Automation of procedures using Python Programming Language. Implementation of APIs using Python micro-services on top of k8s (Kubernetes). Deployment of various applications on Kubernetes (k8s) and Deployed ScyllaDB as an independent cluster and integrated various tools for monitoring, automated alerts, and management tasks.
- Production Cluster Management (ScyllaDB- Addition, Decommission, Repair, Backup).

IBM - IDI Command Line Interface, IBM - DICE, IBM - DRDQ (Data Reconciliation and Data Quality) Engine

Implemented IDI CLI based on click python library. Implementation of Machine Learning Model Loan Use Case using Pandas and PySpark. Enhanced the Model as per RAD-ML (RAD-ML is a proven methodology for developing sellable, reusable, and scalable machine learning assets methodology. Implementation of various automated jobs for data reconciliation using Python. Implementation of DRDQ Engine (Data Reconciliation and Data Quality) using Scala.

Climate Change: Spatio-temporal segmentation and tracking of weather patterns with light-weight Neural Networks

Experimented with CGNet model suitable for lightweight networks based on epochs, activation function, and different loss functions such as Cross Entropy and intersection Over Union. Results showed using contour plots and histograms.

Implemented Deep Learning Models from Scratch

Implemented and learned MLP, MLP Mixer, Resnet 18, Bidirectional GRU encoder-decoder with attention, Transformer with Multiheaded attention, and LayerNorm. Implemented VAE (Variational Auto-encoder). Explored various aspects such as Stop gradient, momentum encoder, and MLP predictor for SimSiam Self-Supervised network. Explored DDPM.

Classifying Handwritten Digits (Modified MNIST), Text Classification

Implemented CNN Ensemble with voting classifier and logistic regression model. Implemented LSTM model, MLP Model, Naive Bayes. Implemented word2vec embeddings and explored pre-trained models and pre-trained embeddings.