DIVYANSHU BHARDWAJ

bhardwaj.di@northeastern.edu | LinkedIn | GitHub | Profile | 857-334-9377 | Boston, MA

EDUCATION:

Northeastern University, Boston *Graduating May 2023*

Master of Science Information Systems, Data Science

Relevant Courses – Deep Learning, Big Data Analytics & Intelligence Systems, ST. Natural Language Processing

University of Mumbai, Mumbai *Jun 2018*

Bachelor of Engineering in Electronics

Relevant Courses – Mathematics, OOP Concepts, Structured Programming Language

TECHNICAL SKILLS:

Languages: Python, R, Java, PL/SQL, PySpark, JSON
Databases: MYSQL, Oracle11g, SSMS, Google Cloud Storage, Redshift, S3, Firebase
Libraries: TensorFlow, PyTorch, OpenCV, NumPy, Pandas, SciPy, scikit-learn, seaborn
Tools: Tableau, Power BI, Spark, JIRA, GIT, GCP, AWS, SageMaker, AWS EC2, Airflow, Docker, Fast API, Streamlit

EXPERIENCE:

Northeastern University, Boston

Graduate Research Assistant: *Sept 2022 - Present*

- Worked on a massive corpus of material to stop modern slavery in the global supply chain with the data shared by businesses
- Extracted PDFs and texts for 50k URLs and reduced its processing time by 80% using multithreading
- Performed data pre-processing, applied LDA that helped to find hidden semantic structure in documents

Graduate Teaching Assistant: (Neural Networks Methods & tools, Intro to Statistics) *Sept 2022 - Present*

- Assisted students with their coursework related to statistics & ML models, improved conceptual skills through assignments
- Supported the professor to mentor and manage 30+ students' performance reviews through dashboards

Motorola Solutions, Chicago

Data Scientist Intern: *May 2022 – Aug 2022*

- Collaborated with cross-functional team to develop solution by identifying relevant data sources and structuring problem
- Optimized the algorithm for correlating incidents reducing the manual effort by 40% for 2M+ records
- Developed a solution to identify events that can be reduced in FM System, 90M+ events a year, overall, 11% reduction
- Implemented collaborative filtering using Matrix Factorization, achieved an RMSE score of 1.58 using ALS technique
- Extracted required data from MySQL server and created dashboards in tableau to improve the business needs

L&T Infotech (LTI), Mumbai, India

Data Scientist: *Jan 2020 - Jun 2021*

- Exercised data mining, web scraping on multiple instances and discovered areas that needed manual effort reduction
- Planned, designed, and constructed a python-based automation for clients, resulting in $150K cost savings annually
- Automated SOX team process using python which saved manual efforts of 120 hours a week and 30% manpower reduction
- Monitored KPIs and DelEx parameters, interfaced with clients to evaluate business needs & proposed convenient solutions
- Developed a query to extract reports from a database with 2M+ rows of data using python

Software Engineer: *Aug 2018 - Dec 2019*

- Enhanced complex SQL queries to efficiently extract large volumes of data from various sources
- Collaborated with a team of 10 members to implement processes that maintained SOX compliance
- Deployed an auto ticket triaging tool which reduced 90% resource efforts that had to work on 24/7 shift

ACADEMIC PROJECTS:

Business Meeting Summarization (GCP Cloud Function, Firebase, FastAPI, Airflow, SQL)

- Orchestrated ETL pipelines and automated video suggestions using Apache Airflow, Established API Endpoints using FastAPI
- Provided the model as a Service & Formulated SQL queries to join from multiple databases and visualize results in Bigquery

Image classifier for the Intracranial Hemorrhage Detection (Keras, TensorFlow, Conv2D)

- Built a CNN model with RSNA dataset that classifies hemorrhages using 3 convolutional layers & 2 fully connected layers
- Applied MaxPooling, Batch Normalization and Dropout techniques to increase the accuracy on validation data to 89.55 %

Yelp Reviews Topic Modeling: (Python, NLP, spacy, Gensim)

- Implemented sentimental Analysis on Topic Modeling to analyze the difference in positive & negative reviews using LDA
- Evaluated feature extraction techniques - count and tf-idf vectorizer for optimal sentiment predictions

Time Series Analysis & Temperature Forecast: (Python, Statsmodel, ARIMA)

- Scrutinized the behavior of global temperatures from 1850-2020 and predicted temperatures for next 25 years
- Optimized Built models based on Random Forest, VAR & SARIMA and compare MAE (Lowest 0.08) to maintain below 1.5℃