

EDUCATION

Northeastern University, Boston, MA, USA – *Khoury College of Computer Sciences*

Sep 2021 – Dec 2023

Masters of Science in Data Science (GPA: **3.9**)

Courses: Statistics & Probability, Data Mining & Unsupervised/Supervised ML, NLP, Deep Learning, Database Management & Processing, DSA, Advanced SQL

Gujarat Technological University, Surat, GUJ, INDIA – *Pacific School of Engineering*

Aug 2016 – Aug 2020

Bachelor's of Engineering in Computer Engineering (CGPA: **8.77**)

Courses: Differentials, Calculus & Linear Algebra, Programming (C, C++, Java, Python, SQL), Data Mining & Business Intelligence, Big-Data, DSA, AI, CVPR

EXPERIENCE

PUMA North America, Inc | Somerville, MA – **Data Scientist**

July 2022 – Dec 2022

Time Series Forecasting – *Python, PyTorch, Tensorflow, Statsmodel, Pandas, Numpy*

- Deployed Time Series Forecasting System for Panel Data in Azure using Python & YML, abating cancelations by **56%** resulting in increased revenue of **\$1M+**
- Developed accurate Univariate & Multivariate Panel Data Time Series & Statistical Models to manage Inventory Stocks & mitigate potential annual losses
- Performed EDA & advanced SQL analytics on **100k+** Products, applied rolling statistics, imputed with mean-interpolation, filtered & curated **top 2k** products
- Employed *ARIMA, SARIMA, Time Series Regression (Random Forests), RNNs/CNNs, etc* for Univariate & Multivariate Cross-Sectional Time Series models
- Implemented an ensemble technique reducing RMSE by **34%** & resulted in **80%** accuracy, constructed interval estimates (**KPIs**) with **95%** confidence levels

Anomaly Detection – *Python, PyTorch, Tensorflow, Statsmodel, Darts, ADTK, sktime*

- Developed an Anomaly Detection system in *Python* to identify outliers, enhancing reliability of analytical reports & reducing manual intervention by **90%**
- Tackled data integrity issue in retail sales database, where intermittent local store register inactivity led to wide data gaps, impacting report accuracies
- Automated data retrieval in 'openpyxl' library, processing **90** days of hourly sales & applying rolling statistics creating Time Series variables with pandas
- Trained Univariate time series models like AR, MA, RNN, fbProphet for each rolling statistic & performed backtest-predictions to flag discrepant estimates
- Used voting across models for different rolling stats, achieving **71%** Accuracy; integrated tkinter & auto_py_to_exe to create executable **APK** with interface

Northeastern University, Khoury College of Computer Science | Boston, MA – **Graduate Teaching Assistant**

Jan 2023 – Dec 2023

CS 7150: Deep Learning -*PyTorch, Tensorflow, Gradescope, Canvaslms*

- Restructured curriculum focused on practicality of neural network theories, including Transformers & designed Diffusion Models homework from scratch
- Fostered 90% increase in student engagement through interactive learning environments & orchestrated panel discussions of trending Research Papers

DS 4400: Machine Learning & Data Mining – *Python, Scikit-learn, Pandas, Numpy*

- Assisted in delivering a comprehensive Machine Learning & Data Mining course, integrating theory with hands-on practice & clarified complex concepts
- Guided experiential learning projects, translating complex concepts into tangible applications and enhanced students' proficiency in predictive analytics

SKILLS

Programming: Python (Advanced), R, C, C++, Java, Advanced SQL (MySQL)

Methodologies: Time Series, DL, LLMs, Supervised/Unsupervised ML [Classification, Regression, Dimensionality Reduction, Clustering]

Algorithms: KNN, Decision Trees, Naive Bayes, SVMs, Logistic Regression, Ensemble Methods, K-Means, Hierarchical Clustering

Libraries: Scikit-Learn, PyTorch, Tensorflow/Keras, Darts, ADTK, sktime, statsr, PySpark

Data Manipulation/Visualization: Pandas, Numpy, Matplotlib, Seaborn, PIL, Geoplotlib, celluloid, Manim, BeautifulSoup, Scrapy, PowerBI

Database/Frameworks/Tools: Advanced Excel, MySQL, NoSQL, Django, Tkinter, auto_py_to_exe, JSON, XML

Cloud Services/Devops: Azure Cloud (ML Studio, DataFactory, Data Lake), Google Cloud (BigQuery), MATLAB, Git, Docker

Blog Publications: [Sentimental Recommendation System](#), [Time Series Prediction Interval](#), [Attention is Not Enough](#)

PROJECTS

Product Sales Analysis (*Python, pandas, matplotlib, Data Modeling & EDA*)

Jan 2022– May 2022

- Processed dataset of 180k+ observations to discover under-lying patterns, comprehend a benchmark & understand consumer behaviour to increase sales
- Executed Data Wrangling & Feature Engineered imperative variable, followed by algorithmic extraction of frequent item-sets & list of best-selling products
- Post extrapolation- filtered out electronic devices to recommend, which on average could probably surge profits by an estimate of 26% in total revenue

Diabetic Classification (*Python, sklearn, scipy, Data Modeling, SML*)

Jan 2022– May 2022

- Curated a comprehensive dataset with 1.8M anthropometric observations spanning diverse Indian states, performing geospatial analysis
- Adopted Grid Search Cross Validation to fine-tune Multi-Layer Perceptron Model's parameters, maximizing model efficiency for data-driven insights
- Addressed challenge of imbalanced data by implementing SMOTE technique, getting substantial leap in accuracy from a modest 13 to an impressive 72%

Visual Grounding - You speak, We'll find it (*Python, PyTorch, Transfer Learning, Deep Learning, Transformers*)

Jan 2023– April 2023

- Designed an efficient & versatile pipeline leveraging flickr30k API for data mining & acquisition, extracting Annotation entities (B-Boxes), Phrases & Images
- Extracted embedding representations using Pre-Trained Vision Transformers & BERT, representing nuanced multi-modal data with exceptional granularity
- Built baseline Transformer Model for B-box prediction with 75% IoU & 58% accuracy & variational models with performance at par or exceeding baseline

Image Search Based Discovery Engine (*Python, Tensorflow, Transfer Learning, Deep Learning*)

Jan 2023 – April 2023

- Crafted a Discovery Engine processing 45k high-res images, drawing deep insights using Transfer Learning with Pre-Trained models ResNet-50 & VGG-16
- Incorporated PCA with Knn, Cosine Similarity & LSH & implemented granular object detection for personalized & targeted product recommendations
- Achieved 66% accuracy for targeted recommendations with high buying propensity using Cosine Similarity with VGG-16, surpassing PCA with Knn with 31%

Neural Machine Translation (*Python, PyTorch, Self-Attention, Transformers, Encoders-Decoders, NLP*)

May 2022 – Aug 2022

- Engineered an efficient data pipeline for audio-derived multilingual data, transforming it into vectorized formats tailored for end-to-end ML pipelines
- Implemented diverse architectures—RNNs (LSTM/GRU) & Transformers (advanced ML techniques) using PyTorch for accurate language translations
- Achieved 56% BLEU & 62% Precision, outperforming baseline models (37% & 45%) emphasizing effectiveness of solution in real-world translation scenarios