SOHAM SANDEEP SHINDE

Boston, MA | (857) 376-8612 | shinde.so@northeastern.edu | linkedin.com/in/sohamsshinde/ | GitHub | Portfolio

EDUCATION

Khoury College of Computer Sciences, Northeastern University, Boston — Sep 2021 - Present
Master of Science in Data Science (GPA : **3.86**/4.0) — Expected Graduation: Dec 2023
Coursework: Supervised & Unsupervised Learning, Algorithms, Deep Learning, Natural Language Processing, Statistics

Don Bosco Institute of Technology, University of Mumbai, India — Aug 2017 – Jun 2021
Bachelor of Engineering in Computer Engineering (GPA : **3.8**/4.0)
Coursework: Big Data Analytics, Cloud Computing, Artificial Intelligence(AI), Soft computing, Software Engineering

WORK EXPERIENCE

Research Assistant | Qimin Yan Lab, Northeastern University | Boston, MA — May 2023 - Aug 2023
- Constructed Equivariant Graph Neural Networks for atomic potentials using group theory and spherical harmonics in PyTorch
- Utilized Crystal Graph-CNNs to predict elasticity and dielectric tensors **improving MAE by 16%** for the Material Project-17
- Developed Material Recommender System employing Bipartite Graph Networks with CSM-based Structure-Motif relations

Data Science Engineer Co-op | SS&C Intralinks | Waltham, MA — Jun 2022 - Dec 2022
- Developed Topic Labeling model utilizing PageRank-Trigram to generate labels with **40% increase** in concept categorization
- Implemented advanced OCR techniques using OpenCV, Microsoft Table-Transformer, and Tesseract, recognizing complex table structures within **5000+ rows** of scanned financial PDFs, amplifying data accessibility and catalyzed insightful analytics
- Enhanced **keyword accuracy** by **30%** through the implementation of ALTOXML, and successfully applied the Longformer model for Named-Entity-Recognition in Chinese text, facilitating improved language support and document understanding
- Deployed production models as Docker containers using Kubernetes, leading to a **40% reduction** in ETL-driven testing time

Khoury Teaching Assistant | Northeastern University | Boston, MA — May 2022 - Dec 2023
- Lead interactive code sessions for **60+ students,** enhancing their understanding in Data Science and Unsupervised Learning

Mobile Analytics Intern | TwinTring LLP | Mumbai, India — Mar 2020 - May 2020
- Engineered a data-driven navigation-networking app for tracking activity data, resulting in **CTR of 98.9%** for biking groups
- Leveraged Tableau dashboards to analyze a customer's usage and recommend personalized routes, boosting user engagement

PUBLICATION

ML-Based Shopping System with Recipe Recommendation 2021 IEEE (ICCICT) (link)
- Designed Recommender System to dynamically suggest recipes from **80k+** diverse ingredients, surpassing traditional methods
- Applied Collaborative-Content based Filtering with Text-Rank tag prediction, improving the precision product personalization
- Integrated Sentiment Analysis on **15K records** with Tableau for real-time analytics, enhancing the retail shopping experience

PROJECTS

Electricity Price Forecasting (LSTM, ARIMA, SARIMA, Time series, Neural Networks) (GitHub)
- Forecasted daily and yearly prices using Timeseries analysis obtained by scrapping generation, consumption, weather data
- Feature engineered candidate variables using sliding window and applied Auto-Regression Differencing for reduced errors
- Optimized LSTM Model, achieving **low MAPE of 9.69%** and surpassing SOTA Kaggle model, with **32% reduced RMSE**

Semantic Segmentation with SWIN Transformers (PyTorch, TensorFlow, Deeplab, Resnet) (GitHub)
- Implemented state-of-the-art using UNET, Transformers and transfer learning by fine-tuning model on **5k+ Cityscapes data**
- Achieved significant improvement in **mIOU score of 63%**, utilizing SWIN attention residual network with ML Perceptron

Question Answering model using BERT and derivatives (BERT, Hugging Face) (GitHub)
- Created a scalable QnA model by leveraging preprocessed Word2Vec, SIF embeddings on **SQuAD v1.1 with 100K+ pairs**
- Achieved **high accuracy of 81% EM** and **84.5% F1-Score** by implementing Distil-BERT ensemble transformer models

SKILLS

Programming:	Python (Pandas, NumPy, Scikit-learn, SciPy, Keras, NLTK, Matplotlib, Plotly), R, Java, C/C++
Databases:	SQL, MySQL, Oracle, Firebase, Postgres, MongoDB
Machine Learning:	Linear Regression, Clustering, Logistic Regression, EDA, Classification, K-Means, Random Forests, Forecasting, Hypothesis Testing, Data Analysis, NLP, Computer Vision
Tools:	PyTorch, TensorFlow, Tableau, Power BI, AWS, Azure, Git, Excel, Scrum Certification (link)