# SOHAM SANDEEP SHINDE

Boston, MA | 857-376-8612 | [shinde.so@northeastern.edu](mailto:shinde.so@northeastern.edu) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

**Northeastern University, Boston, MA**  Sep 2021 - Present
**Khoury College of Computer Sciences**  Expected Graduation: Dec 2023
*Master of Science in Data Science*  **GPA : 3.86**
Related Courses: Supervised & Unsupervised Machine Learning, Algorithms, Deep Learning, Natural Language Processing

**University of Mumbai, Mumbai, India**  Aug 2017 – Jun 2021
**Don Bosco Institute of Technology**
*Bachelor of Engineering in Computer Engineering*  **GPA : 3.84**
Related Courses: Big Data Analytics, Cloud Computing, Artificial Intelligence(AI) & Soft computing, Software Engineering

## SKILLS

| | |
|---|---|
| **Programming Languages:** | Python, R, SQL, Java, C, C++, C# |
| **Frameworks:** | NumPy, pandas, scikit-learn, PyTorch , Keras, TensorFlow, NLTK, Matplotlib, SciPy |
| **Databases:** | MySQL, NoSQL, Oracle, Firebase, PostgreSQL, MongoDB, Hadoop, SQL Alchemy |
| **Machine Learning:** | Regression, Random Forests, Boosted Models, SVM, K-Means, CNN, Neural Networks |
| **Tools & Certifications:** | AWS EC2, S3, Spark, Azure, Tableau, Power BI, Excel, Scrum Fundamentals Certified [link](#) |

## WORK EXPERIENCE

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

**Data Science Engineer Intern** | **SS&C Intralinks** | Waltham, MA  *Jun 2022 – Dec 2022*
- Developed **Topic Labelling** model using **PageRank** and Trigrams to generate high-quality labels and categorize key concepts
- Implemented OCR approaches of **OpenCV**, MS Table-Transformer and Tesseract to extract tabular data from scanned PDFs
- Enhanced quality of Keywords using ALTOXML and utilized Longformer model for **Named-Entity-Recognition** in Chinese
- Deployed autonomous production models as **Docker** containers levering **Kubernetes** for rigorous testing and streamlined ops

**Khoury Teaching Assistant** | **Northeastern University** | Boston, MA  *May 2022 – Present*
- Mentored **60+** students conducting code sessions for **Database Design**, **Data Science Foundation & Unsupervised Learning**

**Mobile Analytics Intern** | **TwinTring LLP** | Mumbai, India  *Mar 2020 – May 2020*
- Developed Android GPS navigation and networking app for tracking activity data, resulting in CFR of **98.9%** for **bike** groups
- Analyzed customer usage data using **Tableau** dashboards and recommended **personalized routes** and relevant ride challenges

## PUBLICATION

**ML-Based Shopping System with Recipe Recommendation** 2021 IEEE (ICCICT) [link](#)
- Designed a **Recommender System** to suggest top 10 recipes based on ingredients in dynamic cart for a dataset of **80k+** values
- Employed Collaborative Filtering suggested associated products, predicting tags using **Text-Rank** and Content-based Filtering
- Conducted **Sentiment Analysis** on **15K** records, improving user shopping experience by embedding Tableau analytics content

## PROJECTS

**Claim Prediction in Travel Insurances** (SMOTE, Random Forest, XGBoost, Flask, REST API) [GitHub](#)
- Developed an **ensemble** using boosted models to classify imbalanced claims data using feature selection and SMOTE analysis
- Utilized Flask framework to deploy the trained model as REST API, with prediction accuracy of **94.69%** and F1-Score of **0.84**

**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
- Achieved a low Mean APE of **9.69%** for LSTM Model, outperforming the 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 mechanism with ML Perceptrons

**Question Answering model using BERT and its 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-BERT** ensemble transformer model