Nidhi Bodar

bodar.n@northeastern.edu | LinkedIn | GitHub |Boston, MA | (857) 317-0223

EDUCATION

Northeastern University, Boston, MA, US **Expected Graduation: Dec 2023**

Khoury College of Computer Sciences, Master of Science in Data Science

Graduate Teaching Assistant: Introduction to Programming for Data Science (DS 5010)

Related Courses: Supervised Machine Learning and Deep Learning, Database Management and Preprocessing, Unsupervised Machine Learning and Data Mining, Natural Language Processing, Introduction to Linear Algebra, Probability and Statistic

Sarvajanik College of Engineering & Technology, Surat, India **May 2021**

Bachelor of Engineering in Information Technology

Related Courses: Data Structures and Algorithms, Big Data Analytics and Distributed Computing, Artificial Intelligence, Business Intelligence and Optimization, Object Oriented Programming and Modelling, Data Compression and Retrieval

TECHNICAL SKILLS

Languages, Databases, Tools:	Python, R, SQL, JavaScript, C, C++, MongoDB, JAVA, MySQL, SQLite, Agile, Jira, Confluence
Python Libraries:	NumPy, Pandas, SciPy, Scikit-learn, Keras, TensorFlow, Matplotlib, NLTK, OpenCV, Statsmodels
Machine Learning Algorithms:	Linear Regression, LightGBM, Discriminant Analysis, Principal Component Analysis, Support Vector Machine, Decision Trees, K-Means Clustering, Ensemble Models, RNN, CNN, LSTM

PROFESSIONAL EXPERIENCE

Fidelity Investments | Data Science Co-op — Boston, US ***Jan 2023 – Jun 2023***

- Applied NLP techniques such as sentence classification, named entity recognition, and the Spacy library to extract valuable insights that significantly reduced the time taken by account executives and MDs to approach prospect clients with Fidelity's products.
- Experimented with three approaches to remove duplicate data, implementing fuzzy logic to reduce the duplication ratio from **18%** to **2%** with a **0.02%** error rate.
- Improved the accuracy of baseline ensemble models by **5%** through hyperparameter tuning using Optuna.
- Collaborated with cross-functional teams to efficiently deploy end-to-end machine learning models. Acquired experience with AWS Sagemaker, Amazon S3 and enabled version control using Git.
- Leveraged Snowflake to extract and analyze data, generating insights that drove business decisions.

Pro start me Technology Private Limited | Data Science Intern — Surat, India ***July 2020 – Dec 2020***

- Scrapped and analyzed tweets using **Twitter API**, identified tweet authors who used third party services as potential customers, stored information about company's competitors to the database. Increased the support team's productivity by connecting their existing knowledge base to an AI-based chatbot and noticed **40%** growth. Showed insights using **Tableau**.

IBL INFOTECH | Machine Learning Intern — Surat, India ***Feb 2019 – June 2019***

- Deployed Naive Bayes model using **Flask** web service for client to predict the probability of a customer purchasing the insurance policy, identified patterns to improve customer retention rate by **20%**. Proposed **A/B testing** approaches.
- Collaborated on-site with data engineers and client team to discuss the automated tasks in machine learning ETL pipeline.

ACADEMIC PROJECTS

Quora Question Pairing Using Natural Language Processing, Northeastern University **July 2022 - Aug 2022**

- Equalized dataset to implement **BERT** and **XGBoost + TF-IDF** model on **word, character and N-gram level** and examined their performance of finding similarity between two questions. Introduced cosine similarity to enhance performance and obtained **0.67 Recall** and **80% accuracy** with character level TF-IDF + XGBoost model.

Employee Turnover Prediction, Northeastern University **Nov 2021 – Dec 2021**

- Incorporated **Random Forest** and **Logistic Regression** to predict the number of employees and used feature importance score to extract significant features affecting the target variable.
- Performed confusion matrix and ROC curve to determine the better prediction model by comparing their accuracy where **RandomForest** model being the best model with **92.4%** accuracy.

Price Listing Predictions and Forthcoming Analysis of Airbnb, Northeastern University **Oct 2021 – Nov 2021**

- Applied data cleaning methods and to overcome the issue of underfitting, implemented **log transformation** on target variable whichresulted in reduced skewness and converted categorical variable into numerical by **one hot encoding method**.
- Performed **Linear, Ridge, Lasso and Gradient Boosting and bagging models** obtained **0.42 RMSE** and **0.87 F1 score.**