

# SHUSHANT GHOSH

Boston, MA | P: +1 8133904691 | [ghosh.shu@northeastern.edu](mailto:ghosh.shu@northeastern.edu) | [LinkedIn](#) | [Portfolio](#)

Data Scientist with **2+** years of experience deploying machine learning models and conducting advanced statistical analysis. Passionate about embedding AI into business solutions. Seeking full time roles from **Spring 2025**, open to relocation.

## EDUCATION

### Northeastern University, Khoury College of Computer Sciences

Boston, MA

Master of Science in Data Science; *GPA: 3.6/4*

Expected December 2024

Relevant Coursework: Machine Learning, NLP, Algorithms, Parallel Processing, Unsupervised Learning, Large Language Models

### KIIT University

Bhubaneswar, India

Bachelor of Technology in Computer Science and Engineering; *GPA: 8.5/10*

June 2017 - May 2021

## WORK EXPERIENCE

### Highradius Technologies

Bhubaneswar, India

Machine Learning Engineer

July 2021- July 2022

- Spearheaded **20+ AI Projects** on **Invoice Payment Date Forecasting** by developing advanced regression models (**XGBRegressor**, **LightGBRegressor**). Owned the end-to-end **ML lifecycle**, from data preprocessing and feature engineering to model optimization and deployment. Achieved **85%** prediction accuracy, boosting **early identification** of delinquent invoices by **40%**.
- Developed **customer segmentation** and **clustering algorithms** using unsupervised learning techniques like **K-means** and **hierarchical clustering**, leading to **25%** improvement in prediction accuracy. Utilized python libraries such as **scikit-learn**, **matplotlib** and **seaborn**, enabling better analysis and more precise targeting strategies. Collaborated with product managers and software engineers to understand **segmentation strategies** for diverse customers.
- Preprocessed and analyzed large-scale client datasets by building efficient **SQL** and **Python scripts**. Conducted thorough **exploratory data analysis** and collaborated with client-side Data Engineers to uncover **patterns** and **anomalies** in the data, ensuring the development of robust **machine learning models** and supporting **data-driven** decision-making for clients.
- Designed and delivered interactive dashboards and **visualizations** using **PowerBI** and **MS Excel** to present machine learning outcomes to **B2B clients** improving **client acquisition**. Collaborated with cross-functional teams of data analysts and business leaders, to perform **A/B testing** and **statistical analysis**, ensuring continuous model improvement and impact communication.
- Managed a team of **5+** entry level engineers, overseeing development and optimization of **ML pipelines** resulting in a reduction of modeling and implementation efforts by **30%**.

### Highradius Technologies

Bhubaneswar, India

Data Science Intern

July 2020 - June 2021

- Assisted Data Scientists and the collections department to implement and deploy efficient **MLOps pipelines** using ML frameworks like **TensorFlow**, **PyTorch**, **scikit-learn**, resulting in a **30%** enhancement in timely revenue collection.
- Automated data preprocessing, feature engineering and selection by scripting modular python functions, coupled with meticulous version control using **Git**, achieving **60%** reduction in implementation efforts.
- Promptly collaborated with **B2B clients** to grasp their **business strategies**, deploying ML models on their **AWS servers**.

## PROJECTS

### AI-Enabled Fintech B2B Invoice Management Application | Highradius Technologies

- Collected and cleaned **1 million** financial records using Python libraries (**Pandas**, **NumPy**). Applied **ETL** processes for data preparation, including handling missing values and normalization.
- Developed a predictive model for partial payments using **Time Series forecasting** (**XGBoost**, **LightGBM**). Conducted **hyperparameter tuning** and cross-validation for optimization.
- Built a **dashboard** to visualize and manage invoices, reducing collection time by **25%**. Utilized **JavaScript**, **React** for front-end, and **Flask** for back-end. Integrated **SQL**, **MongoDB** for databases, and **Tableau** for data visualization. [Project Link](#)

### Financial Advisory Chatbot using NLP and GenAI | Northeastern University

- Developed a financial advisor chatbot deploying **Large Language Models** and implementing **RAG** systems. Utilized **ChatGPT 3.5** to generate a financial Q&A dataset to train the LLM model to answer finance related question and use financial news stored in the database as reference.
- Utilized **PyTorch**, **transformers** architecture and various python libraries to manage **deep learning** tasks, data validation and analysis.

## TECHNICAL SKILLS

**Programming Languages:** Python, R, JavaScript, SQL, Java

**Technologies & Tools:** scikit-learn, SciPy, nltk, spacy, Transformers(BERT, LSTM), TensorFlow, PyTorch, Tableau, Git, MapReduce, MongoDB, AWS, Azure, MS Excel, PowerBI.