Shushant Ghosh

Boston, MA

+1 (813) 390 4691 | ghosh.shu@northeastern.edu | LinkedIn | Portfolio

Data Scientist adept at leveraging machine learning, statistics and programming to transform data into actionable insights. **Available from June 2024. Open to relocate.**

EDUCATION

Northeastern University, Boston, MA	**Expected April 2025**

Khoury College of Computer Sciences

Master of Science in Data Science	**3.50 GPA**

Relevant Coursework: Machine Learning, NLP, Algorithms, Data management and Processing, Large Language Models

Kalinga Institute of Industrial Technology, BBSR, IN	**Jun 2017 - May 2021**

Bachelor of Technology, Computer Science and Engineering

Relevant Coursework: Artificial Intelligence, Data structures & Algorithms, Software Engineering, Agile Project Management

WORK EXPERIENCE

**Associate Software Developer, Highradius Technologies	Odisha, IN**	**Jul 2021- Jul 2022**

- Delivered an average 90% modeling accuracy on **Invoice Payment Date Forecasting** by utilizing advanced regression models such as **XGBRegressor** and **LightGBM**. Conducted **feature engineering** and **feature selection**, and implemented **cross-validation** techniques, which enabled the early identification of **delinquent invoices**. Improved financial forecasting and risk management processes by **40%** for multiple **B2B clients.**
- Performed detailed **customer segmentation** and **clustering** by analyzing **payment patterns** using **Tableau** and **Python** libraries such as sklearn, matplotlib, and seaborn. Leveraged K-means and hierarchical clustering techniques, which led to a **40%** improvement in prediction **accuracy** and allowed for more precise targeting and development of tailored strategies for different customer segments.
- **Preprocessed** and **analyzed** diverse, large-scale client **datasets** using **SQL** for efficient **data extraction** and transformation. **Cleaned** and **normalized** the data followed by **exploratory data analysis** to uncover **patterns** and **anomalies**, ensuring the development of **robust machine learning models** and supporting data-driven decision-making for clients.
- Delivered insightful modeling outcomes and **data visualizations** to **B2B clients** by creating comprehensive **dashboards**, which enhanced **customer acquisition initiatives** by providing clients with a deeper understanding of their data and enabling better-informed strategic decisions.

**Data Science Intern, Highradius Technologies	Odisha, IN**	**Jul 2020 - Jun 2021**

- Assisted **Data Scientists** and the collections department to implement and deploy efficient **MLOps pipelines**, 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.

Supply Chain Forecasting and Fraud Detection | Northeastern University

- Developed and implemented **supply chain** and **fraud detection models** using **Python**, achieving **92%** accuracy for sales and **88%** for fraud detection. **Preprocessed data**, **engineered features**, and handled **missing values**. **Visualized results** with **Seaborn** and **Plotly**, and applied techniques in **time series forecasting**, **model validation**, and **optimization**.

Generative Modeling on Medical Dialogues using GPT neo | Northeastern University

- **Extracted** and **cleansed** LFQA Patient-Doctor Dialogues using **NLP** techniques with **NLTK** and SpaCy, **fine-tuned GPT-Neo 1.3B** with TensorFlow and **Hugging Face**, and developed a generative model for coherent patient responses using **transfer learning** and **attention mechanisms**.

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.

Certifications: Introduction to Data Science in Python (University of Michigan).