

# SHUBHAM GOYAL

☎ (857)370-1627 | Boston, MA | ✉ [goyal.shubh15@gmail.com](mailto:goyal.shubh15@gmail.com) | [🌐 LinkedIn](#) | [🌐 Portfolio Website](#)

## EDUCATION

### Northeastern University, Boston, USA

*Master of Science in Information Systems*

*Sep 2022 — May 2024*

**Relevant Coursework:** Big Data Analytics, Neural Modeling (Deep Learning), Network Science, Graph Databases

### Indian Institute of Technology, Kharagpur, INDIA

*Integrated Bachelor of Science, Master of Science in Geophysics*

*Jul 2014 — May 2019*

## WORK EXPERIENCE

### Verato Inc

Mclean, VA

#### Graduate Data Scientist (Intern)

*Jun 2023 — Aug 2023*

- **Matching-algorithm Enhancement:** Developed phonetic attributes for names using Soundex and Metaphone algorithms, and employed Xgboost to identify name variants with an 87% accuracy. Further integrated into the profile-matching pipeline in Snowflake to improve the accuracy of individual identity resolution
- **Bayesian Modeling for Ethnicity:** Developed a Bayesian probabilistic model to impute ethnicity information by leveraging census and location data, validated results against external and state distributions

### American Express, US Credit and Fraud Risk

Bengaluru, India

#### Senior Data Scientist

*Nov 2021 — Aug 2022*

- **Xgboost for Defaulter Collections:** Spearheaded re-development of collections model to rank order highly collectible defaulters using Xgboost attaining ~15% Gini uplift in benchmark metrics with ~\$1.2MM benefit in collections
- **External Vendor Validation:** Validated multiple external vendor risk attributes by fine-tuning in-house risk models to compare credit-default prediction performance
- **High-Balance Default Case Study:** Adjugged best presenter for in-depth analysis and default preventing recommendations, resulting in saving ~\$400K exposure at default

#### Analyst

*Nov 2019 — Nov 2021*

- **Clustering based Customer Segmentation:** Recognized as Analyst of the Quarter for leading proactive limit increase campaign based on K-means clustering for constrained spend customers resulting in 11% increase in Y-o-Y spending
- **BERT for Transaction Categorization:** Built an Information Extraction pipeline using BERT-based NLP utilities like NER & Topic modeling to create transaction-related variables from bank statements for enhanced risk assessment
- **A/B Test for Risk Strategy:** Executed a 6-month pilot to formalize risk assessment strategy for small businesses resulting in loss save of \$0.2MM, partnering with engineering & compliance teams to execute a phased solution

## TECHNICAL SKILLS

- **Tech Stack:** Python | SQL | PySpark | R | Postgres | Cypher | AWS | GCP | Docker | Airflow | Tableau | FastAPI
- **Frameworks:** Tensorflow | PyTorch | HuggingFace | LangChain | Neo4j | VertexAI | NetworkX | MongoDB | SageMaker

## PROJECTS

### LES Weather Forecasting: Multi-Modal Architecture

*Neural Modeling Project, NEU* ([GitHub](#))

- Developed a classic Encoder-Decoder architecture using ConvLSTM and LSTM for weather forecasting at Lake Michigan
- Used satellite images and meteorological data to create the model attaining a recall of 0.85 on highly imbalanced data

### TravelBud: Travel Itinerary Using GPT3 LLM

*Big Data Project, NEU* ([GitHub](#))

- Developed a travel itinerary-generating application optimized based on user areas of interest, budget, and destination
- Created custom prompt augmented with user's travel data to generate a professional itinerary PDF in a cloud ecosystem

### GrocerEase (Founder): Digitization of Groceries

*Husky Startup Challenge, NEU* ([Deck](#))

- Fine-tuned LayoutMV2 model to extract purchase information from receipt images to efficiently track groceries
- Ranked in Top-5 finalists in student-led startup incubator competing among 20 other startups for \$10K seed funding

### DIME Network Analysis: US Political Contribution Network

*Network Science Project, NEU* ([GitHub](#))

- Researched DIME data to formulate a network among the monetary contributions in US Federal Elections
- Leveraged network science principles to showcase scale-free property, and ideology-based communities among contributors

### VaniVerse: Conversational AI platform

*Big Data Project, NEU* ([GitHub](#))

- Leveraged OPENAI Whisper & GPT3 API to process audio files and generate prompts to converse using Airflow