

SHUBHAM GOYAL

☎ (857)370-1627 | San Francisco Bay Area, CA | ✉ goyal.shubh15@gmail.com | [🌐 LinkedIn](#) | [🌐 Portfolio Website](#)

WORK EXPERIENCE

BILL

San Francisco Bay Area, CA

Senior Data Scientist

Jul 2024 — Present

- Built an autonomous feature-engineering agent using an iterative hypothesis-compute-evaluate loop, reducing model development cycle time by 60% and enabling faster iteration across credit risk baselines
- Designed and deployed a multi-agent underwriting automation system that ingests financial statements, generates structured credit memos, and scores applications directly reducing manual underwriter effort and improving throughput
- Engineered a vector embedding-based semantic search layer over account ledger data using a RAG-style retrieval pipeline, enabling accurate chart-of-accounts categorization at scale for downstream analytics
- Developed and shipped to production a suite of bank risk index models for real-time credit default risk scoring at onboarding and monitoring, delivering ~\$2M in revenue uplift via XGBoost with SageMaker deployment

American Express

Bengaluru, India

Senior Data Scientist

Nov 2021 — Aug 2022

- Led end-to-end rebuild of the collections prioritization model using XGBoost, achieving a ~15% Gini uplift over baseline translating directly to ~\$1.2M in incremental collections recovery
- Developed extensive data processing pipelines in SAS directly run through shell reducing processing time by 80%
- Validated multiple external vendor risk attributes by fine-tuning in-house risk models to compare credit-default prediction
- Collaborated with cross-functional teams to devise a risk-index based on transaction-related variables from bank statements created using BERT-based NLP utilities like NER & Topic Modeling
- Documented and mentored new joiners in on-boarding, knowledge transfer, and ramping up on multiple occasions

Data Scientist

Nov 2019 — Nov 2021

- Designed and ran a 6-month causal inference experiment (test-control) to formalize risk adjudication strategy for non-responsive small business accounts, achieving \$0.2M in loss savings; collaborated with engineering and product on phased rollout
- Adjudged the best presenter for in-depth analysis and default preventing recommendations in a high-balance default case study, resulting in saving ~\$400K exposure at default

Verato Inc

Mclean, VA

Graduate Data Scientist Intern (Python, Snowflake, AWS, Postgres)

Jun 2023 — Aug 2023

- Created phonetic features and used Xgboost to identify name variants with an 87% accuracy to enhance profile matching
- Developed a Bayesian probabilistic model to impute a person's ethnicity information using census and location data

TECHNICAL SKILLS

- **Languages:** Python (Pandas, Numpy, Matplotlib, Seaborn) | SQL | PySpark | R | Postgres | Cypher | VBA | Shell
- **Tools:** AWS | GCP | HuggingFace | Neo4j | Mlflow | Airflow | Vector DBs | SageMaker | Excel | Tableau | Docker | dbt
- **Frameworks:** Tensorflow | PyTorch | LangChain | RAG | NetworkX | Transformers | Scikit-learn | FastAPI | Streamlit

PROJECTS

LES Weather Forecasting: Multi-Modal Architecture (Python, Tensorflow)

([GitHub](#))

- Forecasted precipitation using ConvLSTM and LSTM in Tensorflow on satellite images and meteorological data

TravelBud: Travel Itinerary (Python, RestAPI, AWS, Docker, GPT 3.5, HuggingFace)

([GitHub](#))

- Developed a travel itinerary-generating application using GPT-3.5 with RAG framework on user's travel details

CineSphere: Movie Recommending Chatbot (Python, LangChain, Neo4j, GPT-4)

([GitHub](#))

- Developed a chatbot using GPT4 with multi-agent LangChain Framework and Knowledge Graph over 45K movies

DIME Network Analysis: US Political Contribution Network (Python, SQLite, NetworkX, Gephi)

([GitHub](#))

- Formulated and analyzed a network among the monetary contributions in US Federal Elections using NetworkX & Gephi

EDUCATION

Northeastern University, Boston, USA

Master of Science in Information Systems (Computer Software Engineering)

Sep 2022 — May 2024

Relevant Coursework: Big Data Cloud Applications, 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