SKILLS

Tools: Python, PyTorch, TensorFlow, SQL, R, Power BI, Tabular, Excel, Azure, AWS, Hadoop, Google Cloud, SAS, Git
Analytics: Machine Learning, NLP, Data/Text Mining, A/B testing, Forecasting, Data Visualization, Exploratory Analysis
Generative AI: Prompt Engineering, Sentiment Analysis, Word Embedding, Fine-tuning, HuggingFace, Vector Database

EDUCATION

Ph.D. in Data Science, GPA:3.9 2021 – **Present** – **May, 2025**
M.S. in Applied Statistics

Publication Several Publications written - provided upon request

WORK EXPERIENCE

Microsoft Corporation **July 2019 – Feb 2021**
Data Scientist *Seattle, WA*

- Collaborated with stakeholders to define business problems, applied **Random Forest** to analyze user behavior in Microsoft Community, resulting in actionable insights and achieving commercial cost savings of $620K in FY21.
- Developed data pipelines to collect customer segments from **SQL** server and analyze M365 template subscriptions using **Support Vector Machine**. Led the **AB test** design and execution, optimizing search engine operations.
- Participated in **GenAI** algorithm design to optimize customer subscription strategies. Built interactive **Power BI** dashboards to analyze and assess the impact of marketing touchpoints at a granular customer level.
- Engaged with cross-functional teams to address analytics needs, promoting data-driven decision-making process.

Sina Weibo (NASDAQ: WB) **Jun 2018 – Sep 2018**
Data Scientist Intern *Beijing, China*

- Led algorithm research, prototype development, and strategy plan to target customers for newly launched product.
- Collaborated with stakeholders to define project use cases. Processed 5M+ user data entries from **Apache Hive**, performing data cleaning, handling missing values, and detecting fake accounts through user behavior analysis.
- Applied **XGBoost** to analyze customer features and generate actionable insights for driving decision-making.

China Development Bank **Jun 2016 – July 2017**
Financial Analyst *Lanzhou, China*

- Engaged in customer communication and conducted financial data collection, organization, and categorization in **Excel**.
- Utilized **Lasso Regression** to predict sales for newly launched products and designed dashboards to visualize results.
- Conducted customer segmentation analysis in **Python** to identify potential use cases for personalized recommendations, achieving $20K sales growth and cutting 10% marketing expenses this quarter.

Projects

Stock Price Forecasting during Critical Public Events **Oct 2024**

- Proposed a hybrid model integrating **RNN** and **GNN** using **Pytorch** to capture market volatility driven by critical public events(e.g., presidential election), achieving a 6.6% RMSE improvement in stock price prediction across sectors.
- Employed Large Language Models **FinBERT** and **LLAMA** to generate the sentiment scores and word embeddings, analyzing the impact of news sentiment and semantic information on stock dynamics.
- Integrated **Mean-Variance** method with forecasting models to conduct trading simulations and portfolio optimization.

Hospitalization Rates Forecasting During the Early Stages of the COVID-19 Outbreak **Jun 2023**

- Developed a novel **GNN** to model spatio-temporal dependencies in multivariate time series, enabling accurate hospitalization forecasts across 50 U.S. states during early-stage COVID-19.
- Proposed a **Transfer Learning** architecture to leverage general patterns from existing diseases in emerging disease forecasting, improving RMSE by up to 3.6% over SOTA benchmarks across diverse time configurations.
- Adopted **VADER** and **Sentence-BERT** to learn the impact of public sentiments and policies on hospitalizations.

Text Chatbot for Healthcare Conversation **Apr 2023**

- Designed a text-based chatbot for healthcare conversation using **LangChain**, **Milvus**, and **GPT-4**, enabling real-time medical query handling with accurate domain-specific responses.
- Developed an intuitive user interface using **Gradio**, providing a seamless and interactive platform for query resolution.