Zixiao (Tom) Tan

(919) 225-4404 ◇ zixiaotan2023@gmail.com ◇ Linkedin ◇ GitHub

EDUCATION

Duke University **Durham, NC**
Master of Science in Statistics *Aug 2024 - May 2026*
- **Coursework**: Machine Learning (Python), Bayesian Statistics, Database system(SQL), Casual Inference

Beijing Normal University **Beijing, CN**
Bachelor of Science in Mathematics (GPA: 3.80 /4.00) *Sep 2020 - Jun 2024*
- **Coursework**: Algorithms and Data Structure, Time Series Analysis, Experiment Design

PROFESSIONAL EXPERIENCE

Tencent Company **Shenzhen, China**
Data Scientist Intern *May 2025 - Aug 2025*
- **Goal:** Engineered an autonomous **ChatBI Agent** for Product teams, facilitating data exploration and lightweight ad-hoc analysis.
- **Model:** Deployed a modular Multi-Agent framework utilizing **LangGraph** with a three-layer architecture:
  1. **Data Insight Agent :** Implemented a **RAG** pipeline with **LangChain** to integrated metadata, compliance, and business knowledge bases for context retrieval.
  2. **Data Analysis Agent:** Leveraged OpenAI Function Calling for precise intent classification and Text-to-SQL generation process.
  3. **Visualization Agent:** Developed a rendering module that interfaces with the Xiaoma BI API, transforms query results into line and curve charts.
- **Result:** Integrated **LangSmith** for end-to-end observability and evaluation, achieving 95% query accuracy rate in benchmark.

Foshan Industry and Trade, InC. **Foshan, China**
Data Scientist Intern *Feb 2024 - Jun 2024*
- **Goal:** Built and deployed **AI Chatbot** for customer service utilizing **Large Language Model**, handling complex customer queries and reducing call center operational costs.
- **Model:**
  1. **Knowledge Base:** Designed a custom **Markdown tokenizer** to parse the hierarchical structure of trade documents, processing 100+ files into **80,000+ QA pairs** and vectorizing data using the **bge-zh-v1.5** embedding model.
  2. **Retrieval Optimization:** Implemented a **Hybrid Search** strategy combining **BM25** keyword matching and **ANN vector search** to enhance **Top-k precision**.
- **Results:** Integrated **AI Chatbot** into Wecom, achieving a response accuracy of **95%** and response time of **1.5** seconds.

PROJECT EXPERIENCE

Travelers Insurance Company **Durham, NC**
Travelers University Modeling Competition *Oct 2024 - Dec 2024*
- **Goal:** Developed **Predictive Model** to forecast **80,000** policyholder call counts - the number of times a policyholder is likely to call, optimizing resource allocation and enhancing cost-efficiency in call center operations.
- **Model:** Tackled the challenge of over **50%** zero call counts using **Zero-Inflated Poisson (ZIP)** model, employing a Bernoulli distribution to classify zeros as structural or random and a Poisson distribution to model random zero counts.
- **Results:** Selected **Generalized Linear Model (GLM)** over Boosting methods like **XGboost** and **LightGBM** for its interpretability, achieving a Mean Squared Error (MSE) of 25.02 and best Akaike Information Criterion (AIC) fit.

PUBLICATION

- **Tan, Z**, Li, Z, Liao, Z. Estimating the Impact of Extreme Events on CPI: An EEMD-Fourier Based Method. Chinese Journal of Applied Probability and Statistics. Under Reviewed.
- Yang, K., Liu, T., Wang, Z., Liu, J., Shen, Y., Pan, X., Wen, R., Xie, H., Ruan, Z.,**Tan, Z.**, Zhang, K. (2022). Classifying Drosophila olfactory projection neuron boutons by quantitative analysis of electron microscopic reconstruction. iScience, 104180. https://doi.org/10.1016/j.isci.2022.104180