Cheting Meng (Dakota)

Dakotameng.CM@gmail.com | (470) 923-3934 | Atlanta, GA

Personal Website | [Link] Github | PsyDak-Meng Linkedin | Cheting Meng

Work Experience

Georgia Institute of Technology Atlanta, GA

Graduate Research Assistant Aug 2022 – May 2024

- Crafted Power BI labor statistics dashboard for state DOT to increase operational efficiency by 25%.
- Deployed A/B testing experiment for NSF research on behavioral decision sciences.
- Discovered decision-making bias from data-driven causal inference using data mining (Association rule mining), unsupervised learning (Clustering) and hypothesis testing in R and Python.
- Applied NLP topic modeling (LDA) on over 5k pages of legal claims documents from state DOTs to identify breaches for contract templates.

NEW-TECHEM Co. Ltd. Taipei, Taiwan

Data Scientist Jun 2021 – Dec 2021

- Built ETL pipelines with SQL, Spark and Python to speed up EDA and data processing by 60%.
- Amplified business insights with Power BI and Python visualization dashboards to track key KPIs and reported technical details to executives to enhance decision-making efficiency by 30%.
- Increased order volume by 20% with ML applications on sales prediction (Regressions & NN regressor) and customer churn analysis (gradient boosted decision trees & KNN) in Python.

Education

Georgia Institute of Technology Atlanta, GA

Master of Science in Computational Science and Engineering (*GPA: 3.8)* Aug 2022 – Dec 2024

- Relevant Coursework: Computational Data Analysis (ML), Data Structure & Algorithms, Data Visualization & Analytics (Big data), Time Series Analysis, NLP, Computer Vision

National Taiwan University of Science and Technology Taipei, Taiwan

Bachelor of Science in Civil Engineering Sep 2017 – Jun 2021

Projects

GNN Amazon Recommendation System [Github]

- Predicted dissatisfaction score regression at 0.98 R^2 to build a recommendation system using GNN and multi-task learning in PyTorch. Used BERT embeddings of 36GB amazon review dataset as nodes in GNN.

Healthcare NLP Web Application [Github]

- Retrieved vaccination adverse effects using NLP keyword extraction (NLTK & YAKE!) and predicted them with MLP neural network classifier in PyTorch at 82% top-5 accuracy.
- Built 92% user satisfactory rated frontend using SQL, Flask, HTML, CSS and JavaScript.

Computer Vision AI Application @ *GT 2024 Hacklytics Hackathon* [Github]

- Applied OpenCV hand landmark tracker and developed a CNN-based attention model in TensorFlow to classify hand sign alphabets at 90% accuracy.
- Bridged mutism communication gap by building webpage in Flask enabling textual and audio output.

Vision Transformer Lie Detector [Github]

- Applying ViViT with optical flow intensity as substitute sampling methodology to classify lying videos.
- Enabling classification interpretability with sentiment analysis on highest attention weighted frames.

Large Language Model Evaluation Tool [Github]

-- LLM efficient evaluation through Bayesian Optimization ***[Upcoming Publication]***

- Reducing LLM evaluation time and computation cost by sampling optimal evaluation query/prompt from corpus database through Bayesian Optimization and dense passage retrieval.

Personalized Chatbot: LLM Finetuning & RAG [Video Demo] [Github]

- Efficiently finetuned GPT-neoX 20B model on arxiv-cs-ml dataset while reducing trainable parameters to 0.08% with LoRA in Colab environment.
- Deployed RAG with cosine-similarity retrieval to optimize response and reduce hallucination.

Stock Price Time Series Forecasting [Blog] [Github]

- Predicted daily time series of individual stock price with Generalized Additive Model and ARMA models (ARIMA & ARIMA-GARCH) in R, then formed statistical reports.

Course Registration Database Management System

- Developed database system for course registration using MySQL. Designed an ER-model and translated it to relational schema allowing searching and registration by student Id and course number.

Credit Card Fraud Detection [Link]

- Pipelined machine learning model data processing, training and testing on 5 different algorithms (SVM, Random Forest, MLP, etc.) to classify fraud in Python at 99.95 % test accuracy.

Earthquake Displacement Prediction [Blog] [Link]

- Performed data cleansing, EDA and data transformation on 15 years of sensor data in Python.
- Predicted displacement with multivariate Lasso Regression & Gaussian mixture model and detected structural defects with SVM. Reaching 0.024 mse and 88% accuracy with forward feature selection.

Skills

Programming Languages:	Python, SQL, R, C++, Julia, JavaScript, HTML
Libraries & Frameworks:	Scikit-learn, XGboost, Statsmodels, Scipy, Numpy, Pandas, Matplotlib, Dash, OpenCV, NLTK, PyTorch, TensorFlow, Django, Flask, D3.js
Tools:	MySQL, Spark (SparkML), Databricks, AWS, Github, Docker, Linux, Power BI, Tableau, MS Office (Excel)

Certificates

- IBM Data Science Professional Certificate
- AWS Certified Cloud Practitioner
- Tensorflow Developer Certificate [Link]