Joshua Chen

408-505-5987 | joshua.ycc.chen@gmail.com | linkedin.com/in/joshua-ycc-chen | github.com/Zaanis

Education

UC San Diego — San Diego, CA
Masters of Science in Business Analytics — *Aug. 2023 – Dec. 2024*
- Strong focus on Machine Learning
- Beta Gamma Sigma, GPA: 4.0

UC San Diego — San Diego, CA
Bachelors of Science in Cognitive Science, Specialization in Machine Learning — *Sep. 2021 – June 2023*
- Magna Cum Laude, GPA: 3.966, Major GPA: 4.0

Experience

Data Science Intern — March. 2024 - June 2024
AlphaTrAI — *San Diego, CA*
- Led a team of four, delegated tasks, and managed the project development lifecycle
- Engineered a comprehensive data acquisition pipeline, integrating APIs and web scraping techniques to aggregate data from varied sources and enhancing data diversity by 50%
- Conducted extensive research to enhance data processing techniques and feature generation, developing a streamlined pipeline that improved model performance by 25%
- Implemented machine learning algorithms, including regression and classification models, and evaluated results using A/B testing
- Authored professional analysis reports, delivering actionable insights to clients

Data Science Teaching Assistant — Sep. 2023 - June 2024
UC San Diego - Cognitive Science Department — *San Diego, CA*
- Developed Python scripts to automate assignment grading, enhancing efficiency in the evaluation process
- Mentored 11 students through one-on-one sessions, facilitating a 15% improvement in their academic performance
- Published a tutorial video instructing students on setting up SSH keys for GitHub integration with DataHub and local Jupyter Notebooks, facilitating efficient project management and coding workflow
- Received a 100% positive review from students and commendation from faculty for teaching effectiveness

Projects

Lead Author - **Domain Adaptation in Financial NLP** — Sep. 2024 - Dec. 2024
- Fine-tuned FinBERT for sentiment analysis on the Financial PhraseBank dataset, achieving 86.2% accuracy
- Conducted zero-shot and few-shot learning experiments with LLaMA 2 and GPT-4o, demonstrating competitive accuracy (84.5%) without extensive fine-tuning
- Applied data augmentation techniques to address class imbalance, improving model generalization
- Highlighted trade-offs between computational efficiency and domain-specific performance, underscoring the value of accessible LLMs for small firms or independent analysts
- Proposed future strategies for combining fine-tuning and in-context learning for financial NLP tasks

Lead Data Scientist - **Fraud Detection for Credit Card Transaction** — Mar. 2024 - June. 2024
- Designed machine learning models to detect fraudulent transactions, improving fraud detection accuracy by 20%
- Applied advanced feature engineering, including velocity-based metrics, to enhance model predictive power
- Documented results in a professional report, providing actionable recommendations to stakeholders

Lead Data Scientist - **Restaurant Rating Prediction Using Yelp Reviews** — Jan. 2024 - Mar. 2024
- Secured 1st in a Kaggle competition by developing a classification model to predict restaurant types
- Applied NLP preprocessing techniques and enhanced TFIDF features for improved model performance
- Trained and fine-tuned a DistilBERT model, enhancing accuracy by over 5%
- Boosted model performance by applying training data labels to test predictions, identified via targeted EDA

Technical Skills

Languages: Python, SQL (PostgreSQL, T-SQL, MySQL), R, Java
Methods: A/B Testing, Data Visualization, Statistical Analysis, Machine Learning Algorithms
Developer Tools: Git, Docker, VS Code, RStudio, Jupyter Notebook
Libraries: Transformers, Pandas, NumPy, PyTorch, FinBERT, Django, BeautifulSoup, Selenium, TensorFlow, Keras