

Joshua Chen

408-505-5987 | joshua.ycc.chen@gmail.com | [linkedin.com/in/joshua-ycc-chen](https://www.linkedin.com/in/joshua-ycc-chen) | github.com/Zaanis

EDUCATION

UC San Diego

Masters of Science in Business Analytics

- Strong focus on Machine Learning

San Diego, CA

Aug. 2023 – June 2024

UC San Diego

Bachelors of Science in Cognitive Science, Specialization in Machine Learning

- Magna Cum Laude, GPA: 3.966, Major GPA: 4.0

San Diego, CA

Sep. 2021 – June 2023

EXPERIENCE

Machine Learning Engineer

AlphaTrAI

March. 2024 - Present

San Diego, CA

- Developed a LLM to analyze sentiment from earnings call transcripts, enhancing market prediction accuracy
- Engineered a system to integrate sentiment analysis with 900K rows of historical stock price data and technical indicators, utilizing Snowflake and SQL for efficient ETL processes
- Designed an LSTM model to forecast stock prices using sentiment and historical data, incorporating a custom loss function for enhanced prediction accuracy

Teaching Assistant - COGS108 Data Science in Practice

UC San Diego

Sep. 2023 - Present

San Diego, CA

- Developed a Python script to automate assignment grading, enhancing efficiency in the evaluation process
- Provided one-on-one mentorship to students, contributing to a 10-15% improvement in their grades
- 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

PROJECTS

Restaurant Prediction Using Yelp Reviews | *NLTK, Transformers, Gensim, Keras*

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
- Developed a DistilBERT model, enhancing accuracy by over 5%
- Boosted model performance by applying training data labels to test predictions, identified via targeted EDA

Intuit QuickBooks Case Study | *Python, Sci-Kit*

Jan. 2024 - Mar. 2024

- Conducted EDA on dataset; identified a key column with significantly high response rates, leading to the creation of a new, impactful feature that increased response rate by 15%
- Implemented predictive machine learning models (logistic, random forest, MLP) to identify optimal customer targets, significantly enhancing marketing efforts; ranked 1st in competitive analysis among 32 groups
- Employed rigorous cross-validation techniques to evaluate model performance, ensuring robustness and accuracy in predictions for the prescribed customer targeting strategy

NFL Combine vs Career Length | *XGBoost, Sci-Kit, BeautifulSoup*

April 2023 - June 2023

- Mastered BeautifulSoup for advanced web scraping, creating a unique 8k-entry dataset, published on Kaggle.
- Addressed and cleaned 25% of the dataset containing null values, enhancing data quality and reliability
- Conducted EDA with Seaborn and Matplotlib; identified outliers, uncovered correlations for model development
- Evaluated multiple classifiers, regressors from XGBoost, Sci-Kit Learn, focusing on performance analysis
- Executed hyperparameter tuning, 5-fold cross validation, and grid-searching, achieving a peak accuracy of 70%

TECHNICAL SKILLS

Languages: Python, SQL (PostgreSQL, T-SQL, MySQL), R, Java

Softwares: Snowflake, Excel, Microsoft Windows, Tableau, Tableau Prep Builder, PowerBI, JMP Pro

Developer Tools: Git, Docker, VS Code, RStudio, Jupyter Notebook

Libraries: Sci - Kit Learn, Transformers, Seaborn, BeautifulSoup, MatPlotLib, TensorFlow, PyTorch, XGBoost, NLTK