Rueiyu Chang

michellec92@g.ucla.edu Rueiyuc 0000-0002-1825-0097

Education

University of California, Los Angeles Expected Graduate June 2025
Bachelor of Statistic and Data Science, minor in Data Science in Engineering

Relavent Course/ Project

Predictive Pricing Model for Used Sailboats Using Regression Analysis 2023.12

• Aimed to develop a mathematical model to predict used sailboat prices, incorporating factors for more accurate pricing insights for monohulled and catamaran sailboats. Built ETL pipelines to collect and analyze data from multiple sources, performing data cleaning and considering features like Length, Year, Beam, Average PPP, etc.

• Implemented Random Forest Regression and Linear Regression models to determine key factors affecting sailboat prices, with length and year as the most significant predictors via python. Conducted statistical tests (F-test, T-test, Chi-square, R-squared) to evaluate model accuracy and validate regional pricing impacts for sailboat types.

• Created a comprehensive report for a broker, developing a pricing prediction model with a stronger R^2 of 0.73 for catamarans compared to 0.40 for monohulled sailboats, highlighting key insights such as the impact of regional economic factors like PPP and length as the most significant predictor of prices.

Disaster Tweet Detection Using NLP 2024.06

• Developed a Natural Language Processing (NLP) model to accurately classify disaster-related tweets by utilizing a large dataset of 7,614 tweets labeled as disaster or non-disaster in Kaggle.

• Data preprocessing included tokenization, lemmatization, stop word removal, and TF-IDF vectorization. Three models were implemented: Logistic Regression, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN)with LSTM units. The RNN model, trained with a validation accuracy of 83.8%, effectively captured complex patterns in tweet sequences.

• The RNN with LSTM outperformed other models, minimizing false positives and negatives, proving effective for disaster tweet detection. Future improvements suggested include using larger datasets, attention mechanisms, and integrating multimodal data for real-time applications.

Data Skills

Programming: Python • SQL • R • C++
Data Analytics: Statistical Analysis • A/B Testing • Linear Algebra • Regression Analysis • Time-Series Analysis • Monte Carlo Methods • Data Visualization
Machine Learning: Random Forest • SVM • PCA • Decision Trees • k-NN • Clustering • Neural Networks