ZHUOQUAN CHEN

BROOKLYN, NY | 718-300-0078 | zhuoquan1223@email.com | linkedin.com/in/zhuoquanchen
https://zhuoquan-chen.github.io/portfolio/

I am a Data Scientist, graduated from Brooklyn College in Computer Science in 2020. Skilled in Python and machine learning, with a passion for providing effective data science solutions and application. Team-oriented, solution-driven and engaged. Given my multiple backgrounds, it allows me to translate my creativity and problem solving skills into clear and actionable insights for my team.

SKILLS

Languages / Environments: Python, SQL, Java.
Packages: Pandas, NumPy, Scikit-learn, Matplotlib, Keras, Seaborn, Plotly, Pyspark

PROJECTS

BOX OFFICE PREDICTION

Technologies used: Python / Pandas / Matploylib / Seaborn / Web APIs / Machine Learning Models

Building models for predicting film revenues using data from TMDB Box Office in Kaggle. And analyzing what factors are associated with box office to find ways to increase revenue of films.

1. Applied ELI5 and TfidfVectorizer to analyze the weight of each word in text overview to get information from the audience's expectation.
2. Then extracted the content and transformed to numerical type as new features.
3. Applied logarithm to fix some numerical features with skewing distribution.
4. Integrate the helpful features for building models.

GAMES CLASSIFICATION

Technologies used: Python / Pandas / Matploylib / Seaborn / Web APIs / Machine Learning Models

Scraped game posts from Reddit and developed a Natural Language Processing model that identifies which content of posts belongs to board games and which content of posts belongs to card games.

1. Applied CountVectorizer to tokenize all content.
2. Applied models include Naive Bayes and Gradient Boosting Regression Tree.
3. Applied GridSearch to find the hyperparameters for improving models.
4. Applied Confusion matrix for evaluating the performance of all models.

PREDICTING HOUSE PRICE IN AMES, IOWA

Technologies used: Python / Pandas / Numpy / Matploylib / Seaborn / Machine Learning Models

1. Cleaned dataset, outliers detection, exploratory data analysis and features engineering.
2. Created a linear regression model to find out those features to add the most value to a home, and the things that homeowners could improve in their homes to increase the value.
3. Used lasso regularization optimization for improving models.

EDUCATION

GENERAL ASSEMBLY

Data Science Immersive Student, September 2020 - December 2020

12-week full-time immersive educational program strengthening Data Science skills including: Python, SQL, data cleaning, data visualization, regression models, classification models, web-scraping, APIs, NLP, advanced supervised learning, unsupervised learning, time series analysis, and statistics.

BROOKLYN COLLEGE | NEW YORK, NY

B.S. in Computer Science | Graduated May 2020