# Zhi Wen Huang

Data Scientist

zhiwenhu@buffalo.edu
9174156699
https://www.linkedin.com/in/zhiwenhu/
hzhi54

## Summary

Data Scientist with a background in Computer Science and Engineering. Intrigued by Data + coding and combining both to solve complex problems. Strong expertise in building visualizations, deploying machine learning models, and understanding complex algorithm. Capability to collaborate with stakeholders to gain insights into specific design requirements and accordingly devise software solutions.

## Skills

**Programming:** Python, SQL, Java
**DataAnalysis:** Matplotlib, Seaborn, Numpy, Pandas
**Database:** MongoDB, SQLite
**Machine Learning:** Sklearn, Tensorflow, Statsmodel
**Framework and Tools:** Jupyter, Docker, GIT, GCP

## Awards

Nominated RL Challenge Competition at UB

Google Data Analytics Certificate

## Volunteering

**ReTree the District** ·
*Team Leader*
Buffalo

## Education

**University at Buffalo** Aug. 2019 to Feb. 2021
Master of Science Computer Science

**University at Buffalo** Aug. 2015 to May 2019
Bachelor of Science Computer Science and Engineering

**Metis** Sept. 2021 to Nov. 2021
Data Scientist
Metis is an ACCET accredited 12-week immersive data science bootcamp. Created five data science projects solving challenging problems using data visuals, statistics theories, machine learning techniques and deployed applications.

## Projects

### NLP: Resume Topic Modeling

**Analysis Performed:**

- Acquired data from Kaggle dataset to perform an in depth analysis of resume documents using techniques such as tokenization, lemmatization, parts-of-speech tagging, and entity recognition.
- Performed dimensionality reduction for topic modeling using models such as LDA and NMF to get important words for each topic.
- Performed K-Means clustering technique to fill in labels for each resume and presented graphs using matplotlib and seaborn library.

**Outcomes:**

- Evaluated LDA model performed better results than NMF when choosing the correct topic words for each topic.
- Proposed NLP technique can be useful when recruiters have large quantities of resumes and are trying to decide which resumes belong to certain job titles.

### Music Genre Classification

**Analysis Performed:**

- Acquired data from GTZAN dataset, a public source music file dataset with 10 labeled genres, to extract features of sound files using librosa for depth analysis of different genres.
- Performed classification model such as logistic regression, KNN, Naïve Bayes, decision tree, random forest, and XGBoost with hyper tuning techniques to find the optimal parameters for the models.

**Outcomes:**

- Evaluated classification models and found that Random Forest and XGBoost models are very similar, both performed at 91% accuracy when classifying.
- Proposed that using classification models could help new composed music be categorized by their genres in music applications.

### Used Car Prediction with Regression Model

**Analysis Performed:**

- Gathered over 10,000 data sets with 18 features from Carfax using web scraping tools such as BeautifulSoup and Selenium.
- Performed exploratory data analysis to visualize possible findings to later performed feature engineering ending up with over 30 features.
- Created linear regression models to predict car prices using engineered features and optimized model with regularization techniques and understanding the model bias-variance tradeoff.

**Outcomes:**

- Evaluated linear regression model with cross validation technique with a predictive model of 0.86 R-Sq value.
- Provided insight that indicated car prices go up or down strongly based on mileage, make, engine, and condition of the car.

## Experiences

**University at Buffalo** Buffalo
Master Program Candidate Aug. 2019 to Feb. 2021

- Researched algorithm analysis and parallel algorithm complexity
- Developed solid software development and machine learning skills through hands-on learning approach
- Created various software projects, defining project plans, and solving complex problems while meeting strict delivery deadlines