XINRU HAO

xh2500@columbia.edu| (929)-977-6752 | https://www.linkedin.com/in/xinru-hao-016a43227/ | Seattle, WA

EDUCATION

Columbia University **New York, NY**
M.A. in Statistics, GPA: 3.6/4.0 *Feb 2023*
Relevant Courses: Capital Markets & Investments, Statistical Inference, Statistical Machine Learning, Advanced Data Analysis, Topics in Modern Statistics, Stat Comp & Intro Data Science

Boston University **Boston, MA**
B.A. in Mathematics (Statistics Concentration), Minor in Economics, Cum Laude, GPA: 3.71/4.0 *May 2021*

PROFESSIONAL EXPERIENCE

GYE Zone LLC **Remote**
Data Analyst *Jan 2024 - Present*

- Leveraged **Python**'s NLTK and spaCy libraries to perform NLP analysis on a dataset comprising over 2 million student profiles, successfully extracting key essay features that enhanced model accuracy by 17.4%.
- In collaboration with the Data Science team, fine-tuned a variety of **machine learning** models using Scikit-learn, applying cross-validation and SMOTE to significantly improve model precision and recall.
- Achieved a high-performance predictive model, attaining 92% accuracy on a test set of multi-million data points, thereby increasing the admissions prediction process's effectiveness by 14%.
- Created over 20 interactive **Tableau dashboards** by leveraging SQL for robust data extraction and Python for advanced data preprocessing, markedly enhancing user engagement and optimizing the decision-making workflow.

The Zuckerman Institute at Columbia University **New York, NY**
Data Analyst *Mar 2023 - Nov 2023*

- Enhanced data integrity and optimized **SQL query** performance in the patient database, resulting in a 10% reduction in data loss and a weekly time savings of 4 hours due to increased efficiency.
- Deployed CNN to classify the brain images, allowing for more precise observations of brain patterns and features.
- Leveraged **Python** (PCA and Random Forest models**)** to achieve 86% accurate predictions of brain activity patterns, facilitating medicines decision-making with weekly reports and **Tableau** visualizations.
- Collaborated with experimental teams to enhance imaging techniques, elevating brain examination accuracy by 12% and reducing equipment side effects, while delivering data-driven insights to the R&D team.

New Oriental Education & Technology Group **Beijing, China**
Data Analyst Intern *May 2021 - Aug 2021*

- Conducted an in-depth analysis of online CET course data using **SQL**, spanning 30K+ customers and key sales metrics. Performed Clustering algorithms for a robust customer segmentation, identifying seasonal trends and loyal customers.
- Applied Logistic Regression and SVC models in **Python** to enhance customer conversion rate predictions, improving accuracy from 82.4% to 84.76%, effectively pinpointing sales influencers.
- Collaborated with the Data Science team to access **A/B test** and **T-test** results on new course strategies, delivering insights that contributed to a 15% rise in product page click-through rates.
- Researched 50+ online courses of competitors, optimized website designs, and executed promotional campaigns on social platforms, achieving a 5% increase in conversion rate and 4.92% increase in online course revenue during exam season.

PROJECT EXPERIENCE

Prediction of Telco Subscription Customer Churn *(Python)*

- Explored a classifier solution to predict if a customer will churn or stay on the site with 7000+ raw data.
- Implemented Logistic Regression, KNN Classifier, and Random Forest Classifier by comparing Confusion Matrix, ROC curve, and Precision-Recall curve.
- Applied PCA on the data and fitted the K-means model on the trained data after PCA. Concluded that K-means Clustering is ineffective as classification and ultimately selected the Logistic Regression model, achieving an accuracy of 80.26%.

German Credit Analysis (*Python*)

- Analyzed 1000 German Credit Data to optimize loan decisions with a demographic-focused predictive model.
- Conducted EDA on 20 credit predictors, including age and housing, and preprocessed data by removing duplicates, managing NA values, and implementing advanced One-hot Encoding for improved predictions.
- Achieved a 75.5% prediction accuracy in client credit risk using Logistic Regression with Elastic Net, further refined to 76% through the Random Forest algorithm.

SKILLS

Programming & Database: Python(packages), SQL, R(packages), SAS, JMP
Data Analysis: Statistical Analysis, Data Modeling, Data Visualization, Machine Learning, Deep learning
Tools: Tableau, Microsoft Software Sets, Salesforce, Spark, Power BI
Certifications: FRM(Financial Risk Management) Part I passed