HAOYANG XU

(+86) 13812742769 | haoyxu1@163.com |

Education

Indiana University, Bloomington **Sep. 2018 - Dec. 2021**

Bachelor of Science in Information Systems *Kelley School of Business*

- GPA: 3.76/4.0; Major GPA: 4.0/4.0

Work Experience

Strategic Planning Internship | *Bosch Automotive Products Suzhou Co., Ltd.* **Oct. 2022 – Now**

- Pull daily data from SQL; summarize and update bills of materials using SAP
- Determine the booking volume based on weekly individual material shortage and inventory tables using Power BI.
- Analyze daily stock updates and inventories using SQL and Python.
- Write Python scripts for process automation, e.g, automatically re-formatting customer consumption data and stock trading data, sending pickup notifications to freight drivers, etc.

Data Engineer Internship | *Duxiaoman Financial Technology Company, Robotic Process Automation* **May. – Aug. 2022**

- Data preprocessing: denoise speech data using techniques as MSNE, MCRA, IMCRA, OMLSA, etc.
- Deep learning model applications: applied Speech Recognition model and BERT-based Neural Translation model to convert speech input in English into Chinese sentences.
- Model selection: tune hyperparameters by 10-fold cross-validation on labeled sub-datasets.

Business Analyst Internship | *Bank of China* **Jun. – Aug. 2020**

- Data analysis and interpretation through statistical ML models such as Linear Regression and Decision Trees.
- Implemented the ingestion of staff information data using ETL of SQL queries.
- Transformed disparate data sets into standardized reports for Data Lake Project.
- Rotated through PMO and assisted directors to organize and fulfill business and information needs.

Research Projects

Customer Churn Analysis | *Python, Scikit-Learn, Pytorch* **Jul. 2021 - Sep. 2021**

- Data processing: use matplotlib to plot feature distribution; remove null values and outliers.
- Feature engineering: encode and scale data using OrdinalEncoder and StandardScaler
- Build classifier with Linear Regression, Multilayer Perceptron, Support Vector Machine, XGBoost, and LightGBM
- Over-sampling training set using SMOTE
- Figure out the importance of features and their monotonicity relation with the prediction target

Course Projects

R Implemented Analysis | *R language* **Nov. 2021**

- Data preprocessing and feature selection on original datasets.
- Realized functions: file input, data structure of grid storage, control points for changing shape.
- Led a team of three students; responsible for project management, work distribution and information integration.

Statistical Analysis | *Excel* **Nov. 2020**

- Collected 179 samples to analyze the sale price of used Honda CRV vehicles.
- Calculated statistics (mean, median, std) and constructed a confidence interval for the sale price, mileage and age.
- Built regression model to predict price from mileage and age, testify the regression results from residual distribution.
- Developed a multiple regression model for qualitative variable identification and interpretation.

Deloitte Case Competition | *Excel* **Oct. 2020**

- Constructed different statistical graphics (Pie chart, Bar chart, Line chart) to analyze coffee beans' sale status.
- Explored coffee beans sale trending and visualize relationships between sales volume and month.
- Implemented Monte Carlo simulation model to calculate the store's total profit.
- Assumed 6 decisions by purchasing different amounts deposit to compare total profit by simulation.
- Interpreted results to predict the proper demand amount and deposit amount.

Technical Skills

Python, C#, SQL, R Language, MicroSoft Office, Advanced Excel, HTML, STATA