# Yuqi Han

Email: hanyuqi@wustl.edu | Mobile: 314-229-3691 | LinkedIn | Saint Louis, MO

## EDUCATION

**Washington University in St. Louis** **Saint Louis, MO, USA**

*Master of Information System Management; Master of Engineering Management* | **GPA: 3.91/4.0** *Jan 2020 - Dec 2021*

- Relevant Courses: Foundations of Analytics, Analytics Applications, Decision-Making & Optimization

**University of Electronic Science and Technology of China** **Chengdu, Sichuan, CHN**

*Major: Measurement and Control Technology & Instrumentation* | **GPA: 3.27/4.0** *Aug 2015 - May 2019*

## SKILLS

1. **Language Skills:** Python (NumPy, pandas, matplotlib, scikit-learn, TensorFlow, NLTK, etc.), C, R, SQL, Linux
2. **Tools Skills:** MySQL, MATLAB, Tableau, Hive, Jupyter Notebook, Excel, Microsoft Office, Microsoft Project

## INTERNSHIP EXPERIENCE

**New Channel International Education Group Limited** **Nanjing, CHN**

***Intern,*** *Jinqiu Study Abroad Institute* *Oct 2019 - Jan 2020*

- Analyzed students' current abilities and future development plans based on client needs and student profile.
- Extracted various data from individual students and individual teachers and used Excel to develop class placement plans, educational resource allocation plans and comprehensive study plans for Alevel and IELTS courses.
- Summarized the records and progress of individual students and individual assistants on a weekly basis to create data visualization and reports. Used reports to arrange mock and formal exams for students.

**Integrated Clinical Data Platform Solution Design** **Saint Louis, USA**

***Capstone Project,*** *Centene Corporation* *Sep 2021 - Dec 2021*

- Proposed a solution for Centene Corporation to create a real-time electronic prior authorization system to save close to 40% of the working time of each physician in the system.
- Completed an analysis of the medical data processing workflow: the role of prior authorization, the method of collecting medical data, the role of FHIR in data processing, and the handling of pending cases.
- Identified key stakeholders, road map and timeline for the overall project. Calculated the projected expenditures and revenues for the entire project. The overall ROI of the project is expected to exceed 3.24.

## COMPETITION EXPERIENCE

**Kaggle - Tabular Playground Series - Jul 2021 (Top 11%)** *Jul 2021 - Aug 2021*

- Used 7,111 weather condition records for 2020 to predict weather pollution from January to April 2021.
- Generated additional features such as hour, weekend, time, and sensor displacement through feature engineering.
- Performed EDA, generated distribution of feature and target values, heat map of features, trend analysis of target and features over time using SNS and matplotlib.
- Adjusted hyper parameters of the CatBoost model to improve the prediction accuracy. The prediction resulted in RMSLE scores of 0.14, 0.09, and 0.32 for three target values, respectively, which were 19%, 44%, and 19% lower than the results of linear regression.

**Kaggle - Tabular Playground Series - Aug 2021 (Top 17%)** *Aug 2021 - Sep 2021*

- Used a total of 250,000 loan default-related records to predict loss values for an additional 150,000 data.
- Performed EDA and generated various plots by SNS and matplotlib using the same method.
- Adjusted hyper parameters of the XGBoost model using the optuna framework to improve the prediction accuracy. The prediction resulted in a RMSE score of 7.84 for target values, which was better than the RMSE score of 7.92 using linear regression.

## PROJECT EXPERIENCE

**Analytics Project - Cause of Mortality and Medical Transcript Analysis (Team Leader)** *Oct 2020 - Dec 2020*

- Used mortality data from CDC from 2005-2015. Completed the analysis and visualization of the leading causes of death for different age groups and the trends of leading causes of death overtime using SNS and matplotlib.
- Used a binary classification model to understand the effect of varied factors on accidental, natural and cardiac mortality. For the training data, performed integrating, structuring and cleaning.
- Used logistic regression and random forest models for prediction of deaths. The AUC of predictions using logistic regression and random forest were 0.79, 0.84. Performed feature importance analysis.

**Analytics Project - Flight Delay with Weather (Team Leader)** *Apr 2021 - May 2021*

- Analyzed the relationship between flight delays and weather at the originating airport using over 1,800,000 flights and weather data from the first 30 U.S. airports in 2014. Performed the prediction of flight delays at ATL airports.
- Preformed data pre-processing and generated additional features such as departure time, arrival time, etc. Used Tableau to visualize and analyze the data. Obtained data distribution, histograms, and scatter plots.
- Developed a multi classification model to classify and predict flight delays. The model was built by BP neural network, CART decision tree, and random forest. The log-loss of classification model prediction is 0.586.

**Visualization Project - Global Supermarket Acquisition Analysis (Team Leader)** *Apr 2021 - May 2021*

- Queried the required data from the Global Superstore data set in MySQL and exported it to Excel.
- Using Excel and Tableau for visual analysis of the data, created profit distribution chart for each market and each product, correlation chart between profit and other features, regional product return rate analysis chart.