

# Sixian Chen

New York | 646-206-9416 | sc4724@columbia.edu | linkedin.com/in/sixian-resume | github.com/sixian-C

## EDUCATION

### Columbia University

Master of Science in Applied Analytics

New York

Sep 2021 – Dec 2022

- GPA: 3.9/4.0
- Relevant Coursework: Machine Learning, Advanced Data Analysis, Data Visualization, Tableau, Research Design

### Tamkang University

Bachelor of Information Management

Taiwan

Sep 2016 - June 2020

- GPA: 3.69/4.00
- Relevant Coursework: Business Intelligence Analytics, Relational Database, Statistics & Probability Theory

## PROFESSIONAL EXPERIENCE

### Ningbo Public Information Co., Ltd.

Ningbo, China

Associate Data Analyst

Aug 2020 – June 2021

- Designed and performed **A/B testing** on App banner ad and product page, doubled CTR from 2.3% to 4.6%.
- Helped design an interactive dashboard of company mobile App for monitoring user behavior metrics.
- Utilized Python to clean and analyze structured customer survey data for assessing user experience, wrote a 10-page analysis report, and presented it to high-level managers.

## RELEVANT PROJECTS

### American Express Credit Card Default Prediction

- Loaded 30GB credit card statement data from **AWS S3** and created 1823 features through feature engineering.
- Trained **LightGBM, XGBoost, and CatBoost** models on **AWS SageMaker** to predict customer default.
- Performed model stacking and model blending to improve performance, and achieved a high AUC of 96%.

### Oklahoma Credit Card Fraud Detection

- Utilized Python to create 15 features that capture abnormal transactions based on the R-F-M principle.
- Trained unsupervised learning models (**Auto Encoder, IsolationForest**) using public credit card statements.
- Captured and validated 1% anomalies from over 400,000 transaction records based on summary statistics, greatly reducing man-hours spent on fraud investigations.

### Deployment of Mortgage Default Prediction Model

- Built a machine learning pipeline with Python that automates data extraction, preprocessing, and model training.
- Constructed a **REST API** for the default prediction model and deployed API on the Heroku platform using Git.
- Deployed docker containerized mortgage default prediction model API on **AWS Elastic Container Service**.

### Kaggle Project: Predict Airbnb Rent Price

- Performed data wrangling, and exploratory data analysis by R (**tidyverse, skimr, lubridate, stringr, ggplot2**).
- Successfully reduced mean squared error by 10% through feature creation, aggregation, and variable encoding.
- Built and optimized XGBoost model with cross-validation and grid-search hyperparameter tuning, placed **4th** place among 600 competitors.

### Cryptocurrency Information Dashboard

- Used **Apache Spark Streaming** to process real-time tweets for analyzing trending cryptocurrency hashtags.
- Created ETL pipeline that retrieved real-time cryptocurrency data from Coin-cap API, parsed and stored them into **MongoDB** and **PostgreSQL**.
- Leveraged capabilities of **Metabase** to build an interactive dashboard of cryptocurrency data.

## SKILLS

- Programming: Python (Scikit-learn, NumPy, Pandas, Matplotlib, seaborn), R, SQL, NoSQL
- Machine Learning & Big Data: Regression, Classification, Time Series Analysis, Anomaly Detection, Apache Spark, PostgreSQL, MongoDB
- Data Science & Miscellaneous skills: A/B Testing, ETL, Data Wrangling, Data Visualization, Tableau, Statistics, Experimental Design, Hypothesis Testing, Git, Amazon Web Service, Docker, APIs, Excel.
- Certification: AWS Certified Cloud Practitioner, Neural Networks and Deep Learning (Coursera).