

Xinkai Chen

(412) 708-2025 | xinkaichen97@gmail.com | [LinkedIn](#)

WORK EXPERIENCE

Data Scientist

Elevance Health, Inc.

Chicago, IL

March 2021 – Present

- Built Transformer and LightGBM models to achieve 50% automated prior-approvals with 95% precision
- Performed OCR on medical records and used various NLP techniques to extract, clean, and transform text data
- Developed CNN-based models to provide information surfacing on medical records to cut human costs

Machine Learning Consultant Intern

Hamiltonian Systems, Inc.

Pittsburgh, PA

June – December 2020

- Built a TensorFlow model to predict weekly inventory demands with more than 70% accuracy
- Built models using Random Forest and XGBoost to predict the remaining useful life (RUL) of each item
- Created a Flask API to deploy ML programs on the cloud for the engineers to cut human costs by 95%

Research Assistant

Heinz College, Carnegie Mellon University

Pittsburgh, PA

February – June 2020

- Analyzed 163K Yelp reviews using Structural Topic Model in R to discover sentiments and meaningful topics
- Used OLS regression to evaluate the effect of special occasions (birthdays) on user ratings
- Applied propensity score matching to handle class imbalance and improve causal inference of the study

Data Science Consultant Intern

Deloitte Consulting

Shanghai, China

November 2018 – May 2019

- Built an invoice classification tool for clients using Tensorflow with an accuracy of more than 90%
- Developed automatic invoice verification programs for two companies to reduce human interventions by 95%
- Designed a tool to identify 98% anomalies in daily transaction data for the client using VBA and Python

Data Analyst Intern

Dell Technologies

Shanghai, China

July – October 2018

- Created a Python classification tool for inventory alert level to cut the turnaround time by more than 95%
- Performed data cleaning and aggregation on 4K weekly data of 5 different data sources using VBA and Python
- Built Power BI and Thinkcell dashboards and generated Oracle Smart View reports for weekly reviews

EDUCATION

Carnegie Mellon University

Master of Information Systems Management (Highest Distinction)

Pittsburgh, PA

August 2019 – December 2020

- **Courses:** Deep Learning, Text Mining, ML with Large Datasets, Database Management, Data Visualization

Fudan University

Bachelor of Science in Computer Science

Shanghai, China

September 2015 – July 2019

- **Courses:** Data Mining, Data Communications, Object-oriented Programming, Information Systems

TECHNICAL SKILLS

Programming Languages: Python, R, Java, SQL, Octave, VBA, JavaScript, HTML

Machine Learning Frameworks: Pytorch, TensorFlow, Keras, Scikit-learn, Spark MLlib, H2O

Tools: Spark, AWS, GCP, Jupyter, Hadoop, Databricks, Colab, Git, GPU, Linux, RShiny, Tableau, PowerBI, Docker

PROJECTS

AiBBA: GPT-2 Based Poetry Generation Model

September 2020 – February 2021

- Retrained GPT-2 using 90K poetry corpus on GPU and generated poems that fooled humans 16% of the time
- Created a website for the AI using Streamlit and FastAPI, and published the [world's first AI poetry book](#)

Music Popularity Prediction on One Million Songs

September – December 2020

- Aggregated 1M distributed HDF5 files on AWS S3 and built 5 ML models on AWS EMR using PySpark
- Achieved 72% accuracy using gradient boosting and found 5 most important features based on impurity

Machine Learning-Based Inventory Failure Prediction

August – December 2020

- Built models using Random Forest and XGBoost to predict the remaining useful life (RUL) of each item
- Led 5 people as the project manager to create technical solutions and communicate with stakeholders