Yepu Wang (Male)

Contact Address: 1720 71st St, Brooklyn, NY 11204
Phone Number: 1 9292557022 / Email Address: wangyepu98@gmail.com

EDUCATION BACKGROUND

01/2022-05/2023 **Westcliff University**

- MBA(information technology, professional track, can work 40 hours a week at any location)
- Core courses: Data Analysis & Business Intelligence, E-business Technologies, Front End Web Development, Applied Methods-Information Technology.etc

08/2020-12/2021 **Fordham University**

- Master in Data Science (graduate) GPA:3.8
- Core courses:Deep Learning, Big Data Programming, Machine Learning, Data visualization, Python, Cloud Computing, Natural Language Processing. etc

09/2015-07/2019 **Beijing University of Posts and Telecommunications (BUPT)** GPA:3.5

- Majored in Information Management and Information Systems(undergraduate)
- Core courses: Account, Database, Applied Statistics, Finance, Project Management， C++ .etc

INTERNSHIP EXPERIENCE

10/2019-01/2020 **Tian Feng Securities Co.,Ltd, Data Analyst (Equity Research)**

- Cover Bank industry and write reports on investment strategy and companies.
- Build financial models, performe equity valuation and find quality equities through data analysis.

SKILLS

Language Skill: Chinese (Native); English (Proficient); Japanese (Basic)

Computer skills: Python (sklearn, Pandas, Numpy, Seaborn, Matplotlib, TensorFlow, Keras, Pytorch), SQL, C, Spark, AWS, Google Cloud, Hadoop, MapReduce, Tableau, D3js, Power BI, SPSS

Analysis Technique: Machine Learning (Linear Regression, Logistic Regression, Decision Trees, Random Forest, Gradient Boost, CNN, DNN), Hypothesis Test, A/B Testing

PROJECTS

01/ 2021-05/2021 **Heart Disease Prediction with Logistic Regression [Big Data Programming]**

Implemented Logistic Regression to pinpoint the most heart disease risk factors and predict the overall risk by exploring the Framingham Heart dataset (4241 lines). Trained the dataset via GLM model and improved the model's accuracy by 1% compared with the classical Python script. The algorithms and results used in prediction were organized into documentation for submission.

7/2021-8/2021 **Facial Expression Recognition(wear masks)**

Add masks to around 30,000 human facial expression pictures(Fer2013) to get a new dataset. Build six deep learning models(VGG-16, Mini-Xception, Residual Masking Network, Local Multi-Head, Channel Self-Attention. Res-Netand) and use the new dataset to train these six models respectively. Design a UI interface using PyQt5, which can load the trained DNN models to do facial expression recognition(pictures and videos). Tools: Python

10/2021-12/2021 **US County-level Risk Factors Associated with COVID-19 Exacerbation During Vaccination Era**

Focus on identifying the principal risk factors in predicting COVID-19 infections and mortality rates at the county-level during the early vaccination era. Build six machine learning models(Logistic Regression, SVM, Decision Tree, Random Forest, Neural Network, XGBoost) and compare the efficacy of these established models for the machine learning tasks. Perform risk factor analysis by identifying common principal predictors revealed by the models. Tools: Python

10/2021-12/2021 **Leveraging Sentiment Analysis on 10-k Reports**

Preprocess the data and use the Loughran-McDonald sentiment word lists to perform sentiment analysis on the 10-ks. Evaluate the alpha factors using the cosine similarities, do turnover analysis and calculate Sharpe Ratio of the alphas. Tools: Python

10/2021-11/2021 **Data Visualization of Flight Delays**

Use Tableau to create a story which contains four dashboards to do data visualization of flight and flight delays to find the answers for what reason contributes most on flight delays, which reason of delay has the highest average delay minutes, does airport location or size affect flight delays. Put forward some very useful suggestions for passengers to avoid flight delays.