

EDUCATION**New York University, Center of Data Science****New York, NY**

Master of Data Science

Sep 2022 – May 2024(Expected)

University of Illinois at Urbana-Champaign, College of Liberal Arts and Science**Champaign, IL***Bachelor of Science in Statistics and Economics with Minors in Computer Science and Informatics*

Aug 2019 – May 2022

- GPA and Honors: 3.96/4.00; Magna Cum Laude; Distinction in Economics and Statistics Departments

Nankai University, College of Economics**Tianjin, China***Bachelor of Arts in Economics*

Sep 2017 – June 2019

- GPA:87/100; Teaching Experience: Course Assistant for Calculus I

PROFESSIONAL EXPERIENCE**Tencent Technologies Co.Ltd****Shenzhen, China****Data Scientist Intern in WeTV Team, Platform & Content Group**

Sep 2021 –Dec 2021

- Developed **Logistic Regression**, **Random Forest**, and **Gradient Boosting Tree** models with **Pyspark** to predict customer churn, achieving AUC of 80%
- Automated the machine learning pipeline to perform daily training and prediction to help inform user retention strategies
- Implemented an ETL(Extract, Transform, Load) pipeline to load video data using **Pyspark**, **pandas** and **Airflow**; improved data update frequency from daily to hourly by optimizing table partitioning
- Created an interactive dashboard by internal **BI system** to display **SQL-calculated** metrics quantifying users' activeness and retention

ShareIt Technologies Co.Ltd**Beijing, China****Data Analyst Intern in ShareIt App Data Analysis Team**

May 2021 – Aug 2021

- Developed key metrics quantifying the performance of file transferring and tool functions in ShareIt app, and created **Tableau** dashboards to extract insights based on multi-dimensional data pulled by **SQL** query
- Analyzed the cause of abnormal fluctuations in user activities through hypothesis testing; recommended data-driven solutions to product managers to improve the performances of mobile application
- Researched users' file transfer behaviors on Xiaomi cellphone in the Indonesia marketplace to identify the causes of file transferring failure, and presented improvement opportunities to the engineering team

Intel Corporation**Shenzhen, China****Supply Chain Business Analyst Intern**

Dec 2020 - Feb 2021

- Coordinated with sales agents and manufacturers to maintain supply chain logistics, meet customer demand, and achieve revenue goal
- Leveraged demand prediction and stocks information in SAP system to distribute products in the Chinese market

SELECTED PROJECTS**Image Classification – A CNN-based Deep Learning Project for Classifying COVID-19 Diagnosis Result****May,2021**

- Developed a **CNN** (convolutional neural network) model in **Keras** to predict COVID-19 diagnosis using over 2000 chest x-ray images from Kaggle, achieved 71% accuracy

Heart Disease Diagnosis – A Machine Learning Project for Classifying the type of Heart Disease**Dec,2020**

- Conducted exploratory data analysis, data cleaning, data transformation on human health indices using **pandas** and **seaborn**
- Trained multiple ML models (**KNN**, **decision tree**, **random forest**, **boosted models**) with hyperparameter grid search to predict the level of severity of heart disease; achieving accuracy at 81%

A Database Website Application for Song Search, Rating & Recommendation**July 2020 – Aug 2020**

- Led a group of 4 to build a database website with **CRUD** functions including querying songs, rating songs, adding songs to song lists, recommending friends and songs to users; the website uses **HTML**, **CSS** for frontend, **PHP,SQL** for backend, **matrix-factorization** technique for recommender system

Technical Skills

- Languages: Python, SQL, R, Java, HTML, CSS, PHP, MongoDB, Cypher Query Language
- ML Toolkits: Scikit-Learn, PyTorch, TensorFlow, SciPy, NumPy
- Data Visualization: Matplotlib, Seaborn, Tableau, R Shiny, Plotly, Dash, ggplot
- Software: Microsoft, Airflow, SAP