VALERIA NI

https://www.linkedin.com/in/jingwen-ni/ | 949-556-1200 | nijingwen0@gmail.com | Seattle, WA, 98052

EDUCATION

UNIVERSITY OF CHICAGO Chicago, IL

Master of Arts in Computer & Information Science *Sep 2021- Jun 2023*

- Honor: Nationwide Merit Scholarship 40k

UNIVERSITY OF CALIFORNIA, IRVINE Irvine, CA

Bachelor of Science in Mathematics & Bachelor of Arts in Business Economics *Sep 2017 - Mar 2021*

- Honor: Dean's Honor List in Physical Science School

WORK EXPERIENCE

University of Chicago

Data Scientist Part-Time *Jan 2023 - Jun 2023*

- **Supervised Learning**: Deployed Random Forest, Regression Models, Gradient Boost to predict whether admitted students will accept the offer, fine-tuned model performance based on the confusion matrix and resulted in 94% accuracy
- **Data Preprocessing**: Combined data from multiple sources, fixed duplicated records and handled observations from the missing value using Python packages (e.g Pandas, NumPy, H2O and PySpark)
- **Data Visualization:** Desigend visually appealing and informative dashboards using Tableau, presenting admitted student segmentation and acceptance rate in a user-friendly format, resulting in improved data communication
- **A/B Testing:** Spearheaded and executed the A/B testing to evaluate the effectiveness of different scholarship amounts towards various segments of admitted student groups, resulting in a 7% increase in acceptance rate and a 15% increase in engagement

Union Group USA

Data Scientist Intern *Jun 2022 – Dec 2022*

- **ETL Pipeline**: Preprocessed 200k data from PostgreSQL database to transform data by renaming, filtering, and tokenizing; engineered innovative features based on existing information, and regularly loaded data using RedShift
- **Supervised Learning:** Compared with different model's AUC and chose Random Forest Model (accuracy of 98.2%) to evaluate whether the job seeker can get the offer
- **Marketing Analytics:** Examined pre-and-post ROI uplifts to evaluate event performance, organized prioritization and provided recommendations for future marketing investment regarding event types and geographic regions

ByteDance Technology Co. Ltd.

Data Scientist Intern *Jun 2021 - Sep 2021*

- **Database Management**: Leveraged MySQL to maintain 500k data and extract incremental data based on timestamps, conducting transformation such as data granularity conversion
- **Growth Prediction:** Predicted the magnitude of K12 student and teacher user growth and multidisciplinary question database using linear regression model (accuracy 92.6%), and set ROI goals for next season based on the results
- **Metric Design:** Designed and implemented key performance metrics, including latency, match-rate (MR), click through rate (CTR), return of investment (ROI), mean average precision (MAP)
- **Metric Monitoring:** Built interactive dashboard using Aeolus, detected the abnormal data fluctuation and coordinated with stakeholders to meet the production expectations and business requirements
- **Competitor Analysis:** Conducted competitor analysis by examining competitors' products and strategies, optimized ByteDance product, and provided feedback to the leadership team, resulting in a 15% increase in Net Promoter Score (NPS)

Ernst & Young Global Ltd.

Data Analyst Intern *Jun 2019 - Aug 2019*

- **Derivative Pricing:** Verified the FV of customer self-estimation with F-XSWAP, Black-Scholes Model etc. The PV of future cash flows to calculate the forward exchange rate curve by using the risk-free rate curve and currency swap given by clients
- **FX Rate Interpolation:** Implement the VBA function in Excel to calculate the interpolation of the long-term exchange rate, the exchange rate at the expiration date without risk and the actual value of the client's foreign exchange
- **Web Scraping**: Leveraged Python to automate data collection from Thomson Reuters Eikon, improve 72% results return time on analyzing large dataset (350k) compared to the manual calculation

SELECTED PROJECTS

Recommendation System *Jan 2023 - Apr 2023*

- **Data Collection:** Designed an AWS Lambda scraper, collecting 423 million Amazon reviews and metadata using multithreading and proxy rotation, achieving a high download rate of 12,000 data entries per second into Redshift
- **Deep Learning:** Enhanced the Two-Tower Model with TensorFlow by introducing Neural CF layers for embedding, weighted matrix, and bias vectors, resulting in a remarkable 94% model accuracy
- **Scoring:** Implemented Deep Learning Recommendation Model during the scoring stage to generate feature scores between users and products, effectively curating the top 100 scores to create personalized recommendations for end users

Airbnb Pricing Prediction *Sep 2022 - Dec 2022*

- **Data Preprocessing:** Filled null values, encoded categorical variables, normalized values for regression using Python
- **Model Tuning:** Utilized Learning-Curve and Model Evaluation Metric (RMSE) to check model performance and scalability (e.g Random Forest Regressor, Linear Regressor, XGBoost), used Grid Search and Randomized Search Cross Validation for Random Forest Regressor (Median RMSE Error<20)
- **Model Deployment:** Successfully containerized and deployed the web application on AWS via a streamlined CI/CD pipeline

SKILLS & TOOLS

- **Programming:** Python, SQL, R, HTML/CSS, JavaScript, PHP
- **Tools:** MySQL, PostgreSQL, MongoDB, AWS (S3, EC2, Lambda, EMR, Redshift, Glue, Athena, QuickSight), Tableau, PowerBI, PySpark, Databricks, Docker, Git, MS Office, Google Analytics, MATLAB, Latex