YUN (JESSICA) YAN

(574) 292-8774 • yyan5@alumni.nd.edu • yyyoment.github.io/yan-yun/ • www.linkedin.com/in/yan-yun/

EDUCATION

UNIVERSITY OF NOTRE DAME — Notre Dame, IN
Master of Science in Business Analytics (STEM designated), *Magna Cum Laude* — May 2020
GPA: 3.923/4.0; Merit-based $25,000 Fellowship Recipient (TOP 1%)

XI'AN JIAOTONG UNIVERSITY — Xi'an, China
Bachelor of Economics in Finance — Jul 2019
GPA: 3.6/4.0; Merit-based Scholarship Recipient (TOP 30%)

COPENHAGEN BUSINESS SCHOOL — Copenhagen, Denmark
International Summer University Program — Jun – Aug 2017

PROFESSIONAL EXPERIENCE

BLEND 360 — Columbia, MD
Data Science Associate *(Python, R, Snowflake/SQL, AWS)* — July 2020 – Present
- Target audiences for marketing campaigns by creating profiling analysis, and building and validating predictive models
- Develop machine learning models to improve client's pricing engine and support its customized pricing strategy to various household
- Convert the internal tool package from R to Python to automate the variable imputation, transformation and selection process

IT, ANALYTICS AND OPERATIONS DEPARTMENT, UNIVERSITY OF NOTRE DAME — Notre Dame, IN
Graduate Research Assistant, Machine Learning *(R, Python)* — Jan – May 2020
- Developed machine learning algorithms that are tailored to students' learning goals and are practical in solving real-life problems
- Documented codes, created graphics, texts and instructional slides, performed literature review for machine learning textbook editing

DIDI CHUXING (TOP 2 UNICORN COMPANY IN CHINA) — Beijing, China
Data Scientist Intern, International Business Technology Department *(Python, R, Hive-SQL, Tableau)* — Feb – Jul 2019
- Designed and drove A/B tests, interpreted experimental results and made strategic recommendations based on findings
- Extracted data using Hive-SQL and analyzed driver behavior patterns by metrics, explored and identified reasons behind drivers' long-term idle period, assisted in implementing customized operation strategies towards different driver segments
- Clustered riders' cancellation behaviors into different scenarios and developed Tableau dashboards to monitor cancellation rates
- Automated experimental analyses broken down by hours, order requests and supply-demand ratio; output adopted by the team

ACCENTURE — Xi'an, China
Business Intelligence Intern, QHgas ERP Project Team *(SAP Hana Studio, SAP Lumira Designer, Python)* — May – Dec 2018
- Identified key metrics and visualized the data and designed web interactions with JavaScript, CSS, and various components in SAP Lumira Designer, dashboards highly promoted and used by client's management board for monthly meeting and monitoring
- Built a predictive model and handled seasonality using Python to give a reference purchase amount for the client each month, saved 20% purchase expenses and storage costs for client's logistics department for the next 3 months

CENTER FOR EXPERIMENTAL ECONOMICS IN EDUCATION — Xi'an, China
Research Assistant, Rural Countryside Health Research Group *(R, Stata)* — Apr – Sep 2018
- Cleansed 38K rows of raw data collected from the questionnaire about Chinese village clinics with R and Stata
- Conducted regression modeling on how Chinese medicine education background affects prescription behaviors of doctor

DATA SCIENCE PROJECTS

OSCAR PREDICTION 2020 *(Python, R)*
- Built an XGBoost model to predict Oscar Best Picture 2020 based on movie awards data scraped from IMDb and Rotten Tomatoes, achieving an AUC of 0.77; investigated and plotted feature importance to identify key drivers of winning Oscar
- Performed the text sentiment analysis on the critic reviews of Parasite (Oscar Winner 2020) and predicted the Oscar winning likelihood with a Convolutional Neural Network model on review texts at an accuracy of 78%

SPOTIFY LIKE SONG ANALYSIS *(R, R Shiny)*
- Retrieved audio features of 55,557 songs via Spotify Web API; built a song popularity predictor by lasso and linear mixed model
- Applied Logistic Regression to predict whether a song will be liked based on its musical attributes; deployed it in a R Shiny App

SKILLSET & LEADERSHIP

Tools: Python, R, SQL, Hive, Hadoop, Spark, AWS, R Shiny, Tableau, VBA, SPSS, Stata, SAS, Excel, Google Analytics, HTML, CSS
Skills: Exploratory and Predictive Analytics (Logistic Regression, Random Forest, Boosting Tree), Clustering, Association Rule, Statistical Inference (Bayesian Data Analysis, Linear Mixed Model), Experiment Design, A/B Testing, Web Scraping
Leadership: Student Ambassador, University of Notre Dame (Sep 2019 – Present)