KATIE CAO

DATA SCIENCE & ANALYTICS

CONTACT

- ✉ Lcao18@fordham.edu
- 🏠 New York, NY
- 🔗 linkedin.com/in/katiecao
- 🐙 github.com/katiecao1
- 🌐 katiecao1.github.io

BACKGROUND

Experienced Data Scientist with 3+ years of building predictive analytics solutions to increase efficiency and revenue with a niche focus on behavioral science

EDUCATION

2019

FORDHAM UNIVERSITY [NEW YORK, NY]

Master of Science in Data Science

GPA: 3.8/4.0

2017

MERCYHURST UNIVERSITY [ERIE, PA]

Bachelor of Arts in Finance and Accounting, Double Major

GPA: 3.93/4.0

Summa Cum Laude

Graduated top 5% of the class

TECHNICAL SKILLS

- Data Analysis
- Statistics and Probability
- Data Visualization (Tableau, Looker, PowerBI)
- Database Management
- Python, R, SQL
- Pandas, Scikit-learn, NLTK
- Apache Airflow, BigQuery
- Google Cloud Platform, MongoDB, Google Analytics

AWARDS

- First Prize – 2018 Fordham Consulting Case Competition
- Honorable Mention – 2019 Fordham Data Crunch Basketball March Madness Bracket Prediction

EXPERIENCES

03/2020 – PRESENT   NEW YORK, NY

Data Analyst | automotiveMastermind (part of IHS Markit)

- Perform ad hoc analyses to resolve data issue inquiries from clients
- Automate the extract, transform, load (**ETL**) pipeline for routine bulk updates using SQL and Power Query, **saving 80% time** spent on the process
- Evaluate and monitor reports on databases through MongoDB, real-time analytics dashboards and on Google Cloud Platform environment
- Detect and analyze data errors, system bugs, and collaborate with product managers and software engineers to test new product features and bug fixes

07/2019 – 09/2019   NEW YORK, NY

Data Scientist | Zenon AI

- Facilitated the development of customer segmentation tool to reduce customer churning using **Artificial Neural Network** and **XGBoost** algorithms by using pandas to aggregate and clean 500,000+ rows of data for online trading platform client E*TRADE
- Streamlined client's internal paperwork process by **30%** by building a **Natural Language Processing** tool using Python regular expressions that extract contract terms from sentences

08/2018 – 07/2019   NEW YORK, NY

Data Scientist | Fordham Design Lab – Gabelli School of Business

- Wrote Python scripts to scrape text data from 2k+ research papers and F500 companies' website to perform text analytics and statistical modeling to predict machine learning application in major US health care companies
- Set up the ETL pipeline to clean, aggregate, and visualize 10GB of data into interactive dashboards in order to make recommendation for businesses

08/2017 – 08/2018   NEW YORK, NY

Senior Associate | Ideas42

- Analyzed **A/B Testing** results to validate project hypotheses based on behavioral economics framework
- Collaborated with behavioral scientists and project managers to ensure resource allocation and project timeline
- Streamlined financial reporting process by 50% by writing VBA macros to automate consolidation tasks for each project
- Received an early promotion for outstanding work performance

DATA SCIENCE & MACHINE LEARNING PROJECTS

College Basketball Bracket Prediction   03/2020-05/2020

- Engineered new features and applied **Random Forest, Logistic Regression** algorithms and **GridSearch** to predict 2019 NCAA Men's College Basketball Tournament bracket with 79% accuracy and 0.53 log loss

Employee Turnover Prediction – HR Analytics   06/2020-08/2020

- Performed **Exploratory Data Analysis** and built Random Forest and Gradient Boosting classifiers with **Cross Validation**, **tackling data imbalance using SMOTE** to predict employee attrition

Stock Portfolio Recommendation Using Sharpe Ratio   06/2020-08/2020

- Created a S&P 500 stock portfolio **with a 7% risk-adjusted return** by using **pandas** and **Beautiful Soup** to scrape data from Yahoo Finance, calculating Sharpe Ratio, storing, and querying it from a **PostgreSQL server**