Jack C. Yeung

734-929-8513 | jackyeung99@gmail.com | linkedin.com/in/jackcyeung | jackyeung99.github.io

EDUCATION

Indiana University - Bloomington **August 2024 - May 2026**
Master of Science in Data Science, Applied Data Science in Economic Analysis

Indiana University - Bloomington **August 2020 - July 2024**
Bachelor of Science in Informatics, Minor & Cognate in Psychology

Relevant Coursework: Machine Learning, Applied Machine Learning, Applied Algorithms, Engineering Cloud Computing, Performance Analytics, Network Science, Information Infrastructure, Information Representation

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, Shell, HTML5/CSS, Git
- **Frameworks:** Pandas, Network X, NumPy, SciKit Learn, Flask, SciPy, Scrapy, Selenium, BeautifulSoup, GeoPandas, Pytest, Matplotlib, Seaborn

EXPERIENCE

Center for Complex Networks and Systems Research(CNetS) **June 2023 - Present**
Research Assistant *Bloomington, IN*

- Extended an efficient iterative approach in the Bradley-Terry (BT) model to handle multi-body interactions (Plackett-Luce) using hypergraphs and an expanded likelihood function
- Enabling faster estimation of player strength scores in complex multi-entity competitions, such as Formula One races and or preference data
- Evaluated the model's predictive performance against existing graph-based ranking algorithms, demonstrating a 14x faster convergence on average and improved accuracy across real and synthetic datasets
- Co-authored a paper detailing the methodology and findings, currently in review at a peer-reviewed journal

PROJECTS

Machine Learning Algorithm On Real World Apartment Data **March 2024 - August 2024**

- Employed advanced web scraping techniques with Scrapy to gather real estate data across 16 cities, encompassing 23,000 apartment complexes and over 300,000 individual units
- Conducted exploratory data analysis and visualization using GeoPandas to identify neighborhood price trends and applied K-Means clustering for pattern discovery
- Developed predictive pricing models by tuning hyperparameters and comparing the performance of XGBoost (Extreme Gradient Boosted Trees) against a standard Neural Network
- Utilized TF-IDF(term frequency-inverse document frequency) for feature engineering of amenity text data, improving the models performance by reducing the error by 40%

Cloud Based Applications **August 2024 - December 2024**

- Designed and implemented a key-value store deployable on Google Cloud Platform (GCP) virtual machines, ensuring efficient storage and retrieval
- Recreated and optimized a research-backed matrix multiplication schema using Map-Reduce, leveraging GCP serverless functions for parallel computation, achieving an 8x speed improvement to the standard matrix multiplication
- Developed heuristics for efficient virtual machine allocation, minimizing resource wastage and optimizing cloud infrastructure utilization

Credit Card Fraud Detection Using Ensemble Methods **September 2023 - December 2023**

- Trained and tuned hyper-parameters for multiple classification models, including Support Vector Machine (SVM), Random Forest, and Logistic Regression, to detect fraudulent credit card transactions
- Enhanced predictive performance by implementing ensemble methods, leveraging the combinations of all classifiers to improve accuracy
- Achieved 98% accuracy in fraud detection using a soft voting ensemble on 14 anonymized features

Equipment Rental Interface **January 2024 - May 2024**

- Developed a full-stack equipment rental management system using Flask, HTML, and MySQL to streamline inventory tracking and business operations
- Designed a user-friendly front-end interface with Bootstrap, enhancing usability for employees and customers
- Integrated a Python-Flask backend with a MySQL database to enable real-time updates, efficient data storage, and seamless retrieval of rental information