Jack C. Yeung

jackyeung99@gmail.com | linkedin.com/in/jackcyeung | jackyeung99.github.io

EDUCATION

Indiana University - Bloomington **May 2026**
Master of Science in Data Science, Applied Data Science – Economic Data Analytics *GPA 4.0/4.0*

Indiana University - Bloomington **July 2024**
Bachelor of Science in Informatics

Relevant Coursework: Machine Learning, Applied Machine Learning, Applied Algorithms, Game Theory for Business, Engineering Cloud Computing, Exploratory Data Analysis, Performance Analytics, Network Science

TECHNICAL SKILLS

- **Languages:** Python, SQL, Shell, HTML5/CSS
- **Tools:** Power BI, Git, Shell
- **Libraries:** Scikit-learn, PyTorch, PySpark, Pandas, NumPy, NetworkX, Scrapy, Selenium, Matplotlib, Seaborn, Pytest

EXPERIENCE

CarePlus New Jersey **June 2025 - August 2025**
Data Analyst Intern *Paramus, NJ*

- Developed a Power BI compliance dashboard reconciling reported vs actual telehealth time across two data sources, improving match rates by **10%** and strengthening negotiation readiness with insurers
- Engineered a scalable PySpark pipeline to process **2B+** insurance rate records, integrating billing codes, geolocation, and taxonomy data to isolate accurate in-network rates
- Delivered a market analysis of two major insurers against state averages and competitors, identifying 4 underpaid billing codes and presenting a **26.5%** potential annual revenue uplift to senior leadership

Center for Complex Networks and Systems Research(CNetS) **June 2023 - February 2025**
Research Assistant *Bloomington, IN*

- Developed and optimized probabilistic ranking models to evaluate competitive dynamics (Formula One, elections), improving predictive performance and achieving **14x** faster convergence
- Collaborated with a team in weekly meetings to design and refine experiments, co-authoring a research paper detailing the methodology and findings to a peer-reviewed publication in Physical Review E

PUBLICATIONS

1 Jack Yeung, Daniel Kaiser, and Filippo Radicchi. Efficient inference of rankings from multibody comparisons. *Phys. Rev. E*, 112:014305.

PROJECTS

End-to-End Machine Learning On Real World Apartment Data

- Automated a data retrieval and storage pipeline using Scrapy for advanced web scraping to gather, validate, and store real estate data for over **300,00** units into a SQLite relational database
- Conducted exploratory data analysis (EDA) with GeoPandas for visualizations and spatial mappings of different neighborhoods, providing insights into market analysis and investment decision-making
- Trained and Tuned an XGBoost(Extreme Gradient Boosted Trees) predictive pricing model, incorporating natural language processioning for feature engineering of amenity text data, reducing model error by **55%**, to aide in accurate identification of under-priced apartments

Cloud Based Applications

- Designed and implemented a key-value store deployable on Google Cloud Platform (GCP) virtual machines, applying heuristics for efficient VM allocation to minimize resource wastage and optimize cloud infrastructure utilization.
- 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

Credit Card Fraud Detection Using Ensemble Methods

- 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 in fraud detection by implementing ensemble methods, achieving **98%** accuracy through a soft voting classifier on anonymized features