Yang Wu

Chicago, IL | Princeton, NJ

yangwu@gmail.com | kenwuyang.com/en/post/ | github.com/YangWu1227 | linkedin.com/in/yangwu1227

Personal Profile

I am a graduate of the University of Chicago Master's in the Data Science program. I specialize in statistical learning and machine learning and have four years of expertise in building data science products. I am most impactful in machine learning, data science, Python or R development positions where I can solve challenging problems and contribute to innovative projects, driving results for my stakeholders.

Education

University of Chicago

Chicago, IL

M.S. Applied Data Science — *Sept 2022 - May 2024*

- **Cumulative GPA:** 4.0

Kenyon College

Gambier, OH

B.A. in Statistics and Economics — *Sept 2017 - May 2021*

- **Cumulative GPA:** 3.76 *Magna Cum Laude* and Merit Lists (2018, 2019, 2021)

Work Experience

Home Partners of America

Chicago, IL USA

Data Science Associate / Intern — *June 2023 - Present*

- **Analyze historical delinquency patterns to identify key trends and indicators**, which informs the development of our feature store generating features from transaction data. Integrate these features into an automated detection and monitoring system, proactively alerting operations of potential delinquent accounts and safeguarding our net operating income and overall financial stability.
- **Collaborate with key team members to develop a resident scorecard for segmenting our resident base**, analyzing resident behaviors, communications, payment patterns, issue resolution efficiency, and lease renewals. This initiative facilitates targeted operational improvements, enhancing the resident experience; success is measured by metrics such as reduced delinquency rates, improved rent collection rates, higher renewal rates, and minimized turnover costs (i.e. the expenses related to finding and preparing units for new tenants).
- **Build Python-based applications to surface and deliver analysis results, engineered data science products, and automated detection systems**, freeing up at least 10 to 15 hours per week for the operations team, representing a $20 - 40\%$ reduction in manual efforts and allowing them to focus on high-impact tasks.
- **Developed computer vision models** trained on property images for classification and object detection tasks, enabling automated identification of property features (e.g., pool, yards, flooring materials, countertops), significantly improving property valuation accuracy, directly influencing investment and rent pricing decisions for the leasing business.
- **Collaborated with a senior software developer and data scientist** to co-lead an internal initiative to automate data science product delivery. Leveraged AWS EventBridge to schedule triggers for AWS Lambda, which executed Python-based applications on AWS ECS Fargate. This setup seamlessly integrated with AWS services like Athena, Redshift, and S3, establishing an efficient and cost-effective deployment pattern. The automation allowed our team to concentrate on model and analysis, producing Python scripts that delivered results as reports, visualizations, and spreadsheets for internal stakeholders.
- **Technical Skills:** Deep learning with Keras, AWS SageMaker, AWS Lambda, AWS EventBridge, AWS ECS Fargate, PyTest, Docker, CI/CD with AWS GitHub Actions, Python, SQL

Citizen Data

Washington DC, USA

Jr. Data Scientist / Engineering Intern — *Oct 2021 - July 2022*

- **Built, tested, and deployed end-to-end models**, including clustering and anomaly detection for voter segmentation, social network analysis for precise microtargeting, geotargeting, and lookalike targeting, and text classification for analyzing survey free responses.
- Proposed, built, and deployed our first interactive dashboard, which became an official product line, providing a user-friendly front-end for model outputs, including visualizations, tables, and insights. **Secured our first subscription-based fee, marking a significant revenue milestone and elevating our business model potential.**
- **Developed a three-module Poetry-based Python package** to support in-house data processing and feature engineering tasks. Included custom transformers and pipeline utilities leveraging NLTK and Scikit-learn (for survey text data) and pandas (for structured data). Automated cloud storage and data warehouse operations using the AWS SDK, enhancing the efficiency and accuracy of our data sampling and ingestion processes for modeling tasks from our voter database in Redshift.
- **Developed an R package** to facilitate fast proprietary report tables generation for our social scientists.
- **Built an internal self-service UI tool (Python Dash)**, empowering non-technical team members to produce ad-hoc query results from Redshift, eliminating the need for an expensive SQL builder subscription and delivering substantial cost savings crucial for our small data science consulting business.
- **Technical Skills:** PyTest, SQL, Python, R development, AWS SDK, Model training, Dashboards development to operationalize and visualize model outputs for end-users

References available upon request