STANLEY OU

301-273-8205 | stanleyo@andrew.cmu.edu | Portfolio | LinkedIn | GitHub

EDUCATION

Carnegie Mellon University (CMU) **Pittsburgh, PA**

B.S. in Statistics and Machine Learning, Additional Major in Computer Science (**GPA:** 3.83/4.0) Aug 2022 – May 2026

- **Teaching Assistant**: Probability and Random Processes, Dangerous Ideas in Science and Society
- **Coursework***: Machine Learning, Data Science, Database Systems, Statistical Visualization, Statistical Computing, Probability & Statistical Inference, Regression, Software Engineering, Artificial Intelligence, Data Structures*
- **Extracurriculars:** Statistics and Data Science Mentor, Data Science Club

SKILLS

- Programming Languages: Python, SQL (PostgreSQL, MySQL), R, Scala, C, C++, Java, HTML, CSS, JavaScript
- Technologies: Pandas, NumPy, Sklearn, TensorFlow, PyTorch, Flask, Streamlit, Docker, GitHub, Jupyter, BI, GCP, AWS, Spark
- Proficiencies: Data visualization, Predictive models, LLMs, OOP, Web development, Unit testing, Microsoft Office

PROFESSIONAL EXPERIENCE

Software Engineering Institute | CMU, *Data Science Intern* **Pittsburgh, PA**

- Support the CERT Cyber Risk and Resilience Directorate by collecting, interpreting, and visualizing data from 1000+ insider risk court cases to publish an annual report on trends in insider risk. Sep 2024 - Present
- Enhanced data collection efficiency by 25% through implementing NLP methods, including regular expressions and Named Entity Recognition (NER), to extract over 10 relevant fields from diverse court case documents.
- Develop a publicly available insider risk dashboard web application using Django and Plotly/Dash.

CMU PeeX Lab, *Research Assistant* **Pittsburgh, PA**

- Monitor 1000+ participants in a large-scale field experiment on adblocking's effect on consumer behavior. Aug 2024 - Present
- Debug, test, and develop the study's Chrome browser tracking software using Python, JavaScript, and Scala.
- Implemented data pipelines to extract product purchases from servers into Qualtrics and collect local storage data.
- Derive research insights from thousands of terabytes of participant browsing big data using SQL and Spark.

Ploomber, *Software Engineer Intern* **Remote**

- Used Python, JavaScript, and CI/CD to drive the full-stack development of cloud data products. Jul 2024 – Sep 2024
- Developed and deployed 5 demo applications using Python and data tools such as Streamlit, PostgreSQL, and R-Shiny to guide thousands of current and prospective Ploomber Cloud users.
- Collaborated with a team of engineers using the Agile methodology to design and implement scalable cloud infrastructures, utilizing frameworks such as React, and deploying on AWS.

PA 211, *Data Science Intern* **Lemoyne, PA**

- Supported PA 211's statewide initiative in assisting thousands of people in need of shelter, food, and utilities by developing BI dashboards in DOMO to guide resource allocations & business decisions. Feb 2024 – Jul 2024
- Created ETL pipelines using SQL to integrate data from external software and web-scraped data.
- Developed a multinomial logistic regression model with 75% accuracy in predicting needs.

UMBC Bennett Labs, *Research Intern* **Baltimore, MD**

- Drafted and first-authored a published academic manuscript titled "A Density Functional Theory (DFT) Investigation of Sulfur-Based Adsorbate Interactions on Alumina and Calcite Surfaces". Jun 2021 – Jan 2022
- Ran DFT calculations on a high-performance computing facility, presented the study to a group of 30+ students, and received a first-place junior award from the Washington Academy of Sciences.

RELEVANT PROJECTS

AI Sustainability in Finance Datathon

- Collaborated with a team to research, derive insights, and pitch solutions to a board of judges, culminating in a 3rd place award.
- Optimized a linear regression model to estimate carbon emissions from training AI models based on model features.

NBA Career Outcomes Predictor

- Trained a Random Forest Classifier to predict NBA players' career outcomes from first 4 seasons, achieving 75% test accuracy.
- Performed data cleaning and engineering techniques, including exploratory data analysis, merging datasets, imputing missing values, aggregating player statistics to prepare data for modeling, and conducted hyperparameter optimization.