256.585.7439
ychen.adastra@gmail.com

Yu Chen

New York, NY

ychenxastra.github.io
LinkedIn: yu-chen-astra

EDUCATION

Doctor of Philosophy, Space Science, The University of Alabama in Huntsville **Jun 2018 – Dec 2020**
Master of Science, Space Science, The University of Alabama in Huntsville **Aug 2015 – May 2018**
Bachelor of Science, Atmospheric Science, Nanjing Univ. of Info. Sci. & Tech. **Sep 2011 – Jun 2015**

EXPERIENCE

Center for Space Plasma and Aeronomic Research & Department of Space Science **Huntsville, AL**
Research Scientist Jan 2023 – present
Postdoctoral Researcher Jan 2021 – Dec 2022
Graduate Research Assistant Aug 2016 – Dec 2020

- Processed and **analyzed** over **50GB** of high-resolution raw data using **Python** with specially designed algorithms to identify typical events, improving efficiency and accuracy by **40%**.
- Established a local **database** over **140,000** entries and managed it using **PostgreSQL** and **PySpark**, enabling efficient querying and data transformation and reducing manual data handling.
- Evaluated stakeholders needs and extracted 60+ characteristics per entry, applied time-series and **statistical analyses** to reveal patterns and relationships, deriving insights for strategic decision-making processes.
- Employed **MATLAB**, **Matplotlib**, and **Tableau** for **data visualization**, aiding in understanding metrics and trends.
- Processed and cleaned datasets for **machine learning** applications, trained with different models, fine-tuned inputs, enhancing predictive accuracy by 12%.
- Presented complex data findings in a clear and accessible manner, culminating in **20 articles** in peer-reviewed journals and **18 conference presentations**, ensuring clear **communication** with non-technical stakeholders.
- **Led** two national research grants as Principal Investigator and **collaborated** on multi-million-dollar projects, **instructed** students in an NSF-funded program, contributing to research and educational development.

PROJECTS

PyGS: an open-source Python package for event analysis Python, Numpy, Pandas, Scipy, Matplotlib

- Optimized analyzing models with advanced analytics and theories, refactored and merged **30+** scripts into a **Python library**, resulting in a **100x** performance increase and reducing manual workload from hours to minutes.
- Implemented automation for information extraction and characteristic **visualization** for large and complex datasets, establishing a basis for further analysis and achieving efficient data analysis workflows.
- Released a Python package (**6,500+** lines) on **GitHub**, ensuring adherence to Heliophysics community standards, provided detailed documentation and tutorials, addressed **user concerns**, and promoted its use in the community.

Online & local database of magnetic flux rope (typical events) Python, Matlab, Excel, PostgreSQL, HTML

- Employed **Python** to process large datasets spanning over **50 years** to pinpoint typical events. Developed sophisticated filtering algorithms to process over **200,000** candidates, contributing to the establishment of a robust local database.
- Administered and maintained the **online database** using **HTML** and **Excel** to facilitate efficient and convenient user inquiries, reducing manual event identification time by **80%** and improving data product utility.
- Conducted in-depth analyses including **linear regression** to accurately characterize product properties; delivered parameters in multiple formats, empowering users to extract relevant **insights** tailored to their specific needs.

Analyses of Football Matches and Predicting Results Using Machine Learning Python, Pandas, Scikit-learn, Seaborn

- Analyzed **20+** years of English Premier League data (**11,000+** rows), focusing on Manchester United, and visualized statistics to identify key and controversial factors impacting team performance.
- Developed **machine learning pipelines** with 6 models including **Naive Bayes**, **Decision Tree**, and **Random Forest** to predict match outcomes, selecting the best model based on performance metrics and applicability.
- Enhanced model accuracy using techniques like rolling averages, **feature selection**, and **PCA**, resulting in a 12% improvement; integrated findings with real-world scenarios to explore causes of suboptimal performance.

TECHNICAL SKILLS

Programming Python (Numpy, Pandas, Scipy, Matplotlib, Scikit-Learn, Seaborn, PySpark), MATLAB, C/C++
Tools & Software SQL (MySQL & PostgreSQL), Tableau, Latex, Git, Jupyter Notebook, Excel, Powerpoint, Keynote
Platforms & OS Linux, MacOS, GitHub, AWS