256.585.7439
ychen.adastra@gmail.com

Yu Chen

Data Scientist | New York, NY

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

SUMMARY

Eight years of experience (5 yrs as a quantitative PhD) in large-scale data analysis including processing, visualization, algorithm design, and deriving data-driven insights; proficient in Python and SQL; familiar with statistical models and machine learning; strong communication and collaborative skills, problem solver with self-motivation and curiosity.

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

EXPERIENCE

Center for Space Plasma and Aeronomic Research **Huntsville, AL**

Research Scientist Jan 2023 – present

Postdoctoral Researcher Jan 2021 – Dec 2022

- Processed and analyzed over 50GB of data using Python, leveraged optimized algorithms to enhance the identification of typical events, resulting in a 40% improvement in efficiency and accuracy.
- Established a local database with over 140,000 entries compiled from separate CSV files and managed it using PostgreSQL and PySpark, enabling efficient querying and data transformation and reducing manual workloads.
- 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, assessed the performance, fine-tuned with feature selection, PCA, etc., enhancing predictive accuracy by 12%.
- Led two national grants, collaborated with cross-functional teams on multi-million-dollar projects, and summarized complex findings in 20 articles and 18 top conferences, ensuring clear communication with non-technical stakeholders.

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 the 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 (50+ years) to identify typical events; developed sophisticated filtering algorithms for over 200,000 candidates, reducing visual identification time by 50% and enhancing data product utility.
- Administered and maintained the online database using HTML and Excel, ensuring efficient and convenient user inquiries.
- 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); developed machine learning pipelines with models like Naive Bayes, Decision Tree, and Random Forest to predict match outcomes.
- Assessed model performance and applicability, achieving a 12% improvement in model accuracy through techniques like rolling averages, feature selection, and PCA.
- Integrated findings with real-world scenarios to explore causes of suboptimal performance.

EDUCATION

Doctor of Philosophy, Space Science (Physics), The University of Alabama in Huntsville **Jun 2018 – Dec 2020**

Master of Science, Space Science (Physics), 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**