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 and machine learning models; strong communication and collaborative skills, problem solver with self-motivation and curiosity.

TECHNICAL SKILLS

Programming	Python (Numpy, Pandas, Scipy, Matplotlib, Scikit-Learn, Seaborn, PySpark, NLTK), 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

- Conducted analysis of over 50 years of data using Python, leveraged optimized algorithms to enhance the identification of typical events, resulting in a 50%+ improvement in efficiency and accuracy compared to traditional visual approach.
- Developed an open-source Python package for event studies and automated information extraction as well as characteristic visualization, reducing manual input and user workload from months to minutes.
- Established and managed a database with 140k+ entries using PostgreSQL and PySpark, applied time-series and statistical analyses to reveal their patterns and relationships, deriving insights for strategic decision-making processes.
- Employed MATLAB, Matplotlib, and Tableau for data visualization, aiding in understanding metrics and trends.
- Performed data processing and cleansing for 120k+ data points, trained machine learning models, and applied feature selection, PCA, etc., to enhance predictive accuracy.
- Led 2 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.
- Coordinated the CUWiP, facilitating communication between invited attendees and the conference committee.

PROJECTS

Database of Typical Events & Open-source Package Python, MATLAB, Excel, PostgreSQL, HTML, Matplotlib

- Developed algorithms to catalog over 200k events in a local database, conducted comprehensive analyses to characterize product properties, and extracted 60+ characteristics per entry, enhancing information sharing based on user needs.
- Administered the online database using HTML and delivered parameters in multiple formats including linear regression, empowering users to extract relevant insights tailored to their specific needs.
- Refactored over 30 Python scripts and released a package containing 6,500+ lines of code on GitHub, provided detailed documentation and tutorials, addressed user concerns, and promoted its usage within the community.

Predicting Results of Soccer Matches Using Machine Learning Python, Pandas, Scikit-learn, Seaborn

- Processed and cleansed data from over 20 years of the English Premier League (11,000+ rows) to analyze the impact of controversial factors on match outcomes, extracting valuable insights through data science techniques.
- Employed machine learning 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.

Content-based Recommendation System of IMDb Top 250 Movies Python, Scikit-learn, NLTK, RAKE, Seaborn

- Developed a content-based movie recommend system using Python and NLP techniques.
- Employed Python to scrape data of IMDb Top 250 movies and analyzed relations between rating scores and multiple factors.
- Processed the dataset with tools like TF-IDF Vectorizer and extracted key phrases from plot summaries using RAKE, identifying thematic similarities among movies.

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**