Anderson Banihirwe

https://andersy005.github.io/

Email : axbanihirwe@gmail.com
Github : https://github.com/andersy005
Kaggle: https://www.kaggle.com/andersy005

Education

- **University of Arkansas at Little Rock** — Little Rock, AR
  Bachelor of Science in (Computer) Systems Engineering; GPA: 3.85/4.00 — *Aug 2014 – May 2018*

Experience

- **First Orion** — Little Rock, AR
  Data Scientist — *Dec 2017 – Present*
  - **Data Processing**: Identified patterns and characteristics within First Orion's data warehouses using Dask, Apache Spark, Pandas.
  - **Machine Learning**: Designed and built scoring, predictive machine-learning models, and feature extraction systems with Scikit-learn using First Orion' s proprietary data assets.
- **National Center for Atmospheric Research** — Boulder, CO
  Research Intern — *May 2017 – Aug 2017*
  - **Installation**: Installed Apache Spark v2.2 on both Cheyenne and Yellowstone Supercomputers.
  - **Schedulers**: Cleaned/fixed Spark launch bash scripts that work with the LSF/PBS schedulers.
  - **spark-xarray**: Wrote spark-xarray, a python package that integrates PySpark and xarray for Climate Data Analysis.
  - **Jupyter notebooks contribution**: Contributed Jupyter notebooks and scripts using Apache Spark to NCAR's Coupled Model Intercomparison Project (CMIP) Analysis Platform.
  - **Documentation**: Documented research work at https://ncar.github.io/PySpark4Climate/

Projects

- **spark-xarray**: Open source python library built on top of PySpark - Spark Python API and xarray for climate data analysis. https://github.com/andersy005/spark-xarray
- **deepclimate (WIP)**: A Python library that aims to provide an open-source toolchain for deep-learning use in Atmospheric and Oceanic Sciences. https://github.com/deepclimate/deepclimate
- **climate-learn (WIP)**: Python library that aims to provide machine learning routines for analyzing atmospheric and oceanic data using xarray, dask, numpy, scipy, scikit-learn, matplotlib. https://github.com/deepclimate/climate-learn
- **Advanced Lane Lines Detection**: A pipeline that uses OpenCV to detect lane lines on the road on a series of individual frames and/or a video stream. https://youtu.be/3NnTZ9NR03k

Programming Skills

Languages	Python, SQL, C++, C, Scala
Frameworks/Libraries	Apache Spark, Numpy, Dask, Pandas, Scikit-learn, IPython, Keras
Toolchain	UNIX, Git, LaTeX
Technologies	AWS, GCE

SELECTED PRESENTATIONS

- **PySpark for "Big" Atmospheric Data Analysis**: American Meteorological Society (AMS) 2018 Conference. Austin, TX. Jan 2018 Recorded Presentation
- **PySpark for "Big" Atmospheric and Oceanic Data Analysis**: National Center for Atmospheric Research. Boulder, CO. Aug 2017 Recorded Presentation