Melanie McCord

Williamsburg, VA 23185 | mpmccord99@gmail.com | https://github.com/mpmccord

EDUCATION

William & Mary, Williamsburg, VA
Master of Science, Computer Science, GPA: 3.76 — January 2026
Relevant Coursework: Data Visualization, Deep Learning, Deep Transfer Learning

New College of Florida, Sarasota, FL
Bachelor of Arts, Computer Science, Minor: Statistics — May 2023
Relevant Coursework: Categorical Data Analysis, Machine Learning, Time Series Analysis

TECHNICAL SKILLS

Languages: Python, R, Java, JavaScript
Machine Learning Libraries: PyTorch, HuggingFace Transformers, Scikit-Learn, OpenCV, NLTK, PySpark
Tools: Angular, React, Firebase, HTML, CSS, XML, Pandas, NumPy, Matplotlib, Seaborn, ggplot2, AWS, D3.js
Operating Systems: Windows, Mac, Unix, Linux

RELEVANT EXPERIENCE

Research Intern, NicheVision Forensics, Akron, OH, June 2023 – July 2023
- Evaluated decision tree, logistic regression, and random forest models using Scikit-learn to improve the prediction of the number of contributors from DNA samples.
- Visualized the results in Matplotlib and Seaborn to facilitate understanding of the predictions.
- Achieved > 90% accuracy in the testing set and analyzed the features of the model.

Research Intern, Pacific Northwest National Laboratory, Richland, WA, June 2022 – August 2022
- Implemented part-of-speech tagging using Spacy and object detection using OpenCV and TensorFlow.
- Combined the language and image models to create better datasets by highlighting the portions relevant to annotators.
- Used an R-CNN model and combined it with part-of-speech tagging to improve the dataset quality on AWS.

Research Intern, Penn State University, State College, PA, May 2021 – July 2021
- Analyzed text data using Scikit-Learn and NLTK to predict COVID-19-related misinformation from tweets and news.
- Created a website using a Flask backend to extract the data from a website and detect misinformation from the site.
- Used a bag-of-words tokenization, logistic regression, and decision tree models to detect misinformation.

PROJECTS

On the use of heuristics for generating non-contaminated datasets
- Modified the structure of code from the CodeXGlue using Python AST syntax.
- Analyzed the similarity to original code and impact on performance of CodeT5 from the Transformers library.

Model Compression for Software Engineering Tasks
- Used knowledge distillation to reduce carbon emissions from fine-tuning CodeT5 small on a code generation task.
- Used PyTorch, Intel Neural Compressor library and the Transformers Python library to do this.

Fanfiction vs Classic Literature
- Set timelines, stretch goals and determined methods for a group of 5 to analyze and mine text data using PySpark.
- Ran an LDA analysis in PySpark of the text with different parameters, e.g. without nouns, paragraph vs work, etc.

Visualizing Crime Data in Chicago
- Aggregated Chicago crime data in PySpark to analyze the number of crimes by month and types of crimes.
- Built a website with React for a time series analysis using the preprocessed dataset and visualized the data in D3.js.

Analyzing Chicago Unemployment Data
- Performed a time series analysis in R and ggplot2 and analyzed unemployment data by month in Chicago from the US census.
- Modeled the data using ARIMA and ETS statistical models and analyzed the limitations given the changes from the pandemic.

PUBLICATIONS & PRESENTATIONS

McCord, M., & Hamid, F. (2023). Medical Relevancy of Cancer-Related Tweets and Its Relation to Misinformation. The International FLAIRS Conference Proceedings, 36. https://doi.org/10.32473/flairs.36.133364