Joseph Cappadona

(516) 220-2837 | josephcappadona27@gmail.com | GitHub: josephcappadona | https://jcappadona.com

Education

University of Pennsylvania, School of Engineering and Applied Science *Aug. 2014 – May 2018*
BSE in *Computer Science*, Minor in *Mathematics*
GPA: 3.35

Experience

XaiPient, Machine Learning Engineer *Jul. 2019 – Jun. 2020*
- Implemented and maintained library for Explainable Artificial Intelligence (XAI), wrote corresponding documentation
- Implemented algorithms encompassing attribution methods, adversarial examples, rule-based explanations, counterfactual explanations, and fairness/bias analysis
- Applied XAI library to finance and healthcare datasets, designed visualizations, and presented them to potential clients
- Attended NeurIPS 2019 on behalf of the company in order to network and learn about new research in explainability

University of Pennsylvania, Instructional Technologist, Artificial Intelligence *Jun. – Dec. 2019*
- Built a Python interface for controlling Sphero droids
- Built an iOS application (Swift) to identify, connect to, and extract information from Sphero droids
- Built a GUI using PyGame to demo controlling droids in a virtual environment
- Designed and implemented an assignment in which students use graph algorithms to navigate droids through a maze

Research Projects

Kids Britannica Dataset
- Used BeautifulSoup to scrape kids.britannica.com for three tiers of multimodal articles (125M tokens over 130k articles)
- Used spaCy on article texts to generate sentence statistics such as average parse tree height, average number of entities, and average number of noun phrases

Few-Shot Learning Using Classical Computer Vision
- Extracts image descriptors (KAZE, ORB) via OpenCV and clusters them using K-Means to create a BOVW model
- Aggregates spatial pyramid histograms for images over the generated BOVW vocabulary
- Compares different classification models, feature selectors, and data transformers via k-fold cross validation

Aggregating Insights from Amazon Product Reviews
- Uses word2vec to create embeddings for product features; embeddings are then clustered to group synonymous features
- Generates sentiment scores for an initial set of positive and negative opinion words using VADER sentiment analysis
- Bootstraps domain-specific opinion words based on part-of-speech and syntactic dependency information
- For a given product category, sentiment scores are computed for product feature clusters and displayed in a web application

Information Extraction Using DBSCAN and Layout Analysis
- Extracts text from documents using Tesseract OCR; identifies clusters of text via density-based spatial clustering
- Generates document features based on text properties, spatial properties, and layout properties; pruned via KL divergence
- Model performance is evaluated on the Ghega dataset; full procedure and results are documented in a technical report

Software Projects

Chess Population Analytics (Python, JavaScript)
- A CLI and web application for analyzing chess population data, built with React, Flask, and Seaborn

ML Model Playground (Python, JavaScript)
- A web application for querying machine learning models, built with React and Flask

Machine Learning Notebooks (Python)
- A repository of personally compiled machine learning notebooks in vision, language, RL, and more

Poker Tools (Python, JavaScript)
- A web application for poker study and hand analysis, built with React, Flask, and D3

Teaching

University of Pennsylvania
Crowdsourcing & Human Computation, Teaching Assistant *Jan. 2021 – Present*
Artificial Intelligence, Teaching Assistant *Aug. 2020 – Dec. 2020*
Python Programming, Teaching Assistant *Aug. – Dec. 2015*

Skills/Interests

Languages: Python, JavaScript, Java, Swift/Objective C (iOS), C, Bash, HTML/CSS, Assembly, LaTeX

Libraries:

Scientific:	PyTorch, Tensorflow, Hugging Face, NumPy, Pandas, scikit-learn, NLTK, Spacy, CoreNLP, OpenCV
Visualization:	Matplotlib, Seaborn, Plotly, D3
Scraping:	BeautifulSoup, Scrapy, Selenium, Requests
Web Dev:	Flask, React, REST
Dev Ops:	Vim, Git, AWS, Google Cloud, Docker

Hobbies: Chess, Poker, SSBM, Data Hoarding, Math, Psychology, Philosophy, Meditation, Photography, Hiking