

EXPERIENCE

Data Scientist/Machine Learning Engineer @ Metis (April 2021 to July 2021)

- A 14-week project-based bootcamp covering data analysis, linear regression, machine learning classification, NLP, unsupervised learning, deep learning, and data engineering

Noteworthy projects:

- **Non-fungible token artwork generator and recommendation engine:**
A Streamlit web application that contains an NFT artwork generator that was built from scratch by developing an autoencoder, and an NFT artwork recommendation engine. The recommendation engine uses euclidean distances to measure the similarity between the input artwork and a curated dataset of scraped NFT artworks
- **Convolutional neural network to classify impressionist paintings:**
A CNN using Keras that identifies artists based on their paintings. The model categorizes Impressionist paintings into ten different classes (each class being the artist responsible for the painting). The final model runs with ~67% accuracy, and a validation loss of about 1.0
- **Netflix recommendation engine using unsupervised NLP:**
A content-based Netflix recommendation system using NLP. The recommendations were calculated using the cosine similarities of the text data (movie/tv show descriptions, actors, directors)
- **Classification model to identify poisonous mushrooms:**
A logistic regression model that identifies features of poisonous mushrooms. Found that using only a few features, one could confidently identify poisonous mushrooms (with 99.8% accuracy and perfect precision + recall)

Engineering Linguist @ LinkedIn (April 2020 to April 2021 – 6 month contract extended to 12 months)

- Evaluated output of machine learning classifiers and internal tools
- Performed research, annotated data in various languages (e.g. Spanish, Portuguese, French, German), analyzed data, and calculated evaluation metrics using Google Sheets
- Created instructions for outsourced evaluations and revised existing guidelines
- Identified systematic relationships between concepts like job titles, skills, certificates, etc.
- **Accomplishments:** *Helped an experimental member skill standardizer discover 216 million new member skill mappings and improved coverage of relevant skills across LI ecosystem, *Developed a golden dataset for relationships between skills based on dependency, *Developed a golden dataset for identifying tools (e.g. software, machinery) in member profile text

Data Linguist @ Appen (Jan. 2019 to April 2020)

- Linguistic Annotator for both both the NLU team and the NLP team
- Created dialogues for a Facebook AI assistant over a variety of domains, including weather, holidays, timers
- Used semantic annotation across large volumes of data to label utterances for the AI assistant in domains related to shopping (clothing, furniture, etc)
- Evaluated AI-generated dialogues for clarity and naturalness
- Annotated conversational data for co-reference
- Identified semantic elements in utterances (both user and assistant), including context and intent
- Collaborated with team of linguists to improve annotation strategies
- Adapted to regularly changing conventions for each domain of data

SKILLS

Data Science tools: Python, SQL, Scikit-learn, Keras, NumPy, Pandas, NLTK, SpaCy, Matplotlib, Plotly, Seaborn, Tableau, Google Sheets
Algorithms: Deep learning (neural networks, embeddings, transfer learning), Linear regression, Classification (KNN, logistic regression, Naive Bayes, decision trees, random forest, gradient boosting), NLP (K-means, DBSCAN, Mean Shift, Hierarchical Agglomerative Clustering, LDA, NMF)

Linguistics & Computational

Linguistics: Linguistic annotation, NLG, NLP, conversational structures (pragmatics, prosody, syntax, semantics, morphology, phonology)

EDUCATION

MS, Computational Linguistics,

3.9/4.0 GPA (University of Washington)

Spoken Dialogue Systems, Shallow & Deep Processing for NLP, Advanced Statistical Methods in NLP, & more

BA, Linguistics (University of Colorado Boulder)

Data Science + Machine Learning

Certification (Metis Bootcamp)

Portfolio: github.com/ju-mc
