

WORK EXPERIENCE

Data Scientist (Independent contractor), Shamaym, 2021-22

- Developed NLP topic modelling to classify debriefs, and achieved precision 95% and recall 50%.
- Created NLP text generation to autocomplete, summarize, and auto-suggest using OpenAI.
- Built Python flask app modules with Linux server for deployment into production.

Data Scientist (Independent contractor), Streamline Verify, 2021-22

- Created NLP record linkage algorithm with scikit-learn to deduplicate professional records of medical professionals, and achieved >95% completeness and homogeneity.
- Implemented resampling procedure using the bootstrap method and k-fold cross validation in order to optimally leverage a limited data set.
- Engaged in feature engineering and feature enrichment using Pandas and geocoder to enhance the business value of each insight to be gleaned from my data.

INTERNSHIPS

Data Scientist (Internship), Sefaria, 2020

- Created NLP multi-label classifier using scikit-learn to assign topic labels to text bilingual passages (ancient Hebrew and English), and achieved >95% precision, 36% recall. Designed algorithm specifically for high precision.
- Implemented hierarchical classification using a knowledge graph in order to enrich the insights gleaned from each parent topic and its children.
- Developed an end-to-end NLP pipeline using Pandas, Scikit-Learn, Scikit-MultiLearn, Matplotlib, Seaborn, and Kubernetes in order to automatically provide the potential user with desired topic labels.

Python Developer (Independent contractor), Smrt, 2019

- Created NLP entity extractor using rule-based logic for extraction of key entities from PDF office lease contracts, which achieved accuracy of >90% for several entities and boosted precision and recall by 20%.
- Developed interactive confusion matrix using mplcursors and to evaluate statistics of predictions.
- Created the end-to-end AI portion of the product's web app using Pandas, Scikit-learn, SpaCy, Amazon Web Services, Matplotlib, and Seaborn to enable fine-resolution entity identification.

PERSONAL PROJECTS

BERT Classification: Created modules that prepare data to fine tune BERT for NLP multiclass classification and named entity recognition (NER).

Neural Network: Constructed from scratch without third party libraries. Displays superiority on non-linear dataset.

Quantum Machine Learning: Coding and analysis of loss function for quantum circuits on iris dataset

classification.

Screen Capture: Surveillance system to snap shots of your computer screen in real time, and send them to WhatsApp.

Story Illustrator: Based on the user's input text, this package produces a slideshow. Complete with appropriate images and captions. Published on PyPI available for installation.

EDUCATION

Honors BA Math and Physics, Amherst College, 2004-08

- [Quantifying entanglement thesis](#); outstanding score on comprehensive math exam.

Advanced Talmud, Ohr Somayach, 2008-16

- Gemara with encyclopedic breadth and deep analysis; granted semicha.

CERTIFICATIONS

Data Science and Machine Learning Certificate, IBM, 2018

- Executed data-driven solutions to increase efficiency and accuracy.
- Created data regression models with visualization using predictive data modeling.

Programming Course, Udacity, 2018

- Achieved proficiency in Python, Numpy, Pandas, and Github.
- Implemented knowledge acquired in SQL, HTML, CSS, Javascript.

SKILLS

- Analytics, Pandas, Visualization
- NLP, Scikit-learn, Tensorflow, PyTorch
- Agile, AWS, MongoDB, Django
- Industrious, organized, clean coding
- Selecting and developing ML models
- Learning new technologies quickly