#### YISHAI RASOWSKY

#### **Data Scientist**

## Portfolio | Github | Linkedin | 054 751 6040 | yishairasowsky@gmail.com

As a self-taught software developer and data engineer, I optimize cloud resources to provide the most suitable business solution. I am excited to solve complicated problems which do not have an obvious answer. I am passionate about utilizing ML and NLP algorithms; collecting, assessing, visualizing, and cleaning data; using a framework to connect domains using documented management of the data.

#### **SKILLS**

- Python, Pandas, Natural language processing
- Scikit-learn, Tensorflow, PyTorch
- AWS, MongoDB, Django

- Industrious, organized, clean coding
- Selecting and developing ML models
- Learning new technologies quickly

#### **WORK EXPERIENCE**

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

- Developed NLP topic modelling system to classify debriefs, for precision 95% and recall 50%.
- Developed NLP text generation for autocompletion, summarization, and comment suggestion.

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

Created NLP record linkage deduplication algorithm; achieved >95% completeness and homogeneity.

#### **INTERNSHIPS**

#### Data Scientist (Internship), Sefaria, 2020

- Created NLP multi-label classifier to assign topic labels to text passages.
- >95% precision, 36% recall. Designed algorithm specifically for high precision.
- Dataset was bilingual, in ancient Hebrew and English.
- Leveraged knowledge graph to enable hierarchical classification.
- Used Pandas, Scikit-Learn, Scikit-MultiLearn, MatPlotLib, Seaborn, Kubernetes.

#### Python Developer (Independent contractor), Smrt, 2019

- Created NLP entity extractor for lease contracts using rule-based logic.
- Developed interactive confusion matrix and statistics to evaluate predictions.
- Achieved accuracy of >90% for some entities; boosted precision/recall by 20%.
- Used Pandas, Scikit-learn, SpaCy, Amazon Web Services, MatPlotLib, Seaborn.

#### **PROJECTS**

- <u>BERT Classification</u>: 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.
- <u>Sefer Maker</u>: Utlizied Sefaria's API to automate creation of PDF sefer with translations and commentaries.
- <u>Screen Capture</u>: Surveillance system to snap shots of your computer screen in real time, and send them to WhatsApp.
- <u>Story Illustrator</u>: 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**

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

# **Programming Course, Udacity**

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