

# YISHAI RASOWSKY

Data Scientist

[Portfolio](#) | [Github](#) | [Linkedin](#) | 054 751 6040 | [yishairasowsky@gmail.com](mailto:yishairasowsky@gmail.com)

## SKILLS

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

## WORK EXPERIENCE

### **Data Scientist, Shamaym, 2021**

- 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, Streamline Verify, 2021**

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

## INTERSHIPS

### **Data Scientist, 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, 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

**BERT Classification**: Created modules that prepare data to fine tune BERT for 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.

**Sefer Maker**: Utilized Sefaria's API to automate creation of PDF sefer with translations and commentaries.

**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**

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

### **Advanced Talmud, Ohr Somayach**

- 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.