Data Scientist / AI Researcher / Algorithm Developer

Aviv Kadair I 052-6814195 I [avivkadair@gmail.com](mailto:avivkadair@gmail.com) I [LinkedIn](https://www.linkedin.com/in/aviv-kadair/) I Ramat Gan

**Summary**

* **4 years of experience** in **applied machine learning** and **end-to-end algorithm development**, including innovative, **custom models** built from scratch to address domain-specific challenges.
* Holds an **M.Sc. in Neuroscience** with hands-on experience in **research, signal processing**, statistical modeling, and real-world data analysis.
* Proficientin **Python, SQL, AWS, Git,** and **Linux,** with production experience in libraries such as **Scikit-learn, Pandas, NumPy**, and **Hugging Face Transformers.**
* Strong statistical background, including **regression analysis**, hypothesis testing, feature selection, and **model evaluation** across structured and biological signal datasets**.**
* **Developed models** for diverse applications, including spectroscopic analysis, semantic relation extraction (**NLP**), and computer vision**.**
* Skilled in **data acquisition** and **preprocessing,** including **web scraping**, custom data pipelines, feature engineering, and robust model validation techniques**.**
* **Intelligence Researcher (Arabic) & Tech Lead**, **8200 Unit**

**Experience**

2021 - 2024 **Data Scientist**, Olive Diagnostics

* **Developed spectroscopic algorithms** for the first passive, real-time urinalysis sensor (photodiode array), extracting biomarkers from high-dimensional optical data using smoothing, normalization, and dimensionality reduction.
* Trained and validated **supervised regression** and **unsupervised models** to accurately estimate urine flow rate, volume, and duration from continuous sensor streams.
* **Engineered preprocessing algorithms** to filter noise and fluctuations in real-time signals using FFT, low-pass filtering, and dynamic smoothing techniques.
* Adapted **spectroscopic signal processing** for real-world use by compensating for ambient light through dynamic recalibration, enabling accurate readings outside controlled environments.
* **Performed feature extraction** and **validation** to derive physiologically relevant insights from raw time-series data.
* Built and evaluated **classification models** using established algorithms like **XGBoost** and **logistic regression**, as well as custom **machine learning solutions** tailored to the device’s unique signal characteristics.
* **Implemented automated quality checks** to detect anomalies, sensor drift, and user-specific inconsistencies, ensuring clean and reliable model input.
* Collaborated closely with **hardware** and **firmware engineers** to translate both physiological and physics-based requirements into actionable sensor specifications.
* Aligned **machine learning outputs** with clinically relevant endpoints, working with clinical researchers to validate predictions against gold-standard lab urinalysis.
* Documented **model architectures, training procedures,** and **validation results** to support internal milestones and regulatory submissions.
* Delivered deployable**, real-time algorithms** for **embedded systems**, converting raw optical signals into medically actionable insights in a production environment.
* **Conducted statistical analyses** and **hypothesis testing** to evaluate signal stability, feature relevance, and model performance.

2020 - 2020 **NLP Data Scientist**, Pipl

* **Built a two-layer NLP pipeline** to extract proprietary relationships between entities mentioned in financial news articles.
* **Designed** and **implemented custom web scrapers** to collect labeled and weakly-labeled training data from financial websites.
* Applied **distant supervision techniques** to **generate training data** by aligning known entity relationships with free text, minimizing the need for manual annotation.
* **Conducted exploratory data analysis (EDA)** and **text normalization** to prepare raw financial text for modeling, including tokenization, lemmatization, and custom rule-based preprocessing.
* **Engineered features** for **initial classification** using syntactic patterns and contextual cues to detect the presence of a proprietary relationship.
* **Trained** and **fine-tuned a neural network model** for entity-level extraction using the filtered subset of sentences identified as relevant.
* **Evaluated model performance** with **classification metrics**, iteratively refining both layers to improve relationship extraction accuracy.

**Education**

2017 - 2020 **M.Sc. in Neuroscience**, Tel Aviv University

* **Thesis**: The effects of apoE-haplodeficiency on apoE-driven Alzheimer's Disease phenotype in vivo.
* **Developed a script** to process data from cell imaging, reducing analysis time by 70%.

2012 - 2015 **B.Sc. in Neuroscience**, with honors, University of Sussex

* **Thesis:** In the Eye of the Fly: A Computational Model for Visually Guided Behavior in Fruit Flies.
* Awarded the Chancellor's Excellence Scholarship
* Relevant coursework: **Artificial Neural Networks**, Intelligence in Animals and Machines.

**Trainings**

2020 **Data Science Fellowship**, ITC

* 5-month training focused on real-world applications of ML, DL, and NLP.
* Gained hands-on experience in **EDA, modeling**, and **deployment** using **Python**, with training in **supervised/unsupervised learning**, **model evaluation**, **computer vision algorithms,** and **NLP pipelines.**

**Professional Knowledge**

**Programming Languages & Tools:** Python, SQL, MATLAB, Git, Linux, AWS

**Libraries & Frameworks:** Pandas, NumPy, scikit-learn, Scipy, TensorFlow, Keras, Matplotlib, Seaborn, Hugging Face Transformers, Optuna, MLflow, SHAP, OpenCV

**Concepts & Techniques:** OOP, EDA, Data Visualization, Statistical Analysis, Hypothesis Testing, A/B Testing, Feature Engineering, Model Evaluation, Neural Networks, Time-Series Analysis, Class Imbalance Handling, Data Augmentation, Cross-Validation Strategies, Dimensionality Reduction, Custom Loss Functions, Model Deployment & Inference, Experiment Tracking, Transfer Learning & Domain Adaptation, Explainable AI

**Domains & Applications:** Machine Learning, Deep Learning, Natural Language Processing (NLP), Computer Vision, Web Scraping, Signal Processing, Biological Signal Modelling

**Military Service**

**Intelligence Research in Arabic & Tech Lead**, **8200 Unit**

**Languages**

**Hebrew & English** - Native I **Arabic & Spanish** - Advanced

**Volunteering**

**Youth Counselor,** ELEM – Youth in Distress

* Supported at-risk youth through mentorship and emotional guidance, building trust and promoting healthy coping skills.
* Assisted in community outreach, awareness events, and fundraising efforts.