Erol **Kavvas**

(+1) 530-220-2141 | eskavvas@gmail.com | erolkavvas | erolkavvas

Summary

Data scientist and machine learning researcher with experience developing and deploying scalable, end-to-end AI systems in production. Skilled in building interpretable models, leading cross-functional projects, and optimizing large-scale ML pipelines using cloud infrastructure. A proven track record in providing actionable insights and tools for both research and product teams. Passionate about applying data science to complex real-world problems across industries, from healthcare to consumer technology.

Experience

Envisagenics

New York City, NY

Data Scientist — *July 2022 - Feb 2025*

- Built and deployed *SplicePath*, an Azure cloud-based ML pipeline for modeling exon/transcript impact on pathway activity; adopted by cross-functional oncology teams for experimental design and hypothesis generation of 20+ drug targets.
- Led development of network interpretation layer for *SplicePath* to explain transcript-pathway effects via molecular interaction graphs.
- Designed a no-code interface via Microsoft Fabric for non-technical users and improved internal adoption of *SplicePath* across oncology teams.
- Developed *SpliceImpact*, a transformer-based LLM model for predicting per-residue functional effects from protein sequence; optimized with PySpark and Azure ML.

Integrative Biology and Predictive Analytics Lab

Davis, CA

Postdoctoral Researcher, UC Davis — *Sep 2020 - July 2022*

- Built ML models for personalized diet recommendation using 5,000+ gut microbiome samples and strain-specific metabolic networks.
- Applied ML to multi-omics data in studies on IBA tolerance and bat viral immunity; led ML development on grant-funded research.

Sinopia Biosciences

San Diego, CA

Machine Learning Consultant — *Oct 2019 - Jan 2020*

- Preprocessed and batch-corrected drug response metabolomics data from 1,000+ samples.
- Built ML models to identify metabolic signatures distinguishing drug action.

Systems Biology Research Group

UC San Diego, CA

Graduate Student Researcher, San Diego — *Sep 2015 - Sep 2020*

- Published 4 first-author publications and coauthored 10 papers; cited 900+ times; research featured by UCSD press release and F1000.
- Built pangenome from 1,595 pathogens and trained SVMs to predict drug resistance; identified 33 known and 24 novel resistance genes.
- Pioneered first hybrid FBA–ML framework to model biochemical effects of mutations; matched SOTA performance while improving interpretability and experimental hypotheses for 3 anti-tuberculosis drugs.
- Applied ICA and statistical modeling to link *E. coli* adaptive mutations to transcriptomic and flux-level constraints; identified 6 conserved transcriptomic strategies, 5 regulatory tradeoffs, 4 mutation-flux correlates, and 8 mutation-transcriptome correlates.

Education

University of California, San Diego

San Diego, CA

PhD in Bioengineering, Advisor: Bernhard Ø. Palsson — *2015 - 2020*

Thesis: *Biologically-Interpretable Machine Learning for Microbial Genomics*

University of California, Davis

Davis, CA

B.S. in Civil and Environmental Engineering — *2010 - 2015*

Skills

Programming	Python (Expert), R, SQL, Linux, C, Visual Basic, Fortran, Bash
Machine Learning	PyTorch, scikit-learn, Deep Learning, Large Language Models, Neural Nets, Linear models, SVMs, Interpretable ML, LIME, SHAP, Tree-based Methods, Decomposition Methods, Ensemble Methods, Network Modeling, Recommender Systems, Model Optimization
Data Analysis & Visualization	Pandas, Numpy, Matplotlib, Seaborn, Statistical Methods, A/B Testing, Experimental Design
Big Data Technologies	PySpark, Microsoft Azure (Synapse, ML Studio), Microsoft Fabric, Docker, GPU Optimization
Other Skills	Git, NGS Analysis, Flux Balance Analysis, LaTeX, Matlab, Affinity Designer, Ableton

Invited Talks

Artificial Intelligence in Genomics

Potsdam, Germany

Presenter for <Association of Gene Diagnostics Meeting> — *Sep, 2019*

- Presented on an approach for integrating biochemical mechanisms with machine learning

Tuberculosis and Machine Learning *Oslo, Norway*

PRESENTER FOR <4TH TURNING THE TIDE OF ANTIMICROBIAL RESISTANCE MEETING> *Nov. 2019*

- Presented on a machine learning approach for identifying drug resistance genes in *M. tuberculosis*

Selected Publications (*Full list at my Google Scholar)

2022	**Laboratory evolution reveals unifying systems-level principles of adaptation**, [1st Author]	*Msystems*
2020	**A biochemically-interpretable machine learning classifier for microbial GWAS**, [1st Author][*Github link*]	*Nature Comm.*
2020	**Machine learning with random subspace ensembles identifies antimicrobial...**, [2nd Author]	*PLoS Comp. Bio.*
2018	**Machine learning and structural analysis of *M. tuberculosis* pangenome...**, [1st Author] [*Github link*]	*Nature Comm.*
2018	**Updated and standardized genome-scale reconstruction of *M. tuberculosis*...**, [1st Author] [*Github link*]	*BMC Syst. Bio.*
2018	**Genome-scale metabolic reconstructions of multiple Salmonella strains...**, [2nd Author]	*Nature Comm.*

References

Bernhard Ø. Palsson, Ph.D.
Distinguished Professor
Bioengineering
Jacobs School of Engineering
University of California, San Diego
La Jolla, CA 92093-0411
palsson@ucsd.edu

Martin Ackermann, Ph.D.
Co-Founder and CTO
Envisagenics
30-02 48th Ave, Suite 140,
Long Island City, NY 11101
makerman@envisagenics.com

Ilias Tagkopoulos, Ph.D.
Associate Professor
Computer Science
College of Engineering
University of California, Davis
1 Shields Ave, Davis, CA 95616
itagkopoulos@ucdavis.edu