

# David Tovmasyan

732-668-5138 | dtovm001@gmail.com | linkedin.com/in/daviddtovmasyan | Website: kikossik.github.io

## EXPERIENCE

---

### Machine Learning Engineer

Aug 2024 – Present

*University of Southern California*

*Remote, CA*

- Ophthalmology research at Dr. Penkova's lab, analyzing retinal fundus photography for eye disease detection
- Explored efficient low-cost DL models achieving high accuracy compared to resource-intensive architectures
- Evaluated models on publicly available data for test purposes comprising 5,000+ images, aiming to scale further
- Worked with MobileNet and EfficientNet families, integrating the new ADOPT optimizer, reaching 90% accuracy

### Product Analyst Intern

Jun 2024 – Aug 2024

*ABC Fitness*

*Remote, CA*

- Pioneered a classification model implementation for the largest ABC platform, Ignite, to predict member churn
- Collaborated with ABC Glofox to understand their PyCaret model, developed automated ML pipeline for Ignite
- Engineered a custom ML model, achieving 96% accuracy with an 11% improvement over the PyCaret pipeline
- Tested and trained ML models on 600k+ observations with automated model selection based on f1-score metric
- Analyzed 7+ mil. Azure database records to compare major gym churn rates and evaluate churn across locations
- Created PowerBI dashboards to visualize churn rates across states with a breakdown of the group attributes

### Data Scientist - Research

Jan 2023 – May 2024

*Stony Brook University*

*Stony Brook, NY*

- Research with Prof. Montgomery and Prof. Halada on "How Climate Change Affects Oil and Gas Pipelines"
- Analyzed sea level rise and extreme storm events as key future factors driving pipeline corrosion and cracking
- Deployed an interactive dashboard using D3.js and flask displaying the damage and location of the pipelines
- Preprocessed 11,500 PHMSA pipeline failure records, merging with NOAA storm events by date and location
- Presented research at SBU Economics Conference, presented poster at URECA for Halada's EACC VIP team

### Data Science Intern

Jun 2022 – Aug 2022

*Akai Kaeru*

*Stony Brook, NY*

- Worked with Prof. Klaus Mueller from CS department on Akai Kaeru - AI-powered Data Science platform
- Analyzed 10+ datasets using this software, climate change and oil and gas data, in total of 400k+ observations
- Tested features of the software, constructed 20+ regression and ML models, helped identifying bugs in the software
- Presented findings and insights from data analyses in team meetings, influencing the project direction

### Credit Risk Modeling Intern

Jun 2021 – Aug 2021

*Sberbank*

*Almaty, Kazakhstan*

- Built a Probability of Default model using 2000+ small and micro business loans from the period 2019-2020
- Implemented Logistic Regression with L2 Regularization using k-fold cross-validation for the PD model
- Automated data collection of client credit risk history for LGD model, boosted data entry time efficiency by 44%
- Fixed 4 broken Excel instruments with VBA, continuously supported 2 key applications for merging credit data
- Helped our team with other daily technical problems using Python (Jupyter Notebook) and Excel/VBA

## EDUCATION

---

### University of Southern California

Los Angeles, CA

*Master of Science in Data Science*

- Part-Time + Remote, Open to Full-Time Work
- Ophthalmology research at Dr. Anita Penkova's lab

### Stony Brook University

Stony Brook, NY

*Bachelor of Science in Applied Math & Statistics, Economics | minor in Computer Science*

- Teaching Assistant for ECO 305 (Macroeconomics) - held office hours, proctored exams, graded assignments

## PROJECTS

---

### **New York City Rental Affordability Analysis** | *R, Regression, ggplot2, sf, mapview*

Website: [kikossik.github.io/portfolio/p1-ny-mih/](https://kikossik.github.io/portfolio/p1-ny-mih/)

- Analyzed rental affordability in NYC using spatial and statistical techniques to evaluate MIH policy's effectiveness
- Suggested custom metrics to quantify affordability disparities across boroughs and Community Districts
- Identified 29 districts (49%) showing affordability differences, highlighting significant flaws of the MIH policy
- Sanitized over 80,000 sales data in R (NYC\_Sales), data from NYU Furman center, NYC Open Data

### **Does Climate Change Affect Oil & Gas Infrastructure?** | *d3.js, Python, Flask*

Website: [kikossik.github.io/portfolio/p2-oil-gas-research/](https://kikossik.github.io/portfolio/p2-oil-gas-research/)

- Implemented an interactive dashboard for Hazardous Liquids pipelines (PHMSA data) with D3.js and flask
- Provided tools for users to map specific pipeline failure causes with corresponding damage and location in the US
- Processed 11,500+ observations, merging data from multiple files and converting address names to geo-coordinates
- Presented research at SBU Economics Conference, presented poster at URECA for Halada's EACC VIP team

### **Distributed Database Management System** | *Python, Django, MySQL, Selenium WebDriver, BeautifulSoup*

Website: [kikossik.github.io/portfolio/p3-distributed-db/](https://kikossik.github.io/portfolio/p3-distributed-db/)

- Built a Django-based Distributed Database Management System to handle job postings across MySQL databases
- Developed a custom hash function to partition data evenly across databases, ensuring balanced loads
- Implemented data scraping from LinkedIn with Selenium WebDriver and BeautifulSoup for navigation and parsing
- Built custom commands for bulk import and bulk removal across databases with hash-based partitioning

## TECHNICAL SKILLS

---

**Languages:** Python, SQL, R, MATLAB, JS (d3.js), HTML/CSS

**Developer Tools:** Git, AWS (EC2), Azure, SSMS, DB (NoSQL, MySQL, PostgreSQL, MongoDB)

**Libraries:** Pandas, Geopandas, NumPy, Matplotlib, Seaborn, Sklearn, Tensorflow, Keras, Selenium, BeautifulSoup