David Tovmasyan

(732) 668-5138 | dtovm001@gmail.com | linkedin.com/in/davidtovmasyan | github.com/kikossik

EXPERIENCE

University of Southern California — **Remote, CA**
Machine Learning Engineer — *Aug 2024 - Present*
• Ophthalmology research at Dr. Penkova's lab, focusing on analyzing retinal scans to detect anomalies for early glaucoma

ABC Fitness — **Remote, CA**
Product Analyst Intern — *Jun 2024 - Aug 2024*
• Pioneered churn classification model implementation for the largest ABC platform - Ignite - to predict member churn risk
• Analyzed 7+ mil. instances from Azure database to compare major club churn rates and evaluate churn across locations
• Created a PowerBI dashboard to visualize churn rates across states with a breakdown of churned/non-churned attributes
• Tested and trained ML models on 600k+ observations with automated model selection based on f1-score metric
• Collaborated with ABC Glofox to understand their PyCaret model, developed an automated ML pipeline for ABC Ignite
• Engineered a custom ML model from scratch, achieving 11% improvement in model accuracy over the PyCaret pipeline

Stony Brook University — **Stony Brook, NY**
Data Science - Research — *Jan 2023 - May 2024*
• Conducted research with Prof. Montgomery and Prof. Halada on "How Climate Change Affects Oil and Gas Pipelines"
• Deployed an interactive dashboard using D3.js and flask displaying the damage and location of pipelines (on GitHub)
• Data from PHMSA, data cleaning and preprocessing of 11,500 pipeline failure data implemented in Jupyter Notebook
• Presented research at SBU Economics Conference; presented poster at URECA for EACC VIP team with Prof. Halada

Akai Kaeru — **Stony Brook, NY**
Data Science Intern — *Jun 2022 - Aug 2022*
• Worked with Professor Klaus Mueller from CS department on Akai Kaeru - AI-powered Data Science platform
• Analyzed 8 datasets using this software, climate change and oil and gas data, with 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

Sberbank — **Almaty, Kazakhstan**
Credit Risk Modeling Intern — *Jun 2021 - Aug 2021*
• Built a Probability of Default model in Python 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 client credit data
• Helped our team with other daily technical problems using Python (Jupyter Notebook) and Excel/VBA

EDUCATION

University of Southern California | Viterbi School of Engineering
M.S., Data Science

Stony Brook University | College of Engineering and Applied Sciences
B.S., Applied Math & Statistics, Economics, Minor in Computer Science

ACADEMIC & PERSONAL PROJECTS

Distributed Database Management System | github.com/kikossik/Job-Posting-Distributed-Database-Management-System
• Developed a Django-based Distributed Database Management System for job postings with an interactive platform
• Implemented data scraping from LinkedIn using Selenium WebDriver and BeautifulSoup to get the job postings
• Engineered a partitioning system across three databases using my own hash function to distribute job posting data

Oil and Gas Pipelines Interactive Dashboard | github.com/kikossik/D3js-flask-Interactive-Dashboard
Website URL: kikossik.pythonanywhere.com
• 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 the corresponding damage and location in the US
• Processed 10,000+ observations, merging data from multiple files and converting address names to geo-coordinates

New York City Rental Affordability Analysis
• Led data science in a team of 6 to improve affordability metrics for NYC renters by analyzing rental prices and income
• Sanitized over 80,000 sales data in R (NYC_Sales), data from: NYU Furman center, NYC Open Data
• Responsible for spatial data, e.g. created 300ft buffers around each sales point to include in the regression model

SKILLS

Languages: Python (Pandas, NumPy, Matplotlib, Seaborn, Sklearn, Keras, PyTorch), SQL, JS (d3.js), R, MATLAB
Technologies: AWS (EC2), Azure, SSMS, Django, Flask, PowerBI, DB (SQL, NoSQL, MySQL, MongoDB), Git