Stas Sajin

Lead Data Scientist

San Francisco, California

510-804-5594 | sajin.stanislav@gmail.com | medium.com/@sajin.stanislav | stassajin | stas-sajin

Skills

Programming	Python, R, SQL, Shell Scripting, Scala
Cloud Platform, CI/CD	AWS(S3, EC2, Sagemaker, VPC, RDS, DMS, IAM), Boto, Jenkins, Docker
Machine Learning	Tensorflow, Xgboost, LightGBM, Scikit-Learn, DVC, Pandas, Numpy
Modeling Experience Domain	Dynamic Pricing with Bandits, Lead Scoring, Conversion Modeling, Risk Modeling
Databases, BI, ETL	Dagster, DBT, Airflow, Looker, PostgreSQL, Redshift, Snowflake

Work Experience

Earnest

San Francisco, California

Lead Data Scientist — *Jan. 2019 - Present*

- Core data scientist on the team building machine learning services that support Earnest underwriting infrastructure.
- Developed a Python microservice template. This template allows a person on the team to easily build and deploy Python microservices. It comes with ORM, logging, dependency management, and a variety of automation utilities that enable rapid model deployment.
- Researched, built and deployed an income service model. This model outperformed several 3rd party top tier providers and supports 50% automation in underwriter decision making.
- Researched and built a customer acquisition model that led to a 40% reduction in direct mail marketing costs and outperformed several third-party providers, such as Epic and LendingScience.
- Supervised the work of 3-4 direct reports (mix of Data Engineers and SWE) by scoping out project requirements, performing sprint reviews, retrospectives, postmortems, code reviews, handling promotions and individual mentoring.
- Exposed data from over a dozen microservices in Redshift. Set up the tone for data governance and data documentation. Maintained and optimized the data warehouse.
- Wrote a Python microservice that supports dynamic pricing and scoring. The service is responsible for scoring $10 billion of student loan volume/year.

Senior Data Scientist — *Jul 2017 - Jan. 2019*

- Managed Looker administration, maintenance, and optimization.
- Standardized the process of developing models internally by writing a python package (earnest-ml) which allowed Data Scientists and Engineers to develop models using reproducible workflows.
- Wrote an artifact pipeline that allowed the team to have parity between feature transformation done during training and inference. This cut down model deployment time by 70%.
- Developed and productionalized a personal scoring model that led to a 40% increase in the number of approved applications with a negligible increase in default risk.

Data Scientist — *Jul 2016 - Jul 2017*

- Refactored existing risk scoring models in R and Scala.
- Built a hierarchy of conversion models that guided user flow through our conversion funnel.

Projects

BabyNet

- A neural network trained on baby expressions and spliced videos using EfficientNet as the architecture backbone. The model is pruned extensively using the Lottery Ticket hypothesis technique in order to speed up inference performance
- Developed a pipeline for active labeling, so that the model trains as video streams are labelled
- Several cameras inside the household are used to stream the video feed through RSTP to the model
- The final output are push notification and video streams that are sent to an iPhone app that alerts me about anything that is going on with my baby

Prosper Bot

- A trading bot developed that is used to execute micro-lending transactions
- The bot predicts risk of users on the Prosper platform and automatically makes trades using Prosper API
- The bot has been used for 5 years and has made tens of thousands of transactions averaging 15% return with very little volatility

Education

Binghamton University

Binghamton, NY

Ph.D. Cognitive and Brain Sciences — *Mar. 2011 - 2015*