Boris Dev | boris.dev@gmail.com

Interests

- Process optimization
- Building observability systems
- Knowledge crunching to document a ubiquitous language

Tools

- **Backend engineering:** Linux, Python, Postgres, Flask, Django
- **DevOps:** Docker, Jenkins, Helm Charts, AWS Sagemaker and Lambda
- **Observability:** Splunk, Lightstep for distributed tracing
- **Data science:** SQLAlchemy, PySpark, Pandas, Numpy, PyTorch, Tableau
- **Teamwork:** Inviting feedback with demos and papers (examples).

Job experience

Data Scientist / Technical AI Product Manager, consultant at SimpleLegal, 2022-2023

The SimpleLegal product is a tool for legal departments of large companies to manage the invoice bills from their vendor law firms. Before I arrived, the company had been working on a language AI feature for flagging suspicious line items. For one year, this feature was stuck because of bad performance. They were unable to increase performance even after spending more money on human annotation. My job as tech lead was to increase AI performance. The actions I took resulted in the company launching their first AI feature, with positive feedback from customers and the sales team, as well as an AI Excellence Award.

Managed the data science consultant

- To increase clarity, I helped the data scientist by organizing his analysis of the machine learning metrics into a table, which was continually shared at the weekly stand-ups to help others comprehend our progress.
- In the spirit of shared success, I worked daily as a team member with the data scientist doing exploratory data analysis of the invoice data. This resulted in the following triage: We killed one classifier that was never needed, replaced another with expert rules, stopped the labeling on five whose performance was sufficient, and lastly, narrowed the focus of the labeling effort from eleven to two classifiers. In addition, we identified two culprits of bad quality labeling, 1) convoluted annotation guidelines and 2) missing pre-processing noise filters.
- Periodically, I asked the data scientist to take the lead in explaining complex concepts to the rest of the team, including executives.
- I wrote him a stellar performance review, which highlighted his capacity to articulate complex concepts to executives.

Managed the offshore annotators

- To understand the perspective of the offshore annotators, I labeled several thousand sentences.
- I built a team relationship with the offshore team of annotators. My objective was to increase labeling, quality. I managed them with a spirit of continuous improvement and consensus. If they disagreed, I asked them to challenge the subject-matter expert and I to explain to test our rationale.

Built relationships across the conglomerate

- To teach me, I connected with experts spread across three sister companies. This group included a principle data scientist, two product subject-matter experts, one of which was a founder and VP, one DevOps engineer, one MLOps engineer, and one human annotation expert.
- I probed, got consensus, and acted. After it was agreed that the biggest potential factor to bad performance that had been unexplored to date was *"garbage in, garbage out"*, I stopped the human annotation process and changed the focus from neural network R&D and labeling quantity to learning about the foundational concepts of labeling quality.
- To seek clarity, I asked the company's President to convene a team meeting in the style of the Amazon 6-pager. The team spent the first fifteen minutes reading my high-level design document explaining that the performance blocker was not training data quantity, but quality, as well as other clarifications.
- I assigned Jira issues to engineers. 1:1 meetings with the VP of Engineering. After one engineer who had been helping me left to another job, I stepped in to refactor the inference server.

Backend developer at Sight Machine, 2018-2021

The Sight Machine product is a dashboard tool for process engineers at manufacturing plants to keep track of performance.

I built relationships across silos (customer success, DevOps, data engineers, and product engineers).

- To bring the company's biggest public-facing feature, Recipes, from its embryonic start as a spreadsheet to general release, I shared high-level design papers and gave non-technical explanations to the Product and Customer Support teams. This allowed them to get accurate customer feedback on product technical design decisions.
- I started a new process to ensure toil free and quality sales demos. Engineers must be included in the calendar invites. Twenty-four hours before the scheduled demo, the environment's compute workers are scaled up, the best frontend and backend branches deployed, and lastly, the product manager starts smoke and sanity testing.

- I oversaw the integration of my instrumentation work with the required DevOps work.

I helped increase the engineering team's productivity.

- The Director of Engineering and I started a new process where the engineers were given the autonomy to write their own Jira issues, instead of being delegated Jira issues by the Product team.
- To debug faster, I started the first distributed tracing (Lightstep).
- I simplified the frontend development environment.
- I coached junior engineers.

Data and product engineer tech lead at HiQ Labs, 2015-2018

Our product provided predictions to our customers on whether their employees were about to quit.

Tech lead for scraping

- The puzzle for the CTO and I was to figure out how to get around LinkedIn's bot detection in order to scrape millions of HTML public profiles, the raw data for our prediction pipeline.
- I ran and tracked experiments on different spider configurations.
- I led a junior devops engineer to help me build a Splunk observability system.
- I trained a junior data engineer to help with scraping.

Tech lead to move the company from a monolith to a microservice architecture

- I broke down the coding work.
- I explained to management the microservice architecture using whiteboard diagrams.

Supported data scientists to reduce their toil

- I shadowed the data scientists.
- I migrated the data scientists from Mongo queries to PySpark.
- To scale, I trained a junior data engineer to train the data scientists on the new development environment that I had setup.

Start-up partner and developer at Map Decisions, 2014

We provided a mobile app to automate street sign inspection for the Public Works departments of cities. I built the app.

Developer at Urban Mapping, 2011-2013

We provided a location query and map tiling service to Tableau. I built the company's first observability system and QA system.

Non-software jobs

- Assistant Manager at Gundy's Grill restaurant at Beaver Creek Ski Resort.
- Kids snowboard instructor at Beaver Creek Ski Resort.
- Lead of four analysts at California Business Magazine. We collected financial data on local public companies for newspapers such as the Los Angeles Times. In addition, I provided data visualizations for a prototype of the Wall Street Journal's "Main Street" section.

Open source, papers, and academics

- Why did your language AI feature fail?
- Work papers
- A git bare approach to version control your dot files
- A play Ethereum MEV bot
- Geoscore started as a startup idea to make choropleth maps to rank and visualize the demographics of 85,000 Census tracts (or neighborhoods) based on the user's custom preferences on where she wants to live. Geoscore repo.
- Co-Founder of the ClusterPy library for clustering geographic areas. I taught student researchers in Medellín, Columbia for three weeks in geographic areal data, object-oriented design and UI programming.
- Coached students in the master's degree program, as a PhD student
- Assisted teaching a class in spatial statistics, as PhD student
- PhD dissertation: Assessing Inequality using Geographic Income Distributions,
  2014.
- Entry in Encyclopedia of Human Geography on Spatial Econometrics. Sage Publications. 2009
- Interactive spatio-temporal modelling of health systems
- σ-convergence in the presence of spatial effects
- Integrating Econometric and Input-Output Models in a Multiregional Context