Boris Dev | boris.dev@gmail.com

Interests

- Process optimization to scale human
- Building observability systems to get quick feedback
- Documenting a ubiquitous language to work together

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

Legal departments of big companies to use SimpleLegal software to manage bills from their vendor law firms. The situation before I arrived was that their new language AI feature for flagging suspicious invoice lines was was stuck because of bad performance. My job 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 senior data science consultant

- The technical vision I set for the data scientist was to continually search for "low hanging fruit" to scale back machine learning to launch an MVP. This resulted in scaling back our machine learning training from eleven to two classifiers. The triage was to kill one classifier that was never needed, replace another with expert rules, and stop the labeling on five whose performance was sufficient.
- 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 inference server engineering and QA tasks

- I assigned Jira issues to engineers.

- 1:1 meetings with the VP of Engineering.

- After one engineer left to another company, I stepped in to refactor the inference server.

- I organized the subject-matter expert embarrassment reviews
- To increase clarity, I crunched the knowledge of multiple experts into a working paper, which was continually shared at the weekly stand-ups to help others comprehend our progress.

Managed the offshore annotators

- I overhauled the relationship with our team of offshore annotators from hierarchical to democratic.
- To understand the perspective of the offshore annotators, I labeled several thousand sentences. This lead me to identify convoluted annotation guidelines as the root cause of our poor labeling quality.
- I asked the manager of the annotators to let his workers interact directly with me and the subject-mater through Slack to give provide us with challenging feedback on our guidelines (the shared language document).
- I developed and shared new performance metrics over periodic zoom meetings.

Built relationships across the conglomerate

- 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.
- A VP of Product/Founder at a sister company helped me recognize that past work was missing pre-processing noise filters.
- A annotation expert helped me recognize that we needed a democratic relationship with the annotators to ensure quality.
- 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.

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 led the following initiatives to increase the engineering team's productivity.

- The Director of Engineering and I started a new process where the engineers wrote their own Jira issues, instead of being delegated Jira issues by the Product team. The new technical design autonomy allowed us to meet requirements with cheaper technical implementations.
- To debug faster, I instrumented and built distributed tracing (Lightstep) and managed the DevOps work.
- To simplify on-boarding, I containerized 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 led a junior devops engineer to build a Splunk observability system that was used to track performance and experiments on different spider configurations.
- I trained a junior data engineer to maintain the scraping system.

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

- I coached a data engineer to guide the data scientists into a new microservice development pattern.
- I explained to management the technical vision behind how the microservice architecture using whiteboard diagrams.
- I explained to management the human vision behind how the microservice architecture saving the data scientists from toil.

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.

Misc jobs

- Coached students in the master's degree program, as a PhD student at S.D.S.U.
- Assisted teaching a class in spatial statistics, as PhD student at S.D.S.U.
- Kids snowboard instructor at Vail Resorts, CO.
- Assistant Manager at Gundy's Grill restaurant at Vail Resorts, CO.
- Researcher on building a prototype of the Wall Street Journal's "Main Street" section. California Business Magazine. CA.
- Led analysts collecting regional company ranking for large newspapers. California Business Magazine. CA.
- Counselor for severely emotionally disturbed children at Seneca Family of Agencies. CA.

Writing

- [Why did your language AI feature fail?](#)
- [Work papers](#)
- PhD thesis. [Assessing Inequality using Geographic Income Distributions](#)
- Entry in Encyclopedia of Human Geography on Spatial Econometrics. 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](#)

Open source code

- [A play Ethereum MEV bot](#)
- [A git bare approach to version control your dot files](#)
- [Geoscore](#) website and [repo](#) to rank and visualize 100,000+ neighborhoods.
- Co-Founder of the [ClusterPy library](#) for clustering geographic areas. I taught students in Columbia in geographic data, OO design and UI programming.