

# Nathaniel Budijono

## Resumé

✉ [nathanielbd@gmail.com](mailto:nathanielbd@gmail.com)

📄 [nathanielbd.github.io](https://nathanielbd.github.io)

🌐 [nathanielbd](https://nathanielbd.github.io)

in [www.linkedin.com/in/nathanielbd](https://www.linkedin.com/in/nathanielbd)

## Education

2018–2022 **Bachelor of Science**, *University of Minnesota*, Computer science and Mathematics.

Focuses: machine learning, computational biology, computational chemistry.

- Reviewer, Minnesota Undergraduate Research Academic Journal; Education Director, App Developers Club; Lab Director, Society of Asian Scientists and Engineers; Mentor, Honors program.

## Industry Experience

Jun 2022 – **Data Scientist**, *Infinite Campus*.

- Responsible for all data governance initiatives and machine learning product maintenance for the company. *Docker*, *Kubernetes*.
- Reduced runtime of training and deployment pipeline of machine learned early warning system to forecast late graduation/drop out for millions of students by 10x. *joblib*, *multiprocessing*, *redis*, *Airflow*.
- Engineered monitoring, interpretability, and alerting features to the early warning system deploy pipeline. Automated data validation that previously required manual investigation. *SHAP values*, *Aporia*.

Sep 2021 – **Intern**, *Loyal Cellular Longevity*, Computational biology team.

Apr 2022 Wrote cross-team documentation and worked on corporate website. *Typescript*.

May 2021 – **AI Research Intern**, *Smart Information Flow Technologies (SIFT)*.

- Developed decoding scheme for BERT-based language model to improve joint entity-relation classification. *Huggingface*, *PyTorch*.
- Statistically analyzed sentiment and word embeddings of online groups of interest for publication. *spaCy*.

May 2019 – **Software Engineer Intern**, *Rally Health*, Data engineering team.

- Created a utility for data pipeline to gauge data throughput, monitor job duration, and track which rows failed which data validation. *Scala*, *Datadog*, *Databricks*.
- Created a comprehensive and responsive web dashboard with pipeline health and performance metrics for data pipeline. *Airflow*, *Flask*, *SQLAlchemy*, *nvd3.js*.
- Created an extract-transform-load (ETL) job to process ledger data from incentives program and performed end-to-end testing. *Apache Spark*, *SQL*.

## Research Experience and Publications

May 2020 – **Research Assistant**, *University of Minnesota*, Prof. Chad Myers.

- Created a visualization pipeline for genetic interaction networks from raw combinatorial CRISPR screen data. *Networkx*, *Scipy*.
- Investigated machine learning algorithms for predicting gene function from genetic interaction networks. *PyTorch*, *Scikit-learn*.

- *Ludus: An Optimization Framework to Balance Auto Battler Cards*. **Nathaniel Budijono**, Phoebe Goldman, Jack Maloney, Joseph Mueller, Phillip Walker, Jack Ladwig, and Richard Freedman. AAAI Symposium on Educational Advances in Artificial Intelligence, 2021.

## Awards

- Oct 2021 Winner (1/20, \$500), Acceleration Consortium Hackathon for Scientific Database Management
- Jul 2021 David E. Shaw Research Science and Engineering Undergraduate fellowship
- Sep 2021 Honorable mention (top 5+/1300), 3blue1brown Summer of Math Exposition
- Oct 2020 Finalist (top 10/n, \$1000), Wells Fargo Campus Analytics Challenge
- Mar 2020 Best ML, AI, and Social Good Hacks, Hacktech (Caltech hackathon)