






ILYA NOVIKOV, PH. D.

 [ilyabnovikov](#) |  [ily123](#) |  [ilyanovikov.io](#) |  ibnovikov@gmail.com |  713-876-8284

Skills

Languages & DBs: Python, SQL, MATLAB, HTML, MySQL + SQLite

Libraries & Frameworks: pandas, numpy, sklearn, plotly, Dash

Developer Tools: Jupyter, Git, AWS EC2/RDS/El. Beanstalk, Airflow, vim, \LaTeX

Stats & ML: inferential & descriptive statistics, regression, clustering, graph learning, NNMF

Experience

Data Science Fellow

May 2019 – Aug 2019

Mercury Data Science

Houston, TX


- Consulted on data analytics for an app-based concierge services startup.
- Cleaned raw customer transaction data and engineered relevant features in a SQL database.
- Scraped third-party sources to add meaningful features via Python scripts.
- Identified major customer segments (k-means), and conducted A/B testing finding that apartment tenants who used the app were 20% more likely to renew their lease.
- Uncovered underlying data issues that prevented effective customer churn modeling, and showcased them to the team leader.

Graduate Student Researcher

Aug 2010 – May 2019

Baylor College of Medicine

Houston, TX

- Developed & published RNA Evolutionary Trace method to discover critical functional sites in RNA structures (see projects).
- Traced history of natural mutations in the Human Immunodeficiency Virus to help collaborator prioritize sites for experimental study of the virus.
- Collaborated with industry and team members to develop KniT biomedical literature mining tool, leading to discovery of novel disease-relevant interactions between human proteins.
- Developed novel semantic similarity protein graphs to serve as vehicles for predicting protein functions with graph diffusion and NNMF (see projects, also  presentation at ISCB-2017).
- On daily basis: formulated and tested hypotheses, compiled and cleaned data, transformed & feature engineered data, trained and validated predictive models, visualized and reported results in form of PowerPoint presentations, and written reports.
- Presented technical talks at local and international conferences.

Projects

RNA Evolutionary Trace | *Matlab, Python*

- First-principles algorithm that scans the evolutionary history of RNA molecules and predicts their most important nucleotides with mean AUC=0.83, better than standard method in 16 out of 19 test cases.

GGid | *Python, SQLite, Dash, Heroku App*

- Constructed protein graphs from structured biological data (Gene Ontology), then applied graph learning algorithm (information diffusion) to identify clusters of connected proteins.
- Method captures existing connections well (cross-val AUC=0.87), but overtrains (hold-out validation AUC=0.71).

Education

Baylor College of Medicine

Houston, TX

Ph.D. Biochemistry

May 2019

Texas A&M University

College Station, TX

B.S. Biochemistry

May 2010