Anish Saha

Member of Technical Staff @ Oracle.
Graduate Student @ UIUC.

+1 408.806.8843 saha.anish@gmail.com bit.ly/asaha github.com/anish-saha linkedin.com/in/anish-saha

PROFILE

Ambitious, versatile developer with over two years of experience in machine learning, data science, and software engineering. Currently seeking high impact industry roles to gain experience and technical knowledge.

EDUCATION

University of Illinois, Urbana-Champaign
Champaign, IL • 2021-2022 • GPA: ---
Masters, Computer Science

Stanford University
Stanford, CA • 2019-2021 • GPA: 3.85
Graduate Certificate, Artificial Intelligence

University of California, Berkeley
Berkeley, CA • 2015-2019 • GPA: 3.11
Bachelors, Applied Mathematics
Bachelors, Data Science

PUBLICATIONS

"A Method for Measuring Investigative Journalism in Local Newspapers". Turkel, Eray; Saha, Anish; Owen, Rhett C.; Martin, Gregory J.; Vasserman, Shoshana. *Proceedings of the National Academy of Sciences Jul 2021, 118 (30) e2105155118; DOI: 10.1073/pnas.2105155118.*

TECHNICAL SKILLS

Programming Languages
Python, Java, R, Swift, C / C++ / Objective-C, Javascript, MATLAB, Bash, HTML, CSS, PHP

Frameworks / Libraries
Pytorch, Numpy, Pandas, Scikit-Learn, Scipy, TensorFlow, Matplotlib, NetworkX, Google Natural Language, Facebook Graph, Apache Spark, Amazon S3, Oracle Essbase, and more

Software Tools
Figma, XCode, Android Studio, Tableau, NewRelic, Conviva, Anaconda, Jupyter Notebook, Oracle Smart View

PROJECTS

Social Network Sybil Detection
- Engineered and optimized a two-stage pipeline to detect harmful users within a network graph dataset. Developed Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) for this task.

Neura-Translate
- Trained and optimized a Recurrent Neural Network on text data from the European Parliament Proceedings Parallel Corpus to translate between Spanish and English.

PROFESSIONAL EXPERIENCE

Oracle Corporation
Member of Technical Staff • Jun 2019 - Present
- Backend software developer & machine learning engineer for Oracle EPM Cloud (part of Oracle Cloud Infrastructure)
- Developed feature to view Oracle Essbase application outlines and identify asynchronous application-to-database anomalies within the cloud application layer of Oracle PBCS / EPBCS
- Implemented and maintained various features for SmartView, the Oracle EPM applications plug-in for Microsoft Office
- Currently developing a predictive performance analytics tool, leveraging a data pipeline of user-defined models and Oracle AutoML to deliver useful insights for customers

AT&T Corporation
Software Engineering Intern • May 2018 - Aug 2018
- Software developer for the Open Video Innovation Team
- Developed iOS application to aggregate and analyze scraped text data from Twitter, Reddit, and Facebook to dynamically graph customer sentiment trends on DirecTV NOW
- Created a pipeline to query, parse, preprocess, and analyze terabytes of event-level DirecTV NOW user sessions; developed various data visualizations and predictive models to determine features associated with positive / negative user experiences

Quant Five Engineering
Software Engineering Intern • May 2017 - Aug 2017
- Full-Stack developer for a team developing Safesign, a web application for e-signing important documents securely
- Backend: Involved in development of 2FA, Document Parsing, Account / Database Management, and Biometric Verification
- Frontend: Involved in development of the Profile Creation Workflow, Email Verification, and Mobile Compatibility

RESEARCH EXPERIENCE

Stanford University, Graduate School of Business
Research Assistant • Mar 2020 - May 2021
- Research project (advisors: Prof. Shoshana Vasserman, Prof. Gregory Martin): develop and optimize computational models to analyze linguistic trends in influential journalism
- Created pipeline to parse, aggregate, preprocess, and encode XML data for millions of articles stored in the Newsbank corpus
- Leveraged BERT Entity Recognition, Word2Vec, and other NLP techniques to extract synthetic features from text data
- Utilized document frequency matrices, word embeddings, dynamic LDA influence models, Time Series Analysis, gradient boosting machines, and recurrent neural networks to classify impact scores for works of investigative journalism

UC Berkeley, Division of Data Sciences
Research Assistant • Jan 2018 - May 2018
- Research project: build a web application to analyze data on student majors and classes for over 6000 graduating seniors
- Optimized hyperparameters for t-SNE clustering models; utilized D3.js to create a widget to query and visualize clusters representing anonymized students with various backgrounds
- Configured Amazon S3 database to periodically gather and store precomputed t-SNE output clusters to improve runtime efficiency by over 300% for the data visualization workflow