

# Anish Saha

Software Engineer II @ Amazon.

+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 three 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 · May 2021 - Aug 2022 · GPA: 3.95  
Masters, Computer Science

**Stanford University**  
Stanford, CA · Sep 2019 - Mar 2021 · GPA: 3.85  
Graduate Certificate, Artificial Intelligence

**University of California, Berkeley**  
Berkeley, CA · Aug 2015 - May 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, AWS SDK / Boto3, Oracle Essbase

### Software Tools

Git, Bash, XCode, Android Studio, Tableau, NewRelic, Conviva, Anaconda, Jupyter Notebook, Oracle Smart View, Figma

## PROJECTS

### ClinicalBERT

- Reproduced and optimized a modified BERT pipeline, improving upon Bag-of-Words, BERT, and Bi-LSTM baselines for predicting hospital readmission throughout a dataset of clinical notes from over 40,000 patients (MIMIC-III)

### Neura-Translate

- Engineered and trained a Recurrent Neural Network (RNN) on data from the European Parliament Proceedings Parallel Corpus to translate between Spanish and English.

## PROFESSIONAL EXPERIENCE

### Amazon.com Inc.

#### Software Development Engineer II · Oct 2022 - Present

- Backend engineer for IMDb Growth & Analytics (Amazon Ads)
- Implemented data-driven SEO features to improve IMDb's page rank; helped decrease number of indexing errors by 91% and increased search traffic by 8% month-over-month
- Productionized engagement features; leveraged anonymized user data, proprietary professional network data, and genre cohort models to deliver optimized recommendations

### Oracle Corporation

#### Member of Technical Staff · Jun 2019 - Oct 2022

- Backend / Machine Learning engineer for Oracle EPM (OCI)
- Developed a utility to view Oracle Essbase application outlines and identify asynchronous application-to-database anomalies from the cloud application layer of Oracle PBCS / EPBCS
- Implemented and maintained various features for SmartView, the Oracle Cloud EPM applications plug-in for Microsoft Office
- Developed a predictive analytics tool, leveraging user-defined models and Oracle AutoML to deliver insights to customers

### AT&T Corporation

#### Software Engineering Intern · May 2018 - Aug 2018

- Full-stack engineer for the Open Video Innovation Team
- Developed iOS application to aggregate and analyze scraped data from social media to graph customer sentiment trends
- Created a pipeline to query, parse, preprocess, and analyze terabytes of event-level DirecTV NOW user sessions; developed predictive models to derive insights from 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

## RESEARCH EXPERIENCE

### Stanford University, Graduate School of Business

#### Research Assistant · Mar 2020 - May 2021

- Collaboration with Stanford NLP Group (advisors: Prof. Martin; Prof. Vasserman; Prof. Manning): develop and optimize various models to analyze trends in influential journalism over time
- Created pipeline to parse, aggregate, preprocess, and encode XML data for millions of articles stored in the Newsbank corpus
- Leveraged BERT / GPT Entity Recognition, FastText, and the latest Stanford CoreNLP methodologies to extract features
- Utilized document frequency matrices, LDA influence models, and deep, multi-layer feed-forward neural network models 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 with Bayesian optimization; used D3.js to query and visualize clusters of anonymized students of 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