Anish Saha

Member of Technical Staff @ Oracle.
Research Assistant @ Stanford.

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

PROFILE

Graduated from UC Berkeley in 2019 with a double major in Applied Mathematics and Data Science. Ambitious, versatile developer currently seeking to further my technical knowledge in the fields of data science, machine learning, and artificial intelligence.

EDUCATION

Stanford University
Stanford, CA • 2019 - Present • GPA: 3.85
Artificial Intelligence Professional Graduate Certification Program

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

TECHNICAL SKILLS

PROGRAMMING LANGUAGES
Python, Java, R, Swift, Objective-C, GoLang Javascript, MATLAB, C / C++, SQL, Bash, HTML, CSS, PHP

FRAMEWORKS & TECHNOLOGIES
Git, React.js, Numpy, Pandas, Scikit-Learn, Apache Spark, Scipy, TensorFlow, Pytorch, Google Maps, Google OCR, Google Cloud Natural Language, Matplotlib, NetworkX, Tweepy, PRAW, Facebook Graph, Pytorch, 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.

SNAPCHAT 2.0
- Built a minimalistic iOS SnapChat emulator. Implemented features for sharing images, account management, and notifications.

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 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 INC.
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
Research Assistant • Mar 2020 - Present
- 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 article text data
- Currently evaluating Logistic Regression, Gradient Boosting Machines, Random Forests, Recurrent Neural Networks, and other predictive models to develop an effective classifier for determining the political impact of investigative journalism

UC BERKELEY DIVISION OF DATA SCIENCES
Undergraduate 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 clusters representing anonymized students, grouped by various features
- 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