Anish Saha

Sr. Software Engineer @ Oracle.

+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 five 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
Masters, Computer Science

Stanford University
Stanford, CA · Aug 2019 - Dec 2024
Continuing Studies, Artificial Intelligence

University of California, Berkeley
Berkeley, CA · Aug 2015 - May 2019
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, Scala, Bash, SQL, HTML, CSS, PHP

Frameworks / Libraries
Numpy, Pandas, Scikit-Learn, NLTK, SpaCy, TensorFlow / Pytorch, OpenLLM, FastAPI, Django, Apache Spark / Kafka / Airflow, AWS SDK / CDK, Essbase, Node.js, React.js

Software Tools
Git, Bash, XCode, Android Studio, Tableau, NewRelic, Conviva, Anaconda, Jupyter Notebook, Figma, Docker / Kubernetes

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

Oracle Corporation
Sr. Software Engineer · Sep 2024 - Present
· Backend / ML engineer for Oracle Health & AI (OHAI)
· Working on building out a Generative AI Platform for the Oracle Clinical Digital Assistant application, leveraging EHR data and conversational AI to deliver insights for physicians

Amazon.com Inc.
Software Engineer II · Oct 2022 - Jun 2024
· Backend engineer for IMDb Growth & Analytics (Amazon Ads)
· Involved in initiatives to design and develop services that drive user engagement and reduce AWS cloud infrastructure costs
· Developed services to monitor and improve SEO performance and notification engagement; increased search traffic by 8%, improved CTR by 91%, and decreased AWS costs by 15% YoY
· Productionized a new AI-based email notification system leveraging anonymized user data, time-sensitive trends, and genre cohort models to deliver optimized recommendations

Oracle Corporation
Software Engineer · Jun 2019 - Oct 2022
· Backend / ML engineer for Oracle Cloud EPM Platform (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
· Developed a predictive analytics tool, leveraging user-defined models and Oracle AutoML to deliver insights for 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, process, and deliver predictive insights based on billions of DirecTV NOW user sessions

RESEARCH EXPERIENCE

Stanford University, Graduate School of Business
Research Associate · Jan 2020 - May 2021
· Collaboration with Stanford NLP Group (advisors: Prof. Martin, Prof. Vasserman, Prof. Manning): develop and optimize various models that 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
· Experimented with RNNs, LLMs, gradient boosting, transfer learning, and dynamic LDA influence models to classify and predict impact scores for works of investigative journalism

UC Berkeley, Division of Data Sciences
Research Associate · 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