Abhishek (Adam) Divekar

adivekar@utexas.edu • Google Scholar • GitHub • LinkedIn

Education

The University of Texas at Austin

Master of Science in Computer Science (GPA: 4.0) | Fall 2020 - Spring 2023 (expected)

- CS383C - Advanced Linear Algebra for Computing
- CS394D - Deep Learning
- CS391L - Machine Learning
- CS388 - Natural Language Processing

Veermata Jijabai Technological Institute

Bachelor of Technology in Information Technology (GPA: 8.74 / 10) | 2013 - 2017

Undergraduate thesis: Machine Learning for Anomaly-based Network Intrusion Detection, advised by Dr. Mahesh Shirole.

Experience

Amazon

Applied Scientist - 2 | Oct 2020 – Present, India

- Developed AutoML framework "DRiP" to iteratively maximize AutoML performance while restricted by a user-defined cost budget. Paper accepted at Amazon Machine Learning Conference 2021 (∼750 submissions, acceptance-rate 33%).
- Proposed "GROK" ranking metric to judge the quality of text augmented by Machine-translation and Abstractive Summarization models (BART, T5, PEGASUS). On product-classification tasks, GROK required ∼70% fewer augmented samples to achieve performance of top-beam and rule-based augmentations (Synonym Replacement, Random Deletion, etc).

Research Engineer - 1 | Oct 2019 – Sep 2020, India

- Fine-tuned and deployed AmaBERT (BERT pretrained on Amazon product text) to classify products across 10,000+ browsable categories. Fixed ∼8.5 MM products, improving categorization precision from 62% to 90%. Built automatic model re-training workflow using Apache Spark and HuggingFace.
- Developed low-latency FastText Docker containers to predict UNSPSC codes. Used to correct UNSPSC for ∼100 MM products on amazon.com at 95% precision & 95% recall.

Software Development Engineer - 1 | Jul 2017 – Sep 2019, India

- Built purchase-authentication used by all Kindle devices in Europe. Launched secure Kindle-to-mobile MultiFactor authentication using SMS & Email notifications, CSRF tokens and server-side caching.

Veermata Jijabai Technological Institute

Research Assistant | Apr 2017 – Jun 2017, India

- Summer research at Dr. Mahesh Shirole's lab. Identified class-imbalance vulnerabilities in popular network intrusion dataset and proposed new dataset to handle underrepresented attacks. Paper accepted for Oral Presentation at IEEE ICCCS 2018.

Publications

All publication venues (including industry) follow a double-blind peer-review process.

Conferences

- Abhishek Divekar, Mudit Agarwal and Nikhil Rasiwasia. (2021). **Unsupervised text augmentation using Pretrained Paraphrase Generation.** *(Preprint).*
- Abhishek Divekar*, Gaurav Manchanda*, Prit Raj, Abhishek Das, Karan Tanwar, Akshay Jagatap, Vinayak Puranik, Jagannathan Srinivasan, Ramakrishna Nalam and Nikhil Rasiwasia. (2021). **Squeezing the last DRiP: AutoML for cost-constrained Product classification.** *Proceedings of the 9th Annual conference of Amazon Machine Learning (AMLC). Conference acceptance-rate of 33% out of ∼750 submissions.*
- Abhishek Divekar, Meet Parekh, Vaibhav Savla, Rudra Mishra and Mahesh Shirole. (2018, Oral Presentation). **Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives.** *Proceedings of the 3rd IEEE International Conference on Computer and Communication Systems (IEEE ICCCS).* arXiv:1811.05372

Workshops

- Abhishek Divekar, Vinayak Puranik, Zhenyu Shi, Jinmiao Fu and Nikhil Rasiwasia. (2021, Oral Presentation). **LEAP: LEAf node Predictions in the wild.** *2nd Amazon Selection and Catalog Services Applied Science Workshop.*
- Andrew Borthwick, Abhishek Divekar, Nick Erickson, Fayaz Ahmed Farooque, Oleg Kim, Nikhil Rasiwasia, Ethan Xu. (2021, Oral Presentation). **CPP MultiModal AutoML Corpus and Benchmark.** *Workshop on MultiModal Learning and Fusion, Amazon Machine Learning Conference 2021.*
- Gaurav Manchanda*, Abhishek Divekar*, Prit Raj, Akshay Jagatap, Vinayak Puranik, Jagannathan Srinivasan, Ramakrishna Nalam and Nikhil Rasiwasia. (2020). **Entity Prediction Service: a configurable, end-to-end AutoML system for Product Classification.** *Workshop on Automated Machine Learning, Amazon Machine Learning Conference 2020.*

Invited talks

- Presented work on DRiP AutoML framework at Amazon Research Days 2021 conference.

Awards

First place, Amazon Chennai ML Challenge, 2017

- Kaggle-style competition with ~300 participants. Task was to predict cancellation of KindleUnlimited subscriptions from user purchase & reading history. Transformed time-series problem into classification, thereby increasing dataset from 150k to 3.5 MM samples. Trained RandomForest to predict cancellations with 89.7% F-1.

Skills

Languages

Proficient *(100K+ lines in production)*
Python • Java

Familiar *(Used in work projects)*
Spark SQL • C++ • JavaScript
HTML & CSS

Tools

Data Science
PyTorch • NumPy • Pandas • Dask
Apache Spark • HuggingFace • LaTeX

Software Development & MLOps
Git • Docker • Streamlit

Computer Science

Machine Learning
Automated Machine Learning (AutoML)
Deep Learning
Natural Language Processing
Computer Vision

Projects

Asking the Right Questions: Question Paraphrasing Using Cross-Domain Abstractive-Summarization and Backtranslation *(Abhishek Divekar, Alex Stoken)*

Final project for graduate course CS388 Natural Language Processing at UT Austin.

- Used Abstractive-Summarization model PEGASUS for data augmentation in Question-Answering. Compared results to Backtranslation augmentation (Fairseq EN↔DE WMT'19 News), on NewsQA (in-domain) & BioASQ (cross-domain).
- Trained Bi-LSTM with aligned attention, using 300-dimensional GloVE embeddings. Used PyTorch.

Autonomous agents for realtime multiplayer ice-hockey *(Abhishek Divekar, Jason Housman, Ankita Sinha, Alex Stoken)*

Final project for graduate course CS394D Deep Learning at UT Austin.

- Built autonomous agent to play ice-hockey using image signal from SuperTuxKart videogame (similar to MarioKart).
- Trained multi-headed CenterNet model (with U-Net backend), to predict whether hockey puck was on-screen (classification), puck's x-y coordinates (aimpoint regression) and distance from player (regression). Model made predictions in realtime (avg. 18ms on NVIDIA Tesla V100 GPU for 400×300 images).
- Model predictions used by agent-code to either search and "dribble" puck towards goal, or defend against opposite team.

SearchDistribute: an economical Google Search API *(Abhishek Divekar)*

- Tool to gather datasets of search results from Google, Bing, etc. Able to retrieve ~250K results/day using $5/month VPN connection (120x savings compared to Google Search API).
- Built using Python and Selenium to coordinate multiple PhantomJS browser instances, each connected to a VPN proxy.