

# Abhishek (Adam) Divekar

✉ adivekar@utexas.edu • 📄 Google Scholar • 🔗 LinkedIn • 📞 (+91) 9819712190

## 🧰 EXPERIENCE

### AMAZON SCIENCE | MACHINE LEARNING SCIENTIST-II (2020 - Present, India)

- Owned scientific experimentation for AutoML platform “Entity Prediction Service”. Developed ML framework “DRiP” to iteratively maximize AutoML performance while restricted by a user-defined cost budget. Paper accepted at Amazon ML Conference 2021 (33% acceptance-rate).
- Devised unsupervised function “GROK” to rank the quality of text augmented by Backtranslation and Abstractive-Summarization models (BART, T5, PEGASUS). On product-classification tasks, GROK-based augmentation requires ~70% fewer augmented samples to achieve performance of Beam-Search and rule-based approaches.

### AMAZON SCIENCE | MACHINE LEARNING SCIENTIST-I (2019 - 2020, India)

- Fine-tuned and deployed AmaBERT (BERT pretrained on Amazon product text) to classify products across 10,000+ browsable categories. Fixed categorization of ~8.5 MM Amazon.in 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. Predicted ~500 MM products on Amazon.com at 95% precision & 95% recall.

### AMAZON | SOFTWARE DEVELOPMENT ENGINEER-I (2017 - 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.

### VJTI, MUMBAI | RESEARCH ASSISTANT, DR. MAHESH SHIROLE'S LAB (Apr - Jun 2017, India)

- Introduced a new dataset for improved detection of underrepresented network-attack classes using Machine Learning. Paper accepted for Oral Presentation at IEEE ICCCS 2018 conference.

## 🎓 EDUCATION

### THE UNIVERSITY OF TEXAS AT AUSTIN | MASTER OF SCIENCE IN COMPUTER SCIENCE (OPTION-III)

2020 - 2022 (Expected) • GPA: 4.0

Graduate Coursework:

- Natural Language Processing (CS388): A grade (112%)
- Machine Learning (CS391L): A grade (93%)
- Deep Learning (CS394D): A grade (105%)
- Advanced Linear Algebra for Computing (CS383C): A grade (97%)

### VJTI, MUMBAI | BACHELOR OF TECHNOLOGY IN INFORMATION TECHNOLOGY

2013 - 2017 • GPA: 8.74 (out of 10)

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

## 📄 PUBLICATIONS

### 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: 33% (~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)*.

## 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*.
- 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*.
- 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*.



## INVITED TALKS

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



## HONORS AND 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

Apache Spark • Dask • HuggingFace

#### Software Development & MLOps

Git • Docker • CICD • Streamlit

#### Amazon Web Services (AWS)

SageMaker • Lambda • DynamoDB

Step Functions • Elastic Map Reduce

### COMPUTER SCIENCE

#### Machine Learning

Deep Learning

Natural Language Processing

Automated Machine Learning (AutoML)

Computer Vision

#### Math

Probability & Statistics

Linear Algebra

Multivariate calculus



## PROJECTS

### Asking the Right Questions: Question Paraphrasing Using Cross-Domain Abstractive-Summarization and Backtranslation (*Contributors: [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 (*Contributors: [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 (*Sole contributor*)

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