

## Contact

[www.linkedin.com/in/ardivekar](https://www.linkedin.com/in/ardivekar)  
(LinkedIn)  
[scholar.google.com/citations](https://scholar.google.com/citations)  
(Other)

## Top Skills

Retrieval-Augmented Generation (RAG)  
Information Retrieval  
HuggingFace

## Languages

English (Native or Bilingual)  
Hindi (Native or Bilingual)

## Certifications

Online Learning and Optimization  
Advanced Linear Algebra for Computing  
Reinforcement Learning  
Machine Learning  
Case Studies in Machine Learning

## Honors-Awards

Winner (out of 300 participants), Amazon Machine Learning Challenge, Chennai 2017  
[Invited talk] Squeezing the last DRiP: AutoML for Cost-constrained Product Classification

## Publications

SynthesizRR: Generating Diverse Datasets with Retrieval Augmentation  
CorrSynth: A Correlated Sampling Method for Diverse Dataset Generation from LLMs  
Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives

# Abhishek Divekar

ML Science Lead @Amazon || Prev: @UTAustin  
Bengaluru, Karnataka, India

## Summary

I work at Amazon's Central ML department as a Senior Applied Scientist, where my contributions have led to +\$600MM to Amazon's revenue in 2021-2024, multiple first-author scientific publications at Tier-1 / A\* conferences, and open-sourcing 3 codebases using Foundation Models. Since 2022, my work focuses on LLMs, Retrieval-Augmented Generation, Synthetic Dataset Generation, LLM-as-a-Judge, and AutoML. I enjoy mentoring junior Scientists who are starting out in their journey (I no longer mentor software engineers).

### \*\*Scientific Research (selected):\*\*

External: My research has been published at important NLP venues. (1) "SynthesizRR" and (2) "CorrSynth" were two first-author papers at EMNLP Main conference (~21% acceptance rate). At Amazon, we also submit to Amazon Machine Learning conference (AMLC, ~1500 submissions, 30% acceptance rate). Notable publications include two first-author papers in AMLC Main conference: (3) "DRiP" 2021 Poster (4) "SynthesizRR" 2024 Oral.

### \*\*Production projects (selected):\*\*

(1) Projects where I have been a lead contributor have driven over \$550M for worldwide Amazon businesses like Subscribe and Save (<https://www.amazon.com/b?node=5856181011>).  
(2) An AutoML platform I owned led to the production deployment of 1,000+ ML models across 14 Amazon websites (updating attributes of 1 billion+ products).  
(3) My work on launching Amazon's Rufus chatbot in India has been recognised in leading news outlets such as Times of India, The Hindu and The Economic Times. Rufus is now available to all Amazon customers in India.

### \*\*Open-source contributions:\*\*

Entity Prediction Service - a configurable, end-to-end AutoML system for Product Classification

Squeezing the last DRiP: AutoML for cost-constrained Product classification

(1) I led the development and open sourcing of a codebase to aid research in synthetic dataset generation: <https://github.com/amazon-science/synthesizrr>

(2) (Confidential) an upcoming codebase on foundation models. Despite the title "ML scientist", I have pushed ~200k lines of Python code to production for ML solutions. I regularly lead code and design reviews in tight coordination with software engineering teams.

**\*\*External talks / collaborations:\*\***

(1) I have given invited talks at:

(a) Chapman University

(b) AI/MLSys 2022 conference

(c) Amazon ML Summer School 2022

(2) In 2022, I was interviewed as part of a marketing initiative to recruit ML Scientists to Amazon India: <https://youtu.be/15sa6OeIWJg>

(3) Since 2023, I have maintained an academic collaboration with researchers at The University of Texas at Austin.

**\*\*Hiring:\*\***

As an interviewer, I have conducted 125+ interviews for ML Scientists and Software Engineers.

---

## Experience

Amazon

7 years 7 months

Senior Applied Scientist

December 2021 - Present (3 years 3 months)

Bangalore Urban, Karnataka, India

Tier-1 / A\* research papers:

(1) EMNLP 2024 Main conference (20.8% acceptance rate): Abhishek Divekar and Greg Durrett. "SynthesizRR: Generating Diverse Datasets with Retrieval Augmentation".

(2) EMNLP 2024 Main conference (20.8% acceptance rate): Suhas S Kowshik\*, Abhishek Divekar\*, Vijit Malik. "CorrSynth: A Correlated Sampling Method for Diverse Dataset Generation from LLMs".

Amazon ML Conference (annual internal conference with ~1K submissions and 30% acceptance rate):

- (1) SynthesizRR: Generating Diverse Datasets with Retrieval Augmentation  
(Oral paper, ~10% acceptance rate)
- (2) Squeezing the last DRiP: AutoML for cost-constrained Product  
Classification (Poster paper, ~20% acceptance rate)

#### Applied Scientist

October 2019 - November 2021 (2 years 2 months)

Working on applied research projects at Amazon's India Machine Learning research department.

- (a) AutoML: Designed and implemented of data-processing nodes for an internal AutoML platform, contributing 160K lines of Python code.
- (b) Used BERT-Base model further pre-trained on Amazon data to classify products into 10,000+ browsable categories within the Amazon product taxonomy. Increased classification accuracy over existing multi-task FastText model by ~7% (62% to 69%).
- (c) Developed Docker containers to predict ~500MM products on Amazon.com. Developed Depth-first preprocessing which improved latency by 30-50%.

#### Software Development Engineer

August 2017 - September 2019 (2 years 2 months)

- (a) Designed and implemented multi-device purchase flow used by all Kindle devices in Europe. Launched secure handoff from Kindle to another device using SMS/Email notifications, CSRF tokens and server-side caching, protecting Critical customer data.
- (b) Developed new REST APIs and integration test framework for worldwide Tier-1 payments service, used by internal businesses including Kindle, Alexa and Amazon Prime to carry out customer payment flows. Parallelized calls to downstream services using Java's ThreadPoolExecutor, reducing API latency by 25%.

#### Veermata Jijabai Technological Institute (VJTI)

Research Assistant

April 2017 - June 2017 (3 months)

Mumbai, Maharashtra, India

Research Assistant in Prof. Mahesh Shirole's lab. Was first author on full-length paper which appeared in IEEE ICCCS-2018 conference:

"Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives".

#### Barclays

Software Development Intern  
May 2016 - July 2016 (3 months)  
Pune, Maharashtra, India

Developed prototype to optimize Foreign Exchange transaction time from T+2 days to T+120 seconds (99.93% reduction). Used Ripple blockchain and Node.js.

---

## Education

The University of Texas at Austin  
Master of Science - MS (Thesis track), Computer Science · (August 2020 - May 2024)

Veermata Jijabai Technological Institute (VJTI)  
B. Tech in Information Technology, Computer Science · (2013 - 2017)