

Gaurav Arora

MACHINE LEARNING ENGINEER @ HAPTİK, EX - GOLDMAN SACHS

medium.com/@giganticgemmic | contactgauravforwork@gmail.com | goru001.github.io | goru001 | gaurav-arora-23593220

Experience

Haptik (owned by Reliance Jio)

Mumbai

MACHINE LEARNING ENGINEER - 2

July 2019 - Present

- Working with the ML team on fundamental Conversational-AI problems using Deep Learning
- Built the **Intent Detection** System for Haptik's NLU Engine, owning it from Research to Production
- The performance of new system was bench-marked against 10 data-sets, including 3 open source datasets and it was **25% more accurate** than their previous system, with latency under **150ms**
- The new system is underpinned by key concepts like **Transfer Learning**, **Data-Augmentation**, protection against **Adversarial Queries** and is like an **Auto-ML service** for bot developers

Goldman Sachs

Bengaluru

ANALYST

June 2018 - July 2019

- Worked with the UX and Productivity team on analytics for **Desktop Assistant**, which is a productivity tool used by every employee in the firm to perform activities like Call, IM, Email etc
- My work at GS brought to the fore key metrics about Product Health, User Behavior which were missing earlier

Researchshala

Chandigarh

Co-FOUNDER

Nov. 2016 - Mar. 2018

- Researchshala is an online platform to connect professors to research-interns, helping them with their research projects
- Researchshala had more than 40 research projects and 90+ students working under professors from universities like Carnegie Mellon University, IITs etc. during that time

Goldman Sachs

Bengaluru

INTERN

Jan. 2017 - June 2017

- Worked with the team that built and maintains the firm's standard UI Development Framework - the **UI Toolkit**

Open Source contributions

Natural Language Toolkit for Indic Languages - iNLTK

goru001/inltk

DEEP LEARNING LIBRARY FOR DEVELOPERS

Oct 2018 - PRESENT

- iNLTK aims to provide out of the box support for various NLP tasks that an application developer might need for Indic languages
- iNLTK provides **Data Augmentation**, **Sentence Similarity**, **Sentence Encoding**, **Word Embedding**, **Tokenization** and **Text Generation** utilities for low resource 12 Indic Languages
- iNLTK has **30,000+ downloads**, **450+ stars** on GitHub, and has been widely shared and appreciated on Twitter, LinkedIn, Reddit
- The library is backed by **ULMFiT** Language Models which I had trained using **Fastai** and **Pytorch** libraries, producing SOTA LM perplexity and Classification accuracy in 12 Indic Languages

Code With AI

goru001/code-with-ai

DEEP LEARNING TOOL FOR COMPETITIVE PROGRAMMERS

Dec 2018 - Jan 2019

- A Tool which predicts which techniques one should use to solve a competitive programming problem to get correct answer
- Model for this tool is built using **Transfer Learning** by starting from WikiText103 Language Model pre-trained weights, fine-tuning the language model on current data set and then building classifier on top of fine-tuned language model
- This tool has **3000+ users**, **130+ stars** on GitHub and was also appreciated by Jeremy Howard on Twitter

NLP for Sanskrit

goru001/nlp-for-sanskrit

FIRST EVER SOTA TOKENIZER, LANGUAGE MODEL AND CLASSIFIER IN SANSKRIT

Oct 2018 - Jan 2019

- Dataset: Scraped and Cleaned, 22k Sanskrit Wikipedia articles for Language Model, and Shlokas from sanskrit websites for classifier
- Used **ULMFiT** to train Language Model, built Shloka Classifier on top of pre-trained Language Model using **Transfer Learning**
- Similar repositories for 11 other Indic languages, whose models have been used in iNLTK

Honors & Awards

CODING

Mar. 2020	Fast.ai International Fellow , for contributions to Fast.ai forums	Worldwide
Mar. 2019	Fast.ai International Fellow , for contributions to Fast.ai forums	Worldwide
	Top-17% , Human Protein Atlas Image Classification, Kaggle , for developing Deep Learning	
Dec.2018	model which classified mixed patterns of proteins in microscope images. The competition had 2172 teams, but I participated individually and hence had 100% contribution in the 366 th placed solution	Worldwide
Oct.2017	1st Prize,IEEE-Hackathon , for developing chat-bot to help people with emotional decisions in life	PEC
Feb.2016	Top-100 , among 500,000 students in IT-Olympiad,2016.	India
Oct.2016	2nd-Prize, IEEE-Hackathon , for developing an Augmented reality application to help teachers	PEC
Mar.2016	All India Rank-6 , in IEEE Programming League, among over 1200 undergraduate students.	India
Mar.2016	2nd Rank , CodeWars,a competitive-programming event hosted by IEEE,PEC on CodeChef	PEC

ACADEMIC

Nov.2016	Research Scholarship , for Personal Emotional Doctor - Bot	PEC
2014	All India Rank-885 , JEE-Mains, among 1.4 million candidates	India
2014	1st Rank-Opener,PEC , for best JEE-Mains rank among 600 students of the session 2014-2018.	PEC
2014	1 Lakh Scholarship , from CBSE for 96.4% marks in 12th Boards and 10 CGPA in 10th.	India
2014	Letter of Appreciation , from HRD Ministry,Govt. of India for 96.4% in CBSE-12th exams	India
2011	Catch Them Young , Was among the top-40 students selected from tricity by INFOSYS for 2-week Programming-Basics training on their campus	INFOSYS

Education

International Institute of Information Technology (IIIT - Hyderabad)

Bengaluru, India

ADVANCED CERTIFICATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

2018 - 2019

- This was in-depth and comprehensive AI-ML program for tech professionals. I was certified by the program with "Very good" performance.

Punjab Engineering College

Chandigarh, India

BACHELOR OF TECHNOLOGY (B.TECH.), COMPUTER SCIENCE AND ENGINEERING

2014-2018

- Best JEE-Mains Rank in class
- CGPA : 7.77/10

GMSSS-16

Chandigarh, India

INTERMEDIATE

2014

- 12th standard : 96.4%
- 10th standard(CGPA) : 10/10

Skills & Courses

Mathematics

Discrete Structures for Computer Science, Vector Calculus, Fourier Series and Laplace Transform, Operations Research, Mathematics for Machine Learning: Linear Algebra and Multivariate Calculus (Coursera)

Computer Science

Data Structures and Algorithms, Computer Architecture and Organization, OOP, Microprocessor, DBMS, Operating Systems, Computer Networks, Theory of Computation, Artificial Intelligence, Computer Graphics, Mobile Computing, Fastai: Part 1 and Part 2, DeepLearning.ai by Andrew Ng, Deep Learning by Prof. Mitesh Khapra, IIT Madras

Programming & Web Machine Learning tools

Python, C, C++, Javascript, TypeScript, EcmaScript6, AngularJS, ReactJS, Angular4, Webpack, Django with Python Pytorch, Pandas, Numpy, ScikitLearn, SciPy, Fastai

Extracurricular Activity

MOTIVATIONAL-SPEAKER

2014 - PRESENT

- Given more than 10 motivational lectures at schools, college etc to an audience of about 100, talking about *student-life challenges* and *success-hacks*.