

VIMARSH CHATURVEDI

vimarsh.c@gmail.com | github.com/vimarshc | <https://vimarshc.github.io/jupyterblog/>

OPEN SOURCE CONTRIBUTIONS AND PERSONAL PROJECTS

1. [Deep Learning Dots](#): Youtube Channel to connect the various dots in the Deep Learning Field
2. Deep Learning [POCs and experiments](#) with FastAI, TensorFlow-2, HuggingFace with FastAI, TensorFlow for Swift and Transformers.
3. [Personal Blog](#) for Experiments and thoughts around publications in Deep Learning.
4. [Reproduction](#) of [Regularizing RNNs by Stabilizing Activations](#) by David Krueger, Roland Memisevic
5. [Reproduction](#) of [All you Need is a Good Init](#) by Dmytro Mishkin, Jiri Matas
6. [Reproduction](#) of [Analyzing and Improving Representations with the Soft Nearest Neighbor Loss](#) by Geoffrey Hinton
7. Have actively participated in the [fastai](#) community.
8. <https://github.com/encode/django-rest-framework/pull/5131> Patch for Issue raised for bug in filters.

EDUCATION:

1. **Bachelor of Technology** in Environmental Engineering @**Delhi Technological University (formerly Delhi College of Engineering)**

EXPERIENCE

1. **Lead Engineer (Machine Learning) at INFOEDGE. ([naukri.com](#)) (March 2018 - Present)**

Technologies: FastAI, PyTorch, Pandas, Spark, sklearn, PySpark, Spark, Scala, nltk, numpy, mysql, MongoDB, NLP, Deep Learning, Classification, Associative Modelling, LTR

***Overview:** Worked as part of the Naukri Machine Learning Team in conjunction with different teams to enhance existing pipelines.*

a. Job Search

- i. Created a pipeline to store and update **user preferences** for registered users using **apply behaviour**
 1. Created module to measure entropy of users in job freshness, experience and minimum salary.
 2. Populated user preferences for preferred companies using a *Word2Vec*.
 3. Used association mapping to filter user's *STRONG key skills*.
- ii. Revamped **ranking module for Naukri Search** which caters to **2 cr searches** per week to enhance the **NDCG by 50%, CTR by 125% and reduce zero results by 60%**.
 1. Developed models to predict affinity of queries with *functional area, experience and industry*.
 2. Extended location vocabulary by crawling wikipedia, classifying pages related to India using the **AWD_LSTM(pre-trained on Wikipedia)** setup in *FastAI* and using a pre-trained **BERT NER** model to extract locations. Increased the vocabulary by 15%.
 3. Created a module to map locations to geo-locations to extend location queries.
 4. Performed analysis to understand user behaviour and aspects of user behaviour that were not being served efficiently using *Spark*.
 5. Worked in conjunction with the BigData team and frontend team to stitch user behaviour to understand macro-level preferences and redefine several workflows to capture different aspects of user behaviour.

b. Job Crawling

- i. Created a pre-trained language model (unsupervised pre-training) using **AWD_LSTM**, pre-trained on Wikipedia text and fine-tuned on content for jobs. This was inspired from the *ULMFiT*. This language model was used for different tasks.
- ii. Created module for detecting duplication job descriptions using **Locality Sensitive Hashing** to reduce the processing time and increase **F1 score to 0.92 from 0.7**.
- iii. Created a module for **classifying job pages** for auto-discovery while crawling based on **AWD_LSTM(pre-trained on Wikipedia and job content)** with **F1 score of 0.94** for job page classification and **0.95** for Single job or multiple jobs on single page classification
- iv. Created a module for matching company names with their homepage URLs based on Logistic Regression with **Precision and Recall of 0.9 and 0.7**.
- v. Enhanced accuracy of date parser from 62% to 98.6% and built a classifier to identify *posted date* for a job from html content.
- vi. Created a date parser module to increase the dates being parsed by 500%.

c. **Naukri Hackathon 2020:** Used the Transformer based project **T5** to develop a question-answering setup for extracting key skills from a given job description. The setup was further extended to gauge user preferences for search and predict job titles.

d. **Naukri-Gulf Recruiter Search**

- i. Developed classifiers to predict minimum experience, nationality for search query.
- ii. Responsible for performing exploratory analysis to identify broad trends, areas for improvement and enhance understanding with the **Spark** platform

2. **Developer at Hedgehog Lab (Oct 2016 - March 2018)**

Technologies: Python, Django, Django REST Framework, Vue.JS, Java, AWS, TensorFlow

- a. Primarily responsible for creating backend systems using **Django** and designing **RESTful web services and developing database models**
 - i. Web services architecture was based on **generic views and model based serialization**.
 - ii. **Designed and developed relational database models** based on Django's ORM on top of **Postgres**.
 - iii. Implemented POC for integrating RESTful services with **JSONWebToken Authentication** for enhanced security.
 - iv. Developed deployment process based on **Fabric, Puppet**
- b. Contributed to **establishing company wide development practises**.
 - i. Gave session on recommended usage of **django** and conducted **code reviews shortening development time by 15%**.
 - ii. Contributed **enhancements to company's internal library for developing RESTful APIs** resulting in **reduction of development time by 10%**
 - iii. Gave **beginner & advanced sessions on Git** expanding on everyday practices, under-the-hood workings as a content addressable file system. Named as **point of contact** for Git-related queries
 - iv. Led initiative to onboard development team to **concepts of Machine Learning**.
- c. Mentored 3 interns introducing them to concepts and practises of server side engineering.

3. **Developer Intern at Indiqus Technologies (Aug 2016 - Sept 2016)**

Technologies: angular.js, Node, Express, MongoDB, RabbitMQ, Java, Apache Cloud Stack

- a. Primarily worked on front-end application based on **angular.js** for cloud business solutions based on Apache Cloud Stack
- b. Refactored existing codebase to utilise Services and Factories in Angular.JS to **enhance application response time by 30%**.
- c. Gave **session on backend system architecture** on how web services were coupled with Apache Cloud Stack and RabbitMQ to onboard frontend team in an effort to establish full-stack capabilities.

4. **Data Science Intern at GrayMatter (June 2015 - July 2015):**

Technologies: R, MySql, Shiny, ggplot2, plyr

- a. Trained models using Machine Learning algorithms (Random Forest, K-Means clustering and Neural Networks) in R to build the engine of a recommender system to help targeted sale **products leading to increase in sales by 15%**.
- b. Performed ARMA Time Series Analysis to predict future revenue.

5. **Machine Learning Intern at Freshmonk (December 2014 - February 2015)**

Technologies: Python, NumPy, SciPy, OpenCV, PIL

- a. **Kernel-based clustering in LAB color spaces** to do robust graphic-to-stencil screen conversion for screen printing with the objective of eliminating human intervention.
- b. Achieved automatic vectorization with gap-proof layers using Bezier Spline processing in SVG to refine conversion of bitmap images to vector graphics (OpenCV, Numpy, Scipy)

6. **Research Intern at IIIT-D under Prof. Ojaswa Sharma (Sept 2015 - November 2015)**

Technologies: C++, CGAL, Git

- a. Rewrote and optimised module for reconstruction of 3D objects from 2D contours from MATLAB to C++ reducing **processing time by 60%**
- b. Integrated mesh smoothing algorithms to improve the quality of 3D rendered objects
- c. Modularized code through header files and introduced **version control via git**

7. Teaching Volunteer at Don Bosco India (June 2013 - July 2013):

- a. Taught Maths and English to 7th grade students
- b. Achieved an improvement of 23% in average performance in mathematics

COLLEGE THESIS

1. Thesis I: Time series analysis of air pollutants under Dr. Bharat Jhamnani in R
 - a. Utilised **ARIMA** and **TBATS** model for prediction of future levels of air pollutants

TALKS:

1. **How to use __metaclass__ in everyday life @ PyCon India 2017:** Technical talk introducing Python's metaclass feature with examples of how its used by Django, Django REST Framework, Scrapy.
 - a. <http://github.com/vimarshc/metaclass-talk/>
 - b. <https://www.youtube.com/watch?v=tvaEyEPjWes>

LANGUAGES AND TECHNOLOGIES:

Python, FastAI, PyTorch, TensorFlow, Django, sklearn, nltk, Django REST Framework, Javascript, Java, Vue.JS, Angular.JS, C, C++, openCV, MATLAB, Linux, Unix, Git, CMake, Vagrant, Puppet, Ansible, Fabric, Puppet, AWS, MySQL, Postgres, MongoDB