

## EDUCATION

---

**Rajiv Gandhi Institute of Technology (University of Mumbai)**  
Bachelor in Engineering - Department of Computer Engineering

*August 2017 - June 2021*  
*CGPA - 8.98*

## PUBLICATIONS

---

**Spectral Bloom Filters for Client Side Search** (*Final Draft*)

*Parth Parikh, Mrunank Mistry, Dhruvam Kothari, Sunil Khachane*  
11<sup>th</sup> IEEE Annual IEMCON, November 2020, Vancouver, Canada

## INTERNSHIPS

---

**Indian Institute of Information Technology, Allahabad**  
**Affective Analysis of Project Gutenberg's corpus**

*May 2020 - July 2020*  
*Prof. Uma Shanker Tiwary and Mr. Punit Singh*

- Worked with JeMaS and Doc2Vec models to predict the valence, arousal, and dominance space of each section in a book. Incorporated multiple models such as KNN, Adaboost, and Random Forest to classify the emotions present in Emobank and Gutenberg corpora.
- Implemented a variant of the Discrete Emotions Questionnaire to estimate the mood of any reader and recommend books to improve their mood.

**MLH Fellow** (Major League Hacking, New York)

*September 2020 - December 2020*

**BentoML** - Framework for serving, managing, and deploying ML models. (*commiter*)

- [Added support for URL prefix](#) to allow users to run YataiService behind a reverse proxy server.

Authored sections of a [handbook](#) to help fellows navigate large codebases of any open-source project.

## PROJECTS

---

**Sthir - Spectral Bloom Filters for Client-Side Search**

*June 2020 - October 2020*

Innovated a memory-efficient library to perform client-side searching using the probabilistic data structure - Spectral Bloom Filters. Implemented Okapi BM25 for ranking the documents.

**Detecting air pollution hotspots and identifying source trajectories**

*Jan 2020 - Feb 2020*

Developed a model to detect air-pollution hotspots and predict their forward/backward source trajectories using data obtained from satellites such as ERA5 and Sentinel-5P.

**Anaphora Resolution**

*February 2020 - April 2020*

Designed a BERT model to obtain the contextual word embeddings, and encode the mentions of interest. Designed a logistic regression-based model to predict if a mention-pair creates an anaphoric reference.

**Crossword Solver to solve mini New York Times' crosswords**

*December 2019*

Designed a class to guess clues using databases like Wordnet, Moby's thesaurus and using word2vec models like GloVe and positioned the guesses on the crossword-board using Z3 Theorem Prover (SMT solver).

**Indian Movie Recommendation System**

*September 2019 - November 2019*

Developed a movie recommendation engine using content-based and collaborative filtering approaches with a hybrid recommender system optimizing their recommendations. Curated [The Indian Movie Database\(TIMDB\)](#), currently the largest database available for Indian movies.

**Popup Encyclopedia**

*May 2019*

A browser extension aimed to provide word meanings by double-clicking a word. Optimized it to perform faster than state-of-the-art software like Google Dictionary using offline indexing techniques.

## TECHNICAL SKILLS

---

**Programming Languages**

Proficient in Python, C; Prior experience in Java, Bash, GNU Octave

**Libraries/Frameworks**

Django, Flask, Scikit-learn, Pandas, Numpy, NLTK, Scrappy, Tkinter

**Software Skills**

Linux Shell utilities, MySQL, SQLite, Markdown, Git, L<sup>A</sup>T<sub>E</sub>X, AutoCAD

**Web Technologies**

Javascript, HTML, CSS