Last Updated: December $9^{th}$ 2021

# Parth Parikh

pmparikh@ncsu.edu • https://github.com/pncnmnp • (+1) (984) 810-9749

## Summary

Aspiring Software Engineer who is passionately exploring unique problems in computational geometry, data structure analysis, natural language processing, information retrieval, and affective analysis. Experience working in multidisciplinary, multicultural, and remote team environments. Proven interests in teaching by maintaining a blog to communicate my perspectives in the above mentioned domains. Career supported by pursuit of a Master's Degree (Thesis) in Computer Science.

## Education

**North Carolina State University** *August 2021 - May 2023*
Masters in Computer Science - Department of Computer Science
*Courses: Natural Language Processing, Compiler Construction, Software Engineering*

**Rajiv Gandhi Institute of Technology (University of Mumbai)** *August 2017 - June 2021*
Bachelor in Engineering - Department of Computer Engineering *CGPA - 9.19 (83.70%)*

## Publications

**Spectral Bloom Filters for Client Side Search** *Final Draft*
*Parth Parikh et al.*, 2020: $11^{th}$ IEEE Annual IEMCON, Vancouver, Canada

## Technical Experience

**Research Volunteer - NC State Theory Lab** (Advisor: *Prof. Donald Sheehy*) *August 2021 - Present*
Analyzing a new metric tree data-structure (*GreedyTrees*) and computing Hausdorff distances between subsets in a metric space.

**MLH Fellow** (Major League Hacking, New York) *September 2020 - December 2020*
BentoML - Framework for serving, managing, and deploying Machine Learning models. (*commiter*)

- Added support for URL prefix to allow users to run BentoML's storage and deployment model behind a reverse proxy server.
- Authored the test cases, profilers, debugging, and logging sections of a handbook to help fellows navigate large codebases of any open-source project. *Won the midway MLH hackathon for this handbook.*

**Research Intern - Indian Institute of Information Technology, Allahabad** *May 2020 - July 2020*
**Affective Analysis of Project Gutenberg's corpus** Advisor: *Prof. Uma Shanker Tiwary and Mr. Punit Singh*

- Pioneered models to predict and classify emotions of all the passages in popular books from Project Gutenberg's collection.
- Observed and documented issues affecting the emotion analysis domain - such as skewed datasets, difficulty observing the neutral space, and lack of semantic understanding of Modern English in pre-trained transformer models.
- Implemented *Discrete Emotions Questionnaire* to estimate the mood of a reader and suggest books to improve their mood.

## Projects and Personal Research

**Reversing the 20 Questions Game** *September 2021 - November 2021*
Engineered a transformer-based boolean question-answering model wherein the model chooses an entity at random and the human aims to guess this entity by asking natural language queries.

**LuaNLP - Natural Language Processing Library for Lua** *February 2021 - April 2021*
*Presently, it is one of the largest native libraries for statistical NLP in Lua.* Implemented 14 modules: tokenizers, lemmatization, stemming, parts-of-speech tagger, sentiment analysis, keyword extraction, named-entity recognition, and text summarization.

**Sthir - Spectral Bloom Filters for Client-Side Search** *June 2020 - October 2020*
Pioneered a memory-efficient library to perform client-side searching using the probabilistic data structure - Spectral Bloom Filters. This library produces rankings comparable to Lunr.js but with an 85% decrease in memory footprint.

**Detecting air pollution hotspots and identifying their source trajectories** *Jan 2020 - Feb 2020*
Architectured two models using data obtained from satellites such as ERA5 and Sentinel-5P. Provided unique perspectives in predicting ground pollutant concentration using a geographically weighted regression model.
*This project was submitted to the Indian Space Research Organization for Smart India Hackathon, 2020.*

**Anaphora Resolution** *February 2020 - April 2020*
Designed BERT-based model to obtain contextual word embeddings, encode mentions of interest, and predict anaphoric references. Obtained a 94% accuracy on GAP's test dataset.

**Crossword Solver to solve mini New York Times' crosswords** *December 2019*
Capable of probabilistically solving mini New York Times' crosswords (in under 2 minutes) by guessing clues and positioning them on the grid (an NP-complete problem). Positioned them using the Z3 Theorem Prover (SMT solver).

**Indian Movie Recommendation System** *September 2019 - November 2019*
Curated *The Indian Movie Database*, **currently the largest dataset available for Indian movies**, with over 4500 titles released between 1950 and 2019. Crafted content-based, collaborative filtering, and hybrid models for the dataset.

**Popup Encyclopedia** - To provide word meanings by double-clicking a word. *May 2019*
Optimized the extension to perform faster than state-of-the-art software like Google Dictionary using offline indexing techniques.

## Technical Skills

| | |
|---|---|
| **Programming Languages** | Proficient in Python; Prior experience in Lua, C, Javascript, Java, Bash, GNU Octave |
| **Libraries/Frameworks** | NLTK, Flask, Scikit-learn, Pandas, Numpy, Scrapy, Tkinter, Django |
| **Technologies** | Linux Shell utilities, MySQL, SQLite, Markdown, Git, LaTeX, AutoCAD, HTML, CSS |
| **Web Technologies** | HTML, CSS |
| **Operating Systems** | Ubuntu, MacOS, Debian, Tiny Core Linux, CentOS |

## Teaching

- Authored technical blogs on:
  - **How to moderate an online discourse?** - Inspires ideas by examining points like - where do we draw the line? How does moderation appear under a microscope? What are some good moderation architectures? And, is there a winner?
  - **Approximate Distance Oracles** - A scalable pathfinding data structure.
  - **Pseudocode to Code Generation** - Analyzing works published on this topic by Stanford and Berkeley's NLP labs.
  - **General Guide For Exploring Large Open Source Codebases**
- Guest lectured on the topic of Support Vector Machines in my undergraduate Machine Learning class.