# Parth Parikh

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

## Education

**North Carolina State University** - Masters in Computer Science (GPA: 3.99/4) *Aug 2021 - May 2023*

**University of Mumbai** - Bachelor in Computer Engineering (GPA: 9.19/10) *Aug 2017 - June 2021*

## Publications

**Proximity Search in the Greedy Tree** SIAM Symposium on Simplicity in Algorithms, 2023

**Spectral Bloom Filters for Client Side Search** $11^{th}$ IEEE Annual IEMCON, 2020

## Technical Experience

**Production Engineering Intern - Meta (formerly Facebook)** *May 2022 - Aug 2022*

- Build performance measuring and stress testing tools using Fio which is used by the Linux kernel to test I/O subsystems.
- This helped Meta's Warm Storage team in their transition to a new storage architecture.

**Research Assistant - NC State Theory Lab** Guide: Prof. Donald Sheehy *Aug 2021 - Present*

- Analyzing a new data structure *GreedyTrees* and computing Hausdorff distances between subsets in a metric space.

**MLH Fellow - Major League Hacking** *Sept 2020 - Dec 2020*

- Worked on BentoML (framework for managing and deploying machine learning models) and added support for URL prefix to allow users to run BentoML's storage and deployment model behind a reverse proxy server.

**Research Intern - Indian Institute of Information Technology, Allahabad** *May 2020 - July 2020*
**Affective Analysis of Project Gutenberg's corpus**

- Designed 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.

## Projects and Personal Research

**Reversing the 20 Questions Game** *Sept 2021 - Nov 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 with an accuracy of 78.7%.

**LuaNLP - Natural Language Processing Library for Lua** *Feb 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 - Oct 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 satellite data from ERA5 and Sentinel-5P and submitted them to the *Indian Space Research Organization* for predicting ground pollutant concentration using a geographically weighted regression model.

**Crossword Solver to solve mini New York Times crosswords** *Dec 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** *Sept 2019 - Nov 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.

## Technical Skills

| | |
|---|---|
| **Programming Languages** | Proficient in Python; Prior experience in C/C++, Lua, Javascript, Java, Bash, HTML/CSS |
| **Technologies** | Linux Kernel, CI/CD, CUDA, OpenMP/MPI, MySQL, Hadoop, Git, LaTeX, Django |
| **Relevant Courses** | Parallel Systems, Compiler Construction, Operating Systems, Databases, Graph Theory, Natural Language Processing, Software Engineering, Computer Networks |

## Teaching

- Authored technical blogs on topics such as **Approximate Distance Oracles** (scalable pathfinding data structure), **Pseudocode to Code Generation** (analyzing publications in this domain), and **Moderating an Online Discourse**.
- Teaching Assistant for NC State's CSC 442 - *Introduction to Data Science* course during Fall '22.
- Guest lectured on the topic of Support Vector Machines in my undergraduate Machine Learning class.