

PRATHAM SARAF

India · prathammsaraf@gmail.com · +91 8982597809 · [GitHub](#) · [Kaggle](#) · [Medium](#) · [Portfolio](#) · [Linkedin](#)

EDUCATION

Indian Institute of Information Technology

BTech Computer Science

Bhopal

December 2021 - Present

Relevant Coursework

- Object Oriented Programming
- Discrete Structures
- Advanced Engineering Maths
- Database Management System

WORK EXPERIENCE

LancerNinja

May 2023 - October 2023

Internship Certificate [↗](#)

- Led development of end-to-end AI solutions as Machine Learning Intern, leveraging **OpenAI, LlamaIndex, LangChain**. Rapidly prototyped and deployed models using **FastAPI, AWS, GitHub**.
- Partnered with cross-functional teams to architect customized NLP solutions for financial services and healthcare clients. Incorporated **CromaDB**, **PineconeDB** to develop vector database based solutions.
- Utilized **AWS Lambda** to create functional endpoints for Image Inferencing service. Alongside to building frontend with **ReactJS**

Red Positive

November 2022 - January 2023

Machine Learning Intern

- Collected a total of **370 thousand** sentences in 37 different dialects of India with approx **10 thousand** sentence in each language to build a language detection system
- Used **35 Indian languages** for document translation model which took advantage of parallel corpora
- Built speech detection system for local Indian dialects which has an accuracy of over 90%

Acciolbis

September 2022 - November 2022

Internship Certificate [↗](#) *LOR* [↗](#)

- Improved document retrieval time to under 5 seconds through semantic searching and scoring, and optimized memory storage in **Haystack systems**.
- Employed **GCP** to infer models and evaluated various open source text to image models and their variants.

RESEARCH EXPERIENCE

Indian Institute of Technology, Madras

July 2023- Present

Research Intern under Prof (Dr.) Rupesh Nasre

- Implemented **graph convolutional network** model in **StarPlat** by writing optimized functions for forward and backward passes
- Researched academic papers on utilizing StarPlat DSL for efficient parallelization of graph neural networks on multi-core CPU and GPU systems

Indian Institute of Information Technology, Bhopal

November 2022- August 2023

Research Intern under Prof (Dr.) Bhupendra Singh Kirar

- Developed and implemented Neural Networks for the research paper, which involved utilizing convolutional neural networks and transfer learning techniques.
- Analyzed the performance of the classifier using various evaluation metrics, including accuracy and specificity, and presented the findings in the research paper.

PROJECTS

MultiModalGNN-SentimentAnalysis

April 2023 - June 2023

Colab Notebook [↗](#)

- Developed a **MultiModalGNN-SentimentAnalysis** model using variety of models **Graph Attention Mechanism**, BERT, and VGG16, **Graph Neural Network** to perform sentiment analysis on a multimodal dataset which created a graph with **2060892 edges and 39109 nodes**.
- Created a graph structure using the extracted features and achieved an accuracy of 60% The model utilized an **ensemble of neural networks** to leverage both image and text information for accurate sentiment classification.
Skills Used: Python, NetworkX, DGL, OpenAI, GPT-3, Prompt engineering

Open Domain Chatbot

May 2022 - June 2022

Github link for project [↗](#)

- Developed an engaging and conversational **open domain chatbot** using Python, OpenAI's GPT-3 API, and FastAPI. The chatbot can hold **multi-turn conversations** on a wide range of topics while maintaining context and providing witty and insightful responses. Applied advanced techniques like **prompt engineering** to optimize the chatbot's responses.

- Built a web application with authentication around the chatbot using **FastAPI**, **MongoDB**, and modern web technologies. Includes user login via **Google OAuth** to maintain persistent user chat logs across sessions. Leveraged session middleware, database integration, templating, and dynamic UI updates to provide a smooth user experience.
Skills Used: Python, FastAPI, MongoDB, OpenAI, GPT-3, Prompt engineering

Self driving car simulator

August 2022

Github link for project
Kaggle Notebook

- Gathered **145 thousand photos** which consisted of **3 perspectives** using automobile simulation. It included the braking speed, throttle position, and degree of steering wheel rotation
- Trained model based on **Dave 2 system** which comprised of 5 convolutional layers and 3 fully linked layers; Data augmentation was done as well
Skills: Python, CNN, Data collection, Pytorch

Books recommendation system

May 2022 - June 2022

Github link for project
Kaggle Notebook

- Made use of **2.36 million** book's data with 29 features each and **229 million** user interactions with 4 features each to generate curated recommendations. Employed **MongoDB** to handle user registration and recommendation storage
- Suggested top 100 books using **Nearest Neighbour** and the **tf-idf vector** for document searching. The front end is built with **Flask**.
Skills Used: Python, MongoDB, Flask, Sk-Learn

POSITION OF RESPONSIBILITY		
GNU/Linux Users Club		IIIT Bhopal
<i>ML-OPS lead</i>		August 2022 - November 2023
Google Developer Student Clubs(GDSC)		IIIT Bhopal
<i>Assitant AI-ML lead</i>		August 2022 - November 2023
Kratigence		IIIT Bhopal
<i>Core Team Member</i>		August 2022 - November 2023

SKILLS	
Programming Languages:	Python, C++ ,C, R, Mojo
Platforms:	GitHub (with Github Actions), GitLab, Git
Database Management Systems:	MySQL , PostgreSQL, Mongo-DB , Redis
Web Development:	HTML , CSS (Bootstrap and Tailwind) , JavaScript ,React JS
Backend Libraries / Frameworks:	Flask , Django, FastAPI
Machine Learning Libraries:	Tensorflow , Pytorch , Pytoch Geometric , OPEN-AI GYM, DGL , NetworkX, JAX , Numpy , Dask , Pandas
Cloud Platforms:	AWS (EC2, S3, Lambda), GCP (Compute Engine, Cloud Storage, Cloud Functions), Azure (Virtual Machines, Blob Storage, Functions)
Soft Skills:	Leadership, Communication Skills, Organised
Languges:	German , English