



Aakash Thatte

BTech (2020-2024)

Data Science and Artificial Intelligence
Indian Institute of Technology, Bhilai

✉ thatteaakash@iitbhilai.ac.in

🐙 GitHub Profile

🌐 LinkedIn Profile

EDUCATION

•Indian Institute of Technology, Bhilai

2020-2024

Bachelor of Technology, Data Science and Artificial Intelligence

CGPA: 8.61/10

SKILLS

Software Tools: Python, Rust, Linux, Git, Docker, AWS, Elasticsearch, FastAPI, Streamlit

AI/DL tools: Scikit-learn, PyTorch, HuggingFace, OCR frameworks, vLLM

Data analysis tools: Pandas, Plotly, Spark

Miscellaneous: VectorDBs, RAG tools, reading research papers, Open source integrations

Coursework: Linear algebra, Prob and stats, Big Data, ML, NLP, Parallel computing, Adversarial ML

EXPERIENCE

•Needl AI

May 2024 - Present

Machine Learning Engineer

Remote/Bangalore

- Needl is a platform to turn scattered information from multiple sources into intelligent insights.
- As part of the ML team, I am building Needl's AI layer powering multiple offerings like Search, [AskNeedl](#), auto-mated broker report summaries and bulletins etc, majorly for finance clients.

•Open Source

March 2022 - Present

Contributing to relevant OSS projects

Remote

- Added [static embedding model](#) support to [chonkie](#)(a fast, lightweight chunking library)
- Contributed [summarization](#) and [entity-recall](#) metric in RAGAS
- Implemented BLEU and METEOR metric in UptrainAI, created integrations for marqo in DsPy and other tools.
- Expanded distance metrics in [weaviate vector database version v1.15](#) .
- Crafted diverse example applications for weaviate including a multi-modal search, attendance system, QA etc.
- Developed an approx-KNN based graph pre-training method for GNNs to pytorch geometric.

•Google Summer of Code Contributor @ DBpedia

May 2023 - October 2023

Neural Extraction Framework

Remote

- Designed the first end-2-end neural framework for triple extraction from wikipedia.
- Orchestrated smooth NER, entity-linking, and [relation extraction pipeline](#).
- Conducted rigorous benchmarking and evaluation for entity-linking with Redis-DB, DBpedia lookup, and GENRE.

•Needl AI

May 2023 - July 2023

Machine Learning Intern

Remote

- Engineered a system for table extraction from financial documents like annual, quaterly and 10-K reports.
- Evaluated and benchmarked open source and API based solutions.
- Developed internal evaluation for table detection, OCR. Seamlessly integrated and deployed Nanonets and Amazon Textract APIs, allowing one-click extraction of all tables in a PDF for users.

PERSONAL PROJECTS

•K-Medoids in CUDA

May 2024

Implemented K-Medoids algorithm utilizing parallel programming in GPUs with CUDA

- Achived 10x-15x speedup as compared to pyclustering and scikit learn implementations

•Product reviews analysis

Feb 2024

Performed an in-depth analysis of product reviews

- Sentiment analysis, product sentiment summarization, mining and summarizing negative reviews for faster resolutions, question answering on top of reviews to quickly find issues.

•NER for Low resource languages

November 2023

Developed NER models for Bhojpuri, Maithili and Magahi languages

- Explored BiLSTMs and transformers with CRFs for imbalanced NER dataset and beat the best model on Bhojpuri.

EXTRACURRICULAR

•Inter IIT Sports Meet, Represented IIT Bhilai in the Badminton Team

December 2022, IIT Delhi

ACHIEVEMENTS

•AIR 3953 in JEE Adv 2020 I had secured all India rank 3953.

2020