Yash Sorathiya

sorathiyayash1231@gmail.com | 7096223382 | Linkedin

Ahmedabad, Gujarat, India.

SUMMARY

I am a passionate machine learning enthusiast with hands-on experience in Natural Language Processing (NLP), Deep Learning (DL), and Classical Machine Learning. I recently completed a six-month internship, where I worked on a variety of projects, including computer vision and NLP. I enjoy tackling challenging problems in language technology and am excited to contribute to innovative projects that have a real impact.

Education

Dhirubhai Ambani Institute of Information and Communication Technology **2024-2026**

(DA-IICT)

M.Tech in **ICT,** *specialization in* **Machine Learning (ML)**

Adani Institute of Infrastructure Engineering, **2024**

Gujarat Technological University

Bachelor of Engineering Information and Communication Engineering

CGPA: 8.59

Gujarat Higher Secondary Education Board **2020**

12th Grade 83%

Gujarat Secondary Education Board **2018**

10th Grade 89%

Experience

Machine Learning Intern **February 2024 – July 2024**

Vovance Private Limited

Accomplishments:

- Worked on projects spanning Classical Machine Learning, Computer Vision, Natural Language Processing (NLP), Deep Learning (DL), and Large Language Models (LLM).
- Implemented data preprocessing pipelines to clean and prepare datasets for various machine learning tasks, ensuring data quality and reliability.

- Gained practical experience in building and deploying NLP models to solve real-world problems, including text classification and information extraction.
- Created RESTful APIs using FastAPI to integrate machine learning models with web applications, making them accessible to end users.

Technologies Used: Python, TensorFlow, Keras, Scikit-learn, FastAPI, Pandas, NumPy, sentence-transformers, Langchain, LLamaIndex, BeautifulSoup, Selenium.

Summer Internship Project **August 2023**

Accomplishments:

- Successfully designed, developed, and deployed a fully functional Note-Keeping App using the MERN stack (MongoDB, Express.js, React, Node.js).
- Implemented robust user authentication and authorization mechanisms, including password hashing and token-based authentication, to ensure secure user access.
- Deployed the backend server on Cyclic and the frontend on Vercel, ensuring global accessibility and optimal performance under varying loads.
- Integrated real-time collaboration features, allowing multiple users to work on the same note simultaneously, enhancing the user experience and productivity.

Technologies used: React.js, Node.js, Express.js, MongoDB, JWT, Bcrypt

Skills

Technical Skills: Python, Machine Learning, NLP, Deep Learning, C/C++, Node.js, Express.js, MongoDB, GitHub

Soft Skills: Problem Solving, Quick Learner, highly adaptable, Detail Oriented

Languages: English, Hindi, Gujarati

Projects

Major Project

XML Changelog Bot API

Developed a chatbot API capable of answering queries based on web content fetched from sitemap XML files.

- Implemented an endpoint to upload and process XML URLs, utilizing BeautifulSoup for XML parsing and text extraction from web pages.
- Utilized Langchain for document management and Chroma for storing and retrieving vector embeddings, enabling efficient similarity searches.
- Integrated OpenAI's ChatGPT model for generating context-aware responses based on user queries, utilizing a conversation chain for maintaining context.
- Created a middleware for bearer token authentication to secure the API endpoints, ensuring only authorized users can access the functionalities.
- Enabled CORS to allow access from different origins, facilitating seamless integration with front-end applications.

Technologies used: FastAPI, Langchain, OpenAI API, Chroma, BeautifulSoup, Pydantic, CORS, JWT.

Minor Projects

MCQ.ai : Automated Multiple-Choice Question Generation

Developed an API to generate multiple-choice questions (MCQs) from user-provided documents (PDFs).

- Built a backend application using FastAPI, allowing users to upload PDF documents for processing.
- Utilized Langchain and OpenAI API to extract context from the uploaded documents and generate relevant MCQs.
- Implemented a flexible system where users can specify the number of questions, total options per question, and the number of correct answers, enhancing customization.
- Employed Chroma for effective indexing and retrieval of relevant document sections, improving the accuracy of generated questions.
- Conducted thorough testing to ensure the generated questions were relevant and clear, receiving positive feedback from initial users.

Technologies used: Python, FastAPI, Langchain, OpenAI API, Chromadb, Pydantic, JSON

Semantic Search Engine Using NLP and FAISS Indexing

Developed a semantic search tool utilizing **Sentence Transformers** and **FAISS indexing** for a dataset of 10,000 Dutch words.

- Enabled more effective retrieval of relevant information by implementing a model that captures the contextual meaning of search queries.
- Created interactive demo using Streamlit to demonstrate the search capabilities.
- Integrated advanced indexing techniques to optimize search performance and reduce response times.
- Conducted comparative analysis on the effectiveness of various embedding models to identify the best approach for the dataset.

Technologies used: Python, Pandas, Sentence Transformers, FAISS, Streamlit

Achievements

SSIP Hackathon **Oct,2022**

Accomplishments

- Successfully participated in the SSIP hackathon, clearing the regional round.
- Utilized technologies such as Figma, React to create an interactive and user-friendly web application.
- Collaborated effectively with team members, demonstrating strong teamwork and problem-solving skills.
- Proposed future development recommendations to further improve the project's functionality and impact on Rajkot tourism.

Certifications

Introduction to Machine Learning	NPTEL Online Certification
CS50's Introduction to Programming with Python	Opencourseware
Understanding Visual Code and Machine Learning	MLSA Event