Payal Sonawane

payalsonawane.rmdssoe.it@gmail.com | +91 9096031186 | LinkedIn

Profile

Data Science and AI Engineer with a strong foundation in machine learning, NLP, and large language models. Experienced in developing scalable solutions using Python, SQL, and modern AI technologies. Proven track record of building impactful applications, including phishing detection systems and AI-powered QA platforms.

Education

RMD Sinhgad School Of Engineering Pune, India

Bachelor of Engineering in Information Technology — CGPA: 8.76 *Oct 2020 – July 2024*

Technical Skills

Programming Languages: Python, SQL

Web Development: HTML, CSS, Django

Libraries & Frameworks: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

Machine Learning: Supervised/Unsupervised Learning, Clustering, Classification

AI Technologies: Deep Learning, Natural Language Processing (NLP), Large Language Models (LLM), Retrieval-Augmented Generation (RAG), Groq API

Cloud Platforms: Azure

Experience

CodeSpyder Technologies Pvt. Ltd. April 2024 - April 2025

Data Science Intern

- Cleaned and analyzed large datasets using Python and Pandas; built features to improve ML model accuracy.
- Developed NLP applications using LangChain, vector databases (ChromaDB), and Large Language Models (LLMs).
- Created dashboards and data-driven reports to visualize trends in sales, customer behavior, and performance forecasting.
- **Skills:** Python, ML, NLP, Pandas, ChromaDB, Django, MySQL

Elite Software Dec 2022 – Apr 2023

Python Developer Intern

- Built web apps using Django and Flask with MySQL integration. Automated business logic through scripts.
- Collaborated on database design and performance enhancements.
- **Skills:** Python, Django, SQL, Web Development

Projects

Phishing Detection System Using Machine Learning | *Python, Machine Learning*

- Developed an intelligent model for detecting phishing websites using Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Random Forest.
- Utilized domain-based, address-based, abnormal-based, and HTML/JavaScript-based features to classify URLs as phishing or legitimate.
- Designed a pipeline that accepts a suspicious URL, processes it using a pre-trained dataset, and classifies it based on feature extraction and ML-based classification. Implemented a real-time alert system that notifies users whether the entered URL is a phishing attempt or safe to browse.

AI-Powered RAG System for PDF Question Answering | *Python, Flask, LangChain, ChromaDB, LLM, NLP*

- Developed a Retrieval-Augmented Generation (RAG) system to efficiently search and extract answers from large PDF documents.
- Implemented a pipeline that ingests PDF files, splits text into meaningful chunks using Recursive Character Text Splitter, and stores them in a vector database (ChromaDB).
- Leveraged sentence embeddings and semantic search to retrieve the most relevant document sections matching the user's query.
- Utilized an LLM (Llama 3.3-70b) to generate precise answers based on the retrieved context, ensuring privacy and preventing sensitive data leakage. Designed a user-friendly interface enabling users to upload PDFs, input queries, and receive contextual answers in real-time.