MOHD SAHIL

DATA SCIENTIST

Mobile: +91-8130707399
Email: sahil.md.work@gmail.com
Location: New Delhi, India
LinkedIn: linkedin.com/in/mohd-sahil-profile
Github: github.com/sahil-git99

WORK EXPERIENCE

Rinventix Technologies

Data Scientist | Oct 2023 - Present

- Developed **AI Agents** for conversational systems and task automation in chatbots, voice assistants, and telephony systems, enhancing user interaction and operational efficiency.
- Designed advanced **RAG** solutions for AI Applications, incorporating **Query Expansion, HyDE, Hybrid Search,** and **Reranker** techniques, optimizing retrieval from structured and unstructured data sources (e.g., JSONs, PDFs, websites, books).
- Developed **AI Voice Assistants** for real-time conversational applications, creating an end-to-end flow for **Speech-to-Text (STT), Text-to-Speech (TTS), Voice Activity Detection (VAD),** and integration with LiveKit (**WebRTC** framework) for seamless communication.
- Built **recommendation systems** for E-commerce and File Organization Platforms, leveraging machine learning algorithms to personalize user experiences and improve content discovery.

Aidetic Software

Data Scientist | May 2022 - March 2023
.
- Built a **data pipeline** with **Selenium** and **Scrapy,** optimizing it for large-scale scraping using concurrency and multi-threading.
- Designed a fast and efficient vector search system using **OCR, FAISS, and Transformer-based Model.**
- Designed an end-to-end architecture for Intelligent Video Analytics using **NVIDIA Jetson Device**, enabling real-time analysis and insights from video data.
- Built an accurate age and gender detection model using **CNN** (Convolutional Neural Networks), enhancing the system's capabilities for demographic analysis.
- Utilized **OpenCV** for various **computer vision** applications, including **image processing** and **face detection**, contributing to the overall functionality and accuracy of the system.

PROJECTS

Image Search Engine for Fashion Products

- Built an Image Search Engine using the ResNet50 architecture to enable seamless and accurate visual search functionality.
- Managed large and diverse image datasets by leveraging data augmentation techniques (e.g., rotation, flipping, and cropping) to improve model robustness against real-world variations in product images.

Email/SMS Spam Classification

- Built an interactive Streamlit app to efficiently identify and filter spam messages in email/SMS communications.
- Implemented and optimized a Naive Bayes algorithm, achieving high accuracy and reliability through performance metrics like precision, recall, and F1 score.

A/B Testing over Email Marketing Data

- Conducted A/B tests on email campaigns to optimize open rates, click-through rates (CTR), and conversions.
- Applied statistical methods (e.g., hypothesis testing, t-tests) to ensure reliable and actionable insights.

EDUCATION & TRAINING

- Bachelor of Technology in Computer Science Galgotias College (2016-2020)
- Machine Learning Foundations | Coursera (credential)
- SQL for Data Science | Coursera (credential)

TECH STACK

- **GenAI :** LLM, Foundational Models, AI Agents, RAG, Vector Database, Prompt Engineering, Chatbots, AI Assistant, Knowledge Graph, Langchain, Llama Index, Hugging Face
- **Data Science :** Natural Language Processing, Computer Vision, Machine Learning, Deep Learning, Tensorflow, Pytorch, Scikit-learn, OpenCV
- **Dev :** Python, Django, Flask, DBMS, MySQL, PostgreSQL, Elastic Search, Git, Docker
- **Other :** Web Scraping, Scrapy, Selenium