

MOHD SAHIL

DATA SCIENTIST

Mobile: +91-8130707399

Email: sahil.md.work@gmail.com

Location: New Delhi, India

LinkedIn: www.linkedin.com/in/mohd-sahil-profile

Github: <https://github.com/sahil-git99>

WORK EXPERIENCE

Rinventix Technologies

Data Scientist | Oct 2023 - Present

- **NLP Applications** like intent classification, named entity recognition for chatbots & telephony systems.
- RAG systems for question answering based on knowledge banks (PDFs / Docs / PPT etc), Langchain, OpenAI LLMs, Vector Databases.
- Designed and implemented data pipelines, for content extraction from PDFs ensuring proper ingestion of new information.
- **Generative AI Applications** for image generation using advanced prompting mechanisms. Used models like Dalle, Segmind Epic Realism etc.
- **Speech Applications** like Speech to Text & Text to Speech using OpenAI Whisper, Google APIs etc. Further, did NLP analysis like summarization, report generation etc.

Aidetic Software

Data Scientist | May 2022 - March 2023

- REST API Development using frameworks like Django and Flask.
- Designed a fast and efficient multi modal semantic search system using FAISS and sentence transformers.
- Web scraping using Selenium and Python Scrapy to fetch data from websites for automatic web assistant creation. Optimized the entire data collection pipeline for scraping at scale using multi threading.
- Developed an intelligent social media template search system (templates from Canva) utilizing optical character recognition (OCR) for data capture and the universal sentence encoder for advanced sentence embedding.
- Designed an end-to-end architecture for Intelligent Video Analytics using NVIDIA Jetson Device, enabling real-time analysis and insights from video data.

- Built a highly accurate age and gender detection model using Convolutional Neural Networks (CNNs), 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.

Data Science Associate

Almabetter

October 2021 - March 2022

- Got exposure to ML algorithms like Decision Trees, Random Forest, GBDT etc., DL concepts and NLP techniques.
- Handwriting recognition using R-CNNs as part of a team which was building a custom few shot medical doctor prescription parser.
- Used BERT and other distilled versions of BERT for training custom named entity recognition (NER) models to extract names of drugs from prescriptions.

EDUCATION & CERTIFICATIONS

- Bachelor of Technology in Computer Science
- Machine Learning Foundations by University of Washington | Coursera ([credential](#)).
- SQL for Data Science by University of California | Coursera ([credential](#)).

TECH STACK

- **Gen AI** : LLM, Foundational Models, AI Agents, RAG, Vector Database, Prompt Engineering, Chatbots, AI Assistant, Knowledge Graph, Langchain
- **Data Science** : Natural Language Processing, Computer Vision, Machine Learning, Deep Learning, Web Scraping, Tensorflow, Pytorch, Scikit-learn, OpenCV
- **Dev** : Python, Django, Flask, SQL, Elastic Search, Git, Docker