

# Ankur Kumar

Bangalore, Karnataka • +91 9263025968 • ds.ankurx@gmail.com

## Data Scientist

**Data Scientist with 2 years of industry experience** specializing in Python, Machine Learning, and Natural Language Processing. Proficient in MongoDB, TensorFlow, and Scikit-learn, with hands-on expertise in AWS. Adept at designing and deploying scalable, end-to-end data pipelines and machine learning models, leveraging Docker for containerized deployments in production environments.

## WORK EXPERIENCE

**Webin Infosoft Pvt. Ltd. • Bangalore, Karnataka**

**Oct 2023 - Present**

### Data Scientist • Full-time

- Performed data cleaning, transformation, and analysis on large datasets using **Python (Pandas, NumPy)** and SQL, delivering actionable insights for business decision-making.
- Performed web scraping** and extracted structured data from diverse websites using **BeautifulSoup, Selenium, and Chromium WebDriver** for downstream analytics and machine learning tasks.
- Built and optimized ML pipelines with **feature engineering** and **hyperparameter tuning**, improving model accuracy and reducing latency.
- Built Predictive model and implemented NLP-based solutions** for sentiment analysis, topic modeling, and text classification using Hugging Face Transformers and spaCy.
- Implemented an AI-powered chatbot using Python, FastAPI, and NLP libraries to enable intelligent, real-time user interactions. Integrated the chatbot with external APIs and databases for dynamic responses and used embeddings and vector search for contextual understanding. Ensured scalable deployment with Docker and optimized performance for production use.
- Utilized **PyMongo with MongoDB** for efficient query optimization, schema design, and high-performance data access.
- Designed workflow automation pipelines using **n8n**, integrating with APIs and databases for multi-step automation.
- Explored model fine-tuning or inference optimization like quantization and batching
- Deployed models on **Google Cloud** and **AWS**, using **Cloud Run, EC2, and SageMaker**, ensuring scalability and low-latency inference. This aligns directly with their requirement for **cloud-native ML workloads**.

## SKILLS

**Python:** Matplotlib, Nltk, Numpy, OpenCV, Pandas, Scikit-learn, Seaborn, Spacy, Tensorflow

**Framework:** Fast API, Flask

**Web Scraping:** BeautifulSoup, Chromium Webdriver, Scrapy, Selenium

**Machine Learning:** Ada Boost, Decision Tree, Descriptive Modeling, K-Means, KNN, Linear Regression, Logistic Regression, Naive Bayes, Predictive Modeling, Random Forest, Support Vector Machine

**Deep Learning:** ANN, CNN, RNN

**Generative AI & LLMs:** Hugging Face, LangChain, Ollama, OpenAI GPT, Openrouter, RAG

**Databases:** MongoDB, MySQL, Pymongo

**Cloud & Deployment:** AWS, Docker, Github, Google Cloud Platform, nHost

**Visualization Tool:** MS-Excel, Power BI

**Frontend:** CSS, HTML, React.js, Tailwind CSS, Typescript

**Authentication:** JWT, OAuth

**Automation & Orchestration:** Hasura, n8n (workflow automation), Streamlit

**Soft Skills:** Analytical Thinking, Attention to Detail, Problem-Solving, Time Management

## PROJECTS

### Sweeny-AI (AI-Powered Full Stack Chatbot)

- Developed an **AI-driven chatbot application** combining **full-stack architecture** with intelligent automation.
- Used **Nhost** for backend, database, and authentication; integrated **n8n workflows** for **LLM-powered responses** and multi-step automation.
- Deployed frontend on **Netlify**, ensuring responsive design and real-time chat experience.

- Showcased expertise in blending **AI (LLM + automation workflows)** with **scalable full-stack development**.

## Social Media Multi-agent

- Developed an AI-powered multi-agent system that automates daily LinkedIn posts about trending AI technology topics.
- Integrated multiple agents for topic selection, content generation, and LinkedIn publishing using free large language models via OpenRouter API.
- Implemented content beautification for professional, eye-catching LinkedIn posts.
- Automated daily execution and posting using GitHub Actions, with secure management of API credentials.
- Technologies: Python, REST APIs, GitHub Actions, LinkedIn API, OpenRouter, dotenv, requests.

## RAG With Langchain and Ollama

- Built an end-to-end Retrieval-Augmented Generation (RAG) pipeline using LangChain and the LLaMA2 model via Ollama. Ingested and chunked data from PDFs, text files, and web sources, vectorized them using FAISS with Ollama embeddings, and retrieved relevant content for user queries. Designed a custom prompt template and integrated an LLM to generate context-aware responses.

## Movie Recommender System

- Developed a content-based movie recommendation system using Python, scikit-learn, NLP. Implemented Cosine similarity to enhance the accuracy.

## Diabetes Prediction Model

- Developed a **machine learning model** to predict diabetes risk, achieving **94% accuracy**.
- Conducted **data preprocessing, feature selection, and hyperparameter tuning** to enhance model performance.

## Heart Patient Classification Analysis

- Built a **heart disease classification model** using multiple ML algorithms, achieving **91% precision/recall using hyperparameter tuning**. Performed **data analysis, data preprocessing, and feature engineering** to improve model robustness.

## AirBNB Prediction Model

- Conducted an **end-to-end data analysis** of Airbnb listings to identify pricing trends, customer preferences, and geographical factors.
- Built **machine learning models** to predict property prices, increasing pricing accuracy by **89%**.

## Firststore.io – Modern Grocery Web App

- Developed a **full-stack e-commerce platform** with **React 18, Tailwind CSS, TypeScript, and Express.js backend**.
- Designed mobile-style onboarding, product browsing, categories, cart, and checkout flow.
- Deployed frontend on **Vercel**, ensuring high performance and responsive design.

## Task Tracker Website

- Designed and developed a responsive task management application with user authentication and CRUD functionality for tasks.
- Implemented JWT-based authentication and role-based access control for secure user sessions.
- Utilized ExpressJS to build RESTful APIs and MongoDB to store persistent user and task data.
- Integrated frontend using ReactJS with modular components, form validation, and real-time UI updates.
- Followed MVC architecture and industry-standard coding practices to ensure scalability and maintainability.

## EDUCATION

---

### Full Stack Data Science

iNeuron

Apr 2022 - Jul 2023

### B.Tech In Civil Engineering

West Bengal University Of Technology • Kolkata, West Bengal, India

## CERTIFICATIONS

---

### My Sql - Hackerank

### Technology Virtual Program - Deloitte

### Geodata Processing

### Python - Hackerank

Hacker Rank