Kanhaiya Jha

Noida | +91 9304088627 | knj9304@gmail.com | [Linkedin](#) | [github](#) | [portfolio](#)

Professional Summary

Data Science enthusiast with hands-on experience as a **machine learning and Data Science Intern** skilled in **exploratory data analysis (EDA)** and developing **regression**, **classification**, and **recommendation system** models. Proficient in Python, NumPy, Pandas, Scikit-learn, and deep learning frameworks. Strong problem-solving ability with experience in EDA, data visualization and model fine-tuning.

EDUCATION

Indian Institute of Information Technology (IIIT), Manipur 11/2021-04/2025
Bachelor of Technology in Computer Science Engineering **CPI** – 8.92

EXPERIENCE

Feynn Labs | Data Science & Analytics intern June 2024 – Aug 2024
- Performed Exploratory Data Analysis (EDA) on diverse datasets using **Python libraries** such as Pandas, NumPy, and Matplotlib/Seaborn to identify patterns, trends, and generate insights.
- Cleaned and preprocessed data to address missing values, outliers, and inconsistencies, ensuring high-quality inputs for analysis and modeling, resulting in a 30% increase in model accuracy.
- Generated insightful **reports** that translated complex data analysis into actionable business recommendations.

Lagoon Works | Machine Learning Intern Sep 2024 –April 2025

- Designed and implemented a score-based recommendation system to improve personalized recommendations
- Showcased technical skills by utilizing tools and libraries like Python, Scikit-learn, Pandas, Beautiful Soup and Selenium for data collection and automation
- Built an automated Data Collection pipeline for data gathering and preprocessing, reducing analysis time by 30%.

SKILLS

Languages and Tools	: C++, Python, SQL, Git, GitHub, DVC, MLflow
Libraries & Frameworks	: NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, LangChain
Data Science & Machine Learning	: Data cleaning, EDA, Feature engineering, XGBoost, Supervised & Unsupervised algorithms
Deep Learning	**:** PyTorch, ANN, CNN, RNN, LSTM, Transformers
Mathematics	: Statistics, Probability, [Hypothesis Testing](#)
Tools for Data Analysis	: Excel, Power bi

PROJECTS

Retail Analysis & Customer Segmentation| Python, Pandas, NumPy, Matplotlib | [GitHub](#)
- Developed a customer segmentation model using **RFM** analysis performing extensive **EDA** and feature engineering on 500K+ transactions to identify key customer patterns
- Implemented **RFM Analysis** to segment customers based on purchasing behavior, identifying high-value, average and lost customers to enhance retention strategies

Chat with Website | Python, Langchain, Pinecone, Gemini, streamlit | [GitHub](#)
- Developed an end-to-end conversational application that allows users to interact with any website's content by extracting, embedding, and querying its textual data using a chatbot interface.
- Integrated web scraping (**BeautifulSoup**) with **LangChain** and **Pinecone** to enable semantic search and retrieval-augmented generation for accurate, context-aware responses
- Deployed the Application using Streamlit and Google's API providing a smooth and user friendly interface.

CERTIFICATIONS

- Data Science Job Simulation Certificate | [British Airways](#)
- Earned Badge of Python, Problem Solving, SQL(Basic), SQL(Intermediate) | [Hackerrank](#)
- Data Analytics and Python Certificate | Udemy
- Got a scholarship opportunity after 10th