

# Kanhaiya Jha

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

## 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, NLP** and **recommendation system** models. Proficient in Python, NumPy, Pandas, Scikit-learn, NLP 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**  
Bachelor of Technology in Computer Science Engineering

11/2021-04/2025  
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 – Dec 2024

- 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

<b>Languages and Tools</b>	: C++, Python, SQL, Git, GitHub, DVC, MLflow
<b>Libraries &amp; Frameworks</b>	: NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn
<b>Data Science &amp; Machine Learning</b>	: Data cleaning, EDA, Feature engineering, XGBoost, Supervised & Unsupervised algorithms
<b>Deep Learning</b>	: PyTorch, ANN, CNN, RNN, LSTM
<b>Natural Language Processing</b>	: spaCy, NLTK, Hugging Face Transformers(BERT, RoBERTa, GPT), Text preprocessing
<b>Mathematics</b>	: Statistics, Probability, Hypothesis Testing

## PROJECTS

### AI-Powered Font Recommendation System | RoBERTa, Sentence-BERT, Streamlit | [GitHub](#)

- Developed an AI-powered **Font Recommendation System** that suggests fonts based on **text sentiment, emotion, and keywords**
- Leveraged **NLP models** like **RoBERTa** for **sentiment detection** and **YAKE** for **keyword extraction** to analyze textual content.
- Employed **Sentence-BERT** for generating **vector embeddings** and used **cosine similarity** to find the most relevant fonts.
- Built an **interactive web application using Streamlit** to allow users to input text and receive font recommendations in real-time

### Medilens – Medical Report Analyzer | Gemini API, Python, Hugging Face, Google Generative AI, Streamlit | [GitHub](#)

- Developed an interactive web application that simplifies medical reports by using **Google Gemini's** capabilities to generate human friendly summaries from uploaded report images.
- Translates medical summaries into **multiple languages (Hindi, Tamil)** using custom NLP pipelines.
- Predicts probable diseases from report content and suggests preliminary medication recommendations
- Implemented modular architecture with separate components for disease prediction, translation, paragraph generation, and medicine recommendation.

## 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