

# Himanshu Kumar

Final Year - Indian Institute of Technology, Kharagpur  
☎ +91 89358 28600 | ✉ E-Mail | 🔗 LinkedIn | 🐙 GitHub

## EDUCATION

Indian Institute of Technology, Kharagpur

M.Sc. (5Y) Economics with Micro Spl. in Artificial Intelligence and Applications

Kharagpur, India

2020 - 2025 (Expected)

## EXPERIENCE

LLM Trainer

Oct 2024 - Dec 2024

Soul AI

Remote

- Assessed outputs from multiple **LLMs** in both **STEM** and non-STEM fields, focusing on **accuracy** and **correctness**
- Reviewed **code execution** of AI systems, evaluating **tool calls**, **function parameters**, and **API integrations**
- Designed and optimized **prompts** using techniques such as **zero-shot**, **few-shot**, and **chain-of-thought prompting**

Artificial Intelligence Intern

May 2024 - Jul 2024

International Operations AI BI Team - Schneider Electric

Bangalore, India

- Initiated the development of a **Knowledge Bot** to support **SAP ABAP developers**, focusing on AI knowledge
- Researched **open-source LLMs** such as **llama-3-70b**, **gemma-7b-it**, and **mixtral-8x7b** for **SAP ABAP** code
- Designed an **ETL** pipeline to extract data from tables via **SAP OData Services API**, for LLM fine-tuning purposes

Artificial Intelligence Research Intern

Sep 2023 - Nov 2023

Prof. Priyadarshi Patnaik, IIT Kharagpur

Kharagpur, India

- Developed a **Vulnerability to Viability (V2V) Information System**, comprising data from across **35+** countries
- Fine-tuned **llama-2-7b** on **500+** fisheries survey using **PEFT**, achieving significant relevance in fishery queries
- Implemented a custom **RAG** system using **Sentence Transformers** and **FAISS**, indexing **1k+** fishery reports

## PROJECTS

Loan Default Prediction | Predictive Analysis

May 2023 - Jun 2023

- Carried out **EDA** and **feature engineering** on defaulters dataset with over **65,000** records and over **30** features
- Implemented **Random Forest** algorithm; **oversampled data** using **SMOTE** to address the **class imbalance**
- Achieved **F1-score** of **0.81** for the trained **classifier** after **hyper-parameter tuning**, using **RandomSearchCV**

Customer Churn Rate | Predictive Analysis

Jan 2023 - Feb 2023

- Analyzed **price sensitivity** to identify **churn factors** among **SMEs**, improving retention and profitability
- Performed **EDA** to address **skewed consumption patterns** and price-churn correlations for insights
- Improved **Random Forest** accuracy by **15%** with **feature engineering**, achieving **92%** churn prediction accuracy

Airline Reviews Analysis | NLP

Jun 2022 - Jul 2022

- Analyzed **1,000** Skytrax reviews, performed **sentiment analysis** using **VADER**, finding **53.8%** positive sentiment
- Trained **RandomForest** and **XGBoost** on **50,000** rows, achieving **85.09%** accuracy (**AUC: 0.56**)
- Visualized insights with **Word Clouds** and **Pie Charts**; key features were **route**, **origin**, and **duration**

## COMPETITIONS

Data Analytics 2023-24

Mar 2024 - Apr 2024

General Championship

IIT Kharagpur

- Analyzed **85k+** rows and **300+** columns of **SDOH** data across **5** domains, by normalization and **MICE** imputation
- Utilized **K-Means**, **BIRCH**, **Agglomerative Clustering** to group **85k+** census tracts on **SDOH** characteristics
- Developed a **multi-level scoring model**, and a **Gemini-powered** recommendation engine using **few-shot learning**

## TECHNICAL SKILLS

Languages: Python, SQL

Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, BeautifulSoup, NLTK, LlamaIndex, Transformers

Softwares and Tools: Visual Studio Code, Anaconda, PowerBI, Microsoft Excel

Frameworks: PyTorch, TensorFlow, LangChain, OpenAI