# Ankit Kumar

## Data Scientist/ML Engineer

An applied economics post-graduate with a knack for creating data-intensive applications. Adept at tackling end-to-end machine learning, NLP, and computer vision challenges. Skilled in utilizing key technologies like PyTorch, YOLO, Transformers, LangChain, AWS, LLMs, Docker and more to deliver robust solutions.

**Phone: +918757170724** **Email: ank.ak2635@gmail.com** GitHub Portfolio

## SKILLS

- **Languages & DBs:** Python, R, SQL, MongoDB
- **Analysis & Visualization:** Excel, Tableau, Power BI, Matplotlib, Seaborn, ggplot, Pandas, Numpy
- **Libraries:** Scikit-learn, PyTorch, TensorFlow, NLTK, HuggingFace, LangChain
- **Machine Learning:**Machine Learning, Deep Learning, NLP, Computer Vision, Image Processing, Feature Engineering, Web Scraping, MLOps
- **APIs & Version Control:** Flask, Tornado, Streamlit, Git, GitHub
- **Others:** Statistics, Econometrics, OpenCV, YOLO, OCR, AWS, Linux, Cuda, LLMs, Docker

## PROJECTS

### FREELANCE PROJECTS

**1. Data Analysis AI**

Developed an GenAI solution enabling Chat-GPT style interactions for in-depth analysis of domain-specific data.

- Tailored and **fine-tuned the GPT-3.5** LLM with domain-specific data to ensure reliability and accuracy in responses.
- Implemented a **Langchain** agent to automate the generation of pandas codes and query underlying datasets based on user input.
- Enhanced solution to generate dynamic visualizations such as **plots, charts, and tables**, alongside textual insights.

**2. Table Localiser**

Table Detector is a Python program designed for a specific use case. It efficiently processes images containing data tables, accurately localizing the table, header, and columns within.

- Trained the 'yolov8n' object detection model from scratch using custom data, emphasizing a tailored approach for the specific use case.
- Implemented fine-tuning strategies involving data augmentation techniques and hyperparameter tuning to enhance the model's accuracy and localization capabilities.
- The fine-tuned model achieved outstanding results, boasting an accuracy and precision score of 96% and 97%, respectively, with a mAP50-95 of 89%.
- Engineered a Tornado web application to provide an interactive user interface for localizing data tables in new images.

### PERSONAL PROJECTS

**1. Chat with PDFs**

Hosted Python web app enabling users to upload multiple PDFs and engage in ChatGPT like conversations based on the PDF content.

- Designed and implemented a user-friendly web application hosted on **Streamlit**, serving as the primary interface for PDF upload and conversation initiation.
- Utilized the **PdfReader** library to extract textual content from uploaded PDFs, optimizing data retrieval for subsequent processing.
- Implemented the **CharacterTextSplitter** component to segment extracted text into smaller, more manageable chunks.
- Leveraged the **'BAAI/bge-small-en'** model from **HuggingFace** to embed segmented text chunks, resulting in the creation of a comprehensive vector store.
- Utilized **Langchain** and **ChatOpenAI** to dynamically construct conversation chains, enhancing the conversational experience by utilizing the embedded vector store.

**2. Customer Churn Predictor**
A Python-based flask web application focused on predicting customer churn within a targeted company.

- Developed a Customer Churn Prediction Web Application using Python, hosted on **AWS Elastic Beanstalk.**
- Utilized the "**Telco Customer Churn"** dataset from IBM, prioritized business logic for customer retention, **emphasizing a high recall score** to minimize false negative predictions.
- Implemented a comprehensive machine learning workflow, including EDA, feature engineering, and addressing target imbalance.
- Compared eight classification models, fine-tuning the Random Forest model for optimal performance with a **95% accuracy score** and a remarkable **98% recall score**.

## WORK EXPERIENCE

**1. Thomson Reuters** — Bengaluru
**Correspondent: North America Resources** — 07/2022 - 07/2023

- Conducted vigilant monitoring of North American resource stocks encompassing commodities such as oil, gold, and various metals.
- Employed a proactive approach to track and dissect emerging trends within companies that influenced stock movements.
- Performed comprehensive analysis of macroeconomic developments, assessing their potential ramifications on the resource sector.
- Crafted industry-focused news pieces encompassing topics such as deals, quarterly earnings, production statistics, and more.

**2. JPAL South Asia** — Remote
**Data Analyst (Intern)** — 06/2022 - 07/2022

- Spearheaded the cleaning, transformation, and in-depth analysis of raw primary survey data, pivotal for research on behavioral nudges.
- Led and mentoring a team of college juniors, successfully orchestrating the creation of a comprehensive analysis roadmap.
- Engineered a sub-group-wise Nutritional Deficiency Index derived from the dataset, enhancing insights into nutritional trends and disparities.

## EDUCATION

**Christ University** — Bengaluru
Master of Arts in Applied Economics | CGPA: 8.2 / 10 — 2020-2022

## ACHIEVEMENTS

**Research Under Review for Publication: Nutritional Deficiency Index**
Conducted a data-intensive research to assess state-wise nutritional deficiencies in India and devised evidence-based policy recommendations for their mitigation.

**Guest Lecturer: Empowering Data-Driven Research**
As a Guest Lecturer at Christ University, I offered valuable guidance to researchers, sharing expertise on enhancing research projects through data-driven methodologies.

## LANGUAGES

English, Hindi, Nagpuri