# 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 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

## 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**.
- Modularized deployment tasks, including Data Ingestion, Data Transformation, Model Training, **Flask Application** setup, Prediction Pipeline, and successful deployment on AWS Elastic Beanstalk.

**3. Licence Plate Reader**

Vehicle License Plate Recognition using Computer Vision.

- Demonstrated proficiency in utilizing pre-trained **YOLOv8** and **SORT** models to successfully detect and track vehicles within provided video.
- Developed and trained a personalized model to accurately identify license plates, showcasing understanding of machine learning techniques.
- Executed an image processing pipeline to crop and manipulate license plate images, followed by text extraction using the **EasyOCR** library.
- Employed optimization strategies to enhance **text recognition accuracy**, contributing to the creation of a structured data frame for efficient information management.
- Leveraged data interpolation techniques to intelligently fill gaps within the data frame, resulting in a polished video output with smooth transitions and consistent visual quality.

**4. SQL: Danny's Diner Use-Case Resolution**
Leveraged PostgreSQL to successfully address the renowned Danny's Diner case study, showcasing my adept SQL skills.

- Demonstrated expertise in SQL through the application of foundational concepts including SELECT, GROUP BY, and aggregation functions.
- Progressed to more sophisticated techniques such as JOINs, Common Table Expressions (CTEs), CASE statements, DENSE_RANK, and Window functions.
- Skillfully utilized SQL to analyze and extract insights from the dataset, offering actionable recommendations for business optimization.

## 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.

**3. Sri Satya Sai Annapoorna Trust** — Remote
**Application Developer (Intern)** — 05/2021 - 06/2021

- Demonstrated adeptness in application development on the Zoho Creator platform, yielding impactful solutions.
- Engineered the initial iteration of the Employee Attendance application, showcasing technical prowess and problem-solving skills.
- Innovatively proposed and integrated a Leave Tracker application into the Employee Attendance system, enhancing the application's
- functionality and user experience.

## EDUCATION

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

**St. Xavier's College** — Ranchi
Bachelor of Arts in Economics | CGPA: 6.4 / 10 — 2017 - 2020

## 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