Yash Kumar Roy

• 8192090180 • yashkumarroy164@gmail.com • Linkedin • GitHub • Portfolio

Skills

- Python, Machine Learning, NLP, Gen AI, MySql, Excel, Tensorflow
- Data Analysis, Predictive Analysis, Data Visualization

Education

Liverpool John Moores University | U.K. Apr 26

Master of Science : Data Science

IIIT | Bangalore Jun 25

Executive Post Graduation Programme : Data Science & AI

GBPIET | Uttarakhand Aug 19 - Jun 23

B.Tech in CSE

Professional Experience

AI Variant Jul 23 - Mar 24

Data Science Intern

- Designed and deployed machine learning models using Python (Scikit-learn, Pandas, NumPy) to solve real-world business problems, analyzing datasets with over **250,000 records**.
- Created interactive dashboards and data visualizations that improved stakeholder understanding and accelerated decision-making by **30%**.
- Enhanced model accuracy and performance through regression models, feature engineering, and statistical analysis, improving forecast precision by **17%**.
- Delivered predictive analytic solutions that increased resource allocation efficiency by **15%**, reducing operational costs for the client by an estimated **10%**.

Projects

RAG Based Document Question Answering System using LlamaIndex | April 25

- Build a RAG using OpenAI's **GPT-3.5 Turbo** and **HuggingFace Transformers** for vector embedding.
- Utilized **LlamaIndex** for document ingestion, chunking, vector storage, and semantic querying over PDFs.
- Implemented a custom Q&A pipeline capable of natural language querying over document content with LLM-powered responses.

Telecom Churn Prediction | Aug 24

- Built a logistic regression model on a datasets of **7,043 customers** to predict churn.
- Engineered relevant features and handled data Preprocessing using Pandas and Scikit-learn, reducing model noise by **20%**.
- Achieved **93% prediction accuracy**, supporting a churn reduction strategy with the potential to retain **~25% more customers**.

Bankruptcy Risk Prediction | Mar 24

- Developed a Random Forest model on a datasets of **6,000+ firms**.
- Applied outlier handling and hyperparameter tuning (GridSearchCV), improving model performance by **13%**.
- Delivered a robust model with **91% accuracy**, enabling early identification of high-risk companies and reducing potential financial loss exposure.