Om Kuman

Aspiring Data Scientist

omkuman@gmail.com | 8329278508 | Pune | Om Kuman | GitHub | Portfolio Website

EDUCATION

Masters of Computer Science, *Indira College of Commerce and Science* 2024 – present | Pune, India

Bachelor of Computer Science., *Indira College Of Commerce And Science* 2021 – 2024 | Wakad, India
CGPA-8.59

PROJECTS

1) Diamond Price Prediction 2024 – present

- **Developed an end-to-end Diamond Price Prediction system** using **pandas** for data extraction, cleaning, and feature engineering, processing over **50,000 rows** of data from relational databases **MySQL** to ensure high-quality preparation for analysis
- **Performed EDA** with interactive visualizations to uncover patterns and key factors like carat, cut, clarity, and color impacting diamond prices.
- **Implemented and optimized machine learning models** using **XGBoost** and **scikit-learn**, achieving **95% accuracy** and validating results with MAE and RMSE metrics.
- Tracked and managed machine learning experiments using **MLflow** for performance monitoring, logging and reproducibility.
- **Built and deployed a Flask-based web application** for real-time diamond price forecasting, offering users an intuitive interface for predictions.
- **Containerized the application with Docker** and integrated a **CI/CD pipeline** to automate testing, building, and deployment, ensuring scalability and seamless updates.
- **Techbologies Used**: Pandas,Scikit-learn, Docker, Flask, Matplotlib, Seaborn

2) Automation - Machine Learning Algorithm Selection, 05/2023 – 07/2023
Project to Automate Machine Learning Algorithm Selection

- Addresses **time-consuming** task of manually testing every algorithms.
- **Automates** process for efficiency and user-friendliness.
- Reduces **manual effort** in selecting **optimal machine learning algorithm** in less than **20 sec.**
- Provides **streamlined solution** for data scientists.
- **Technology Used:** *Flask, Python, CSS, HTML*

3) Diabetes Risk Prediction - Machine learning 09/2022 – 10/2022

- Obtained dataset from Kaggle.com of **10,000+** rows, performed the necessary **data preprocessing** using **Pandas.**
- Fitted the model using **ML Algorithms**, and obtained the predictions for the risk of diabetes with **97% Accuracy** using **Scikit Learn.**
- **Flask** server was then utilized to interface the model for test users. Used **CSS** and **HTML** for the same.

COURSES

1) Data Science - Python Programming,, 09/2023 – 01/2024
Ethan's Tech | Powered By Wipro DICE ID.

- I've mastered the basic fundamentals of **data science, machine learning, and deep learning,** looking forward to gain knowledge in **NLP ,** and now eager to expand my knowledge in these fields.
- I gained a thorough understanding of the fundamentals of Python programming and, with practice, was able to obtain a good command in python.

HARD SKILLS

Python: Pandas, Matlplotlib, seaborn, sklearn, EDA.

SQL: Big 6 Statements, Intermediate queries

Machine Learning: Random Forest, Xgboost , Decision Tree

Statistics: Descriptive , Inferential.

Tools: Power BI, Jupyter Notebook, VSCode

Database: MySQL

Deep learning: Activation Functions, Optimizers, Loss Functions

CERTIFICATES

3+ intercollege coding competitions — Best Team Award