Gold Price Analysis and Prediction

End-to-End Data Analysis and Machine Learning Project

■ Project Overview

This project analyzes historical gold price data in India from 2019 to 2025 and builds a baseline machine learning model to understand and predict price trends. It demonstrates a complete data science workflow from data cleaning to model evaluation.

■ Problem Statement

Gold prices fluctuate due to various economic and market factors. This project aims to analyze historical price movements and evaluate how well a machine learning model can learn and predict unseen historical data.

■ Dataset Description

The dataset contains monthly gold price data with the following attributes: Year, Month, and Gold Price (INR per 10 grams). The data spans from 2019 to 2025 and is structured for time-series analysis.

■ Data Cleaning

Data cleaning involved checking for missing values, validating data types, and ensuring consistency. A Date column was created by combining Year and Month, and the data was sorted chronologically.

■ Exploratory Data Analysis (EDA)

EDA was performed to uncover trends, seasonality, and volatility in gold prices. Line plots, yearly averages, monthly patterns, and rolling averages were used for analysis.

■ Feature Engineering

Key features such as lag values, rolling averages, monthly price changes, and percentage changes were engineered to capture historical patterns. A target variable was created to predict the next month's gold price.

■ Model Building

Linear Regression was implemented as a baseline machine learning model. A time-based train-test split was applied to avoid data leakage and simulate real-world prediction scenarios.

■ Model Evaluation

Model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Predicted prices were compared visually with actual historical prices to validate model behavior.

■ Key Learnings

This project strengthened understanding of data cleaning, exploratory analysis, feature engineering, time-series validation, and regression modeling. It highlights the importance of correct evaluation before performing future forecasting.

■ Conclusion

The project demonstrates a complete and structured data analysis and machine learning pipeline. It is suitable for Data Analyst and entry-level Machine Learning roles and serves as a strong foundation for future AI-focused enhancements.