Bank Term Deposit Prediction Project

Project Overview
This project focuses on predicting whether a bank client is likely to subscribe to a term deposit. The goal was not just to achieve good performance, but to build a reliable and realistic machine learning workflow that mirrors how such problems are handled in real-world financial settings.

Exploratory Data Analysis (EDA)
The project began with a careful exploration of the dataset to understand customer behavior, feature distributions, and potential data quality issues. Visual analysis helped identify meaningful trends and relationships that could support predictive modeling. A key decision during EDA was removing the *duration* feature from modeling, as it is only available after the call is completed and would otherwise cause data leakage.

Handling Categorical Data
Instead of removing or imputing unknown values, they were retained as a separate category. This approach allows the model to learn from uncertainty itself, which reflects real operational conditions where incomplete information is common.

Preprocessing & Feature Engineering
Appropriate encoding techniques were applied to categorical variables, while numerical features were prepared to ensure compatibility across different models. The preprocessing steps were designed to be consistent, reproducible, and suitable for future deployment.

Modeling Approach
Several machine learning models were trained and compared, including Random Forest, XGBoost, and LightGBM. Each model was carefully tuned using hyperparameter optimization techniques to improve generalization and performance.

Evaluation
Model performance was primarily evaluated using the ROC-AUC metric. The final models achieved an ROC-AUC score of approximately 80%, indicating a strong ability to distinguish between subscribing and non-subscribing clients.

Ensembling & Pipeline
To further enhance performance, a stacking ensemble was implemented by combining Random Forest, XGBoost, and LightGBM models. The complete workflow — from preprocessing to prediction — was encapsulated within a single pipeline, ensuring consistency and ease of reuse.

Model Persistence & Local Demo
The trained models and preprocessing pipeline were saved using joblib for future use. A local demonstration of the application was performed to validate the end-to-end workflow and confirm that predictions behaved as expected under different inputs.

Conclusion
This project demonstrates a complete, industry-aligned machine learning workflow — from thoughtful EDA and leakage prevention to advanced ensemble modeling. While live deployment could further enhance usability, the project is fully complete from a modeling and analytical perspective and is ready for future extension.