



Diamond Price Prediction using Machine Learning

This project aims to predict diamond prices based on physical characteristics using machine learning, offering insights for consumers and jewelers.

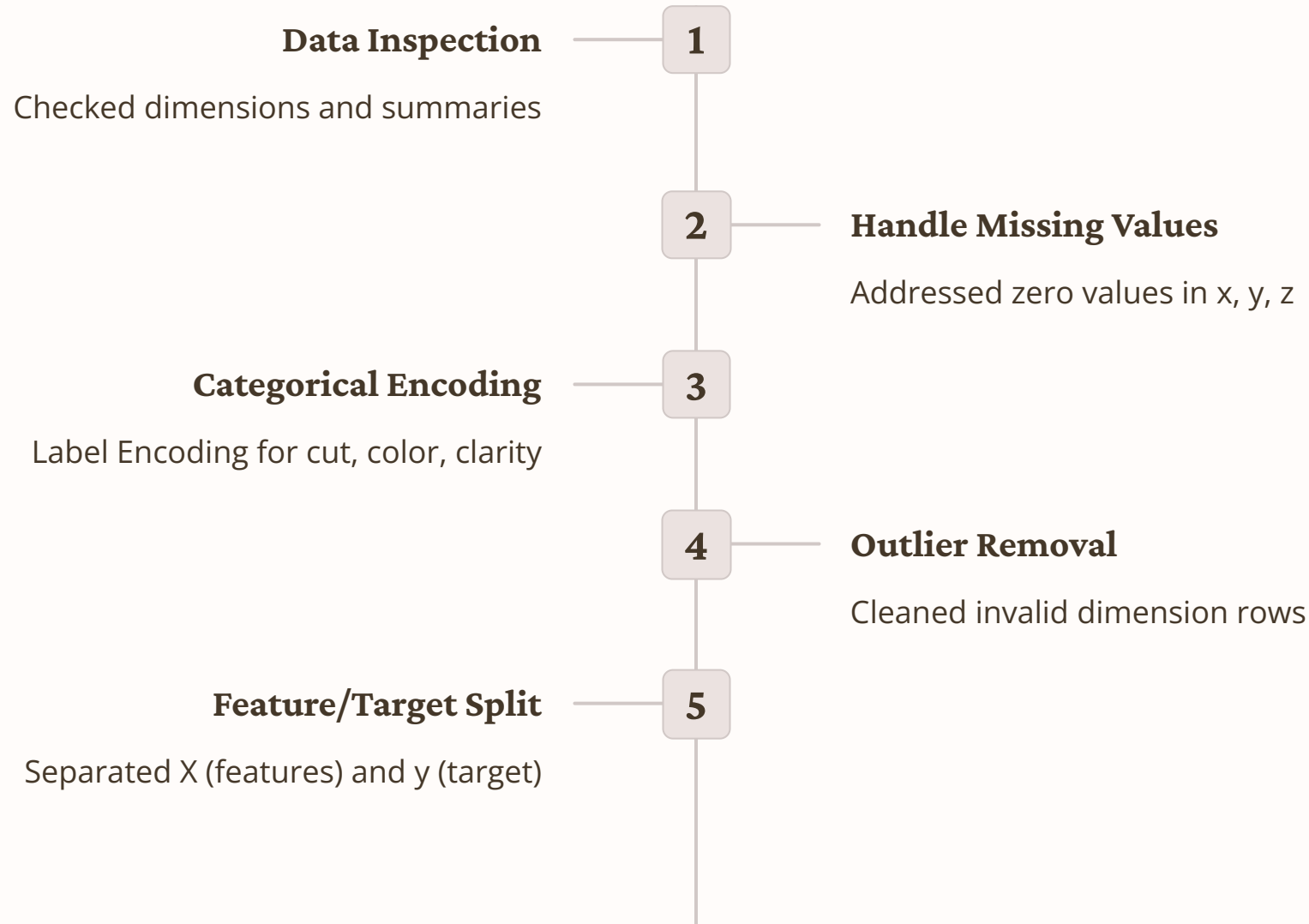
Dataset Overview: Diamonds from Kaggle



Key Attributes:

- Carat: Diamond weight
- Cut: Quality (Fair to Ideal)
- Color: Hue (J to D)
- Clarity: Clearness measurement
- Dimensions: Depth, table, x, y, z
- Price: Target variable

Preprocessing Steps: Data Transformation



Rigorous preprocessing is essential for robust model performance and accurate predictions.

Modeling: Random Forest Regressor



Data Split

Train/test sets created



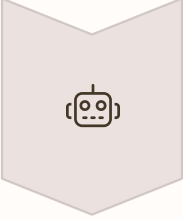
Model Initialization

RandomForestRegressor() from sklearn



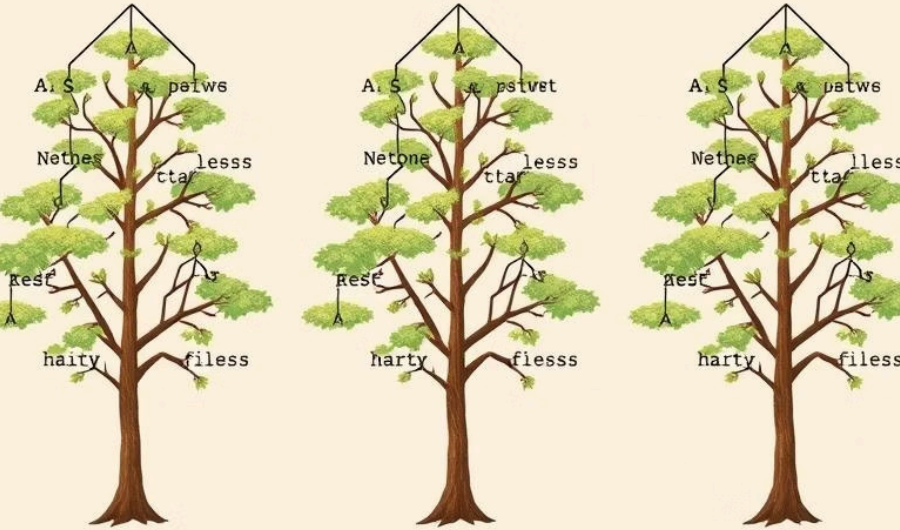
Training

Fitted on X_train, y_train



Prediction

Generated y_pred using X_test



Data fees:
5.442.3.10.1344
50:26,20,1054

Ramoture:s
2.068,94,1018
2.075,94,1010

L:B IPles:
7.062.25.1.3 6:78
1275,:10,08. 5273)

Evaluation: Performance Metrics



R² Score

Coefficient of Determination: prediction accuracy



Mean Absolute Error (MAE)

Average absolute difference between actual and predicted values



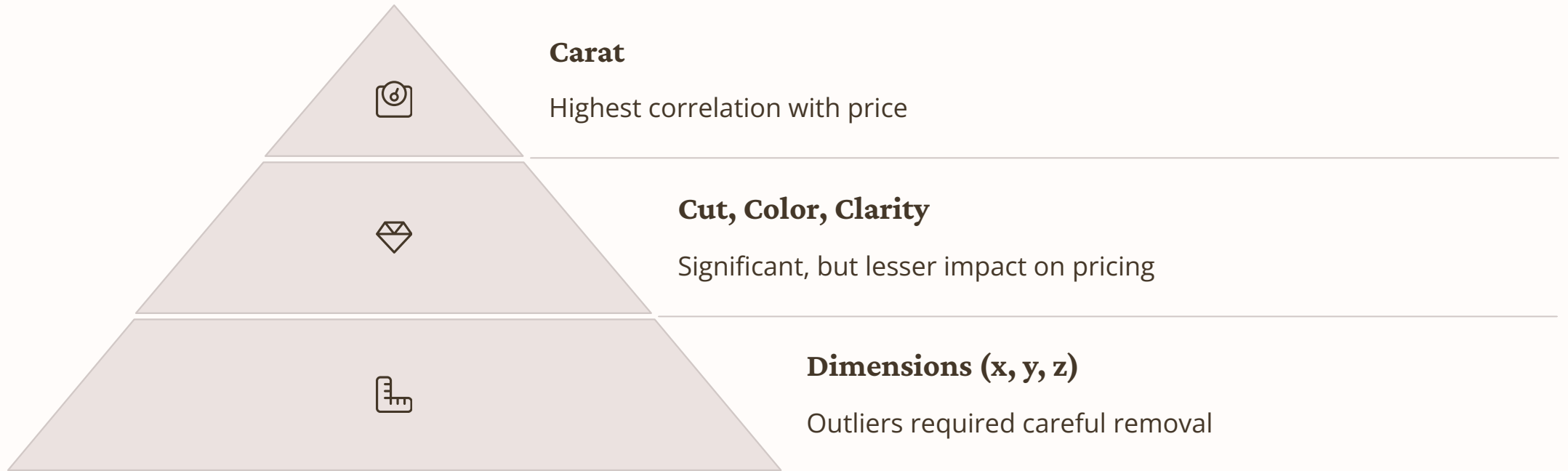
Mean Squared Error (MSE)

Average of squared differences, penalizes large errors

```
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
print("R2 Score:", r2_score(y_test, y_pred))
print("MAE:", mean_absolute_error(y_test, y_pred))
print("MSE:", mean_squared_error(y_test, y_pred))
```

Model achieved a high R² score (>0.95), demonstrating strong predictive ability.

Key Findings: Feature Importance



Random Forest outperformed simpler models due to its robust ensemble learning capabilities.



Conclusion: Practical Applications

Effective Prediction

Random Forest model accurately predicts diamond prices.

Data Quality Impact

Cleaning and encoding are crucial for performance.

Market Value Estimation

Supports consumers and jewelers in valuation.