Techniques of Feature Engineering

🔹 1. Handling Missing Values

- **Imputation**:
    - Mean/Median/Mode substitution
    - Forward/Backward fill (time-series)
    - KNN or regression-based imputation
- **Dropping** rows/columns (if too many missing values).

🔹 2. Encoding Categorical Variables

- **Label Encoding** → Assigns integer labels (for ordinal categories).
- **One-Hot Encoding (OHE)** → Creates dummy variables (for nominal categories).
- **Target / Mean Encoding** → Replace categories with average of target variable.
- **Binary Encoding / Hash Encoding** → Useful for high-cardinality features.

🔹 3. Feature Scaling (Normalization/Standardization)

- **Min-Max Scaling** → Rescales features to [0,1].
- **Standardization (Z-score)** → Centered at 0 with unit variance.
- **Robust Scaler** → Uses median & IQR (robust to outliers).

🔹 4. Transformation Techniques

- **Log Transform** → Reduce skewness (e.g., income, sales).
- **Box-Cox / Yeo-Johnson** → Normalize distribution.
- **Binning (Discretization)** → Convert continuous values into categories (e.g., age groups).

🔹 5. Feature Creation

- **Polynomial Features** → x,x2,x3x, x^2, x^3 etc. for capturing non-linearity.
- **Interaction Features** → Combining features (e.g., product of variables).
- **Domain-specific features** → e.g., extracting year/month/day from a date.

🔹 6. Feature Selection (Reduce Dimensionality)

- **Filter Methods**: Correlation, Chi-Square, ANOVA.
- **Wrapper Methods**: Recursive Feature Elimination (RFE).
- **Embedded Methods**: Lasso (L1), Ridge (L2), Decision Tree feature importance.
- **PCA / SVD / t-SNE** → Dimensionality reduction.

🔹 7. Outlier Handling

- **Z-score / IQR method** → Cap or remove extreme values.
- **Transformation** → Log, Box-Cox.
- **Clipping / Winsorization**.

🔹 8. Time-Series Feature Engineering

- **Date/Time Extraction** → Year, Month, Day, Hour, Weekday, etc.
- **Lag Features** → Previous time steps as features.
- **Rolling Statistics** → Moving average, rolling std, cumulative sum.
- **Seasonality Indicators** → Holidays, weekends, quarter.

🔹 9. Text Feature Engineering

- **Bag of Words (BoW)**
- **TF-IDF (Term Frequency-Inverse Document Frequency)**
- **Word Embeddings (Word2Vec, GloVe, BERT)**
- **N-grams** for context-based features.

🔹 10. Feature Reduction / Regularization

- Drop irrelevant/redundant features.
- Use **regularization (L1, L2)** to avoid overfitting.
- Apply **Autoencoders** for deep learning feature extraction.

Key Idea: Feature Engineering is iterative and depends on:

- Data type (categorical, continuous, text, time-series).
- Domain knowledge.
- Model being used (e.g., tree-based models need less scaling than linear models).