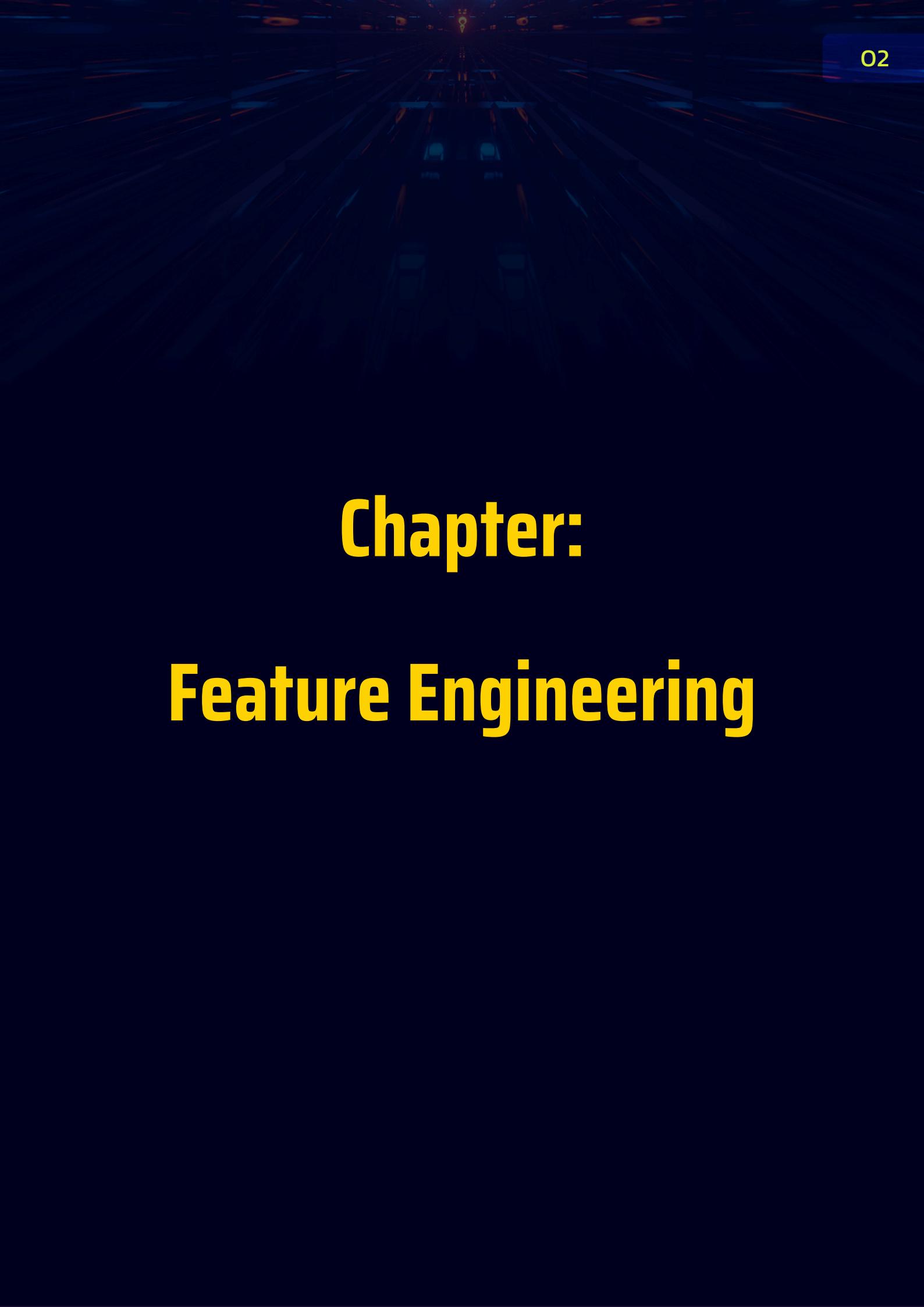




MACHINE LEARNING TAKEAWAYS





Chapter:

Feature Engineering

3 Ways of Doing Feature Engineering

Feature engineering can be done in two ways: business understanding and statistics.

- 1 Feature cleaning & Transformation:** Improving data quality by cleaning and transforming features to make them suitable for analysis
 - a. Handling miss data
 - b. Remove duplicates
 - c. Treat Outliers
 - d. Scaling
 - e. Normalization
 - f. Encoding
- 2 Feature Selection:** Choosing the most relevant features from the dataset to improve model performance and reduce complexity.
- 3 Feature Creation:** Generating new features from existing data to enhance the predictive power of the model.

Feature Selection Using Correlation

- 1** Correlation in feature engineering feature selection: Measures the relationship between features to identify and remove redundant or irrelevant ones.
- 2** Scenarios where you should not use correlation for Feature Selection
 - Non-Linear Relationship
 - Outliers
 - Categorical variables
 - Correlation vs causation.

Feature Selection Using Variance Inflation Factor (VIF)

- 1** Variance Inflation Factor (VIF) measures how much a feature's variance is increased due to multicollinearity, which is when features are highly correlated with each other.
- 2** High VIF means the feature is highly correlated with others, which can cause problems in the model.

VIF: Practical Implementation (Salary Prediction)

- 1 Multicollinearity may not affect the model's accuracy directly, but model interpretation and the stability of coefficients will be negatively impacted.
- 2 Detect features with high Variance Inflation Factor (VIF) indicating multicollinearity. → Eliminate the feature with the highest VIF → Recompute VIF for the remaining features → Repeat the process until all features have acceptable VIF values.