

Shaun Chandraa Shin
23373005

Task-7

In machine learning with scikit-learn, common pitfalls like inconsistent preprocessing and data leakage can degrade model performance. Preprocessing stages must be applied consistently to both training and test data and pipelines help automate this process. Data leakage occurs when test data influences model training, leading to overly optimistic results; it can be avoided by ensuring feature selection considers only training data. Additionally, controlling randomness via random state parameters ensure reproducibility, crucial for consistent results. Using pipelines and managing randomness properly helps avoid these issues and improves model quality.