



MAIN FLOW SERVICES & TECHNOLOGIES

Data Science with Python Internship: Task 5



JULY 8, 2024

SAKSHI RAJESH BHAVSAR

bsakshi2019@gmail.com

Main Flow Services & Technologies (batch: 25May-25July 2024)

Notes from task 5:

Feature Engineering and Selection

Description:

Engineer new features and select relevant features for model training.

Responsibility:

- 1. Generate meaningful features from existing data.**
 - 2. Use techniques like PCA or feature importance to select the most important features.**
- Optimize feature sets for improved model performance.**

Feature Engineering

Definition:

Feature engineering is the process of using domain knowledge to create new features from raw data that make machine learning algorithms work better.

Importance:

- ❖ Enhances model performance by providing more relevant information.
- ❖ Helps in capturing domain-specific knowledge which raw data might not explicitly present.
- ❖ Simplifies model complexity by transforming data into more informative and non-redundant forms.

Feature Selection

Definition:

Feature selection is the process of selecting a subset of relevant features for use in model construction. This reduces the dimensionality of the data, which can lead to better model performance and reduced overfitting.

Importance:

- ❖ Reduces the risk of overfitting by eliminating irrelevant or redundant features.
- ❖ Improves model accuracy by focusing on the most informative features.
- ❖ Enhances model interpretability by simplifying the model.
- ❖ Decreases training time and computational costs.