## **Feature Engineering**

Feature engineering is a crucial step in the process of heart attack prediction analysis. It involves transforming and selecting relevant features (variables) to improve the predictive performance of a model. In this section, we will outline different approaches for feature engineering on the following variables: 'age', 'sex', 'cp', 'trtbps', 'chol', 'fbs', 'rest\_ecg', 'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', and 'target'.

### Approach Name and Procedure

Here's a table that summarizes the approach names and procedures for feature engineering on each of the variables:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Approach Name** | **Procedure** |
| age | Binning | Grouping age ranges into bins for categorization. For example, creating bins like 'Young', 'Middle-Aged', 'Senior'. |
| sex | One-Hot Encoding | Creating binary columns for each gender category (e.g., 'male' and 'female'). |
| cp | One-Hot Encoding | Creating binary columns for each chest pain type (e.g., 'typical angina', 'atypical angina', 'non-anginal pain', 'asymptomatic'). |
| trtbps | Normalization | Scaling blood pressure values to have a mean of 0 and a standard deviation of 1. |
| chol | Normalization | Scaling cholesterol levels to have a mean of 0 and a standard deviation of 1. |
| fbs | One-Hot Encoding | Creating binary columns for fasting blood sugar status (e.g., 'fasting sugar < 120 mg/dl' and 'fasting sugar > 120 mg/dl'). |
| rest\_ecg | One-Hot Encoding | Creating binary columns for different resting electrocardiographic results (e.g., 'normal', 'ST-T wave abnormality', 'left ventricular hypertrophy'). |
| thalach | Normalization | Scaling maximum heart rate achieved to have a mean of 0 and a standard deviation of 1. |
| exang | Binary Encoding | Converting exercise-induced angina into binary values (e.g., 'yes' to 1 and 'no' to 0). |
| oldpeak | Normalization | Scaling ST depression induced by exercise to have a mean of 0 and a standard deviation of 1. |
| slope | Ordinal Encoding | Assigning ordinal values to the slope of the peak exercise ST segment (e.g., 'upsloping' as 1, 'flat' as 2, 'downsloping' as 3). |
| ca | One-Hot Encoding | Creating binary columns for the number of major vessels colored by fluoroscopy (e.g., '0 vessels', '1 vessel', '2 vessels', '3 vessels'). |
| thal | One-Hot Encoding | Creating binary columns for different types of thalassemia (e.g., 'normal', 'fixed defect', 'reversible defect'). |
| target | N/A | The target variable does not require feature engineering; it is already in the desired binary format (e.g., '0' for no heart attack and '1' for a heart attack). |

|  |  |  |
| --- | --- | --- |
|  |  |  |

These feature engineering techniques aim to make the data suitable for machine learning algorithms and improve the predictive performance of the heart attack prediction model. Depending on the specific dataset and the machine learning algorithm you use, you may choose different feature engineering strategies.