FEATURE ENGINEERING

Trevor Lindsay

LEARNING OBJECTIVES

‣ Understand the concept of feature engineering and apply it towards a machine learning problem

WHICH MODEL PERFORMS THE BEST?

ACTIVITY: KNOWLEDGE CHECK

COMPLETE THE FOLLOWING TASKS

EXERCISE

1. Using the raw features in the *titanic_train* dataset, train a logistic regression, random forest, GBDT and KNN model to predict if a passenger survived.

2. Which model performs the best? What is the most appropriate metric for this task?

WHAT IS FEATURE ENGINEERING?

WHAT IS FEATURE ENGINEERING?

‣ Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models

‣ Feature engineering depends on:

  ‣ The performance measures you've chosen

  ‣ The framing of the problem (classification? regression?)\

  ‣ The predictive models you're using

  ‣ The raw data you have selected and prepared

THE SECRET TO FEATURE ENGINEERING

DOMAIN KNOWLEDGE

FEATURE ENGINEERING IN PRACTICE

FEATURE ENGINEERING IN PRACTICE

```python
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

from sklearn import (
    metrics,
    linear_model,
    ensemble,
    neighbors,
)

df = pd.read_csv('../../data/titanic_train.csv')
```