

[Sign up](#)[Sign in](#)

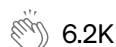
# Fundamental Techniques of Feature Engineering for Machine Learning

All required methods for comprehensive data preprocessing with Pandas examples.



Emre Rençberoğlu · [Follow](#)

Published in Towards Data Science · 14 min read · Apr 1, 2019



6.2K



17





## Introduction

What is a feature and why we need the engineering of it? Basically, all machine learning algorithms use some input data to create outputs. This

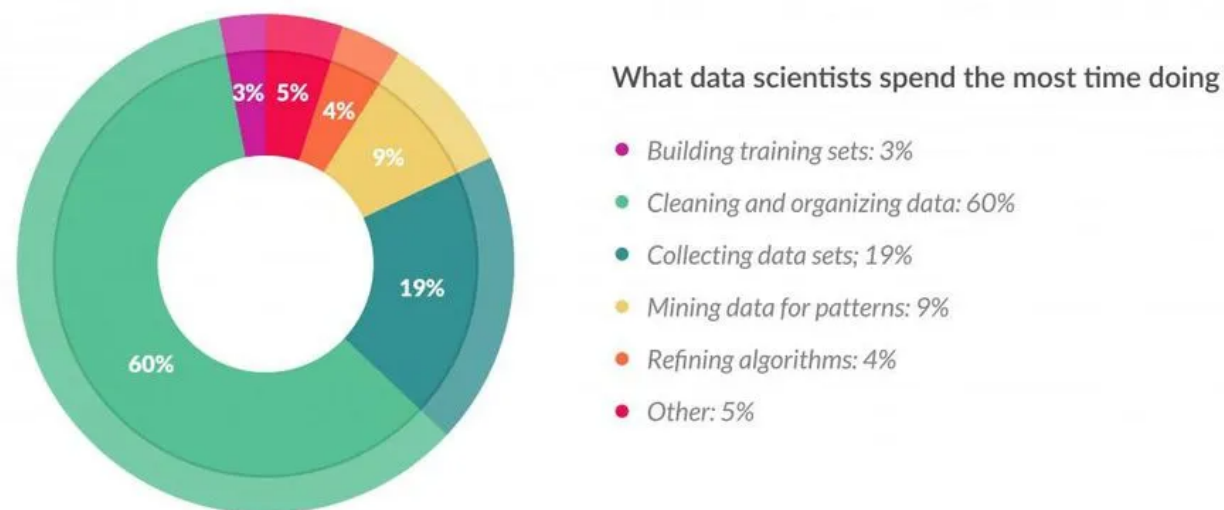
input data comprise features, which are usually in the form of structured columns. Algorithms require features with some specific characteristic to work properly. Here, the need for **feature engineering** arises. I think feature engineering efforts mainly have two goals:

- Preparing the proper input dataset, compatible with the machine learning algorithm requirements.
- Improving the performance of machine learning models.

*The features you use influence more than everything else the result. No algorithm alone, to my knowledge, can supplement the information gain given by correct feature engineering.*

— Luca Massaron

According to a survey in Forbes, data scientists spend **80%** of their time on **data preparation**:



Source: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>

This metric is very impressive to show the importance of feature engineering in data science. Thus, I decided to write this article, which summarizes the main techniques of feature engineering with their short descriptions. I also added some basic python scripts for every technique. You need to import **Pandas** and **Numpy** library to run them.

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

Some techniques above might work better with some algorithms or datasets, while some of them might be beneficial in all cases. This article does not aim to go so much deep in this aspect. Tough, it is possible to write an article for every method above, I tried to keep the explanations brief and informative. I think the best way to achieve expertise in feature engineering is practicing different techniques on various datasets and observing their effect on model performances.

### **List of Techniques**

- 1.Imputation
- 2.Handling Outliers
- 3.Binning
- 4.Log Transform
- 5.One-Hot Encoding