**3 Fundamental Processes in Feature Engineering**

Presenting data patterns to models the right way



Photo by [Acton Crawford](https://unsplash.com/@acton_crawford?utm_source=medium&utm_medium=referral" \t "_blank) on [Unsplash](https://unsplash.com/?utm_source=medium&utm_medium=referral" \t "_blank)

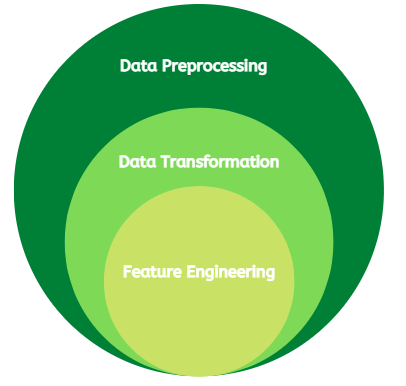
**Introduction**

This post explains the three crucial processes in feature engineering (FE) that you need to know to properly present data patterns to machine learning (ML) models.

Feature engineering is the process of modifying existing features to enhance the ability of a model to learn from the data.

FE offers tangible improvement in model accuracy without significantly increasing computational time and cost.

FE is a subset of data transformation, a critical element of data preprocessing.



The layers of data preprocessing (image by author)

In a recent article, I discussed data transformation in detail. However, FE stands out as a subfield. The link to the article is displayed below:

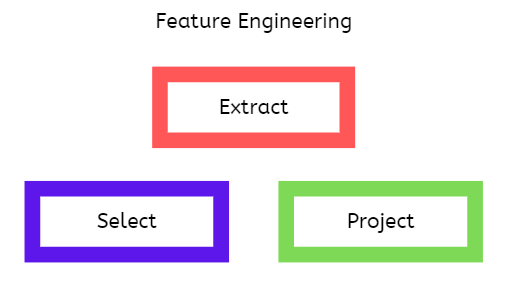
**[Three Critical Elements of Data Preprocessing — Part 3](https://towardsdatascience.com/three-critical-elements-of-data-preprocessing-part-3-6a7da681ae16" \t "_blank)**

[The backbone of modeling in data science.](https://towardsdatascience.com/three-critical-elements-of-data-preprocessing-part-3-6a7da681ae16" \t "_blank)

[towardsdatascience.com](https://towardsdatascience.com/three-critical-elements-of-data-preprocessing-part-3-6a7da681ae16" \t "_blank)

The main objective of this article is to discuss the components of data transformation that are FE. This provides a better understanding of the data preprocessing step in the data science project life cycle.

**3 Fundamental Processes in Feature Engineering**



Fundamental processes of feature engineering (image by author)

Now, let’s delve into the main processes in FE:

1. **Feature Extraction**

This is the process of generating new features from existing ones. It is highly domain-specific and relies heavily on your knowledge of the subject area. The main idea is to create new features that enable an ML model to learn better from the data. For example, when predicting the power output from a wind turbine, creating a windspeed magnitude feature from the raw X and Y direction windspeeds gives better model accuracy.

In addition, many ML models work with numerical data, data consisting of integers or decimals. Hence, we need to encode the raw categorical data, data consisting of strings to make them usable to the models. For example, a status variable can have “ON” and “OFF” categories.

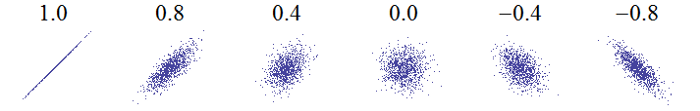
There are more advanced methods of feature extraction based on feature learning techniques. This approach is more data-driven, generic, and scalable. Some examples include [autoencoders](https://machinelearningmastery.com/autoencoder-for-classification/" \t "_blank) and [clustering](https://towardsdatascience.com/how-to-create-new-features-using-clustering-4ae772387290" \t "_blank).

2. **Feature Selection**

This is the process of choosing the most relevant features for the training process. Feature selection methods fall under three main categories namely wrapper, filter, and embedded methods. An in-depth discussion on feature selection can be found [here](https://neptune.ai/blog/feature-selection-methods" \t "_blank).

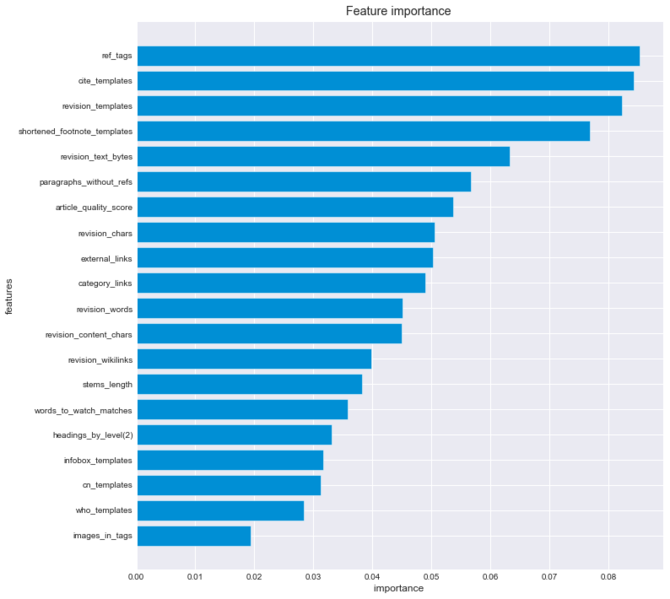
Some measures of a feature’s relevance include:

*Correlation analysis:*The correlation coefficient measures the relationship between two variables and takes a value between -1 and +1. A positive correlation means both variables move in the same direction (that is, as one increases, the other increases, and vice versa). In addition, the larger the magnitude of the coefficient the stronger the correlation between the variables. In feature selection, features having a higher correlation with the target variable are chosen because they have a higher predictive power.



Sample correlation plot. [Original photo](https://commons.wikimedia.org/wiki/File:Correlation_examples.png) cropped by [w:User:Imagecreator](https://commons.wikimedia.org/wiki/File:Correlation_examples_1.png), [CC0](https://creativecommons.org/publicdomain/zero/1.0/deed.en), via Wikimedia

*Feature importance:*Some tree methods such as random forests and gradient-boosting algorithms provide feature importance scores that show the effect of each feature on the target prediction. These scores may be used the choose the most relevant features. More details can be found [here](https://towardsdatascience.com/understanding-feature-importance-and-how-to-implement-it-in-python-ff0287b20285" \t "_blank).



Sample feature importance ranking. Photo by [0xkaywong](https://commons.wikimedia.org/wiki/File:Wikireliability_original-research_feature-importance.png), [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0) via Wikimedia

*Mutual information:*This measures the reduction in the uncertainty of one variable based on the knowledge of another variable. A reduction in uncertainty results from having more information about the variable. Features with high mutual information scores are considered more relevant and are chosen for ML modeling. More details can be found [here](https://guhanesvar.medium.com/feature-selection-based-on-mutual-information-gain-for-classification-and-regression-d0f86ea5262a).

3. **Feature Projection**

This is the process of mapping high-dimensional data to a lower-dimensional space. It typically involves reducing the number of features fed to an ML algorithm. This is beneficial for many reasons. One is to reduce the complexity of the resulting model hence, reducing the chance of overfitting. Another is to reduce computational time and effort while not significantly affecting the model's accuracy.

There are two main classes of feature projection techniques:

*Linear projection:*These methods employ a linear combination of the features and do not capture interactions between two or more features. Some examples include linear discriminant analysis (LDA) and principal component analysis (PCA). More details can be found [here](https://bitsandbrains.io/2018/09/25/linear-dimensionality-reduction.html" \t "_blank).

*Non-linear projection:* These methods are more complex and described by non-linear equations. Some examples include kernel principal component analysis (KPCA) and principal curves. More details can be found [here](https://golden.com/wiki/Nonlinear_dimensionality_reduction_(NDR_or_NLDR)-6834P" \t "_blank).

**Conclusions**

In this article, we covered the three fundamental processes in feature engineering, a subfield of data transformation. The processes are feature extraction, selection, and projection. Examples of different methods employed in these processes were provided including some resource links.