Feature Engineering

ASHRAE - Great Energy Predictor III

Date: **09-11-2019**

Venue: **Internshala, Gurugram**
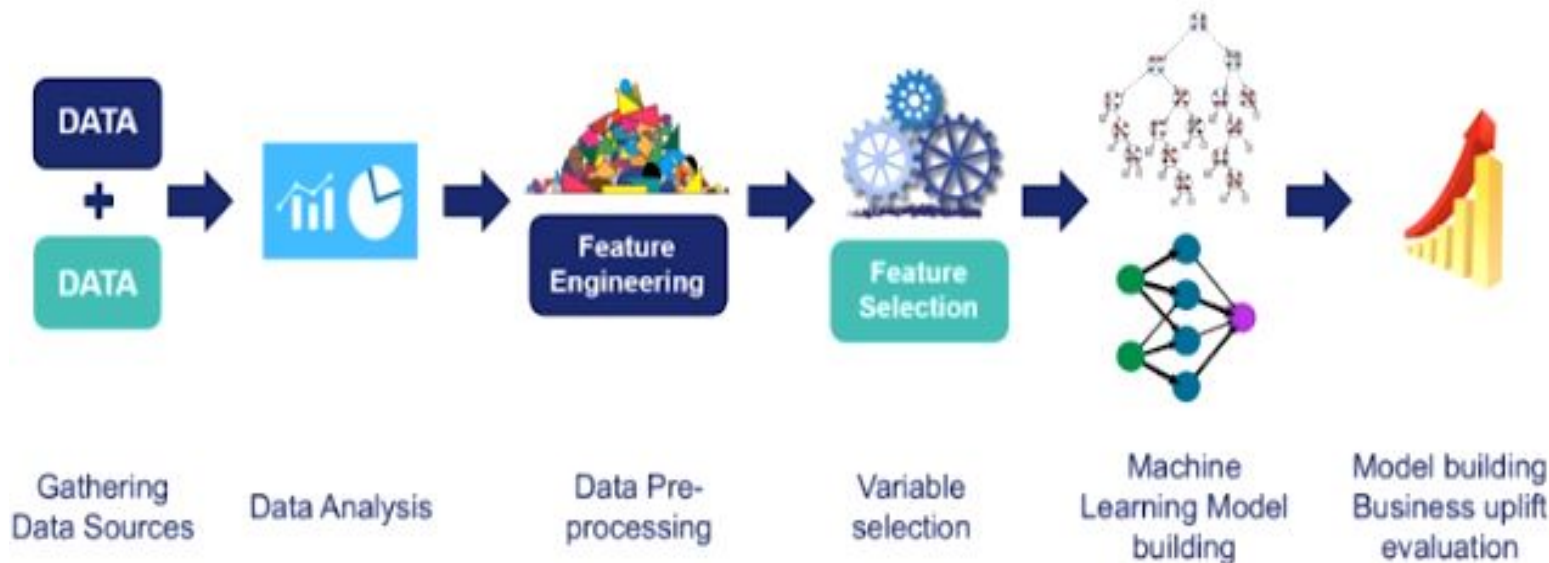
"More data beats clever algorithms but better data beats more data".

-*Peter Norvig*

A look at the Data Science Pipeline....

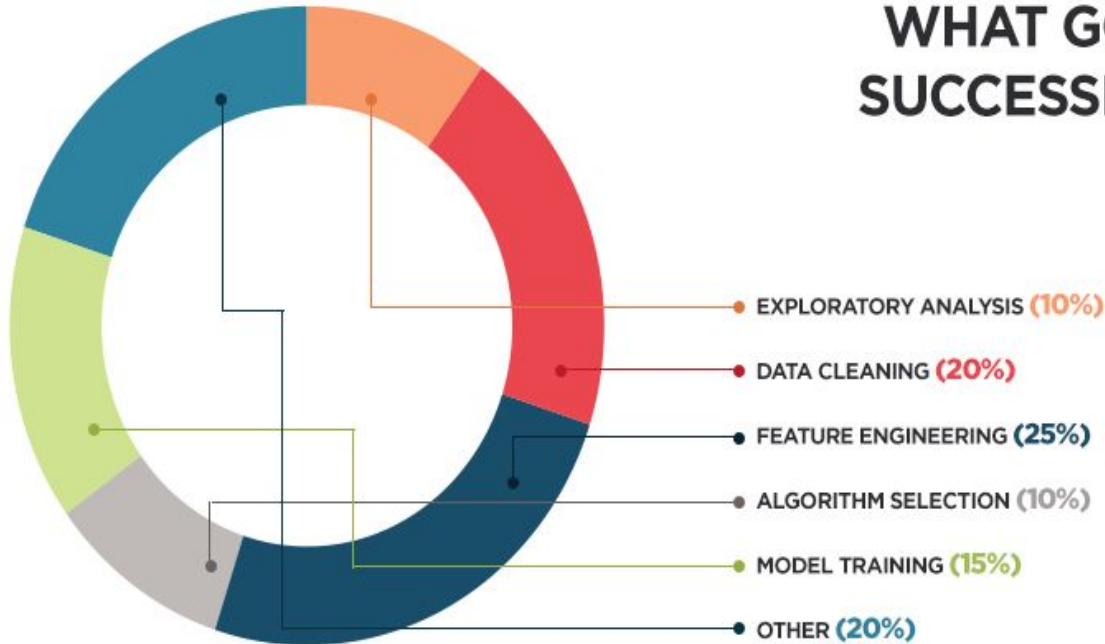| Gathering Data Sources | Data Analysis | Data Pre-processing | Variable selection | Machine Learning Model building | Model building Business uplift evaluation |

Feature Engineering is domain specific

- Feature Engineering is the process of finding the optimal set of features that should be given as input to the machine learning model.

- Feature Engineering is *domain specific* thus requires intuition, smart decisions, creativity and lots of trials and errors.

- *Takes up the most of the time* spent on a machine learning problem.

WHAT GOES INTO A SUCCESSFUL MODEL

EXPLORATORY ANALYSIS (10%)

DATA CLEANING (20%)

FEATURE ENGINEERING (25%)

ALGORITHM SELECTION (10%)

MODEL TRAINING (15%)

OTHER (20%)

DATA SCIENCE PRIMER
by EliteDataScience

kaggle DAYS
MEETUP

INTERNSHALA kaggle LOGICAI Devfolio

Important Techniques of Feature Engineering

Some of the fundamental techniques used for feature engineering in machine learning are listed below:

- *Feature Representation*
- *Feature Transformation*
- *Feature Extraction*

Feature Representation

- Feature representation is the way of representing a same feature in a different way so as to leverage it's predicting power.

- Feature representation generally includes :
  - *Categorical Encoding*
  - *Date-Time Engineering*
  - *Feature Mapping*

- It is an impactful way of getting your data in an ideal format for your machine learning algorithm.

Feature Transformation

- Feature transformation is the process of transforming the existing features into an optimal format.

- Feature Transformation generally includes:

    - *Variable Transformation*

    - *Discretisation*

    - *Feature Scaling*

- Feature transformation comes handy when you want to change the distribution of a feature in order to interact better with other features.

Feature Extraction

- Feature Extraction is the process of creating new features from the existing data by using various operations.

- Feature Extraction generally includes:
  - *Indicator variables*
  - *Interaction variables*

- Feature Extraction explicitly requires some domain knowledge to extract meaningful relationships between the predictors and the target variable.

Automated Feature Engineering

- **Featuretools** is an open source python package used to create features automatically from multiple tables of structured and related data.
- Automated feature engineering mostly focuses on finding relationships between variables and possible transformations that could be applied to the predictors.

Featuretools

Feature Engineering Best Practices

- Read **articles, blogs and research papers** for gathering some inside information about the problem.

- Bring in some **external data** if available, to use for the problem in hand which will significantly affect your model performance.

- Interact with the **domain experts** or colleagues working in the same field to validate the relationships you used for creating features for your model.

Feature Engineering Resources

- **Github links**
  - Awesome-feature-engineering

  - Feature-engineering-and-feature-selection

  - Featuretools - Automated Feature Engineering

- **Articles**

  - Discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it

  - Feature-engineering-best-practices

Thank you Everyone
&
Happy Kaggling!

Ishaan Jain