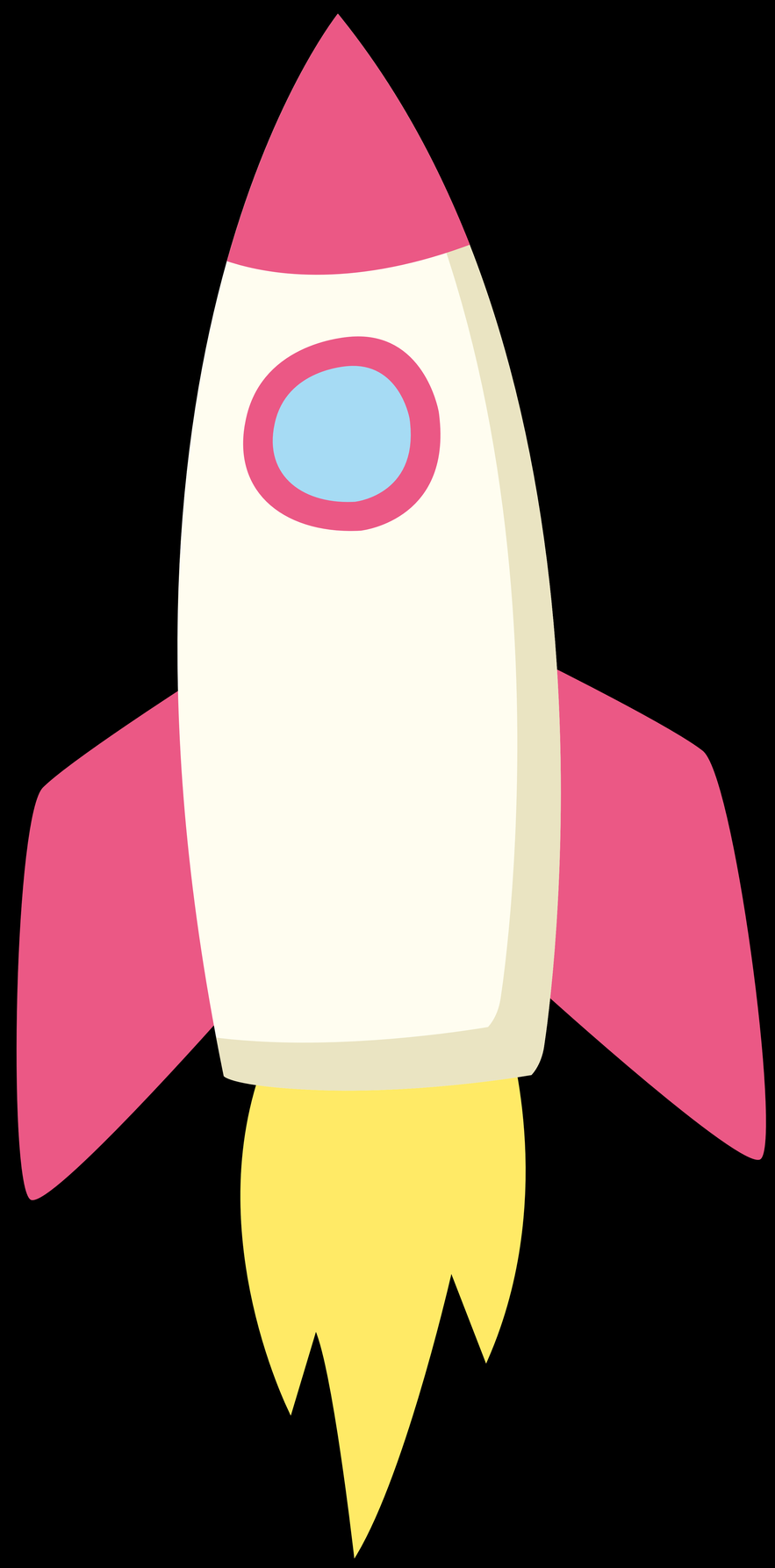
GUIDE TO BUILD END TO END MACHINE LEARNING PROJECT



Data cleaning and formatting

In data cleaning, inaccurate records are identified and removed from a dataset, table or database. The process involves identifying unfinished, unreliable, inaccurate, or irrelevant data and then restoring, remodeling, or removing the dirty or crude data.

Exploratory data analysis

Using summary statistics and graphical representations, EDA refers to the process of identifying patterns in data, spotting anomalies, testing hypotheses, and checking assumptions.

The purpose of this approach is to summarize, visualize, and become intimately familiar with the important characteristics of a data set.

Feature engineering and selection

The well-known concept of "grabage in - garbage out" applies 100% to any task in machine learning.

Feature extraction and feature engineering: Transformation of raw data into features suitable or modeling

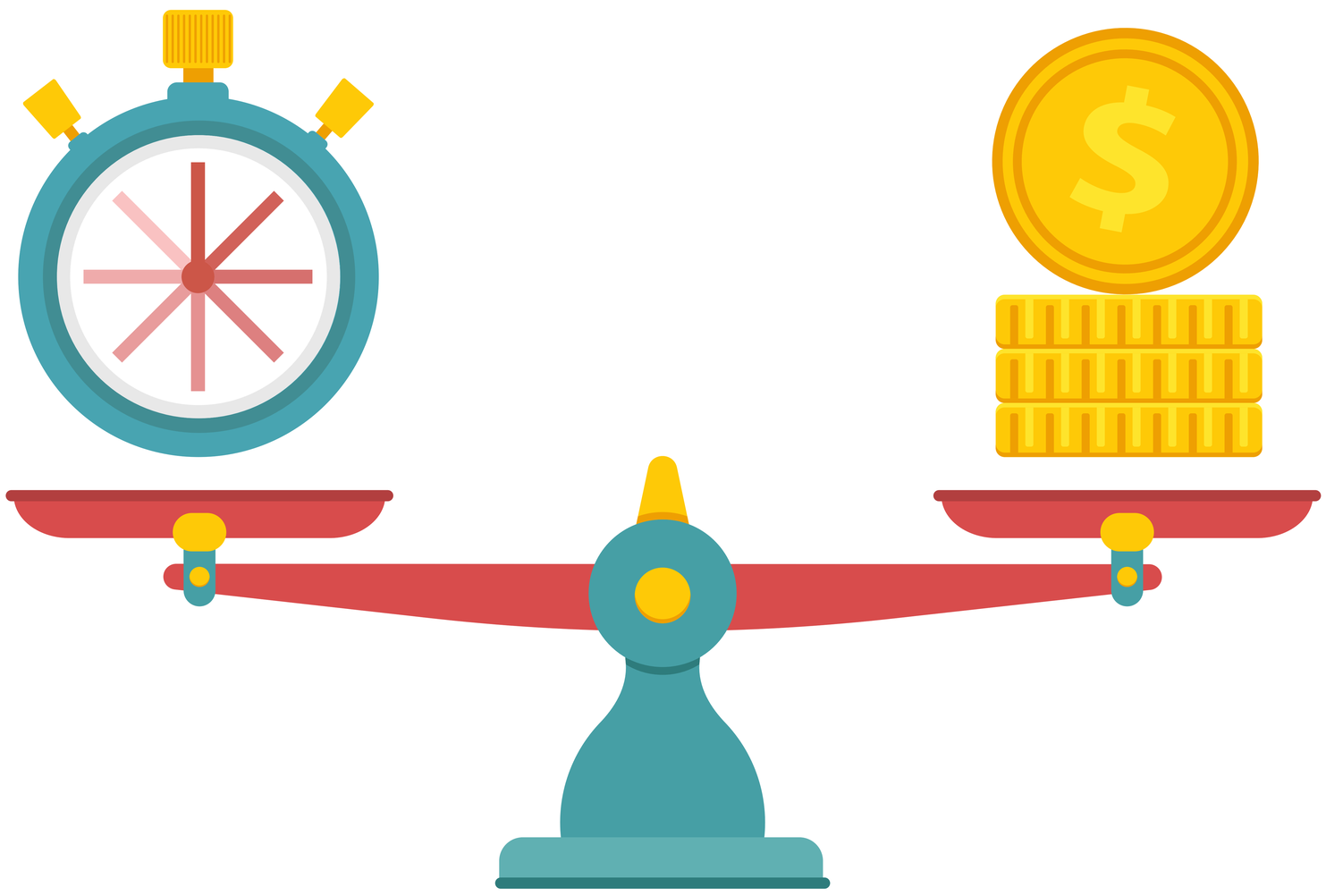
Feature Transformation: Transformation of data to improve the accuracy of algorithm

Feature selection: Removing unnecessary features.

Compare multiple Algorithms

Multiple machine learning algorithms should be compared consistently.

ML projects often result in multiple good models to choose from. There will be differences in the performance characteristics of each model.



Hyperparameter tuning

Hyperparameters cannot be directly trained from data as they are not model parameters.

A hyperparameter optimization or tuning problem in machine learning involves selecting a set of optimal hyperparameters for a learning algorithm.

The hyperparameter values are used to control the learning process.

Evaluate the models

The evaluation of models is an integral part of the model development process. Our data is used to find the best models that represent our data and how well the chosen model will performll perform in the future.

When evaluating model performance with the data used for training, overoptimistic and overfit models are easily generated.

Holdout and cross-validation are two methods for evaluating models in data science

Deploy the model

It is fundamentally important to be able to interpret complex models in machine learning. Interpretability helps debug the model by analyzing what it really considers important.

You deploy your model so that you can make predictions from a machine learning model available to others, such as users, management, or other systems. Model deployment is closely related to ML systems architecture.

Documentation

- Reproducibility
- Ensure Successful project completion
- Documentation helps ensure consent and expectations. it helps to tell the narrative for decisions made, and how you or the client responded to different situations.
- In the same manner, it is important to record information that can help support the proper treatment plan and the reasoning for such services.

TODAY'S QUIZ

In the following type of feature selection method, we start with an empty feature set

A forward feature selection

B backword feature selection

C Both A and B

D None of the above

Comment your answer and why? 👇👇

Like

Save