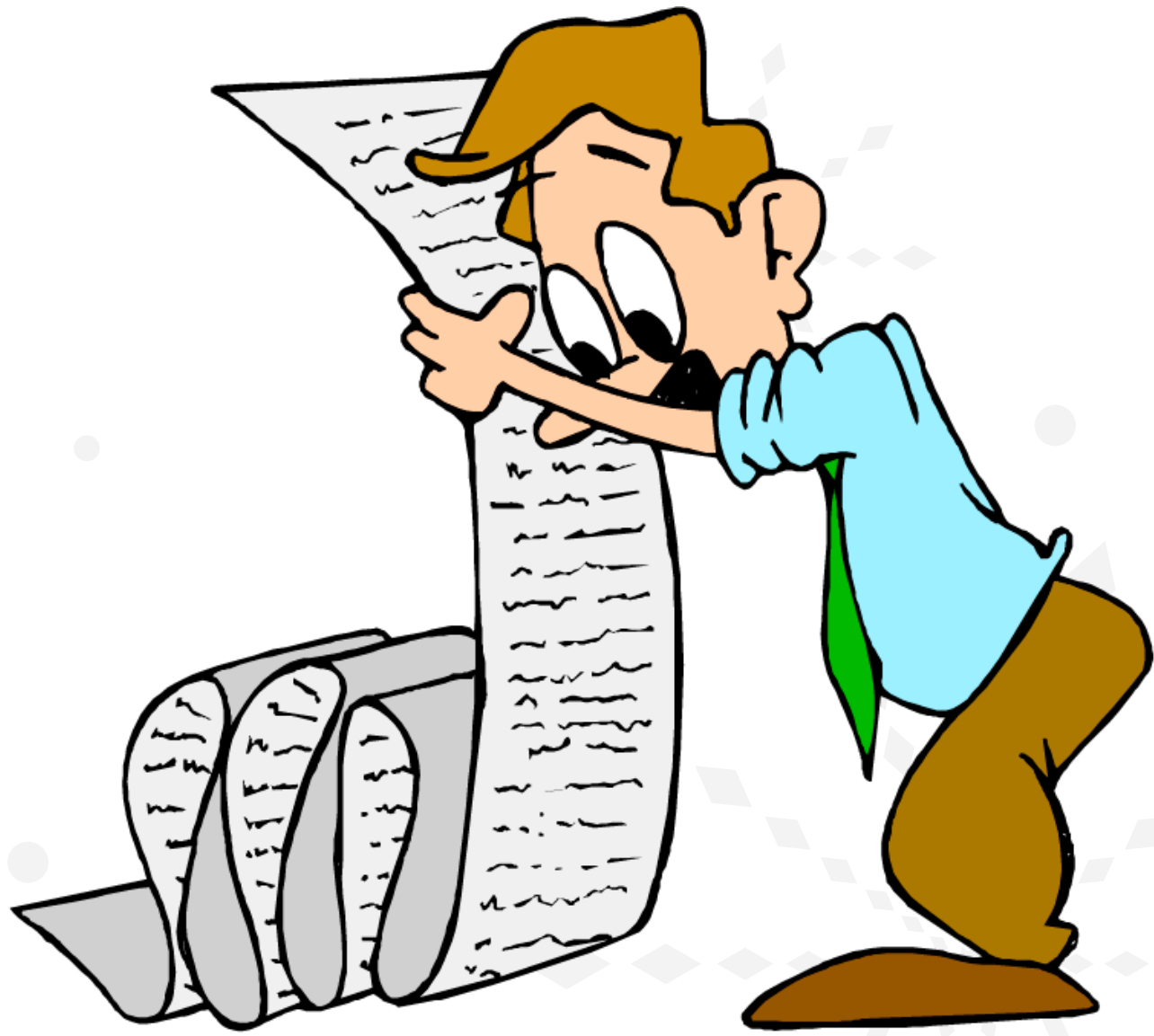


DIFFERENT THINGS TO DO IN

**DATA**

**PRE-PROCESSING**

**STEP**



@learn.machinelearning

# Things to do (Depends on the problem)

- Check for missing values
- Dropping of unwanted columns
- Removing the Noisy Data
  - Binning Method
  - Regression @learn.machinelearning
  - Clustering
  - Outliers
- Check for Inconsistent values
- Check for categorical data
- Check for Duplicate values
- Feature Aggregation
- Feature Encoding (Encoding categorical values or numeric values or text)
- Train / Validation / Test Split

# Things to do (Depends on the problem)

- Data Transformation
  - Normalization
  - Attribute Selection
  - Discretization
  - Concept Hierarchy Generation
- Data Reduction @learn.machinelearning
  - Data Cube Aggregation
  - Attribute Subset Selection
  - Numerosity Reduction
  - Dimensionality Reduction
- Feature engineering
- Skewness/Kurtosis
- Feature correlation
- Imbalance Dataset (Data imputation)
- Feature scaling

# Things to do (Depends on the problem)

- Handling Dates
- Multicollinearity
- Feature selection
- Domain knowledge @learn.machinelearning
- Own ideas in Feature engineering and feature selection