



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Supervised Learning with Python

Dr. Tulasi Prasad Sariki
Associate Professor, SCOPE
VIT, Chennai

www.learnersdesk.weebly.com

<https://github.com/stulasiprasad/Machine-Learning-with-Python>

Outline

- What is Supervised Learning?
- Tools used
- Categorization of Supervised Learning Algorithms
 - Regression
 - Classification
- Implementation steps
 - Regression
 - Classification

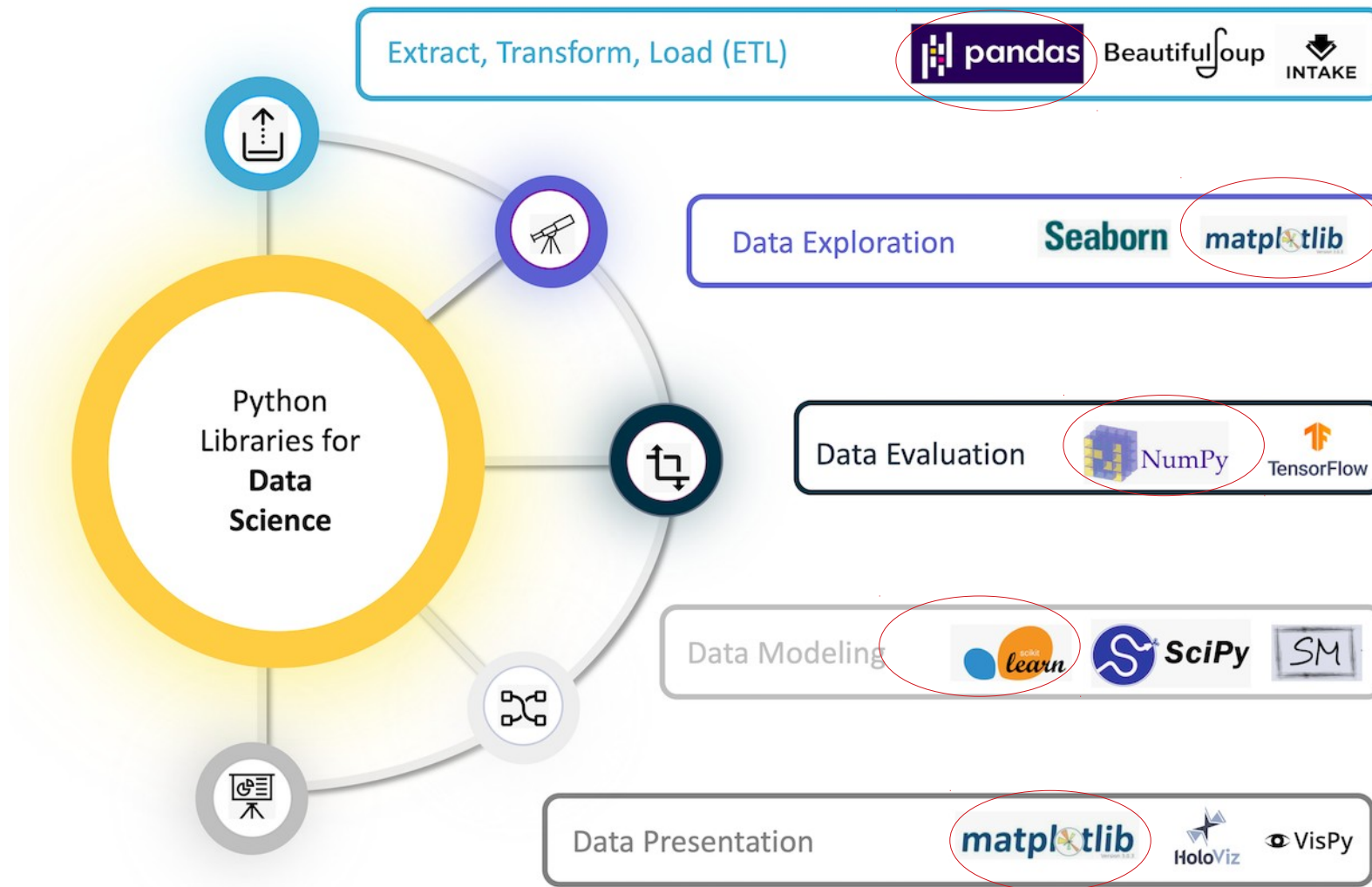
What is Supervised Learning?

- *Supervised learning* is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.
- In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal).
- A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

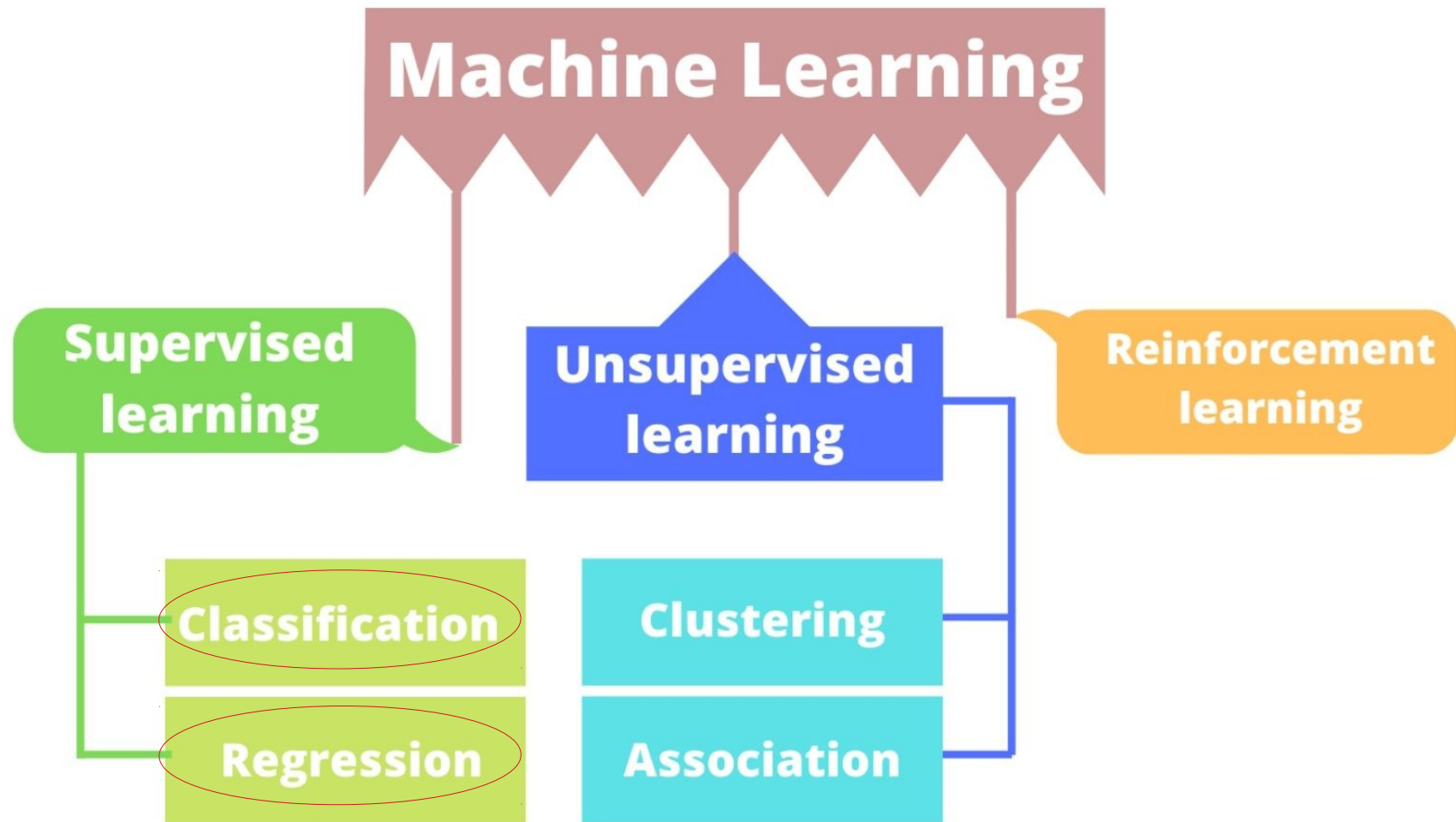
Tools and Libraries used

- We use **Colab** from Google: Colaboratory or Colab , allows us to write and execute Python in your browser, with
 - Zero configuration required
 - Free access to GPUs
 - Easy sharing
- **Python libraries:**
 - numpy
 - pandas
 - matplotlib
 - sklearn

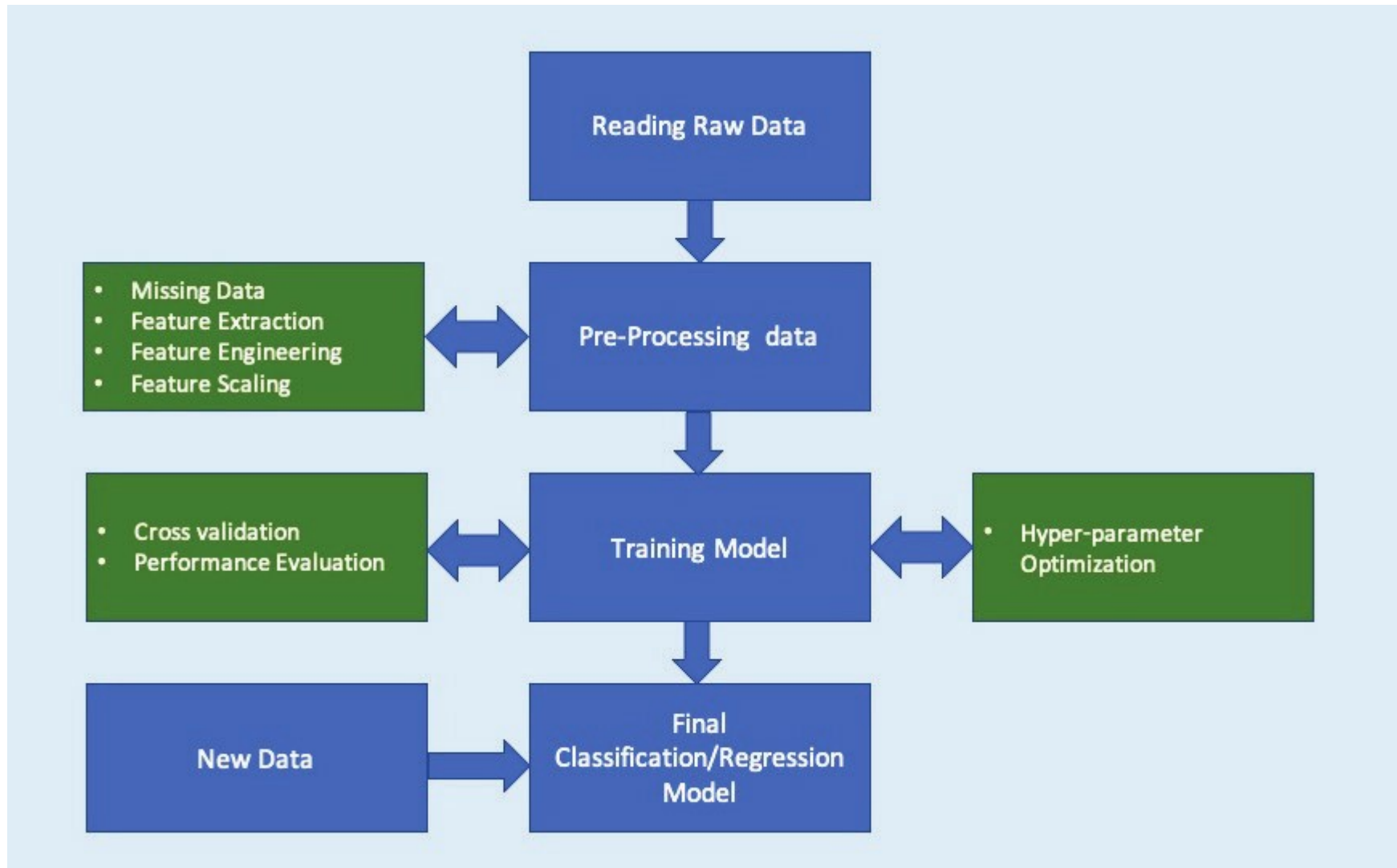
Tools and Libraries used



Categorization of Supervised Learning Algorithms



Implementation steps



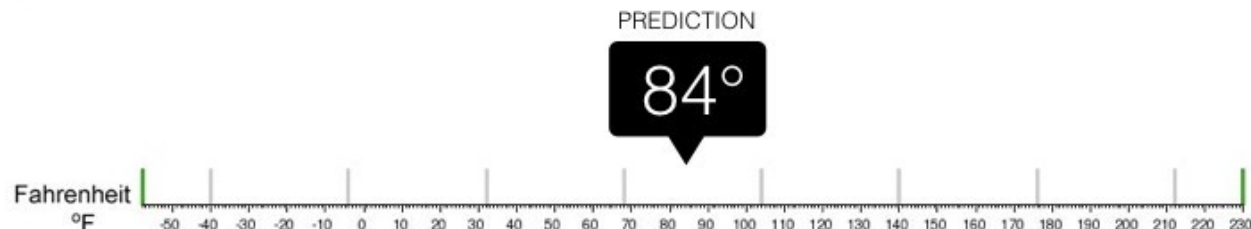
Regression

- A regression problem is when the output variable is a real or continuous value, such as “temperature” .
- Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points.



Regression

What is the temperature going to be tomorrow?



Implementation steps

1. Load and describe the data

2. Exploratory Data Analysis

- Exploratory data analysis - numerical
- Exploratory data analysis - visual
- Analyse the target variable
- Compute the correlation

3. Pre-process the data

- Dealing with missing values
- Treatment of categorical values
- Remove the outliers
- Normalise the data

Implementation steps

4. Split the data

5. Choose a Baseline algorithm

- defining / instantiating the baseline model
- fitting the model we have developed to our training set
- Define the evaluation metric
- predict scores against our test set and assess how good it is

6. Refine our dataset with additional columns

7. Test Alternative Models

8. Choose the best model and optimise its parameters

- Gridsearch

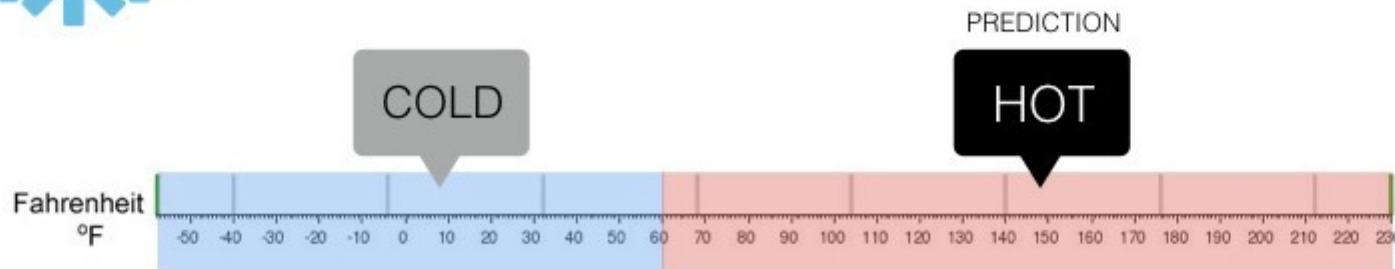
Classification

- A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
- A classification model attempts to draw some conclusion from observed values.



Classification

Will it be Cold or Hot tomorrow?



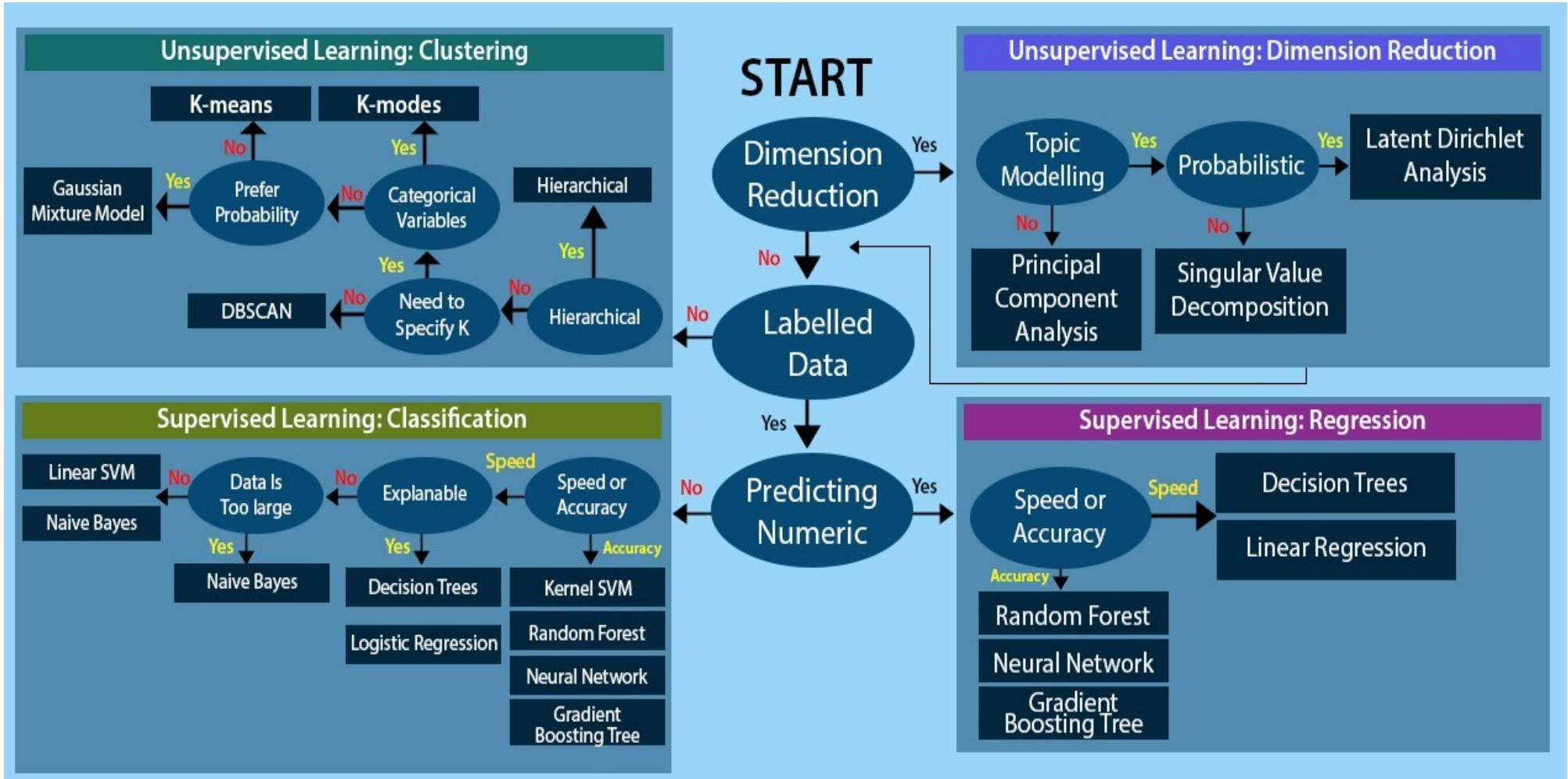
Implementation steps

1. Load the data
2. Exploratory data analysis
 - Analyse the target variable
 - Check if the data is balanced
 - Check the co-relations
3. Split the data
4. Choose a Baseline algorithm
5. Train and Test the Model

Implementation steps

6. Choose an evaluation metric
7. Refine our dataset
8. Feature engineering
9. Test Alternative Models
10. Ensemble models
11. Choose the best model and optimize its parameters

Machine Learning Algorithms Cheat Sheet



References

- https://en.wikipedia.org/wiki/Supervised_learning
- <https://pandas.pydata.org/>
- <https://numpy.org/>
- <https://matplotlib.org/>
- [https://scikit-learn.org/stable/https://scikit-learn.org/stable/](https://scikit-learn.org/stable/)
- <https://www.datasciencecentral.com/group/ai-deep-learning-machine-learning-coding-in-a-week>



Happy Learning!