**Beat The Heat with Machine Learning Cheat Sheet**

**Machine Learning at a Glance!!**

**Machine Learning, not a straight path**

If you are a beginner and just started machine learning or even an intermediate level programmer, you might have been stuck on how do you solve this problem. **Where do you start? and where do you go from here?**

In Machine Learning, there’s no single solution that can fit all and multiple solutions to a problem can exist. With lots of varieties of algorithms, choosing the right algorithm…

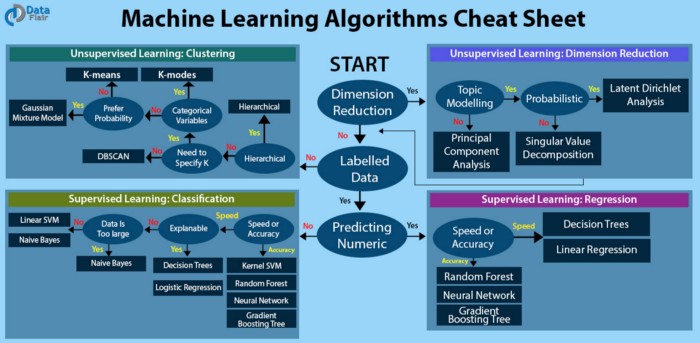
Don’t worry! in this article, we will be simplifying your approach in Machine Learning with a cheat sheet that you can use to select the right algorithm suited for your problem.

**Factors to consider while Choosing algorithm -**

There are several factors that can affect the decision of choosing the right algorithm. Some problems are specific and require unique approaches. For Example, a recommendation system is used to solve a very specific kind of problem. While some type of problems are open and they require trial and error method. Supervised Learning, classification, and regression are open types of problems.

* What you want to do with the data — Do you want to perform classification, regression or clustering?
* **Size:**The size of your data set, whether it is large or small is important in choosing algorithms.
* **Quality:**How much variation is in your data set, is the data set balanced or not.
* **Nature of data:**Do we have labelled data? How is the input and output of the model represented?
* **Time Availability:**How much time you have to build and train the model. Some models can be built quicker but you might have to sacrifice some accuracy.
* **Speed or Accuracy:**   
  For production-ready models, you may need your model to have as much accuracy as it can and sometimes you just only need a fast working model that can compute things faster.

**Guide to Cheat Sheet**



To use the cheat sheet you just have to look at the labels on the chart as decisions and move towards the arrow that answers the question. For example,

* If you want to reduce the number of dimensions and don’t need topic modelling then go with PCA.
* If you want to predict the numeric value of some variable and accuracy is your preference then you should try Random Forrest, neural networks or Gradient boosting tree.
* If you don’t have labelled data and want to perform clustering then you can go with k-means clustering.

**Choosing the right algorithm**

It is worth mentioning that even an experienced data scientist can’t tell which algorithm will perform best without trying different algorithms. This cheat sheet may not be the only solution for your problems and there may be multiple paths for the same task. This cheat sheet is only hoped to provide you with some guidance on what algorithms can be used based on the known factors.

**Types of Machine Learning Algorithms**

[**Make the Next-to-last mistake**](https://data-flair.training/blogs/machine-learning-algorithms/)

**Supervised Learning**

Supervised learning algorithms involves direct supervision of operation. We teach or train the machine using data, which means that the data is labelled with the right answer. We use an algorithm to analyse the training data and learn the function that maps inputs with their outputs. The function can then be used to predict output of unknown inputs by generalising from training data. Supervised learning is basically used for two types of problems.

* **Classification:** In a classification problem, the objective is to find the category of input data. For example, classifying an image as a “dog” or a “cat”.
* **Regression:** In a regression problem, the output is a real value. Where we try to predict the value of a variable based on the input.

**Semi-Supervised Learning**

Supervised learning requires labelled data, which can be challenging to find or generate if someone else didn’t work on a similar project. In a semi-supervised approach, we use some of the labelled data with unlabelled data.

As you can see, the data is not fully labelled that is why this is called semi-supervised learning. The model’s accuracy is improved by using labelled data with unlabelled data.

**Unsupervised Learning**

Unsupervised learning is used for unlabelled data. The machine has to discover patterns, similarities and differences that lies in the data without any supervision. The perform clustering and reducing the number of dimensions.

* **Clustering:** According to some criteria and similarities the data is grouped into one or more clusters. For example, grouping customers with their purchasing behaviour.
* **Dimension reduction:** Some of the features or dimensions of data may be irrelevant that are not needed for model training. With some algorithms, we can reduce the dimensions and irrelevant features. This process is called Dimensionality reduction.

**Reinforcement Learning**

Reinforcement learning optimizes the behaviour of an agent, based on the feedback from the environment. An agent rewards the machine when it makes the right decisions and penalties for its bad decisions. This learning doesn’t require us to collect data and then clean data. It is a self-sustained system that tries to improve itself in the real world. AlphaGO, a computer program based on reinforcement learning beats the best Go player in the world.

**EndNote**

Machine Learning problems can be solved in numerous ways, and you can choose algorithms based on several factors like accuracy, objective, size, and nature of data. You can refer the cheat sheet and get a head start in building the model. Once, you have implemented the solution and got results then you can further explore different algorithms to see what is the best algorithm that is suited for that particular problem.