**What is Machine Learning**

Machine Learning is a method of data analysis that automates analytical model building. It’s part of AI based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention

There’s many different ways you can go about machine learning specifically with certain algorithms and functions within a program

**Types of Machine Learning**

* **Supervised**
  + Uses a known dataset that includes desired inputs and outputs
  + The algorithm must find a method to determine how to arrive at those inputs and outputs
  + There are three types of supervised machine learning:
    - **Classification**: the machine learning algorithm must draw a conclusion from observed values and determine in which category new observations belong.
    - **Regression**: the machine learning program estimates the relationships between variables. Regression analysis focuses on one dependent variable and a series of other changing variables, making it useful for prediction and forecasting.
    - **Forecasting**: making predictions about the future based on the past and present data, commonly used to analyze trends.
* **Semi-supervised**
  + Uses both labeled (information that has meaningful tags so that the algorithm can understand the data) and unlabeled data
  + Semi-supervised machine learning algorithms can learn to label unlabelled data.
* **Unsupervised**
  + Studies data to identify patterns
  + There is no answer key or human operator to provide instruction
  + The machine determines the correlations and relationships by analyzing available data
  + In an unsupervised learning process, the machine learning algorithm is left to interpret large data sets and address that data accordingly
  + The algorithm tries to organize that data in some way to describe its structure
  + There are two types on unsupervised machine learning:
    - **Clustering**: grouping sets of similar data (based on defined criteria). It’s useful for segmenting data into several groups and performing analysis on each data set to find patterns.
    - **Dimension reduction**: Reduces the number of variables being considered to find the exact information required.
* **Reinforcement**
  + Focuses on regimented learning processes, where a machine learning algorithm is provided with a set of actions, parameters and end values.
  + By defining the rules, the machine learning algorithm then tries to explore different options and possibilities, monitoring and evaluating each result to determine which one is optimal.
  + Teaches the machine trial and error
  + Learns from past experiences and begins to adapt its approach in response to the situation to achieve the best possible result.

Given that we know our inputs (team data) and desired output (probability of each team winning a game), we can narrow our focus on supervised machine learning algorithms. With that in mind, we should look at different supervised machine learning algorithms:

* **Naïve Bayes Classifier Algorithm (Supervised Learning - Classification)**
  + Based on Bayes’ theorem and classifies every value as independent of any other value. It allows us to predict a class/category, based on a given set of features, using probability.
* **Support Vector Machine Algorithm (Supervised Learning - Classification)**
  + Analyze data used for classification and regression analysis.
  + Filters data into categories, which is achieved by providing a set of training examples, each set marked as belonging to one or the other of the two categories. The algorithm then works to build a model that assigns new values to one category or the other.
* **Linear Regression (Supervised Learning/Regression)**
  + The most basic type of regression, allows us to understand the relationships between two continuous variables.
* **Logistic Regression (Supervised learning – Classification)**
  + Focuses on estimating the probability of an event occurring based on the previous data provided.
  + Used to cover a binary dependent variable, that is where only two values, 0 and 1, represent outcomes.
* **Decision Trees (Supervised Learning – Classification/Regression)**
  + A flow-chart-like tree structure that uses a branching method to illustrate every possible outcome of a decision.
  + Each node within the tree represents a test on a specific variable – and each branch is the outcome of that test.
* **Random Forests (Supervised Learning – Classification/Regression)**
  + Combines multiple algorithms to generate better results for classification, regression and other tasks.
  + The algorithm starts with a ‘decision tree’ (a tree-like graph or model of decisions) and an input is entered at the top. The input travels down the tree, with data being segmented into smaller and smaller sets, based on specific variables.
* **K-Nearest Neighbors (Supervised Learning)**
  + Estimates how likely a data point is to be a member of one group or another.
  + It essentially looks at the data points around a single data point to determine what group it is actually in.

**Resources**

**Different types of machine learning and their applications:**

* <https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html#:~:text=There%20are%20four%20types%20of,%2Dsupervised%2C%20unsupervised%20and%20reinforcement.>

**Videos for beginners:**

* Machine Learning with Scikit-learn - Data Analysis with Python and Pandas p.6
  + <https://www.youtube.com/watch?v=7eh4d6sabA0>
* Python Machine Learning Tutorial (Data Science)
  + <https://www.youtube.com/watch?v=BpPJxtOk8uw>