

## Getting Started

**Machine Learning** is a program that tries to find patterns in data and then uses those patterns for either predicting or recognizing something similar. Those predictions are not always accurate; however, the more time and data you provide to a machine learning algorithm, the more accurate it becomes.

**Data** is crucial for any machine learning algorithm. Without data, the algorithm cannot learn anything. The quality of data can drastically alter the results, thus it's important to always make sure that the data that is being used is trustworthy. A machine learning algorithm is able to use that data to find **patterns** that could be used for identifying if other data fall under those patterns as well as make any predictions based on those patterns. **Accuracy** is a way to verify if the program is making actual predictions and not some random guesses. The higher the accuracy, the better performance of the model is.

Machine learning is a subset of **Artificial Intelligence**. Machine learning algorithms specifically learn from data rather than some other AI algorithms that learn from computation.

### ML Application #1

One of the applications, where machine learning was used is handwriting recognition. It's incredibly difficult to teach a computer if not impossible to recognize human writing without any machine learning algorithm. With the machine learning algorithm we don't need to explicitly teach a computer how to recognize human writing. We just need to provide data so that the machine learning algorithm could train a model that could recognize human writing based on the patterns that it observed through training.

### ML Application #2

Another example of an application where machine learning is used is predicting whether an email is spam or not. With this example, again it's difficult to teach a computer if a program is a spam or not as not all emails are the same. However, if you provide the program enough data, it will be able to pick up patterns that could be used for predicting whether an email is spam or not.

Every row in a data set is called an "**observation**". However, every column in a data set is called a "**feature**". Features are important as they allow us to group data based on a certain characteristic. For example, an observation could be a list with the following contents: 3.2, 15, 1450, "Junior". Features help us organize this data by setting a category for each item. For example, 3.2 in this case is the student's GPA, 15 is number of credit hours they are taking, and 1450 is their SAT score, and etc. We could use several features to predict a certain feature of a data if the algorithm was able to find patterns among those features. Features could be considered as **quantitative data** or **qualitative data**. Quantitative data is data that is numeric. For example, GPA is quantitative data. Qualitative data is data that is conceptual. For example, student's year in college is considered conceptual data as there's 4 categories: Freshman, Sophomore, Junior, and Senior.

**My interest in Machine Learning**

I am interested in learning more about Machine Learning as I believe that's where the future is headed, and I would love to be a part of that. It always fascinated me how computers could learn things such as reading human handwriting or tell if the person has an arrhythmia or not. Seeing such great impact on the society, made my want to explore the field of machine learning more deeply.