



# A SIMPLE GUIDE TO ASSIST YOU IN SELECTING THE BEST MACHINE LEARNING ALGORITHM FOR BUSINESS STRATEGY

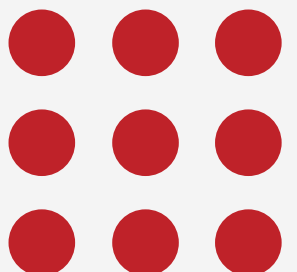
An Academic presentation by  
Dr. Nancy Agnes, Head, Technical Operations, Phdassistance  
Group [www.phdassistance.com](http://www.phdassistance.com)  
Email: [info@phdassistance.com](mailto:info@phdassistance.com)





# Today's Outline

- Introduction
- The volume of training data
- Accuracy and/or Understandability of the Results
- Training or Speed time



# Introduction

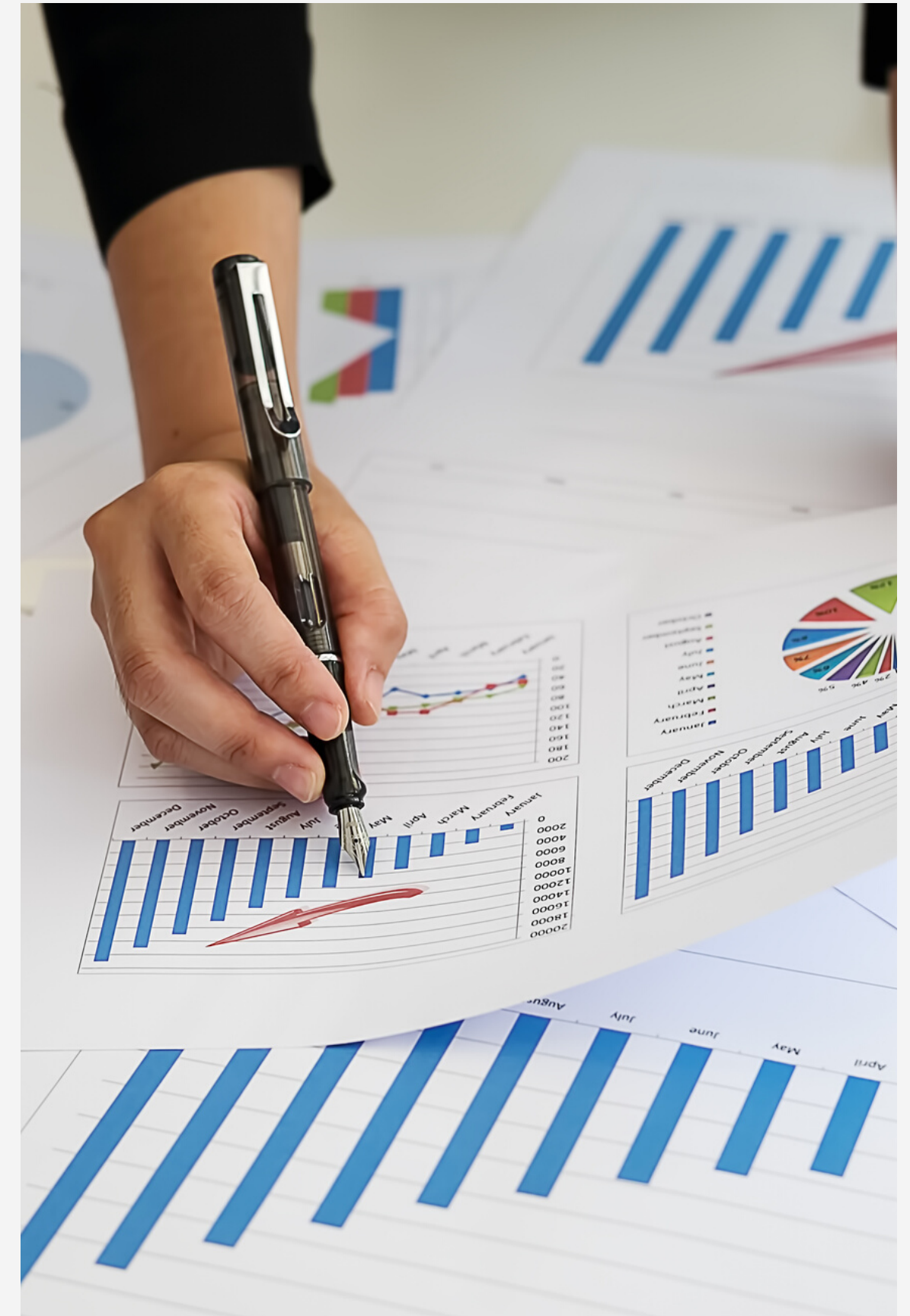
- The scientific discipline of [machine learning](#) enables computers to learn without explicit programming.
- One of the most intriguing technologies that has ever been developed is machine learning.
- A machine-learning algorithm is a computer program that has a specific way of changing its own settings based on feedback from previous predictions made using the data set.





# The volume of training data

- To generate accurate predictions, it is typically advised to acquire a sizable amount of data. Data accessibility, however, is frequently a limitation.
- Select algorithms with high bias/low variance, such as Linear regression, Naive Bayes, and Linear SVM, if the training data is less or if the dataset, such as genetics or text data, has fewer observations but more features.
- KNN, decision trees, and kernel SVM are examples of low bias/high variance algorithms that can be used if the training data is sufficiently large, and the number of observations is more than the number of features.





# Accuracy and/or Understandability of the Results

- An accurate model is one that predicts response values for given observations that are reasonably close to the actual response values for those observations.
- Flexible models provide increased accuracy at the expense of low interpretability, whereas restrictive models (like linear regression) have highly interpretable algorithms that make it simple to comprehend how each individual predictor is related to the result (Lee & Shin, 2020) .
- Some algorithms are said to as restrictive because they only yield a limited number of mapping function shapes.

**contd..**



- For instance, because it can only produce linear functions, like the lines, linear regression is a limited strategy.
- Some algorithms are referred to as flexible because they may provide a broader variety of mapping function forms. For instance, KNN with  $k=1$  is particularly versatile since it will consider every input data point to produce the output function for the mapping.
- The goal of the business problem will determine which algorithm to apply next. Restrictive models are preferable if inference is the desired outcome since they are much easier to understand.
- If greater accuracy is the goal, flexible models perform better. The interpretability of a method generally declines as its flexibility does (Zou, 2020) .



# Training or Speed time

- Typically, longer training times correspond to increased accuracy. Algorithms also take longer to train on massive training data. These two elements dominate in real-world applications when [choosing an algorithm](#).
- Simple to use and quick to run are algorithms like Naive Bayes, Linear Regression, and Logistic Regression.
- It takes a long time to train the data for algorithms like SVM, which requires parameter adjustment, neural networks with a fast convergence rate, and random forests.



**contd..**

# Linearity

- Many algorithms are based on the idea that classes can be divided along a straight line (or its higher-dimensional analog). Support vector machines and logistic regression are two examples. The underlying premise of linear regression algorithms is that data trends are linear.
- These algorithms work well when the data is linear. However, since not all data is linear, we also need alternative algorithms that can deal with large-scale, intricate data structures. Kernel SVM, random forest, and neural nets are examples.
- The best technique to determine linearity is to fit a straight line, conduct a logistic regression, or use a support vector machine (SVM) and look for residual errors. Higher error indicates nonlinear data that would require sophisticated algorithms to fit.

**contd..**





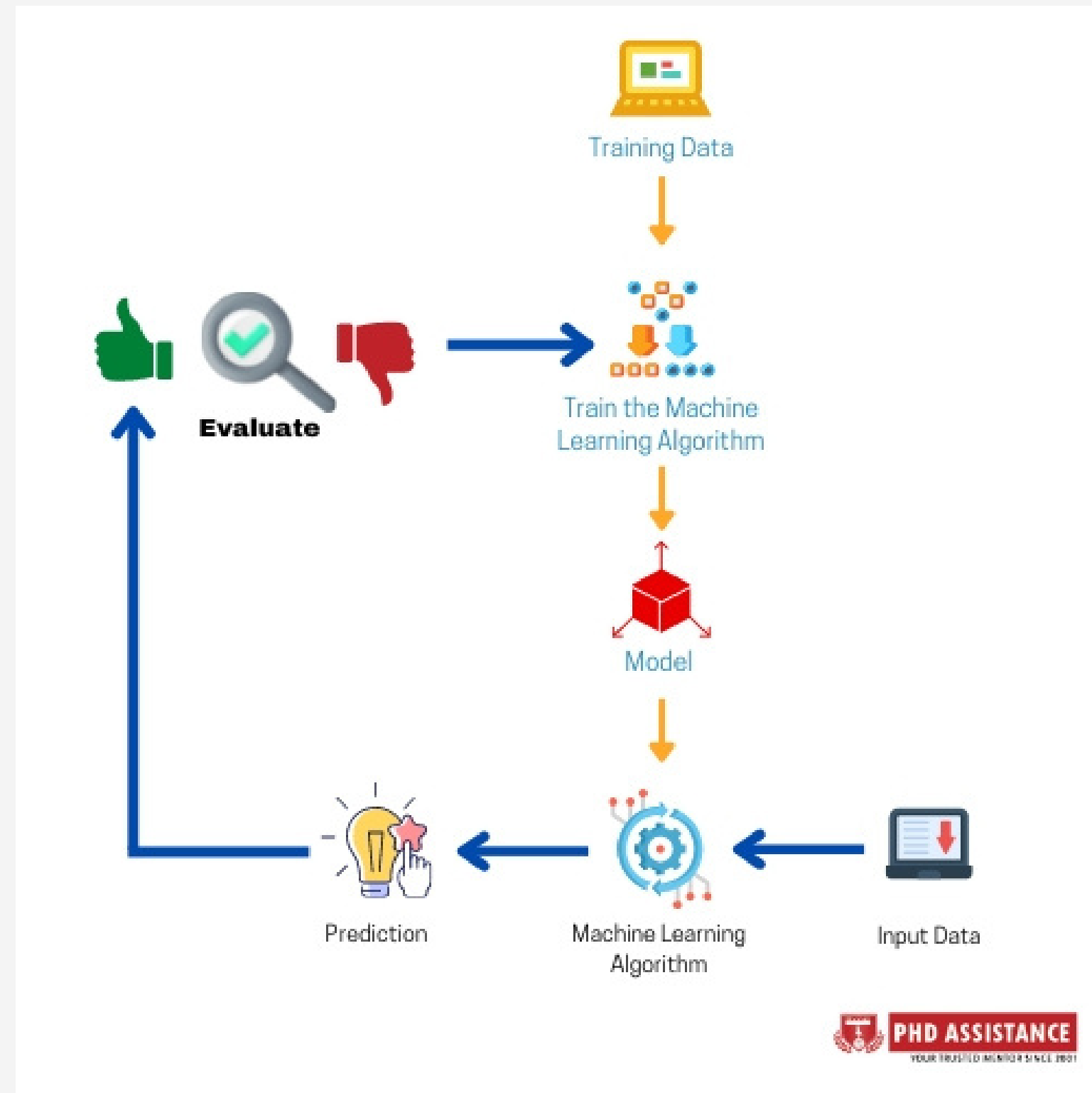
# Number of features

- The dataset could contain numerous features, not all of which are necessarily important and relevant. The number of features for some types of data, such as genomics or [textual data](#), can be relatively high in comparison to the number of data points.
- Some learning algorithms can become sluggish due to a large number of features, making training time unworkable. When there are few observations and a lot of features in the data, SVM works better.
- To minimize dimensionality and choose key features, feature selection methods like PCA should be utilized (Baig et al., 2020) .
- When the training data's output variables match the input variables, supervised learning algorithms are used. In order to map the link between the input and output variables, the algorithm evaluates the input data and learns a function.

**contd..**

- Additional categories for supervised learning include [regression](#) , classification, forecasting, and anomaly detection. When the training data lacks a response variable, unsupervised learning algorithms are employed.
- These algorithms look for inherent patterns and hidden structures in the data. Unsupervised learning algorithms include clustering and dimension reduction techniques. Regression, classification, anomaly detection, and clustering are briefly explained in the infographic below, along with examples of possible applications for each.
- The guidelines will be of great assistance in shortlisting a few algorithms, but it can be challenging to determine which algorithm would be the most effective for your situation. It is advised to work in iterations as a result. Test the input data with each of the shortlisted options to see which is best, and then assess the algorithm's performance (Sarker, 2021) .

**contd..**



**Figure 1:** Steps involved in evaluating the accuracy using ML algorithms



- Therefore, to create a flawless solution to a real-world problem, one must be well-versed in applied mathematics and be cognizant of business requirements, stakeholder concerns, and legal and regulatory requirements.
- For more blogs in selecting the right type of algorithm for various applications can contact PhD assistance anytime for the best service.
- At PhD assistance, we offer the highest excellence [thesis writing assistance](#) in the most ethical manner. The finest PhD thesis writing services are offered by our professional experts.

# Contact Us

---



**UNITED KINGDOM**  
**+44 7537144372**



**INDIA**  
**+91-9176966446**



**EMAIL**  
**[info@phdassistance.com](mailto:info@phdassistance.com)**



**PHD ASSISTANCE**

**YOUR TRUSTED MENTOR SINCE 2001**