* **Why selected the specified algorithm?**

We have selected logistic regression algorithm. Whenever we need to predict whether a student graduates or drops out of a degree, then the target variable comprises of two values i.e pass and fail. Therefore we can solve classification of problem statements which is a supervised machine learning technique using logistic regression.

* **About algorithms in model builing:-**

Finding a good algorithm is one of the most important steps in machine learning. There are various existing models and algorithms are there to use. Our job is to find an appropriate algorithm from the variety of options over there. There are three types of Machine Learning Models – Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

T[ypes of Machine Learning](https://www.analyticsvidhya.com/machine-learning/?utm_source=blog&utm_medium=commonly-used-machine-learning-algorithms) Algorithms:-

1. Supervised Learning

How it works: This algorithm consists of a target/outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using this set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, [Decision Tree](https://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/), [Random Forest](https://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/), KNN, Logistic Regression etc.

2. Unsupervised Learning

How it works**:**In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering populations in different groups, which is widely used for segmenting customers into different groups for specific interventions. Examples of Unsupervised Learning: Apriori algorithm, K-means.

3. Reinforcement Learning:

How it works**:** Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

List of Common Machine Learning Algorithms

Here is the list of commonly used machine learning algorithms. These algorithms can be applied to almost any data problem:

1. Linear Regression
2. Logistic Regression
3. Decision Tree
4. SVM
5. Naive Bayes
6. kNN
7. K-Means
8. Random Forest
9. Dimensionality Reduction Algorithms
10. Gradient Boosting algorithms
    1. GBM
    2. XGBoost
    3. LightGBM
    4. CatBoost

There are various algorithms available for various purposes of the model. Some algorithms are suited for dealing with text, some for images, and much more. Choose an appropriate algorithm to build your model after analyzing the dataset and aim to build the model. We can implement our model to distinguish fruits with Linear Regression since we are dealing with 3 independent variables to predict the outcome.

* **Process followed for model building:-**

Following is the stepwise procedure followed for model building

1. Understanding the problem
2. Data Collection and Data Preprocessing
3. Exploratory Data Analysis
4. Feature Scaling
5. Cross-Validation
6. Choose appropriate model
7. Train the model
8. Overfitting and Underfitting
9. Evaluate the model
10. Hyperparameter tuning
11. Prediction

* **Technologies used for model building :-**

**Common Tools for the Model Building Phase :**

**R and PL/R :**  
They were described earlier in the model planning phase, and PL/R is procedural language for PostgreSQL with R. Using this approach means that R commands can be executed in the database.

[Octave](https://www.geeksforgeeks.org/basic-operations-in-octave/)**:**  
It is free software programming language for computational modeling, has some of functionality of Matlab. Because it is freely available, Octave is used in major universities when teaching machine learning.

**WEKA :**  
It is free data mining software package with an analytic workbench. The functions created in WAKA can be executed within the java code.

[Python](https://www.geeksforgeeks.org/python-programming-language/)**:**  
It is programming language that provides toolkits for machine learning and analysis, such as scikit-learn, [NumPy](https://www.geeksforgeeks.org/python-numpy/), [scipy](https://www.geeksforgeeks.org/data-analysis-with-scipy/" \t "_blank), [Pandas](https://www.geeksforgeeks.org/pandas-tutorial/), and related data visualization using matplotlib.

[SQL](https://www.geeksforgeeks.org/sql-tutorial/)**:**  
SQL in database implementations, such as MADlib, provides an alternative to memory desktop analytical tools.

**MADlib :**  
It provides an open-source machine learning library of algorithms that can be executed in the database, for [PostgreSQL](https://www.geeksforgeeks.org/what-is-postgresql-introduction/)or Greenplum.

* **Accuracy of machine learning algorithms:-**

Machine learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

* **Hyperparameters used in models:-**

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), a hyperparameter is a [parameter](https://en.wikipedia.org/wiki/Parameter) whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are derived via training. Hyperparameters can be classified as model hyperparameters, that cannot be inferred while [fitting the machine to the training set](https://en.wikipedia.org/wiki/Model_fitting) because they refer to the [model selection](https://en.wikipedia.org/wiki/Model_selection) task, or algorithm hyperparameters, that in principle have no influence on the performance of the model but affect the speed and quality of the learning process. An example of a model hyperparameter is the topology and size of a neural network. Examples of algorithm hyperparameters are [learning rate](https://en.wikipedia.org/wiki/Learning_rate) and batch size as well as mini-batch size. Batch size can refer to the full data sample where mini-batch size would be a smaller sample set.

**Github link:-**

https://github.com/SakshiManjrekar/1\_BinaryBrains\_1

**Google colab link:-**

https://colab.research.google.com/drive/1Y1NMeB0uoYkY62XJ3OpW-GMZaLC7fBd2?usp=sharing