I. THEORETICAL FEASIBILITY OF LOGISTIC REGRESSION

A. *Machine learning:*

Machine learning is the idea that enables computer programs or systems to learn and improve automatically without human involvement. There is no need for explicit programming [5]. Machine learning algorithms use historical data to learn and improve and predict outcomes more accurately Machine learning is mainly of four types:

Supervised Machine learning: In this machine learning algorithm, only labeled data used for training and predicting the accurate outcome [6].

Unsupervised machine learning: This algorithm involves training unlabeled data. It analyzes the whole dataset to draw a meaningful connection.

Semi-Supervised machine learning: Both labeled and unlabeled used for training, usually a small amount of labeled data and large amounts of unlabeled data [7].

Reinforcement Learning: This algorithm is about taking appropriate action to maximize reward in a specific situation. It is used to find the best possible behavior or approach for a particular situation through different software and machines

Logistic Regression:

Logistic regression is a very well-known machine learning algorithm borrowed from the statistic. It is a supervised machine learning technique used to analyze a dataset with one or more independent variables that predict the outcome of the categorical dependent variables. It converts outcomes using the logistic or sigmoid function to return a probability value and this value can be mapped into two or more discreet groups. The logistic regression can be classified into three types Binary logistic regression, Multinomial logistics regression, Ordinal Logistic regression [8].

How logistic regression works:

In general, logistic regression is a supervised classification algorithm. For a classification problem, the target variable (or output) is y, it can only use discrete values for a certain set of functions, X. The logistic regression model uses a more complex cost function which is also known as the sigmoid function or logistic function. The sigmoid function or logistic function is a function that is identical to an "S"-shaped curve when graphically drawn. It takes values from 0 to 1 and "cuts" them to the top and bottom edges, and sticks them to 0 or 1. Logistic regression restricts the cost function between 0 and 1. Like Linear regression, it uses an equation as the representation.

$$y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$$

The function above contains variables b0 and b1. These variables are called weights or coefficient values. b0 describes the bias or intercept and b1 represents the coefficient. These weights are learned and trained from the collected dataset. The outcome of the formula will generate a percentage or probability and it will be mapped through discrete classes. The specified separation of these two groups is known as the boundary of decisions [9]. You can see figure 11.

Advantages of Logistic Regression:

Logistic Regression is a very simple executable machine learning algorithm. It provides excellent training efficiency. The training model using logistic regression doesn't require high computational power**.** Unlike decision trees and support vector machines, the logistic regression enables models to be updated easily to represent new data [10]. The predicted parameters (trained weights) show the significance of each characteristic. There is also a positive or negative direction of the association. Hence, we can use logistic regression to figure out the relationship within the features. It produces well-calibrated predictions and outcomes and this is an advantage over other models that only produce final classification as outcomes. Moreover, Logistic regression is less likely to overfitting into a low dimensional dataset with adequate numbers of training instances. It is very quick and easy to execute thus it is considered as a benchmark model to measure performance. If the dataset has features that can be linearly separable then logistic regression provides excellent accuracy. It has a close association with neural networks. Due to a simple probabilistic explanation,

It takes less time for training. It is less complex than any other complex algorithms such as the artificial neural network. It is also known as multinomial logistic regression [10].

Limitations of Logistic Regression:

Logistic regression should not be used when the number of observations is lesser than the number of features. Therefore, it can lead to overfitting. It creates linear boundaries therefore Solving Non-linear cannot be possible. It is hard to found data in the real-world that linearly separable. Moreover, Logistic Regression cannot capture complex relationships. Hence powerful and complex algorithms like Neural networks can easily outperform logistic regression. It needs average or no multicollinearity within independent variables. Furthermore, Logistic regression requires the log odds to be connected linearly with the independent variables (log(p/(1-p)) [11].

B. Framework

Factors	Why it matters
Ease of Use	It ensures efficiency when performing the task. As it makes sure simple syntax that easy to understand and handle the error.
Available Library	By using different libraries different job or task can be accomplished
Visualization capability	Visualization capability enables information and data into visual context including the map, graph. It makes it easy to understand and analyze the data.
Computational speed	Good computational speed lets you accomplish your task within a short time. It makes it easy to work on a big dataset.
Community support	Community support is an advantage, it helps to clear knowledge and learn fundamentals and you can get a solution to any problem regarding your project
Cost	Less cost provides flexibility to the developer
Statistical Correctness	It helps to produce accurate statistics during data analysis.
System Integration	It brings together all the subsystem for accurate data analysis
Consistency	It enables developers to do more assumption and predictions