II Theoretical Feasibility of Logistic Regression

A. Machine Learning

Machine learning is an artificial intelligence also can be called AI technology that allow programs the facility to learn and develop from practice spontaneously without being programmed directly. Machine learning efforts on the expansion of computer programs that can view, use data and learn on their own. Health diagnosis, image recognition, estimation, grouping, correlation with learning, regression, for example, etc. Smart systems based on machine learning algorithms are able to learn from previous data or historical knowledge. Machine learning will also aid me in doing the project of cryptocurrency. There are three types of machine learning language. They are supervised, unsupervised and reinforcement. The most common model for conducting machine learning tasks is supervised learning. It is commonly used for details where there is an effective mapping of data input-output. In this case, the dataset is named, implying that the algorithm directly defines the characteristics and carries out predictions or classifications accordingly. The extraordinary style written by [1]. In the case of an unsupervised learning algorithm, there is no clear marking of the data in various groups, that is, no labels. By discovering implicit patterns, the model is able to learn from the results. Based on their densities, architectures, comparable segments, and other related characteristics, unsupervised learning algorithms classify the results. This format is written by [2]. In order to achieve goals, reinforcement learning encompasses more fields of Artificial Intelligence that allow machines to communicate with their complex environment. Through this in a given sense, computers and virtual agents are able to determine the optimal behavior. An exceptional format from [3].

B. Logistic Regression

Logistic regression is a statistical model that, while several more complex extensions exist, uses a logistic function to model a binary dependent variable in its basic form. Logistic regression or can be known as logit regression in regression analysis calculates the parameters of a logistic model as a form of binary regression. To analyse the relationship of categorical or continuous independent variables with one dichotomous dependent variable, logistic regression analysis is used. This is in contrast to study of linear regression in which continuous variable is the dependent variable. Logistic regression is mainly used for calculating the binary event for the cryptocurrency prediction. Like to visualise the graph for how much the cryptocurrency have been used and how often it is developing. It predicts the likelihood of yes or no outcomes. This helps to reduce the risk of loss and to enhance to increase the profits [3].

C. How does Logistic Regression Works

As a representation logistic regression uses an equation, very much like linear regression. Input values in x-axis are combined linearly to predict the output value by using weights or coefficient values in y-axis. Data fits into the model of linear regression, which is then acted upon by a logistic function that predicts the dependent goal categorical variable. A threshold can be set to predict which class the data belongs to. The obtained calculated likelihood is divided into groups based on this threshold. A great perception by [4].

D. Advantages of Logistic Regression

Logistic regression is simpler to introduce, analyse and train very effectively. It does not make any claims regarding class distributions in feature space. Multiple classes or multinomial regression and a natural probabilistic view of class predictions can easily be generalized. It provides not only a measure of how suitable a predictor is as in coefficient size, but also its association direction either positive or negative. Classifying unknown records are very easy. Good consistency for several basics sets of data and when the dataset is linearly separable, it performs well. It may view model coefficients as feature significance indicators. Logistic regression is less likely to over-fit, but in high-dimensional datasets it can over-fit. Regularisation approaches like L1 and L2 can be considered to prevent over-fitting in these situations. The predicted factors provide suggestion about the status of each feature like if the result will be positive or negative. Logistic regression is often used as a benchmark model to evaluate output, rather than automatically beginning with a complex model, as it is relatively fast and simple to apply. I have found this in here [5].

E. Limitations of Logistic Regression

Logistic Regression should not be used if the number of observations is smaller than the number of features, otherwise it will lead to over-fitting. It is constructing linear boundaries. The presumption of linearity between the dependent variable and the independent variables is the main drawback of Logistic Regression. Only for predicting discrete functions can it be used. Therefore, logistic regression's dependent variable is bound to the discrete number range. With logistic regression, non-linear issues can't be overcome since it has a linear decision surface. In real-world situations, linearly detachable data is seldom found. Average or no multicollinearity between independent variables is

required by Logistic Regression. Using logistic regression, it is difficult to obtain complex relationships. More efficient and compact algorithms such as Neural Networks will easily outperform this algorithm. Independent and dependent variables are connected linearly in linear regression. But logistic regression demands that the log odds be connected linearly to independent variables. This part is taken from here [5].

F. Framework

Factors	Why It Matters?
Visualisation Capability	It can be easily comprehended to visualise the scrutinize data via the graph.
Libraries	Libraries are a collection of useful functions that remove the need for codes to be written from scratch. In designing computer learning, data science, data visualization, image and object editing software.
Ease of Use	The libraries are already built in the software so used do not need to code instead they can import the libraries and can plot.
Memory Consumption	The quantity of memory that R and Python will use through their execution. As any application depends on the underlying memory to store instances of variables, this definition is self-explanatory.
Computational Speed	This is required because it shows how fats the tool can regain and construe commands.
Community Support	The platform for this tools R and Python has digitally built a large core of users around a body of expertise.
System Integration	The process of bringing the component sub-systems together into one system.