**Logistic Regression:**

Logistic regression calculates the likelihood of an event occurring, such as ‘has voted’ or ‘has not voted’, based on a set of independent variables. As the resulting outcome is a probability, the dependent variable has a range of 0 to 1. A logit transformation is applied to the odds in logistic regression, which is the probability of success divided by the probability of failure. A logit transformation is applied based on the odds, which is the ‘probability of success divided by the probability of failure’. This is also referred to as log odds or the natural logarithm of odds. [1]

Binary logistic regression is a regression model where binary values are the target variables, and they take only two values, 0 or 1. It is the most used regression model in readmission prediction. The response variable is dichotomous in nature, meaning that there are only two possible outcomes. [2]

**Formula:**

* logit(pi) = 1/(1+ exp(-pi))
* ln(pi/(1-pi)) = Beta\_0 + Beta\_1\*X\_1 + … + B\_k\*K\_k

Here, logit(pi) is the response variable and ‘x’ is an independent variable. The coefficient is estimated by Maximum Likelihood Estimation (MLE). This test the different values of the beta value by performing multiple iterations to optimize the perfect value of the log odds. The iterations produce a log likelihood function, and this regression method finds ways to maximize the function for the best parameter. After the coefficients are found, each observation’s conditional probabilities are calculated with log and added together to produce a predicted probability. [1]

In the binary classification, the value 0 is predicted with a probability less than 0.5 and the value 1 is predicted with a probability more than 0. The Hosmer-Lemeshow test is one of the best methods to check how good the model predicts the variables. [1]

**Use of Logistic Regression in Machine Learning:**

This type of regression belongs in the list of ‘supervised machine learning’ models. It also attempts to differentiate between categories, i.e., it is a discriminative model. This algorithm cannot generate any form of information of the class it tries to predict. [1]

In machine learning, the factor of logistic regression maximizing the log likelihood function changes. The negative log likelihood is used as a loss function to find a global maximum with gradient descent. [1]

**Use Cases:**

Since logistic regression is used for problems in classification and predictions. One of the cases is fraud detection. These models can assist teams in identifying data anomalies that predict fraud. Certain characteristics may be more closely associated with fraudulent activities, which can help banks and other financial institutions protect their customers. When conducting data analysis around business performance, SaaS-based companies have begun to adopt these practices to eliminate fake user accounts from their datasets. [1]

**Implementation of Logistic Regression in Packet Header Intrusion Detection:**

To prevent zero-day attacks, intrusion detection systems are to satisfy complex requirements, to be manageable, reliable, and durable. An anomaly-based detection model is developed with a statistical method, combining with a binary logistic regression approach. The model introduced, called the Layer based Anomaly Detection (LbAD) can detect the remote-to-user and user-to-root attacks by observing the degree of field values statistically within the three layers in the OSI model, which are the data link, network, and transport layer. [3]

**References:**

1. <https://www.ibm.com/topics/logistic-regression#:~:text=Logistic%20regression%20estimates%20the%20probability,bounded%20between%200%20and%201>
2. <https://www.sciencedirect.com/topics/computer-science/binary-logistic-regression#:~:text=Binary%20logistic%20regression%20(LR)%20is,or%20not%20readmitted%20(0)>
3. <https://ieeexplore.ieee.org/abstract/document/7491570>