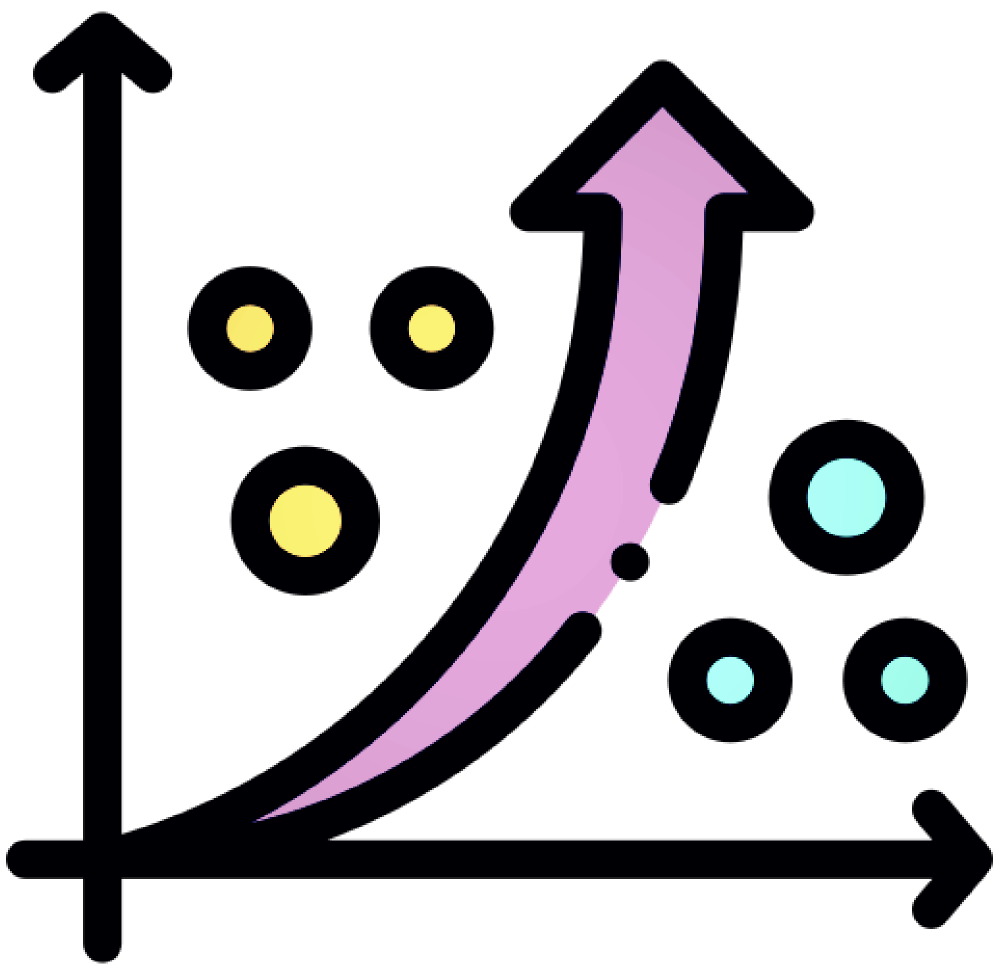
Logistic Regression



DATARANCH.org

VISUALIZE | ANALYZE | CAPITALIZE

What is Logistic Regression?

Logistic regression is a statistical model used to analyze the relationship between a dependent variable (usually binary, meaning it can take one of two values) and one or more independent variables.

It is commonly used for classification problems, such as predicting the probability of an event occurring or not occurring based on a set of independent variables.

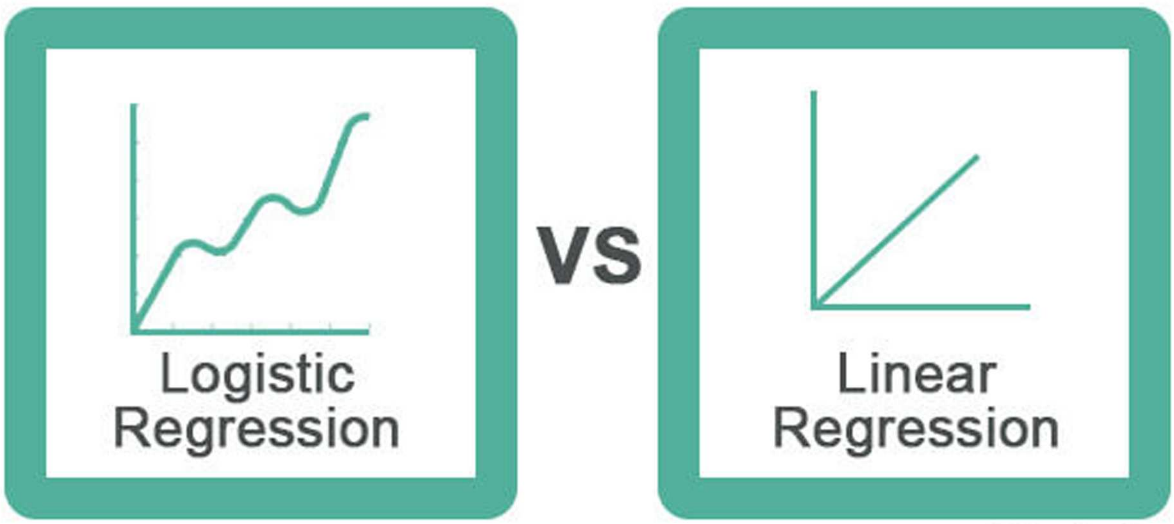
Usage of Logistics Regression

Fraud detection: Logistic regression models can help teams identify data anomalies, which are predictive of fraud.

Disease prediction: In medicine, this analytics approach can be used to predict the likelihood of disease or illness for a given population.

Churn prediction: Specific behaviors may be indicative of churn in different functions of an organization. Logistics regression can be applied to predict churn in areas like sales, HR, etc.

Linear vs Logistic Regression

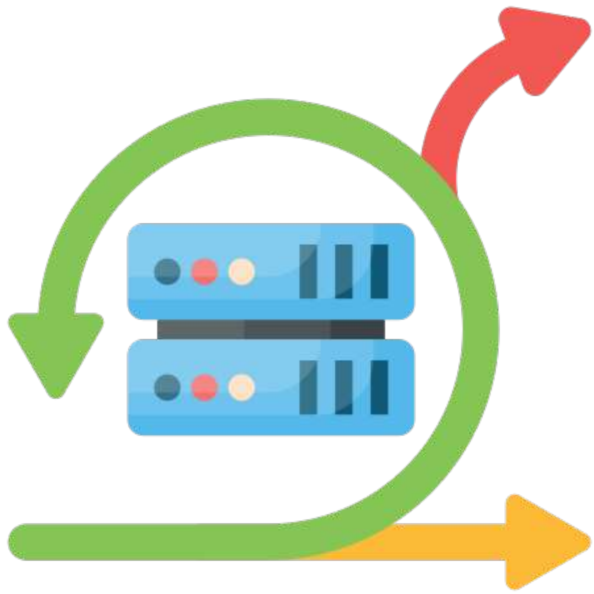


The main difference between linear and logistic regression is the type of response variable.

A linear regression model is used when the response variable takes on a continuous value such as price, height, age and distance. Conversely, a logistic regression model is used when the response variable takes on a categorical value such as yes or no, win or not win.

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

How Logistic Regression Works?



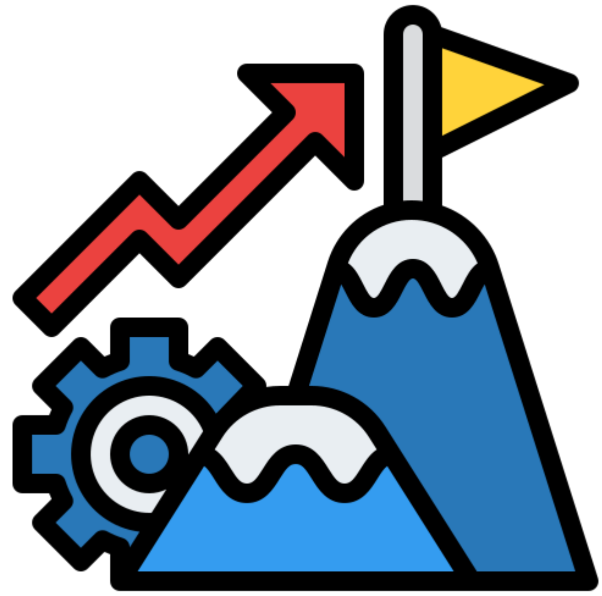
The logistic regression model uses a sigmoid or logistic function to map the input variables to a probability value between 0 and 1. The sigmoid function is given by:

p = 1 / (1 + e^(-z))

where p is the predicted probability of the event occurring, e is the mathematical constant, and z is a linear combination of the input variables and their associated coefficients.

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

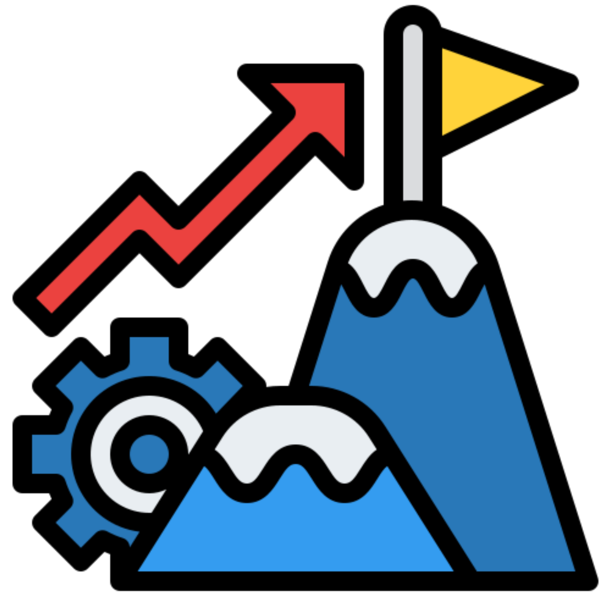
How to Build a Logistics Regression Model



1. Collect and preprocess the data: Collect a dataset that includes both the predictor variables (also known as features) and the dependent variable. Preprocess the data to remove missing values and outliers, normalize the data, and perform other data cleaning steps as necessary.

2. Choose the features: Select the features that are most relevant to the problem being addressed. This can involve using domain knowledge, statistical tests, or feature selection techniques.

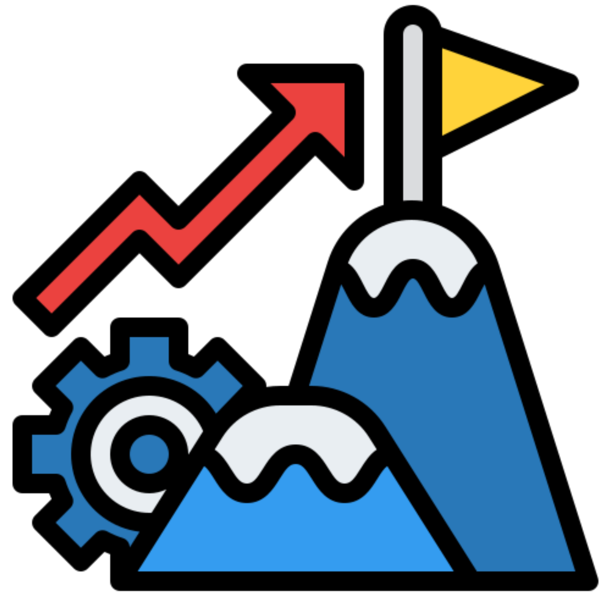
How to Build a Logistics Regression Model



3. Fit the model: Estimate the coefficients of the logistic regression model using maximum likelihood estimation. This involves finding the set of coefficients that maximizes the likelihood of the observed data.

4. Evaluate the model: Evaluate the performance of the model using appropriate metrics such as accuracy, precision, recall, and F1 score. Use techniques such as cross-validation to ensure that the model is not overfitting the data.

How to Build a Logistics Regression Model



5. Use the model to make predictions: Once the model has been trained and evaluated, it can be used to make predictions on new data. To make a prediction, the values of the predictor variables are input into the model, and the sigmoid function is used to estimate the probability of the event occurring.

Drawbacks of Logistics Regression

a. Linear decision boundaries: Logistic regression models can only represent linear decision boundaries between classes. If the classes are not linearly separable, then the model may not perform well.

b. Sensitive to outliers: Logistic regression models are sensitive to outliers, which can have a large influence on the estimated coefficients and predictions.

c. Assumes independence of predictor variables: Logistic regression assumes that the predictor variables are independent of each other. If there are correlations among the predictor variables, then the coefficients may be unstable and the model may not perform well.

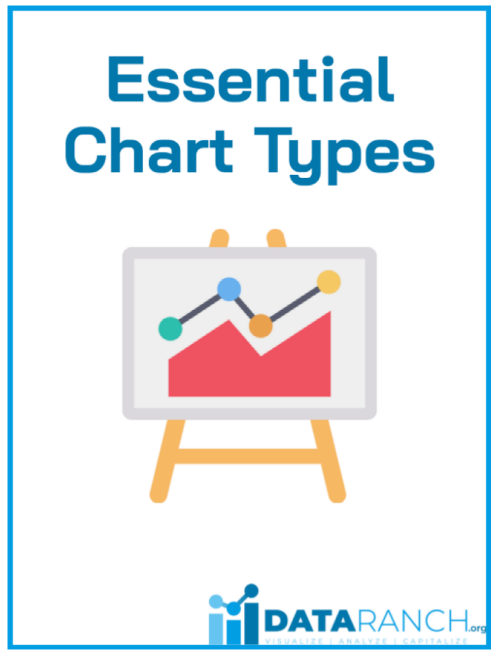
Drawbacks of Logistics Regression

d. May not work well with small sample sizes: Logistic regression requires a relatively large sample size to estimate the coefficients accurately. If the sample size is small, then the model may not perform well.

e. Cannot handle non-linear relationships: Logistic regression assumes a linear relationship between the predictor variables and the logit of the dependent variable. If there are non-linear relationships, then the model may not capture them.

f. May overfit the data: Logistic regression models can overfit the data if too many predictor variables are included or if the model is too complex. This can result in poor performance on new data.

Follow **#DataRanch** on LinkedIn for more...



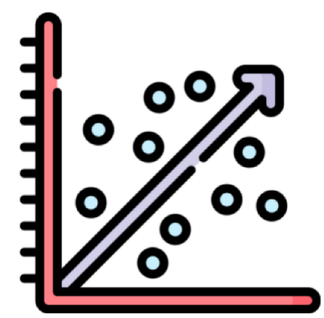
DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

Follow **#DataRanch** on LinkedIn for more...

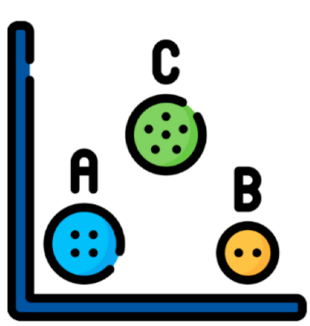
What is Supervised Learning?

What is Unsupervised Learning?

Regression Analysis

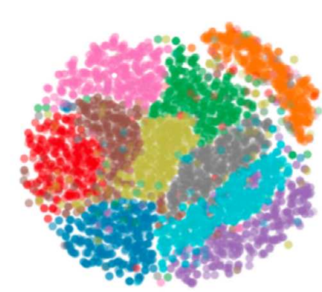


Clustering



Principal Component Analysis

t-Distributed Stochastic Neighbour Embedding (t-SNE)



DATARANCH.org

VISUALIZE | ANALYZE | CAPITALIZE

info@dataranch.org

linkedin.com/company/dataranch