**Machine Learning Algorithms**

# **LR = LogisticRegression()**

Logistic regression is a type of regression analysis used for predicting the probability of an event occurring, especially in binary classification problems where the output can be one of two classes (e.g., spam or not spam, malignant or benign, fake or true).

**Logistic Regression Model:** Logistic regression models the probability that a given instance belongs to a particular category. The logistic function (also called the sigmoid function) is used to map any real-valued number into a value between 0 and 1, which can be interpreted as a probability.

**Usage**

Once you've created an instance of the logistic regression model using LogisticRegression(), you typically use it in the following steps:

* Data Preparation:
  + Load and preprocess your dataset. This may involve tasks such as handling missing values, scaling features, or encoding categorical variables.
* Splitting Data:
  + Split your dataset into training and testing sets. The training set is used to train the model, and the testing set is used to evaluate its performance on unseen data.
* Training:
  + Use the fit method to train the logistic regression model on your training data. This involves adjusting the model's parameters to minimize the difference between the predicted probabilities and the actual class labels.

#Logistic Regression model

from sklearn.linear\_model import LogisticRegression

LR = LogisticRegression()

**LR.fit(xvect\_train, y\_train)** #xvect\_train is the feature matrix of the training set and y\_train is the corresponding vector of labels

pred\_lr = LR.predict(xvect\_test)

#Accuracy score for Logistic Regression model

LR.score(xvect\_test, y\_test)

#Print classification report. The classification\_report function builds a text report showing the main classification metrics

print(classification\_report(y\_test, pred\_lr))

Here, xvect\_train is the feature matrix of the training set, and y\_train is the corresponding vector of labels.

pred\_lr = LR.predict(xvect\_test) make the model predict on testing set.

LR.score(xvect\_test, y\_test) here, we evaluate the performance of the trained model on the testing set to see how well it generalizes to new, unseen data. Here, xvect\_test is the feature matrix of the testing set and y\_train in the label matrix of the testing set.

# **Decision Tree Classifier model**

A Decision Tree Classifier is a type of machine learning algorithm that falls under the category of supervised learning. It is used for both classification and regression tasks. In the context of classification, the algorithm makes decisions based on the input features to assign a data point to one of the predefined classes.

**Decision Tree Structure:**A Decision Tree is a tree-like model where each internal node represents a decision based on the value of a specific feature. The branches leaving each node represent the possible outcomes of the decision, and the leaves represent the predicted class labels.

* Training Process :During the training process, the Decision Tree algorithm recursively splits the data based on feature values to create a tree that can make accurate predictions. The splitting process is guided by criteria such as Gini impurity or information gain, depending on the specific implementation.
* Decision Tree Structure: A Decision Tree is a tree-like model where each internal node represents a decision based on the value of a specific feature. The branches leaving each node represent the possible outcomes of the decision, and the leaves represent the predicted class labels.

In my code, Decision Tree Classifier is trained on a dataset, and its performance is evaluated on a testing set. The classification report includes metrics such as precision, recall, and F1-score for each class. The model is then used to make predictions on a user-input news article in the manual\_testing\_dt function.

# **Gradient Boosting Classifier model**

A Gradient Boosting Classifier is an ensemble machine learning algorithm that combines the predictions of multiple weak learners (typically decision trees) to create a strong predictive model. It belongs to the family of boosting algorithms, which sequentially builds a series of weak models and adjusts them to correct errors made by the previous ones. Gradient Boosting is particularly powerful and widely used for both regression and classification tasks.

GradientBoostingClassifier(random\_state=0). The random\_state parameter is set for reproducibility.

This algorithm builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage n\_classes\_ regression trees are fit on the negative gradient of the loss function, e.g. binary or multiclass log loss. Binary classification is a special case where only a single regression tree is induced.

# **Random Forest classifier model**

Random forest (RF) is an advanced form of decision trees (DT) which is also a supervised learning model. RF consists of large number of decision trees working individually to predict an outcome of a class where the final prediction is based on a class that received majority votes. The error rate is low in random forest as compared to other models, due to low correlation among trees.

# Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. In the context of classification, SVM is particularly powerful for tasks where the goal is to separate data into different classes based on their features. SVM finds the optimal hyperplane that maximally separates the instances of different classes in the feature space.

Key Concepts of SVM:

* Hyperplane:

In a two-dimensional space, a hyperplane is a line that separates two classes. In higher dimensions, it becomes a plane or a hyperplane.

* Support Vectors:

Support vectors are the data points that are closest to the hyperplane and have a significant impact on determining the optimal hyperplane.

* Margin:

The margin is the distance between the hyperplane and the nearest data point from either class. SVM aims to maximize this margin to achieve better generalization.

* Kernel Trick:

SVM can handle non-linear decision boundaries by mapping the input features into a higher-dimensional space using a kernel function.

# How exactly the fake news is detected from real news?

One of the critical advantages of machine learning is its ability to learn from past data and improve its accuracy over time. Training the algorithm with labeled data allows it to recognize patterns in the text and images that indicate false news. The algorithm can also be updated regularly, ensuring it stays up-to-date with the latest types of false news.

Labeling code  
One hot

The front page of the Daily Express is dominated by an image of Aimee Goold, an intensive care nurse from Nottinghamshire.She posted a picture online of herself at the end of her shift, her face a mass of marks, sores and red patches because of the tight-fitting mask she'd been wearing for 13 hours. "Bruised, battered and exhausted," the Express says, "Aimee Goold begged people to 'remember this face' and stay at home."The Daily Mirror leads on complaints by some NHS staff of the lack of personal protective equipment (PPE) - which it illustrates with a picture on its front page of three nurses wearing improvised gowns made out of binbags.Others are resorting to cutting up hospital curtains or using bits of plastic as makeshift masks, the Mirror says, or even reusing old scrubs.The Guardian focuses on the first 10 doctors in the UK named as having died from Covid-19, all of whom were from ethnic minorities.It reports that the British Medical Association has asked the government to "urgently investigate" what it calls an extremely disturbing and worrying trend. The chairman of the BMA, Dr Chaand Nagpaul, tells the paper: "At face value, it seems hard to see how this can be random."The Daily Telegraph says ministers now believe the cost of the crisis here could be the economy "slumping by more than 14% in the next three months."That is sparking concerns - it says - that tens of thousands of Britons may die prematurely from the "worst recession in history", with the lockdown ultimately proving "more lethal to the health of the country" than the coronavirus itself.The Daily Mail reports that officials believe normal life will have stay on hold until a vaccine becomes available, in about 18 months.Ministers want to lift some restrictions within weeks it says - but nonetheless the country will have to get used to a "new normal" until effective medication is available.The Daily Telegraph's take is even more gloomy - "Social distancing 'could be indefinite'" is its headline.It reports people will be asked to consider whether social interactions - such as returning to an office or visiting relations - are necessary, as part of a plan to live with the virus for "months or even years."The Times however, has spoken to the British scientist leading what it calls one of the world's most advanced efforts to find a vaccine - who says it could be ready as soon as September. Sarah Gilbert, professor of vaccinology at Oxford University, tells the paper she is "80% confident" the vaccine being developed by her team will work.Several papers focus on the health of the prime minister. The Sun says Boris Johnson's fiancee Carrie Symonds sent him daily letters and scans of their unborn child while he was in intensive care to "lift his spirits."As Mr Johnson gets better, the Times says, he's needed things to do as he rests. The paper reports his activities have included doing puzzles - like Sudoku - and watching films, such as the Lord of the Rings trilogy and Withnail and I.Of course - because of the lockdown - it's not just the prime minister who suddenly has more time on his hands.The Daily Mail says it's meant huge numbers of people turning to gardening and DIY projects - and massively overloading B&Q's website in the process.At one point yesterday, it reports, would-be purchasers of home improvements were told they would have to wait more than a hour to get online, as more than 300,000 people were simultaneously trying to buy products.The environment is at least benefiting though, the Times reports. It says cities including London, Leeds, Oxford and Glasgow have seen big falls in nitrogen dioxide levels. One expert tells the paper the change is so profound it's the equivalent of everyone switching to electric cars.