CUISINE CLASSIFICATION:

Classification algorithms help us determine the label or class of a data point based on its features. In our case, we will work with a dataset containing information about different cuisines. By analyzing the ingredients, we aim to predict the cuisine of a dish.Classification is a powerful technique in machine learning that enables us to categorize data into distinct classes or categories. By applying classification algorithms and leveraging data-cleaning techniques, we can predict the cuisine of a dish based on its ingredients.In this article, we explored binary and multiclass classification, supervised learning, and the importance of data cleaning and balancing in classification tasks. We also discussed the use of various classification algorithms and evaluation metrics to assess model performance.Before feeding the data into a machine learning model the first step would be to transform the data into a quantitative form that a machine learning model is able to use. The easiest way to do this would be to use one-hot encoding, a technique where a list is converted to a vector of 1s and 0s depending on whether an item is present.Like logistic regression, it can quickly learn a linear separation in feature space for two-class classification tasks, although unlike logistic regression, it learns using the stochastic gradient descent optimization algorithm and does not predict calibrated probabilities.

In this tutorial, you will discover the Perceptron classification machine learning algorithm.

After completing this tutorial, you will know:

- The Perceptron Classifier is a linear algorithm that can be applied to binary classification tasks.
- How to fit, evaluate, and make predictions with the Perceptron model with Scikit-Learn.
- How to tune the hyperparameters of the Perceptron algorithm on a given dataset.

The Accuracy can be increased using more advanced NLP techniques and ML/DL models but however my purpose was to give a basic understanding of doing this project for beginners so that you can start working on a text data and build classification model.

To understand the goals of this analysis, we have cooked up the two high-level questions below:

1. Can machine learning be used to demonstrate learning and successfully classify cuisines based on ingredients?
2. Can machine learning be used to cluster similar cuisines and identify cuisines suitable for fusion recipes?

1) CAN MACHINE LEARNING BE USED TO DEMOSTRATE LEARNING AND SUCCESSFULLY CLASSIFY CUISINES BASED ON INGREDIENTS?

To answer question 1, we can rely on the test accuracy increases that we generated by tuning models and comparing to human performance. Modern machine learning techniques can, in fact, help classify recipes into different cuisine groups. As trained data scientists we feel obligated to mention that accuracy is often an overused metric (because it does not factor in unequal costs of false positives and false negatives). In this case, however, we are indifferent to incorrect classifications of cuisines.

2) CAN MACHINE LEARNING BE USED TO CLUSTER SIMILAR CUISINES AND POTENTIALLY IDENTIFY CUISINES SUITABLE FOR FUSION RECIPES?

To answer question 2, we can rely on two outputs: cluster similarity and mis-classifications. Using the K-Means cluster plot above we see that **Southern US and Spanish recipes** are in different clusters but considered relatively close. Similarly, **Greek and Moroccan cuisines** are considered in different clusters but relatively similar. These recommendations seem reasonable but we can take it a step further and leverage the confusion matrix in an unorthodox way.

The study intends to use the wealth of available recipe data to construct prediction algorithms for predicting flavour, popularity, and nutritional information, using a variety of machine learning methods, including as regression, classification, and deep learning.Support vector machine and associative classification, two basic classification approaches in data science, were used to examine the relationship between different recipes and their ingredient sets and gave a thorough understanding of other cuisines, ingredient patterns, and the key components of a good recipe.Some model classification and predictions are there in my cuisine classification code.

These are as follows:

1)Data preparation: Once you have a good understanding of the problem, the next step is to prepare your data. This includes collecting and preprocessing the data and splitting it into training, validation, and test sets. In this step, the data is cleaned, preprocessed, and transformed into a format that can be used by the classification algorithm.

- X: It is the independent feature, in the form of an N*M matrix. N is the no. of observations and M is the number of features.
- y: An N vector corresponding to predicted classes for each of the N observations.

2)Feature Extraction: The relevant features or attributes are extracted from the data that can be used to differentiate between the different classes.

Suppose our input X has 7 independent features, having only 5 features influencing the label or target values remaining 2 are negligibly or not correlated, then we will use only these 5 features only for the model training.

3)Model Selection: There are many different models that can be used for classification, including logistic regression, decision trees, support vector machines (SVM), or neural networks. It is important to select a model that is appropriate for your problem, taking into account the size and complexity of your data, and the computational resources you have available.

4)Model Training: Once you have selected a model, the next step is to train it on your training data. This involves adjusting the parameters of the model to minimize the error between the predicted class labels and the actual class labels for the training data.

5)Model Evaluation: After training the model, it is important to evaluate its performance on a validation set. This will give you a good idea of how well the model is likely to perform on new, unseen data.

SUMMARY:

In conclusion, classification is a fundamental task in machine learning, involving the categorization of data into predefined classes or categories based on their features.However my purpose was to give a basic understanding of doing this project for beginners so that you can start working on a text data and build classification model.An interesting exception would be to explore configuring learning rate and number of training epochs at the same time to see if better results can be achieved.