INTRODUCTION

This analysis's goal is to assess the model's accuracy by running a KNN (k-nearest neighbors) study on simulated data. The sklearn. Datasets module's make blobs method was used to create the data, which was divided into three clusters with centers at (2, 4), (6, 6), and (1, 9).

METHODS

The train test split function from the sklearn.model selection module was used to divide the data into a training set (80% of the data) as well as a test set (20% of the data). Using the KNeighbors Classifier classes from the sklearn.neighbors module, a KNN classifier with k=5 was built on the training set of data. The precision of the model was determined by utilizing the accuracy score method from the sklearn. metrics package, which was used to predict all labels of the test data using the classifier.

The matplotlib.pyplot module was then utilized to plot the findings. The decision function of the KNN classifier was first superimposed on the plot of the simulated data, which had distinct colors for each class.

RESULTS

The KNN classifier's accuracy on the test data was determined to be 0.9667, a sign that the model was able to classify the majority of the test data appropriately.

The KNN classifier's decision border and the plot of a simulated data. The plot demonstrates that the classifier was able to differentiate the three groups reasonably successfully, however there are certain spots along the boundaries between classes in which the classifier is less sure.

DISCUSSION

With the simulated data, the KNN classifier performed effectively, earning a highly accurate score of 0.9667. The data was generated with very well clusters, so it may be simpler to categorize than real-world data with more intricate and overlap class boundaries.

The KNN classifier's decision boundary shows that it is certain in different parts of the feature space. Given that KNN classification is dependent on the distances to a k nearest neighbors in the training set, this might be caused by the density of training points in such areas.

Overall, the analysis's findings show how well KNN classification performs on hypothetical data and offer a helpful illustration of how it actually functions.