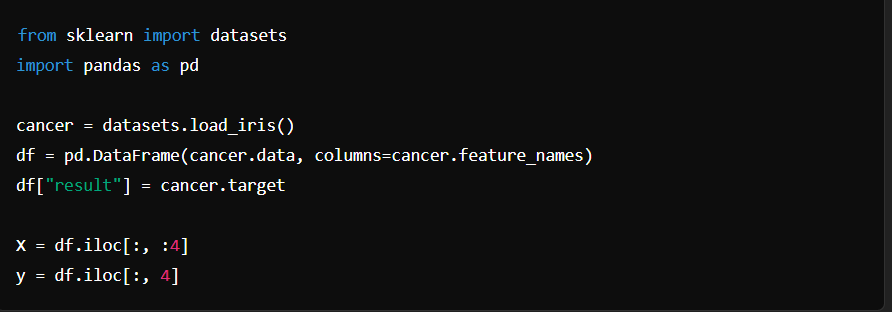
**PROJECT REPORT**

**PROJECT - AUTOML**

**Problem Statement**-The quality of performance of a Machine Learning model heavily depends on its hyperparameter settings. Given a dataset and a task, the choice of the machine learning (ML) model and its hyperparameters is typically performed manually. Develop an automated hyperparameter optimization (HPO) system using AutoML techniques that can efficiently identify the best hyperparameter configuration for a given machine learning model and dataset.

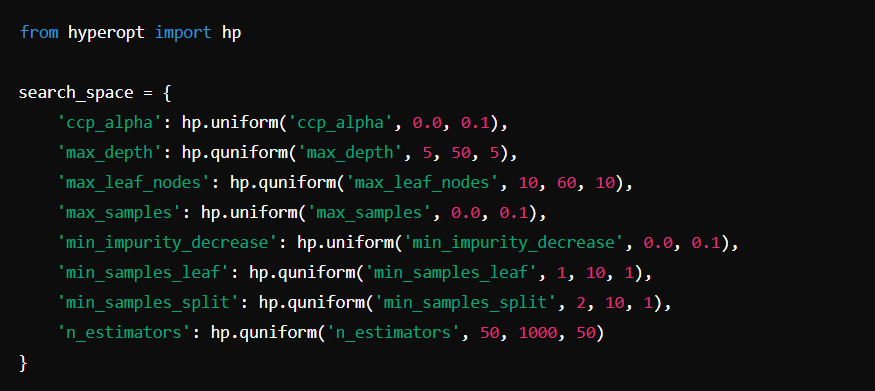
## PROJECT  
  
The objective of this project is to optimize the hyperparameters of a `RandomForestClassifier` using a combination of Kernel Density Estimation (KDE) and the Hyperopt library. By utilizing a probabilistic model-based approach, we aim to efficiently explore the hyperparameter space and identify the configurations that maximize the model's accuracy.

#DATA PREPERATION



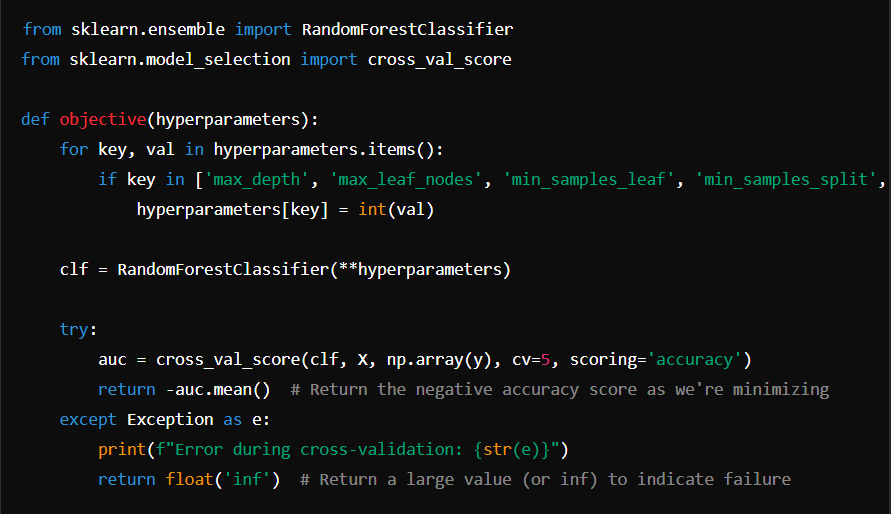
#SEARCH SPACE

The hyperparameter search space for the RandomForestClassifier was defined using the Hyperopt library. The chosen hyperparameters and their respective ranges are:



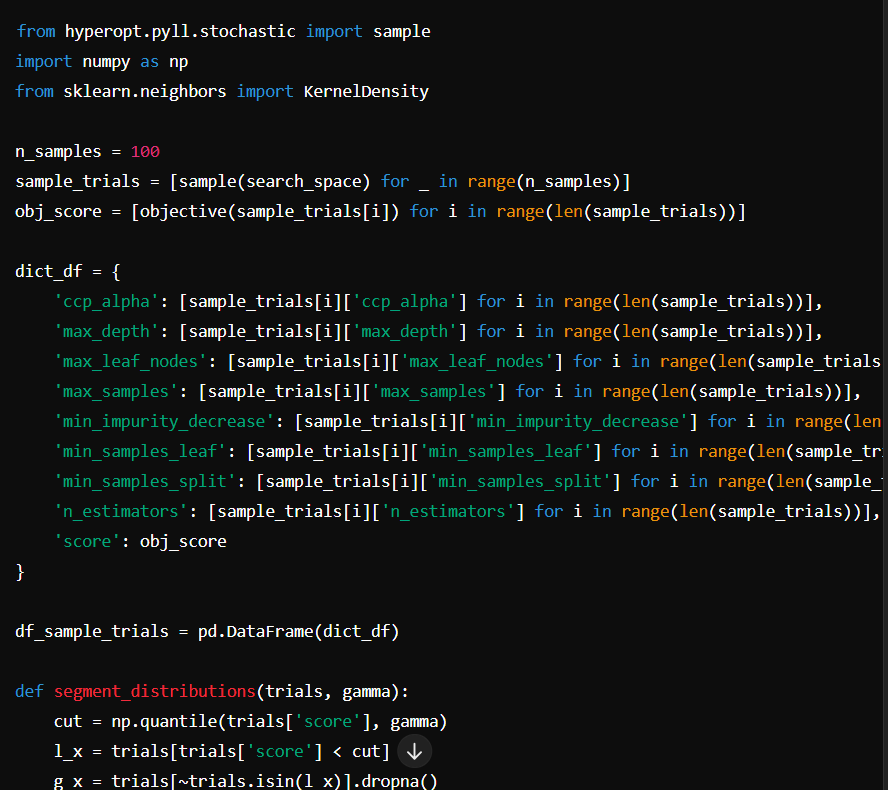
#OBJECTIVE FUNCTION

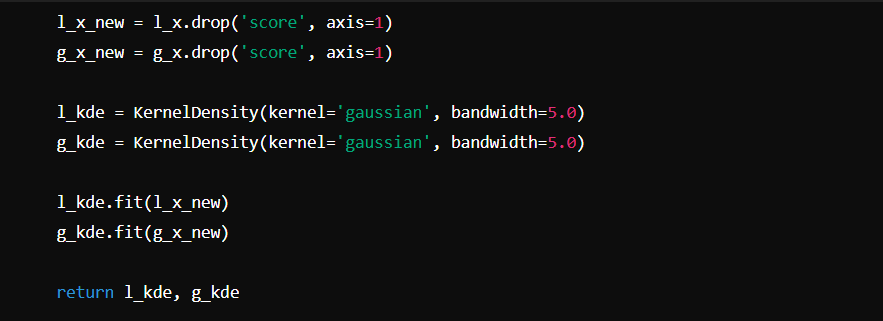
The objective function evaluates the performance of the RandomForestClassifier with given hyperparameters using cross-validation. The aim is to maximize the accuracy score.



# INITIAL SAMPLING AND SEGMENTATION

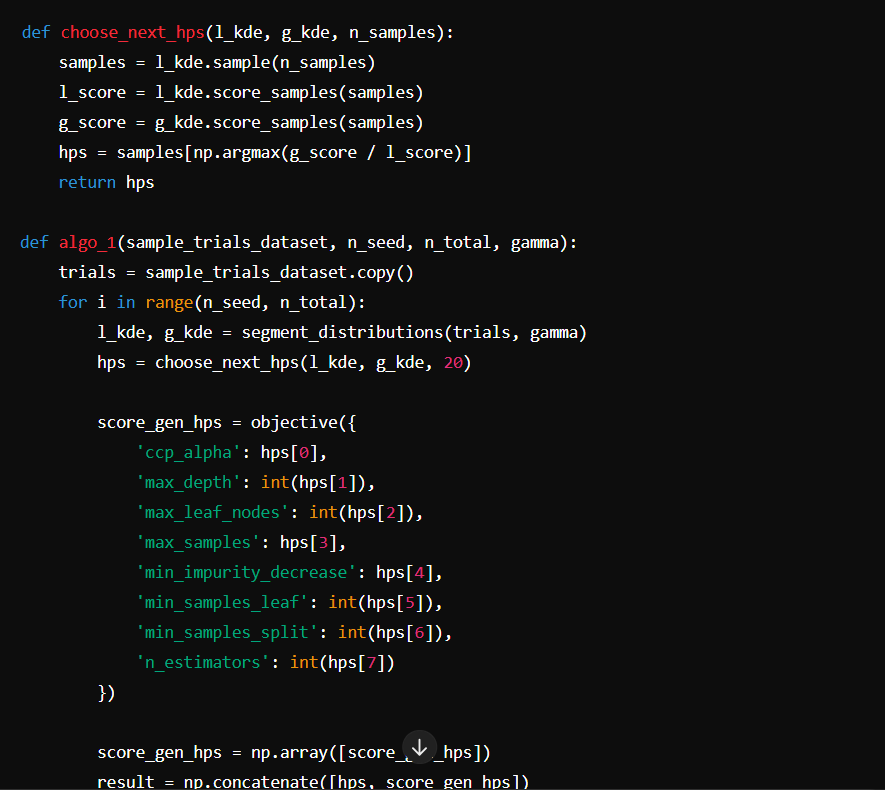
We generated an initial set of hyperparameter samples using the Hyperopt library and evaluated them using the objective function. These samples were then segmented into l(x) and g(x) distributions based on a quantile cutoff.

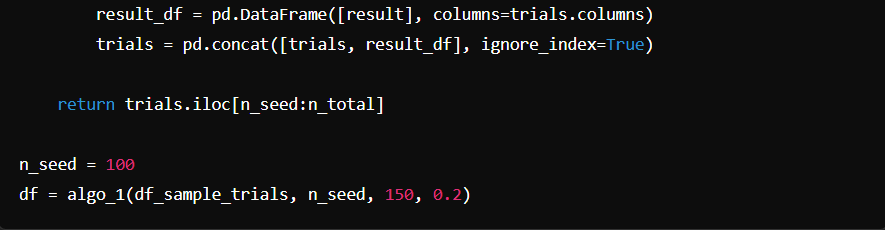




# HYPERPARAMETER SELECTON ALGO

The algo\_1 function iteratively selects new hyperparameters based on the KDE of the l(x) and g(x) distributions. It evaluates these hyperparameters using the objective function and appends the results to the trials DataFrame.





# RESULT AND CONCLUSION

The implementation of the above algorithm allowed for the efficient exploration of the hyperparameter space for the RandomForestClassifier. By iteratively selecting hyperparameters that maximize the ratio of g(x) to l(x), we were able to identify configurations that improve the model's accuracy. The approach demonstrates the effectiveness of probabilistic model-based optimization in hyperparameter tuning.