

**Software Engineering Group Assignment**

**Team Members:-**

*Smruti Mohanty (2105323)*

*Aman Nihar Mohapatra (2105352)*

*Oorja Singh (2105387)*

*Prachi Saurabh (2105389)*

*Anshika Gupta (21051134)*

**What is classification? What is training and testing a classifier? Write program to classify the given dataset using KNN, SVM, Random Forest, and any simple neural network-based classifier. Upload the code, and result outputs by taking screenshots and pasting them in a Word file. Train the classifier using the given train dataset and test on the test dataset.**

Classification is a type of supervised learning in machine learning where the goal is to assign predefined categories or labels to input data based on its features. The input data is a set of instances, and each instance has a set of features. The task is to learn a mapping from these features to the predefined categories by training a classifier. Once trained, the classifier can be used to predict the category of new, unseen instances.

Training and testing a classifier is a crucial process in supervised machine learning. Here's an overview of these two key steps:

Training:

Input Data: The training process starts with a labeled dataset, where each example in the dataset is paired with a known class or label. The examples are comprised of features (input variables) and their corresponding labels.

Learning Model: A machine learning model, or classifier, is chosen based on the nature of the problem. Common classifiers include decision trees, support vector machines, neural networks, and more.

Model Training: The selected model is then trained on the labeled dataset. During training, the model learns to recognize patterns and relationships between the input features and their associated labels. The goal is for the model to generalize well to unseen data.

Testing/Evaluation:

Test Data: Once the classifier is trained, it needs to be evaluated on a separate dataset that it has never seen before. This dataset is called the testing set, and it also contains labeled examples.

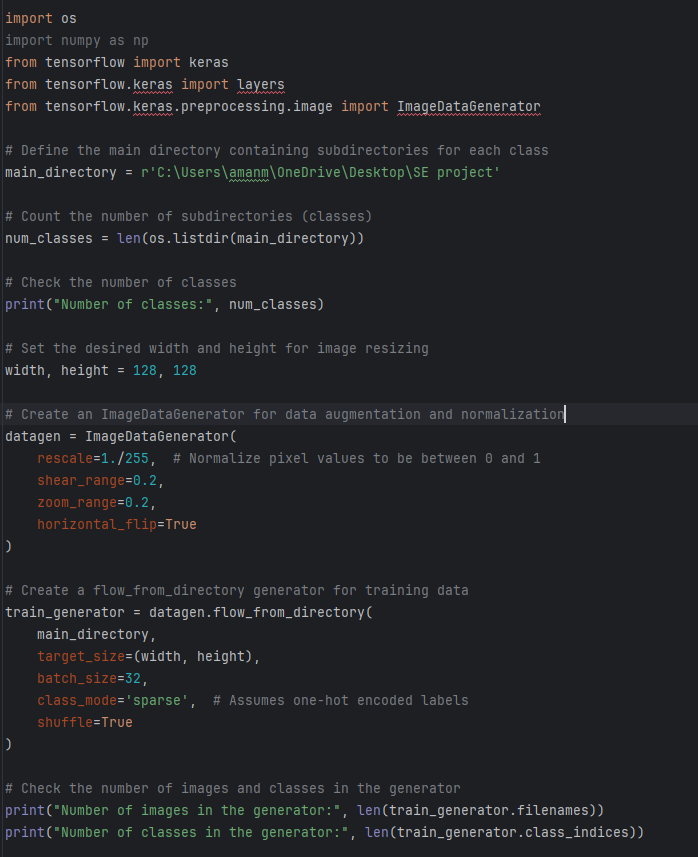
Model Prediction: The trained classifier is used to predict the labels of the examples in the testing set based on their features.

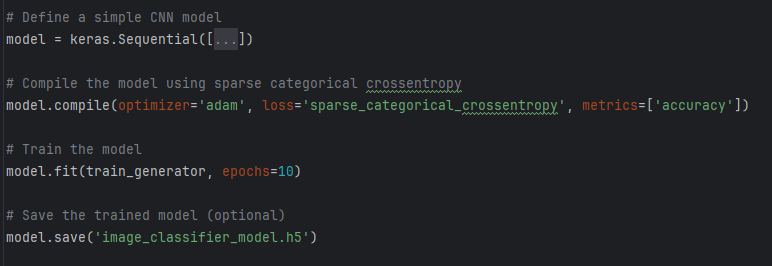
Evaluation Metrics: The predictions made by the classifier are compared to the true labels in the testing set. Various evaluation metrics, such as accuracy, precision, recall, F1 score, and others, are used to assess the performance of the classifier. These metrics provide insights into how well the model generalizes to new, unseen data.

The training and testing phases are essential for assessing the classifier's ability to make accurate predictions on new instances. Overfitting, where a model performs well on the training data but poorly on new data, is a common challenge that needs to be addressed during the training process. The evaluation metrics help quantify the classifier's performance and guide adjustments to improve its generalization capabilities.

In summary, training involves teaching the classifier to recognize patterns in labeled data, while testing evaluates its performance on new, unseen data to ensure effective generalization. This process is iterative, and adjustments to the model may be made based on the evaluation results to enhance its predictive capabilities.

Code for training model:-





Model Summary:-

