**Hindi Letter Recognition**

**System**

A Final Report Submitted

for the Course of

**Assignment 2**

In

First year – second Semester of

MCA

specialization

In

**Machine Learning**

By

**Jharna Yadav**

**Amity online**

1. **Introduction**

* Image recognition and categorization is a very impactful tool in today era. It’s used cases are found in many fields like healthcare, marketing, transportation and e-commerce.
* Image recognition technique is used in facial recognition and other security systems.
* In healthcare its major use is in cancer detection which is very helpful for doctors to diagnose the disease.
* Many e-commerce sites use this technique for better costumer experience.

1. **Problem statement**

In the assignment the question is provided by the institute with specified techniques.

1. **Methodology**
2. **Data Collection –** Kaggle website is used for dataset. Hindi letter images dataset was available on the site.
3. **Data Preprocessing –** It includes converting the images into csv format so that we can use that csv file to complete the task. Two dimensionality reduction techniques i.e., PCA and TSNE are specified in our problem statement for dimensionality reduction.
4. **Building Model –**As specified in our problem statement we use SVC (Support Vector machine) and Neural network to train our model.
5. **Method used**

* This model uses PCA and TSNE as dimensionality reduction technique.
* This project uses two algorithms i.e., Neural network and Support Vector Machine (SVM)

1. **Steps**

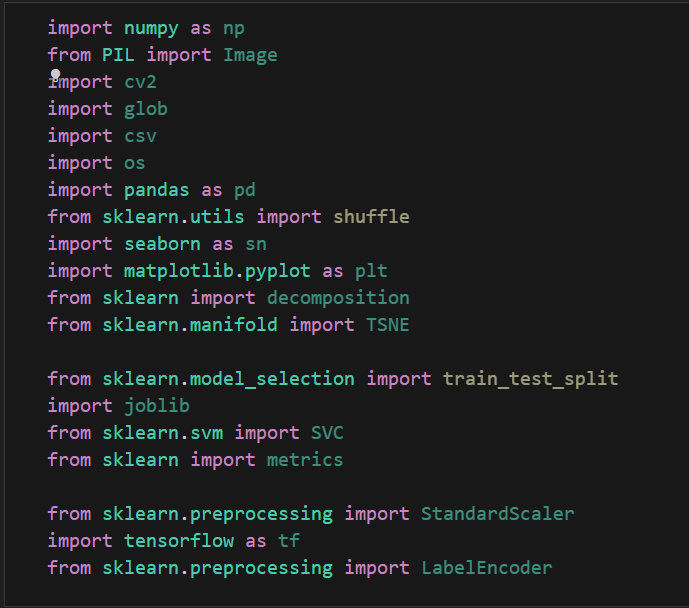
* Importing all the important packages.
* Data pre-processing: Converting the images into pixel data and writing them to a csv file. Reading the pixel data csv file. Selecting the 500 images of each letter.
* Separating labels and features.
* Splitting test and train data and encoding labels
* Mapping the labels with their actual names.
* Scaling the data.
* Applying PCA as dimensionality reduction technique and converting data to pandas-data-frame.
* Applying TSNE as dimensionality reduction technique on PCA data frame and converting data to pandas-data-frame.
* Training the model with SVC and Neural network.
* Hyperparameter tunning and plotting the graph and selecting the one with highest accuracy.
* Additionally, I trained two more models using TSNE and PCA only Compared their accuracy.
* Plotting graph of different accuracies.

1. **Implementation**

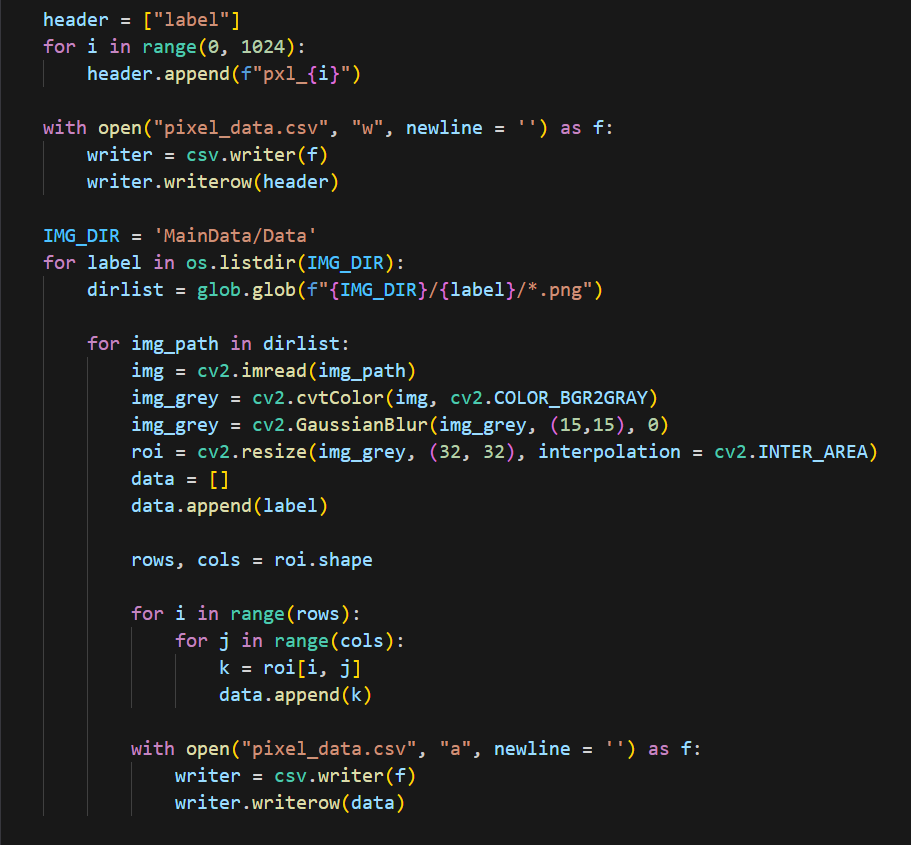
**Collecting Dataset**

[**https://www.kaggle.com/datasets/ujjawalsharma20/hindi-letters**](https://www.kaggle.com/datasets/ujjawalsharma20/hindi-letters)

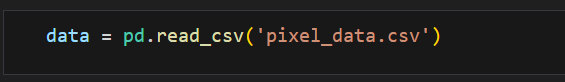
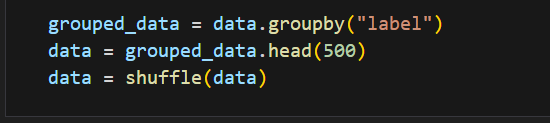
**Importing all the required packages**

****

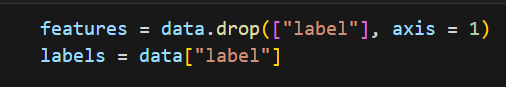
**Converting images data to pixel csv**

****

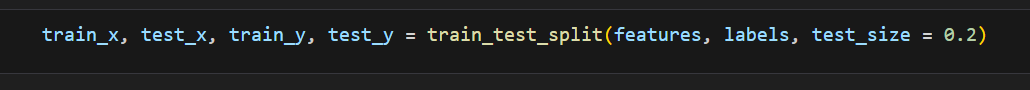
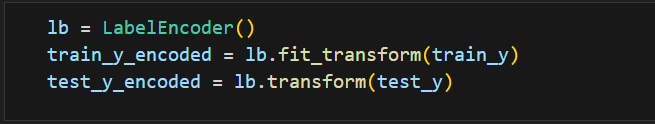
**Reading the pixel data csv file. Selecting the 500 images of each letter**

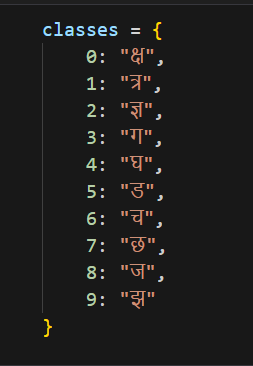
**Separating labels and features**

****

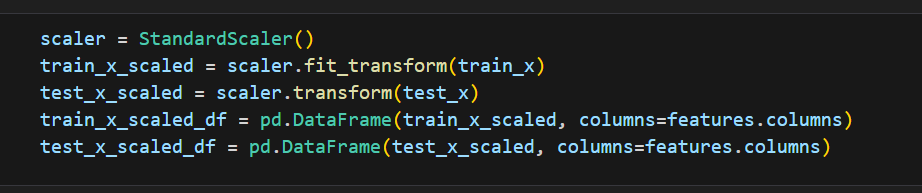
**Splitting test and train data and encoding labels**

**** ****

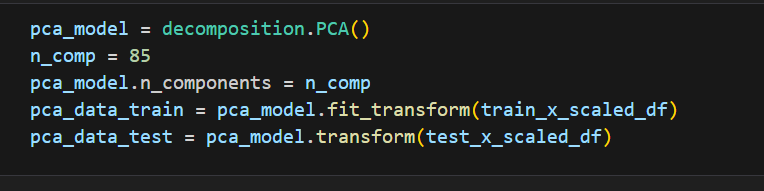
**Mapping the labels with their actual names**

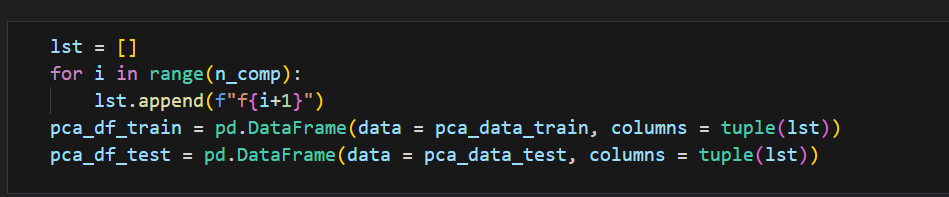
****

**Scaling the data**

****

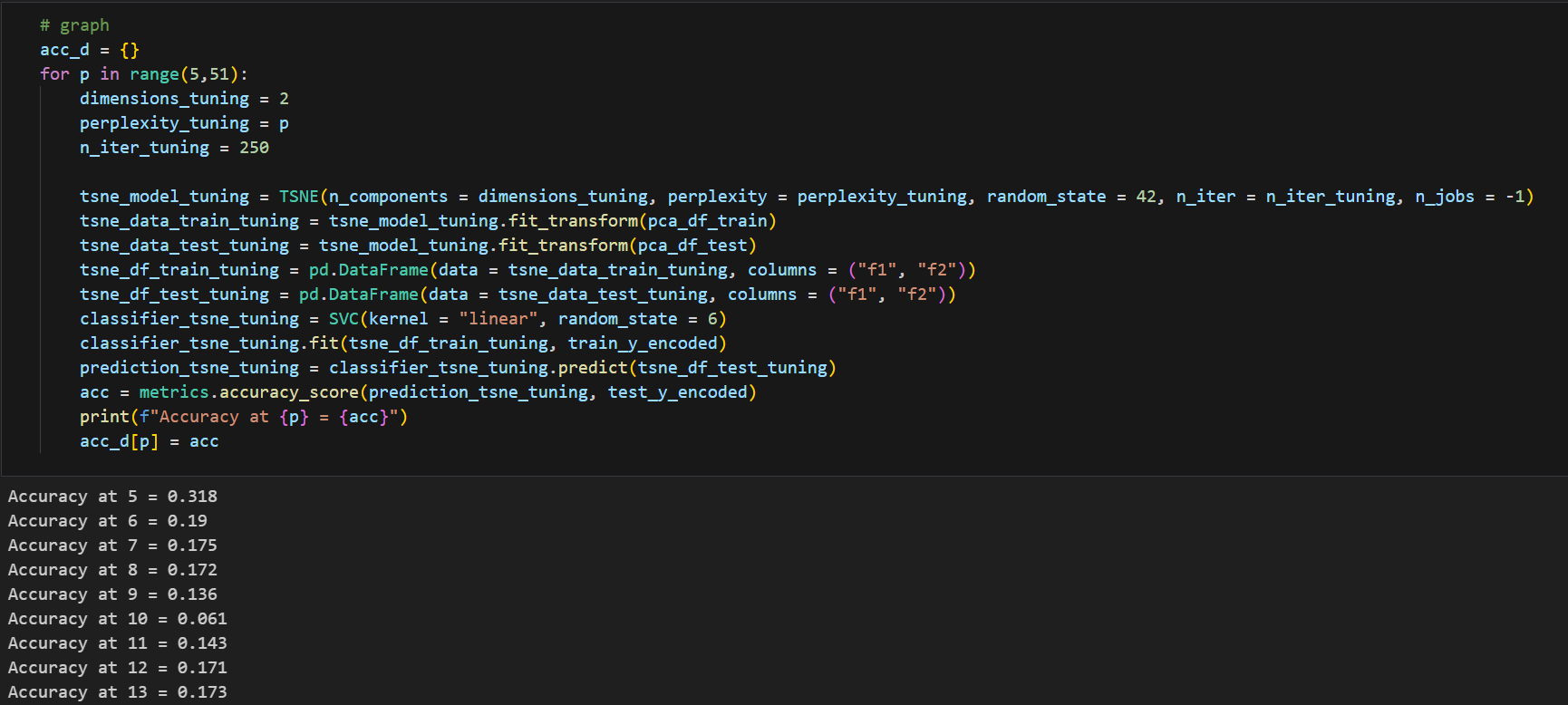
**Applying PCA as dimensionality reduction technique and converting data to pandas-data-frame**

****

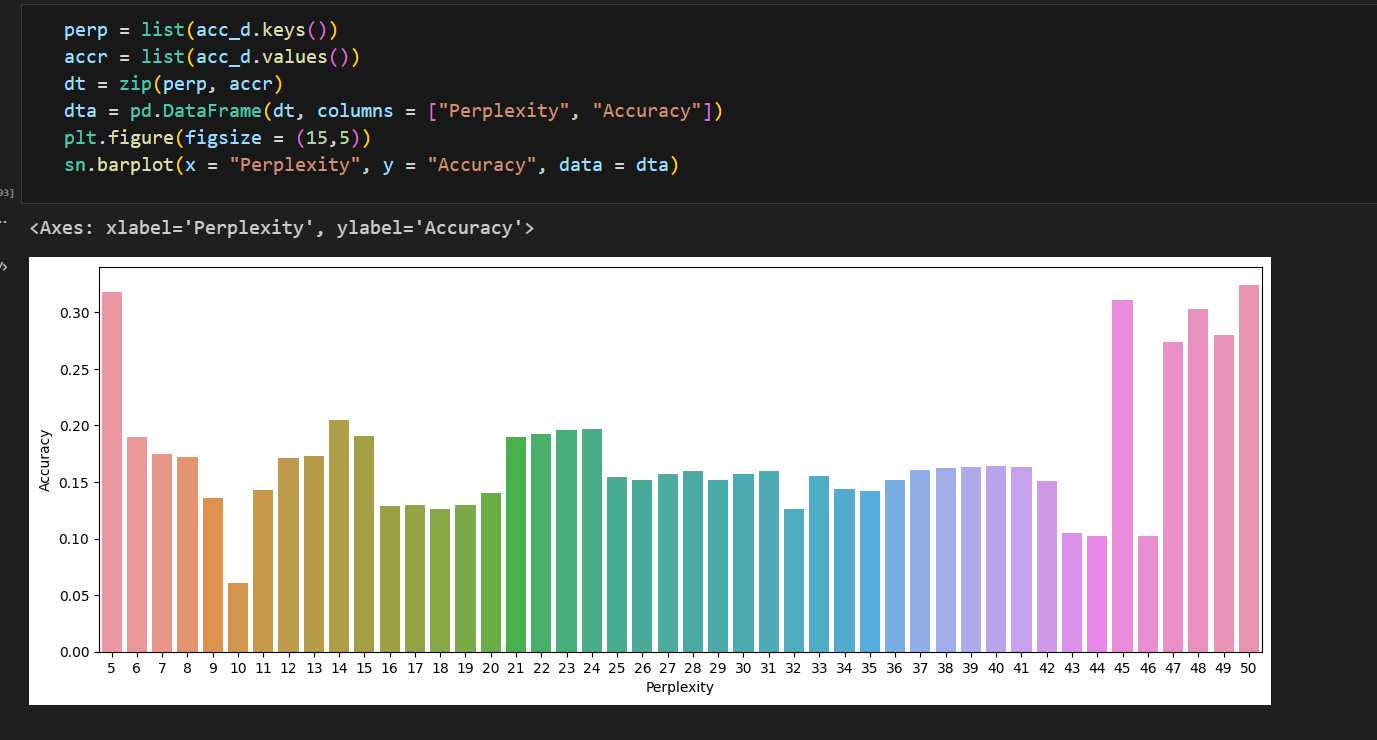
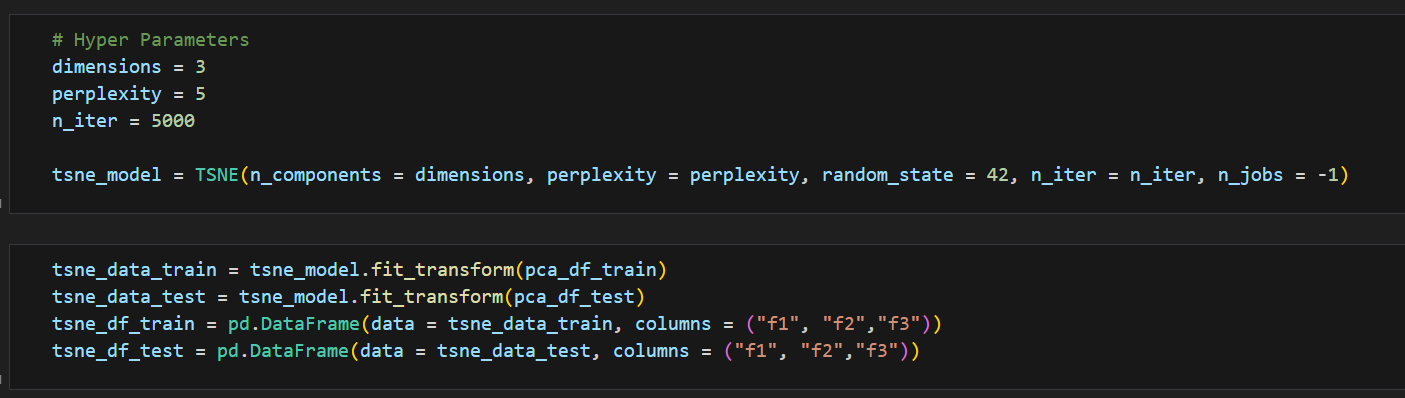
****

**Applying TSNE as dimensionality reduction technique on PCA data frame and converting data to pandas-data-frame.**

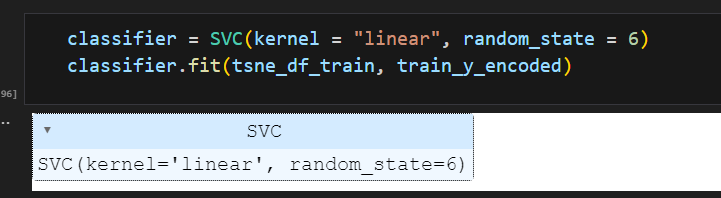
Hyper Parameter tuning for SVC:

****

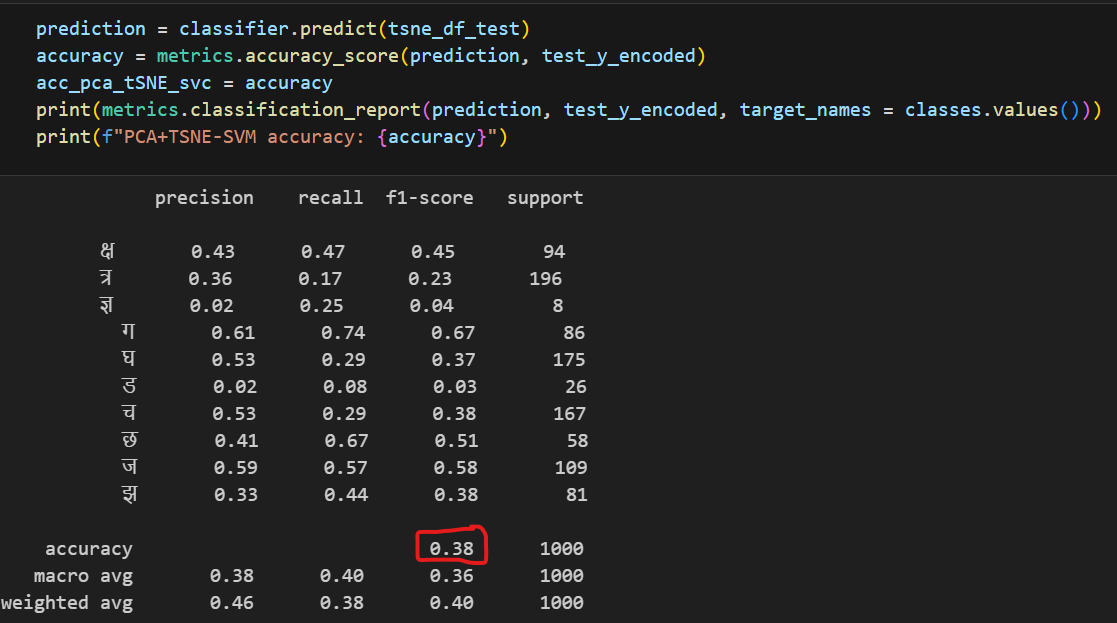
Plotting the graph and selecting the one with highest accuracy

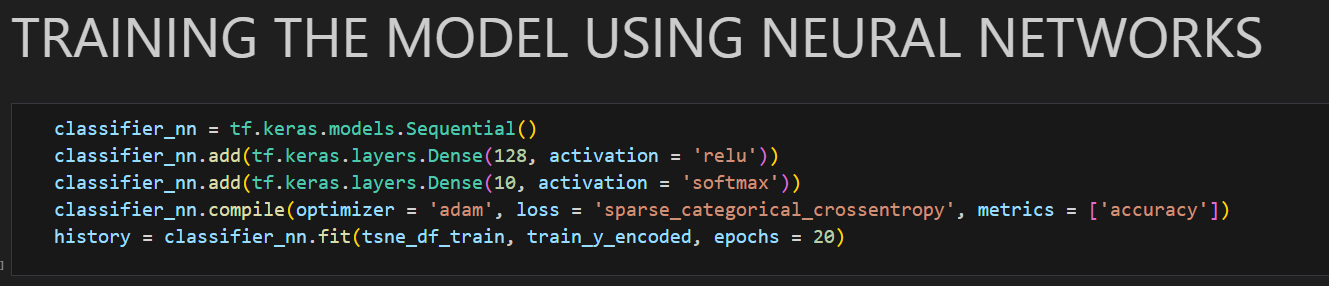
**** 

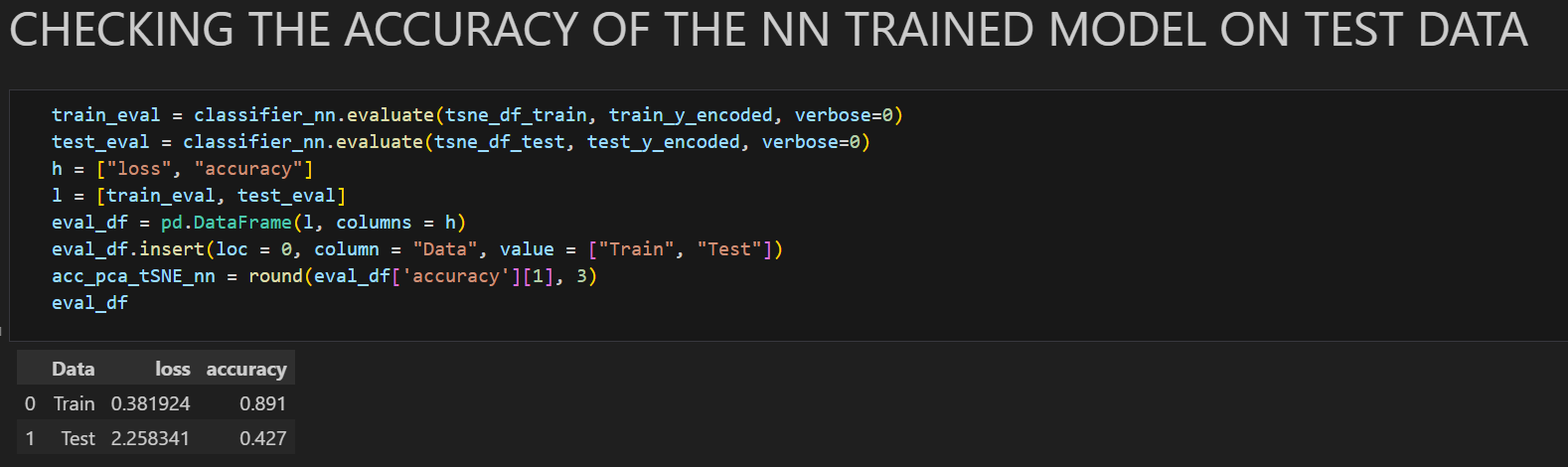
**Training the model with SVC and Neural network.**

****

Checking the accuracy:

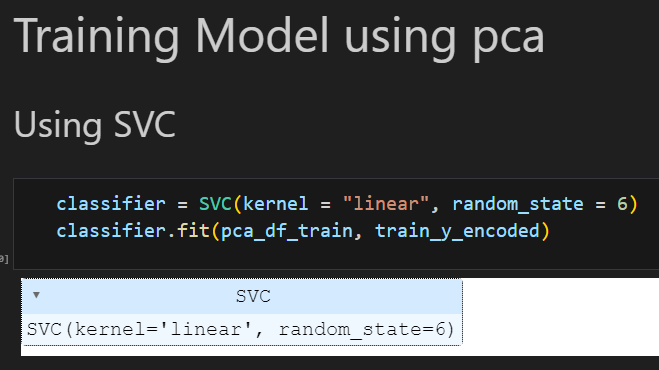


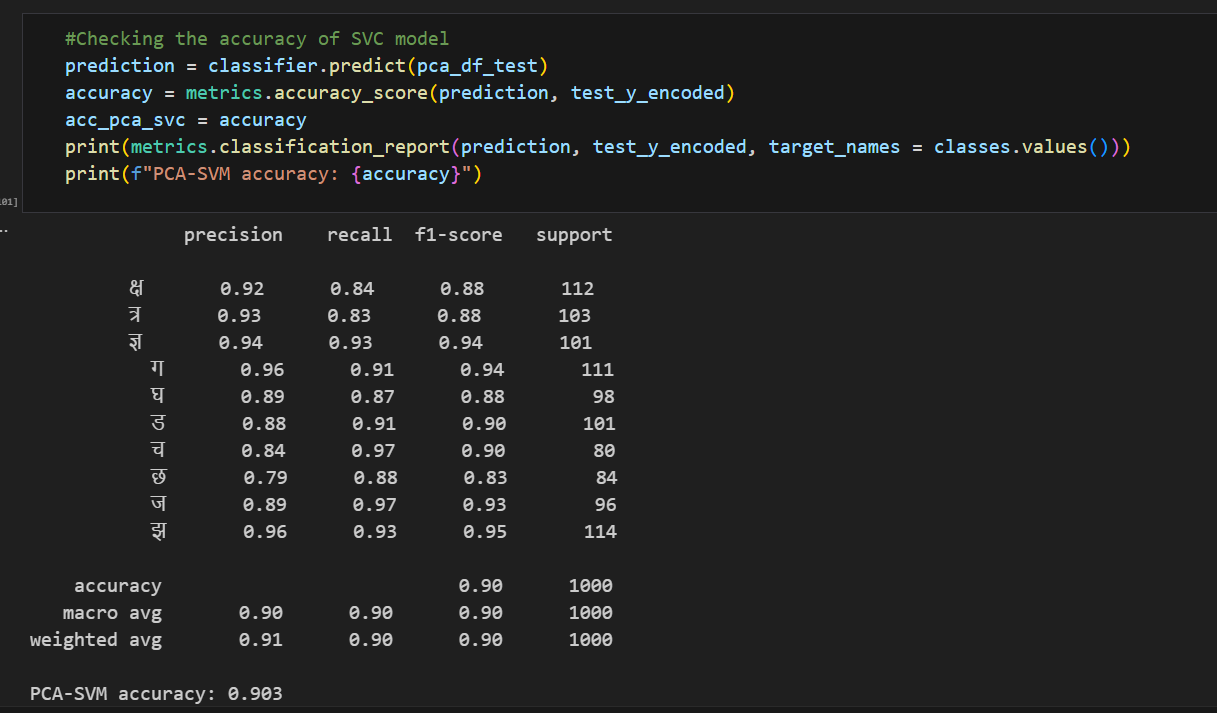




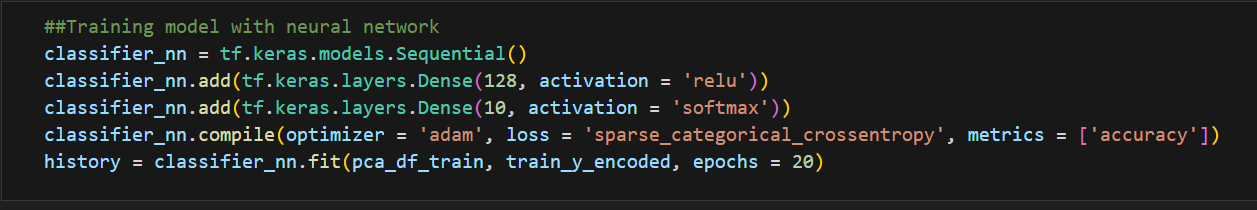
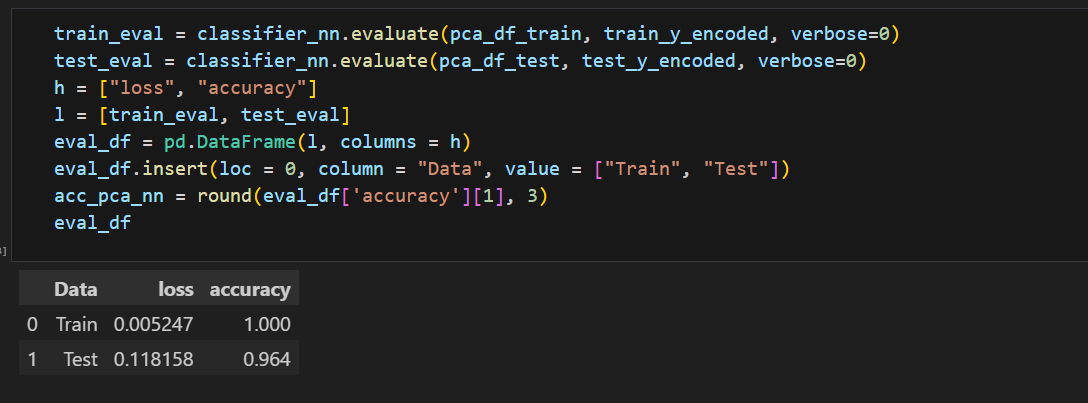
**Accuracy = 42.7%**

**Additionally, I trained two more models using TSNE and PCA separately**

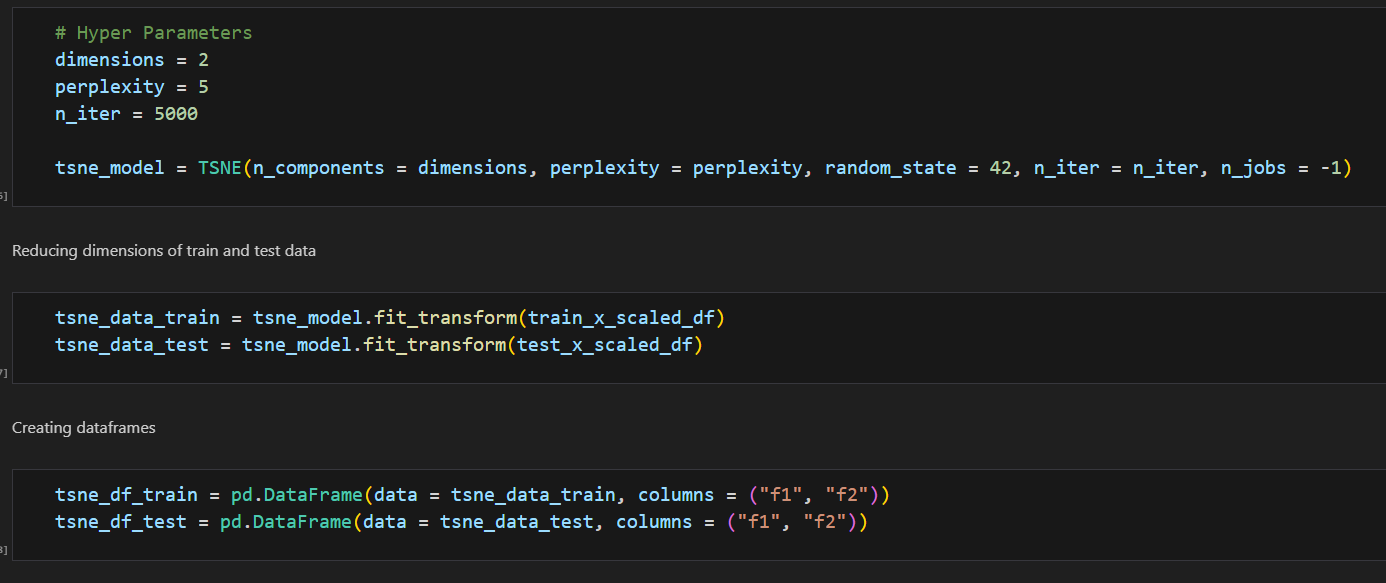
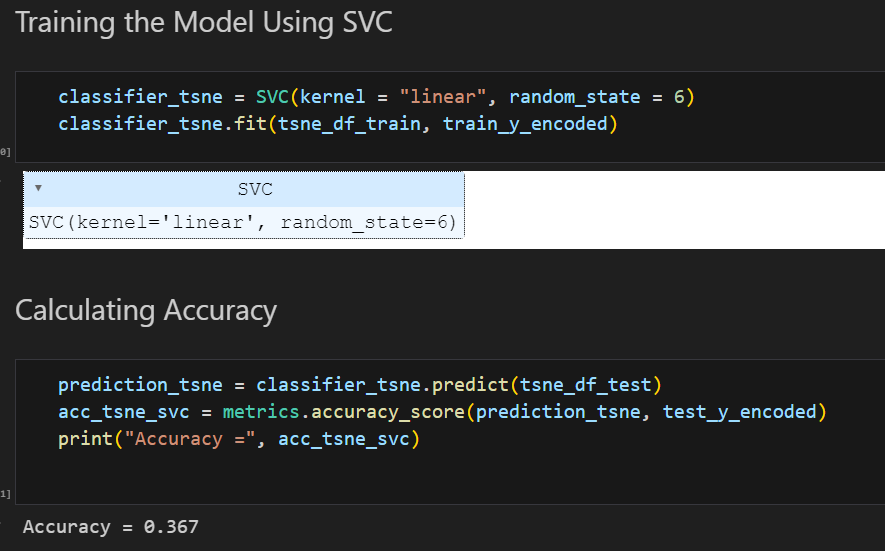
****

****

**Accuracy = 90.3%**

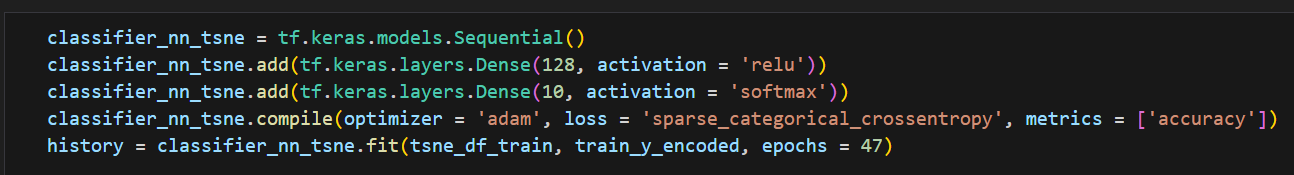
**** ****

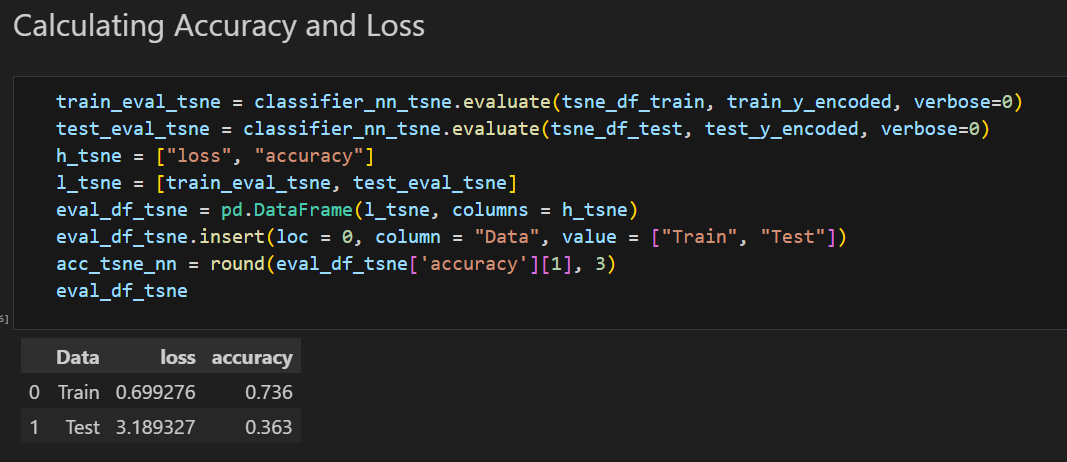
**Accuracy = 96.4%**

Using TSNE only:**** 

**Accuracy = 36.7%**

Using NN after tsne:

****

****

**Accuracy = 36.3%**

**Compared their accuracies of all the trained models and plotting graph**

****

1. **Conclusion:**

* Here we found that when we use PCA and TSNE both as dimensionality reduction technique and train our model through NN and SVC the accuracy is quite low i.e., around 40%.
* When we use TSNE dimensionality reduction tech. alone and train our model the accuracy is still around 40%.
* When we use PCA dimensionality reduction tech. alone and train our model it gives us very accurate results. The accuracy in this case is 97% with neural network.