

Report on Classification Model classifying chemical compound into two categories either as Musk compounds or Non musk compounds.

Preprocessing step:-

In dataset , two columns are present as strings which should be converted into numeric format as machine learning models only process numerical values and not string vaalues so i have used Bag of Words techniques for converting strings to numeric format.

molecule_name and conformation_name are converted into numeric format as they are present in string format.

I have break the whole dataset into train and test with ratio 80:20 by using train_test_split function.

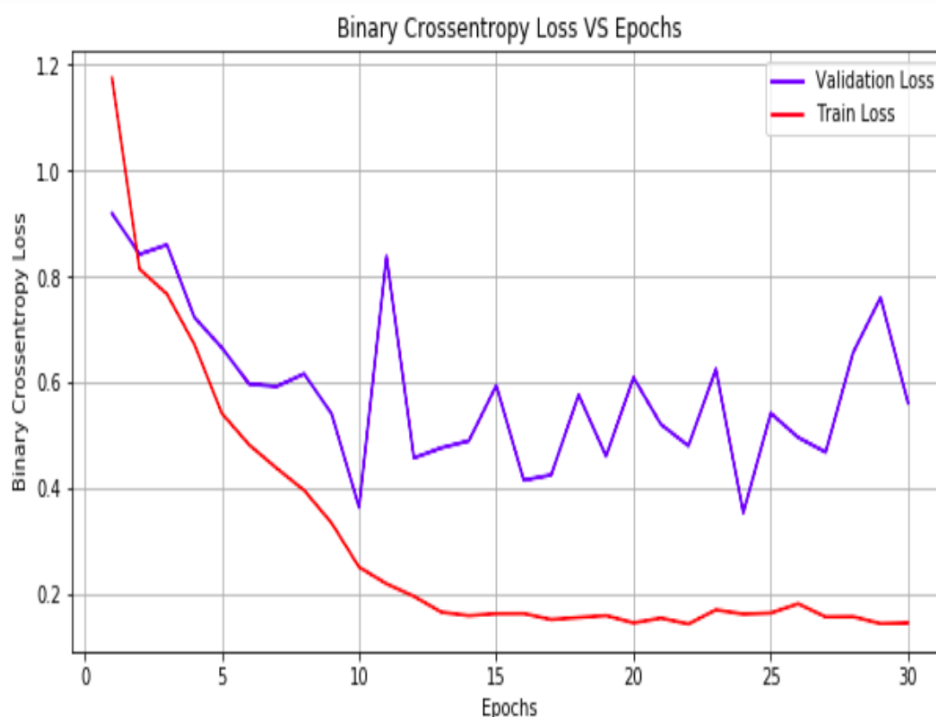
Missing Values:-

No missing values are present.

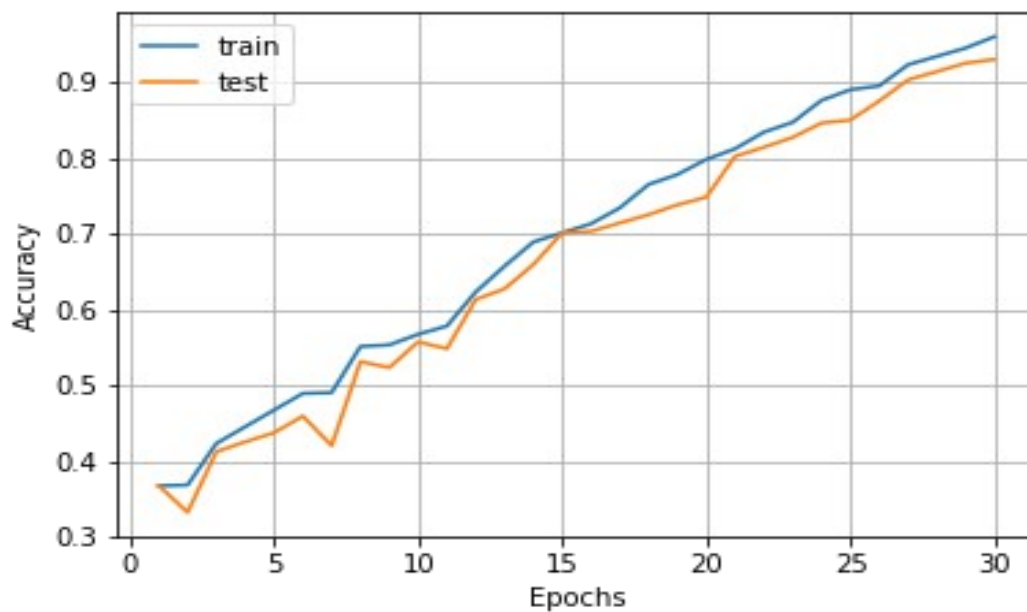
Model:-

I have used Covolutional neural network for modelling as the hierarchical structure and powerful feature extraction capabilities from document makes CNN a very robust algorithm for various classification tasks.

Loss Graph



Accuracy graph:-



Classification Report:-

Validation accuracy-- 0.92

Loss -- 0.56

Precision -- 0.85

Recall -- 0.82

F1 score -- 0.93