**REPORT**

Internshala Assignment Problem.

Chemical Compound Classification.

The Model:

The model is made using keras's Sequential API, it consists of 3 dense hidden layers and an output layer with one unit. The hidden layers utilize the RELU activation function and the output layer has the sigmoid activation function.

PREPROCESSING:

The dataset had 170 columns with 'ID', 'conformation\_name' being unique features and 'class' as target ('molecule\_name' feature was not used at it mentioned if the compound was musk or not). Now, with 166 features, Principal Component Analysis was carried out and plotted for cumulative explained variance vs number of components. For 25 components, cumulative variance was about 90%. Hence, features were reduced to 25.



Finally, the model was trained on the Training set and predicted on Validation set (80:20)





The Predicted metrics on Validation set are,

