Write-up.

Data in hand:

- Smiles Representation of the Chemical Compounds. from ZINC-15 database

Objective — Generate 3 new Molecule.

Hint - given - VAE

Step - 1

Data Preprocessing:-

The Smiles data had to be numerically Represented. hence. characters were one hot encoded.

For Simplicity only elements with max 20 characters were chosen from ZINC15Database.

And. 35 classes of characters (.unique).

Converting them on the fly to one hot encoding using data generator.

Model:

Encoder. → The Role was to generate a LabelVector of length 256.

1D convolution was used on the data. (Non. unfazed to. variable Smile length).

Decoder: was an LSTM model.

Reconstructing the one hot encoded output.

VAE loss: Sparse_Softmax cross entropy was used.

Training: → Trained for 6 epochs gave an accuracy of 96%.

by collecting the means from encoder new latent space was Created generated and Decoded with Smile characters.