School of Computer Science Engineering and Technology

|  |  |
| --- | --- |
| Course- BTech | Type- Elective |
| Course Code- CSET335L | Course Name-Deep Learning |
| Year- 2023 | Semester- Even |

**Lab Assignment 10.1 – Model Compression**

This Lab focused on solving of problem with compressing a deep learning model. Make any small deep learning model and perform the following tasks:

1. Convert the weights of the model to zero and one. Any weight smaller than zero, convert it to zero. Any thing greater than zero should be one.
2. Repeat the previous task by converting the weights to minus 1 and plus 1.
3. Perform quantization and trim the weights to smaller precision values.
4. Perform quantization to integer values and compute the distribution of decimals values. At runtime compute the weights values and perform classification.
5. Perform clustering and replace the weights of the model via its centroid.

You can choose any dataset you like.