COURSE CODE: 16B11BI612   MAX. MARKS: 25

COURSE NAME: Datawarehousing and Mining for Bioinformatics

COURSE CREDITS: 03   MAX. TIME: 1.5HR

Note: All questions are compulsory. Carrying mobile phone during examinations will be treated as a case of unfair means.

1. Explain the meaning of the following terms in the context of data mining: (1.5 *6=9)
   (a) Unique rule
   (b) Consecutive rule
   (c) Null rule
   (d) Field overloading
   (e) Entity Identification Problem
   (f) Curse of dimensionality

2. With reference to the work of Ahmadi et al. 2015, explain the following: (10)
   (a) Why is it important to identify the inhibitors of P2X7 receptors? (2)
   (b) What are the samples in this dataset and which features are used to describe the samples? Make a hypothetical data matrix. (2)
   (c) Explain the methodology followed by them. Did it perform better than the traditional QSAR study? (5)
   (d) Will you use PCA for your future research based on machine learning and for what purpose? (1)

3. Enlist and elaborate the various ways of handling noisy data. (3)

4. Make your own hypothetical matrix and explain the meaning of eigen vectors and eigen values. (3)