GROUP 82

MEMBERS : 1)JYOTIRMOY ROY(180020044) 2)ESHA KOLTE(180020031)

ABSTRACT:

1. **Domain** : Biostatistics and Data Science in Bioengineering

2. **Libraries used**:matplotlib,pandas,tensorflow,sklearn,seaborn

3. **Implementation** :The projects will deal with analysis of a set of protein levels obtained from various meningioma patients. Meningioma or brain cancer are basically of three kinds : Groups I, Group II and Group III, out of which Group I is more benign and the other two are more lethal. The aim of the project will be to identify the set of proteins which can act as a potential biomarker based on their differential expression in various patient groups. We plan to use the data set to train various machine learning models like random forest, Naive Bayes, KNN and try to figure out which one of them works the best based on precision, recall and AUC values. Other than this we also plan to visualise the data using heat maps, boxplots and use PCA and PLSDA plots to identify the clusters formed

4. **Small rough timeline for 2 weeks**: In the first week, we will do the data visualisation and feature selection and in the second week, using the selected features we will train various machine learning models.

5. **What other things you need to learn that weren't a part of our course**: sklearn to apply machine learning algorithms. In case we are not able to catch up with machine learning as of now, we plan to use orange software as backup.

6. **Motivation**: The peptides that distinguish grade I meningioma from higher grades are still not properly discovered. In other diseases, normal statistical methods like fold change criteria and p values help in finding the differentially expressed proteins for that disease. However in meningioma samples, the variations in the samples are so much that normal statistical tests do not give reliable results. So we plan to use unsupervised clustering in the form of PCA and then use the topmost peptides in designing the ML model to separate a severe patient from a mild patient.

7. **References**
i)AVAILABILITY OF DATASET:
This dataset is already available with me as apart of my project under Prof. Sanjeeva Srivastava(BSBE,IITB)
ii)Dunn J, Ferluga S, Sharma V, Futschik M, Hilton DA, Adams CL, Lasonder E, Hanemann CO. Proteomic analysis discovers the differential expression of novel proteins and phosphoproteins in meningioma including NEK9, HK2 and SET and deregulation of RNA metabolism. EBioMedicine. 2019 Feb;40:77-91. doi: 10.1016/j.ebiom.2018.12.048. Epub 2018 Dec 26. PMID: 30594554; PMCID: PMC6412084.