CBIT- Winter UpSkilling January 2025 - Assignment-1

Google sheet Link:

https://docs.google.com/spreadsheets/d/1bJil4Eq9TXIclZjYP-sLV8Rwbf_Bk1eDZQuw6Fe0-pY/edit?usp=sharing.

All the Students are informed to complete the below work today

1. Load your working link in the above google sheet, give me edit access to dlsrinivasareddy_it@cbit.ac.in, tprathima_it@cbit.ac.in, asirisha_it@cbit.ac.in
2. Upload your code in GitHub by creating your own account with your college official email. Make them to public under CBIT Winter Upskilling Tag.
3. Simulate a dataset for Classification with 3 Labels, 6 attributes with interaction among them, 10000 instances and use Uniform random function
4. Simulate a dataset for regression 6 attributes with interaction among them, 10000 instances and use normal random function
5. Simulate a dataset for Clustering 6 attributes with interaction among them, 10000 instances and use normal random function.
6. Build a Non Linear Regression model and get 90% Accuracy for the above simulated data in question2. Split the data into 70% and 30% respectively for Training and Testing for model building
7. Build any five Classifiers for the above dataset in question1 and compare the metrics of these built models. Metrics need to build are Accuracy, Precision, Recall, F1Score. Split the data into 70% and 30% respectively for Training and Testing for model building
8. Load CSV, EXCEL, EXCELS, JSON format datasets from local host and Websites into Colab interface and split them into Train and test data.
9. Load different datasets from SkLearn library into Colab interface and split them into train and test data. Apply Normalization techniques to convert input data into standardize.
10. Load any data from website/Local host and apply the below data wrangling techniques
    a. Split b) merge c) select rows by conditions d) select columns by conditions e) Rename columns f) Min, max, avg, varience g) describe h) size I) shape j) find unique values k) find missing values and fill them with average value/ mode value l) replace attribute value name with other name m)delete columns n) delete some rows o) join datasets(left, right) p) groups values by rows q) loop over columns r) group rows by time s) concatenating datasets t) append rows u)append columns w) reorder columns using pandas lib

11. Load the Images and audio files. Covert audio file into spectrogram. Apply it for a length of one minute telugu song. Convert a colour image into gray scale image.
12. Load any numeric data and apply Ridge and Lasso regressions to get 90% accuracy
13. Load Digits dataset and apply clustering models to get its metrics
14. Implement the techniques to handle imbalanced dataset to balanced dataset with binary class data
15. Apply KNN classifier on appropriate data. Find best neighbourhood size, Indices of neighbourhood, distances of neighbourhood by taking a random new tuple.
16. Find support vectors and their probabilities by applying SVM on any binary classification dataset. Also find the best model to generate support vectors. Get the metrics for this trained model. Represent its graph also.
17. Load a mixed type features data and apply appropriate model. Predict class for any new tuple and its probabilities.

Time Line: 11:59pm of 07.01.2025

Marks Allotted: 10