****

**Work Integrated Learning Programmes Division**

**M.Tech (Data Science and Engineering) Machine Learning DSECLZG565**

**Second Semester, 2022 -23**

**Assignment 1 – PS15 CSM (Conventional and Social Media Movies)**

|  |  |  |
| --- | --- | --- |
| **Name** | **Bits ID** | **Contribution** |
| M THEJASWINI | 2021SC04822 | 100% |
| ISHWAR CHANDRA TIWARI | 2021SC04556 | 100% |
|  |  | 100% |
|  |  | 100% |

**PART A**

**(5-marks) Research**

Select the research paper of your choice.

Attach the chosen paper along with the assignment submission.

Write a synopsis and find below pointers:

3. Paper Contribution

4. Data Pre-processing

5. Machine Learning Activity

6. Result analysis with metrics used from paper

7. Exploratory Data Analysis / Visualization

**PART B**

**(15 – marks) Dataset-based Implementation**

Refer to the dataset mapped against your group.

Use python based APIs and perform the following three classes of activities.

**EDA 1**. Perform Exploratory Data Analysis to gather insight from the dataset. Write your inference about the analysis learned from visualizations (minimum 3) [3]

**Classification**. Any of the Logistic Regression / SVM / Decision Tree/ Naïve Bayes/KNN/ANN.

Justify your design choices at each step:

Write as a markdown cell in jupyter notebook at the beginning of each subsection.

**1. Perform and explain necessary pre-processing / feature engineering on this dataset [0.5]**

**2. Perform the Machine Learning activity. Explain the choice of target attribute, classification type, model selected with reason [1.5]**

**3. Quantify and explain the quality of your ML model. Explain the choice of evaluation metric [1.5]**

**4. Your observation about the results (Hint: comment on the problem statement and conclude the effectiveness of the machine learning activity) [0.5]**

**Regression**. Any of the Linear Regression (any of Gradient / Stochastic / MiniBatch)/linear basis models/KNN/Locally weighted regression/ any of the regularization techniques).

Justify your design choices at each step:

Write as a markdown cell in jupyter notebook at the beginning of each subsection.

1. Perform and explain necessary pre-processing / feature engineering on this dataset [0.5]

2. Perform the Machine Learning activity. Explain Attributes of interest, Regularization type with reason, model selected with reason [1.5]

3. Quantify and explain the quality of your ML model. Explain the choice of evaluation metric [1.5]

4. Your observation about the results (Hint: comment on the problem statement and conclude the effectiveness of the machine learning activity) [0.5]

**Ensemble ML**.

Justify your design choices at each step:

Write as a markdown cell in jupyter notebook at the beginning of each subsection.

**1. Perform and explain necessary pre-processing / feature engineering on this dataset [0.5]**

**2. Perform the Machine Learning activity. Explain Attributes of interest, base classifier chosen with reason, model selected with reason [1.5]**

**3. Quantify and explain the quality of your ML model. Explain the choice of evaluation metric [1.5]**

**4. Your observation about the results (Hint: comment on the problem statement and conclude the effectiveness of the machine learning activity) [0.5]**