Class Project

From the Movies Database
[https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset] extract
a subset of its 9,000 movies identified by the column C in the file
"links_small.csv".

Take Average Ratings as the decision attribute (with 10 values: 0.5, 1, 1.5,
2, ..., 5). The values of this decision attribute for all 9,000 movies should be
computed from the data available in the file "ratings.small".

Propose 4 new classification attributes and add them to your decision table.
Explain why these new attributes are correlated in your opinion with
customer ratings of the movies.

Build several classifiers from the data available in your decision table (with
and without your four additional attributes) using Orange or Weka and
compare their performance. Use F-score for that purpose. Clearly, it is
expected that the classifier built from the decision table with four additional
attributes will have a greater F-score than the classifier built without these
attributes.

Finally, find several action rules (using Lisp-Miner) showing
recommendations on how the ratings of some movies could be improved.