German University in Cairo
Department of Computer Science
Dr. Mohamed Karam Gabr

CSEN 933 Artificial Intelligence, Winter Term 2025
Project 2: Support Vector Machine

1. **Project Description:** In this assignment, you are asked to perform a case study on a business problem of your choice (or continue on the one of a previous assignment). Such a study is to be performed with the aim of implementing a Support Vector Machine model as taught in class. If you wish to continue on the dataset previously used, you will have to state that, alongside rewriting the points below again. If you wish to use a new dataset, these requirements are to be met:

   - Defined the problem on which the case study will be performed.
   - Find an appropriate data set for the previously defined problem on Kaggle.
   - Assess which of the attributes of the data set shall be considered as the information to be operated on.
   - Clarification of the reasoning behind the decision of using either item-item or content-base filtering (or both).
   - Provide the code for the Support Vector Machine in Python.

   This is to be fulfilled through the two deliverables discussed in further details bellow.

2. **Groups:** You may work in groups *at most* five, Team members are allowed to be from different tutorial groups. If a team from assignment 1 wished to be changes, you are required to work on a new dataset. No two teams should use the same report from the previous assignment.

3. **Deliverables:**

   a) Source Code:

   - You can work on any IDE you like, preferable Kaggle (for easier dataset import).
   - You should implement a code section showing the analysis of the chosen attributes for the predictive model to be based on, showing all chosen preprocessing techniques. **[3 marks]**
   - You should implement a code section showing the analysis for the attribute to be considered at the output to be predicted. The analysis is performed similarly to the previous step. **[3 marks]**
   - You should implement a code section for the usage of the Support Vector Machine model as per the input and output attributes. **[5 marks]**
   - Your implemented Support Vector Model can be the predefined class from scikit learn, more information in this link `https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html`. Or it can be from the implementation that was given in Assignment 6 code solution, with taking care of changing whatever is needed for the model to run accurately. **[2 marks]**

- You should implement code section for representing the calculated output and the comparison of it with the true values by calculating the accuracy by any accuracy measure. Usage of the helper functions in Assignment 6 code solution will do, other ways are accepted too. **[2 marks]**

b) Project Report: If you already did this in assignment 1, you are allowed to use it again for this assignment.

- You should provide a discussion about the problem of choice, The challenges why this problem is important, and the methodology used in assessing such a problem. **[3 marks]**
- You should provide a discussion about the reasoning behind the choices made for the data set of Kaggle, in order for it to be taken as the data set for this case study. **[3 marks]**
- You should provide a discussion about the importance of the attributes which are chosen as the attributes upon which the predictive model is based, alongside the attributes to be predicted. **[4 marks]**
- You should provide a reasoning for each preprocessing technique chosen for each of the chosen attributes for prediction. **[4 marks]**
- You should provide a discussion about the conclusions documenting the end results and probabilities produced by the constructed Markov Model showcasing whether the achieved results are satisfactory or not. **[3 marks]**
- If using the previous dataset, you are required to add a part that compares between the results from the Markov Model, and the Support Vector Machine results. You ar also required to declare which had better results with your dataset and the reason for it. **[3 marks]**

4. Submission

- The source code and the report are due by Sunday, November 30, at 23:59.
- Directions on how to submit will be provided in due time.