COSC 2820/2815 Advanced Programming for Data Science

Assignment 3: NLP Web-based Data Application

Milestone II: Web-based Data Application

Assessment Type	Individual assignment. Submit online via Canvas→Assignments→Assignment 3→ Milestone II: Web-based Data Application. Marks are awarded for meeting requirements as closely as possible. Clarifications/updates may be made via announcements/relevant discussion forums.
Due Date	Week 12, October 17th 11:59 PM
Marks	10

1. Overview

In Milestone I you built various machine-learning models using different document vector representations to recommend a clothing item review. In Milestone II, you will develop an online shopping website based on the Flask web framework. The website allows online shoppers to browse clothing items and lets shoppers create new reviews. This website will make use of one machine learning model that you trained in Milestone I. It will be used to create a recommendation label to assign to newly created reviews. This feature will help to aid in creating labels for item recommendations for other website users and improve the user experience of online shoppers.

Data source: https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews
Note: The dataset has been modified for the course, hence, it is not in its original state.

2. Learning Outcomes

This assessment relates to the following learning outcomes of the course:

- CLO 5: Document and maintain an editable transcript of the data pre-processing pipeline for professional reporting;
- CLO 6: Build small to medium-scale data-driven applications using a Web development framework.

3. Assessment Details

In this milestone, you are required to build an online shopping website using Flask.

Clothing Data

To test and demonstrate the functionality of the website, you will need some clothing review data. The dataset could be downloaded from the Canvas (Assignment 3 – Milestone II). Two new columns "**Clothes Title**" and "**Clothes Description**" are artificially generated to add information to the website. The rest of the data is the same as Assignment 3 Milestone I. If you want, you can also create any additional artificial data (e.g. images) for the purpose of display (this is optional).

Website Layout and Design

The layout and design of the website is flexible. You can refer to the exercises in the material from Week 10 and 11, or an online shopping website (such as theiconic.com.au)

Minimum Functional Requirements

There are minimum functional requirements of the developed website for online shoppers and store owners/developers, respectively. Feel free to include other functionalities that suit you.

Functionality for Online Shopper

Item Search

The online shopping website will allow online shoppers to effectively search for an item job listings that are of their interest. The search could be based on the categories, which could be used as keyword for search. Upon user entering a keyword string, the develop system should return a message saying how many matched clothing items have matched, and it also returns a list of item previews that are relevant to the keyword string. When users click on an item preview, they can see the full details and review of the item. You need to design a search algorithm that:

- support to search keyword strings in similar forms. For example, if users enter the keyword strings "dress" or "dresses", the search results from these two keyword strings will be the same.
- Apart from the above requirement, you have the flexibility to use simple string matching-based methods, or any other language models, or build your own models to evaluate the relevance of a clothing item to the entered keyword string.

Functionality for Buyers and Employees

Creating labels for item recommendations

The critical aspect of the online shopping website is to recommend similar items to new customers which depends on the positive recommendation of items from previous customers. The website allows buyers (customer who have bought an item) to add a review about the item. When creating a new review, the buyer can enter information such as title, review description, rating, etc.

When a new review is created, using the review description (and/or review title), the website should generate a binary label whether the customer recommends the item or not ('0' or '1' for not recommended and recommended respectively). Though this happens behind the screen, for this assignment, the label/recommendation is generated by the model which is shown to the buyer/customer. The reviewer can choose a different response if the buyer does not find the response from the classification model suitable, i.e., they can override the value suggested by the website. Upon confirmation, the review should be included on the website and be accessible via URL.

You have the flexibility to choose the language model and the classification model for the task.

Note: You do not need to create any customer login for the shopping website.

Video Demonstration

The requested functionalities should be demonstrated through a video presentation. The video should not exceed 3.5 minutes and must cover the following three requirements:

1. Clothes browsing demonstration
2. Creating a new review
3. Displaying the newly created review

You may use any tool to record the video. If you are unable to find another suitable tool, Microsoft Stream is recommended for recording the presentation.

4. Marking Guidelines

Marking Criteria

The marking of this milestone will consider both the design, implementation and demonstration of the website. **Please refer to the rubric on the assignment page for mark allocation details**

5. Submission

The final submission of this milestone must include:

- The Python source code

- The .mp4 file for the video demonstration
- A README.txt file with your student number and any important details.
- Any other files/data that are required to locally run the website. Therefore, make sure all necessary files to run your code are included in your submission. If these files are greater than 50MB package them in a zip file and upload them to OneDrive. Provide OneDrive URL in the README.txt file. Remember to provide access to the OneDrive folder. If the team is unable to access the file then the submission will receive '0' mark for the assignment.
- Package these file files in a folder, named after your student ID, zip the folder with the same name (i.e., **<student_number>.zip**) and upload Canvas.
- All source code will be passed to a plagiarism checker to ensure academic integrity.

Assessment declaration

When you submit work electronically, you agree to the assessment declaration:
https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/assessment/assessment-declaration

Late Submission Penalty

Late submissions will incur a 10% penalty on the total marks of the corresponding assessment task per day or part of day late. Submissions that are late by 5 days or more are not accepted and will be awarded zero, unless special consideration has been granted. Granted Special Considerations with a new due date set more than 2 weeks after the original due will automatically result in an equivalent assessment in the form of a practical test with interview, assessing the same knowledge and skills of the assignment (location and time to be arranged by the instructor). Please ensure your submission is correct (all files are there, compiles, etc.), re-submissions after the due date and time will be considered as late submissions.

6. Academic integrity and plagiarism (standard warning)

Academic integrity is about the honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e. directly copied), summarized, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods,
- provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites.

If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate reference, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity