CS 534 Homework Assignment 5 (100 Points)
Due: 11:59 p.m. on 07/19/2022

Homework Objective: The goal of this assignment is to help you master the materials of Lecture 8 and 9.

Homework Deliverables: Submit your answers and solutions in a *zip* file that includes all the write-ups in a *pdf* file and your *python codes* to Canvas.

Notes:
1. Each answer needs to be explained.
2. The implementation should be in Python if You need.

Project Questions:

Question 1. After reviewing the lecture materials and videos from Week 8, please do the following:

(1) Describe and explain three major differences between Model-based Reinforcement Learning (MBRL) and Model-free Reinforcement Learning (MFRL). In EACH RL approach (i.e., MBRL and MFRL respectively), please provide and describe one real-world problem that you can find on the Internet or any resources respectively.

(2) Describe and explain three major differences between Passive Reinforcement Learning (PRL) and Active Reinforcement Learning (ARL). In EACH RL approach (i.e., PRL and ARL respectively), please provide and describe one real-world problem that you can find on the Internet or any resources respectively.

Note that all the above examples that you will provide should not be the same as the one that you have already posted in your weekly discussion forums. You need to provide the citation or the website for that real-world problem that you find. Specifically, in your answers, for each approach (MBRL, MFRL, PRL, and ARL, respectively), you need to include the contextual background of that problem, the main issues of the problem, the particular method in that approach that can solve the problem and explain how/why that method in that approach can be used to solve that problem.

Question 2. Given the following two simple documents:

Doc 1 = "The cat chased a rat."
Doc 2 = "A big rat chased the big dog."

Please generate and show the two document vectors (i.e., Doc 1 and Doc 2) by using each of the following methods respectively.

  i.   Bag of Words
 ii.   Bag of 2-grams
iii.   Bag of 3-grams
 iv.   TF-IDF

(a) Compute the cosine similarity for the above two documents using each of the above text vectorization methods respectively. Show all the necessary steps to demonstrate the computational process. **No points** are given if only the answer is provided.

(b) Based upon our lessons, use the two below documents and the NLTK/Spacy/Sklearn libraries of your interest and choice to create and submit a Python program to complete the following tasks.

Doc 1 = "Mr Jeremy put on a macintosh, and a pair of shiny shoes; he took his fishing rod and basket, and set off with enormous hops to the place where he kept his boat. The boat was round and green, and very like the other lily-leaves. It was tied to a water-plant in the middle of the pond."

Doc 2 = "Peter never stopped running or looked behind him till he got home to the big fir-tree. He was so tired that he flopped down upon the nice soft sand on the floor of the rabbit-hole and shut his eyes. His mother was busy cooking; she wondered what he had done with his clothes. It was the second little jacket and pair of shoes that Peter had lost in a week!"

i.   Pre-process the above two documents and then display the cleaned documents respectively on the output.
ii. Generate the two TF-IDF vectors respectively and then display the two TF-IDF vectors on the output.
iii. Compute their cosine similarity and display the value on the output.
iv. Discuss if these two documents are similar or not. Why and Why Not?

Question 3. The goal is to build a deep learning model that utilizes Neural Network Models, RNN and LSTM, to predict the appropriate sentiment label for each row of online shopping review data contained in the 'Amazon Fine Food Reviews' dataset obtained from Kaggle (https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews).

The data is split 70% train data and 30% using the Train-Test-Split function. The below table show the best performance that were obtained on the testing dataset.

TABLE 6: RNN model Performance

Target Class Value	Precision	Recall	F1-score	Support
0	0.70	0.65	0.67	25918
1	0.94	0.95	0.95	152536
Macro Avg	0.82	0.80	0.81	178454
Accuracy				0.91

TABLE 7: LSTM model Performance

Target Class Value	Precision	Recall	F1-score	Support
0	0.78	0.70	0.74	25918
1	0.95	0.97	0.96	152536
Macro Avg	0.87	0.84	0.85	178454
Accuracy				0.93

Now your task is to go through the NLP process that we discussed in our lectures & videos and then build your own RNN and LSTM models respectively to beat the above performance. Any step that you think will do a better job than the above is welcome. You need to include all the steps, the result tables, and the codes in your submission.

Grading Criteria: Your answers must be complete and clear.