MIS780 – Advanced Artificial Intelligence for Business

Trimester 2, 2023

Assessment 1 (Individual) – Data Analysis and Report

DUE DATE AND TIME: **Due** 7th August 2023 8:00PM AEST
PERCENTAGE OF FINAL GRADE: 35%
WORD LIMIT: Equivalent to 1500 words (1 Jupyter Notebook)
Word count is only an indication of the workload. Word limit is not applied for this assignment.

Learning Outcome Details

Unit Learning Outcome (ULO)	Graduate Learning Outcome (GLO)
ULO 1: Appraise the suitability of major artificial intelligence and advanced machine learning concepts to solve business problems	GLO1: Discipline-specific knowledge and capabilities
ULO 2: Design and develop artificial intelligence solutions for multifaceted business problems	GLO5: Problem solving

Assessment Feedback

Students who submit their work by the due date will receive their marks and feedback on CloudDeakin 3 weeks after the due date.

Extensions

No extensions will be considered unless a written request is submitted and negotiated with **the Unit Chair before the due date and time**. Extension request form must be filled via Cloud Deakin – Assessment – Extension Request (https://www.deakin.edu.au/students/studying/assessment-and-results/assignments), which is accompanied by appropriate documentary evidence for the extension. Submissions after the due date/time without an approved extension will be considered late.

Extensions are only granted in extreme circumstances, such as ongoing health, personal hardship or work-related problems. Temporary illnesses, normal work pressures, multiple assignments due at the same time, failure to keep backups, technology failure, etc are **not** reasons for an extension. Extension request **after** the assignment due date should be submitted visa **Student Connect** following Deakin procedures https://www.deakin.edu.au/students/studying/assessment-and-results/special-consideration.

Assignment Objectives

This assignment aims for students to learn how to analyse data relating to a business problem and propose artificial intelligence solutions based on machine learning and data mining techniques. The report will discuss and interpret the results. In particular, students will learn to:

- Articulate problems and solutions in business terms.

- Prepare data for different analytics tasks.
- Develop and justify sentiment analysis and topic models.
- Assess and report valuable insights to business.

Case Study Description

Social Analytics is a company specialized in gather and find meaning in data gathered from social channels to support business decisions. Social analytics is usually hired by marketers to track online conversations about products and companies. In the role of data analysts working at Social Analytics, you are appointed to analyse a large data set containing tweets posted by customers about Airlines in America. You are provided with a sample of more than 14k tweets (accessible via Cloud Deakin).

Your task is to use Python – Jupyter Notebook to process and explore the provided data. In particular, you are to generate some insights and provide answers to these questions of interest:

A. Rank the popularity of airlines based on the number of tweets posted for each airline.

B. Identify the most popular states where customers are located for each airline. (*Hint: use the* tweet *location for states information, e.g., CA, MA, TX. Tweets with missing location information can be ignored*).

C. Is it possible to predict tweet's sentiments using machine learning approach based on the provided raw data set? If yes, demonstrate the procedure and evaluate the performance with two different machine learning models.

D. Apply lexicon-based sentiment analysis to the tweets and then compare the proportions of positive and negative sentiments expressed by customers toward the top 3 airlines with most tweets. Among them, which airlines received more positive sentiments than negative sentiments from customers? (*Hint: determine the sentiments of the tweets generated by customers using lexicon-based sentiment analysis approach, compute the portions of positive vs. negative tweets, compare the proportions positive vs negative tweets between the top 3 companies.)*

E. What are the commonly mentioned issues that customers feel negative about the airlines? (*Hint: apply topic modelling to negative tweets of all airlines. Use Part-of-Speech tagging to extract only nouns for topic modelling. Remove highly frequent/infrequent words for meaningful topics discovery.*)

Task and Deliverables:

- **Executive Summary**: Define your problem in business term and present your proposed approaches. Present your major findings and explain how they help to address the business problem. Cross-reference with other report sections for support.
- **Data Exploration**: Process and explore the characteristics of the attributes the provided data set. Use table or figure to support answering questions (A) and (B).
- **Sentiment analysis**: Use machine learning based sentiment analysis to answer question (C). Use lexicon-based sentiment analysis to answer question (D).
- **Topic modelling**: Use text-processing techniques to process and prepare textual data for topic modelling. Use LDA to explore topics discussed in the text reviews. Carry out experiments and demonstrate how an appropriate topic number is determined for your model. Interpret the discovered topics and answer question (E).

- **Practical implication**: Based on the discovered insights from your analysis, provide some recommendations to the businesses on how to better support customers.

Submission Instructions

See CloudDeakin for more info about this assignment, especially the assignment template and the assessment rubric.

The assignment must be prepared using the provided assignment template (.ipynb file) using Jupyter Notebook. Your assignment should contain all necessary codes and ready to run. If you use any new python package, ensure that you include installation code in your .ipynb file. All python codes should be ready to execute without any further modification on Google Colab.

Upon completion of the assignment, execute all python codes and then generate a PDF file. Your files should be named as your firstname_lastname_MIS780A1 (e.g. John_Smith_MIS780A1.pdf and John_Smith_MIS780A1.ipynb).

You are to submit your assignment (both the PDF file and the source .ipynb file) in the individual Assignment Dropbox in the MIS780 CloudDeakin unit site on or before the due date. Do **NOT** zip the files. Any submission contained in a zip file will not be marked.

Notes

- You are allowed to use any sample code provided in the lab materials or online resources. However, you must modify/customize such sample code to your own assignment (e.g. rename variables, labels, titles; restructure code flow, modify chart types, colour and symbols.). References and citations must be provided where appropriate.
- Any work you submit may be checked by electronic or other means for the purposes of detecting collusion and/or plagiarism
- Feel free to discuss concepts and ideas with peers but remember your submission must be your own work. **Be careful not to allow others to copy your work**. Submissions, whose python codes are significantly similar (e.g. mostly identical except for only some variable names), are subjected to investigation for potential copying issue. The authors of such submissions may also be asked to present their work to an academic panel if necessary.
- You must keep a backup copy of every assignment you submit, until the marked assignment has been returned to you. In the unlikely event that one of your assignments is misplaced, you will need to submit your backup copy.
- When you are required to submit an assignment through your CloudDeakin unit site, you will receive an email to your Deakin email address confirming that it has been submitted. You should check that you can see your assignment in the Submissions view of the Assignment dropbox folder after upload, and check for, and keep, the email receipt of the submission. You are responsible for submitting the correct documents for the correct unit, in the required content or format. Should you wish to correct your submission, you can resubmit with any applicable penalties. You will not be able to submit, resubmit or correct your submission after the 5 day lateness period (or your extension deadline).

- ***Penalties for late submission:*** The following marking penalties will apply if you submit an assessment task after the due date without an approved extension: 5% will be deducted from available marks for each day up

to five days, and work that is submitted more than five days after the due date will not be marked. You will receive 0% for the task. 'Day' means working day for paper submissions and calendar day for electronic submissions. The Unit Chair may refuse to accept a late submission where it is unreasonable or impracticable to assess the task after the due date.

Acknowledgement:

This assignment was developed based on a data set and case study published on Kaggle.com.
https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment