



3032ICT / 1117ICT

Big Data Analytics and Social Media

**Assignment Milestone 2 Specifications**

## Instructions

- **Structure:** Milestone 2 continues your Assignment case study that you started in Milestone 1. Please refer back to Milestone 1 for the setting of the case study and advice on how to complete this Assignment. Milestone 2 also includes a video presentation (see details below).
- **Due:** See course site on Learning@Griffith for the due dates of each milestone.
- **Late Submissions:** An assessment item submitted after the due time on the due date set by the Course Convenor, without an approved extension, will be penalised. Assessment items submitted after the due time on the due date will be penalized at a rate of 5 percent (%) for each calendar day the assessment item is late. Assessment items submitted more than seven calendar days after the due date will be awarded zero marks.
- **Extensions:** If for any valid reason (e.g., being sick) you need an extension, you must apply for an extension by the due date of the milestone through this online form: <https://www.griffith.edu.au/students/assessment-exams-grades/assessment-applications>

## **Milestone 2**

## Data Selection & Exploration (continued)

2.1) Use the Spotify API to extract data about your artist/band.

For example:

- How many years have they been active?
- How many albums & songs have they published?
- With whom have they often collaborated?
- What are the prevalent features of their songs (e.g., valence)?

How does the Spotify data compare to the information you collected from other sources in Step 1.1 (Milestone 1)? (=> *Lab 2.2*)

[2 marks]

2.2) Retrieve data relevant to your artist/band from YouTube. Which videos have the highest number of views and likes? Do you see a correlation between views and likes? (Your dataset may contain hundreds of videos, so it's OK if you choose only a subset of those to get their statistics, in order to avoid hitting the rate-limit. However, you should get statistics for at least 5 videos.) (=> *Lab 3.2*)

[2 marks]

## Text Pre-Processing

2.3) Perform text pre-processing and create a Term-Document Matrix for your Twitter data.

What are the 10 terms occurring with the highest frequency? How are they different to your answer for Step 1.4 (Milestone 1)? (=> *Lab 2.2*)

[2 marks]

## Social Network Analysis

2.4) Perform centrality analysis by detecting degree centrality, betweenness centrality, and closeness centrality. Explain how relevant the results are to your artist/band. What are the actual degree, betweenness, and closeness centrality scores for your artist/band node in the network? Compare these scores to the scores for related artists. (=> *Lab 3.1*)

[4 marks]

- 2.5) Perform community analysis with the Girvan-Newman (edge betweenness) and Louvain methods. Explain how relevant the results are to your artist/band. Perform the community analysis also for related artists. Is their community structure similar?

(=> Lab 3.2)

[4 marks]

## Machine Learning Models

- 2.6) Use sentiment analysis to identify how the public reacts to events and/or topics related to your artist/band. Provide a summary of public opinions (emotions, reactions). (=> Lab 5.2)

[2 marks]

- 2.7) Build a decision tree and evaluate its performance in predicting whether a song is by your artist/band. (=> Lab 5.2)

[2.5 marks]

- 2.8) Use LDA topic modelling to identify some terms that are closely related to your artist/band. Find at least 3 significant groups of words that can be meaningful to your analysis. Explain your findings. (=> Lab 5.1)

[2 marks]

## Visualisation

- 2.9) Visualise your Twitter actor network in Gephi, with the node size determined by the number of followers for that actor. What insights can you extract from the visualisation?

*(This question is a little more difficult. Skip it if you're unsure and come back later. Hint: Look at the vosonSML documentation. No further hints will be provided for the question.)*

[2 marks]

- 2.10) Create at least three charts from your datasets using Tableau and combine them together into a dashboard. Describe each chart in your dashboard and **why you chose** to include it. Explain **the functionality of your dashboard** and **what insights you can obtain from it**.

[2.5 marks]

## **Video Presentation**

To complete your Milestone 2 submission, you will need to record a video presentation of minimum 3 minutes and **maximum 5 minutes** duration. You should use PowerPoint slides or similar to show the results from your report. **You will also need to record yourself while you are presenting and show your student ID at the beginning.** Your presentation will need to cover your work from both Milestones (i.e., Milestone 1 and Milestone 2). In the video, you should answer the following questions:

### **Evaluation**

- Briefly introduce yourself (show your student ID) and your artist/band. What data have you collected (search terms, search parameters, amount of data)?

[1 mark]

- What are the findings of your social media analytics?

[2 marks]

- How could you refine your social media analytics?

For example:

- Could you use different data sources?
- Could you choose different parameters?
- Can you think of ways to obtain more relevant data?

[2 marks]