

**Question 1:**

Explain why data cleaning is important in text analytics.

(3 marks)

**Question 2:**

Recent studies have been focusing on the increasing relevance of the social networks in the public image of politicians and political parties. Several of them have been focusing on the opinion on candidates to presidency and on how this opinion changes. You are planning to develop an application accessing a set of tweets containing the name of a specific politician and for each tweet determine if it is about the politician and if it is positive or negative. The application input will be the name of a politician, and the output will be the tweets referred to the politician, indicating for each of them, if the politician polarity is positive or negative.

a. Discuss about the application described in the scenario.

(2 marks)

b. Discuss any challenges that you might face to find the name of a specific politician.

(3 marks)

c. Tweets also present several difficulties when applying natural language tools. Describe any issues or problems (at least 3) related to processing of tweets using relevant examples.

(6 marks)

d. Discuss how to address the issues described in Question 2(c).

(6 marks)

----- THE END -----