<u>Solutions - Week 2 Assignment</u>

1. Social network APIs
   a. **Allow users to programmatically fetch data from a social network**
   b. Allow users to customise the look and feel of their homepage on a social network
   c. **Limit the number of requests a user can make in a given period of time**
   d. Limit the number of times a user can make a given request

2. Data from social network APIs is returned in which of the following formats?
   a. NDCG
   b. **JSON**
   c. HTTP
   d. **XML**

3. Which of the following statements are true about MongoDB and MySQL?
   a. **MySQL is a relational database management system**
   b. **MongoDB is a NoSQL database management system**
   c. phpMyAdmin is a Graphical User Interface for MongoDB
   d. **phpMyAdmin is a Graphical User Interface for MySQL**

4. If you were to create a new Facebook application today, which of the following versions of the Graph API would your application be able to access?
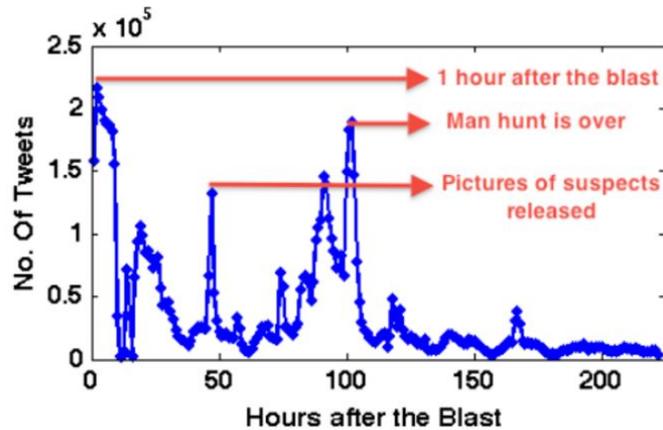   a. v2.7
   b. **v2.10**
   c. v2.6
   d. v2.9

5. In python, the **body** of a loop is defined by which of the following syntactic constructs?
   a. **indentation**
   b. curly braces
   c. square braces
   d. round braces

6. The objective of using k-fold cross validation during classification tasks is to
   a. Find the k most important features in the feature set
   b. Reduce the number of features by a factor of k
   c. Split the feature set into k classes to avoid overfitting
   d. **Gain confidence on the accuracy score obtained during classification**

7. Which of the following is true about the classification experiments performed on the Hurricane Sandy image dataset released by the Guardian?
   a. User-based features performed better than tweet-based features
   b. **Tweet-based features performed better than user-based features**

    c. Decision Trees performed better than Naive Bayes
    d. Naive Bayes performed better than Decision Trees

8. Look at the following graph which plots the number of tweets about the Boston blasts against the number of hours since the Boston blasts. Which of the following statements is true in the context of the graph?
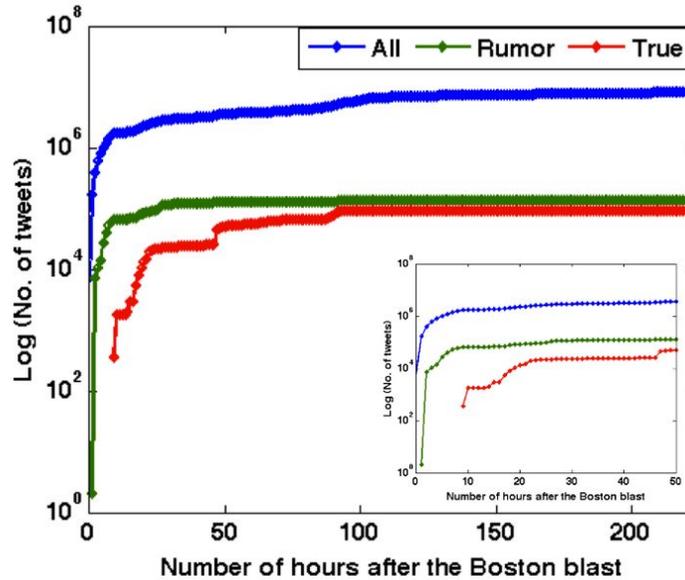

    a. More than 200,000 tweets were tweeted in the hour after the blast
    b. Less than 50,000 tweets were tweeted in the 100th hour after the blast
    c. Tweet volume is correlated with real world events
    d. Tweet volume is not correlated with real world events

9. Which of the following are popular crowdsourcing platforms for collecting ground truth data?
    a. Gephi
    b. Amazon Mechanical Turk
    c. Crowdflower
    d. Pittpatt
    e. All of the above

10. Which of the following values of Cronbach's Alpha indicate an acceptable level of inter-annotator agreement?
    a. 0.5
    b. 0.6
    c. 0.7
    d. 0.8

11. The advantage of requesting user feedback on browser extensions like TweetCred is that
    a. **Feedback can be used to improve the classification efficiency of the underlying model.**
    b. Feedback can be used to blacklist users who disagree with the model.
    c. Asking for feedback has been shown to positively impact extension usage.
    d. Feedback can be used to cherry pick users who agree with the model and thus increase its perceived relevance.

12. Which of the following do you need in order to successfully create an app and access token on Facebook?
    a. Valid profile picture
    b. **Valid phone number**
    c. Valid age proof
    d. Valid ID proof
    e. Valid home address

13. Which of the following rumor / rumors was / were spread during the Hurricane Sandy event?
    a. 8 year old boy killed
    b. **McDonalds in Virginia flooded**
    c. **Shark spotted on waterlogged highway**
    d. 1 dollar per retweet for helping victims

14. What is NDCG?
    a. Technique for collecting ground truth data for classification tasks
    b. Technique for calculating credibility of user generated content
    c. Metric for calculating the confidence in classification accuracy
    d. **Metric for calculating the quality of ranking**

15. Which of the following Facebook API would you use to retrieve data from Facebook?
    a. Facebook OAuth API
    b. Firehose API
    c. **Graph API**
    d. Streaming API

Look at the following graph, *G*, which shows the cumulative distribution of the number of tweets talking about the Boston blast with respect to the number of hours after the Boston blast.

16. After how many hours of the Boston blast did true information tweets appear on the

network?

    a. 100
    b. 50
    c. 1
    d. 9

17. After approximately how many hours of the Boston blast did the number of tweets spreading true information become equal to the rumour tweets?
    a. 50
    b. 95
    c. 100
    d. 9

18. Which of the following features was the most influential in determining the credibility of a tweet in the Boston blast dataset?
    a. The number of words in a tweet
    b. The number of retweets of a tweet

c. **The number of characters in a tweet**
d. The number of URLs in a tweet

The following graph shows the distribution of tweet sources for three different kinds of tweets collected during the Boston blasts. *NA* stands for tweets that are neither true nor fake.
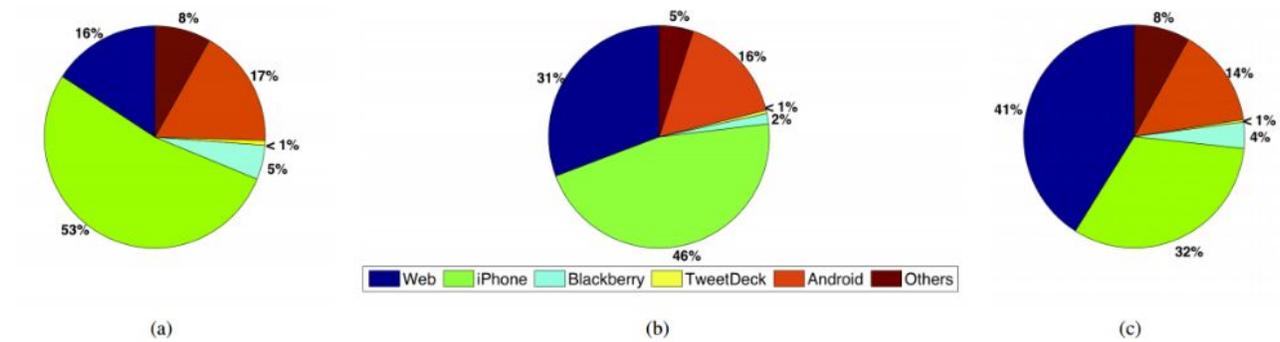

Fig. 7.   Tweet source analysis for the three categories of tweets: (a) Fake (b) True (c) NA.

19. What percentage of fake tweets were tweeted using mobile devices?
    a. 45
    b. 55
    c. 65
    d. 75

20. Most *NA* tweets were tweeted using which of the following platforms?
    a. Web
    b. iPhone
    c. TweetDeck
    d. Android