

Data Science Quiz

u19cs012@coed.svnit.ac.in [Switch account](#)

 Draft saved

Your email will be recorded when you submit this form

* Required

A Bloom filter guarantees no *

- ☒ false negatives
- ☐ false positives and false negatives
- ☐ false positives or false negatives, depending on the Bloom filter type
- ☐ false positives

In text mining, if an association between two concepts has 7% support, it means that 7% of the documents had both concepts represented in the same document. *

- ☒ True
- ☐ False



*

Ram has a Hadoop cluster with 20 machines under default setup (replication 3, 128MB input split size). Each machine has 500GB of HDFS disk space. The cluster is currently empty (no job, no data). Ram intends to upload 5 Terabyte of plain text (in 10 files of approximately 500GB each), followed by running Hadoop's standard WordCount1 job. What is going to happen?

- ☐ The data upload fails at a later stage: the disks are full
- ☐ WordCount fails: too many input splits to process.
- ☒ WordCount runs successfully.
- ☐ The data upload fails at the first file: it is too large to fit onto a DataNode

Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning. *

- ☒ True
- ☐ False

Which of the following techniques would perform better for reducing dimensions of a data set? *

- ☒ Removing columns which have too many missing values
- ☐ Removing columns which have high variance in data
- ☐ Removing columns with dissimilar data trends
- ☐ None of these



What are DGIM's maximum error boundaries? *

- ☐ DGIM always underestimates the true count; at most by 25%
- ☒ DGIM either underestimates or overestimates the true count; at most by 50%
- ☐ DGIM always overestimates the count; at most by 50%
- ☐ DGIM either underestimates or overestimates the true count; at most by 25%

What is the value of the Jaccard index when the two sets are disjoint? *

- ☐ 3
- ☐ 1
- ☒ 0
- ☐ 2

Given a stream of text, Named Entity Recognition determines which pronoun maps to which noun. *

- ☐ True
- ☒ False



*

Player Name	Age	Club	Weight	Foot
Craig Cathcart	27	Watford	172lbs	Right
Andre Gray	27	Watford	134lbs	Right
Kiko Femenía		Watford	209lbs	Right
Tom Cleverley	23	Watford	168lbs	Right
Nathaniel Chalobah	24	Watford	192kg	Left
Adam Masina	31	Watford	174lbs	Right

Identify the error(s) in the above dataset.

- ☐ Data is missing
- ☐ Typo error
- ☐ Unit of measurement is wrong
- ☒ All of the above

Imagine, you have 1000 input features and 1 target feature in a machine learning problem. You have to select 100 most important features based on the relationship between input features and the target features. Do you think, this is an example of dimensionality reduction?

- ☒ Yes
- ☐ No

Clear selection

Submit

Clear form

This form was created inside of Sardar Vallabhbhai National Institute of Technology, Surat. [Report Abuse](#)



Google Forms

