

CS381 Data Analytic Regression Practice Quizz

Question 1: Fill in the blank with either "higher" or "lower" or "bigger" or "smaller" in the following statements

1. The IDF of words in emails sent by the presidents of US should be _____ than the average IDF of words in emails sent by all US citizens.
2. Speech from President Trump's should have a _____ IDF than that from President Obama.
3. When considering only all historical speeches from all the past Presidents of United States, the average IDF from President Trump should be _____ than the average when stops words are not excluded.
4. When considering only all historical speeches from all the past Presidents of United States, the set of stops word from President Trump should be _____ than that of the average from all the other presidents

Question 2: Write down the results from the statements.

1. Results from running tokenizer of the sentence: Good morning Americans, have a great day!

Answer:

2. Results from running stemming from the words: jumping, jumped, jump

Answer:

3. Results after removing stopwords of the sentence: Cristiano Ronaldo was born on February 5, 1985.

Answer:

Question 3: The 8 actual values of a target variable is [0, 0, 1, 0, 1, 1, 1, 1]. What is the entropy of the target variable of this data set?

Answer: