**Quiz – Data Preprocessing II**

1. Manakah yang BUKAN cara tepat untuk mengatasi imbalanced data
2. Undersample data majority
3. Oversample data minority
4. SMOTE data minority
5. Oversample data majority
6. Apa yang dilakukan dalam proses tokenization
7. Menghilangkan karakter khusus
8. Mengubah kalimat menjadi kumpulan kata
9. Mengubah kata berimbuhan menjadi kata dasar
10. Menghapus kata-kata yag sering muncul
11. Manakah kumpulan kata di bawah ini yang termasuk ke dalam kategori Stopwords dalam Bahasa Indonesia
12. [“kambing”, “saya”, “dalam”, “rumah”]
13. [“rumah”, “cluster”, “milenial”, “idaman”]
14. [“baca”, “koran”, “di”, “kamar]
15. [“saya”, “dengan”, “kamu”, “ke”]

Informasi:

Diketahui Dokumen A berisi 100 kata dan kata “Trenggiling” muncul sebanyak 12 kali.  
Dari total keseluruhan dokumen sebanyak 1,000,000 dokumen, dokumen yang berisi kata “Trenggiling” muncul di 300,000 dokumen.

1. Berapakah nilai IDF untuk kata “Trenggiling”
2. 0.3
3. 0.12
4. 1.52
5. 0.52
6. Berapakah bobot TF-IDF kata “Trenggiling” pada Dokumen A
7. 0.06
8. 0.04
9. 0.46
10. 0.16
11. Berapakah nilai TF untuk kata “Trenggiling” pada Dokumen A
12. 0.3
13. 0.12
14. 1.52
15. 0.52
16. Algoritma yang digunakan pada metode SMOTE sebagai solusi data imbalaced adalah
17. Naïve Bayes
18. Decision Tree
19. K-Nearest Neighbor
20. Support Vector Machine
21. Steeming merupakan salah satu tahapan dalam proses text pre-processing. Steeming bertujuan untuk mentransofrmasikan kata menjadi kata dasarnya (root) dengan menghilangkan semua imbuhan kata (affixes) meliputi awalan kata (prefixes), sisipin kata (infixes), akhiran kata (suffixes) dan atau menghilangkan awalan dan akhiran kata (confixes) pada kata turunan
22. False
23. True
24. Beberapa hal yang dikerjakan pada tahap Case Folding, kecuali
25. Mengubah text menjadi lowercase
26. Menghapus angka
27. Menghapus tanda baca
28. Membuat potongan kata dari kalimat
29. Hasil dari proses stemming pada kalimat “rahmat memberikan pengajaran data science”
30. “rahmat beri ajar data science”
31. “rahmat beri pelajaran data science”
32. “rahmat berikan pelajaran data science”
33. “rahmat berikan pengajaran data science”