Topik -1 adalah yang terbesar dan merujuk pada tweet outlier yang tidak dikaitkan dengan topik apa pun yang dihasilkan. Dalam kasus ini, kami akan mengabaikan Topik -1

Pilih topik ke 6

[('orangorang', 0.040055753884755546),

('sesungguhnya', 0.03887516544841604),

('kafir', 0.026382378549590677),

('beriman', 0.022119899598126394),

('bertakwa', 0.018679661943296868),

('rasulnya', 0.01695860359995458),

('yang', 0.01597920851898845),

('ayatayat', 0.015376373978138805),

('allah', 0.014380471962524182),

('mukmin', 0.014178543398340652)]

Untuk topik diatas ada ‘'sesungguhnya'’,’ 'sesungguhnya'’ bisa dilihat ini lebih ke topik klasifikasi orang ada yang beriman,kafir dan mukmin

kesimpulan :

untuk bertopic pada datasheet terjemah alquran ini ada beberapa preprocesing yang harus lebih diperhatika kembali karena banyak topik yang dihasilkan pada topic count -1 itu sebanyak 21667, artinya ada outlier yang tidak dikaitkan dengan topik apa pun

intertopic distance map

ada bebearap topik Tampaknya jelas dan terpisah dari topik lainnya. dan ada juga topik-topik ini agak berdekatan, mengindikasikan adanya potensi tumpang tindih atau kesamaan dalam tema yang mendasarinya.

topic word scores

ini menampilkan istilah yang dipilih untuk beberapa topik dengan membuat diagrambatang dari scor c-Tf-IDF

masih banyak hal yang harus diperhatikan karena preprocesing data yang dilakukan sangat sederahana jadi masih perlu bimbingannya

terimakasih