

Algoritma :

1. Pemanggilan teks atau korpus Al-Qur'an
2. Lalu masuk tahap preprocessing
 - Tokenisasi
Ternyata disaat dilakukan token, disini bisa langsung melakukan part of speech tagging
 - Stemming
3. Setelah itu cari TF-IDF
4. Udah itu masuk kedalam klasifikasi ternyata klasifikasi ada seleksi fitur, ada 3 jenis fitur :
 - Fitur kata
Jadi kata terakhir yang mengacu pada tanda ('.', '?', '!') adalah penting
 - Fitur POS
Jadi memilih kata yang ambiguitas, yang kapan fitur POS dimungkinkan
 - Fitur Affix & Orthografis
Fitur ini digunakan untuk menandakan kata-kata yang tidak dikenal.

Set pola fitur kaya yang digunakan dalam percobaan dan simbolnya

Fitur kata	$W_{-3}, W_{-2}, W_{-1}, W_0, W_{+1}, W_{+2}, W_{+3}$
Fitur POS	$P_{-3}, P_{-2}, P_{-1}, P_0, P_{+1}, P_{+2}, P_{+3}$
Awalan (Prefiks)	$S_1, S_1S_2, S_1S_2S_3, S_1S_2S_3S_4$
Akhiran (Sufiks)	$S_n, S_{n-1} S_n, S_{n-2} S_{n-1} S_n, S_{n-3} S_{n-2} S_{n-1} S_n$
Fitur kata biner	Semua huruf besar, semua huruf kecil dan berisi angka
Panjang kata	Integer

5. Setelah itu kita cari masing-masing dari klasifikasi
 - KNN
Algoritma
 1. Menentukan parameter k (jumlah tetangga paling dekat).
 2. Menghitung kuadrat jarak eucliden objek terhadap data training yang diberikan
 3. Mengurutkan hasil no 2 secara ascending (berurutan dari nilai tinggi ke rendah)
 4. Mengumpulkan kategori Y (klasifikasi nearest neighbor berdasarkan nilai k)
 5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksikan kategori objek.
 - NAIVE BAYES
Algoritma
 1. Menghitung jumlah kelas / tabel
 2. Menghitung jumlah kasus per kasus
 3. Kalikan semua variabel kelas
 4. Bandingkan hasil per kelas

6. lalu masuk ke confusion matriks untuk menghitung akurasi terhadap data tersebut
7. setelah dapat nilai ACC ke masing-masing metode klasifikasi
8. lanjut ke penggabungan metode klasifikasi ke dua metode klasifikasi menggunakan algoritma voting mayoritas (strategi kombinasi) dengan satu set fitur setelah itu hitung kembali nilai ACCnya.