Algoritma :

1. Pemanggilan teks atau korpus Al-Qur’an
2. Lalu masuk tahap preprocssing

* Tokenisasi

Ternyata disaat dilakukan token, disini bisa langsung melakukan part of speech tagging

* Stemming

1. Setelah itu cari TF-IDF
2. 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 dogunakan 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, W0,W+1,W+2,W+3 |
| Fitur POS | P-3,P-2,P-1,P0,P+1,P+2,P+3 |
| Awalan (Prefiks) | S1,S1S2, S1S2S3, S1S2S3S4 |
| Akhiran (Sufiks) | Sn,Sn-1 Sn, Sn-2 Sn-1 Sn, Sn-3 Sn-2 Sn-1 Sn |
| Fitur kata biner | Semua huruf besar, semua huruf kecil dan berisi angka |
| Panjang kata | Integer |

1. Setelah itu kita cari masing-masing dari klasifikasi

* KNN

Algoritma

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak eucliden objek terhadapndata 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 dipredisikan kategori objek.

* NAIVE BAYES

Algoritma

1. Menghitung jumlah kelas / tabel
2. Menghitung jumlah kasus perkasus
3. Kalikan semua variabel kelas
4. Bandingkan hasil perkelas

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.