1. Pendahuluan

Latar Belakang

Indonesia adalah negara dengan populasi terbesar ke empat dan menjadi negara terbesar untuk populasi muslim dunia. Sebagaimana seharusnya, Al-Qur'an menjadi pedoman untuk hidup para muslimin. Al-Qur'an diturunkan kepada nabi Muhammad SAW, Al-Qur'an sudah ditulis 1400 tahun yang lalu dan hingga sekarang umat islam mempelajari isi dari Al-Qur'an, bahkan dimulai dari kecil [7]. Al-Qur'an memiliki 114 Surat, 6000 lebih ayat, dan 128219 kata, merujuk ke data *Buckwalter*. Dikarenakan Al-Qur'an menggunakan bahasa Arab, yang dimana bahasa Arab terbilang rumit[6]. Bagi orang yang tidak mengerti atau tidak mempelajari bahasa Arab terasa susah untuk mempelajari Al-Qur'an. Bagi muslimin mempelajari Al-Qur'an adalah sebuah keharusan, tetapi banyak dari orang tidak beragama muslim pun ikut untuk mempelajari Al-Qur'an[11]. Dikarenakan Al-Qur'an memiliki banyak kata, alangkah lebih baik jika proses pembelajaran Al-Qur'an dibantu dengan memanfaatkan *Natural La- ngue Procesing* (NLP).

Klasifikasi teks Arab dapat menjadi tugas yang menantang karena sifat bahasa Arab yang kaya dan kompleks[3]. Salah penelitian yang berjudul *Do Words with Certain Part of Speech Tags Improve the Performance of Arabic Te- xt Classification?* bahwa klasifikasi tulisan Arab menghasilkan nilai akurasi sebesar 90% dengan menggunakan *single word orthography* atau satu kata ortografi dalam bahasa arab yang sudah ada *Part-Of-Speech Tagging*[2].

Pada penelitian ini, penulis akan menggunakan metode klasifikasi *Naive Bayes* dan *Random Forest* dengan menggabungkan *Term Frequency Inverse Document Frequency* sebagai pembuatan *part-of-speech tagging*. Sistem yang dibangun diharapkan dapat menghasilkan nilai akurasi yang baik.

Topik dan Batasannya

Pada penelitian kali ini akan dilakukan klasifikasi kata yang terdapat di Al-Qur'an dengan menggunakan *Naive Bayes* dan *Random Forest* yang dimana kata tersebut merupakan bagian dari ayat-ayat Al-Qur'an yang sudah di potong menjadi kata.

Klasifikasi menggunakan satu kata ortografi, kata unik yang digunakan atau *Number of feature* yang digunak- an hanya 5000, *Term Frequency Inverse Document Frequency* berguna untuk membatu proses klasifikasi. Semua data dalam *dataset* akan digunakan penelitian ini hanya dilakukan sampai menghasilkan nilai akurasi dan akan melakukan Evaluasi.

Tujuan

Tujuan dari penelitian ini adalah untuk mencari tahu seberapa akurat klasifikasi *Part-Of-Speech Tagging* meng- gunakan metode *Naive Bayes* dan *Random Forest* dengan *Term Frequency Inverse Document Frequency* dapat di- terapkan pada bahasa Arab khususnya dalam Al-Qur'an.

Organisasi Tulisan

Selanjutnya akan dijelaskan apa itu , *Part-Of-Speech Tagging*, *Term Frequency Inverse Document Frequency Naive Bayes*,*Random Forest* , *flowchart* yang menunjukkan sistem yang dibangun serta data apa saja yang akan dipakai, lalu diakhiri dengan evaluasi serta analisis dari hasil penerapan sistem yang telah dibangun.