

LAPORAN
PEMROSESAN BAHASA ALAMI

TUGAS 2: MEMBANGUN POSTAGGER BAHASA INDONESIA



Oleh:

Aditya Alif Nugraha 1301154183

PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS TELKOM
BANDUNG
2018

POS Tag

POS Tag adalah suatu teknik yang digunakan untuk mengidentifikasi kelas kata yang ada dalam suatu kalimat. POS Tag bisa digunakan untuk Information Extraction, Question-Answering, Speech Recognition, Parser, dan lainnya untuk tahap pemrosesan awal.

Metode

Ada 3 metode yang digunakan untuk membangun model POS Tagger pada tugas ini, yaitu:

1. Baseline
2. Neural Network
3. HMM-Viterbi

Baseline

Baseline adalah metode yang paling simpel dan mudah diimplementasikan untuk membangun model POS Tagger. Baseline hanya memberi tag pada kata dengan mengambil tag yang paling umum untuk kata tersebut. Dan jika ada kata baru yang tagnya belum dihitung sebelumnya, kata tersebut akan diberi tag sebagai noun.

Kelebihan dari metode baseline yaitu mudah diimplementasikan, akurasi cukup baik untuk klasifikasi per kata, dan lainnya. Sedangkan kekurangan dari metode baseline yaitu akurasi buruk untuk klasifikasi kata pada kalimat, dan lainnya.

Neural Network

Neural Network atau yang dalam bahasa Indonesia disebut jaringan syaraf tiruan merupakan jaringan dari unit pemrosesan kecil yang saling terhubung antar layer/lapisannya.

Kelebihan dari metode Neural Network adalah akurasi cenderung lebih baik, cocok untuk dataset dengan ukuran besar, dan banyak lagi. Sedangkan kekurangan dari Neural Network yaitu training time yang cenderung lama.

Viterbi

Viterbi adalah salah satu metode dalam HMM dimana berbasis sequence. Metode ini memperhatikan keterkaitan antar kata sebelumnya. Sehingga metode ini bagus dan dapat membedakan kata-kata yang memiliki ambiguitas tag.

Viterbi membangun tabel emission dan transisi untuk menghitung probabilitas kemunculan tag selanjutnya given kata yang diberikan. Kemudian mengambil probabilitas tertinggi untuk mengambil kelas kata tersebut.

Kelebihan dari metode ini adalah cocok untuk memprediksi kalimat karena dapat memahami konteks dalam keseluruhan kalimat. Kekurangan dari metode ini adalah running time yang lama apalagi bila kalimat memiliki kata yang panjang.

Hasil Pengujian

Metode Baseline

Baseline Method Accuracy: 85.44117647058823 %
saya menyukai makan nasi
['PRP', 'VB', 'VB', 'NN']

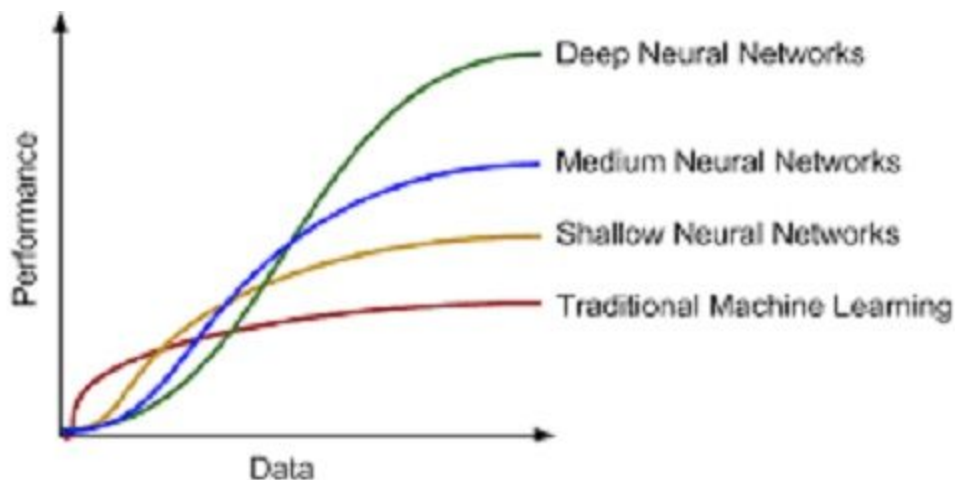
Metode Neural Network

Neural Network Accuracy:
0.9264705882352942
saya menyukai makan nasi
['PRP' 'VB' 'VB' 'NN']

Metode HMM-Viterbi

Viterbi Algorithm Accuracy:
63.382352941176464
saya menyukai makan nasi
['PRP', 'VB', 'NN', 'NN']

Dari hasil akurasi diatas, didapatkan bahwa metode Neural Network memiliki akurasi yang terbaik. Hal ini dikarenakan neural network adalah metode yang baik untuk mengklasifikasi data dengan jumlah besar.



Dari gambar diatas dapat dilihat neural network dengan jumlah layer yang banyak dan data yang banyak dapat menghasilkan performansi yang sangat baik. Sedangkan untuk algoritma machine learning yang tradisional, menghasilkan performansi yang kurang baik jika dihadapkan dengan data yang besar.