

LAPORAN
PEMROSESAN BAHASA ALAMI
TUGAS I: LANGUAGE MODELING



Oleh:
Aditya Alif Nugraha 1301154183

PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS TELKOM
BANDUNG
SEPTEMBER 2018

Keterangan Artikel

Sumber Artikel : Tempo.co

Topik : Nasional

Jumlah Artikel : 420

Teknik Pengambilan : *Scraping* menggunakan kode sumber terbuka

Alasan:

- Tempo merupakan portal berita yang terkenal sehingga isi dari berita tersebut terjamin fakta
- Artikel berita dipilih karena memiliki struktur kalimat yang baik
- Topik nasional dipilih karena topik tersebut bersifat umum sehingga memiliki kumpulan kata-kata yang lebih kaya

Analisis Hasil Pengujian

Prediksi Kemunculan Kata

Kata	Alasan	Keluaran
sssss	Mencoba kata awal yang akan dihasilkan	ketua
joko	Ingin mengetahui joko siapa yang akan keluar	widodo
presiden	Ingin mengetahui apakah model dapat mengetahui presiden Republik Indonesia saat ini berdasarkan artikel	joko
saya	Mencoba kata bersambung	kira
akan		memberikan
melakukan		gerakan
hahaha	Mencoba kata tidak relevan	bersyukur
wkwkwk		bersyukur
i	Mencoba bahasa lain selain Indonesia	nomor
gracias		bersyukur
hola		bersyukur

Dari hasil diatas, dapat dilihat bahwa kata-kata yang berada di berita akan menghasilkan keluaran yang cocok. Dikarenakan probabilitas benar-benar dihitung sesuai dengan corpusnya.

Sedangkan untuk kata-kata yang tidak ada di berita akan menghasilkan kata keluaran yang sama. Hal tersebut karena model akan mengambil kata yang sama karena hasil perhitungan probabilitas yang sama.

Perplexity

Kalimat	Nilai Perplexity	Analisa
presiden joko widodo akan meresmikan bendungan baru	0.026702420009071373	Kalimat tersebut cocok dengan berita. Dan juga struktur kalimat tersebut cukup baik.
saya suka makan nasi merah dan minum susu biru	0.0007817649932282991	Kalimat tersebut tidak relevan dengan artikel yang digunakan untuk membangun model
tim badminton indonesia berhasil meraih medali emas terbanyak pada asian games 2018	0.026642774431505686	Kalimat tersebut memiliki struktur kalimat yang cukup baik.
rupiah melemah menjadi diatas 15 ribu	0.0017905284420307258	Kalimat tersebut singkat dan tidak baku
saya bingung kalimat apalagi yang akan digunakan untuk evaluasi perplexity	0.0075782856747257914	Kalimat tersebut tidak relevan tetapi memiliki struktur yang cukup baik.