

LAPORAN TUGAS PEKAN 5 NLP : POSTagging

CLARISA HASYA YUTIKA | 1301174256 | IF 41 GAB01

1. Deskripsi Masalah

Membuat POSTagger sederhana berbasis 3 metode :

- Metode baseline
- Metode klasifikasi konvensional non-sekuensial (Algoritma Decision Tree)
- Metode HMM-Viterbi

Terdapat 50 kalimat untuk training, dan 10 kalimat untuk testing.

2. Perancangan Sistem & Analisis

• Read File TSV

```
df = pd.read_csv("train.01.tsv", sep="\t", header=None)
df = df.astype(str)
```

• Kalimat Uji

```
['menteri pertahanan as dijadwalkan mengunjungi india .',
'tata power menyuplai batu bara pada tahun 2000 .',
'pemerintah hati-hati dalam mengelola bumh .',
'perusahaan baru tersebut mencanangkan target perolehan laba bersih .',
'menteri pertahanan mengunjungi pangkalan udara .',
'menurut laporan sekretaris perusahaan , laba bersih meningkat .',
'transaksi penjualan barang mewah tahun 2007 turun .',
'transaksi memperkirakan inflasi akan meningkat dibanding tahun lalu .',
'kenaikan tarif didorong oleh target laba bersih yang meningkat .',
'makanan dari luar negeri tidak bisa masuk pasar lokal .']
```

• Hasil Tagging pada metode Baseline

```
'NNP', 'NN', 'NNP', 'VB', 'VB', 'NNP', 'Z'
'NNP', 'NNP', 'VB', 'NN', 'IN', 'NN', 'CD', 'Z'
'NN', 'JJ', 'IN', 'VB', 'NN', 'Z'
'NN', 'JJ', 'PR', 'VB', 'NN', 'NN', 'NN', 'Z'
'NNP', 'NN', 'VB', 'NN', 'NN', 'Z'
'IN', 'NN', 'NNP', 'NN', 'Z', 'NN', 'VB', 'Z'
'NN', 'NN', 'NN', 'NN', 'CD', 'VB', 'Z'
'NNP', 'VB', 'NN', 'MD', 'VB', 'VB', 'NN', 'CC', 'Z'
'NN', 'NN', 'VB', 'IN', 'NN', 'NN', 'SC', 'VB', 'Z'
'NN', 'IN', 'NN', 'NEG', 'MD', 'VB', 'NN', 'JJ', 'Z'
```

• Hasil Tagging pada metode Klasifikasi

```
'IN', 'NN', 'NNP', 'VB', 'VB', 'NNP', 'Z'
'CD', 'NNP', 'VB', 'NN', 'IN', 'NN', 'CD', 'Z'
'NN', 'JJ', 'IN', 'VB', 'NN', 'Z'
'NN', 'JJ', 'PR', 'VB', 'NN', 'NN', 'NN', 'Z'
'IN', 'NN', 'VB', 'NN', 'NN', 'Z'
'IN', 'NN', 'NNP', 'NNP', 'Z', 'NN', 'VB', 'Z'
'NN', 'NN', 'NN', 'NN', 'CD', 'NNP', 'Z'
'IN', 'VB', 'NN', 'MD', 'VB', 'VB', 'NN', 'CC', 'Z'
'NN', 'NN', 'VB', 'IN', 'NN', 'NN', 'SC', 'VB', 'Z'
'NN', 'IN', 'NN', 'NEG', 'MD', 'VB', 'NN', 'JJ', 'Z'
```

• Hasil Tagging pada metode HMM-Viterbi

```
'NNP', 'NNP', 'NNP', 'VB', 'VB', 'NNP', 'Z'
'NNP', 'NNP', 'VB', 'NN', '<start>', 'NN', 'CD', 'Z'
'NN', 'JJ', 'NN', 'VB', 'NN', 'Z'
'NNP', 'JJ', 'PR', 'VB', 'NN', 'NN', 'NN', 'Z'
'NNP', 'NNP', 'VB', 'NN', 'NN', 'Z'
'<start>', 'NN', 'NNP', 'NNP', '<start>', 'NN', 'VB', 'Z'
'NN', 'NN', 'NN', 'NN', 'CD', 'VB', 'Z'
'NNP', 'VB', 'NN', '<start>', 'VB', 'VB', 'NN', 'CC', 'Z'
'NN', 'NN', 'VB', '<start>', 'NN', 'NN', '<start>', 'VB', 'Z'
'NN', '<start>', 'NN', 'NEG', '<start>', 'VB', 'NN', 'JJ', 'Z'
```

• Akurasi pada metode Baseline

Accuracy Baseline : 0.948051948051948

• Akurasi pada metode Klasifikasi

Accuracy Classification: 0.8831168831168831

• Akurasi pada metode HMM-Viterbi

Accuracy HMM Viterbi : 0.8311688311688312

3. Kesimpulan

Kesimpulan yang didapat dari 3 metode yang dijalankan, akurasi tertinggi didapatkan dari metode baseline, dan yang paling rendah didapatkan dari metode HMM-Viterbi. Hal ini dapat terjadi karena pada metode baseline, kata-kata dari kalimat uji akan dicari dari corpus, jika terdapat kata yang sama, maka tag kata tersebut di assign tag yang paling sering muncul dari kata tersebut, jika tidak ada kata yang sama, maka tag kata tersebut akan di assign tag yang paling sering muncul dari corpus.

Sedangkan pada metode klasifikasi dengan algoritma decision tree mendapat akurasi tertinggi kedua setelah baseline. Hal ini dapat terjadi dalam proses pemodelan terdapat aturan yang ambigu, sehingga terdapat beberapa kesalahan.

Kemudian pada metode HMM-Viterbi mendapatkan akurasi paling rendah

LAPORAN TUGAS PEKAN 5 NLP : POSTagging

CLARISA HASYA YUTIKA | 1301174256 | IF 41 GAB01

dibandingkan 2 metode lainnya. Hal ini dapat terjadi pada saat proses decoding Viterbi dan backtracking, terdapat nilai 0, sehingga tag yang didapat adalah 'start'.