

**Nama** : Muhammad Azhar Rasyad  
**NIM** : 0110217029  
**Program Studi** : Teknik Informatika

### Ujian Akhir Semester Natural Language Processing

1. (15 poin) Hitung jarak Levenshtein dari kata Belanda dan Holland

**Jawaban :**

A	7							
D	6							
N	5							
A	4							
L	3							
E	2							
B	1							
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

Tahap 1 Inisialisasi (Levenshtein Distance = 2)

A	7							
D	6							
N	5							
A	4							
L	3							
E	2							
B	1	2	3	4	5	6	7	8
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

A	7							
D	6							
N	5							
A	4							
L	3	4	5	4	5	6	7	8
E	2	3	4	5	6	7	8	9
B	1	2	3	4	5	6	7	8
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

A	7							
D	6							
N	5	6	7	6	7	6	5	6
A	4	5	6	5	6	5	6	7
L	3	4	5	4	5	6	7	8
E	2	3	4	5	6	7	8	9
B	1	2	3	4	5	6	7	8
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

A	7	8	9	8	9	8	7	6
D	6	7	8	7	8	7	6	5
N	5	6	7	6	7	6	5	6
A	4	5	6	5	6	5	6	7
L	3	4	5	4	5	6	7	8
E	2	3	4	5	6	7	8	9
B	1	2	3	4	5	6	7	8
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

### Tahap 2 Mengisi Isi Baris

A	7	8	9	8	9	8	7	6
D	6	7	8	7	8	7	6	5
N	5	6	7	6	7	6	5	6
A	4	5	6	5	6	5	6	7
L	3	4	5	4	5	6	7	8
E	2	3	4	5	6	7	8	9
B	1	2	3	4	5	6	7	8
#	0	1	2	3	4	5	6	7
	#	H	O	L	L	A	N	D

### Tahap 3 Jarak Belanda dan Holland adalah 6

2. (15 poin) Apa perbedaan open class dengan close class? Sebutkan 5 contoh kata dalam percakapan bahasa Indonesia yang termasuk open class yang 10 tahun lalu belum ada!

**Jawaban :** Open Class merupakan sekumpulan kata-kata baru yang dibuat seiring berjalannya waktu atau kata lama yang mengalami perubahan yang mudah diubah dari adanya kata benda, kata sifat, dan kata kerja. Sedangkan Close Class merupakan sekumpulan kata-kata yang tidak dapat berubah karena mengandung fungsi gramatikal bahasa tertentu, contohnya dari kata penghubung dan kata depan seperti kata dan, atau, tapi, ke, di, dan sebagainya.

Berikut 5 contoh kata dalam percakapan bahasa Indonesia yang termasuk Open Class dalam 10 tahun lalu yaitu mantul, santuy, kuy, mager, dan baper.

3. (15 poin) Buat sebuah kalimat lengkap (antara 7 hingga 15 kata) yang bercerita tentang salah satu teknologi yang terkait NLP. Beri tagging sesuai dengan tagset dari INACL.

**Jawaban :** Contoh kalimat : Kecerdasan buatan merupakan teknologi yang belakangan ini sedang berkembang pesat di Indonesia. Berikut tagset dari INACL dari kalimat di atas :

Kecerdasan\_NNO buatan\_NNO merupakan\_VBL teknologi\_NNO yang\_PRR belakangan\_NNO ini\_ART sedang\_ADK berkembang\_VBI pesat\_ADJ di\_PPO Indonesia\_NNP .\_SYM

4. (30 poin) Diberikan sebuah model HMM berikut :

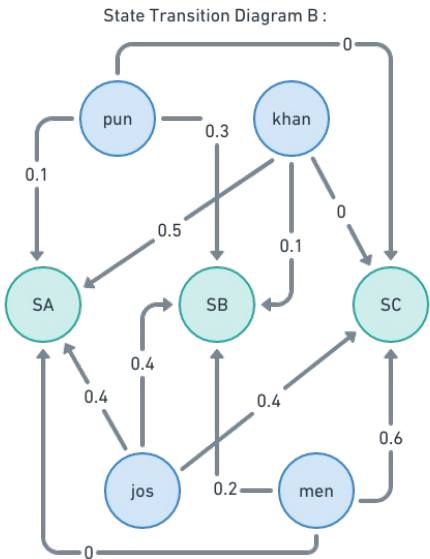
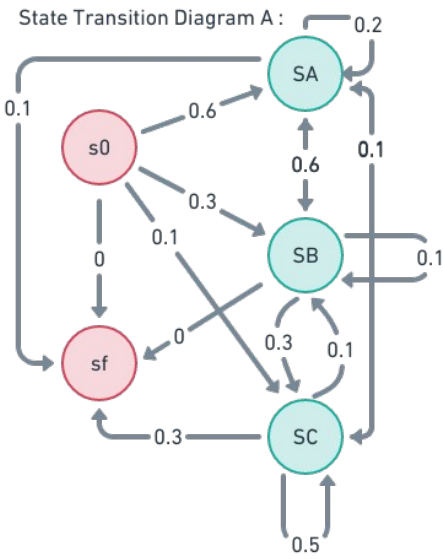
A :

	SA	SB	SC	sf
s0	0.6	0.3	0.1	0
SA	0.2	0.6	0.1	0.1
SB	0.6	0.1	0.3	0
SC	0.1	0.1	0.5	0.3

B :

	SA	SB	SC
pun	0.1	0.3	0
khan	0.5	0.1	0
jos	0.4	0.4	0.4
men	0	0.2	0.6

(a) gambarkan diagram transisi statenya.



(b) Hitung berapa peluang kemunculan kata “khan jos pun men” menggunakan forward algorithm.

Jawaban :

	SA	SB	SC
khan	0.5	0.1	0
jos	0.4	0.4	0.4
pun	0.1	0.3	0
men	0.1	0.2	0.6

Output Probabilities

	SA	SB	SC
s0	0.6	0.3	0.1

Input State Probabilities

Initialization Output khan		
$\alpha_1(1) =$	$0.6 * 0.5 =$	0,3
$\alpha_1(2) =$	$0.3 * 0.1 =$	0,03
$\alpha_1(3) =$	$0.1 * 0 =$	0

	SA	SB	SC
khan	0,3	0,03	0
jos			
pun			
men			

Recursion Output jos		
$\alpha_2(1) =$	$[(0,3 * 0,2) + (0,03 * 0,6) + (0 * 0,1)] * 0,4 =$	0,0312
$\alpha_2(2) =$	$[(0,3 * 0,6) + (0,03 * 0,1) + (0 * 0,1)] * 0,4 =$	0,0732
$\alpha_2(3) =$	$[(0,3 * 0,1) + (0,03 * 0,3) + (0 * 0,5)] * 0,4 =$	0,0156

	SA	SB	SC
khan	0,3	0,03	0
jos	0,0312	0,0732	0,0156
pun			
men			

Recursion Output pun		
$\alpha_3(1) =$	$[(0,0312 * 0,2) + (0,0732 * 0,6) + (0,0156 * 0,1)] * 0,1 =$	0,005172
$\alpha_3(2) =$	$[(0,0312 * 0,6) + (0,0732 * 0,1) + (0,0156 * 0,1)] * 0,3 =$	0,00828
$\alpha_3(3) =$	$[(0,0312 * 0,1) + (0,0732 * 0,3) + (0,0156 * 0,5)] * 0 =$	0

	SA	SB	SC
khan	0,3	0,03	0
jos	0,0312	0,0732	0,0156
pun	0,005172	0,00828	0
men			

Recursion Output men		
$\alpha_4(1) =$	$[(0,005172 * 0,2) + (0,00828 * 0,6) + (0 * 0,1)] * 0 =$	0
$\alpha_4(2) =$	$[(0,005172 * 0,6) + (0,00828 * 0,1) + (0 * 0,1)] * 0,2 =$	0,00078624
$\alpha_4(3) =$	$[(0,005172 * 0,1) + (0,00828 * 0,3) + (0 * 0,5)] * 0,6 =$	0,00180072

	SA	SB	SC
khan	0,3	0,03	0
jos	0,0312	0,0732	0,0156
pun	0,005172	0,00828	0
men	0	0,00078624	0,00180072

Berikut Peluang Kemunculan kata “khan jos pun men”

	SA	SB	SC	$\Sigma$
khan	0,3	0,03	0	0,33
jos	0,0312	0,0732	0,0156	0,12
pun	0,005172	0,00828	0	0,013452
men	0	0,00078624	0,00180072	0,00258696

(c) (bonus) Apa state yang paling mungkin untuk kata tersebut (POS Tag dari kalimat tersebut)? Bisa dicari menggunakan viterbi algorithm.

Jawaban :

	SA	SB	SC
khan	0.5	0.1	0
jos	0.4	0.4	0.4
pun	0.1	0.3	0
men	0.1	0.2	0.6

Output Probabilities

	SA	SB	SC
s0	0.6	0.3	0.1

Input State Probabilities

Initialization Output khan			
$v1(1) =$	$0,6 * 0,5 =$	0,3	
$v1(2) =$	$0,3 * 0,1 =$	0,03	
$v1(3) =$	$0,1 * 0 =$	0	

	SA	SB	SC
khan	0.3	0.03	0
jos			
pun			
men			

Recursion Output jos			
$v2(1) =$	$\text{Max}((0,3 * 0,2); (0,03 * 0,6); (0 * 0,1)) * 0,4 =$	0,024	
$v2(2) =$	$\text{Max}((0,3 * 0,6); (0,03 * 0,1); (0 * 0,1)) * 0,4 =$	0,072	
$v2(3) =$	$\text{Max}((0,3 * 0,1); (0,03 * 0,3); (0 * 0,5)) * 0,4 =$	0,012	

	SA	SB	SC
khan	0,3	0,03	0
jos	0,024	0,072	0,012
pun			
men			

Recursion Output pun			
$v3(1) =$	$\text{Max}((0,024 * 0,2); (0,072 * 0,6); (0,012 * 0,1)) * 0,1 =$	0,00432	
$v3(2) =$	$\text{Max}((0,024 * 0,6); (0,072 * 0,1); (0,012 * 0,1)) * 0,3 =$	0,00432	
$v3(3) =$	$\text{Max}((0,024 * 0,1); (0,072 * 0,3); (0,012 * 0,5)) * 0 =$	0	

	SA	SB	SC
khan	0,3	0,03	0
jos	0,024	0,072	0,012
pun	0,00432	0,00432	0
men			

Recursion Output men			
$v4(1) =$	$\text{Max}((0,00432 * 0,2); (0,00432 * 0,6); (0 * 0,1)) * 0 =$	0	
$v4(2) =$	$\text{Max}((0,00432 * 0,6); (0,00432 * 0,1); (0 * 0,1)) * 0,2 =$	0,0005184	
$v4(3) =$	$\text{Max}((0,00432 * 0,1); (0,00432 * 0,3); (0 * 0,5)) * 0,6 =$	0,0007776	

	SA	SB	SC
khan	0,3	0,03	0
jos	0,024	0,072	0,012
pun	0,00432	0,00432	0
men	0	0,0005184	0,0007776

Berikut Peluang Kemunculan kata “khan jos pun men” dengan viterbi algorithm

	SA	SB	SC	Max
khan	0,3	0,03	0	0,3
jos	0,024	0,072	0,012	0,072
pun	0,00432	0,00432	0	0,00432
men	0	0,0005184	0,0007776	0,0007776

State yang paling mungkin untuk kata tersebut (POS Tag dari kalimat tersebut) berdasarkan viterbi algorithm adalah

SA – SB – SA – SC atau SA – SB – SB – SC

5. (15 poin) Berikan penjelasan tentang istilah berikut :

- (a) **Perceptron** merupakan algoritma dasar untuk neural network sederhana yang meniru cara kerja syaraf manusia dengan bekerja dari impuls-impuls yang diberikan oleh neuron disebelahnya, dan berfungsi untuk mengklasifikasikan suatu tipe pola tertentu dengan pemisahan secara linear.
- (b) **Overfitting** merupakan keadaan ketika suatu data training berbeda jauh dengan data test sehingga menimbulkan bias yang sangat jauh ataupun sekumpulan data yang memiliki kriteria tertentu namun ada beberapa data yang berbeda jauh dari kriteria tersebut.
- (c) **Feature Engineering** merupakan cara agar mesin dapat mengenali suatu model dari ciri model tersebut sehingga akurasi dalam mengenali suatu model tersebut dapat meningkat.

6. (10 poin) Jelaskan tentang MFCC dari apa yang kamu ketahui!

**Jawaban :** MFCC atau Mel Frequency Cepstral Coefficient merupakan metode dalam pengenalan audio untuk melakukan feature extraction sehingga mendapatkan suatu parameter dan informasi mengenai ciri dari audio tersebut.

7. (bonus poin) Apa perbedaan antara LSTM dan GRU?

**Jawaban :** LSTM atau Long Short Term Memory merupakan arsitektur yang mampu memproses data secara berurut karena dapat mengingat urutan yang lebih panjang dengan adanya 3 gate yaitu forget gate, input gate, dan output gate, sedangkan GRU atau Gated Recurrent Unit merupakan gabungan dari forget gate dan input gate sehingga menjadi update gate yang menjadikan komputasinya lebih sederhana dari LSTM dan melatih dengan lebih cepat, namun keduanya masih memiliki kelebihan dan kekurangannya masing-masing.