

LAMPIRAN

Lampiran 1. *Preprocessed dataset*

Label	<i>Tweet</i>	N chars	N words	N slang	N english	N repeat	N abbrev	Split
-1	lagu bosan apa yang aku save ni huhuhuhuhuhuuuuu	65	8	0	0	1	0	<i>train</i>
-1	kita lanjutkan saja diam ini hingga kau dan aku mengerti tidak semua kebersamaan harus melibatkan hati	102	16	0	0	0	0	<i>test</i>
1	doa rezeki tak putus inna haa zaa larizquna maa lahu min na fadesungguhnya ini ialah pemberian kami kepada kamu p	113	20	0	0	0	0	<i>train</i>
1	makasih loh ntar kita bagi hasil aku 99 9 sisanya buat kamu	59	12	0	0	0	0	<i>train</i>
0	barusan liat tulisan di belakang truk rela injek kopling demi kamu bisa shopping	80	13	1	0	0	0	<i>train</i>
-1	cerita dia macam ni mamat ni hari2 aku nampak keluar masuk cimb ceria ja tapi hari ni aku tengok muka dia masam aku p	117	24	0	0	0	0	<i>train</i>
1	ya aku akan menjadi satu satunya bukan nomor satu tetapi satu satunya	69	12	0	0	0	0	<i>train</i>
-1	aku aja capek sama diriku sendiri apalagi kamu maaf ya '	56	11	1	0	0	0	<i>test</i>
-1	aku cuba nak amalkan sikap pergi mampus tapi sayang aku gagal sebab aku jenis	110	19	0	0	0	0	val

	mudah kesian dan mudah memaafkan							
0	pernah pas masih ngajar privat gt adek kakak muridnya adek aku weekend ini mau ke spore miss pernah ke s	104	20	0	0	0	1	<i>train</i>
.....								
-1	kalau kita tahu cinta tu sakit mesti takramai orang pilih untuk bercinta um	75	13	0	0	0	0	<i>train</i>
0	jadi kita harus mkn dulu di kfc sebesar 60 ribu dpt struk promo trs klo mau ditukarkan hrs beli super	101	20	1	0	0	5	<i>test</i>
0	tante kapan nikah akun rp panggil calon istri kenalin ke tante masukin cincin ke jarinya ini tant	97	17	0	0	0	1	<i>test</i>
-1	pelik tengok orang zaman sekarang ni dulu kemain beria tunjuk effort dekat kita last ii dia kongsi perasaan tu kat org lain jgk	127	23	1	0	0	1	<i>train</i>
0	menikah itu belum tentu harus sudah pernah pacaran jika dia sudah mapan dan juga serius sama kamu kena	102	18	0	0	0	0	<i>test</i>
1	jangan membandingkan kehidupanmu dengan kehidupan orang lain karena setiap dari kita	115	15	0	0	0	0	<i>test</i>

	memiliki karunia masing masing							
0	sini uname lu ntar gua follow	29	6	2	1	0	0	<i>test</i>
1	apapun yg telah kamu lakukan apapun kesalahanmu kamu akan selalu menemukan kata maaf dalam hati seorang ibu botkece	115	18	1	0	0	1	<i>train</i>
1	3 cara untuk ingat semula apa yang kita dah hafal	49	10	1	0	0	0	<i>train</i>
1	semoga allah permudahkan urusan aku amp orang sekeliling aku yang tengah stress tak kisah pasal assignment ke kerja lain ke	123	20	0	0	0	0	<i>train</i>



Lampiran 2. Contoh Hasil *Token* 34 Kata Uji

Kategori	Kata	mBERT <i>tokens</i>	#mBERT	IndoBERT <i>tokens</i>	#IndoBERT
Tidak baku/ <i>slang</i>	gue	gu ##e	2	gue	1
Tidak baku/ <i>slang</i>	lo	lo	1	lo	1
Tidak baku/ <i>slang</i>	gak	ga ##k	2	gak	1
Tidak baku/ <i>slang</i>	udah	ud ##ah	2	udah	1
Tidak baku/ <i>slang</i>	bgt	bg ##t	2	bgt	1
Tidak baku/ <i>slang</i>	wkwk wk	wk ##w ##k ##w ##k	5	wkwkwk	1
Tidak baku/ <i>slang</i>	mantul	man ##tul	2	mant ##ul	2
Tidak baku/ <i>slang</i>	baper	ba ##per	2	baper	1
Tidak baku/ <i>slang</i>	gabut	gab ##ut	2	gab ##ut	2
Tidak baku/ <i>slang</i>	anjir	an ##jir	2	anj ##ir	2
Singkatan	yg	yg	1	yg	1
Singkatan	dgn	d ##gn	2	dgn	1
Singkatan	krn	kr ##n	2	krm	1
Singkatan	tdk	td ##k	2	tdk	1
Singkatan	sdh	sd ##h	2	sdh	1
Singkatan	utk	ut ##k	2	utk	1
Singkatan	klo	kl ##o	2	klo	1
Singkatan	jgn	jg ##n	2	jgn	1
Singkatan	pdhl	pd ##hl	2	pd ##hl	2
Singkatan	bgmn	bg ##mn	2	bgm ##n	2
<i>Code-mixing</i> (EN)	love	love	1	love	1
<i>Code-mixing</i> (EN)	happy	happy	1	happy	1
<i>Code-mixing</i> (EN)	aweso me	aw ##eso ##me	3	aw ##eso ##me	3

<i>Code-mixing</i> (EN)	really	really	1	real ##ly	2
<i>Code-mixing</i> (EN)	please	please	1	please	1
<i>Code-mixing</i> (EN)	thanks	thanks	1	thanks	1
<i>Code-mixing</i> (EN)	followers	followers	1	followers	1
<i>Code-mixing</i> (EN)	online	online	1	online	1
Karakter berulang	hahaha ha	ha ##hah ##aha	3	hahahaha	1
Karakter berulang	huhuh uhu	hu ##hu ##hu ##hu	4	huh ##uhu ##hu	3
Karakter berulang	bangett	bang ##ett ##t	3	bangett ##tt	2
Karakter berulang	iyaaaa	i ##ya ##aa ##a	4	iyaa ##aa	2
Karakter berulang	okeeee	ok ##ee ##ee	3	oke ##eee	2
Karakter berulang	suka a	suka ##aa ##a	3	suka ##aaa	2

Lampiran 3. Tokenisasi per-tweet

No	Tweet	Label name	N words	mBERT subwords	IndoBERT subwords	mBERT subword per word	IndoBERT subword per word
1	lagu bosan apa yang aku save ni huhuhuhuhuhuuuuu	Negatif	8	25	20	3,125	2,5
2	kita lanjutkan saja diam ini hingga kau dan aku mengerti tidak semua kebersamaan harus melibatkan hati	Negatif	16	23	16	1,4375	1
3	doa rezeki tak putus inna haa zaa larizquna maa lahu min na fadesungguhnya ini ialah pemberian kami kepada kamu p	Positif	20	34	33	1,7	1,65
4	makasih loh ntar kita bagi hasil aku 99 9 sisanya buat kamu	Positif	10	15	10	1,5	1
5	barusan liat tulisan di belakang truk rela injek kopling demi kamu bisa shopping	Netral	13	19	14	1,4615	1,0769
6	cerita dia macam ni mamat ni hari2 aku nampak keluar masuk cimb ceria ja tapi hari ni aku tengok muka dia masam aku p	Negatif	24	31	25	1,2917	1,0417
7	ya aku akan menjadi satu satunya bukan nomor satu tetapi satu satunya	Positif	12	12	12	1	1
8	aku aja capek sama diriku sendiri apalagi kamu maaf ya '	Negatif	10	15	10	1,5	1
9	aku cuba nak amalkan sikap pergi mampus tapi sayang aku gagal sebab aku jenis mudah kesian dan mudah memaafkan	Negatif	19	27	22	1,4211	1,1579

10	pernah pas masih ngajar privat gt adek kakak muridnya adek aku weekend ini mau ke spore miss pernah ke s	Netral	20	26	22	1,3	1,1
.....							
10797	kalau kita tahu cinta tu sakit mesti takramai orang pilih untuk bercinta um	Negatif	13	20	15	1,5385	1,1538
10798	jadi kita harus mkn dulu di kfc sebesar 60 ribu dpt struk promo trs klo mau ditukarkan hrs beli super	Netral	19	30	21	1,5789	1,1053
10799	tante kapan nikah akun rp panggil calon istri kenalin ke tante masukin cincin ke jarinya ini tant	Netral	17	25	18	1,4706	1,0588
10800	pelik tengok orang zaman sekarang ni dulu kemain beria tunjuk effort dekat kita last ii dia kongsi perasaan tu kat org lain jgk	Negatif	23	32	28	1,3913	1,2174
10801	menikah itu belum tentu harus sudah pernah pacaran jika dia sudah mapan dan juga serius sama kamu kena	Netral	18	23	18	1,2778	1
10802	jangan membandingkan kehidupanmu dengan kehidupan orang lain karena setiap dari kita memiliki karunia masing masing	Positif	15	22	16	1,4667	1,0667
10803	sini uname lu ntar gua follow	Netral	6	8	7	1,3333	1,1667
10804	apapun yg telah kamu lakukan apapun	Positif	18	26	21	1,4444	1,1667

	kesalahanmu kamu akan selalu menemukan kata maaf dalam hati seorang ibu botkece						
10805	3 cara untuk ingat semula apa yang kita dah hafal	Positif	9	12	9	1,3333	1
10806	semoga allah permudahkan urusan aku amp orang sekeliling aku yang tengah stress tak kisah pasal assignment ke kerja lain ke	Positif	20	26	23	1,3	1,15



Lampiran 4. Hasil Ablasi mBERT, IndoBERT, dan *TF-IDF SVM*

<i>Model</i>	<i>Class weight</i>	bs	lr	<i>Epoch</i>	<i>Accuracy</i>
mBERT	Ya	16	3,00E-05	2	0,63
mBERT	Tidak	16	3,00E-05	4	0,6438
IndoBERT	Ya	32	3,00E-05	4	0,6984
IndoBERT	Tidak	32	3,00E-05	1	0,7105
<i>TF-IDF SVM</i>	Ya (balanced)	N/A	N/A	N/A	0,63
<i>TF-IDF SVM</i>	Tidak	N/A	N/A	N/A	0,6198

<i>Model</i>	<i>Class weight</i>	<i>Precision (macro)</i>	<i>Recall (macro)</i>	<i>F1 (macro)</i>	<i>Precision (weighted)</i>	<i>Recall (weighted)</i>	<i>F1 (weighted)</i>
mBERT	Ya	0,6125	0,6235	0,6143	0,6359	0,63	0,63
mBERT	Tidak	0,6265	0,6382	0,6308	0,6481	0,6438	0,6447
IndoBERT	Ya	0,6852	0,6983	0,6907	0,7022	0,6984	0,6993
IndoBERT	Tidak	0,6973	0,6954	0,6957	0,7102	0,7105	0,7098
<i>TF-IDF SVM</i>	Ya (balanced)	0,611	0,5993	0,6043	0,6262	0,63	0,6271
<i>TF-IDF SVM</i>	Tidak	0,6288	0,545	0,5601	0,6202	0,5513	0,597

Lampiran 5. Log Pelatihan mBERT dan IndoBERT

<i>Model</i>	<i>Case</i>	<i>Epoch</i>	<i>Loss</i>	<i>Acc</i>	<i>Precision Macro</i>	<i>Recall Macro</i>	<i>F1 Macro</i>	<i>Precision Weighted</i>	<i>Recall Weighted</i>	<i>F1 Weighted</i>
mBERT	CW	1	1,0287	0,4728	0,5462	0,5389	0,4735	0,6046	0,4728	0,4525
mBERT	CW	2	0,8037	0,6636	0,6428	0,6457	0,6424	0,6655	0,6636	0,6632
mBERT	CW	3	0,5611	0,6462	0,6332	0,6399	0,6316	0,6584	0,6462	0,6483
mBERT	CW	4	0,3759	0,6428	0,63	0,6487	0,6342	0,6577	0,6428	0,6453
mBERT	noCW	1	0,9701	0,6416	0,6444	0,5837	0,5979	0,6438	0,6416	0,6284
mBERT	noCW	2	0,7628	0,6543	0,6475	0,6025	0,6161	0,6514	0,6543	0,6444
mBERT	noCW	3	0,5326	0,6509	0,6328	0,6279	0,6278	0,6499	0,6509	0,6483
mBERT	noCW	4	0,3563	0,6543	0,6358	0,6416	0,6384	0,6569	0,6543	0,6554
IndoBERT	CW	1	0,8471	0,7087	0,6996	0,7048	0,6974	0,7195	0,7087	0,7103
IndoBERT	CW	2	0,5379	0,7098	0,6943	0,709	0,7001	0,7164	0,7098	0,7115
IndoBERT	CW	3	0,2937	0,7017	0,6909	0,71	0,6968	0,7116	0,7017	0,7028
IndoBERT	CW	4	0,1587	0,7145	0,7011	0,7139	0,7066	0,7183	0,7145	0,7154
IndoBERT	noCW	1	0,818	0,7168	0,704	0,703	0,7033	0,7165	0,7168	0,7165
IndoBERT	noCW	2	0,5236	0,7191	0,7058	0,6981	0,7012	0,7181	0,7191	0,7179
IndoBERT	noCW	3	0,2904	0,711	0,6985	0,6915	0,6946	0,7097	0,711	0,71
IndoBERT	noCW	4	0,1487	0,7064	0,6927	0,6902	0,6909	0,7066	0,7064	0,706

