

LAMPIRAN

Lampiran 1. Hasil Tokenisasi Lengkap 830 Kata Unik Leksikon Bahasa Bali pada Ketiga Model

Lampiran ini menyajikan hasil tokenisasi lengkap untuk seluruh 830 kata unik leksikon bahasa Bali pada ketiga model. Tanda ‘|’ memisahkan antar sub-token, prefiks ‘##’ menandai kelanjutan sub-kata pada WordPiece (IndoBERT dan mBERT), dan prefiks ‘_’ menandai batas awal kata pada SentencePiece (XLM-RoBERTa). Kolom Jumlah Token menampilkan jumlah token masing-masing model dengan urutan IndoBERT / mBERT / XLM-RoBERTa.

Tabel 1. Hasil Tokenisasi 830 Kata Unik Leksikon Bahasa Bali pada Ketiga Model

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (I/m/X)
1	abot	ab ##ot	abo ##t	_a bot	2/2/2
2	abra	ab ##ra	ab ##ra	_a bra	2/2/2
3	abu	abu	abu	_abu	1/1/1
4	acep-acepan	ace ##p - ace ##pan	ace ##p - ace ##pan	_a cep - ace pan	5/5/5
5	acepan	ace ##pan	ace ##pan	_a ce pan	2/2/3
6	adah	ada ##h	ada ##h	_a dah	2/2/2
7	adeng-adeng	aden ##g - aden ##g	aden ##g - aden ##g	_a deng - a deng	5/5/5
8	adep	ade ##p	ad ##ep	_ad ep	2/2/2
9	adopsi	adopsi	ado ##psi	_ado psi	1/2/3
10	adu	adu	ad ##u	_a du	1/2/2
11	aduh	aduh	ad ##uh	_a duh	1/2/2
12	adung	adu ##ng	ad ##ung	_a dung	2/2/2
13	ae	ae	ae	_a e	1/1/2
14	aeng sajan	ae ##ng saja ##n	ae ##ng saja ##n	_a eng _saja n	4/4/4
15	agem-ageman	age ##m - age ##man	age ##m - age ##man	_ agem - agem an	5/5/5
16	ageng	ageng	agen ##g	_a geng	1/2/2
17	ajan	aja ##n	ajan	_ajan	2/1/1
18	ajerih	aje ##rih	aj ##eri ##h	_a jer ih	2/3/3
19	ajin	aji ##n	aj ##in	_aj in	2/2/2
20	ajumang	aj ##uman ##g	aj ##uman ##g	_a ju mang	3/3/3
21	aktip	ak ##tip	akt ##ip	_akti p	2/2/2
22	alep	al ##ep	ale ##p	_a lep	2/2/2
23	alesan	al ##esan	ales ##an	_ale san	2/2/2
24	alih	alih	alih	_ali h	1/1/2
25	alihpati	alih ##pati	alih ##pati	_ali h pati	2/2/3
26	alo-alon	al ##o - al ##on	al ##o - al ##on	_a lo - al on	5/5/5
27	alon-alon	al ##on - al ##on	al ##on - al ##on	_al on - al on	5/5/5
28	aluh	al ##uh	al ##uh	_a luh	2/2/2
29	alus	alus	alus	_alus	1/1/1
30	aman	aman	aman	_aman	1/1/1
31	ambet	ambe ##t	amb ##et	_amb et	2/2/2
32	ambil	ambil	amb ##il	_ambil	1/2/1
33	ampak	amp ##ak	amp ##ak	_ampak	2/2/1
34	ampakang	amp ##akan ##g	amp ##akan ##g	_ampak ang	3/3/2
35	ampura	amp ##ura	amp ##ura	_am pura	2/2/2
36	anak medagang	anak med ##agang	anak med ##agan ##g	_anak _med a gang	3/4/4
37	ancem	anc ##em	anc ##em	_an cem	2/2/2
38	ancema	anc ##ema	anc ##ema	_ ance ma	2/2/3

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
39	anceman	anc ##eman	anc ##eman	_ ance man	2/2/3
40	andeg	and ##eg	and ##eg	_and eg	2/2/2
41	ane mela	ane mel ##a	ane mela	_a ne _mela	3/2/3
42	ane nembak	ane nem ##bak	ane nem ##bak	_a ne _ne mbak	3/3/4
43	ane nukung	ane nuk ##ung	ane nuk ##ung	_a ne _nuk ung	3/3/4
44	ane wanen	ane wan ##en	ane wan ##en	_a ne _wan en	3/3/4
45	angen	ang ##en	ange ##n	_angen	2/2/1
46	antem	ant ##em	ante ##m	_ante m	2/2/2
47	anti	anti	anti	_anti	1/1/1
48	antiang	anti ##ang	anti ##ang	_anti ang	2/2/2
49	antisipasi	antisipasi	anti ##si ##pasi	_an tisipasi	1/3/2
50	antos	anto ##s	ant ##os	_an tos	2/2/2
51	antuk	ant ##uk	ant ##uk	_an tuk	2/2/2
52	anuraga	anu ##raga	anura ##ga	_anu raga	2/2/2
53	aon	ao ##n	aon	_aon	2/1/1
54	apa	apa	apa	_apa	1/1/1
55	api	api	api	_api	1/1/1
56	apresiasi	apresiasi	apres ##ias ##i	_apre si asi	1/3/3
57	apui	ap ##ui	ap ##ui	_ap ui	2/2/2
58	arta	art ##a	arta	_ar ta	2/1/2
59	artine	arti ##ne	arti ##ne	_arti ne	2/2/2
60	asih	asih	asi ##h	_as ih	1/2/2
61	asil	asi ##l	asi ##l	_as il	2/2/2
62	ati	ati	ati	_ati	1/1/1
63	aturang	aturan ##g	at ##uran ##g	_aturan g	2/3/2
64	awak	awak	aw ##ak	_awak	1/2/1
65	awake	awak ##e	aw ##ake	_awak e	2/2/2
66	awinan	aw ##inan	aw ##inan	_a win an	2/2/3
67	awu	aw ##u	aw ##u	_a wu	2/2/2
68	ayu	ayu	ay ##u	_ay u	1/2/2
69	baan	ba ##an	baan	_baan	2/1/1
70	baang	ba ##ang	baan ##g	_ba ang	2/2/2
71	baat	ba ##at	ba ##at	_baat	2/2/1
72	bagia	bagi ##a	bagi ##a	_bagi a	2/2/2
73	bakuh	baku ##h	baku ##h	_bak uh	2/2/2
74	balik	balik	balik	_balik	1/1/1
75	ban	ban	ban	_ban	1/1/1
76	bangga	bangga	bang ##ga	_bangga	1/2/1
77	banggaang	bangga ##ang	bang ##gaan ##g	_bangga ang	2/3/2
78	bangke	bangk ##e	bang ##ke	_bang ke	2/2/2
79	bangkit	bangkit	bang ##kit	_bangkit	1/2/1
80	bani	bani	bani	_bani	1/1/1
81	banyol	bany ##ol	ban ##yol	_banyo l	2/2/2
82	bareng-bareng	bareng - bareng	bare ##ng - bare ##ng	_bare ng - bare ng	3/5/5
83	bas	bas	bas	_bas	1/1/1
84	bati	bat ##i	bati	_bati	2/1/1
85	baud	bau ##d	bau ##d	_baud	2/2/1
86	baya	baya	bay ##a	_baya	1/2/1
87	bayu	bayu	bay ##u	_bay u	1/2/2
88	bea	bea	be ##a	_be a	1/2/2
89	bebanyolan	beban ##yol ##an	be ##ban ##yol ##an	_beban yo lan	3/4/3
90	bebas	bebas	bebas	_bebas	1/1/1
91	becat	bec ##at	beca ##t	_be cat	2/2/2
92	becik	bec ##ik	bec ##ik	_be cik	2/2/2
93	becikan	bec ##ikan	bec ##ikan	_be ci kan	2/2/3
94	begal	beg ##al	beg ##al	_be gal	2/2/2
95	beh	beh	be ##h	_beh	1/2/1
96	bek	bek	bek	_bek	1/1/1
97	belah	belah	bela ##h	_ belah	1/2/2
98	belat	bela ##t	bela ##t	_be lat	2/2/2

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
99	beler	bel ##er	bele ##r	_be ler	2/2/2
100	belog	bel ##og	belo ##g	_be log	2/2/2
101	bencana	bencana	ben ##cana	_bencana	1/2/1
102	bench	ben ##ch	bene ##h	_bene h	2/2/2
103	benges	beng ##es	ben ##ges	_ben ges	2/2/2
104	benjep	ben ##je ##p	ben ##je ##p	_be nje p	3/3/3
105	benyah	beny ##ah	ben ##yah	_ben yah	2/2/2
106	biakta	biak ##ta	bia ##kta	_bi ak ta	2/2/3
107	biji	biji	bij ##i	_bij i	1/2/2
108	bincuh	bin ##cu ##h	bin ##cu ##h	_bin cu h	3/3/3
109	bingung	bingung	bing ##ung	_bingung	1/2/1
110	biseka	bis ##eka	bis ##eka	_bis eka	2/2/2
111	bocok	boc ##ok	boc ##ok	_bo cok	2/2/2
112	bodo	bod ##o	bodo	_bodo	2/1/1
113	bogbog	bog ##bo ##g	bog ##bo ##g	_bog bog	3/3/2
114	bom	bom	bom	_bom	1/1/1
115	bonggan	bong ##gan	bong ##gan	_bon ggan	2/2/2
116	brana	bra ##na	bra ##na	_bra na	2/2/2
117	buduh	bud ##uh	budu ##h	_budu h	2/2/2
118	bunga	bunga	bunga	_bunga	1/1/1
119	bungkah	bungk ##ah	bung ##kah	_ bung kah	2/2/3
120	bungkem	bungk ##em	bung ##kem	_ bung kem	2/2/3
121	buta	buta	buta	_but a	1/1/2
122	buung	bu ##ung	bu ##ung	_bu ung	2/2/2
123	buungang	bu ##ungan ##g	bu ##ungan ##g	_bu un gang	3/3/3
124	cacarang	cacar ##ang	caca ##rang	_ca ca rang	2/2/3
125	cacep	cac ##ep	cac ##ep	_ca cep	2/2/2
126	cager	ca ##ger	cage ##r	_ca ger	2/2/2
127	cang	can ##g	cang	_can g	2/1/2
128	cegah	cegah	cega ##h	_ce gah	1/2/2
129	ceteng	cet ##eng	cet ##eng	_ce teng	2/2/2
130	cetik	cet ##ik	cet ##ik	_ce tik	2/2/2
131	cidera	cidera	cid ##era	_ci dera	1/2/2
132	cocok	cocok	coco ##k	_cocok	1/2/1
133	corah	cor ##ah	cora ##h	_co rah	2/2/2
134	dadalan	dada ##lan	dada ##lan	_dada lan	2/2/2
135	dadian	dadi ##an	dadi ##an	_da dian	2/2/2
136	dagang	dagang	dag ##ang	_da gang	1/2/2
137	dakep	dak ##ep	da ##kep	_da kep	2/2/2
138	dakepa	dak ##ep ##a	da ##kep ##a	_dake pa	3/3/2
139	daki	dak ##i	daki	_da ki	2/1/2
140	dalih	dalih	dali ##h	_da lih	1/2/2
141	daliha	dalih ##a	dali ##ha	_da liha	2/2/2
142	dame	dam ##e	dame	_dame	2/1/1
143	dawa	daw ##a	da ##wa	_dawa	2/2/1
144	daya	daya	daya	_daya	1/1/1
145	demen	dem ##en	dem ##en	_de men	2/2/2
146	diman	diman	dima ##n	_di man	1/2/2
147	doktrin	doktrin	dok ##tri ##n	_doktrin	1/3/1
148	dot	dot	dot	_dot	1/1/1
149	dueg	du ##eg	due ##g	_du eg	2/2/2
150	duka	duka	du ##ka	_du ka	1/2/2
151	dukung	dukung	du ##kung	_du kung	1/2/2
152	dukungan	dukungan	du ##kung ##an	_dukungan	1/3/1
153	duleg	dul ##eg	dul ##eg	_du leg	2/2/2
154	dum	dum	dum	_dum	1/1/1
155	dumogi	dum ##ogi	dum ##ogi	_dum ogi	2/2/2
156	duratmaka	dur ##atm ##aka	dura ##tma ##ka	_dura tma ka	3/3/3
157	durgama	dur ##gam ##a	dur ##gama	_dur gama	3/2/2
158	duri	duri	dur ##i	_du ri	1/2/2
159	durusang	dur ##usan ##g	dur ##usan ##g	_du rus ang	3/3/3

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
160	efektip	efek ##tip	efekt ##ip	_efekt ip	2/2/2
161	ejuk	ej ##uk	ej ##uk	_e juk	2/2/2
162	ejuka	ej ##uka	ej ##uka	_e juk a	2/2/3
163	ekslusip	eks ##lus ##ip	eks ##lusi ##p	_eks lusi p	3/3/3
164	eling	el ##ing	eli ##ng	_el ing	2/2/2
165	empati	empati	empat ##i	_empati	1/2/1
166	enggal	engga ##l	eng ##gal	_e nggal	2/2/2
167	engkab	engk ##ab	eng ##kab	_eng kab	2/2/2
168	engsap	eng ##sa ##p	eng ##sa ##p	_eng sap	3/3/2
169	enjep	enj ##ep	en ##je ##p	_en je p	2/3/3
170	entel	ente ##l	ente ##l	_en tel	2/2/2
171	enye	en ##ye ##h	en ##ye ##h	_en ye h	3/3/3
172	gabeng	gab ##eng	gaben ##g	_gabe ng	2/2/2
173	gagah	gagah	gaga ##h	_gaga h	1/2/2
174	gagal	gagal	gagal	_gagal	1/1/1
175	galah	gala ##h	gala ##h	_ga lah	2/2/2
176	ganggas	gangg ##as	gang ##gas	_gang gas	2/2/2
177	ganggu	ganggu	gang ##gu	_ ganggu	1/2/2
178	gati	gat ##i	gat ##i	_gati	2/2/1
179	gaya	gaya	gaya	_gaya	1/1/1
180	gebug	geb ##ug	geb ##ug	_ge bug	2/2/2
181	gede	gede	ge ##de	_gede	1/2/1
182	gedeg	gede ##g	ge ##deg	_gede g	2/2/2
183	gegawan	geg ##awan	ge ##gawa ##n	_geg awan	2/3/2
184	geguritan	geg ##guri ##tan	ge ##gur ##itan	_ge guri tan	3/3/3
185	gelem	gel ##em	gele ##m	_gele m	2/2/2
186	geleng	gel ##eng	gelen ##g	_gelen g	2/2/2
187	gelis	geli ##s	gel ##is	_geli s	2/2/2
188	gelur	gel ##ur	gel ##ur	_gel ur	2/2/2
189	geluran	gel ##uran	gel ##uran	_gel uran	2/2/2
190	gemes	gem ##es	gem ##es	_ge mes	2/2/2
191	genep	gen ##ep	gene ##p	_gene p	2/2/2
192	geni	gen ##i	geni	_geni	2/1/1
193	gerak	gerak	gera ##k	_gerak	1/2/1
194	gertak	ger ##tak	gert ##ak	_ger tak	2/2/2
195	getap	get ##ap	geta ##p	_get ap	2/2/2
196	getep	get ##ep	get ##ep	_get ep	2/2/2
197	gubeg	gub ##eg	gu ##be ##g	_gu beg	2/3/2
198	gugua	gug ##ua	gu ##gua	_gu gua	2/2/2
199	gulem	gul ##em	gul ##em	_gul em	2/2/2
200	guluk	gul ##uk	gul ##uk	_gu luk	2/2/2
201	gumi	gum ##i	gum ##i	_gumi	2/2/1
202	guncang	gunc ##ang	gun ##cang	_gu ncang	2/2/2
203	guncangan	gunc ##angan	gun ##cang ##an	_gu ncang an	2/3/3
204	ha	ha	ha	_ha	1/1/1
205	harapan	harapan	harapan	_harapan	1/1/1
206	hasil	hasil	hasil	_hasil	1/1/1
207	hebat	hebat	heb ##at	_hebat	1/2/1
208	heboh	heboh	heb ##oh	_heb oh	1/2/2
209	heroik	hero ##ik	heroi ##k	_hero ik	2/2/2
210	hukum	hukum	hukum	_hukum	1/1/1
211	hukuma	hukum ##a	hukum ##a	_hukum a	2/2/2
212	idep	ide ##p	ide ##p	_ide p	2/2/2
213	idup	id ##up	id ##up	_id up	2/2/2
214	ila	ila	ila	_il a	1/1/2
215	ilang	ilang	ilang	_ilang	1/1/1
216	indah	indah	indah	_indah	1/1/1
217	indik	ind ##ik	indi ##k	_indik	2/2/1
218	inget	inget	inge ##t	_inget	1/2/1
219	inggi	ing ##gi ##h	inggi	_inggi	3/1/1
220	inguh	ing ##uh	ing ##uh	_ing uh	2/2/2

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
221	interogasi	inter ##ogasi	intero ##gasi	_inte roga si	2/2/3
222	iraga	ira ##ga	ira ##ga	_i raga	2/2/2
223	irup	ir ##up	ir ##up	_i rup	2/2/2
224	iseng	iseng	ise ##ng	_is eng	1/2/2
225	isi	isi	isi	_isi	1/1/1
226	istimewa	istimewa	isti ##me ##wa	_istimewa	1/3/1
227	iteh	it ##eh	it ##eh	_i teh	2/2/2
228	iwasin	iw ##asin	iwa ##sin	_ iwas in	2/2/3
229	jagat	jagat	jag ##at	_jag at	1/2/2
230	jagra	jag ##ra	jag ##ra	_jag ra	2/2/2
231	jangkep	jangk ##ep	jang ##kep	_ja ng kep	2/2/3
232	jantengang	jant ##engan ##g	jan ##teng ##ang	_ja nten gang	3/3/3
233	jawab	jawab	jawab	_jawab	1/1/1
234	jaya	jaya	jaya	_ja ya	1/1/2
235	jegeg	je ##ge ##g	jeg ##eg	_jeg eg	3/2/2
236	jekeh	jej ##eh	jej ##eh	_jej eh	2/2/2
237	jekehina	jej ##ehi ##na	jej ##eh ##ina	_je je hina	3/3/3
238	jelas	jelas	jelas	_jelas	1/1/1
239	jele	jel ##e	je ##le	_je le	2/2/2
240	jemak	jem ##ak	je ##mak	_je mak	2/2/2
241	jengah	jeng ##ah	jen ##gah	_jen gah	2/2/2
242	jit	ji ##t	ji ##t	_ji t	2/2/2
243	juang	juang	juan ##g	_ju ang	1/2/2
244	juk	juk	ju ##k	_ juk	1/2/2
245	junjung	jun ##jung	jun ##jung	_jun jung	2/2/2
246	jurit	juri ##t	juri ##t	_juri t	2/2/2
247	juru selamat	juru selamat	juru selama ##t	_juru _selamat	2/3/2
248	kabania	kab ##anian	ka ##bani ##an	_kaba nian	2/3/2
249	kabecikan	kab ##eci ##kan	ka ##bec ##ikan	_ka bec ikan	3/3/3
250	kabenengan	kab ##ene ##ngan	ka ##benen ##gan	_ka ben engan	3/3/3
251	kabinawa	kabin ##awa	kabi ##na ##wa	_kabin awa	2/3/2
252	kablinger	kab ##ling ##er	ka ##bling ##er	_kab lincer	3/3/2
253	kacorahan	kac ##ora ##han	ka ##cor ##ahan	_ka co rahan	3/3/3
254	kaganggu	kaga ##ng ##gu	ka ##gang ##gu	_ka ganggu	3/3/2
255	kagantung	kaga ##nt ##ung	ka ##gant ##ung	_ka gan tung	3/3/3
256	kagiat	kag ##iat	ka ##gia ##t	_ka gia t	2/3/3
257	kailangan	kail ##angan	kai ##lengan	_kailangan	2/2/1
258	kajejehin	kaj ##ej ##ehi ##n	kaj ##eje ##hin	_kaj e je hin	4/3/4
259	kajerihan	kaj ##eri ##han	kaj ##eri ##han	_kaj eri han	3/3/3
260	kalah	kalah	kalah	_kalah	1/1/1
261	kalahanga	kalah ##anga	kalah ##anga	_kala hanga	2/2/2
262	kalahin	kalah ##in	kalah ##in	_kalah in	2/2/2
263	kalahina	kalah ##ina	kalah ##ina	_kalah ina	2/2/2
264	kalaina	kala ##ina	kala ##ina	_kala ina	2/2/2
265	kalin	kali ##n	kali ##n	_kali n	2/2/2
266	kangen	kangen	kang ##en	_kang en	1/2/2
267	kapademan	kap ##adem ##an	kap ##ade ##man	_kapa de man	3/3/3
268	kapal	kapal	kapal	_kapal	1/1/1
269	kapanggil-panggil	kapang ##gg ##il - panggil	kap ##ang ##gil - pang ##gil	_ka pang gil - pang gil	5/6/6
270	kapedihan	kap ##edi ##han	kap ##edi ##han	_ka pedi han	3/3/3
271	kapelihan	kap ##elih ##an	kapel ##ihan	_kape lihan	3/2/2
272	kapengaruh	kap ##engaruh	kap ##eng ##aru ##h	_ka pengar uh	2/4/3
273	kapracayain	kap ##rac ##aya ##in	kap ##raca ##yai ##n	_ka prac aya in	4/4/4
274	karuan	karuan	kar ##uan	_kar uan	1/2/2
275	karuetan	kar ##ue ##tan	kar ##uet ##an	_karu etan	3/3/2
276	karugian	kar ##ugian	kar ##ugi ##an	_karu gian	2/3/2
277	karya	karya	karya	_karya	1/1/1
278	kasub	kasub	kas ##ub	_ka sub	1/2/2
279	kasugihan	kas ##ugihan	kas ##ugi ##han	_kasu gihan	2/3/2
280	kasungkawan	kas ##ung ##kawan	kas ##ung ##kawa ##n	_kas ung kawan	3/4/3

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
281	katangkep	katan ##g ##kep	kata ##ng ##kep	_kata ng kep	3/3/3
282	kateror	kate ##ror	kater ##or	_katero r	2/2/2
283	katincapin	kat ##inc ##api ##n	kati ##nca ##pin	_kati nca pin	4/3/3
284	katos	kat ##os	kato ##s	_ka tos	2/2/2
285	katuduh	kat ##udu ##h	kat ##udu ##h	_katu duh	3/3/2
286	katul	kat ##ul	kat ##ul	_ka tul	2/2/2
287	kausakan	kaus ##akan	ka ##usa ##kan	_kau sakan	2/3/2
288	kauyak	kau ##yak	ka ##uya ##k	_kau yak	2/3/2
289	keadilan	keadilan	ke ##adilan	_keadilan	1/2/1
290	keaduk	ke ##aduk	ke ##adu ##k	_ke a duk	2/3/3
291	keamanan	keamanan	keamanan	_keamanan	1/1/1
292	kebecikan	keb ##eci ##kan	ke ##bec ##ikan	_ke bec ikan	3/3/3
293	kebeleran	keb ##ele ##ran	ke ##bele ##ran	_ke bele ran	3/3/3
294	kebencian	kebencian	ke ##ben ##cian	_ke ben cian	1/3/3
295	kebenehan	keb ##ene ##han	ke ##bene ##han	_ke bene han	3/3/3
296	kebenyahan	keb ##enya ##han	ke ##ben ##yah ##an	_ke ben yahan	3/4/3
297	kebus	keb ##us	ke ##bus	_ke bus	2/2/2
298	kecag	kec ##ag	ke ##ca ##g	_ke ca g	2/3/3
299	kedasina	keda ##sin ##a	ked ##asi ##na	_ke das ina	3/3/3
300	kedek	ked ##ek	ked ##ek	_ke dek	2/2/2
301	kejam	kejam	ke ##jam	_ke jam	1/2/2
302	keji	keji	ke ##ji	_ke ji	1/2/2
303	kekerasan	kekerasan	kek ##eras ##an	_kekerasan	1/3/1
304	kekewehan	kek ##ewe ##han	kek ##ew ##ehan	_ke ke we han	3/3/4
305	kekraasan	kek ##rasa ##n	kek ##rasa ##n	_ke kra san	3/3/3
306	kekuatan	kekuatan	kekuatan	_kekuatan	1/1/1
307	keleb	kel ##eb	ke ##leb	_ke leb	2/2/2
308	kemadina	kem ##adin ##a	kem ##adi ##na	_ke ma dina	3/3/3
309	kemaling	kem ##aling	kemal ##ing	_ke ma ling	2/2/3
310	kematian	kematian	kematian	_kematian	1/1/1
311	kematiang	kematian ##g	kematian ##g	_kematian g	2/2/2
312	kena	kena	ken ##a	_kena	1/2/1
313	kenal	kenal	ken ##al	_kenal	1/2/1
314	kendel	kend ##el	kende ##l	_ken del	2/2/2
315	keneh	ken ##eh	ken ##eh	_kene h	2/2/2
316	kenen	ken ##en	ken ##en	_ken en	2/2/2
317	kenin	ken ##in	ken ##in	_ke nin	2/2/2
318	kenjel	ken ##je ##l	ken ##jel	_ken jel	3/2/2
319	kenyamanan	kenyamanan	kenya ##mana ##n	_kenyamanan	1/3/1
320	kenyat	kenya ##t	kenya ##t	_ke nya t	2/2/3
321	kenyehan	keny ##ehan	ken ##ye ##han	_ke nye han	2/3/3
322	kenyel	keny ##el	ken ##yel	_ke ny el	2/2/3
323	kenyem	keny ##em	ken ##ye ##m	_ke ny em	2/3/3
324	kepentingan	kepentingan	kep ##enti ##ngan	_kepentingan	1/3/1
325	kerahayuan	kerah ##ayu ##an	ker ##aha ##yuan	_ke raha yuan	3/3/3
326	kerangkeng	kerang ##ken ##g	ker ##ang ##ken ##g	_ke rang ke ng	3/4/4
327	keren	keren	keren	_keren	1/1/1
328	kerep	ker ##ep	ker ##ep	_ke rep	2/2/2
329	kesan	kesan	kesan	_kesan	1/1/1
330	kesempatan	kesempatan	kes ##empatan	_kesempatan	1/2/1
331	kese pian	kese pian	kes ##ep ##ian	_kese pian	1/3/2
332	ketat	ketat	ket ##at	_ketat	1/2/1
333	ketegangan	ketegangan	kete ##gang ##an	_kete gangan	1/3/2
334	kewanenan	kewan ##enan	kew ##anen ##an	_ke wane nan	2/3/3
335	keweh	kew ##eh	kew ##eh	_ke we h	2/2/3
336	kilangan	kil ##angan	kila ##ngan	_ki langan	2/2/2
337	kimudang	kim ##udang	kim ##uda ##ng	_ki mu dang	2/3/3
338	kirig	kiri ##g	kiri ##g	_ki rig	2/2/2
339	kompak	kompak	kom ##pak	_kompa k	1/2/2
340	konflik	konflik	konflik	_konflik	1/1/1
341	konspirasi	konspirasi	kon ##sp ##iras ##i	_konspira si	1/4/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (I/m/X)
342	korban	korban	korban	__korban	1/1/1
343	korban jiwa	korban jiwa	korban jiwa	__korban __jiwa	2/2/2
344	kual	kual	ku ##al	__ku al	1/2/2
345	kuang	kua ##ng	ku ##ang	__ku ang	2/2/2
346	kuasa	kuasa	kuasa	__kuasa	1/1/1
347	kuat	kuat	kuat	__kuat	1/1/1
348	kuciwa	kuc ##iwa	ku ##ci ##wa	__ku ci wa	2/3/3
349	kukuh	kuku ##h	ku ##ku ##h	__kuku h	2/3/2
350	kuri	kur ##i	kuri	__kuri	2/1/1
351	kusakan	kus ##akan	kus ##akan	__ku sakan	2/2/2
352	kutang	kuta ##ng	ku ##tang	__ku tang	2/2/2
353	lacur	lac ##ur	lac ##ur	__la cur	2/2/2
354	langgar	langg ##ar	lang ##gar	__lang gar	2/2/2
355	lantang	lantang	lan ##tang	__la ntang	1/2/2
356	larang	larang	lara ##ng	__la rang	1/2/2
357	larangan	larangan	lara ##ngan	__la rangan	1/2/2
358	lawan	lawan	law ##an	__lawan	1/2/1
359	lebih	lebih	lebih	__lebih	1/1/1
360	lebih luung	lebih lu ##ung	lebih luu ##ng	__lebih __lu ung	3/3/3
361	lebih melah	lebih melah	lebih mela ##h	__lebih __me lah	2/3/3
362	lebihan	lebih ##an	lebih ##an	__lebih an	2/2/2
363	ledang	led ##ang	leda ##ng	__le dang	2/2/2
364	lekina	lek ##ina	lek ##ina	__lekin a	2/2/2
365	lempuyengan	lemp ##uy ##engan	lem ##pu ##yen ##gan	__le mpu y engan	3/4/4
366	lengh	leng ##eh	leng ##h	__leng h	2/2/2
367	lengkara	lengk ##ara	len ##g ##kara	__leng kara	2/3/2
368	lenyap	lenyap	len ##ya ##p	__len yap	1/3/2
369	liang	liang	liang	__li ang	1/1/2
370	lincah	lincah	lin ##cah	__lin cah	1/2/2
371	liwat	li ##wat	li ##wat	__li wat	2/2/2
372	luar biasa	luar biasa	luar biasa	__luar __biasa	2/2/2
373	luung	lu ##ung	luu ##ng	__lu ung	2/2/2
374	maan	ma ##an	maan	__maan	2/1/1
375	maang	ma ##ang	maan ##g	__ma ang	2/2/2
376	mabaya	mab ##aya	mab ##aya	__ma baya	2/2/2
377	mael	ma ##el	mae ##l	__ma el	2/2/2
378	maengkeb	maen ##g ##keb	mae ##ng ##ke ##b	__mae ng ke b	3/4/4
379	magetih	mag ##eti ##h	mag ##eti ##h	__mag et ih	3/3/3
380	maguna	mag ##una	mag ##una	__mag una	2/2/2
381	makatang	maka ##tang	maka ##tang	__maka tang	2/2/2
382	makenyem	make ##nye ##m	maken ##ye ##m	__maken y em	3/3/3
383	makomitmen	mak ##omi ##tm ##en	mak ##omi ##t ##men	__ma kom it men	4/4/4
384	makrana	mak ##rana	mak ##rana	__mak rana	2/2/2
385	makudang-kudang	mak ##udang - kuda ##ng	mak ##uda ##ng - kuda ##ng	__maku dang - ku dang	5/6/5
386	makutang	mak ##utang	mak ##utan ##g	__maku tang	2/3/2
387	maling	maling	mali ##ng	__mal ing	1/2/2
388	malinga	maling ##a	mali ##nga	__mal inga	2/2/2
389	malingang	maling ##ang	mali ##ngan ##g	__mal ing ang	2/3/3
390	mamelit	mam ##elit	mam ##eli ##t	__mame lit	2/3/2
391	mamongor	mam ##ongo ##r	mam ##ongo ##r	__ma mon gor	3/3/3
392	manah	mana ##h	mana ##h	__man ah	2/2/2
393	mapiguna	map ##igu ##na	mapi ##guna	__map i guna	3/2/3
394	mapikenoh	map ##ike ##no ##h	mapi ##ken ##oh	__ma pik en oh	4/3/4
395	maplanting	map ##lan ##ting	map ##lanti ##ng	__ma plant ing	3/3/3
396	marti	mart ##i	marti	__marti	2/1/1
397	masaat-saatang	masa ##at - saat ##ang	masa ##at - saat ##ang	__masa at - saat ang	5/5/5
398	masalah	masalah	masalah	__masalah	1/1/1
399	masemaya	mas ##ema ##ya	mase ##maya	__mase maya	3/2/2
400	masiat	masi ##at	masia ##t	__mas iat	2/2/2
401	mastiang	mas ##tian ##g	mast ##ian ##g	__masti ang	3/3/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
402	mastu	mas ##tu	mast ##u	_mas tu	2/2/2
403	masuryak	mas ##ury ##ak	masu ##rya ##k	_masu ry ak	3/3/3
404	matakön	matak ##on	mata ##kon	_mata kon	2/2/2
405	matatu	mata ##tu	mata ##tu	_mata tu	2/2/2
406	matembak	mat ##embak	mate ##mba ##k	_mate mbak	2/3/2
407	mati	mati	mati	_mati	1/1/1
408	matiang	mati ##ang	mati ##ang	_mati ang	2/2/2
409	matunda	mat ##unda	mat ##unda	_mat unda	2/2/2
410	matur suksma	matur suk ##sma	mat ##ur suk ##sma	_matur _suk sma	3/4/3
411	matur suksma	matur ##suk ##sma	mat ##urs ##uks ##ma	_matur suk sma	3/4/3
412	mayit	may ##it	may ##it	_may it	2/2/2
413	mebakti	meb ##akti	me ##bak ##ti	_me bak ti	2/3/3
414	mecapatan	mec ##apatan	mec ##apat ##an	_me ca patan	2/3/3
415	medil	medi ##il	medi ##il	_med il	2/2/2
416	megetih	meg ##eti ##h	meget ##ih	_meget ih	3/2/2
417	mekalin	mek ##alin	me ##kali ##n	_me kali n	2/3/3
418	mekedeng	mek ##ede ##ng	me ##kede ##ng	_me ke deng	3/3/3
419	mela	mel ##a	mela	_mela	2/1/1
420	melah	melah	mela ##h	_me lah	1/2/2
421	melakukan	melakukan	melakukan	_melakukan	1/1/1
422	meliah	melia ##h	meli ##ah	_meli ah	2/2/2
423	melihang	mel ##ihan ##g	meli ##hang	_meli hang	3/2/2
424	memedi	meme ##di	meme ##di	_me medi	2/2/2
425	menci	menc ##i	men ##ci	_men ci	2/2/2
426	mencurigakan	mencurigakan	men ##curi ##ga ##kan	_mencuri ga kan	1/4/3
427	mendep	mend ##ep	mende ##p	_mende p	2/2/2
428	mengekeb	mengek ##eb	menge ##ke ##b	_menge ke b	2/3/3
429	mengerikan	mengerikan	menge ##rikan	_menge rikan	1/2/2
430	mengkeb	mengk ##eb	men ##g ##ke ##b	_meng ke b	2/4/3
431	menyahn	menya ##hin	men ##yah ##in	_meny ahin	2/3/2
432	mepadu	me ##pad ##u	me ##padu	_me padu	3/2/2
433	merangin	mera ##ng ##in	mera ##ngi ##n	_me rang in	3/3/3
434	mesaing	mes ##aing	mesa ##ing	_mesa ing	2/2/2
435	mesuang	mes ##uang	mes ##uang	_mes u ang	2/2/3
436	metatu	meta ##tu	meta ##tu	_met atu	2/2/2
437	mik	mik	mi ##k	_mik	1/2/1
438	milu	mil ##u	mil ##u	_mil u	2/2/2
439	misi	misi	misi	_misi	1/1/1
440	mogbög	mog ##bö ##g	mo ##gb ##og	_ mog bog	3/3/3
441	mogi	mog ##i	mo ##gi	_mo gi	2/2/2
442	molihang	mol ##ihan ##g	mol ##ihan ##g	_moli hang	3/3/2
443	molihin	mol ##ihin	mol ##ihin	_mol ihin	2/2/2
444	mua	mu ##a	mua	_mua	2/1/1
445	mudah	mudah	mudah	_mudah	1/1/1
446	muji	muj ##i	mu ##ji	_mu ji	2/2/2
447	muka	muka	mu ##ka	_muka	1/2/1
448	mukak	muka ##k	mu ##kak	_muka k	2/2/2
449	musuh	musuh	musuh	_musuh	1/1/1
450	muungang	mu ##ungan ##g	muun ##gang	_muun gang	3/2/2
451	naan	na ##an	naa ##n	_na an	2/2/2
452	naenang	na ##enang	na ##enan ##g	_na en ang	2/3/3
453	nagih	nag ##ih	nag ##ih	_nag ih	2/2/2
454	nagih pelih	nag ##ih pelih	nag ##ih peli ##h	_nag ih _peli h	3/4/4
455	nahan	nah ##an	nah ##an	_na han	2/2/2
456	nak belög	nak bel ##ög	nak belo ##g	_nak _be log	3/3/3
457	nak lengh	nak leng ##eh	nak lenge ##h	_nak _lenge h	3/3/3
458	nakal	nakal	nak ##al	_na kal	1/2/2
459	nakonang	nak ##onan ##g	nakon ##ang	_nakon ang	3/2/2
460	nakonin	nak ##onin	nakon ##in	_nakon in	2/2/2
461	nampi	nam ##pi	nam ##pi	_na mpi	2/2/2
462	nangkep	nang ##kep	nang ##kep	_nang kep	2/2/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
463	nekat	nekat	neka ##t	_ne kat	1/2/2
464	nembak	nem ##bak	nem ##bak	_ne mbak	2/2/2
465	nengil	neng ##il	nen ##gil	_nen gil	2/2/2
466	neraka	neraka	nera ##ka	_neraka	1/2/1
467	nerima	ner ##ima	neri ##ma	_ne rima	2/2/2
468	ngabih	ngab ##ih	nga ##bi ##h	_nga bih	2/3/2
469	ngacep	nga ##ce ##p	nga ##ce ##p	_nga cep	3/3/2
470	ngaenang	nga ##enang	nga ##enan ##g	_nga en ang	2/3/3
471	ngagah	nga ##ga ##h	nga ##gah	_nga gah	3/2/2
472	ngajiang	ngaji ##ang	nga ##jian ##g	_nga ji ang	2/3/3
473	ngakuin	ngaku ##in	nga ##kui ##n	_nga ku in	2/3/3
474	ngakunin	ngaku ##ni ##n	nga ##kun ##in	_nga ku nin	3/3/3
475	ngamanin	nga ##mani ##n	ngam ##anin	_nga man in	3/2/3
476	ngamatiang	nga ##mati ##ang	ngam ##atia ##ng	_nga mati ang	3/3/3
477	ngampak	nga ##mp ##ak	ngam ##pak	_ng ampak	3/2/2
478	ngancurang	ngan ##cu ##rang	ngan ##cura ##ng	_ngan cu rang	3/3/3
479	ngantiang	ngan ##tian ##g	nganti ##ang	_nganti ang	3/2/2
480	ngantosin	ngan ##to ##sin	ngan ##tos ##in	_ng anto sin	3/3/3
481	ngarep	nga ##rep	nga ##re ##p	_nga rep	2/3/2
482	ngargain	nga ##r ##gai ##n	nga ##rga ##in	_nga rga in	4/3/3
483	ngastawa	nga ##st ##awa	nga ##sta ##wa	_nga s tawa	3/3/3
484	ngebayain	ngeb ##aya ##in	ng ##eba ##yai ##n	_nge baya in	3/4/3
485	ngedeng	nged ##eng	ng ##eden ##g	_nge deng	2/3/2
486	ngedoktrin	nged ##ok ##trin	ng ##edo ##kt ##rin	_nge dok trin	3/4/3
487	ngeh	ng	ng ##eh	_ng eh	1/2/2
488	ngelanggar	ngel ##anggar	ng ##ela ##ngga ##r	_nge lang gar	2/4/3
489	ngelemahin	ngel ##emah ##in	ng ##elem ##ah ##in	_ng elem ahin	3/4/3
490	ngelindungin	ngel ##indungi ##n	ng ##elin ##dungi ##n	_ng eli ndung in	3/4/4
491	ngelumpuhang	ngel ##umpuhan ##g	ng ##elu ##mp ##uhan ##g	_nge lump u hang	3/5/4
492	ngelumpuhin	ngel ##ump ##uhi ##n	ng ##elu ##mp ##uhi ##n	_nge lump u hin	4/5/4
493	ngematiang	ngem ##atian ##g	ng ##ema ##tian ##g	_nge mati ang	3/4/3
494	ngematiang raga	ngem ##atian ##g raga	ng ##ema ##tian ##g rag ##a	_nge mati ang _raga	4/6/4
495	ngendukang	nge ##nd ##ukan ##g	ng ##end ##ukan ##g	_ng en duk ang	4/4/4
496	ngeng	nge ##ng	ng ##eng	_ng eng	2/2/2
497	ngerebeda	nger ##eb ##eda	ng ##ere ##bed ##a	_ng ere beda	3/4/3
498	ngeri	ngeri	ng ##eri	_ng eri	1/2/2
499	ngerti	ngerti	ng ##erti	_ngerti	1/2/1
500	ngeyehin	nge ##ye ##hin	ng ##eye ##hin	_ng eye hin	3/3/3
501	nggih	ngg ##ih	ng ##gi ##h	_ nggih	2/3/2
502	ngidih	ng ##idih	ng ##idi ##h	_ng id ih	2/3/3
503	ngidih pelih	ng ##idih pelih	ng ##idi ##h peli ##h	_ng id ih _peli h	3/5/5
504	ngimpasin	ng ##imp ##asin	ng ##im ##pasi ##n	_ng impa sin	3/4/3
505	ngindarin	ngin ##dari ##n	ng ##inda ##rin	_ng inda rin	3/3/3
506	nglaksanayang	ngl ##aksana ##yang	ng ##lak ##sana ##yang	_ng laksana yang	3/4/3
507	nglanggar	ngl ##anggar	ng ##lang ##gar	_ng lang gar	2/3/3
508	ngoopin	ng ##oo ##pin	ngo ##opi ##n	_ngo o pin	3/3/3
509	ngoyong	ng ##oyong	ngo ##yong	_ngo yong	2/2/2
510	nguda	ng ##uda	ngu ##da	_ng uda	2/2/2
511	nguliang	ng ##ulia ##ng	ngu ##lian ##g	_ng uli ang	3/3/3
512	ngulungang	ng ##ulung ##ang	ngu ##lung ##ang	_ng ulungan g	3/3/3
513	nguotin	ng ##u ##otin	ngu ##oti ##n	_ng uoti n	3/3/3
514	ngusir	ng ##usi ##r	ngu ##sir	_ngu sir	3/2/2
515	ngwantu	ng ##wan ##tu	ng ##wan ##tu	_ng wan tu	3/3/3
516	nilai	nilai	nilai	_nilai	1/1/1
517	nincapang	nin ##cap ##ang	nin ##cap ##ang	_ni nca pang	3/3/3
518	ning	ning	ning	_ning	1/1/1
519	ningkatin	ning ##kati ##n	ning ##kati ##n	_ning kat in	3/3/3
520	nongos	nong ##os	nong ##os	_non gos	2/2/2
521	nongsok	nong ##so ##k	nong ##sok	_no ng sok	3/2/3

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
522	norok	nor ##ok	nor ##ok	—no rok	2/2/2
523	norong	nor ##ong	nor ##ong	—no rong	2/2/2
524	nraka	nr ##aka	nr ##aka	—n raka	2/2/2
525	nukung	nuk ##ung	nuk ##ung	—nuk ung	2/2/2
526	nulud	nu ##lu ##d	nul ##ud	—nu lud	3/2/2
527	nulungin	nu ##lung ##in	nul ##ung ##in	—nu lung in	3/3/3
528	numbuk	num ##buk	num ##buk	—nu mbuk	2/2/2
529	nunas	nun ##as	nuna ##s	—nuna s	2/2/2
530	nunas ampura	nun ##as amp ##ura	nuna ##s amp ##ura	—nuna s —am pura	4/4/4
531	nutup	nut ##up	nut ##up	—nut up	2/2/2
532	nuunin	nu ##uni ##n	nu ##uni ##n	—nu u nin	3/3/3
533	nuut	nu ##ut	nu ##ut	—nu ut	2/2/2
534	nyalahang	nyala ##han ##g	nya ##lah ##ang	—nya la hang	3/3/3
535	nyanan	nya ##nan	nya ##nan	—ny anan	2/2/2
536	nyandang	nya ##nd ##ang	nya ##ndang	—nya ndang	3/2/2
537	nyanggra	nya ##ng ##gr ##a	nya ##ng ##gra	—ny ang gra	4/3/3
538	nyawab	nyawa ##b	nya ##wa ##b	—nyawa b	2/3/2
539	nyebabang	ny ##ebab ##ang	nye ##bab ##ang	—nye ba bang	3/3/3
540	nyebak	ny ##ebak	nye ##bak	—nye bak	2/2/2
541	nyeh	ny ##eh	nye ##h	—nye h	2/2/2
542	nyerang	ny ##erang	nye ##rang	—nye rang	2/2/2
543	nyeret	ny ##eret	nye ##ret	—nyere t	2/2/2
544	nyidang	nyi ##dang	ny ##idan ##g	—ny i dang	2/3/3
545	nyita	nyi ##ta	ny ##ita	—ny ita	2/2/2
546	nyokong	nyok ##ong	ny ##okon ##g	—n yokong	2/3/2
547	nyorog	ny ##oro ##g	ny ##oro ##g	—nyo rog	3/3/2
548	nyurug	ny ##urug	nyu ##ru ##g	—ny u rug	2/3/3
549	oke	oke	ok ##e	—o ke	1/2/2
550	olas	ol ##as	ola ##s	—o las	2/2/2
551	olih	oli ##h	oli ##h	—oli h	2/2/2
552	orang	orang	orang	—orang	1/1/1
553	oranga	orang ##a	orang ##a	—orang a	2/2/2
554	padem	pad ##em	pad ##em	—pa dem	2/2/2
555	pademang	pad ##emang	pad ##eman ##g	—pa de mang	2/3/3
556	padet	pad ##et	pad ##et	—pa det	2/2/2
557	padidi	padi ##di	padi ##di	—padi di	2/2/2
558	pageh	page ##h	page ##h	—page h	2/2/2
559	pahlawan	pahlawan	pa ##lita ##wan	—pah lawan	1/3/2
560	paid	pai ##d	paid	—paid	2/1/1
561	pakeling	pake ##ling	pak ##eling	—pake ling	2/2/2
562	paling gede	paling gede	paling ge ##de	—paling —gede	2/3/2
563	palsu	palsu	pal ##su	—palsu	1/2/1
564	pamastu	pam ##ast ##u	pam ##ast ##u	—pama stu	3/3/2
565	panelas	panel ##as	panel ##as	—panel as	2/2/2
566	pangarga	pang ##arga	pang ##ar ##ga	—panga rga	2/3/2
567	pangkat	pangkat	pangkat	—pangkat	1/1/1
568	pangrejek	pang ##re ##jek	pang ##rej ##ek	—pang re jek	3/3/3
569	panik	panik	pani ##k	—panik	1/2/1
570	panten	pant ##en	pan ##ten	—pan ten	2/2/2
571	pantes	pant ##es	pan ##tes	—pa ntes	2/2/2
572	panyelehan	pan ##ye ##leh ##an	pan ##yel ##ehan	—pa ny ele han	4/3/4
573	panyokong	pan ##yo ##kon ##g	pan ##yok ##ong	—pan yokong	4/3/2
574	parodi	par ##odi	par ##odi	—parodi	2/2/1
575	pasti	pasti	pasti	—pasti	1/1/1
576	pastu	past ##u	past ##u	— pastu	2/2/2
577	patakon	pat ##ako ##n	patak ##on	—pata kon	3/2/2
578	pateh	pat ##eh	pate ##h	—pate h	2/2/2
579	patuh	patuh	pat ##uh	—pa tuh	1/2/2
580	patut	patut	pat ##ut	—patut	1/2/1
581	peceng	pec ##eng	pec ##eng	—pec eng	2/2/2
582	pedalem	pedal ##em	pedal ##em	—pedal em	2/2/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
583	pedesang	pede ##sang	pe ##des ##ang	—pede sang	2/3/2
584	pedidi	ped ##idi	pe ##did ##i	—pedi di	2/3/2
585	pedih	pedih	pe ##di ##h	—pedi h	1/3/2
586	peduli	peduli	pe ##du ##li	—peduli	1/3/1
587	pelaku	pelaku	pela ##ku	—pelaku	1/2/1
588	pelaku kecorahan	pelaku kec ##ora ##han	pela ##ku ke ##cor ##ahan	—pelaku —ke co rahan	4/5/4
589	pelih	pelih	pe ##h	—pe h	1/2/2
590	pelihang	pelih ##ang	pe ##hang	—pe hang	2/2/2
591	pembedilan	pemb ##edi ##lan	pe ##mbe ##dila ##n	—pembe di lan	3/4/3
592	pemberontakan	pemberontakan	pe ##mber ##onta ##kan	—pembe ron takan	1/4/3
593	penangkapan	penangk ##epan	pena ##ng ##kep ##an	—pena ng ke pan	2/4/4
594	pengecut	pengecut	peng ##ecu ##t	—penge cut	1/3/2
595	pengeng	pengen ##g	peng ##eng	—penge ng	2/2/2
596	penjara	penjara	pen ##jara	—penjara	1/2/1
597	penting	penting	penting	—penting	1/1/1
598	penuh	penuh	penuh	—penuh	1/1/1
599	penularan	penularan	pen ##ular ##an	—penu lar an	1/3/3
600	penyelidikan	penyelidikan	pen ##yel ##idi ##kan	—penyelidikan	1/4/1
601	perang	perang	perang	—perang	1/1/1
602	perasaan	perasaan	pera ##saan	—perasaan	1/2/1
603	percaya	percaya	percaya	—percaya	1/1/1
604	percayina	perca ##ina	per ##cay ##ina	—perc ay ina	2/3/3
605	perdamaian	perdamaian	perda ##mai ##an	—per dama ian	1/3/3
606	peringatan	peringatan	perin ##gatan	—peringatan	1/2/1
607	perjuangan	perjuangan	per ##juang ##an	—perjuangan	1/3/1
608	pertempuran	pertempuran	pertempuran	—pertempuran	1/1/1
609	pesan	pesan	pesa ##n	—pesan	1/2/1
610	pesu	pes ##u	pes ##u	—pesu	2/2/1
611	petaken	petak ##en	peta ##ken	—petak en	2/2/2
612	petakon	petak ##on	peta ##kon	—petak on	2/2/2
613	peteng	pet ##eng	pete ##ng	—pe teng	2/2/2
614	pipis	pipis	pi ##pis	—pipi s	1/2/2
615	pis	pis	pis	—pis	1/1/1
616	pisan	pis ##an	pisa ##n	—pisan	2/2/1
617	pisuna	pis ##una	pis ##una	—pis una	2/2/2
618	pokus	pok ##us	po ##kus	—pokus	2/2/1
619	polih	poli ##h	poli ##h	—poli h	2/2/2
620	positip	pos ##itip	pos ##iti ##p	—posi tip	2/3/2
621	prabiya	prab ##iya	pra ##biy ##a	—pra biya	2/3/2
622	pracaya	prac ##aya	praca ##ya	—praca ya	2/2/2
623	pradnya	prad ##nya	pra ##dny ##a	—prad nya	2/3/2
624	pradnyan	prad ##nya ##n	pra ##dny ##an	—prad nya n	3/3/3
625	pragat	prag ##at	praga ##t	—prag at	2/2/2
626	praja	praja	pra ##ja	—pra ja	1/2/2
627	prakara	prak ##ara	pra ##kara	—pra kara	2/2/2
628	pralanang	pra ##lanan ##g	pra ##lanan ##g	—pra lan ang	3/3/3
629	prasida	pras ##ida	pra ##sida	—pra sida	2/2/2
630	provokasina	prov ##okasi ##na	pro ##vo ##kasi ##na	—provok asi na	3/4/3
631	puja	puj ##a	pu ##ja	—puja	2/2/1
632	puji	puji	pu ##ji	—pu ji	1/2/2
633	pul	pul	pul	—pul	1/1/1
634	pungkur	pungk ##ur	pun ##gku ##r	— pung kur	2/3/3
635	pura	pura	pura	—pura	1/1/1
636	purug	pur ##ug	puru ##g	—pu rug	2/2/2
637	puruh	pur ##uh	puru ##h	—pur uh	2/2/2
638	purun	pur ##un	puru ##n	—pur un	2/2/2
639	puyung	puy ##ung	puy ##ung	—pu yung	2/2/2
640	racun	racun	rac ##un	—racun	1/2/1
641	raga	raga	rag ##a	—raga	1/2/1
642	rage	rag ##e	rage	— rage	2/1/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
643	rahajeng	rah ##aj ##eng	raha ##jeng	__raha jeng	3/2/2
644	rahayu	rahayu	raha ##yu	__raha yu	1/2/2
645	rajaheng	raja ##hen ##g	raja ##heng	__raja heng	3/2/2
646	rarisang	ra ##ris ##ang	ra ##ris ##ang	__ rari sang	3/3/3
647	ras	ras	ras	__ras	1/1/1
648	rejek	rejek	re ##jek	__re jek	1/2/2
649	rereh	rer ##eh	rer ##eh	__re re h	2/2/3
650	rireh	ri ##rih	ri ##ri ##h	__ri rih	2/3/2
651	rubuh	rub ##uh	ru ##bu ##h	__rub uh	2/3/2
652	rubuhang	rub ##uhan ##g	ru ##buhun ##g	__rub u hang	3/3/3
653	rudeg	rud ##eg	rude ##g	__rude g	2/2/2
654	rungu	run ##gu	rung ##u	__rung u	2/2/2
655	sabda	sabda	sa ##b ##da	__sab da	1/3/2
656	sadia	sad ##ia	sad ##ia	__sa dia	2/2/2
657	saham	saham	saha ##m	__saham	1/2/1
658	sahasa	sah ##asa	saha ##sa	__sa hasa	2/2/2
659	sahut	sahut	sah ##ut	__sa hut	1/2/2
660	sajaan	saja ##an	saja ##an	__saja an	2/2/2
661	sajan	saja ##n	saja ##n	__saja n	2/2/2
662	sakit	sakit	sakit	__sakit	1/1/1
663	salam	salam	salam	__salam	1/1/1
664	salut	salut	salut	__salut	1/1/1
665	samping	samping	samping	__samping	1/1/1
666	sanget	sang ##et	sange ##t	__sang et	2/2/2
667	sangkala	sangka ##la	sang ##kala	__sang kala	2/2/2
668	sanglir	sang ##lir	sang ##lir	__sang lir	2/2/2
669	sangsara	sang ##sara	sang ##sara	__sang sara	2/2/2
670	sangsaya	sang ##saya	sang ##say ##a	__sang saya	2/3/2
671	santai	santai	santa ##i	__santa i	1/2/2
672	santi	santi	santi	__sa nti	1/1/2
673	sap	sap	sap	__sap	1/1/1
674	sapa	sapa	sap ##a	__sapa	1/2/1
675	saran	saran	saran	__saran	1/1/1
676	sareng	sar ##eng	sar ##eng	__sa reng	2/2/2
677	sasar	sas ##ar	sasa ##r	__sa sar	2/2/2
678	saur	sau ##r	sau ##r	__sa ur	2/2/2
679	saut	sau ##t	saut	__sau t	2/1/2
680	sawa	saw ##a	saw ##a	__sawa	2/2/1
681	sayaga	saya ##ga	saya ##ga	__saya ga	2/2/2
682	sayang	sayang	saya ##ng	__sayang	1/2/1
683	sebet	seb ##et	sebe ##t	__se bet	2/2/2
684	sebut	sebut	se ##but	__sebut	1/2/1
685	seda	sed ##a	seda	__seda	2/1/1
686	sedeng	sed ##eng	sede ##ng	__se deng	2/2/2
687	sedia	sedia	sed ##ia	__sedia	1/2/1
688	sedih	sedih	sed ##ih	__sedih	1/2/1
689	seduk	sed ##uk	sed ##uk	__se duk	2/2/2
690	segan	segan	sega ##n	__se gan	1/2/2
691	seger	seger	seg ##er	__se ger	1/2/2
692	segeran	segera ##n	segera ##n	__segera n	2/2/2
693	sekaa	sek ##aa	seka ##a	__se kaa	2/2/2
694	sekadi	sek ##adi	seka ##di	__sek adi	2/2/2
695	sekar	sekar	seka ##r	__se kar	1/2/2
696	seken	seken	sek ##en	__se ken	1/2/2
697	semangat	semangat	sem ##anga ##t	__semangat	1/3/1
698	sempat	sempat	sempat	__sempat	1/1/1
699	seneng	seneng	sen ##eng	__seneng	1/2/1
700	senggang	sengg ##ang	seng ##gang	__se nggang	2/2/2
701	sengit	sengit	seng ##it	__seng it	1/2/2
702	sengker	sengk ##er	seng ##ker	__seng ker	2/2/2
703	senjata	senjata	senjata	__senjata	1/1/1

No	Kata Bali	IndoBERT	mBERT	XLm-RoBERTa	Jumlah Token (l/m/X)
704	sepatutne	sepatu ##tn ##e	sep ##atu ##tne	__sepatu tne	3/3/2
705	sepi	sepi	sep ##i	__se pi	1/2/2
706	seragam	seragam	sera ##gam	__ser agam	1/2/2
707	seram	seram	sera ##m	__se ram	1/2/2
708	serang	serang	seran ##g	__se rang	1/2/2
709	serangan	serangan	serangan	__serangan	1/1/1
710	serem	serem	serem	__serem	1/1/1
711	siap	siap	sia ##p	__siap	1/2/1
712	siep	si ##ep	sie ##p	__sie p	2/2/2
713	siepanga	si ##epang ##a	sie ##pang ##a	__sie panga	3/3/2
714	sigap	sigap	sig ##ap	__sig ap	1/2/2
715	simpati	simpati	sim ##pati	__simpati	1/2/1
716	simpatik	simpatik	sim ##pati ##k	__simpati k	1/3/2
717	sing	sing	sing	__sing	1/1/1
718	sing aman	sing aman	sing aman	__sing __aman	2/2/2
719	sing kalah	sing kalah	sing kalah	__sing __kalah	2/2/2
720	sing runguanga	sing run ##gu ##anga	sing rung ##uang ##a	__sing __rung u anga	4/4/4
721	singkaon	singk ##a ##on	sing ##kao ##n	__sing ka on	3/3/3
722	sirna	sirna	sir ##na	__si rna	1/2/2
723	sisi	sisi	sisi	__sisi	1/1/1
724	siteng	site ##ng	siten ##g	__site ng	2/2/2
725	sokong	sok ##ong	soko ##ng	__so kong	2/2/2
726	sokongan	sok ##ongan	soko ##ngan	__sokongan	2/2/1
727	soleh	soleh	sole ##h	__sole h	1/2/2
728	soleh ne	soleh ne	sole ##h ne	__sole h __ne	2/3/3
729	solehne	soleh ##ne	sole ##hne	__sole hne	2/2/2
730	solidaritas	solidaritas	solida ##rita ##s	__solidari tas	1/3/2
731	sopan	sopan	so ##pan	__so pan	1/2/2
732	soroh	sor ##oh	sor ##oh	__so roh	2/2/2
733	stres	stres	str ##es	__stres	1/2/1
734	striker	striker	striker	__strike r	1/1/2
735	suap	suap	sua ##p	__sua p	1/2/2
736	suarga	suar ##ga	sua ##rga	__sua rga	2/2/2
737	sugih	sugi ##h	su ##gi ##h	__sug ih	2/3/2
738	sukeh	suk ##eh	suk ##eh	__suke h	2/2/2
739	suksma	suk ##sma	suk ##sma	__suk sma	2/2/2
740	sumangsaya	sum ##angsa ##ya	suma ##ngs ##aya	__su mang saya	3/3/3
741	sungkawa	sungk ##awa	sung ##kawa	__sung kawa	2/2/2
742	sungsut	sung ##su ##t	sung ##sut	__sung sut	3/2/2
743	super	super	super	__super	1/1/1
744	suryak	surya ##k	surya ##k	__sur yak	2/2/2
745	suryakan	surya ##kan	surya ##kan	__sur ya kan	2/2/3
746	suud	su ##ud	suu ##d	__su ud	2/2/2
747	suung	su ##ung	suu ##ng	__su ung	2/2/2
748	sweca	sw ##eca	sw ##eca	__swe ca	2/2/2
749	taanang	ta ##anan ##g	ta ##anan ##g	__ta anang	3/3/3
750	tabrak	tab ##rak	tab ##rak	__ta brak	2/2/2
751	taenang	tae ##nan ##g	tae ##nang	__ta en ang	3/2/3
752	tahan	tahan	tahan	__tahan	1/1/1
753	tahanan	tahanan	tahan ##an	__tahanan	1/2/1
754	taken	take ##n	taken	__taken	2/1/1
755	takon	tak ##on	tako ##n	__tako n	2/2/2
756	takonina	tak ##onin ##a	tako ##nina	__tako nina	3/2/2
757	takut	takut	tak ##ut	__takut	1/2/1
758	takutina	takut ##ina	tak ##uti ##na	__takut ina	2/3/2
759	tanggung jawab	tanggung jawab	tang ##gung jawab	__tanggung __jawab	2/3/2
760	tangkepa	tangk ##ep ##a	tang ##kep ##a	__tang ke pa	3/3/3
761	tantangan	tantangan	tantang ##an	__tantangan	1/2/1
762	tatkala	tatkala	tat ##kala	__tat kala	1/2/2
763	tatu	tat ##u	tat ##u	__tatu	2/2/1

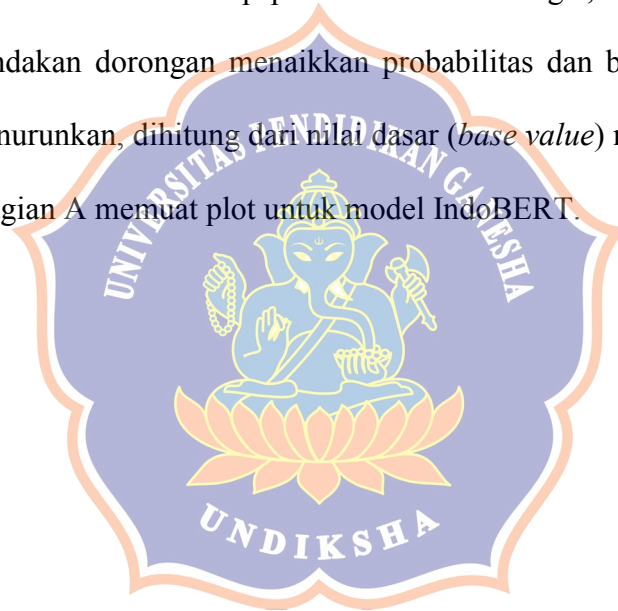
No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
764	tawah	tawa ##h	ta ##wah	_ta wah	2/2/2
765	tawah ne	tawa ##h ne	ta ##wah ne	_ta wah _ne	3/3/3
766	tegang	tegang	tega ##ng	_te gang	1/2/2
767	teges	teg ##es	te ##ges	_te ges	2/2/2
768	tegt	teg ##te ##g	te ##gte ##g	_te gte g	3/3/3
769	teja	te ##ja	tej ##a	_te ja	2/2/2
770	tekep	tek ##ep	tek ##ep	_te kep	2/2/2
771	telah	telah	telah	_telah	1/1/1
772	telas	tel ##as	tela ##s	_tela s	2/2/2
773	tempur	tempur	temp ##ur	_temp ur	1/2/2
774	tenang	tenang	tenan ##g	_tenang	1/2/1
775	tenggelam	tenggelam	ten ##gge ##lam	_tenggelam	1/3/1
776	tengkejut	tengk ##ej ##ut	ten ##g ##kej ##ut	_teng ke jut	3/4/3
777	terkenal	terkenal	terkenal	_terkenal	1/1/1
778	teror	teror	ter ##or	_teror	1/2/1
779	terora	teror ##a	ter ##ora	_teror a	2/2/2
780	tersangka	tersangka	ter ##sang ##ka	_tersangka	1/3/1
781	tersinggung	tersinggung	ter ##sing ##gung	_tersi nggung	1/3/2
782	tertarik	tertarik	ter ##tari ##k	_tertarik	1/3/1
783	tertentu	tertentu	tertentu	_tertentu	1/1/1
784	teteg	tet ##eg	tete ##g	_te te g	2/2/3
785	tetep	tetep	tete ##p	_tetep	1/2/1
786	tiang	tiang	tian ##g	_ti ang	1/2/2
787	tileh	til ##eh	til ##eh	_til eh	2/2/2
788	tincap	tin ##cap	tin ##cap	_tin cap	2/2/2
789	tingalin	ting ##alin	ting ##alin	_ting alin	2/2/2
790	tingkat	tingkat	tingkat	_tingkat	1/1/1
791	tingkatina	tingkat ##ina	tingkat ##ina	_tingkat ina	2/2/2
792	titang	titi ##ang	tit ##ian ##g	_titi ang	2/3/2
793	tiwas	ti ##was	ti ##was	_ti was	2/2/2
794	tiyang	ti ##yang	ti ##yang	_tiyang	2/2/1
795	tolih	tol ##ih	tol ##ih	_to lih	2/2/2
796	tompok	tom ##pl ##ok	tom ##plo ##k	_to mp lok	3/3/3
797	tragedi	tragedi	tra ##ged ##i	_tragedi	1/3/1
798	trauma	trauma	trauma	_trauma	1/1/1
799	tresna	tr ##es ##na	tres ##na	_tres na	3/2/2
800	trima	trima	trim ##a	_tri ma	1/2/2
801	truna	tr ##una	tru ##na	_tru na	2/2/2
802	tsing aman	ts ##ing aman	ts ##ing aman	_t sing _aman	3/3/3
803	tuduha	tud ##uha	tud ##uh ##a	_tu du ha	2/3/3
804	tuduhan	tuduhan	tud ##uhan	_tuduhan	1/2/1
805	tulung	tul ##ung	tul ##ung	_tu lung	2/2/2
806	tumpas	tump ##as	tum ##pas	_tum pas	2/2/2
807	tunda	tunda	tun ##da	_tund a	1/2/2
808	tusing	tus ##ing	tus ##ing	_tu sing	2/2/2
809	tusing aman	tus ##ing aman	tus ##ing aman	_tu sing _aman	3/3/3
810	tusing runguanga	tus ##ing run ##gu ##anga	tus ##ing rung ##uang ##a	_tu sing _rung u anga	5/5/5
811	tuun	tu ##un	tu ##un	_tu un	2/2/2
812	tuunang	tu ##unan ##g	tu ##unan ##g	_tu un ang	3/3/3
813	ulung	ulu ##ng	ulu ##ng	_u lung	2/2/2
814	undag	und ##ag	und ##ag	_un dag	2/2/2
815	unduk	und ##uk	und ##uk	_und uk	2/2/2
816	upaya	upaya	upaya	_upaya	1/1/1
817	usan	usa ##n	usan	_ usan	2/1/2
818	uyak	u ##yak	uy ##ak	_u yak	2/2/2
819	wajib	wajib	wa ##ji ##b	_wajib	1/3/1
820	wanen	wan ##en	wan ##en	_wan en	2/2/2
821	wangde	wang ##de	wang ##de	_wang de	2/2/2
822	waspada	waspada	wasp ##ada	_was pada	1/2/2
823	wates	wat ##es	wat ##es	_wat es	2/2/2

No	Kata Bali	IndoBERT	mBERT	XLM-RoBERTa	Jumlah Token (l/m/X)
824	wenang	wen ##ang	wen ##ang	_ wenang	2/2/2
825	wikan	wika ##n	wi ##kan	_wika n	2/2/2
826	wong	wong	wong	_wong	1/1/1
827	wow	wow	wow	_wow	1/1/1
828	yang	yang	yang	_yang	1/1/1
829	yuda	yud ##a	yu ##da	_y uda	2/2/2
830	yudha	yudha	yu ##dha	_y udha	1/2/2



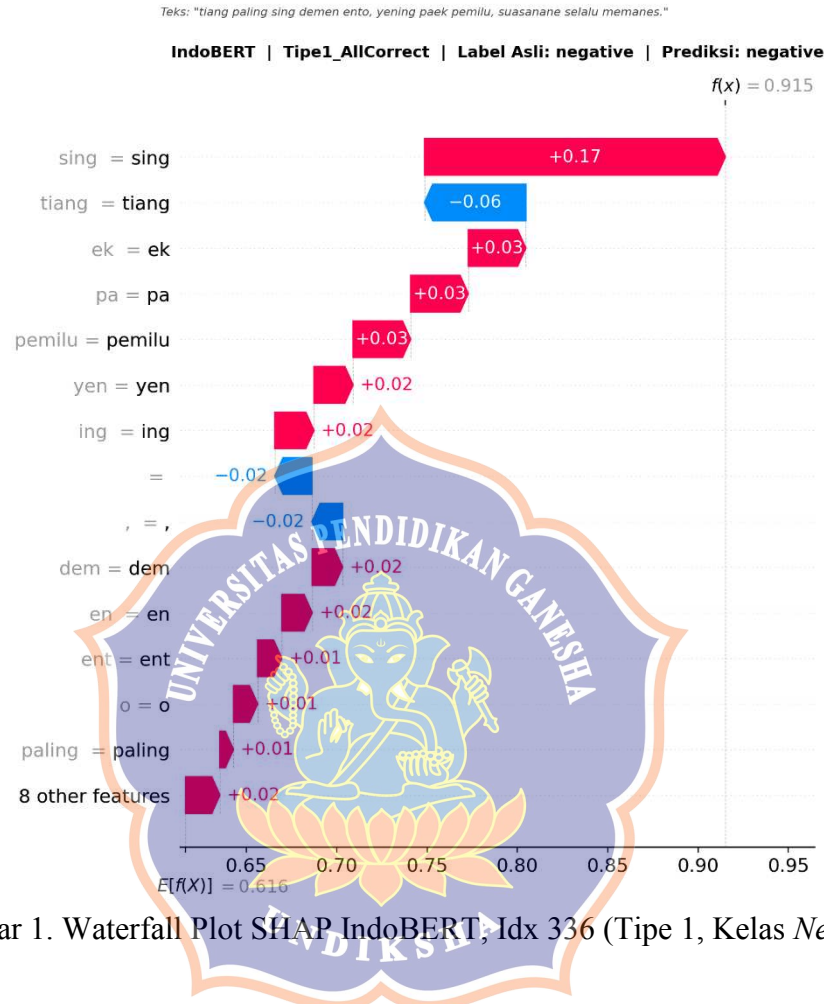
Lampiran 2. *Waterfall Plot* SHAP Lokal

Lampiran ini menyajikan seluruh *waterfall plot* SHAP lokal untuk dua belas sampel pada ketiga model. Setiap *waterfall plot* menampilkan token dalam bentuk yang telah dinormalisasi tanpa penanda *subword* ('##' pada WordPiece dan '_' pada SentencePiece) untuk menjaga keterbacaan. Token yang tampak sebagai potongan kata, misalnya 'nga' atau 'jeng', merupakan sub-token hasil pemecahan kata oleh *tokenizer* masing-masing model. Nilai pada setiap plot menunjukkan kontribusi SHAP token terhadap probabilitas kelas target, dengan bar berwarna merah menandakan dorongan menaikkan probabilitas dan bar biru menandakan dorongan menurunkan, dihitung dari nilai dasar (*base value*) menuju nilai prediksi akhir $f(x)$. Bagian A memuat plot untuk model IndoBERT.



A. Model IndoBERT

Tipe 1 (*All Correct*)

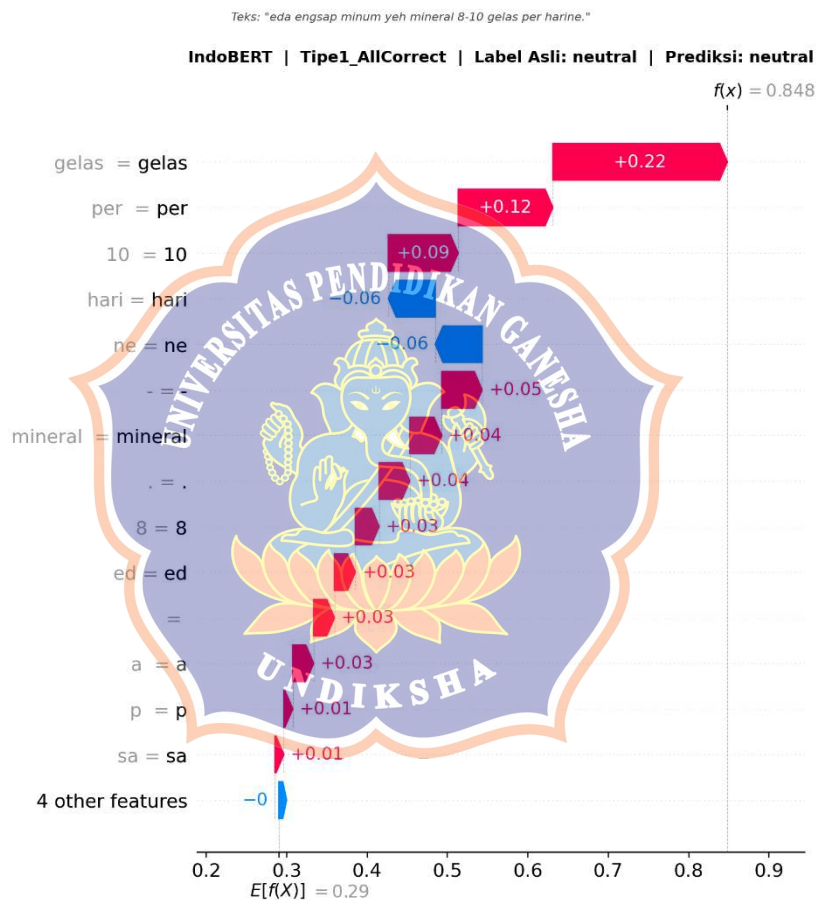


Gambar 1. Waterfall Plot SHAP IndoBERT, Idx 336 (Tipe 1, Kelas *Negative*)

Kalimat (Idx 336): “*tiang paling sing demen ento, yening paek pemilu, suasanane selalu memanes.*”

Pada Gambar 1, model IndoBERT mengevaluasi data Idx 336 dengan label aktual *negative* dan memprediksinya secara tepat sebagai *negative*. Penalaran model dimulai dari nilai dasar $E[f(x)] = 0,616$ dan mencapai nilai prediksi akhir $f(x) = 0,915$. Penguat utama prediksi negatif adalah token *sing* (+0,17). Beberapa kata bahasa Bali terpecah menjadi sub-token, seperti *paek* menjadi *pa* (+0,03) dan *ek* (+0,03), *yening* menjadi *yen* (+0,02) dan *ing* (+0,02), serta *demen* menjadi *dem* (+0,02) dan *en* (+0,02). Meskipun *demen* secara mandiri bermakna positif (suka),

kehadiran token negasi *sing* di awalnya secara kontekstual membalik arah sentimen menjadi negatif. Akumulasi kontribusi positif sub-token tersebut, ditambah gabungan fitur lainnya (*8 other features*, +0,02), mengatasi reduksi kecil dari token *tiang* (-0,06), sehingga mendorong probabilitas akhir mendekati satu untuk kelas negatif.

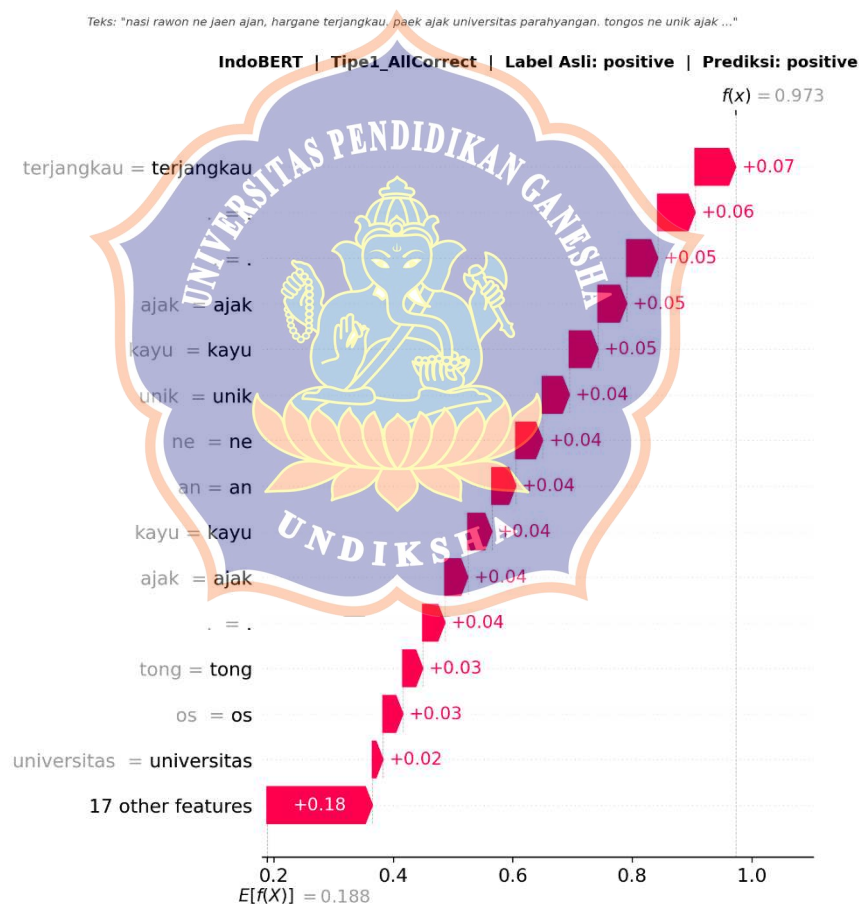


Gambar 2. Waterfall Plot SHAP IndoBERT, Idx 300 (Tipe 1, Kelas *Neutral*)

Kalimat (Idx 300): "eda engsap minum yeh mineral 8-10 gelas per harine."

Pada Gambar 2, IndoBERT mengevaluasi data Idx 300 dengan label aktual *neutral* dan memprediksinya secara tepat sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,29$ dan mencapai $f(x) = 0,848$. Teks ini bersifat informatif berupa saran kesehatan, dan model mengenalinya melalui kemunculan satuan

ukuran dan kuantitas numerik. Kontributor terbesar ke arah netral adalah token *gelas* (+0,22), disusul penunjuk satuan *per* (+0,12) dan angka *10* (+0,09). Kata larangan *eda* (jangan) terpecah menjadi *ed* (+0,03) dan *a* (+0,03) yang secara akumulatif tetap mendukung klasifikasi netral. Penurunan sempat dipicu token *hari* (-0,06) dan *ne* (-0,06), namun dorongan dari kata objektif seperti *mineral* (+0,04) dan tanda hubung (+0,05) menjaga akumulasi akhir tetap mengarah ke sentimen netral.



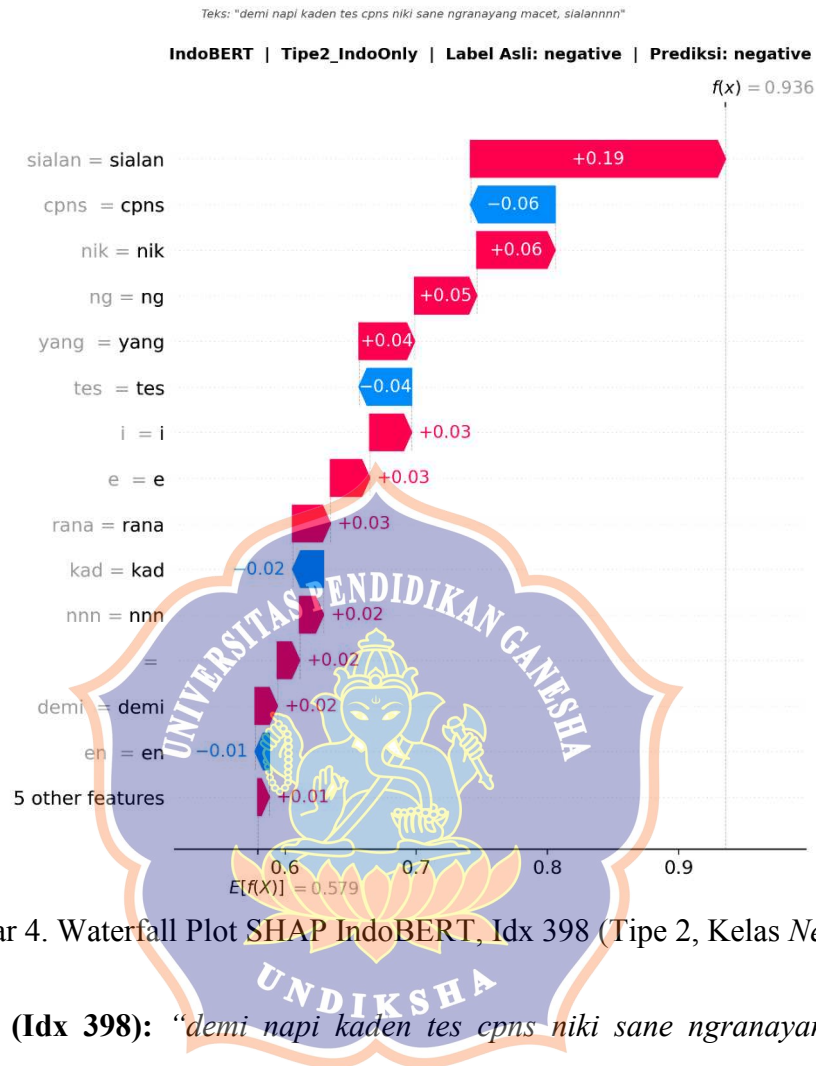
Gambar 3. Waterfall Plot SHAP IndoBERT, Idx 272 (Tipe 1, Kelas *Positive*)

Kalimat (Idx 272): "nasi rawon ne jaen ajan, hargane terjangkau. paek ajak universitas parahyangan. tongos ne unik ajak kayu kayuan."

Pada Gambar 3, IndoBERT mengevaluasi data Idx 272 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,188$ dan mencapai $f(x) = 0,973$. Model mengenali kombinasi kata sifat positif lintas bahasa Indonesia dan Bali. Kontribusi penarik terbesar diberikan kata *terjangkau* (+0,07), kata *unik* (+0,04), dan kata hubung *ajak* (+0,05). Kata penunjuk tempat *tongos* terpecah menjadi *tong* (+0,03) dan *os* (+0,03) yang keduanya berkontribusi positif. Kekuatan utama prediksi terletak pada akumulasi gabungan fitur pendukung lainnya (17 *other features*, +0,18). Tanpa kontribusi negatif yang signifikan, nilai SHAP naik konsisten hingga mencapai probabilitas akhir untuk sentimen positif.



Tipe 2 (IndoBERT Only Correct)

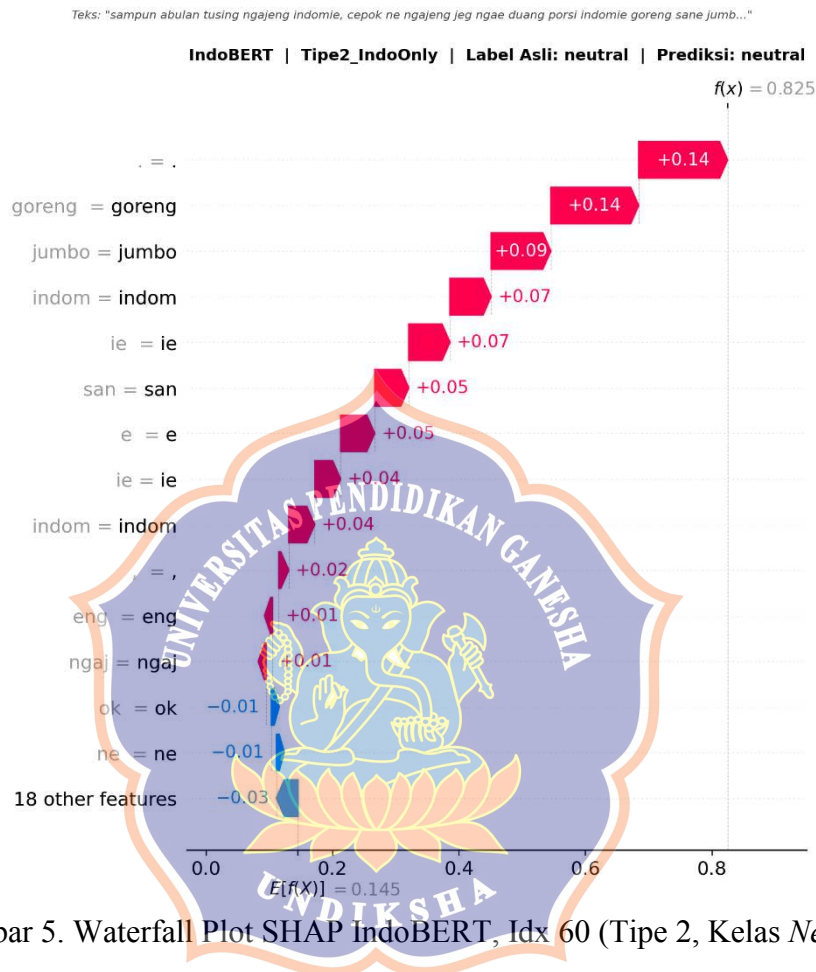


Gambar 4. Waterfall Plot SHAP IndoBERT, Idx 398 (Tipe 2, Kelas *Negative*)

Kalimat (Idx 398): “demi napi kaden tes cpns niki sane ngranayang macet, sialannnn”

Pada Gambar 4, IndoBERT mengevaluasi data Idx 398 dengan label aktual *negative* dan memprediksinya secara tepat sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,579$ dan mencapai $f(x) = 0,936$. Keberhasilan ini didorong secara dominan oleh leksikon umpatan eksplisit *sialan* (+0,19). Token pecahan dari kata *ngranayang*, yaitu *ng* (+0,05) dan *rana* (+0,03), serta potongan *niki* menjadi *nik* (+0,06) memberikan dorongan tambahan ke arah kelas negatif. Meskipun

terdapat fitur penahan seperti *cpns* (-0,06) dan *tes* (-0,04), muatan emosi negatif di akhir kalimat tetap dominan sehingga prediksi tepat pada kelas negatif.

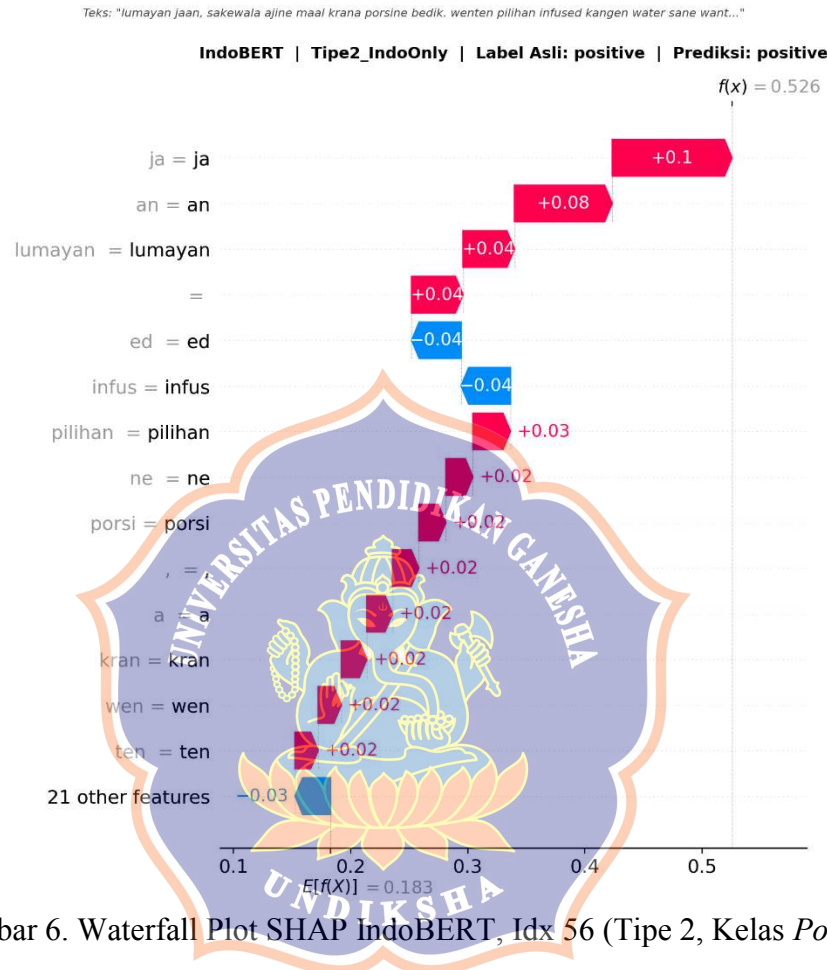


Gambar 5. Waterfall Plot SHAP IndoBERT, Idx 60 (Tipe 2, Kelas *Neutral*)

Kalimat (Idx 60): “sampun abulan tusing ngajeng indomie, cepok ne ngajeng jeg ngae duang porsi indomie goreng sane jumbo.”

Pada Gambar 5, IndoBERT mengevaluasi data Idx 60 dengan label aktual *neutral* dan memprediksinya secara tepat sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,145$ dan mencapai $f(x) = 0,825$. Model mengenali sifat informatif dan objektif teks ulasan makanan ini. Fitur penanda utama ke arah netral adalah tanda baca titik (+0,14) dan kata komoditas objektif seperti *goreng* (+0,14), ukuran *jumbo* (+0,09), serta potongan nama produk *indom* (+0,07 dan +0,04).

Dominasi fitur faktual ini menaikkan nilai SHAP signifikan, mengabaikan reduksi minor dari token *ok* (-0,01) dan *ne* (-0,01), sehingga prediksi tepat pada kelas netral.



Gambar 6. Waterfall Plot SHAP IndoBERT, Idx 56 (Tipe 2, Kelas *Positive*)

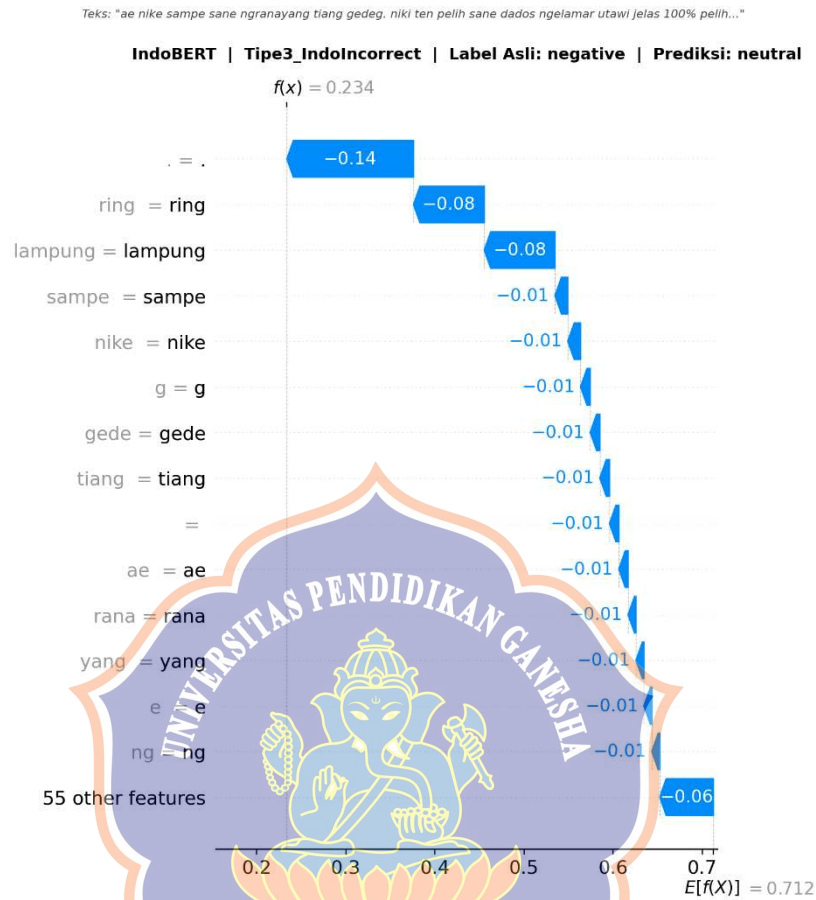
Kalimat (Idx 56): "lumayan jaan, sakewala ajine maal krana porsine bedik. wenten pilihan infused kangen water sane wantah antuk 4 orang"

Pada Gambar 6, IndoBERT mengevaluasi data Idx 56 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,183$ dan mencapai $f(x) = 0,526$. Keputusan ini ditopang oleh pecahan kata penguat sentimen positif *jaan* (enak), yang terurai menjadi *ja* (+0,1) dan *an* (+0,08). Leksikon positif lain seperti *lumayan* (+0,04) dan kata *pilihan* (+0,03) turut berkontribusi. Meskipun muncul token penurun seperti *ed* (-0,04) dan

infus (-0,04) yang merepresentasikan keluhan harga atau porsi, bobot kepuasan rasa pada paruh awal kalimat mempertahankan nilai akhir pada zona klasifikasi positif.



Tipe 3 (IndoBERT Incorrect)

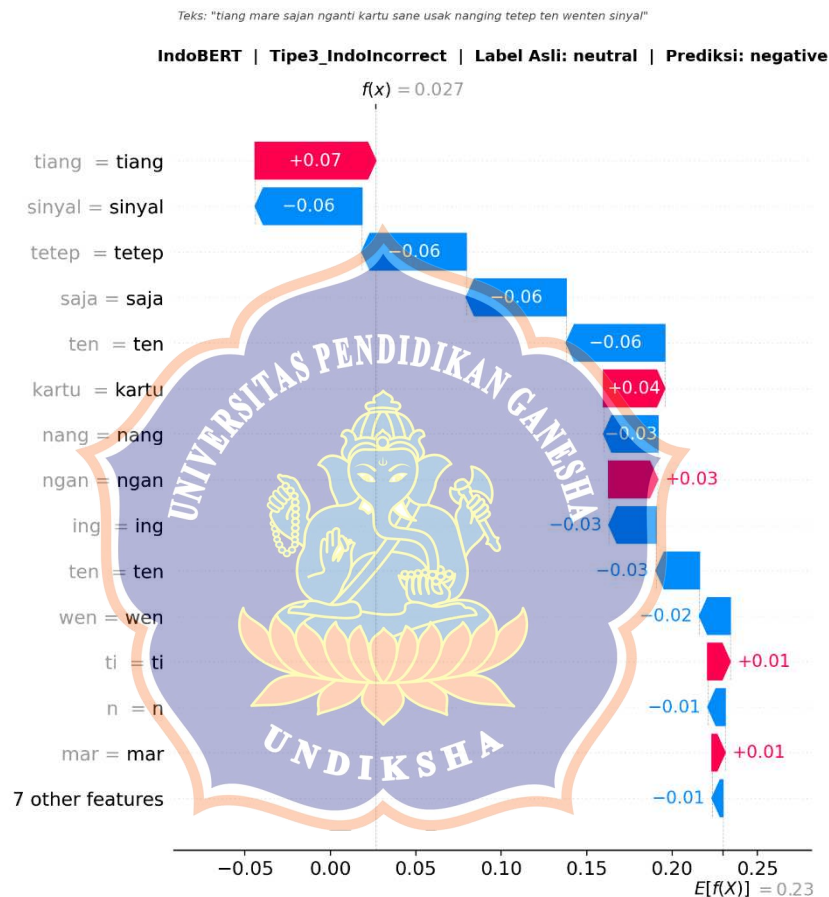


Gambar 7. Waterfall Plot SHAP IndoBERT, Idx 27 (Tipe 3, Kelas Negative)

Kalimat (Idx 27): "ae nike sampe sane ngranayang tiang gedeg. niki ten pelih sane dados ngelamar utawi jelas 100% pelih ring pihak sane di duur sane engken kaden cara ipun makoordinasi. sesek bane tiang yadiastun kasus puniki timpal tiang sane ring lampung."

Pada Gambar 7, IndoBERT mengevaluasi data Idx 27 dengan label aktual *negative*, tetapi memprediksinya sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,712$ dan turun ke $f(x) = 0,234$. Pola atribusi mengindikasikan bahwa penurunan dipicu oleh tanda baca titik (-0,14) serta token *ring* (-0,08) dan *lampung* (-0,08) yang kurang relevan dengan muatan emosi kalimat. Fragmentasi kata penanda emosi negatif seperti *ngranayang* (menjadi *ng*, *rana*, *yang*, masing-masing

sekitar -0,01) dan *gedeg* (marah, menjadi *g* dan *gede*, masing-masing sekitar -0,01) memperlemah deteksi sentimen. Akumulasi penurunan ini menggeser prediksi ke kelas netral, yang mengindikasikan bahwa muatan kemarahan pada teks tidak terdeteksi secara memadai oleh model.

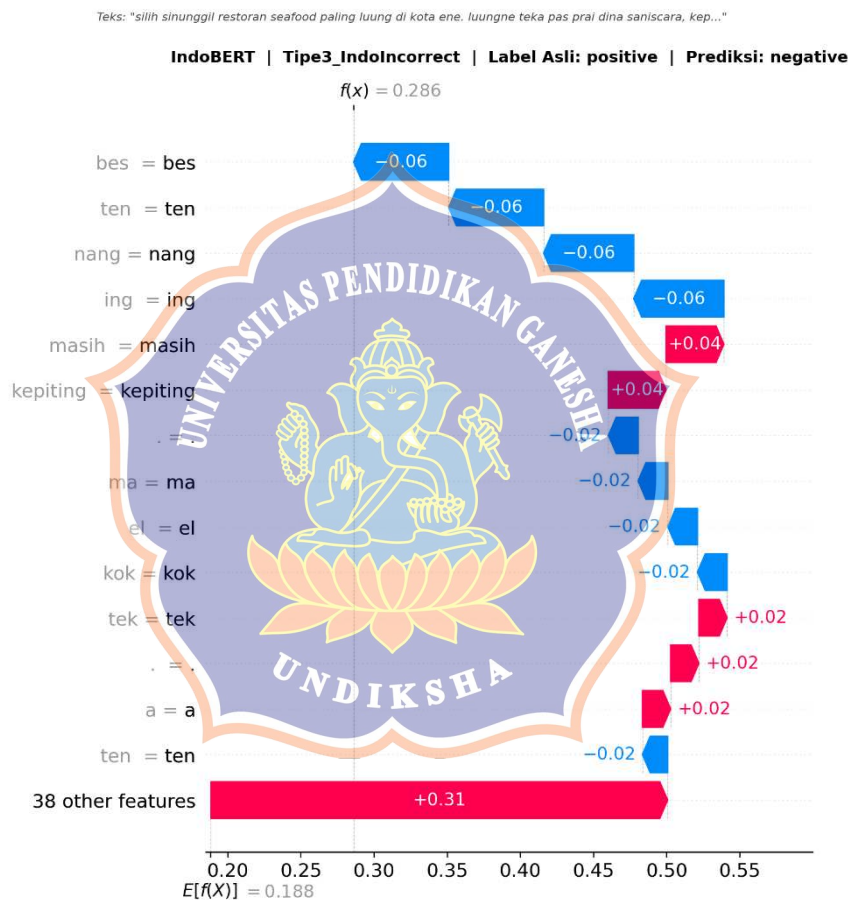


Gambar 8. Waterfall Plot SHAP IndoBERT, Idx 149 (Tipe 3, Kelas *Neutral*)

Kalimat (Idx 149): “*tiang mare sajan nganti kartu sane usak nanging tetep ten wenten sinyal*”

Pada Gambar 8, IndoBERT mengevaluasi data Idx 149 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,23$ dan turun ke $f(x) = 0,027$. Pola atribusi mengindikasikan bahwa leksikon yang lazim berasosiasi dengan keluhan fasilitas mendorong prediksi ke

arah negatif. Token *sinyal* (-0,06), *tetep* (-0,06), kata penolak *ten* (-0,06), serta fragmentasi konjungsi *nanging* menjadi *nang* (-0,03) dan *ing* (-0,03) memangkas probabilitas kelas netral. Meskipun token *tiang* (+0,07) dan *kartu* (+0,04) memberi dorongan positif, dominasi kelompok token keluhan menekan nilai akhir mendekati nol, sehingga prediksi bergeser ke sentimen negatif.



Gambar 9. Waterfall Plot SHAP IndoBERT, Idx 139 (Tipe 3, Kelas *Positive*)

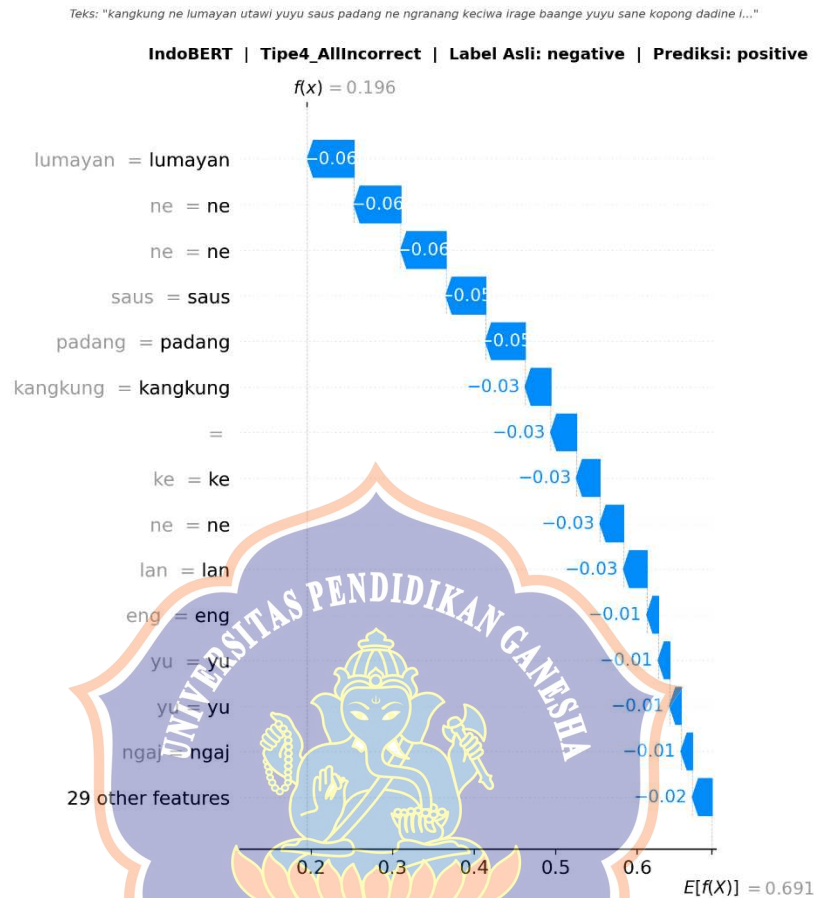
Kalimat (Idx 139): "silih sinunggil restoran seafood paling luung di kota ene. luungne teka pas prai dina saniscara, kepiting masih luung mara teka. pengarga bersaing, ten mudah mula nanging ten bes mael kok."

Pada Gambar 9, IndoBERT mengevaluasi data Idx 139 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar

$E[f(x)] = 0,188$ dan mencapai $f(x) = 0,286$. Model sempat memperoleh dorongan naik dari gabungan fitur (*38 other features*, +0,31) yang menangkap frasa pujian *paling luung* (paling bagus). Namun, kemunculan token penurun di paruh akhir kalimat menahan laju nilai SHAP, yaitu *bes* (-0,06), kata penolak *ten* (-0,06), serta fragmentasi *nanging* menjadi *nang* (-0,06) dan *ing* (-0,06). Pola atribusi mengindikasikan adanya ambiguitas leksikal, di mana rangkaian kata penolak dan konjungsi pada teks pujian ini cenderung ditafsirkan model sebagai sanggahan, sehingga nilai akhir tertahan di bawah ambang kelas positif.



Tipe 4 (*All Incorrect*)

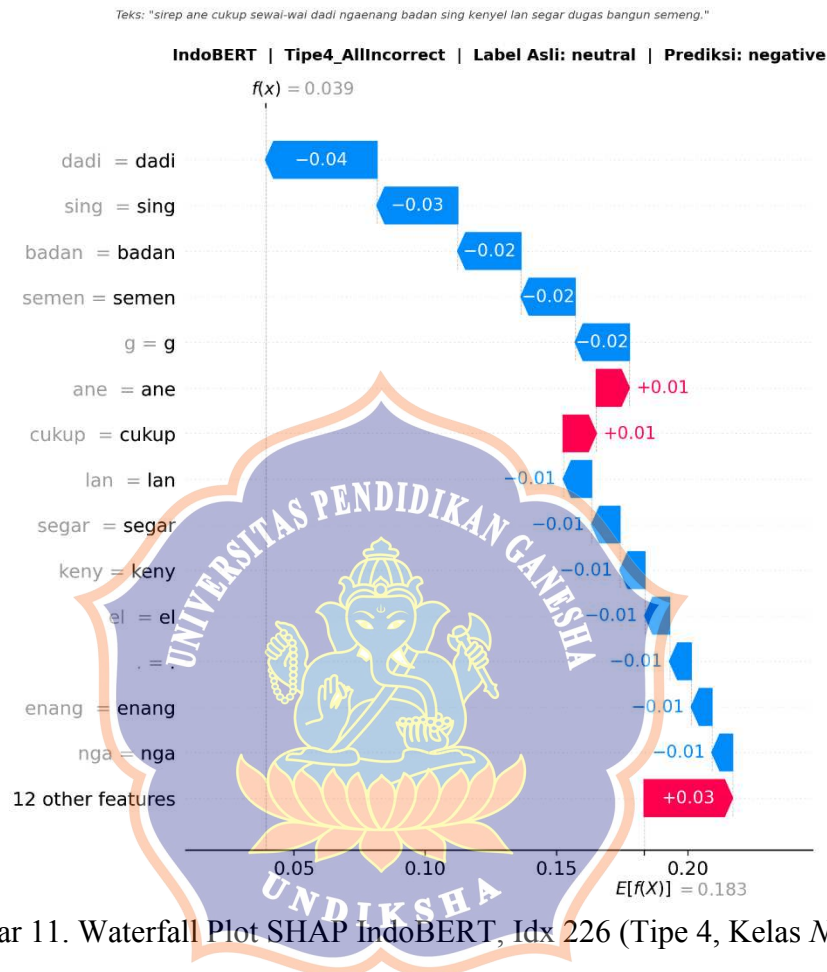


Gambar 10. Waterfall Plot SHAP IndoBERT, Idx 2 (Tipe 4, Kelas *Negative*)

Kalimat (Idx 2): "kangkung ne lumayan utawi yuyu saus padang ne ngranang keciwa irage baange yuyu sane kopong dadine irage ten ngajeng yuyu ne lan ke uliang."

Pada Gambar 10, IndoBERT mengevaluasi data Idx 2 dengan label aktual *negative*, tetapi memprediksinya sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,691$ dan turun ke $f(x) = 0,196$. Pola atribusi mengindikasikan bahwa paruh awal kalimat yang bernada netral-positif terkait menu makanan memberi bobot penurunan terhadap kelas negatif. Token *lumayan* (-0,06), partikel *ne* (-0,06), serta nama hidangan *saus* (-0,05) dan *padang* (-0,05) berkontribusi menurunkan probabilitas kelas aktual. Akibatnya, leksikon negatif di akhir kalimat seperti

keciwa dan *kopong* kehilangan dominasinya, sehingga nilai akhir bergeser ke arah sentimen positif.

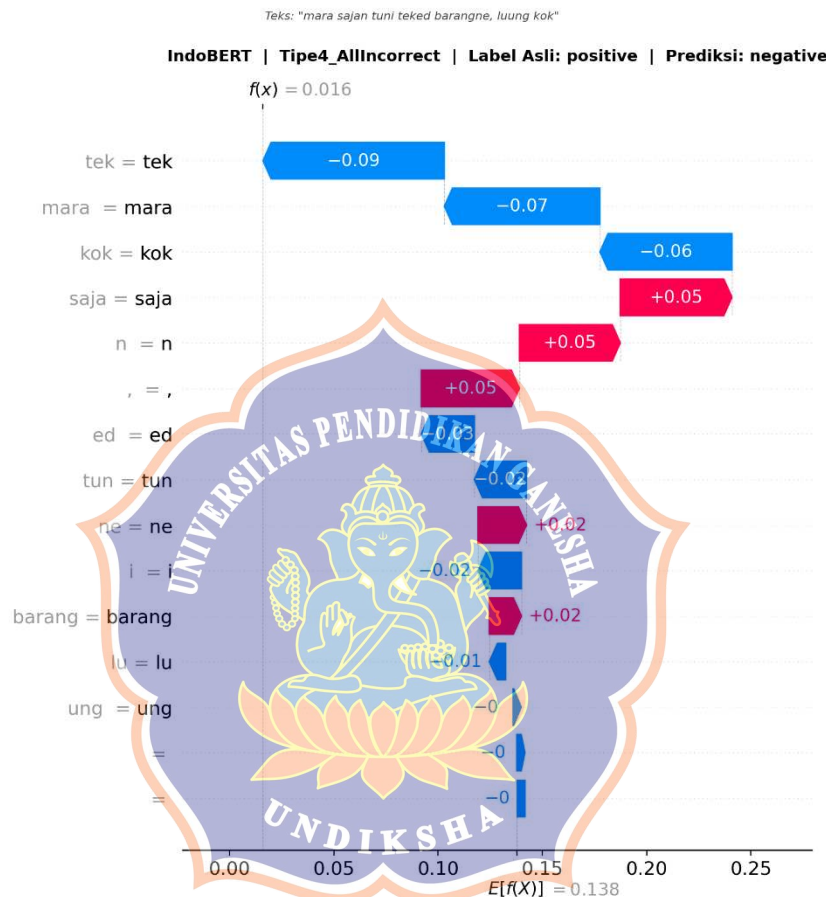


Gambar 11. Waterfall Plot SHAP IndoBERT, Idx 226 (Tipe 4, Kelas *Neutral*)

Kalimat (Idx 226): “sirep ane cukup sewai-wai dadi ngaenang badan sing kenyel lan segar dugas bangun semeng.”

Pada Gambar 11, IndoBERT mengevaluasi data Idx 226 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,183$ dan turun ke $f(x) = 0,039$. Pola atribusi mengindikasikan kesulitan model menafsirkan struktur negasi. Kata penolak *sing* (-0,03), yang dalam konteks berpasangan dengan *kenyel* (sehingga bermakna tidak lelah), cenderung ditafsirkan sebagai penanda negatif. Ditambah kontribusi penurunan dari *dadi* (-0,04), *badan* (-

0,02), serta pecahan *semeng* menjadi *semen* (-0,02) dan *g* (-0,02), akumulasi nilai SHAP merosot dan mengeliminasi pengaruh positif dari kata *cukup* (+0,01), sehingga prediksi bergeser ke sentimen negatif.



Gambar 12. Waterfall Plot SHAP IndoBERT, Idx 363 (Tipe 4, Kelas *Positive*)

Kalimat (Idx 363): “*mara sajan tuni teked barangne, luung kok*”

Pada Gambar 12, IndoBERT mengevaluasi data Idx 363 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,138$ dan turun ke $f(x) = 0,016$. Pola atribusi mengindikasikan bahwa fragmentasi kata turut melemahkan deteksi sentimen positif. Token pecahan dari kata *teked*, yaitu *tek*, menjadi penurun terbesar (-0,09), disusul *mara* (-0,07) dan *kok* (-0,06). Kata penguat *sajan* terpecah menjadi *saja* (+0,05) dan *n* (+0,05) yang

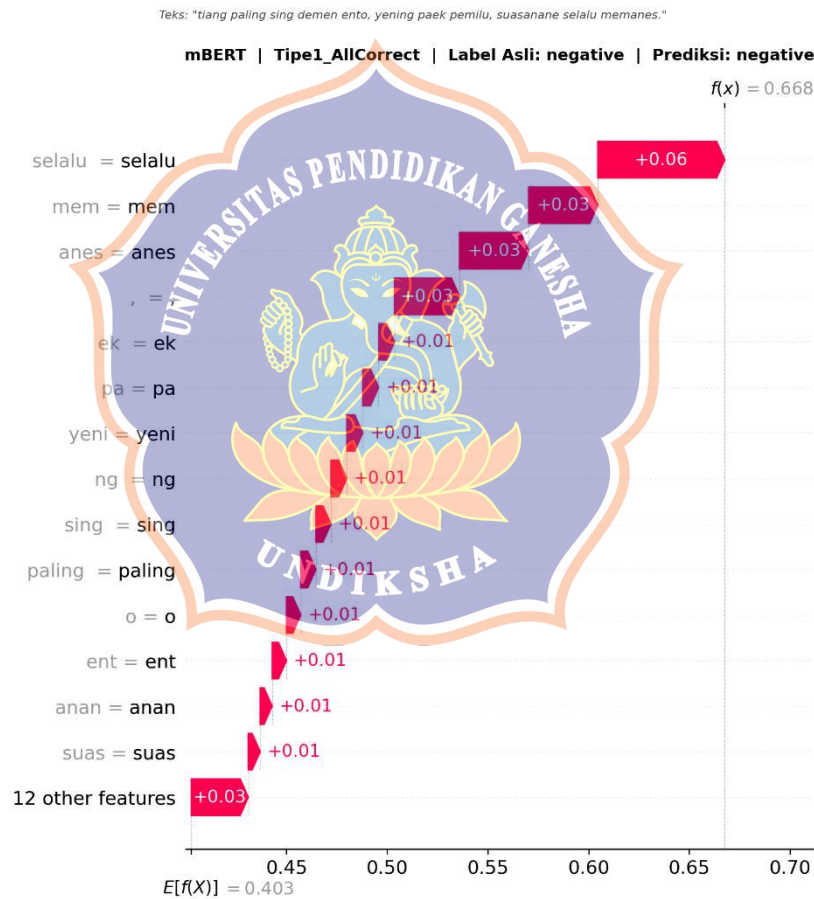
mendorong positif, tetapi dampaknya teranulir oleh rentetan token penurun. Kata kunci sentimen positif *luung* (bagus) terpecah menjadi *lu* (-0,01) dan *ung* (-0,00), sehingga ikut menahan pergerakan probabilitas di bawah ambang kelas positif.



B. Model mBERT

Bagian ini memuat seluruh *waterfall plot* SHAP lokal untuk model mBERT pada dua belas sampel. Ketentuan pembacaan plot sama dengan Bagian A: token ditampilkan tanpa penanda *subword*, bar merah menandakan dorongan menaikkan probabilitas, dan bar biru menandakan dorongan menurunkan, dihitung dari nilai dasar (*base value*) menuju nilai prediksi akhir $f(x)$.

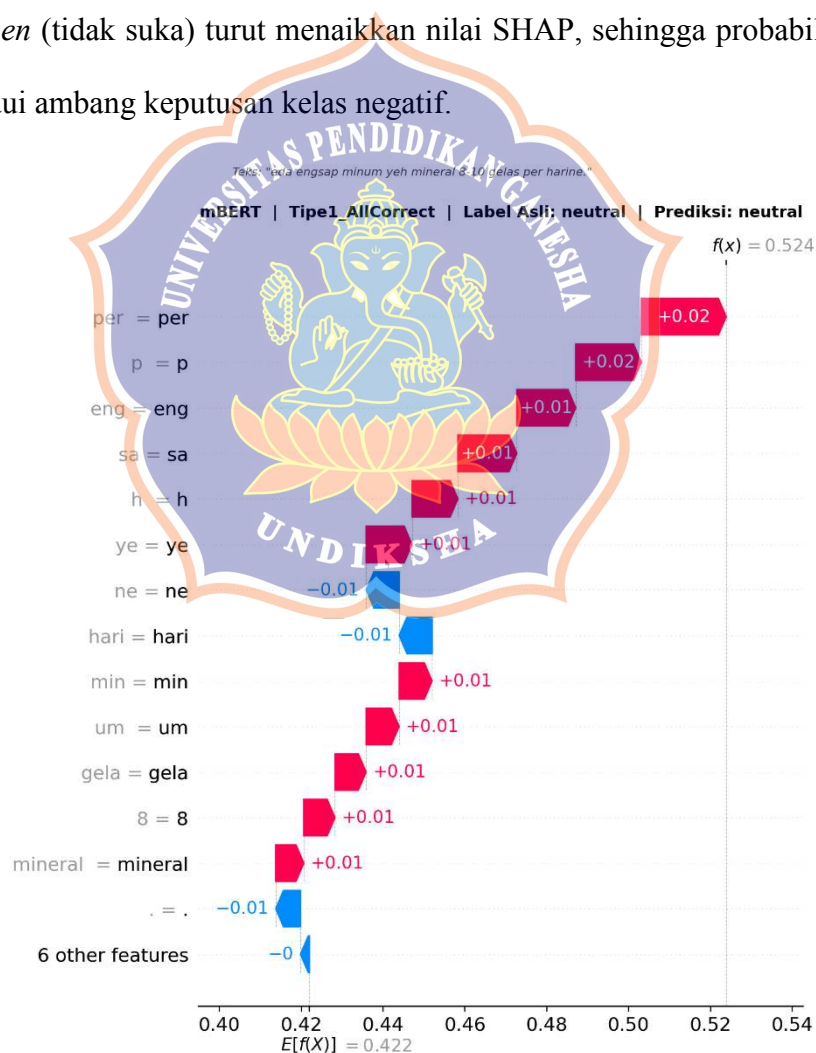
Tipe 1 (*All Correct*)



Gambar 13. Waterfall Plot SHAP mBERT, Idx 336 (Tipe 1, Kelas *Negative*)

Kalimat (Idx 336): “*tiang paling sing demen ento, yening paek pemilu, suasanane selalu memanes.*”

Pada Gambar 13, mBERT mengevaluasi data Idx 336 dengan label aktual *negative* dan memprediksinya secara tepat sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,403$ dan mencapai $f(x) = 0,668$. Keberhasilan ini didorong oleh pengenalan kata keterangan dan leksikon bermuatan negatif. Token *selalu* memberi kontribusi positif terbesar (+0,06). Penegasan emosi negatif diperkuat oleh pecahan kata *memanes* (memanas) menjadi *mem* (+0,03) dan *anes* (+0,03). Kehadiran kata penolak *sing* (+0,01) dan penguat *paling* (+0,01) yang merujuk frasa *sing demen* (tidak suka) turut menaikkan nilai SHAP, sehingga probabilitas akhir melampaui ambang keputusan kelas negatif.

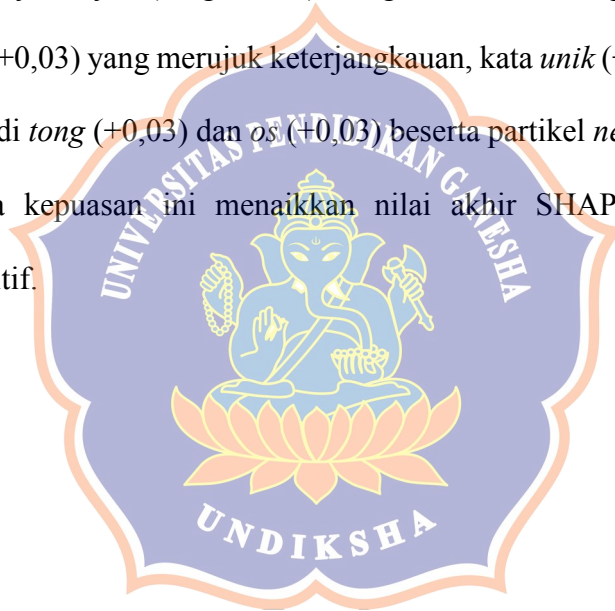


Gambar 14. Waterfall Plot SHAP mBERT, Idx 300 (Tipe 1, Kelas *Neutral*)

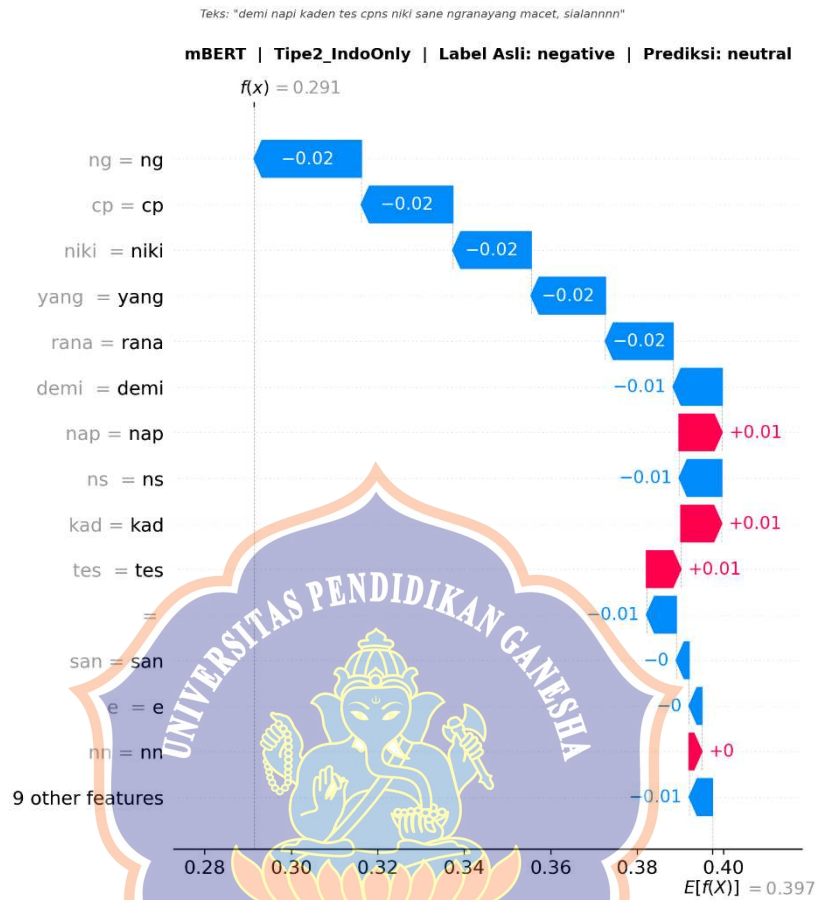
Kalimat (Idx 300): “eda engsap minum yeh mineral 8-10 gelas per harine.”

Kalimat (Idx 272): “nasi rawon ne jaen ajan, hargane terjangkau. paek ajak universitas parahyangan. tongos ne unik ajak kayu kayuan.”

Pada Gambar 15, mBERT mengevaluasi data Idx 272 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,178$ dan mencapai $f(x) = 0,755$. Kenaikan probabilitas yang signifikan ini dipicu kemunculan kata sifat bermakna pujian dan kepuasan. Fitur kolektif 20 *other features* menyumbang +0,13, yang merepresentasikan frasa penguat positif *jaen ajan* (sangat enak). Pengenalan sentimen positif diperkuat oleh token *harga* (+0,03) yang merujuk keterjangkauan, kata *unik* (+0,02), serta pecahan *tongos* menjadi *tong* (+0,03) dan *os* (+0,03) beserta partikel *ne* (+0,03). Kombinasi fitur penanda kepuasan ini menaikkan nilai akhir SHAP menuju klasifikasi sentimen positif.



Tipe 2 (IndoBERT Only Correct)

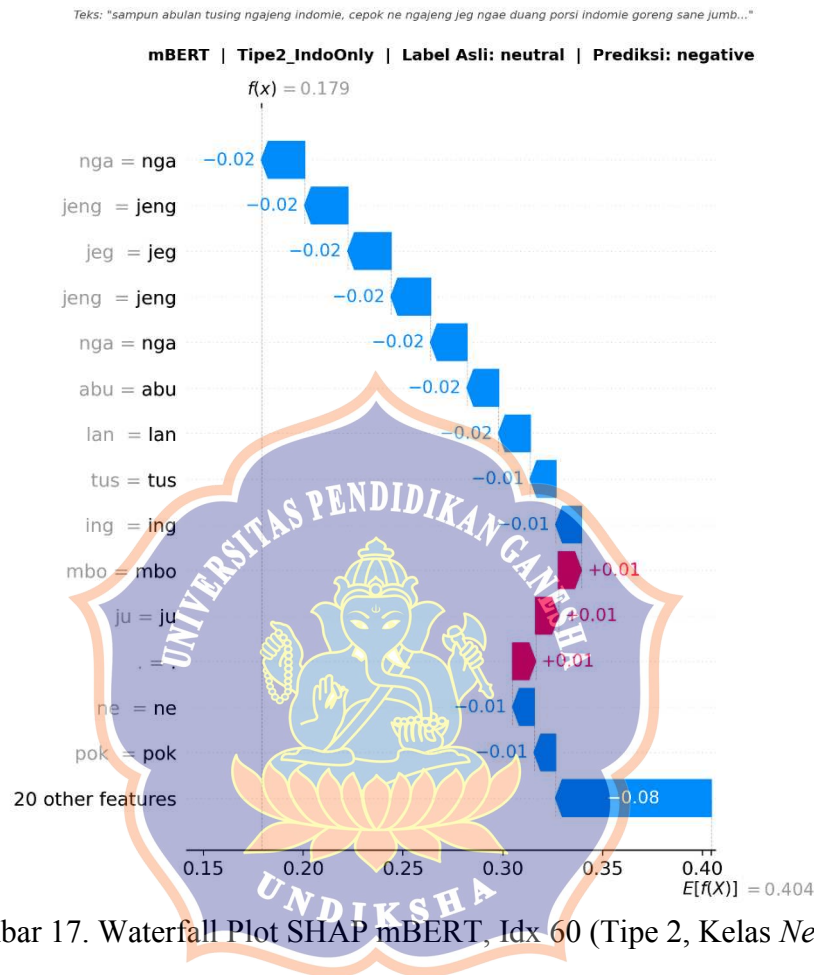


Gambar 16. Waterfall Plot SHAP mBERT, Idx 398 (Tipe 2, Kelas *Negative*)

Kalimat (Idx 398): “demi napi kaden tes cpns niki sane ngranayang macet, sialannnn”

Pada Gambar 16, mBERT mengevaluasi data Idx 398 dengan label aktual *negative*, tetapi memprediksinya sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,397$ dan turun ke $f(x) = 0,291$. Pola atribusi mengindikasikan bahwa pemecahan kata pada kosakata bahasa Bali dan singkatan melemahkan deteksi negatif. Kata *ngranayang* terpecah menjadi *ng* (-0,02), *rana* (-0,02), dan *yang* (-0,02), sementara *cpns* menjadi *cp* (-0,02) dan *ns* (-0,01). Sub-token ini menurunkan probabilitas kelas aktual. Akibatnya, kata bermuatan emosi negatif seperti

sialannnn dan *macet* kehilangan dominasinya, sehingga nilai akhir tertahan di bawah ambang klasifikasi yang tepat.

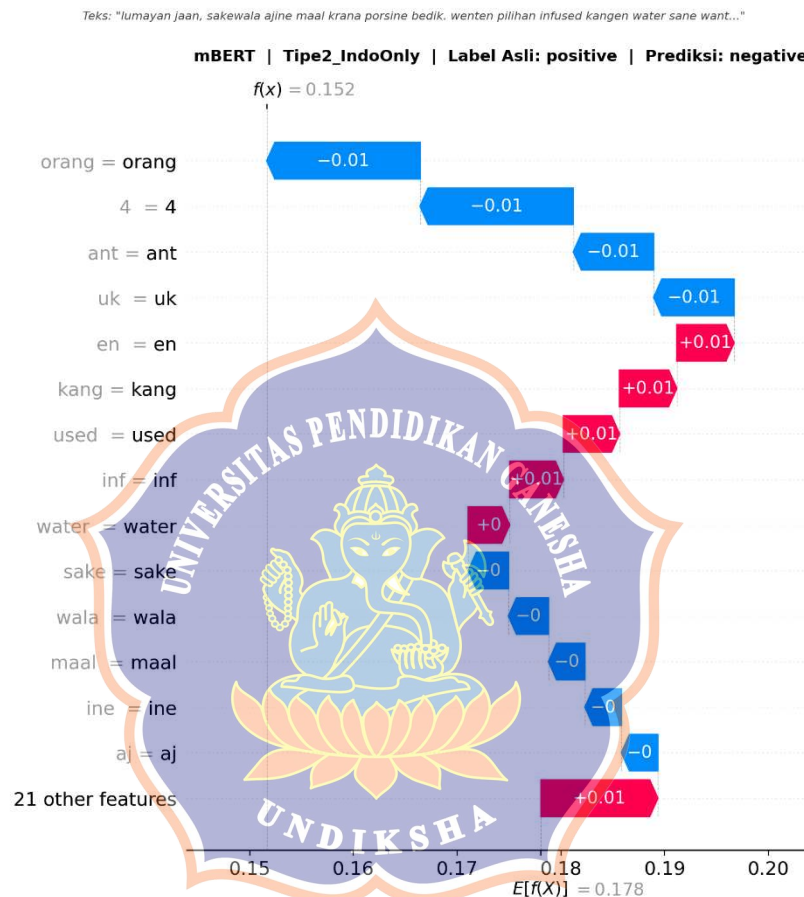


Gambar 17. Waterfall Plot SHAP mBERT, Idx 60 (Tipe 2, Kelas *Neutral*)

Kalimat (Idx 60): “*sampun abulan tusing ngajeng indomie, cepok ne ngajeng jeg ngae duang porsi indomie goreng sane jumbo.*”

Pada Gambar 17, mBERT mengevaluasi data Idx 60 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,404$ dan turun ke $f(x) = 0,179$. Pola atribusi mengindikasikan akumulasi penurunan dari kelompok *20 other features* (-0,08) serta fragmentasi pada kosakata bahasa Bali. Kata *ngajeng* (makan) yang muncul dua kali terpecah konsisten menjadi *nga* (-0,02) dan *jeng* (-0,02). Penurunan diperparah sub-token *jeg* (-0,02),

pemecahan *abulan* menjadi *abu* (-0,02) dan *lan* (-0,02), serta *tusing* menjadi *tus* (-0,01) dan *ing* (-0,01). Sebaran nilai penurun ini menekan probabilitas kelas netral, sehingga prediksi bergeser ke kelas negatif.



Gambar 18. Waterfall Plot SHAP mBERT, Idx 56 (Tipe 2, Kelas Positive)

Kalimat (Idx 56): "lumayan jaan, sakewala ajine maal krana porsine bedik. wenten pilihan infused kangen water sane wantah antuk 4 orang"

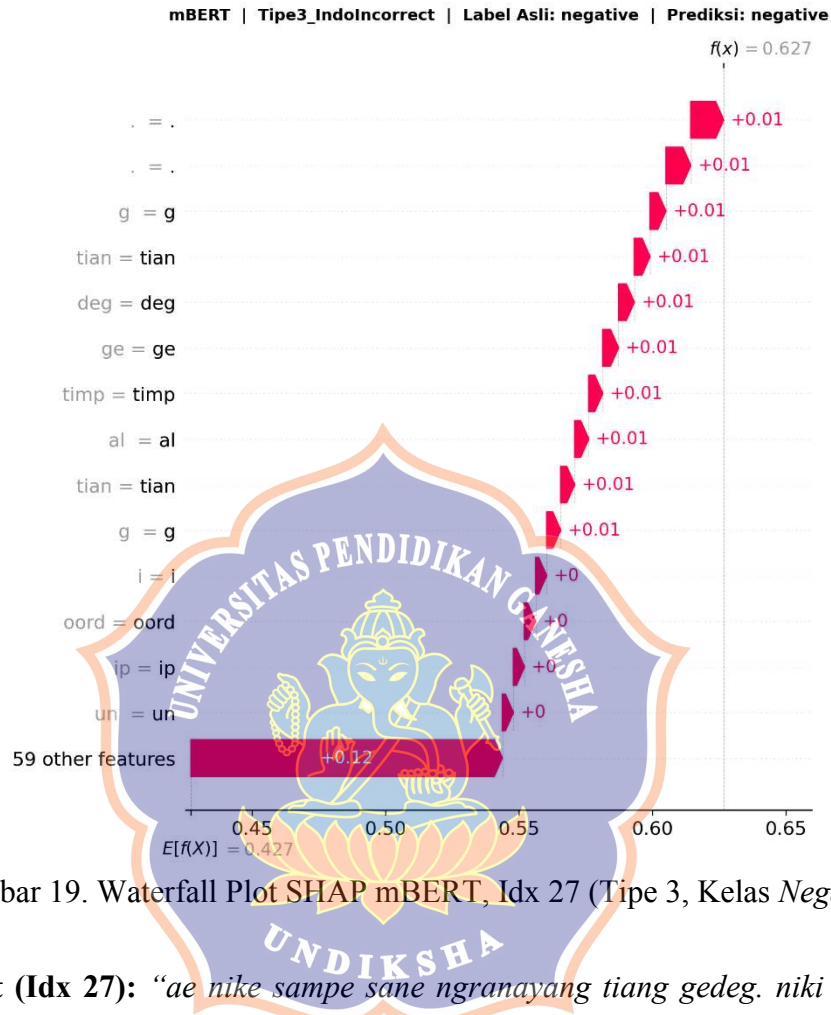
Pada Gambar 18, mBERT mengevaluasi data Idx 56 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,178$ dan turun ke $f(x) = 0,152$. Pola atribusi mengindikasikan bahwa klausa transisi bernada keluhan ekonomi (*sakewala ajine maal krana porsine bedik*) memotong sentimen positif utama. Meskipun istilah asing *infused kangen water*

terpecah menjadi *inf* (+0,01), *used* (+0,01), *kang* (+0,01), dan *en* (+0,01) yang mendorong positif, kontribusi itu teranulir oleh token penurun seperti *orang* (-0,01), *4* (-0,01), *ant* (-0,01), dan *uk* (-0,01). Bobot frasa penguat *lumayan jaan* di awal kalimat tidak dapat dipertahankan, sehingga teks salah diklasifikasikan sebagai negatif.



Tipe 3 (IndoBERT *Incorrect*)

Teks: "ae nike sampe sane ngranayang tiang gedeg, niki ten pelih sane dados ngelamar utawi jelas 100% pelih..."

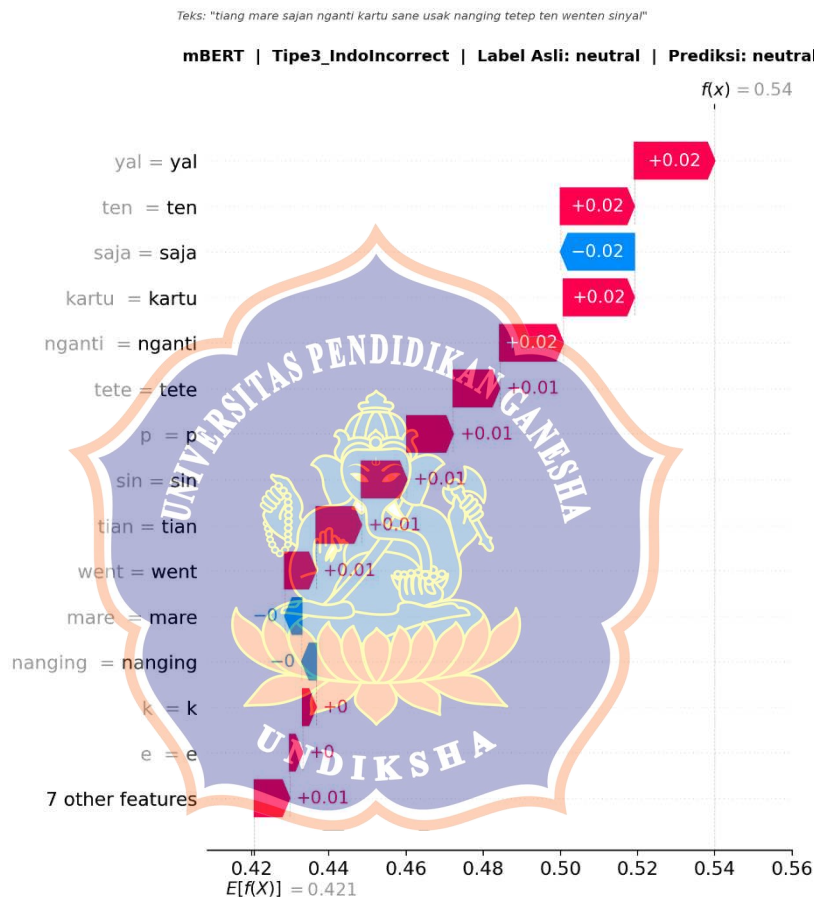


Gambar 19. Waterfall Plot SHAP mBERT, Idx 27 (Tipe 3, Kelas *Negative*)

Kalimat (Idx 27): "ae nike sampe sane ngranayang tiang gedeg. niki ten pelih sane dados ngelamar utawi jelas 100% pelih ring pihak sane di duur sane engken kaden cara ipun makoordinasi. sesek bane tiang yadiastun kasus puniki timpal tiang sane ring lampung."

Pada Gambar 19, mBERT mengevaluasi data Idx 27 dengan label aktual *negative* dan memprediksinya secara tepat sebagai *negative*. Pada sampel ini mBERT benar, sementara IndoBERT keliru. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,427$ dan mencapai $f(x) = 0,627$. Keberhasilan ini didorong secara kolektif oleh kelompok *59 other features* yang menyumbang +0,12. Peningkatan didukung oleh potongan token pembentuk kata negatif, yaitu *ge* (+0,01) dan *deg*

(+0,01) yang menyusun *gedeg* (marah), serta *tian* (+0,01) dan *g* (+0,01) yang membentuk *tiang* (saya). Tanda baca titik (+0,01) di akhir kalimat turut mendukung, sehingga akumulasi mendorong nilai keputusan melampaui nilai dasarnya menuju klasifikasi yang tepat.

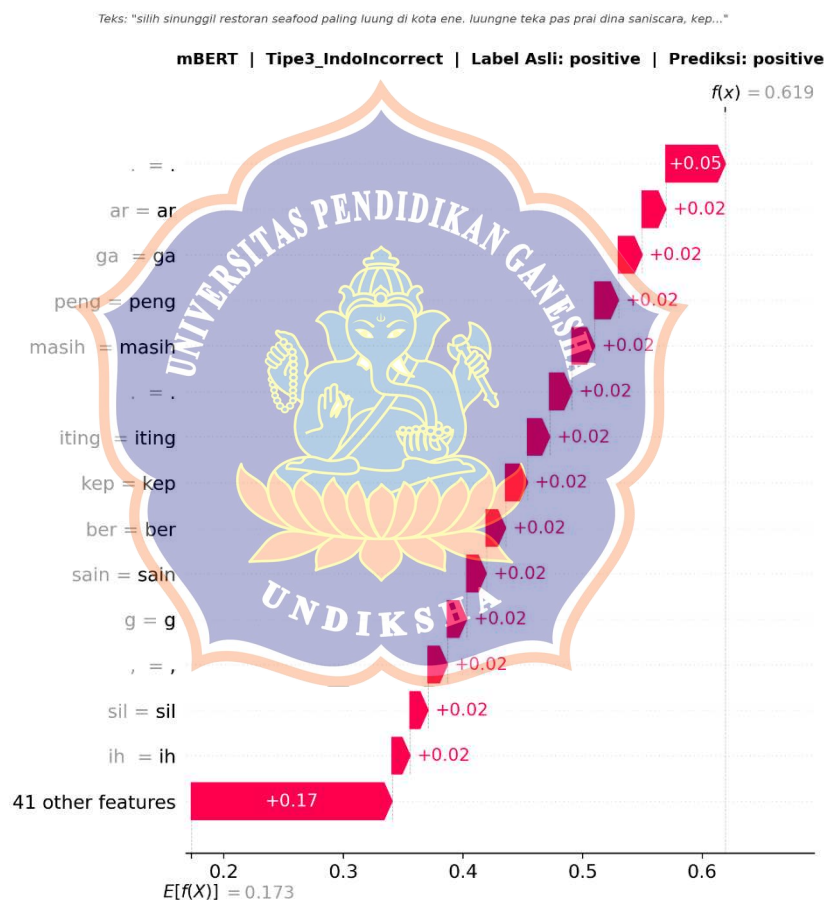


Gambar 20. Waterfall Plot SHAP mBERT, Idx 149 (Tipe 3, Kelas *Neutral*)

Kalimat (Idx 149): “*tiang mare sajan nganti kartu sane usak nanging tetep ten wenten sinyal*”

Pada Gambar 20, mBERT mengevaluasi data Idx 149 dengan label aktual *neutral* dan memprediksinya secara tepat sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,421$ dan mencapai $f(x) = 0,54$. Hasil ini dipengaruhi kombinasi kontribusi fitur yang berkompetisi secara berimbang. Penguat kelas

target didorong oleh *yal* (+0,02) dari kata *sinyal*, kata negasi *ten* (+0,02), kata benda *kartu* (+0,02), dan kata kerja *nganti* (+0,02). Beberapa token lain seperti *went*, *tian*, *sin*, dan *tete* memberi dorongan kecil ke atas. Di sisi lain, token *saja* (-0,02) dari kata *sajan* menarik nilai ke bawah. Kontribusi dengan magnitudo kecil dan merata ini menjaga nilai akhir stabil di area tengah (0,54), sehingga model menempatkannya pada kategori netral.



Gambar 21. Waterfall Plot SHAP mBERT, Idx 139 (Tipe 3, Kelas *Positive*)

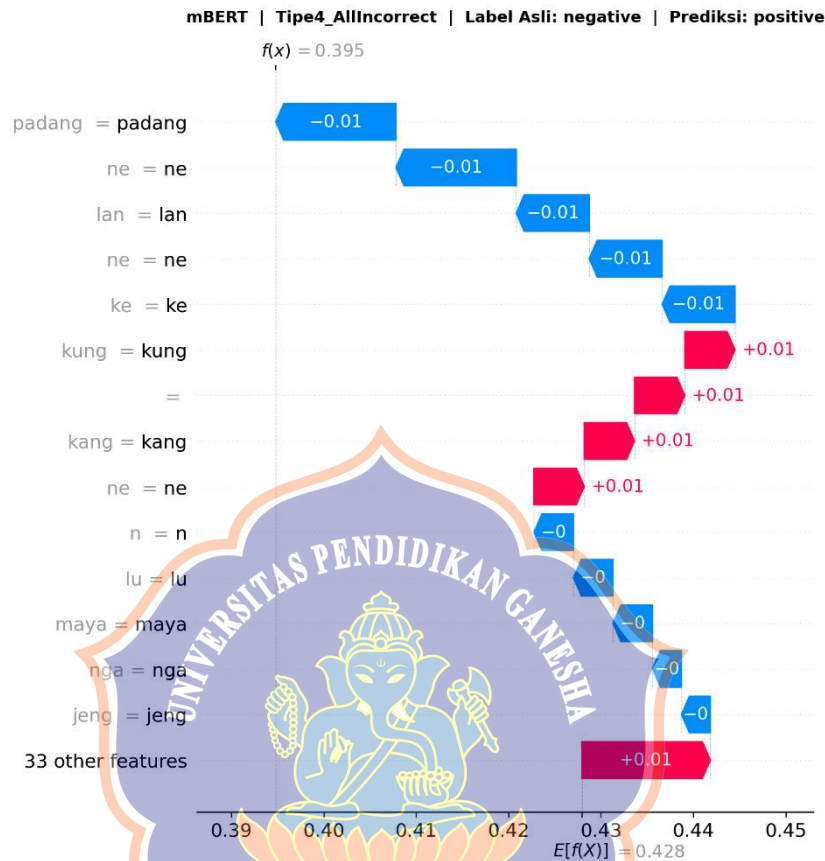
Kalimat (Idx 139): "silih sinunggil restoran seafood paling luung di kota ene. luungne teka pas prai dina saniscara, kepiting masih luung mara teka. pengarga bersaing, ten mudah mula nanging ten bes mael kok."

Pada Gambar 21, mBERT mengevaluasi data Idx 139 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,173$ yang tergolong rendah, namun naik hingga $f(x) = 0,619$. Model mengekstrak sinyal positif yang tersebar merata di sepanjang kalimat. Dorongan awal yang besar disumbangkan kelompok *41 other features* (+0,17). Peningkatan diperkuat oleh rentetan token kata dan sub-kata dengan kontribusi seragam +0,02, meliputi *ar*, *ga*, *peng*, *masih*, *iting*, serta pecahan *sil* dan *ih* pembentuk kata *silih*. Tanda baca titik di akhir memberi kontribusi +0,05. Distribusi kontribusi positif yang konsisten ini mengakumulasi bobot semantik kalimat hingga mencapai klasifikasi positif.



Tipe 4 (*All Incorrect*)

Teks: "kangkung ne lumayan utawi yuyu saus padang ne ngranang keciwa irage baange yuyu sane kopong dadine i..."

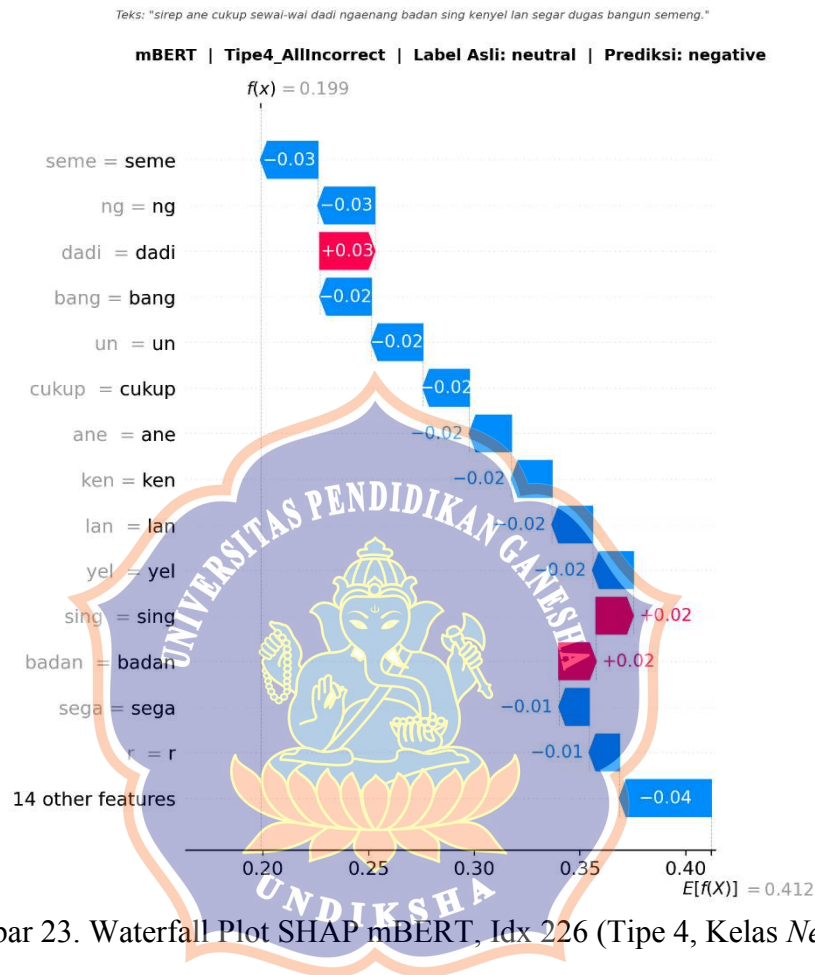


Gambar 22. Waterfall Plot SHAP mBERT, Idx 2 (Tipe 4, Kelas *Negative*)

Kalimat (Idx 2): "kangkung ne lumayan utawi yuyu saus padang ne ngranang keciwa irage baange yuyu sane kopong dadine irage ten ngajeng yuyu ne lan ke uliang."

Pada Gambar 22, mBERT mengevaluasi data Idx 2 dengan label aktual *negative*, tetapi memprediksinya sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,428$ dan turun ke $f(x) = 0,395$. Pola atribusi mengindikasikan bahwa pemecahan kata bahasa Bali melemahkan deteksi sentimen negatif (seperti kata *keciwa*). Terdapat kontribusi penurunan dari token *padang* (-0,01), *ne* (-0,01), *lan* (-0,01), dan *ke* (-0,01), namun tidak cukup kuat membawa prediksi ke kelas negatif karena ditahan dorongan positif dari *kung* (+0,01) dan *kang* (+0,01) pembentuk kata

kangkung, token *ne* (+0,01), serta kelompok 33 *other features* (+0,01). Akibatnya, nilai akhir tertahan di 0,395 dan model salah memprediksi.

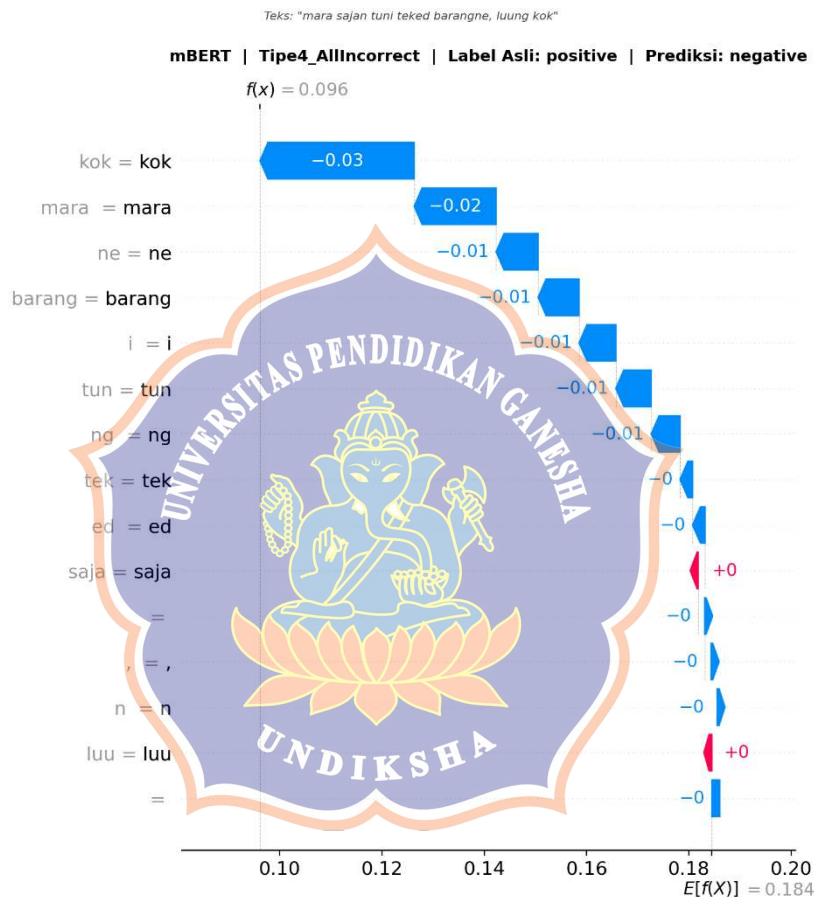


Gambar 23. Waterfall Plot SHAP mBERT, Idx 226 (Tipe 4, Kelas *Neutral*)

Kalimat (Idx 226): “sirep ane cukup sewai-wai dadi ngaenang badan sing kenel lan segar dugas bangun semeng.”

Pada Gambar 23, mBERT mengevaluasi data Idx 226 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,412$ dan turun ke $f(x) = 0,199$. Pola atribusi mengindikasikan bahwa fragmentasi sub-kata memberi sinyal bias negatif. Penurunan dipimpin potongan *seme* (-0,03) dan *ng* (-0,03) penyusun kata *semeng* (pagi), diikuti pecahan *bang* (-0,02) dan *un* (-0,02) dari *bangun*. Kata *cukup* (-0,02), *ane* (-0,02), *lan* (-0,02), serta

ken (-0,02) dan *yel* (-0,02) dari kata *kenyel* (lelah) turut menurunkan probabilitas. Meskipun *dadi* (+0,03), *sing* (+0,02), dan *badan* (+0,02) memberi dorongan naik, kelompok *14 other features* (-0,04) menekan nilai akhir ke 0,199, sehingga prediksi keliru ke kelas negatif.



Gambar 24. Waterfall Plot SHAP mBERT, Idx 363 (Tipe 4, Kelas *Positive*)

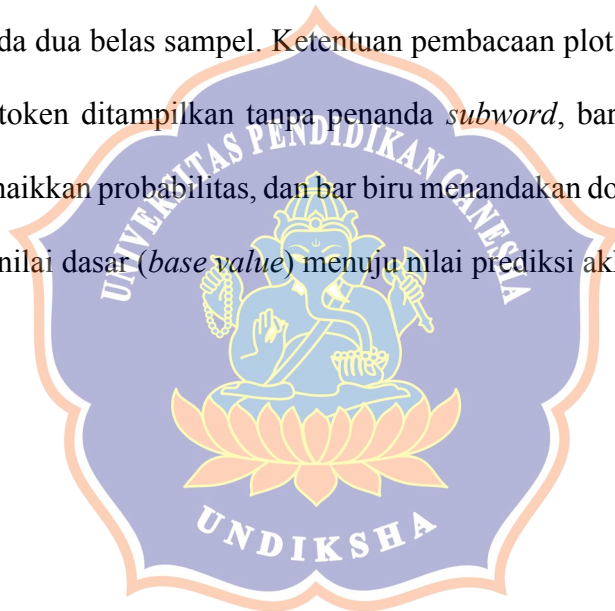
Kalimat (Idx 363): “mara sajan tuni teked barangne, luung kok”

Pada Gambar 24, mBERT mengevaluasi data Idx 363 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,184$ dan turun ke $f(x) = 0,096$. Hampir seluruh token dominan menarik nilai ke bawah. Penurunan terbesar berasal dari token penegas khas Bali *kok* (-0,03), disusul kata keterangan waktu *mara* (-0,02). Kata benda *barang* (-0,01) serta

pecahan *tun* (-0,01) dan *ng* (-0,01) ikut mengikis nilai prediksi. Kata bermakna positif *luung* (bagus) terpecah menjadi *luu* (+0,00) dan *ng* (-0,01) sehingga kehilangan bobot semantik positifnya. Ketiadaan gaya dorong positif yang berarti membuat nilai akhir menyusut ke 0,096, sehingga model gagal mengenali sifat positif teks.

C. Model XLM-RoBERTa

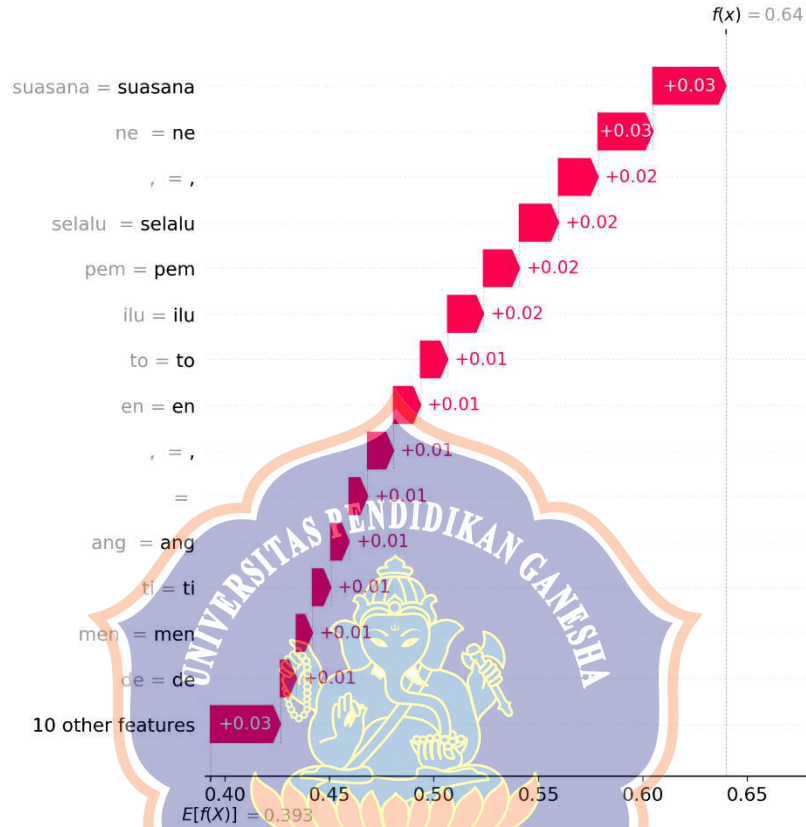
Bagian ini memuat seluruh *waterfall plot* SHAP lokal untuk model XLM-RoBERTa pada dua belas sampel. Ketentuan pembacaan plot sama dengan bagian sebelumnya: token ditampilkan tanpa penanda *subword*, bar merah menandakan dorongan menaikkan probabilitas, dan bar biru menandakan dorongan menurunkan, dihitung dari nilai dasar (*base value*) menuju nilai prediksi akhir $f(x)$.



Tipe 1 (*All Correct*)

Teks: "tiang paling sing demen ento, yening paek pemilu, suasanane selalu memanes."

XLM-RoBERTa | Tipe1_AllCorrect | Label Asli: negative | Prediksi: negative

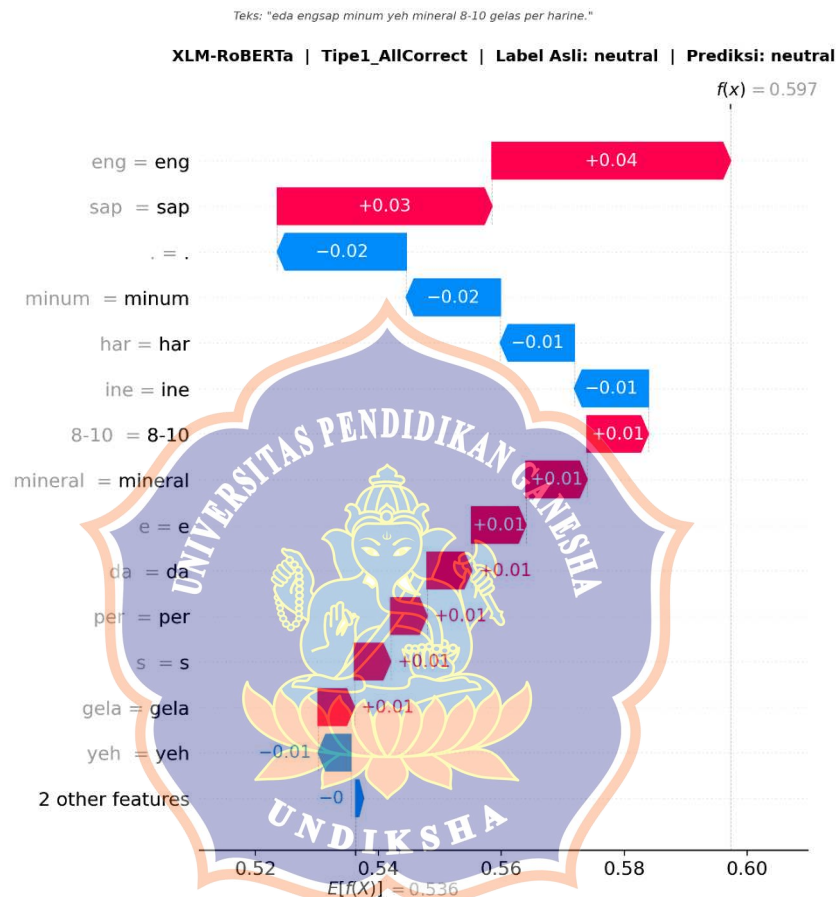


Gambar 25. Waterfall Plot SHAP XLM-RoBERTa, Idx 336 (Tipe 1, Kelas Negative)

Kalimat (Idx 336): "tiang paling sing demen ento, yening paek pemilu, suasanane selalu memanes."

Pada Gambar 25, XLM-RoBERTa mengevaluasi data Idx 336 dengan label aktual *negative* dan memprediksinya secara tepat sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,393$ dan mencapai $f(x) = 0,64$. Hampir seluruh fitur memberikan dorongan searah untuk menaikkan nilai prediksi ke kelas target. Kontribusi dipimpin merata oleh token *suasana* (+0,03) dan *ne* (+0,03) yang membentuk sub-frasa *suasanane*. Pengaruh ini diperkuat oleh *selalu* (+0,02) serta pecahan *pem* (+0,02) dan *ilu* (+0,02) yang merujuk kata *pemilu*. Komponen lain

seperti potongan token *tiang* dan *demen*, bersama kelompok *10 other features* (+0,03), mendorong nilai probabilitas melampaui ambang keputusan untuk sentimen negatif.

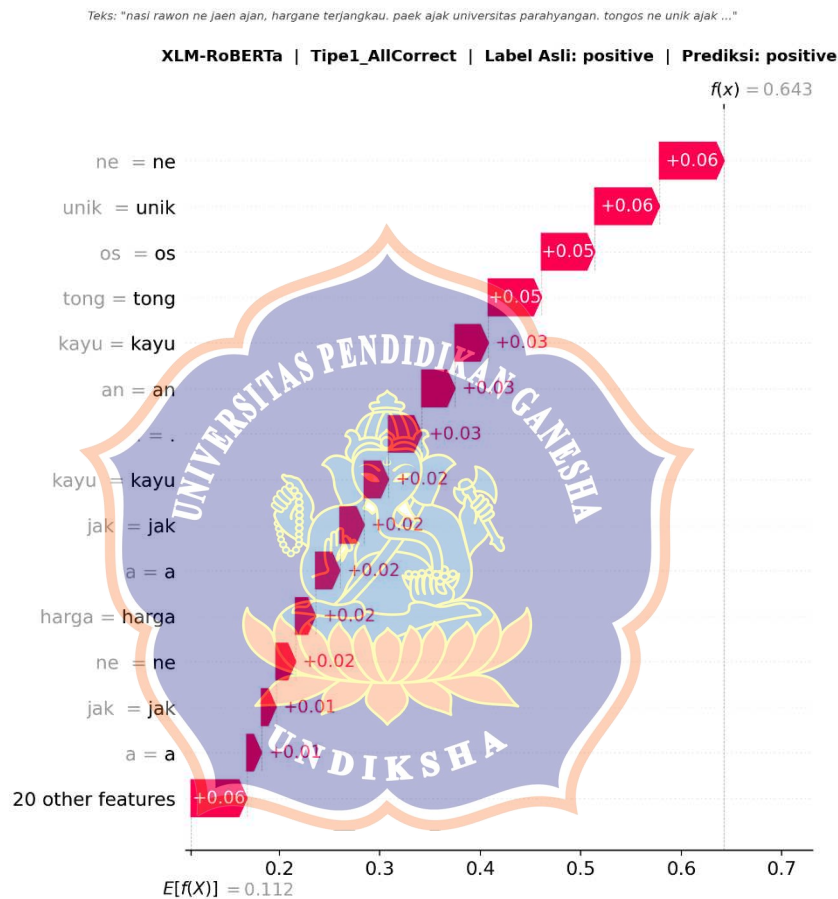


Gambar 26. Waterfall Plot SHAP XLM-RoBERTa, Idx 300 (Tipe 1, Kelas *Neutral*)

Kalimat (Idx 300): "eda engsap minum yeh mineral 8-10 gelas per harine."

Pada Gambar 26, XLM-RoBERTa mengevaluasi data Idx 300 dengan label aktual *neutral* dan memprediksinya secara tepat sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,536$ dan mencapai $f(x) = 0,597$. Peningkatan probabilitas didorong paling kuat oleh pecahan token *eng* (+0,04) dan *sap* (+0,03) yang membentuk kata *engsap* (lupa). Penguatan minor disumbangkan token angka *8-10*

(+0,01) dan kata *mineral* (+0,01). Di sisi lain, terdapat gaya penahanan dari tanda baca titik (-0,02) dan kata *minum* (-0,02). Keseimbangan kontribusi yang tidak ekstrem ini menjaga nilai prediksi di sekitar nilai dasar, sehingga model menetapkan teks pada sentimen netral.



Gambar 27. Waterfall Plot SHAP XLM-RoBERTa, Idx 272 (Tipe 1, Kelas Positive)

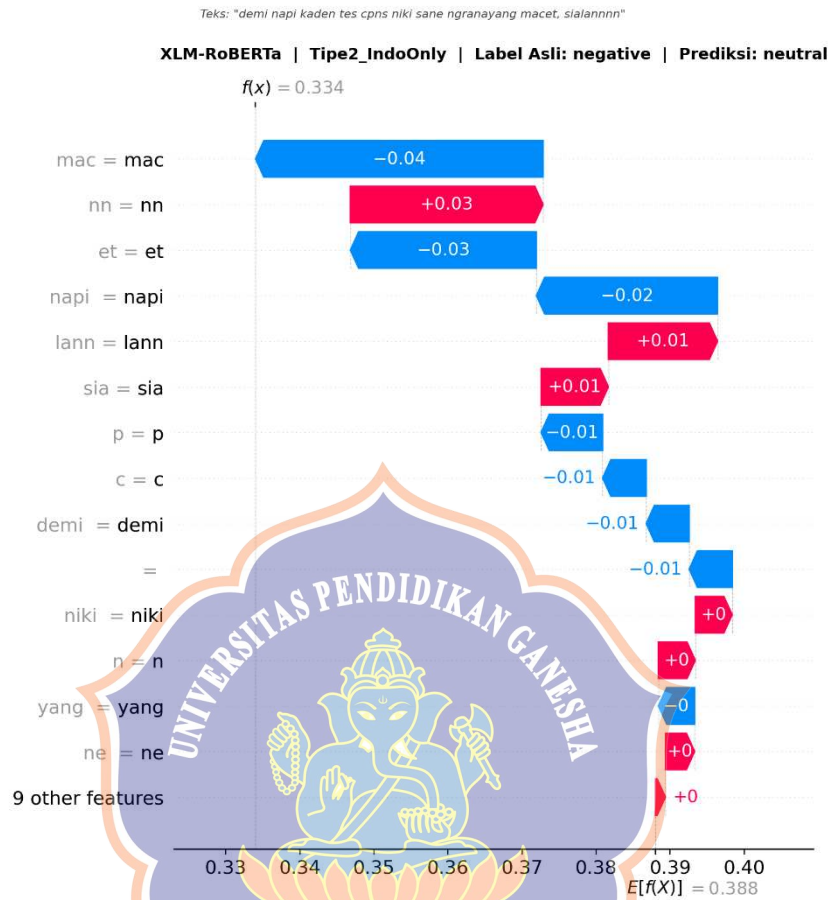
Kalimat (Idx 272): "nasi rawon ne jaen ajan, hargane terjangkau. paek ajak universitas parahyangan. tongos ne unik ajak kayu kayuan."

Pada Gambar 27, XLM-RoBERTa mengevaluasi data Idx 272 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,112$ dan naik signifikan hingga $f(x) = 0,643$. Kenaikan ini didorong akumulatif oleh kemunculan kata bermakna leksikal positif.

Kontribusi terbesar diberikan token penegas *ne* (+0,06) dan kata sifat *unik* (+0,06). Potongan *os* (+0,05) dan *tong* (+0,05) yang menyusun kata *tongos* (tempat) menjadi pendorong berikutnya. Fitur penunjang seperti *kayu* (+0,03), *an* (+0,03), serta kelompok *20 other features* (+0,06) memperkuat prediksi, sehingga model mengenali sentimen positif pada teks.



Tipe 2 (IndoBERT Only Correct)

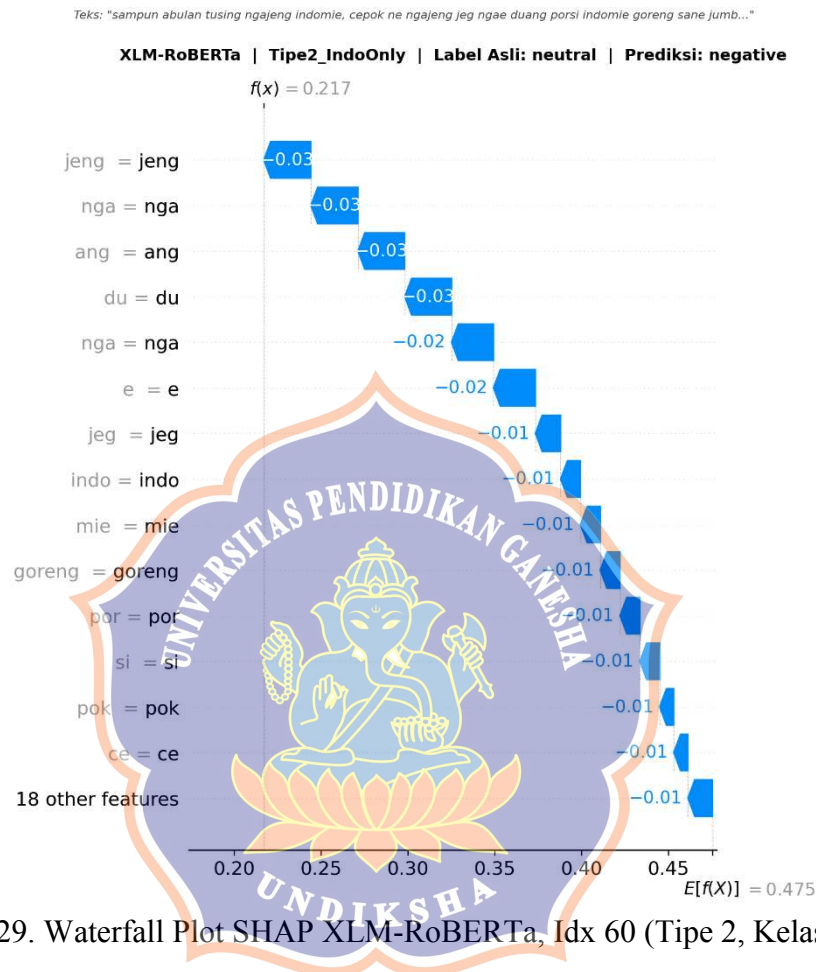


Gambar 28. Waterfall Plot SHAP XLM-RoBERTa, Idx 398 (Tipe 2, Kelas Negative)

Kalimat (Idx 398): "demi napi kaden tes cpns niki sane ngranayang macet, sialannnn"

Pada Gambar 28, XLM-RoBERTa mengevaluasi data Idx 398 dengan label aktual *negative*, tetapi memprediksinya sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,388$ dan turun ke $f(x) = 0,334$. Pola atribusi mengindikasikan reduksi probabilitas oleh beberapa token sub-kata. Penurunan terbesar berasal dari token *mac* (dari kata *macet*) sebesar -0,04, diikuti *et* (-0,03), serta kata tanya *napi* (-0,02). Di sisi lain, token *nn* (+0,03) serta pecahan *sia* (+0,01) dan *lann* (+0,01) dari kata *sialannnn* mencoba mendorong ke kelas target. Akumulasi keseluruhan

menurunkan nilai ke 0,334, sehingga model gagal mengidentifikasi teks sebagai negatif.

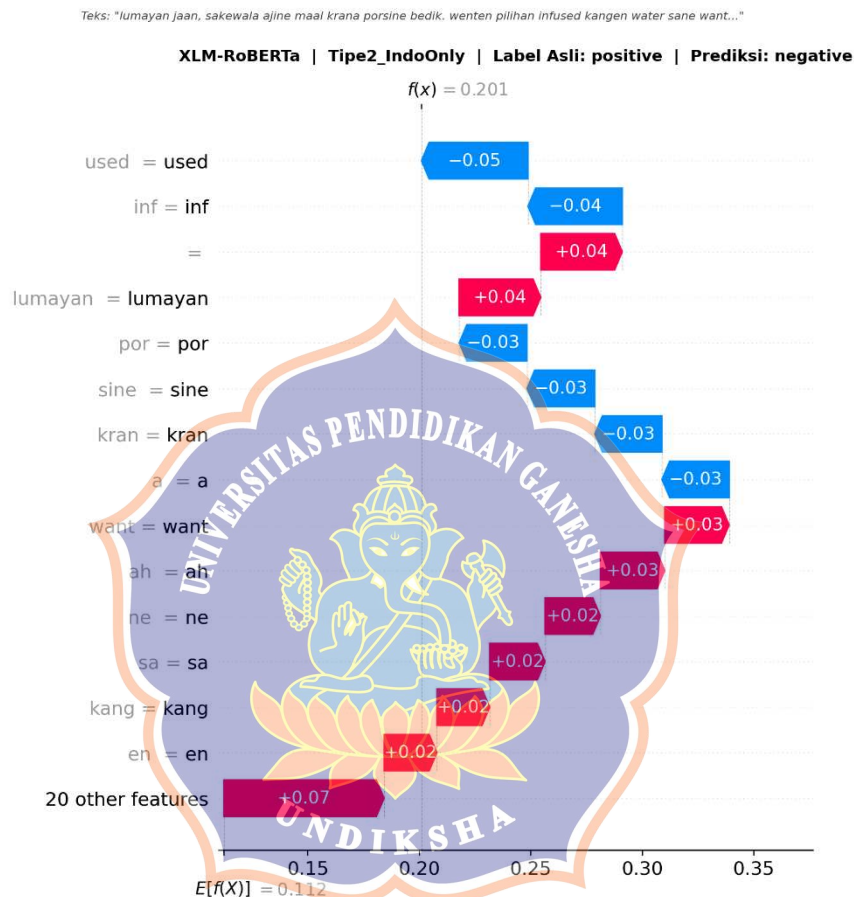


Gambar 29. Waterfall Plot SHAP XLM-RoBERTa, Idx 60 (Tipe 2, Kelas *Neutral*)

Kalimat (Idx 60): “sampun abulan tusing ngajeng indomie, cepok ne ngajeng jeg ngae duang porsi indomie goreng sane jumbo.”

Pada Gambar 29, XLM-RoBERTa mengevaluasi data Idx 60 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,475$ dan turun ke $f(x) = 0,217$. Hampir seluruh token dominan memberikan kontribusi negatif. Penurunan didorong merata oleh pemotongan sub-kata dari aktivitas porsi makan, seperti *jeng* (-0,03), *nga* (-0,03), *ang* (-0,03), dan *du* (-0,03) yang berasal dari frasa *ngae duang porsi*. Kata penegas khas Bali *jeg*

turut menurunkan nilai sebesar -0,01. Total akumulasi token ini, termasuk kelompok *18 other features* (-0,01), membelokkan prediksi dari yang seharusnya netral menjadi kelas negatif.



Gambar 30. Waterfall Plot SHAP XLM-RoBERTa, Idx 56 (Tipe 2, Kelas Positive)

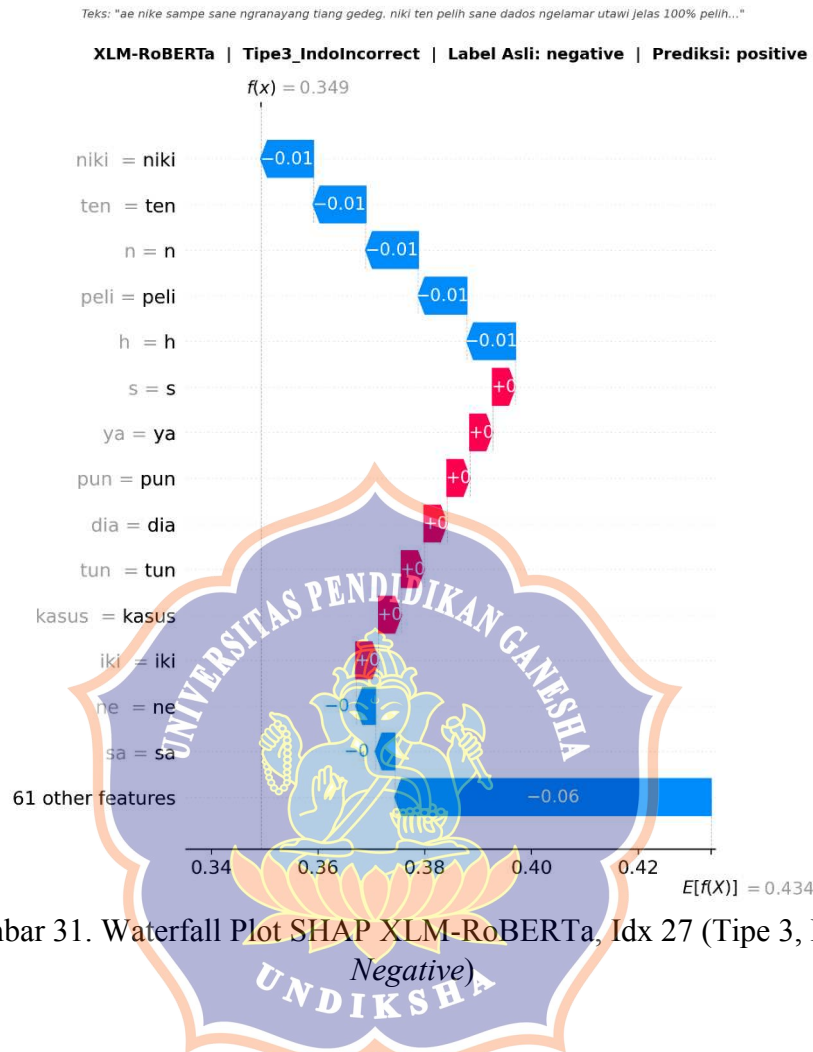
Kalimat (Idx 56): "lumayan jaan, sakewala ajine maal krana porsine bedik. wenten pilihan infused kangen water sane wantah antuk 4 orang"

Pada Gambar 30, XLM-RoBERTa mengevaluasi data Idx 56 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,112$ dan mencapai $f(x) = 0,201$. Pola atribusi menunjukkan kompetisi kontribusi yang ketat. Sentimen positif sudah didorong token *lumayan* (+0,04) serta kelompok *20 other features* (+0,07). Namun, nilai akhir tertahan oleh

token negatif dari kata serapan asing, seperti *used* (-0,05) dan *inf* (-0,04) yang merujuk istilah *infused kangen water*. Ditambah potongan *por* (-0,03) dan *kran* (-0,03), akumulasi nilai negatif ini membatasi kenaikan, sehingga model gagal menembus ambang keputusan kelas positif.



Tipe 3 (IndoBERT *Incorrect*)

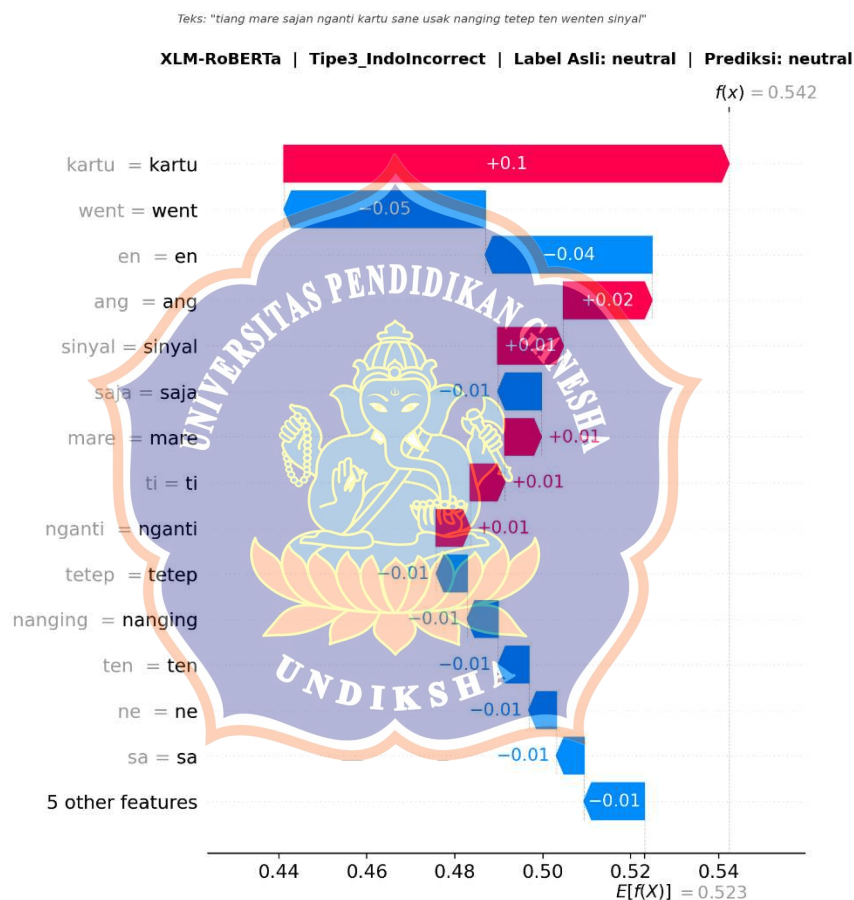


Gambar 31. Waterfall Plot SHAP XLM-RoBERTa, Idx 27 (Tipe 3, Kelas Negative)

Kalimat (Idx 27): "ae niki sampe sane ngranayang tiang gedeg. niki ten pelih sane dados ngelamar utawi jelas 100% pelih ring pihak sane di duur sane engken kaden cara ipun makoordinasi. sesek bane tiang yadiastun kasus puniki timpal tiang sane ring lampung."

Pada Gambar 31, XLM-RoBERTa mengevaluasi data Idx 27 dengan label aktual *negative*, tetapi memprediksinya sebagai *positive*. Pada sampel ini XLM-RoBERTa keliru, sedangkan mBERT benar. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,434$ dan turun ke $f(x) = 0,349$. Pola atribusi mengindikasikan nilai probabilitas sempat ditekan oleh kelompok *61 other features* (-0,06) serta token

penolak seperti *niki* (-0,01), *ten* (-0,01), serta *pelih* (-0,01) dan *h* (-0,01) dari kata *pelih* (salah). Namun, dorongan positif kecil dari token transisi di bagian tengah menahan nilai akhir pada 0,349. Nilai ini tetap memicu kesalahan keputusan, yang mengindikasikan bahwa pada konfigurasi ini model kesulitan mengekstrak emosi negatif kuat seperti kata *gedeg* pada struktur kalimat tersebut.

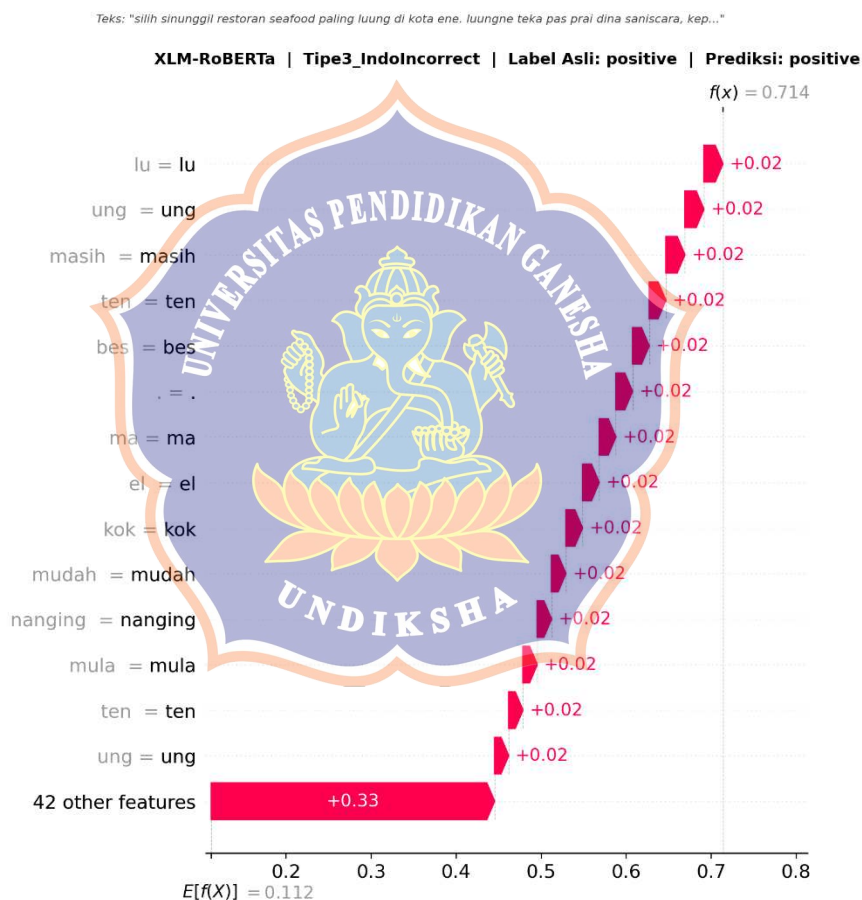


Gambar 32. Waterfall Plot SHAP XLM-RoBERTa, Idx 149 (Tipe 3, Kelas *Neutral*)

Kalimat (Idx 149): “tiang mare sajan nganti kartu sane usak nanging tetep ten wenten sinyal”

Pada Gambar 32, XLM-RoBERTa mengevaluasi data Idx 149 dengan label aktual *neutral* dan memprediksinya secara tepat sebagai *neutral*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,523$ dan mencapai $f(x) = 0,542$. Plot memperlihatkan

dinamika kompetisi fitur yang kontras. Token objek *kartu* memegang kendali utama mendorong prediksi ke atas dengan kontribusi tunggal tinggi (+0,1). Sebaliknya, dorongan itu dinetralisasi oleh kata penolak *went* (-0,05) dan *en* (-0,04) yang membentuk kata *wenten* (ada). Kombinasi token penguat minor seperti *ang* (+0,02) dan *sinyal* (+0,01) dengan penahan seperti *tetep* (-0,01) dan *nanging* (-0,01) menjaga nilai akhir stabil di koridor sentimen netral.



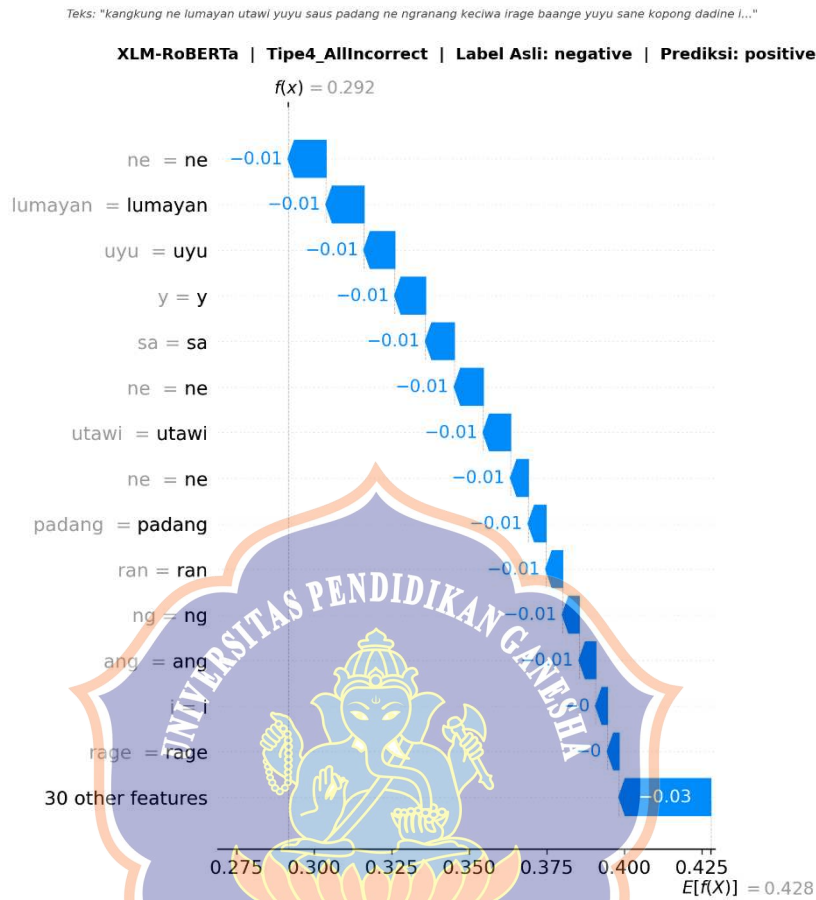
Gambar 33. Waterfall Plot SHAP XLM-RoBERTa, Idx 139 (Tipe 3, Kelas Positive)

Kalimat (Idx 139): "silih sinunggil restoran seafood paling luung di kota ene. luungne teka pas prai dina saniscara, kepiting masih luung mara teka. pengarga bersaing, ten mudah mula nanging ten bes mael kok."

Pada Gambar 33, XLM-RoBERTa mengevaluasi data Idx 139 dengan label aktual *positive* dan memprediksinya secara tepat sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,112$ dan naik hingga $f(x) = 0,714$. Gaya dorong terbesar didominasi kelompok 42 *other features* (+0,33). Penguatan diperkuat oleh rentetan token bahasa Bali dengan kontribusi merata +0,02, meliputi *lu* dan *ung* (pembentuk kata *luung* atau *luungne*, bagus), serta kata pendukung seperti *masih*, *ten*, *bes*, *mula*, dan *nanging*. Akumulasi fitur positif ini menaikkan probabilitas dari titik dasar menuju estimasi akhir yang akurat.



Tipe 4 (*All Incorrect*)

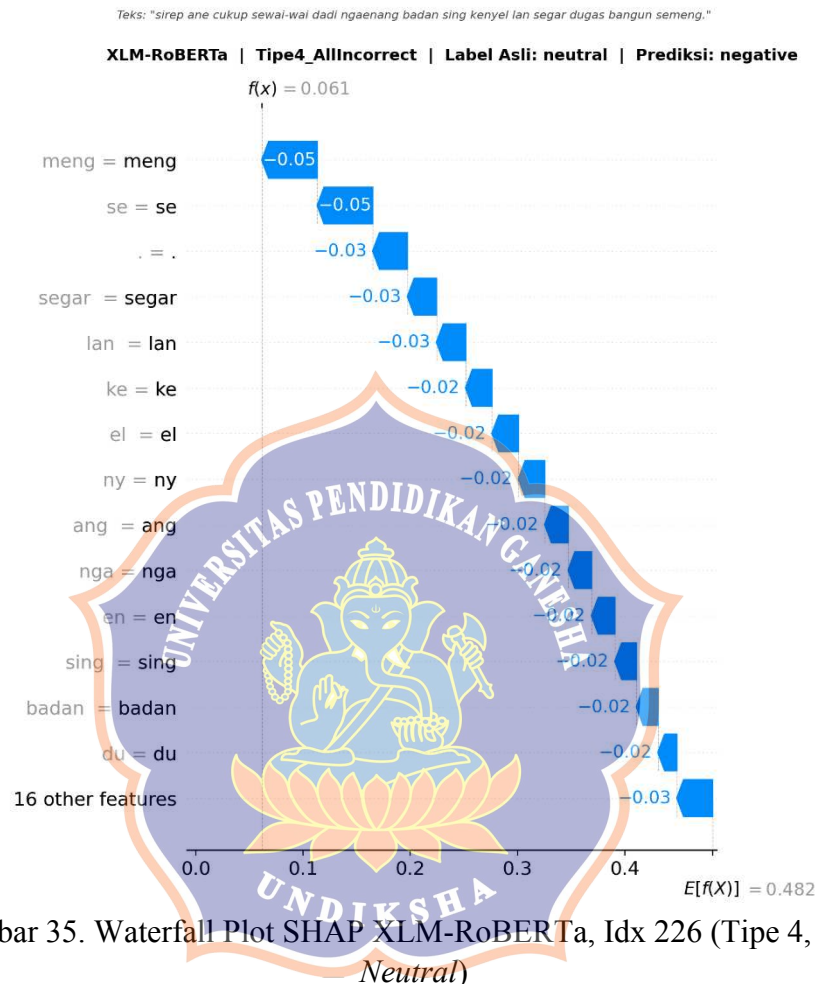


Gambar 34. Waterfall Plot SHAP XLM-RoBERTa, Idx 2 (Tipe 4, Kelas *Negative*)

Kalimat (Idx 2): “*kangkung ne lumayan utawi yuyu saus padang ne ngranang keciwa irage baange yuyu sane kopong dadine irage ten ngajeng yuyu ne lan ke uliang.*”

Pada Gambar 34, XLM-RoBERTa mengevaluasi data Idx 2 dengan label aktual *negative*, tetapi memprediksinya sebagai *positive*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,428$ dan turun ke $f(x) = 0,292$. Seluruh fitur memberikan kontribusi penurunan. Token penurun terbesar berasal dari kelompok *30 other features* (-0,03). Token individu seperti *ne* (-0,01), kata *lumayan* (-0,01), potongan *uyu* (-0,01) dari kata *yuyu*, serta *y* (-0,01) dan *sa* (-0,01) memberi penurunan seragam. Pola penurunan tanpa token penguat dominan ini mengindikasikan

ketidakpastian model dalam memetakan batas keputusan kelas pada teks dengan kepadatan kosakata bahasa Bali yang tinggi.

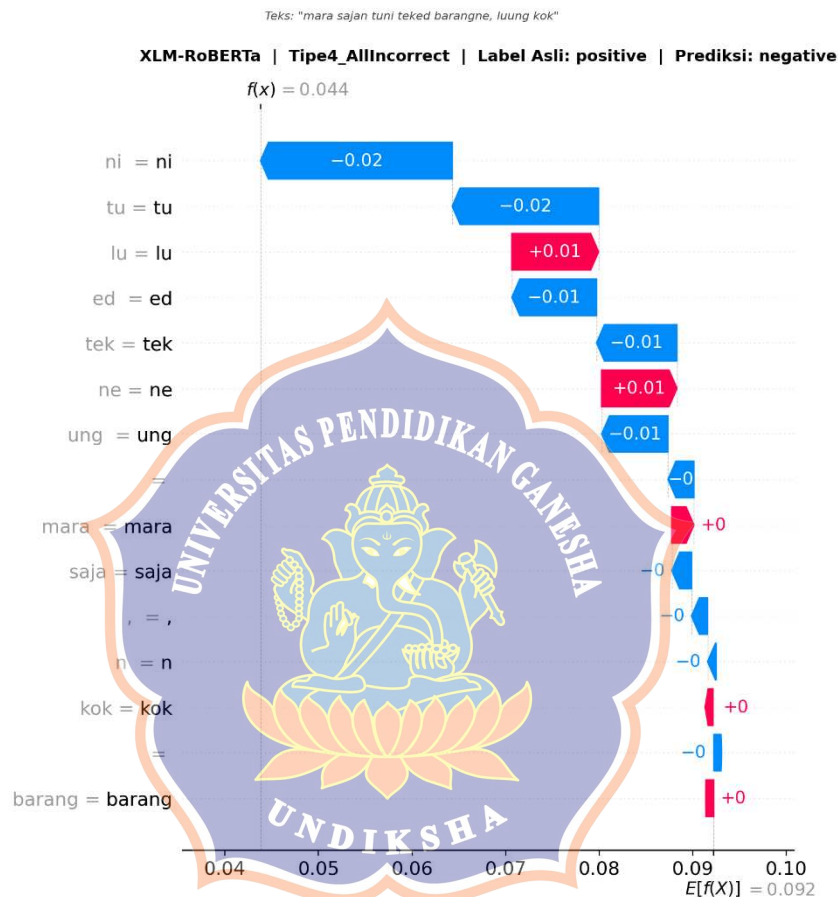


Gambar 35. Waterfall Plot SHAP XLM-RoBERTa, Idx 226 (Tipe 4, Kelas *Neutral*)

Kalimat (Idx 226): “sirep ane cukup sewai-wai dadi ngaenang badan sing kenyel lan segar dugas bangun semeng.”

Pada Gambar 35, XLM-RoBERTa mengevaluasi data Idx 226 dengan label aktual *neutral*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,482$ dan turun ke $f(x) = 0,061$. Seluruh token dominan secara kompak memberikan kontribusi negatif. Penurunan dipimpin token *meng* (-0,05) dan *se* (-0,05), diikuti tanda baca titik (-0,03), serta kata *segar* (-0,03) dan *lan* (-0,03). Fitur pendukung lain seperti *sing* dan *badan* turut memangkas probabilitas

masing-masing sekitar -0,02. Akumulasi gaya penolak tanpa token penguat ini menekan nilai ekspektasi awal, sehingga menghasilkan nilai akhir yang sangat rendah dan meleset dari label aslinya.



Gambar 36. Waterfall Plot SHAP XLM-RoBERTa, Idx 363 (Tipe 4, Kelas Positive)

Kalimat (Idx 363): "mara sajan tuni teked barangne, luung kok"

Pada Gambar 36, XLM-RoBERTa mengevaluasi data Idx 363 dengan label aktual *positive*, tetapi memprediksinya sebagai *negative*. Penalaran dimulai dari nilai dasar $E[f(x)] = 0,092$ dan turun ke $f(x) = 0,044$. Pola atribusi mengindikasikan bahwa kata leksikal positif *luung* terpecah menjadi *lu* (+0,01) dan *ung* (-0,01). Dampak positif dari *lu* (+0,01) dan *ne* (+0,01) teranulir oleh kontribusi penurunan dari

ni (-0,02) dan *tu* (-0,02) yang berasal dari kata *tuni*. Kombinasi pemotongan token ini menyebabkan akumulasi nilai akhir berada di bawah ambang keputusan untuk sentimen positif.



Lampiran 3. Riwayat Hidup Penulis

RIWAYAT HIDUP PENULIS

Penulis bernama I Komang Deny Supanji, lahir di Karangasem pada tanggal 27



Desember 1989. Penulis merupakan anak ketiga dari tiga bersaudara, putra dari pasangan I Komang Oka dan Ni Made Sukaningsih. Pendidikan dasar ditempuh di SD Negeri 1 Karangasem dan diselesaikan pada tahun 2003. Penulis kemudian melanjutkan pendidikan di SMP Negeri 2 Amlapura dan lulus pada tahun 2005, serta menyelesaikan pendidikan menengah atas di SMA Negeri 2 Amlapura pada jurusan Ilmu Pengetahuan Alam (IPA) pada tahun 2008. Pada tahun yang sama, Penulis diterima di Universitas Pendidikan Ganesha Singaraja pada Program Studi S1 Pendidikan Teknik Informatika dan menyelesaikan studi pada tahun 2012 dengan memperoleh gelar Sarjana Pendidikan (S.Pd.).

Setelah menyelesaikan pendidikan sarjana, Penulis berkarier sebagai guru Pegawai Negeri Sipil (PNS) di SMK Negeri 1 Denpasar sejak tahun 2014 dan hingga saat ini mengajar pada program keahlian Desain Komunikasi Visual. Dorongan untuk terus meningkatkan kompetensi profesional membawa Penulis melanjutkan studi pada Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha, pada tahun 2022. Komunikasi terkait penelitian tesis ini dapat disampaikan melalui surel : deny@student.undiksha.ac.id.