



# Analisis Sentimen Ulasan Aplikasi eFootball Pada Google Play Store Menggunakan Multinomial Naïve Bayes dan SVM

Mohammad Umar Sasongko<sup>1\*</sup>, Bambang Irawan<sup>2</sup>

<sup>1\*,2</sup>Program Studi Teknik Informatika, Universitas Muhadi Setiabudi Brebes, Indonesia

\*Email : [moh.umar717@gmail.com](mailto:moh.umar717@gmail.com)

## Abstract:

*In the digital era, the use of mobile applications is increasing, so it is important to understand user satisfaction and dissatisfaction with the applications used. One of the popular mobile games in Indonesia is eFootball 2026, which has a lot of reviews from players on the Google Play Store. The large number of reviews allows sentiment analysis to be carried out to find out user opinions on the quality of the application. This study aims to analyze the sentiment of eFootball application user reviews using the Multinomial Naive Bayes method and Support Vector Machine. Review data is processed through the stages of text preprocessing, feature extraction using TF-IDF, and class imbalance handling with SMOTE. Model evaluation was carried out using accuracy, precision, recall, and F1-score metrics. The results showed that the Multinomial Naive Bayes method produced an accuracy of 76.72%, while the Support Vector Machine obtained an accuracy of 74.92%, with relatively balanced precision, recall, and F1-score values. Based on these results, it can be concluded that the Multinomial Naive Bayes method has a better performance in analyzing the sentiment of eFootball app reviews on the Google Play Store and can be used as a basis for evaluation for future app development.*

**Keywords:** Sentiment Analysis; eFootball; Multinomial Naive Bayes; SVM

## 1. PENDAHULUAN

Dalam era digital saat ini, penggunaan aplikasi *mobile* mengalami peningkatan yang sangat pesat, terutama di kalangan masyarakat urban yang semakin bergantung pada perangkat pintar untuk hiburan sehari-hari. Hal ini menjadikan pemahaman mendalam terhadap kepuasan dan ketidakpuasan pengguna sangat krusial bagi pengembang aplikasi guna meningkatkan kualitas produk dan mempertahankan loyalitas pengguna (R & Effendy, 2025). *Google* menyediakan layanan *Google Play Store* sebagai platform distribusi utama konten digital, yang mencakup beragam kategori seperti *game*, aplikasi utilitas, film, musik, dan buku elektronik, dengan jutaan unduhan harian di seluruh dunia (Nouval et al., 2026). Salah satu fitur paling berharga di *Play Store* adalah sistem *rating* dan ulasan, di mana pengguna dapat menyampaikan opini jujur mereka mengenai pengalaman menggunakan produk tersebut, mulai dari kelebihan fitur hingga keluhan teknis (Lundam & Darmawan, 2024).

*Google Play Store* menampilkan berbagai aplikasi game dengan tingkat popularitas tinggi, di mana predikat "aplikasi terbaik" sering ditentukan oleh kombinasi jumlah unduhan, *rating* bintang rata-rata, dan volume ulasan positif (Nurian et al., 2023).

---

Kategori *game* olahraga, khususnya sepak bola digital, mendominasi *chart* unduhan berkat daya tarik globalnya. eFootball 2026, yang dikembangkan oleh Konami sebagai evolusi dari seri *Pro Evolution Soccer* (PES), merupakan salah satu *mobile game* paling populer di Indonesia dan dunia, dengan mekanisme permainan realistis, grafis berkualitas tinggi, mode *multiplayer online*, serta pembaruan musiman yang menarik jutaan penggemar sepak bola. *Game* ini memungkinkan pengguna untuk membangun tim impian, bertanding secara *real-time*, dan mengikuti *event* global, sehingga menghasilkan basis pengguna aktif yang masif. Pada Januari 2026, eFootball telah mencapai ratusan juta unduhan di *Google Play Store*, dengan ribuan ulasan baru setiap hari yang mencerminkan *feedback* autentik dari berbagai kalangan pemain.

Ulasan pengguna eFootball 2026 di *Google Play Store* sering kali membahas aspek krusial seperti performa grafis, kestabilan server saat mode *online*, keseimbangan *gameplay*, pembelian *in-app*, serta respons pengembang terhadap bug. *Feedback* negatif umum meliputi lag pada perangkat *low-end*, ketidakseimbangan tim *pro player*, atau keterlambatan update fitur baru, sementara ulasan positif menyoroti realisme fisika bola dan komunitas kompetitif. Volume ulasan yang begitu besar sering mencapai ratusan ribu menghasilkan data teks berharga yang dapat dianalisis untuk mengidentifikasi tren kepuasan pengguna secara keseluruhan. Aplikasi ini telah memperoleh lisensi resmi dari otoritas terkait dan mengadopsi konsep *open platform* untuk integrasi dengan layanan lain seperti pembelian tiket *event* olahraga atau *merchandise* klub sepak bola, yang semakin memperluas ekosistemnya. Dengan lebih dari 100 juta unduhan global saat ini, eFootball 2026 menjadi subjek ideal untuk analisis mendalam karena tingkat penggunaannya yang tinggi di Indonesia.

Analisis sentimen merupakan cabang dari text mining dan pemrosesan bahasa alami (*Natural Language Processing*) yang bertujuan untuk mengidentifikasi serta mengklasifikasikan opini, emosi, dan sikap subjektif yang terkandung dalam suatu teks (Ranjan & Bhaba, 2020). Penelitian sebelumnya menunjukkan efektivitas algoritma Naive Bayes dalam mengklasifikasikan sentimen ulasan aplikasi di *Google Play Store*, di mana mayoritas ulasan positif mencerminkan kepuasan pengguna terhadap fitur layanan aplikasi finansial seperti Dana (Nawulansih et al., 2025). Artikel *open source* juga memperlihatkan bahwa Multinomial Naive Bayes mencapai akurasi 77% pada ulasan aplikasi mobile banking, menunjukkan konsistensi performa pada berbagai jenis aplikasi (Amirkhalili & Wong, 2024). Fokus utama analisis sentimen adalah menentukan polaritas sentimen, yaitu positif, negatif, atau netral, terhadap suatu objek tertentu berdasarkan informasi tekstual yang tersedia (Sarimole, 2024). Metode ini banyak digunakan untuk menganalisis opini publik pada media sosial, forum daring, serta platform berbasis ulasan seperti *Google Play Store* karena mampu mengolah data dalam jumlah besar secara efisien (Setiawan & Kaburuan, 2023).

"Dalam kajian ulasan aplikasi Liputan6.com, nilai AUC dan akurasi SVM menunjukkan bahwa metode ini efektif dalam memisahkan sentimen positif dan negatif, menguatkan peran machine learning dalam evaluasi opini pengguna (Jurnal et al., 2025). *Google Play Store* menyediakan sistem rating dan ulasan yang memungkinkan pengguna untuk memberikan penilaian terhadap aplikasi yang mereka gunakan (Zuhdi & Prasetyo, 2025). Rating biasanya ditampilkan dalam bentuk bintang, sedangkan ulasan berupa teks berisi

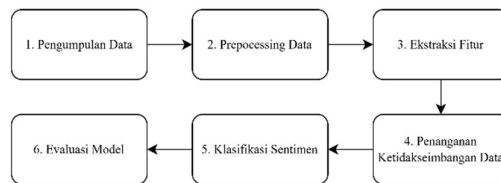
pengalaman, kritik, maupun saran pengguna. Informasi tersebut memiliki peran penting karena tidak hanya menjadi bahan evaluasi bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan fitur, tetapi juga menjadi referensi bagi calon pengguna sebelum memutuskan untuk mengunduh dan menggunakan aplikasi tertentu (Ranjan & Bhaba, 2020). Oleh karena itu, analisis sentimen terhadap ulasan di *Google Play Store* dapat memberikan gambaran objektif mengenai persepsi pengguna secara luas.

Dalam konteks game mobile, khususnya game olahraga seperti sepak bola digital, ulasan pengguna sering kali mencerminkan aspek teknis dan non-teknis, seperti kualitas grafis, stabilitas server, mekanisme *gameplay*, sistem pembelian dalam aplikasi, serta dukungan pengembang. *eFootball 2026* sebagai game sepak bola digital populer memiliki volume ulasan yang sangat besar dan beragam, sehingga analisis sentimen menjadi pendekatan yang tepat untuk mengidentifikasi kecenderungan opini pengguna terhadap kualitas permainan secara keseluruhan.

Metode *machine learning* banyak diterapkan dalam analisis sentimen, di antaranya *Multinomial Naive Bayes* dan *Support Vector Machine*. *Multinomial Naive Bayes* dikenal sebagai algoritma klasifikasi yang sederhana namun efektif untuk data teks, sedangkan *Support Vector Machine* memiliki kemampuan yang baik dalam memisahkan kelas dengan margin optimal (Anggara et al., 2025). Penggunaan kedua metode ini diharapkan mampu menghasilkan klasifikasi sentimen yang akurat serta memberikan wawasan yang bermanfaat bagi pengembang dalam mengevaluasi dan meningkatkan kualitas aplikasi *eFootball 2026* di masa mendatang.

## 2. METODOLOGI PENELITIAN

Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data hingga evaluasi performa model klasifikasi. Alur tahapan penelitian dapat dilihat di gambar 1.



**Gambar 1.** Alur Tahapan Penelitian

### 2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi *eFootball 2026* yang diperoleh dari *Google Play Store*. Pengumpulan data dilakukan menggunakan teknik *web scraping* dengan bantuan bahasa pemrograman *Python* (Syarifudin et al., 2025). Data yang dikumpulkan berupa teks ulasan pengguna yang merepresentasikan opini dan pengalaman pengguna terhadap aplikasi *eFootball 2026*. Dataset ini kemudian digunakan sebagai bahan utama dalam proses analisis sentimen.

## 2.2. Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan data teks agar dapat diolah oleh algoritma *machine learning*. Tahapan *preprocessing* yang diterapkan dalam penelitian ini meliputi:

1. *Text Cleaning*, yaitu menghapus karakter non-alfabet, tanda baca, dan simbol yang tidak relevan.

**Tabel 1.** *Text Cleaning*

Sebelum	Sesudah
Game sangat kompetitif jadi sangat menghibur <a href="https://play.google.com/store/games">https://play.google.com/store/games</a>	Game sangat kompetitif jadi sangat menghibur

2. *Case Folding*, untuk mengubah seluruh huruf menjadi huruf kecil.

**Tabel 2.** *Case Folding*

Sebelum	Sesudah
Game Sangat Kompetitif jadi sangat menghibur	game sangat kompetitif jadi sangat menghibur

3. *Tokenization*, yaitu memecah teks menjadi unit kata.

**Tabel 3.** *Tokenization*

Sebelum	Sesudah
game sangat kompetitif jadi sangat menghibur	"game" "sangat" "kompetitif" "jadi" "sangat" "menghibur"

4. *Stopword Removal*, yaitu menghapus kata-kata umum yang tidak memiliki makna signifikan.

**Tabel 4.** *Stopword Removal*

Sebelum	Sesudah
"game" "sangat" "kompetitif" "jadi" "sangat" "menghibur".	"game" "sangat" "kompetitif" "jadi" "sangat" "menghibur"

5. Normalisasi *slangword*, untuk mengubah kata tidak baku menjadi bentuk baku.

**Tabel 5.** Normalisasi *Slangword*

Sebelum	Sesudah
"game" "kompetitif" "jadi" "sgt" "menghibur".	"game" "kompetitif" "jadi" "sangat" "menghibur"

6. *Stemming*, untuk menghilangkan dan mengembalikan kalimat yang telah diubah ke bentuk aslinya. Penghapusan sufiks dan prefiks menggunakan library sastrawi.

**Tabel 6.** *Stemming*

Sebelum	Sesudah
Tolong di perbaiki untuk pemakaian bantuan cerdas , karena saat musuh menggunakan bantuan cerdas , sering kali jaringan berubah jadi jelek pemain terasa berat, sering patah patah	tolong perbaiki pakai bantu cerdas karena saat musuh guna bantu cerdas sering kali jaring ubah jadi jelek main rasa berat sering patah patah

Tahapan *preprocessing* ini bertujuan untuk mengurangi *noise* pada data dan meningkatkan kualitas fitur yang dihasilkan.

### 2.3. Ekstraksi Fitur

Setelah tahap *preprocessing*, data teks diubah menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kemunculan kata dalam dokumen dan keseluruhan *dataset*, sehingga kata yang bersifat informatif memiliki bobot yang lebih besar dibandingkan kata yang sering muncul namun kurang bermakna.

### 2.4. Penanganan Ketidakseimbangan Data

*Dataset* ulasan pengguna umumnya memiliki ketidakseimbangan jumlah data antar kelas sentimen. Oleh karena itu, pada penelitian ini diterapkan metode *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk menyeimbangkan jumlah data pada kelas minoritas. Penerapan SMOTE dilakukan pada data latih agar model tidak bias terhadap kelas mayoritas.

### 2.5. Klasifikasi Sentimen

Proses klasifikasi sentimen dilakukan menggunakan dua algoritma *machine learning*, yaitu:

1. *Multinomial Naive Bayes*, yang merupakan metode klasifikasi probabilistik dan banyak digunakan pada data teks karena kesederhanaan serta efisiensinya.

2. *Support Vector Machine*, yang bekerja dengan mencari hyperplane optimal untuk memisahkan kelas sentimen positif dan negatif.

Kedua algoritma ini digunakan untuk membandingkan performa klasifikasi sentimen terhadap ulasan aplikasi eFootball 2026.

## 2.6. Evaluasi Model

Tahap evaluasi merupakan tahapan untuk mengetahui kinerja algoritma klasifikasi yang digunakan dalam penelitian ini berdasarkan hasil pengujian pada data uji. Evaluasi dilakukan dengan memanfaatkan *confusion matrix* dan metrik evaluasi yang dihasilkan dari proses pemodelan menggunakan *library scikit-learn*, yaitu *accuracy*, *precision*, *recall*, dan F1-score.

*Confusion matrix* digunakan untuk menggambarkan hasil klasifikasi model ke dalam empat kategori, yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*, sehingga performa model dapat dianalisis secara lebih menyeluruh. Adapun penjelasan metrik evaluasi yang digunakan adalah sebagai berikut:

1. *Accuracy*, digunakan untuk mengukur tingkat ketepatan model dalam memprediksi label sentimen dengan membandingkan jumlah prediksi yang benar terhadap seluruh data uji.
2. *Precision*, digunakan untuk mengetahui tingkat ketepatan prediksi model terhadap kelas positif, yaitu perbandingan antara jumlah prediksi positif yang benar dengan seluruh prediksi positif yang dihasilkan model.
3. *Recall*, digunakan untuk mengukur kemampuan model dalam mengenali seluruh data yang termasuk dalam kelas positif, yaitu perbandingan antara jumlah prediksi positif yang benar dengan seluruh data aktual positif.
4. F1-score, merupakan nilai rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk menilai keseimbangan performa model dalam mengklasifikasikan sentimen.

## 3. HASIL DAN PEMBAHASAN

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu *Accuracy*, *Precision*, *Recall*, dan F1-Score, baik dalam skema *macro average* maupun *weighted average*. Metrik-metrik ini digunakan untuk memberikan gambaran yang lebih komprehensif terhadap kemampuan model, khususnya dalam menangani distribusi kelas yang tidak seimbang.

### 3.1. Scraping Data

Berikut hasil *scraping data* ulasan terbaru menggunakan *python* yang diambil dari *Google Play Store* berjumlah 5000 data:

reviewid	username	review	rating	thumbsUpCount	reviewCreatedVern	tanggal
5483c3-b7c5-4314-a163-c479	Pengguna Google	mantap gamenya . rekomendasi	5	0	1022	14/01/2026 12.47
13a493e-1150-421e-8c8e-9493	Pengguna Google	WOY TOLONG LAH LOGIKA NGAPA SIH KITA KLO LAWAN AI ITU GAK BISA MEN	5	0	1022	14/01/2026 12.37
651183e-79fd-4116-9704-d6d1	Pengguna Google	nice	5	0	1022	14/01/2026 12.36
83ccae8-940e-47d8-9e39-4e1e	Pengguna Google	game gblck jaringan bagus tapi patah patah pas kita yg bawr boia jaringan ny	1	0	1022	14/01/2026 12.32
22c4697e-744b-492b-93c1-ad	Pengguna Google	minta pembaruan trusssss?????	1	0	1021	14/01/2026 12.27
6251741-e837-4884-a095-85f1	Pengguna Google	keren!	5	0	1022	14/01/2026 12.27
9e8505c-330e-4795-9a8c-39	Pengguna Google	Konami minimal tambahn fita kalau sryal kita ada gangguan langsung di pau	1	0	1022	14/01/2026 12.19
1020fa12-38a8-452b-ba2b-11	Pengguna Google	wait rnya kaya anj!	1	0	1011	14/01/2026 12.18
4f3a115-4237-496a-ba36-8b1e	Pengguna Google	Konami stress beri musuh yang susah terus baru ovisi 3 ketemu dikit 1 pacal!	1	0	1022	14/01/2026 12.15
744120e-6c38-4755-9850-49	Pengguna Google	saya suka	5	0	1022	14/01/2026 12.14
9b7a0c7-9493-48d1-a8a5-9f1e	Pengguna Google	coba bikin karir manajer dan karir pemain plir dibaca	5	0	1022	14/01/2026 12.08
5796c07-08e4-46d2-c17a-c5e1	Pengguna Google	Zhuo sekali update sudah ni game gas jiar ni game	1	0	1021	14/01/2026 12.08
4930189-8853-4292-ba5d-4868	Pengguna Google	tolong tambahn mode game world cup, ASEAN cup, dan aling boar lainnya c	4	0	1022	14/01/2026 11.47
ed214e8-4239-481c-ba00-ca0d	Pengguna Google	tolong dilah perbaruan semua hal yang baru seperti patch rda baru, sejal	5	0	1023	14/01/2026 11.47
4547915-4266-4341-84d3-aa5d	Pengguna Google	Konami agride bagusse itu smart assist bodo bikin game mase main aja bodoh l	1	0	1022	14/01/2026 11.46
42364aa-470c-454a-bc72-1e1e	Pengguna Google	sering update terus	4	0	1021	14/01/2026 11.39
7223b72-6972-4c7e-8038-511e	Pengguna Google	sangat bagus	5	0	1022	14/01/2026 11.36
03ae04f7-3192-4da4-bba2-0ca1	Pengguna Google	Semoga di efootball idz var	5	0	1022	14/01/2026 11.33
56e367d5-340c-4008-a995-066e	Pengguna Google	ngleg banget anjir abis mt	1	0	1022	14/01/2026 11.31
56e21be0-536d-4761-b485-7d	Pengguna Google	game lai,	1	0	1022	14/01/2026 11.29
bae7420c-4e5a-40e7-ba43-28	Pengguna Google	game yang bagus	5	0	1022	14/01/2026 11.19

Gambar 2. Hasil Scraping Data

### 3.2. Preprocessing Data

Tahap selanjutnya adalah *preprocessing data*, yang bertujuan untuk membersihkan dan menyiapkan data teks agar dapat diolah oleh algoritma klasifikasi. Proses preprocessing meliputi *case folding*, *tokenizing*, penghapusan *stopword*, dan *stemming*. Tahapan ini sangat penting untuk mengurangi *noise* pada data dan meningkatkan kualitas fitur yang dihasilkan. Semua *preprocessing* data dapat dilihat pada beberapa gambar dibawah ini:

```

1 stop_words = set(stopwords.words('indonesian'))
2
3 def preprocess_text(text):
4     # Case folding
5     text = text.lower()
6
7     # Hapus URL
8     text = re.sub('http\S+', '', text)
9
10    # Hapus angka dan tanda baca
11    text = re.sub('\d+', '', text)
12    text = text.translate(str.maketrans('', '', string.punctuation))
13
14    # Tokenization
15    tokens = word_tokenize(text)
16
17    # Ganti slangword
18    tokens = [slang_dict[word] if word in slang_dict else word for word in tokens]
19
20    # Stopword removal
21    tokens = [word for word in tokens if word not in stop_words]
22
23    return " ".join(tokens)
24
25 df['clean_text'] = df['review'].astype(str).apply(preprocess_text)
26 df[['review', 'clean_text']].head()
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Gambar 3. Tahap Preprocessing

### 3.3. Fitur Ekstraksi

Setelah proses preprocessing, data teks diekstraksi menjadi bentuk numerik menggunakan metode ekstraksi fitur berbasis teks. Representasi fitur ini digunakan sebagai input bagi algoritma Multinomial Naive Bayes dan Support Vector Machine dalam proses klasifikasi sentimen.

```
1 tfidf = TfidfVectorizer(  
2     max_features=5000,  
3     ngram_range=(1,2)  
4 )  
5  
6 X = tfidf.fit_transform(df['clean_text'])  
7 y = df['label']  
8
```

**Gambar 4.** Fitur TD-IDF

Dengan pendekatan ini, kata-kata yang sering muncul namun kurang informatif akan memiliki bobot yang lebih rendah, sedangkan kata-kata yang bersifat lebih spesifik terhadap sentimen tertentu akan memiliki bobot yang lebih tinggi.

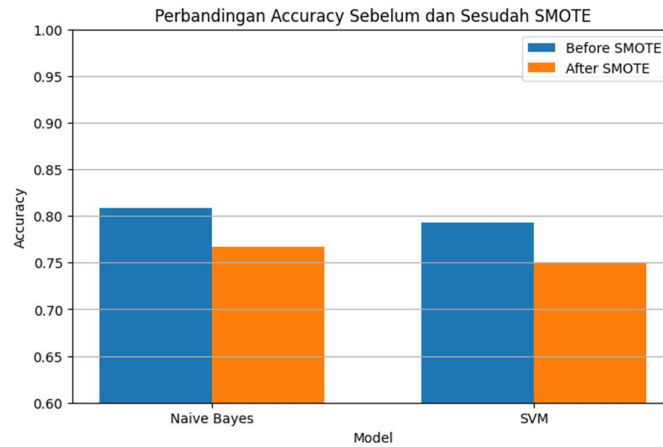
### 3.4. Penanganan Ketidakseimbangan Data

*Dataset* ulasan cenderung memiliki distribusi kelas yang tidak seimbang antara sentimen positif dan negatif. Oleh karena itu, diterapkan metode SMOTE pada data latih untuk menyeimbangkan jumlah data pada masing-masing kelas dan mencegah bias model terhadap kelas mayoritas.

```
1 # ===== SEBELUM SMOTE =====  
2  
3 # Naive Bayes  
4 nb_before = MultinomialNB()  
5 nb_before.fit(X_train, y_train)  
6 y_pred_nb_before = nb_before.predict(X_test)  
7  
8 # SVM  
9 svm_before = LinearSVC()  
10 svm_before.fit(X_train, y_train)  
11 y_pred_svm_before = svm_before.predict(X_test)  
12  
  
1 # ===== SESUDAH SMOTE =====  
2  
3 # Naive Bayes  
4 nb_after = MultinomialNB()  
5 nb_after.fit(X_train_smote, y_train_smote)  
6 y_pred_nb_after = nb_after.predict(X_test)  
7  
8 # SVM  
9 svm_after = LinearSVC()  
10 svm_after.fit(X_train_smote, y_train_smote)  
11 y_pred_svm_after = svm_after.predict(X_test)
```

**Gambar 5.** SMOTE

Dan hasil perbandingan dari sebelum dan sesudah menggunakan metode SMOTE sebagai berikut:



**Gambar 6.** Perbandingan Sebelum dan Sesudah SMOTE

### 3.5. Klasifikasi Sentimen

Tahap klasifikasi sentimen dilakukan dengan menggunakan dua algoritma *machine learning*, yaitu *Multinomial Naive Bayes* dan SVM. Kedua algoritma dilatih menggunakan data latih yang telah diproses untuk mengklasifikasikan ulasan ke dalam kelas sentimen positif dan negatif.

```

1 nb = MultinomialNB()
2 nb.fit(X_train_smote, y_train_smote)
3
4 y_pred_nb = nb.predict(X_test)
5
6 print("=== Naive Bayes ===")
7 print("Accuracy:", accuracy_score(y_test, y_pred_nb))
8 print(confusion_matrix(y_test, y_pred_nb))
9 print(classification_report(y_test, y_pred_nb))
10
*** === Naive Bayes ===
Accuracy: 0.767268862911796
[[250  70]
 [149 472]]
      precision    recall  f1-score   support

     0       0.63     0.78     0.70     320
     1       0.87     0.76     0.81     621

   accuracy          0.77     941
  macro avg          0.75     941
 weighted avg          0.79     941
    
```

**Gambar 7.** Hasil Menggunakan *Multinomial Naive Bayes*

```

1 svm = LinearSVC()
2 svm.fit(X_train_smote, y_train_smote)
3
4 y_pred_svm = svm.predict(X_test)
5
6 print("=== SVM ===")
7 print("Accuracy:", accuracy_score(y_test, y_pred_svm))
8 print(confusion_matrix(y_test, y_pred_svm))
9 print(classification_report(y_test, y_pred_svm))
10

```

```

... === SVM ===
Accuracy: 0.7492029755579172
[[237  83]
 [153 468]]

```

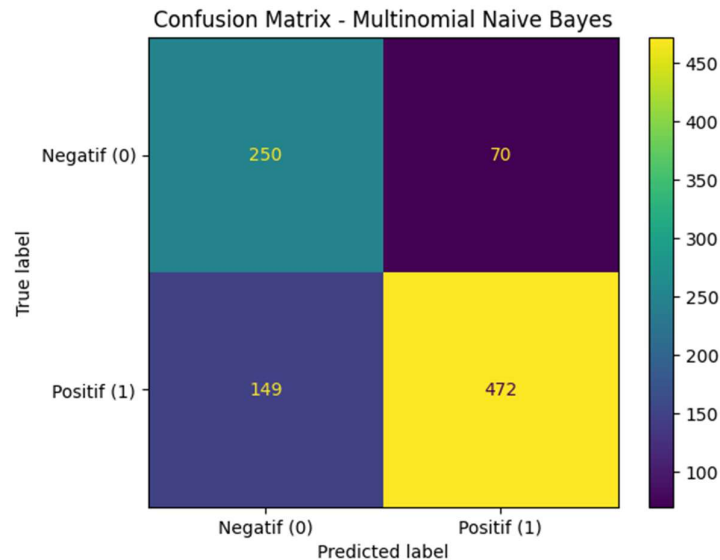
	precision	recall	f1-score	support
0	0.61	0.74	0.67	320
1	0.85	0.75	0.80	621
accuracy			0.75	941
macro avg	0.73	0.75	0.73	941
weighted avg	0.77	0.75	0.75	941

**Gambar 8.** Hasil Menggunakan SVM

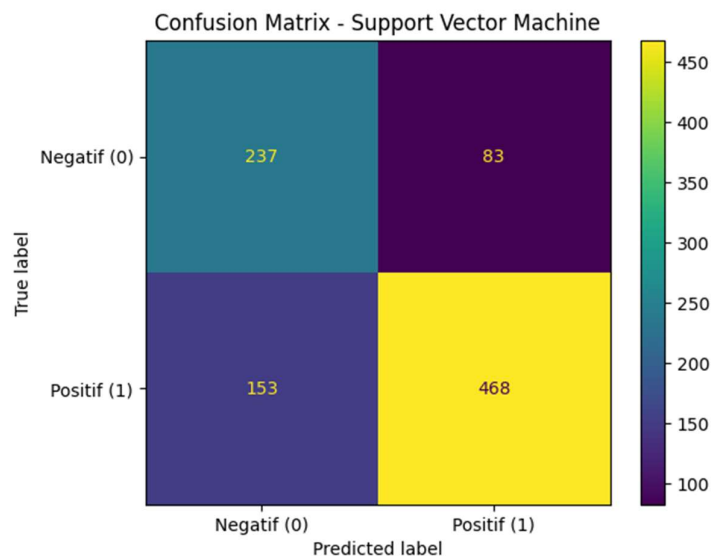
### 3.6. Evaluasi Model

Hasil evaluasi digunakan untuk membandingkan kinerja kedua algoritma serta menentukan model yang memiliki performa terbaik dalam analisis sentimen ulasan pengguna aplikasi eFootball 2026.

Berdasarkan hasil pengujian yang telah dilakukan, diperoleh nilai evaluasi untuk masing-masing algoritma klasifikasi sentimen, yaitu *Multinomial Naive Bayes* dan *Support Vector Machine*.



**Gambar 9.** Confussion Matrix Naive Bayes



**Gambar 10.** *Confussion Matrix SVM*

menampilkan hasil pengujian model berupa nilai accuracy, precision, recall, dan F1-score yang digunakan sebagai dasar analisis kinerja masing-masing algoritma.

Berdasarkan hasil tersebut, diperoleh ringkasan nilai evaluasi seperti yang ditunjukkan pada Tabel 3.1

**Tabel 3.1** Hasil Evaluasi Model

Model	Naive Bayes	SVM
<b>Accuracy</b>	0.767269	0.749203
<b>Precision (Macro)</b>	0.748708	0.728529
<b>Recall (Macro)</b>	0.770657	0.747124
<b>F1-Score (Macro)</b>	0.753552	0.733120
<b>Precision (Weighted)</b>	0.787777	0.767181
<b>Recall (Weighted)</b>	0.767269	0.749203
<b>F1-Score (Weighted)</b>	0.772150	0.754077

Berdasarkan tabel hasil evaluasi, model *Multinomial Naive Bayes* menunjukkan performa yang lebih unggul dibandingkan SVM pada seluruh metrik pengujian. Nilai *accuracy Multinomial Naive Bayes* sebesar 0,767269 lebih tinggi dibandingkan SVM sebesar 0,749203, yang menunjukkan bahwa *Naive Bayes* memiliki tingkat ketepatan prediksi yang lebih baik terhadap data uji.

Pada metrik *precision (macro)* dan *recall (macro)*, *Multinomial Naive Bayes* juga memperoleh nilai yang lebih tinggi, masing-masing sebesar 0,748708 dan 0,770657, dibandingkan dengan SVM yang memperoleh nilai 0,728529 dan 0,747124. Hal ini menunjukkan bahwa *Multinomial Naive Bayes* memiliki kemampuan yang lebih baik dalam mengklasifikasikan kedua kelas sentimen secara seimbang tanpa dipengaruhi oleh distribusi kelas.

Nilai F1-score (macro) pada *Multinomial Naive Bayes* sebesar 0,753552 juga lebih tinggi dibandingkan SVM sebesar 0,733120, yang menandakan bahwa keseimbangan antara *precision* dan *recall* pada model *Multinomial Naive Bayes* lebih optimal.

Selain itu, pada metrik weighted average, *Multinomial Naive Bayes* kembali menunjukkan performa yang lebih baik dengan nilai *precision*, *recall*, dan F1-score masing-masing sebesar 0,787777, 0,767269, dan 0,772150. Hasil ini menunjukkan bahwa model *Multinomial Naive Bayes* lebih stabil dalam menangani distribusi kelas pada *dataset* ulasan pengguna aplikasi eFootball 2026.

#### **4. KESIMPULAN**

Pada studi ini, *dataset* ulasan aplikasi dianalisis menggunakan TF-IDF untuk ekstraksi fitur dan dibandingkan performa *Multinomial Naive Bayes* dan *Support Vector Machine*, menunjukkan bahwa metode tersebut efektif untuk mengklasifikasikan sentimen pada ulasan aplikasi mobile (Rhomaningtias & al., 2025). Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap ulasan pengguna aplikasi eFootball 2026 di *Google Play Store* berhasil dilakukan menggunakan algoritma *Multinomial Naive Bayes* dan *Support Vector Machine*. Tahapan penelitian yang meliputi preprocessing data, ekstraksi fitur menggunakan TF-IDF, penanganan ketidakseimbangan data dengan SMOTE, serta proses klasifikasi sentimen terbukti mampu menghasilkan model dengan performa yang baik (Dwi et al., 2024).

Hasil evaluasi menunjukkan bahwa algoritma *Multinomial Naive Bayes* memberikan performa yang lebih unggul dibandingkan SVM pada seluruh metrik pengujian. *Multinomial Naive Bayes* memperoleh nilai *accuracy* sebesar 76,72%, dengan nilai *precision*, *recall*, dan F1-score yang lebih tinggi baik pada pendekatan macro maupun *weighted average*. Hal ini menunjukkan bahwa *Multinomial Naive Bayes* lebih efektif dan stabil dalam mengklasifikasikan sentimen ulasan pengguna pada *dataset* yang digunakan.

Sementara itu, SVM juga menunjukkan performa yang cukup baik dengan nilai akurasi sebesar 74,92%, namun sedikit lebih rendah dibandingkan *Multinomial Naive Bayes*. Perbedaan performa ini dipengaruhi oleh karakteristik data teks serta kesesuaian algoritma terhadap representasi fitur TF-IDF.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan *machine learning* dapat digunakan secara efektif untuk menganalisis sentimen ulasan pengguna aplikasi eFootball 2026 di *Google Play Store*. Hasil analisis sentimen ini dapat menjadi bahan evaluasi bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan pengalaman pengguna. Penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih besar, menambahkan kelas sentimen netral, serta mencoba metode berbasis *deep learning* guna meningkatkan performa klasifikasi.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

## REFERENSI

- Amirkhalili, Y., & Wong, H. Y. (2024). *Banking on Feedback: Text Analysis of Mobile Banking iOS and Google App Reviews*.
- Anggara, M. B., Informasi, F. T., & Bandung, U. B. (2025). *PERBANDINGAN NAÏVE BAYES DAN SVM DALAM ANALISIS SENTIMEN*. 20, 32–42.
- Dwi, P., Maulidiana, R., Vitianingsih, A. V., Kacung, S., & Maukar, A. L. (2024). *Comparative Analysis of SVM and NB Algorithms in Evaluating Public Sentiment on Supreme Court Rulings*. 13, 230–236.
- Jurnal, S., Informasi, S., Tika, Y., Sholiha, R., Asjad, L., Nabilah, M., Teknik, F., Informasi, S., Bina, U., Informatika, S., & Tangerang, K. (2025). *Analisis Sentimen Aplikasi Liputan6 . Com pada Ulasan Pengguna di Google Playstore dengan Menggunakan Algoritma Support Vector Machine ( Svm ) dan Naive Bayes*.
- Lundam, I. S., & Darmawan, J. B. B. (2024). *Analisis Sentimen Ulasan Aplikasi Gojek Di Google Play Store Menggunakan Multinomial Naive Bayes*. 2024(November), 149–156.
- Nawulansih, D. F., Santi, N. C., Aristia, I., Informatika, T., Nahdlatul, U., Sunan, U., Bojonegoro, G., Informasi, S., Nahdlatul, U., Sunan, U., & Bojonegoro, G. (2025). *Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store: Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique Sentiment Analysis of DANA Application Reviews on Google Play Store: Implementation of Support Vector Machine an*. 5(9), 2660–2671.
- Nouval, M., Habibi, F. M., Rahmi, A., Dawam, M., Bittaqwa, A., Agustianto, R., & Hasan, F. N. (2026). *Analisis Sentimen Ulasan Aplikasi Maxim di Google Play Store Menggunakan Algoritma Support Vector Machine*.
- Nurian, A., Sari, B. N., Karawang, U. S., & Timur, T. (2023). *Analisis sentimen ulasan*

- pengguna aplikasi google play menggunakan naïve bayes. 11(3), 829–835.*
- R, S. M. M., & Effendy, Y. S. (2025). *Analisis Sentimen Ulasan Game Simulator Indonesia di Google Play Store Menggunakan Algoritma Naive Bayes. 7(2), 169–178.* <https://doi.org/10.35580/variansiunm336>
- Ranjan, S., & Bhaba, H. (2020). *Comparative Sentiment Analysis of App Reviews. 1–10.*
- Rhomaningtias, L., & al., et. (2025). Analisis Sentimen Ulasan Aplikasi SMILE Indonesia Menggunakan Naïve Bayes dan SVM. *HOAQ: Jurnal Teknologi Informasi, 4(November), 63–72.* <https://publikasi.uyelindo.ac.id>
- Sarimole, F. M. (2024). *Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine. 5(3), 783–790.*
- Setiawan, N. R., & Kaburuan, E. R. (2023). *Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM. 12, 105–116.*
- Syaifudin, A., Rafsanjani, A., Fithri, D. L., Studi, P., Informasi, S., Teknik, F., & Kudus, U. M. (2025). *Analisis Sentimen Ulasan Pengguna Aplikasi KitaLulus pada Google Play Store dengan menggunakan algoritma Support Vector Machine ( SVM ) Sentiment Analysis of User Reviews of the KitaLulus Application on Google Play Store using the Support Vector Machine (. 14(Cv), 2519–2530.*
- Zuhdi, H. N., & Prasetyo, B. (2025). *Sentiment Analysis On Ipusnas Application Reviews In Google Play Store Using Naive Bayes Classifier Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier. 5(1), 12–19.*