

ANALISIS SENTIMEN ULASAN APLIKASI SPEEDCASH DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Oleh :
Naufal Ainurrochman

Program Studi Informatika
Universitas Muhammadiyah Sidoarjo
2026



Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang semakin pesat telah mendorong pertumbuhan layanan keuangan digital di Indonesia. Kemajuan teknologi tersebut memberikan kemudahan bagi masyarakat dalam melakukan berbagai aktivitas transaksi secara cepat, praktis, dan efisien melalui aplikasi berbasis mobile. Aplikasi keuangan digital tidak hanya berfungsi sebagai alat transaksi, tetapi juga menjadi bagian penting dalam mendukung aktivitas ekonomi masyarakat di era digital saat ini.

Salah satu aplikasi layanan keuangan digital yang cukup dikenal adalah . Aplikasi ini menyediakan berbagai layanan transaksi seperti pembayaran tagihan, pembelian pulsa, transfer uang, dan layanan pembayaran lainnya. Sebagai aplikasi yang tersedia di Google Play Store, SpeedCash menerima berbagai ulasan dan penilaian dari pengguna. Ulasan tersebut berisi pengalaman pengguna terhadap kualitas layanan aplikasi, baik berupa komentar positif maupun negatif.

Jumlah data yang besar tentu sulit dianalisis secara manual, sehingga diperlukan teknik analisis sentimen untuk mengelompokkan opini masyarakat menjadi sentimen positif, netral, dan negatif. Dalam penelitian ini, saya membandingkan dua metode klasifikasi yang banyak digunakan dalam analisis sentimen, yaitu Naïve Bayes Multinomial dan Support Vector Machine linear. Perbandingan ini dilakukan karena performa kedua metode dapat berbeda pada karakteristik data yang berbeda. Melalui penelitian ini diharapkan dapat diketahui metode yang memberikan hasil terbaik sekaligus memberikan gambaran mengenai persepsi masyarakat terhadap kualitas aplikasi SpeedCash.

Rumusan masalah

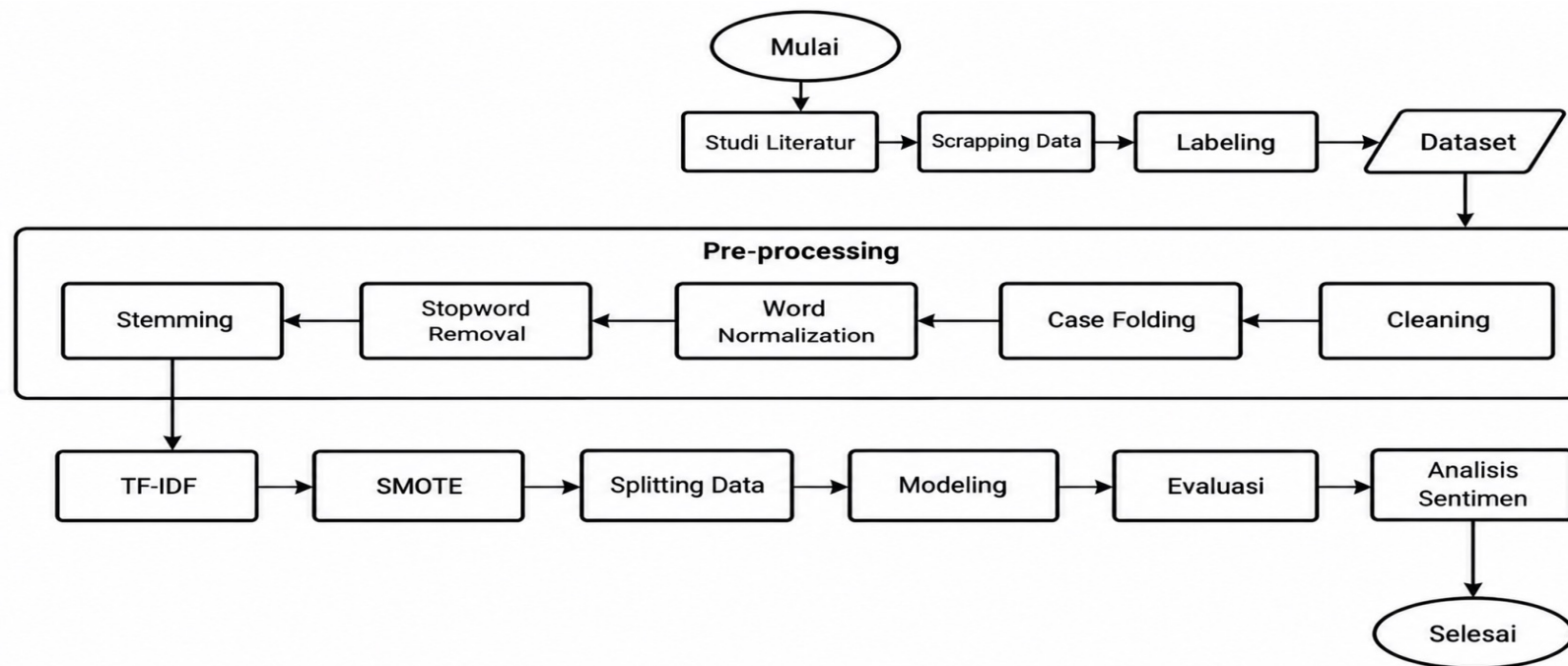
1. Bagaimana proses pengumpulan dan pengolahan data ulasan yang berkaitan dengan aplikasi SpeedCash pada platform Google Play Store
2. Bagaimana implementasi metode Naïve Bayes dan Support Vector Machine (SVM) guna mengklasifikasikan sentimen ulasan pengguna pada pelayanan aplikasi SpeedCash?
3. Bagaimana komparasi kinerja metode Naïve Bayes dan Support Vector Machine (SVM) guna melakukan analisis sentimen ulasan aplikasi SpeedCash berdasarkan metrik evaluasi akurasi, *precision*, *recall*, dan *f1-score*?
4. Metode klasifikasi manakah yang menunjukkan performa paling optimal dalam mengklasifikasikan sentimen pengguna terhadap kualitas pelayanan aplikasi SpeedCash?

Distribusi Sentimen

Total 30.967 data dan dilakukan labelling otomatis menggunakan lexicon yang memiliki 11.929 Positif, 3.718 Negatif, 15.319 Netral terlabel secara otomatis, kemudian dilakukanya preprocessing melalui case folding, cleansing, tokenizing, stopwords removal, stemming, kemudian split data, kemudian pembobotan ke numerik guna TF-IDF (Term Frequency-Inverse Document Frequency), dan pelatihan Naïve Bayes dan SVM, serta evaluasi dan perbandingan Accuracy, F1-Score, Confusion Matrix.

Alur penelitian

Pada penelitian ini, proses dimulai dari scraping data tweet, kemudian dilakukan preprocessing, pelabelan sentimen, pembobotan TF-IDF, pembagian data latih dan data uji, serta oversampling pada data latih. Selanjutnya dilakukan pelatihan menggunakan metode Naïve Bayes dan SVM, kemudian dievaluasi menggunakan Accuracy, Precision, Recall, dan F1-Score untuk mengetahui metode dengan performa terbaik



Split Data

- Dataset dibagi menjadi data latih dan data uji menggunakan tiga skema, yaitu 80:20, 70:30, dan 60:40.
- Tujuan penggunaan tiga skema adalah untuk menganalisis pengaruh proporsi data latih terhadap performa model klasifikasi.
- Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model pada data yang belum pernah dilihat sebelumnya.
- Untuk mengatasi ketidakseimbangan kelas (imbalanced data), diterapkan teknik Oversampling berbasis SMOTE.
- Oversampling hanya diterapkan pada data latih setelah proses splitting, sedangkan data uji tetap dibiarkan asli.
- Langkah ini bertujuan menjaga keseimbangan distribusi kelas selama pelatihan dan mencegah terjadinya data leakage pada proses pengujian.

Split Data SVM

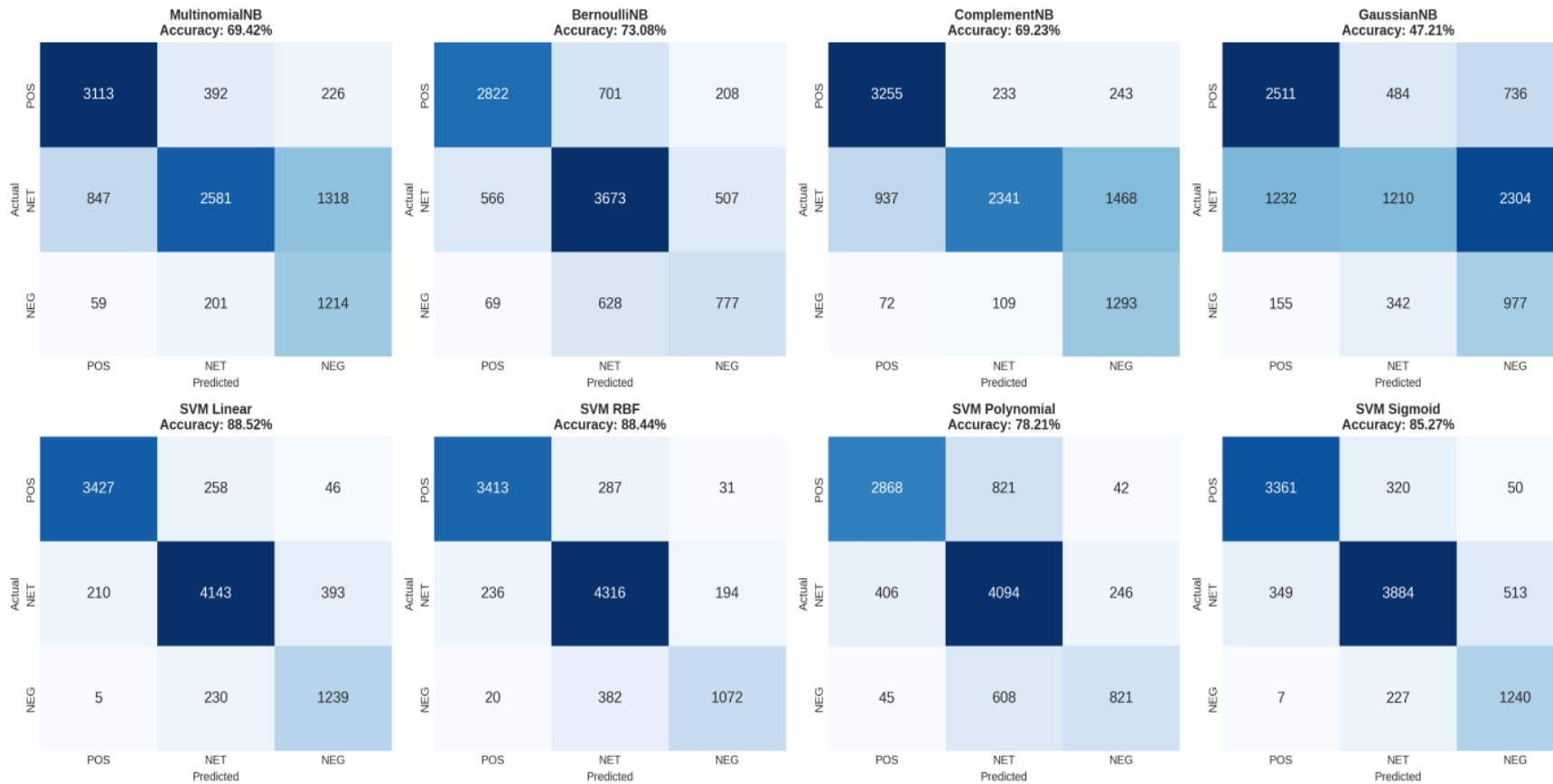
Rasio	Model	Accuracy	Precision	Recall	F1-Score
60:40	SVM Linear	88,52%	88,89%	88,52%	88,64%
60:40	SVM RBF	88,44%	88,41%	88,44%	88,35%
60:40	SVM Sigmoid	85,27%	85,90%	85,27%	85,42%
60:40	SVM Polynomial	78,21%	78,72%	78,21%	77,95%
70:30	SVM Linear	88,85%	89,16%	89,16%	88,96%
70:30	SVM RBF	89,07%	89,03%	89,07%	88,99%
70:30	SVM Sigmoid	85,54%	86,12%	85,54%	85,69%
70:30	SVM Polynomial	78,71%	79,20%	78,71%	78,49%
80:20	SVM Linear	88,44%	88,78%	88,44%	88,56%
80:20	SVM RBF	88,93%	88,89%	88,93%	88,85%
80:20	SVM Sigmoid	84,69%	85,37%	84,69%	84,87%
80:20	SVM Polynomial	78,50%	79,17%	78,50%	78,30%

Split Data Naive Bayes

Rasio	Model	Accuracy	Precision	Recall	F1-Score
60:40	Bernoulli NB	73,08%	73,34%	73,08%	73,14%
60:40	Multinomial NB	69,42%	74,34%	69,42%	69,70%
60:40	Complement NB	69,23%	76,61%	69,23%	69,14%
60:40	Gaussian NB	47,21%	56,10%	44,21%	46,70%
70:30	Bernoulli NB	73,47%	73,77%	73,47%	73,57%
70:30	Multinomial NB	69,40%	74,57%	69,40%	69,76%
70:30	Complement NB	69,07%	76,95%	69,07%	68,99%
70:30	Gaussian NB	46,90%	56,73%	46,90%	46,64%
80:20	Bernoulli NB	73,47%	73,83%	73,47%	73,61%
80:20	Multinomial NB	69,09%	76,66%	69,09%	69,53%
80:20	Complement NB	68,91%	76,66%	68,91%	68,85%
80:20	Gaussian NB	46,40%	56,19%	46,40%	46,15%

Confusion Matix Naïve Bayes dan SVM 60% Latih : 40 % Uji

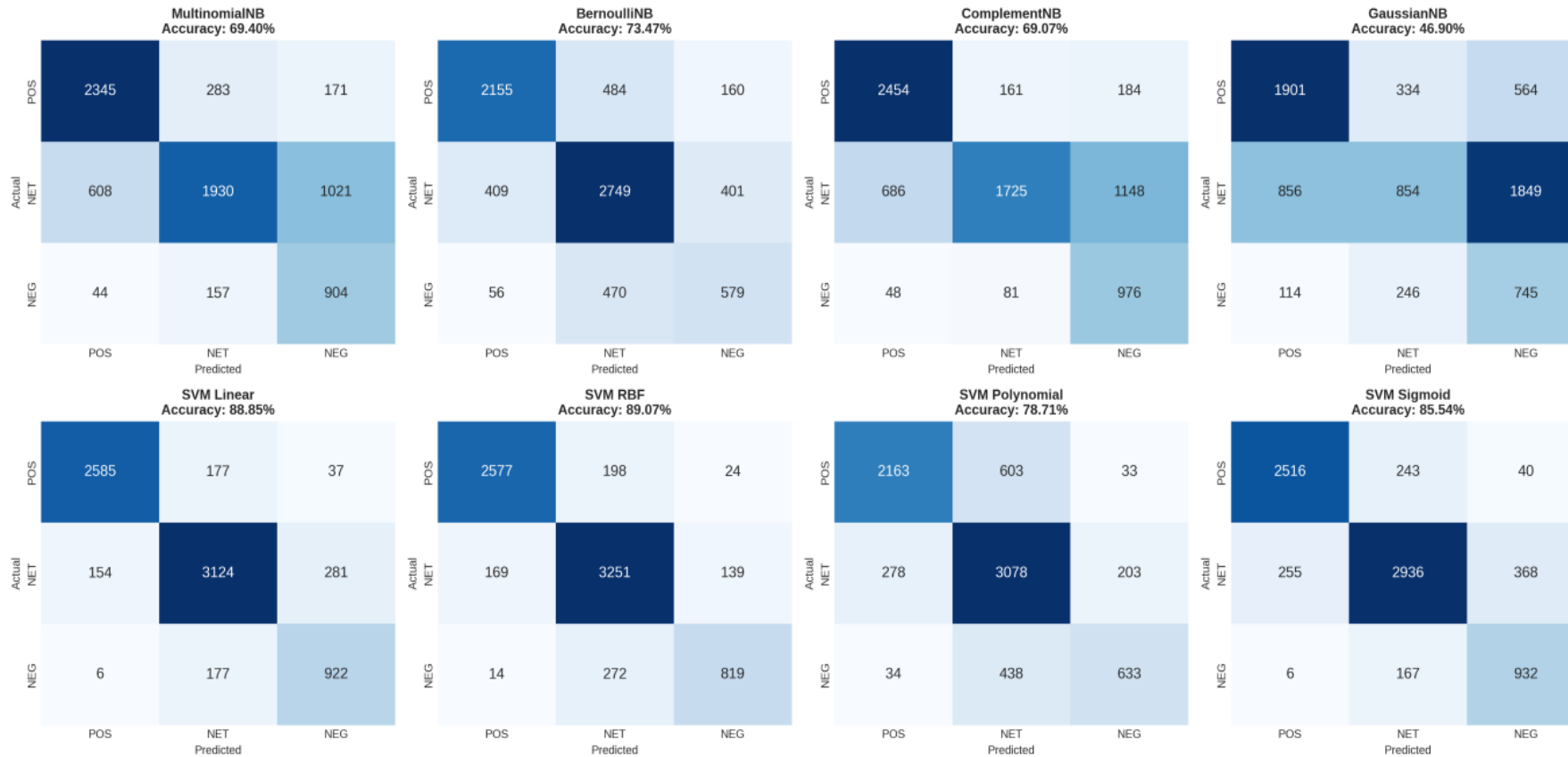
Confusion Matrix 8 Model



Confusion matrix menunjukkan bahwa SVM Linear paling baik mengenali kelas Positif (3.427 dari 3.731 data uji diklasifikasikan dengan benar) dan kelas Netral (4.143 dari 4.746 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 1.239 dari 1.474 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (2.822 dari 3.731 data uji diklasifikasi dengan benar) dan kelas Netral (3.673 dari 4.746 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 777 dari 1.474 data yang diklasifikasikan dengan benar dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral.

Confusion Matix Naïve Bayes dan SVM 70% Latih : 30 % Uji

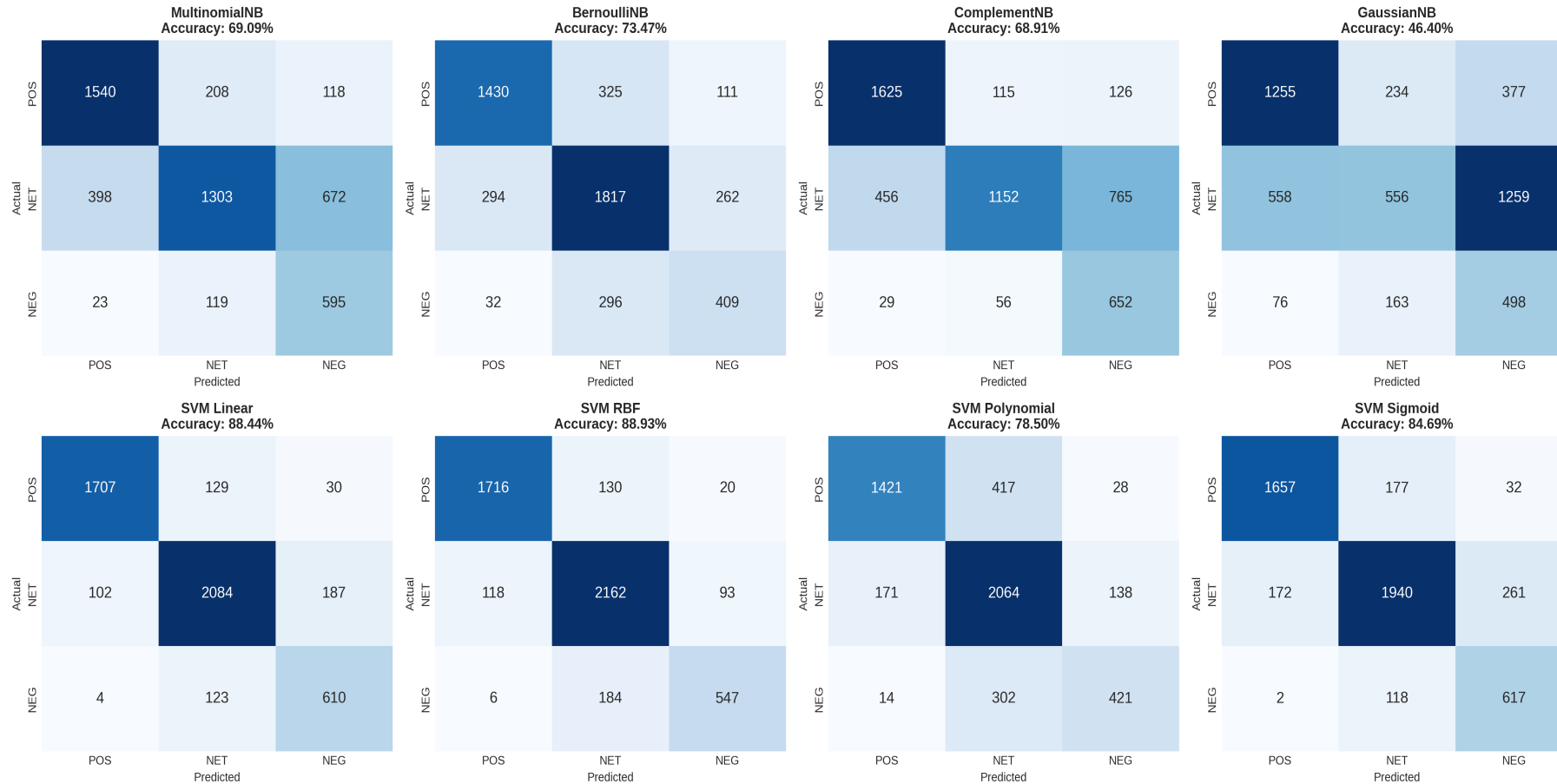
Confusion Matrix 8 Model



diklasifikasikan dengan benar) dan kelas Netral (3.251 dari 3.559 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 819 dari 1.105 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (2.155 dari 2.799 data uji diklasifikasi dengan benar) dan kelas Netral (2.749 dari 3.559 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 579 dari 1.105 data yang diklasifikasikan dengan benar dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral.

Confusion Matix Naïve Bayes dan SVM 60% Latih : 40 % Uji

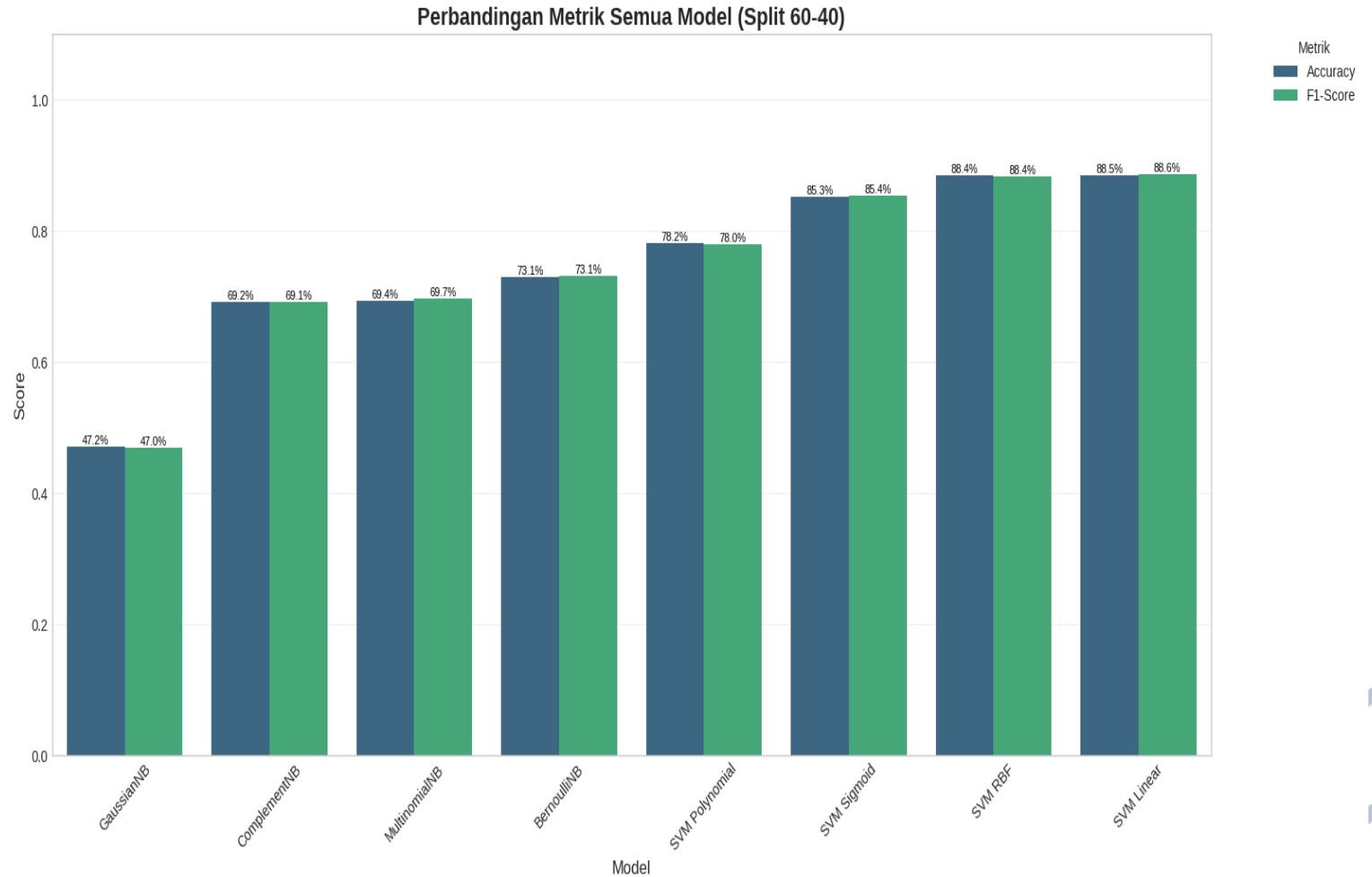
Confusion Matrix 8 Model



Confusion matrix menunjukkan bahwa SVM RBF paling baik mengenali kelas Positif (1.716 dari 1.866 data uji diklasifikasikan dengan benar) dan kelas Netral (2.162 dari 2.373 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 547 dari 737 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (1.430 dari 1.866 data uji diklasifikasi dengan benar) dan kelas Netral (1.817 dari 2.373 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 409 dari 737 data yang diklasifikasikan dengan benar dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral.

Perbandingan F1-Score dan Accuracy Naïve Bayes dan SVM 60% Latih :40% Uji

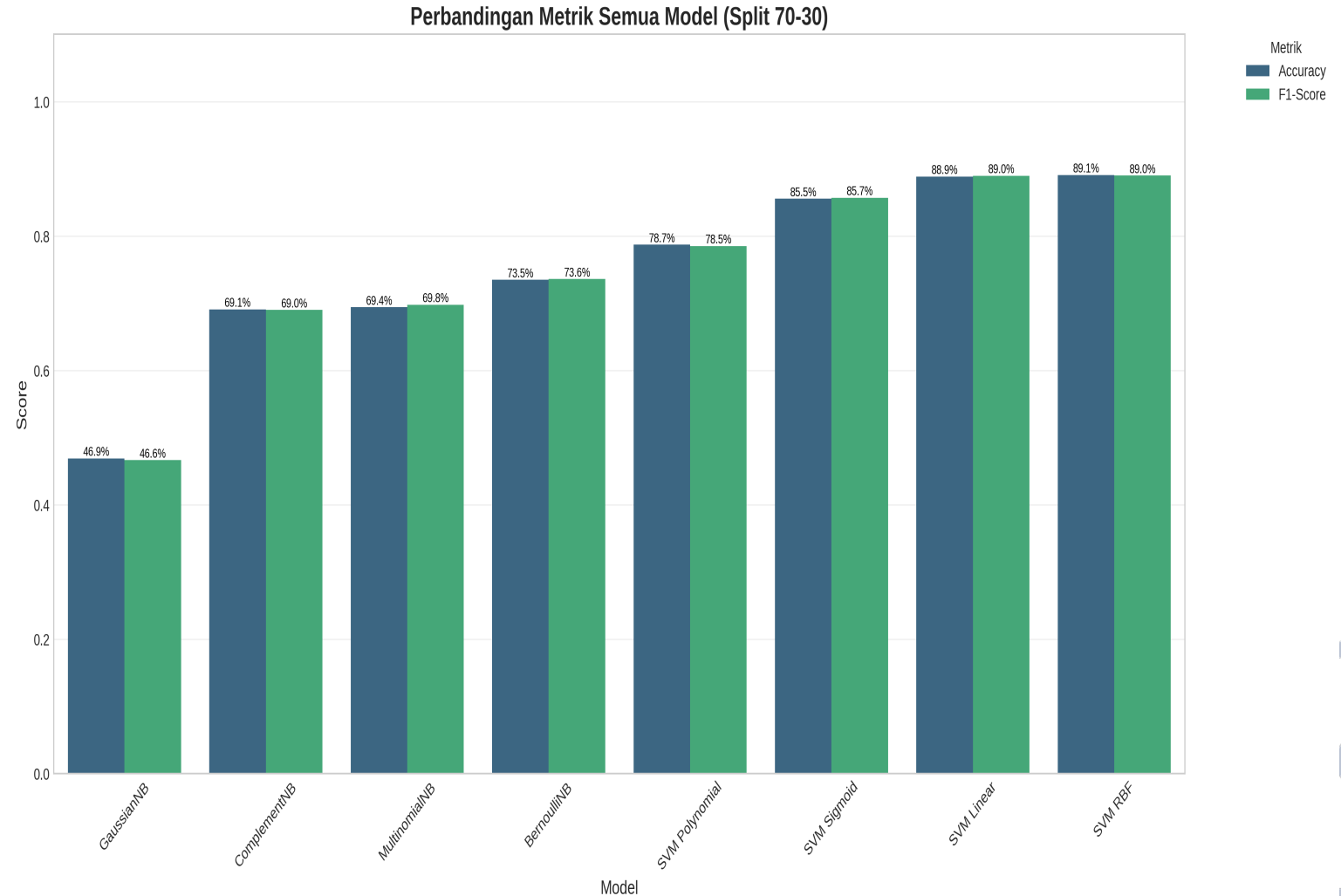
seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (Linear) memperoleh F1-Score tertinggi sebesar 88,64% sekaligus Accuracy tertinggi sebesar 88,52%. SVM (RBF) menghasilkan F1-Score sebesar 88,35% dan Accuracy sebesar 88,44%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 85,42% dan Accuracy sebesar 85,27%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 77,95% dan Accuracy sebesar 78,21%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,14% dan Accuracy sebesar 73,08%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,70% dan Accuracy sebesar 69,42%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 69,14% dan Accuracy sebesar 69,23%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,97% dan Accuracy sebesar 47,21%. Berdasarkan confusion matrix



Perbandingan F1-Score dan Accuracy Naïve Bayes dan SVM

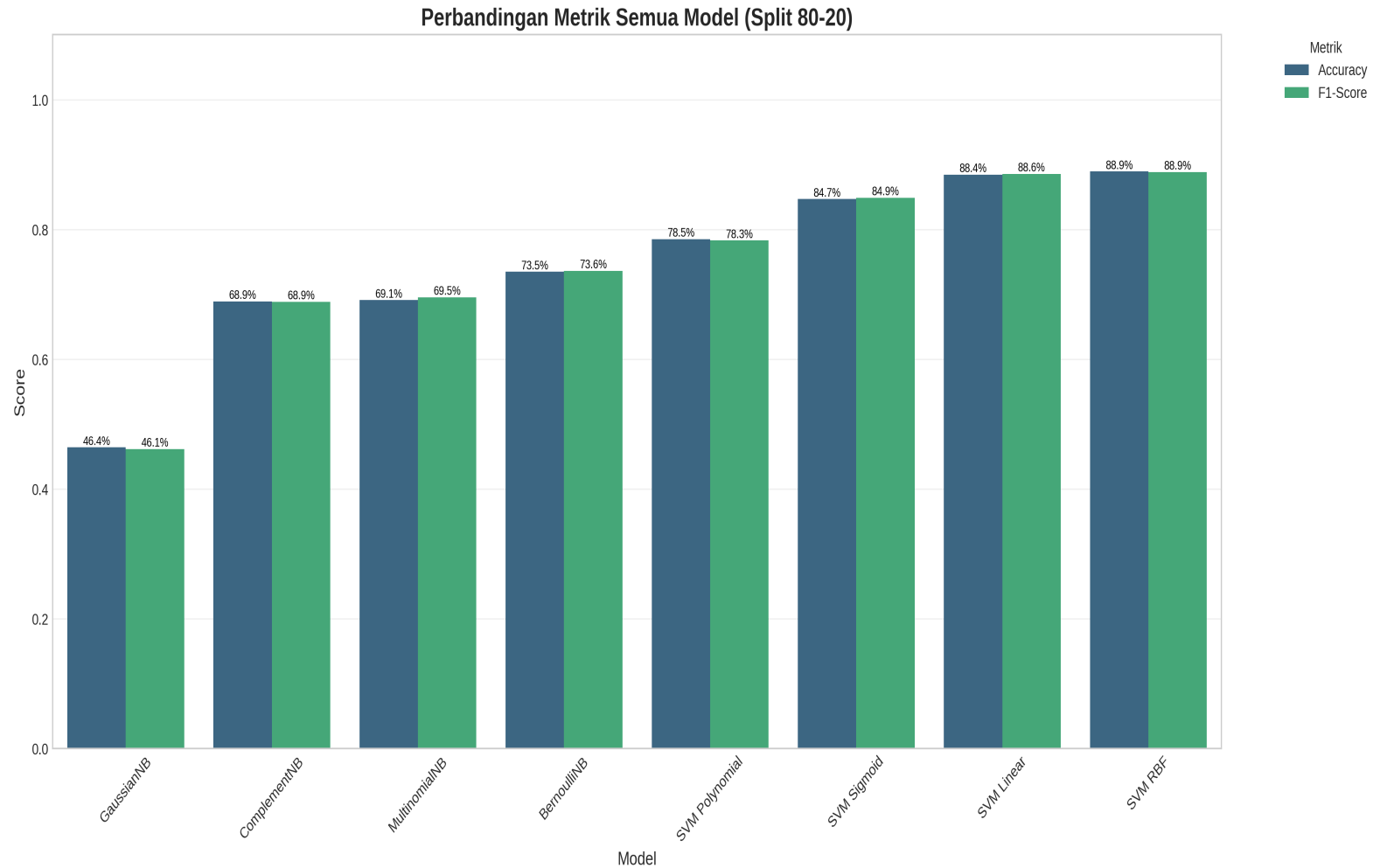
70% Latih :30% Uji

seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (RBF) memperoleh F1-Score tertinggi sebesar 88,99% sekaligus Accuracy tertinggi sebesar 89,07%. SVM (Linear) menghasilkan F1-Score sebesar 88,96% dan Accuracy sebesar 88,85%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 85,69% dan Accuracy sebesar 85,54%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 78,49% dan Accuracy sebesar 78,71%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,57% dan Accuracy sebesar 73,47%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,76% dan Accuracy sebesar 69,40%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 68,99% dan Accuracy sebesar 69,07%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,64% dan Accuracy sebesar 46,90%. Berdasarkan confusion matrix



Perbandingan F1-Score dan Accuracy Naïve Bayes dan SVM 80% Latih :20% Uji

seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (RBF) memperoleh F1-Score tertinggi sebesar 88,85% sekaligus Accuracy tertinggi sebesar 88,93%. SVM (Linear) menghasilkan F1-Score sebesar 88,56% dan Accuracy sebesar 88,44%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 84,87% dan Accuracy sebesar 84,69%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 78,30% dan Accuracy sebesar 78,50%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,61% dan Accuracy sebesar 73,47%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,53% dan Accuracy sebesar 69,09%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 68,85% dan Accuracy sebesar 68,91%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,15% dan Accuracy sebesar 46,40%. Berdasarkan confusion matrix





- 
- 15

Kesimpulan

Hasil klasifikasi analisis sentiment pengguna aplikasi speedcash cukup beragam. Sesuai data yang ada ulasan Netral mendominasi ulasan yang ada pada google play store. Dengan 11.864 ulasan lebih pengguna aplikasi speedcash berkomentar Netral, dan terdapat 9.328 ulasan lebih pengguna yang berkomentar Positif, serta lebih dari 3.684 ulasan berkomentar negatif. Pengujian dilakukan terhadap delapan model, yaitu empat kernel Support Vector Machine (SVM) dan empat varian Naïve Bayes, pada tiga rasio pembagian data (60%:40%, 70%:30%, dan 80%:20%) dengan menerapkan Synthetic Minority Over-sampling Technique (SMOTE) pada data latih. Hasil penelitian menunjukkan bahwa metode SVM secara konsisten memberikan kinerja yang lebih baik dibandingkan Naïve Bayes pada rasio pengujian (60%:40%), Performa terbaik diperoleh SVM Linear dengan F1-Score sebesar 88,64% dan Accuracy 88,52% , sedangkan untuk rasio (70%:30% dan 80%:20%) di dominasi SVM RBF dengan F1-Score tertinggi yaitu 88,99% dan Accuracy 89,07%

