



ANALISIS SENTIMEN PADA ULASAN APLIKASI BIBIT DI PLAY STORE DENGAN METODE NAÏVE BAYES, SUPPORT VECTOR MACHINE, C4.5 DAN K-NEAREST NEIGHBOR

¹Antika Zahrotul Kamalia, ²Andi Al Zaroni, ³Miftah Wangsadanureja

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

antika.kamalia@pelitabangsa.ac.id

Abstrak

Tujuan penelitian ini mencoba menganalisis sentimen yang ada pada ulasan aplikasi bibit dengan menerapkan metode Naïve Bayes, Support Vector Machine (SVM), C4.5, K-Nearest Neighbor (KNN). Hasil nilai accuracy, recall, precision akan menjadi acuan dalam menentukan metode mana yang paling tepat dalam menganalisis sentimen aplikasi bibit. Dalam penelitian ini menguji 464 dataset yang terbagi menjadi 310 data sentimen positif dan 154 sentimen negatif. Evaluasi model menggunakan 4 Fold Cross Validation, dan pengujian pertama tidak menggunakan Particle Swarm Optimization dan yang kedua menggunakan Particle Swarm Optimization. Dalam penelitian ini algoritma naïve bayes memiliki nilai accuracy, recall dan precision tertinggi dalam melakukan analisis sentimen ulasan aplikasi bibit..

Kata Kunci: Analisa Sentimen, Naïve Bayes, Support Vector Machine(SVM), C4.5, K-Nearest Neighbor(KNN), Particle Swarm Optimization, Bibit..

Abstract

This study tries to analyze the sentiments that exist in the Bibit application review by applying the Naïve Bayes, Support Vector Machine (SVM), C4.5, K-Nearest Neighbor (KNN) methods. The results of the accuracy, recall, precision values will be a reference in determining which method is the most appropriate in analyzing the sentiment of Bibit applications. In this study, we tested 464 datasets which were divided into 310 positive sentiment data and 154 negative sentiment data. The evaluation of the model uses 4 Fold Cross Validation, and the first test does not use Particle Swarm Optimization and the second uses Particle Swarm Optimization. In this study, the Naïve Bayes algorithm has the highest accuracy, recall and precision values in analyzing the sentiment analysis of the Bibit application.

Keywords : *Sentiment analysis, Naïve Bayes, Support Vector Machine(SVM), C4.5, K-Nearest Neighbor(KNN), Particle Swarm Optimization, Bibit.*

1. Pendahuluan

Investasi menjadi kegiatan yang mulai di minati masyarakat, hal ini terbukti dengan munculnya aplikasi – aplikasi investasi salah satunya adalah Bibit. Bibit adalah aplikasi reksadana populer di Indonesia, saat laporan ini dibuat Bibit sudah di download lebih dari 5 juta di play store dengan rating 4,5 dan berada di urutan ke 8 pada kategori gratis – populer di keuangan. Reksadana merupakan wadah untuk menghimpun dana dari investor yang nantinya akan di kelola oleh manajer investasi. Walaupun aplikasi bibit sudah di awasi OJK (Otoritas Jasa Keuangan), masih banyak masyarakat yang ragu untuk berinvestasi di aplikasi bibit. Hal ini disebabkan karna bibit tidak memiliki bentuk fisik

dan image yang kuat dari masyarakat. Image sebuah aplikasi bisa didapat dari sentimen pengguna terhadap aplikasi tersebut. Ulasan pada sebuah aplikasi dapat menjadi acuan untuk mengetahui sentimen pengguna lain tentang aplikasi tersebut. Untuk mengetahui sentimen pengguna pada ulasan bisa dengan menggunakan opinion mining (Analisis Sentimen). Opinion mining banyak digunakan untuk menganalisis secara otomatis opini pelanggan terhadap produk dan pelayanannya (Rozi, 2013). Opinion mining merupakan cabang dari penelitian text mining, cara kerjanya dengan menggali informasi dari data berbentuk tekstual. Text mining dalam analisis sentiment mampu mengevaluasi perasaan terhadap suatu pernyataan (Zhang, 2011). Dalam analisis sentimen diperlukan adanya klasifikasi, klasifikasi digunakan untuk mengelompokkan tanggapan mana yang memiliki sentimen positif dan sentimen negative (Pamungkas, 2021).

Penelitian yang dilakukan oleh Fajar Sodik Pamungkasa, dan Iqbal Kharisudin. yang berjudul, “Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter” Berdasarkan tingkat akurasi dengan menggunakan evaluasi model 10-Fold Cross Validation, diperoleh SVM memiliki akurasi 90,01%, Naïve Bayes memiliki akurasi 79,20% dan KNN memiliki akurasi 62,10%. Kesimpulan bahwa algoritma SVM memiliki akurasi yang lebih tinggi dari pada Naive Bayes dan KNN dengan akurasinya sebesar 90,01%. Sedangkan pada penelitian yang dilakukan oleh Fattya Ariani dan Andi Taufik yang berjudul, “Perbandingan Metode Klasifikasi Data Mining Untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar” dengan menambahkan fitur seleksi Particle Swarm Optimizatin di dapatkan nilai accuracy algoritma C4.5 sebesar 96,50%, Support Vector Machine sebesar 89,66%, dan algoritma Naïve Bayes sebesar 95,85%. Algoritma C4.5 mendapatkan nilai accuracy tertinggi, sehingga algoritma C4.5 lebih baik dalam memprediksi tingkat kepuasan pelanggan telkomsel prabayar.

Setiap algoritma klasifikasi pasti memiliki keunggulan dan kelemahannya masing-masing dalam mengklasifikasikan sentimen, diantara support vector

machine, naive bayes, C4.5 serta K-nearest neighbor algoritma mana yang paling tepat mengklasifikasikan data ulasan pada play store. Selain itu bagaimana performance hasil klasifikasi data tersebut.

Berdasarkan penjabaran di atas di perlukan metode yang paling tepat untuk mengklasifikasikan sentimen yang ada pada ulasan aplikasi bibit. Agar analisis sentimen bisa dilakukan dengan optimal dan calon pengguna bisa yakin pada analisa sentimen pada ulasan aplikasi bibit ini. Adapun nilai yang akan dijadikan sebagai acuan metode terbaik yaitu accuracy, recall dan precision.

2. Landasan Pemikiran

Analisis sentimen atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistic dan teks mining yang memiliki tujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Analisa sentimen dalam ulasan berfokus pada analisis orientasi sentimen dari ulasan, yang menunjukkan bahwa pengguna mengekspresikan sentimen positif, negatif atau netral terhadap produk, jasa atau acara (Liu, 2012).

Penelitian yang dilakukan oleh Fajar Sodik Pamungkasa, dan Iqbal Kharisudin. yang berjudul, “Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter” Berdasarkan tingkat akurasi dengan menggunakan evaluasi model 10-Fold Cross Validation, diperoleh SVM memiliki akurasi 90,01%, Naïve Bayes memiliki akurasi 79,20% dan KNN memiliki akurasi 62,10%. kesimpulan bahwa algoritma SVM memiliki akurasi yang lebih tinggi dari pada Naive Bayes dan KNN dengan akurasinya sebesar 90,01%

pada penelitian yang dilakukan oleh Fattya Ariani dan Andi Taufik yang berjudul, “Perbandingan Metode Klasifikasi Data Mining Untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar” dengan menambahkan fitur seleksi Particle Swarm Optimizatin di dapatkan nilai accuracy algoritma C4.5 sebesar 96,50%, Support Vector Machine sebesar 89,66%, dan algoritma Naïve Bayes sebesar

95,85%. Algoritma C4.5 mendapatkan nilai accuracy tertinggi, sehingga algoritma C4.5 lebih baik dalam memprediksi tingkat kepuasan pelanggan telkomsel prabayar.

Pada penelitian yang dilakukan oleh Nuraeni Herlinawati, Yuri Yuliani, Siti Faizah, Windu Gata, Samudi Samudi yang berjudul “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine”. Dalam penelitian ini membahas perbandingan naïve bayes dan support vector machine dalam melakukan analisis sentimen terhadap 1007 dataset yang diambil dari ulasan aplikasi zoom cloud meetings di play store. Hasilnya Support Vector Machine menghasilkan nilai akurasi dan auc yang sedikit lebih unggul daripada Naïve Bayes.

Pada penelitian yang dilakukan oleh Aaputra, S. A., Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain. (2019). Dengan judul “Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization”. Yang membahas analisis sentiment pada ulasan aplikasi E-wallet dengan menggunakan algoritma naïve bayes dan seleksi fitur particle swarm optimization (PSO). Dalam penelitian ini dapat peningkatan akurasi yang cukup signifikan saat menggunakan seleksi fitur Particle Swarm Optimization.

3. Metode Penelitian

Pada penelitian ini dimulai dengan melakukan web scraping pada playstore dengan menggunakan python dengan memasukkan link aplikasi bibit. Dataset lalu di preprocessing yang terbagi 3 tahap, cleansing, normalisasi dan proses dokumen (transform case, tokenizing, filtering, stopwords, dan stemming). Proses pengklasifikasian data dibagi dalam 2 skema pengujian. Pengujian pertama dengan menerapkan algoritma Naïve Bayes, Support Vector Machine, C4.5 dan K-Nearest Neighbor. Pengujian kedua hampir sama dengan pengujian pertama, bedanya pada pengujian kedua ditambahkan seleksi fitur Particle Swarm Optimization. Hasil pengujian akan di evaluasi menggunakan Confusion Matrix untuk mencari perbandingan nilai accuracy, recall, precision algoritma yang diuji baik sebelum dan sesudah menambahkan fitur seleksi Particle Swarm

Optimization. Nilai accuracy, recall, precision akan menentukan metode klasifikasi mana yang paling tepat dalam melakukan analisis sentimen ulasan aplikasi bibit.

3.1 Objek Penelitian

Dalam penelitian ini penulis mencoba mencari algoritma klasifikasi mana yang paling tepat untuk melakukan analisis sentimen terhadap ulasan pada play store. Dalam penelitian ini penulis mengambil studi kasus ulasan aplikasi bibit. Adapun algoritma klasifikasi yang digunakan merupakan algoritma populer yang biasa digunakan dalam melakukan text mining. Algoritma klasifikasi yang dimaksud yaitu Naïve Bayes, Support Vector Machine, C4.5 dan K-Nearest Neighbor. Untuk meningkatkan nilai akurasi dalam penelitian ini juga menambahkan seleksi fitur particle swarm optimization.

3.2 Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah kualitatif, data berbentuk text yang di ambil dari ulasan-ulasan pengguna aplikasi bibit yang ada pada web google play store.

3.3 Data yang Digunakan

Data yang digunakan dalam penelitian ini adalah data ulasan yang diambil melalui situs <https://play.google.com> dengan teknik web scraping. Teknik web scraping adalah teknik yang digunakan untuk mengekstrak sejumlah besar data dari situs web dimana data yang sudah di ekstraksi di simpan ke sebuah file berformat .csv yang kemudian di olah pada rapidminer untuk dilakukan tahap preprocessing.

3.3.1 Data yang Digunakan

Pengumpulan data menggunakan teknik web scraping melalui bantuan situs web scraping webharvy. Data yang diambil merupakan ulasan pengguna aplikasi bibit. Parameter yang disediakan diantaranya id aplikasi, lang, country, sort, count, filter_score_with. Proses pengambilan data pada penelitian dilakukan pada tanggal 19 januari 2022 jam 20:00 dengan memasukkan id aplikasi bibit “com.bibit.bibitid” berdasarkan newest (terbaru) Bahasa Indonesia (id) jumlah 500. Dari hasil web scraping penulis hanya mengambil data content (ulasan) dan score.

3.3.2 Preprocessing

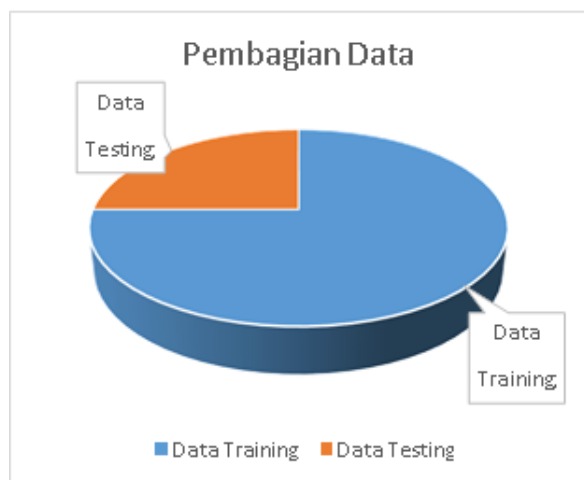
Preprocessing merupakan tahap ke dua dalam text mining yang bertujuan untuk mendapatkan data yang bersih, tahap preprocessing pada penelitian ini terdiri dari cleansing, normalisasi dan pemberian label (labeling).

3.3.3 Text Transformation

Text transformation merupakan tahap ke tiga dalam text mining yang bertujuan untuk membuat data yang berupa text siap di proses data mining, tahap text transformation pada penelitian ini terdiri dari transform case, Tokenizing, filtering, stopwords & stemming.

3.3.4 Data Training dan Data Testing

Karna penelitian ini menggunakan 4 folds cross validation maka data testing dan data training akan otomatis dipilihkan oleh komputer sesuai dengan nilai folds yang dipakai. Berhubung penelitian ini menggunakan 4 folds cross validation maka data training akan berjumlah $\frac{3}{4}$ (75%) dari dataset dan data testing berjumlah $\frac{1}{4}$ (25%) dari dataset. Dalam dataset penelitian ini terdapat 464 data, sehingga data training berjumlah 348 dan data testing berjumlah 116.

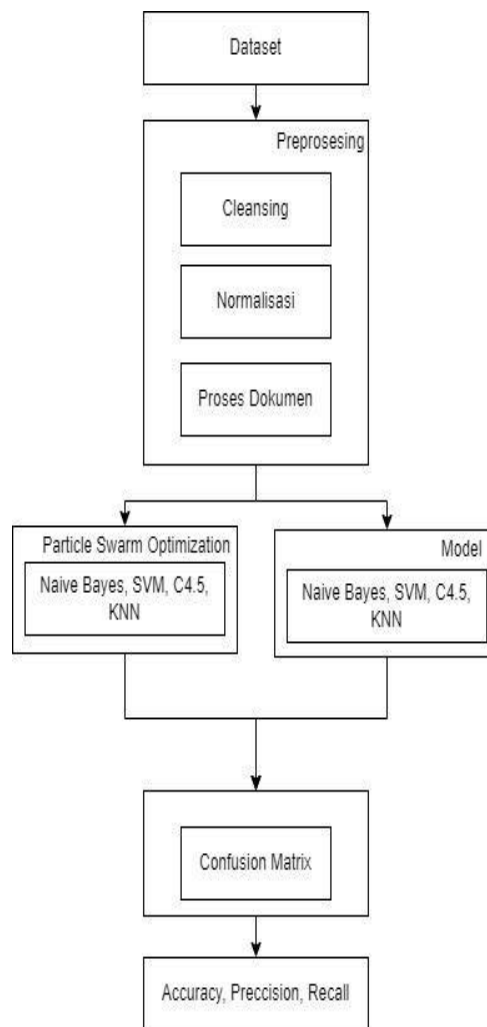


Gambar 1. Data Training dan Data Testing

3.3.5 Metode yang digunakan

Penelitian ini menggunakan algoritma naïve bayes, Support Vector Machine, C4.5 & K Nearest Neighbor dengan menambahkan metode seleksi fitur yaitu Particle Swarm Optimization.

Gambar dibawah ini merupakan desain metode yang akan digunakan pada penelitian penulis:



Gambar 2. Metode Penelitian

3.3.6 Evaluasi

Cara mengetahui hasil evaluasi yaitu dngan menggunakan operator performance yang telah di sediakan rapidminer, untuk hasil yang ingin di cari yaitu accuracy, precision dan recall.

3.4 Perangkat Penelitian

Adapun software dan hardware yang digunakan peneliti dalam membantu proses penelitian ini sebagai berikut:

1. Perangkat Lunak (software)

| No | Perangkat Lunak | Spesifikasi |
|----|-----------------|--|
| 1 | Sistem Operasi | Windows 10 |
| 2 | Tools | RapidMiner Studio versi 9.10.000 Microsoft Office |

2. Perangkat Keras (hadwere)

| No | Perangkat Keras | Spesifikasi |
|----|-----------------|---------------------------|
| 1 | Device | ACER E5-475g |
| 2 | Prossessor | Intel CORE I5 7200U 2.5GH |
| 3 | GPU | NVIDIA GeForce 940MX |
| 4 | Ram | 8 GB |

4. Hasil dan Pembahasan

4.1 Dataset

Dataset pada penelitian ini di dapat menggunakan teknik web scraping data pada 19 januari 2022 pada jam 20:00 WIB melalui situs https://colab.research.google.com/drive/1AmELtETDYabUIN-eLl5nlbG-_espuWO?usp=sharing dengan id aplikasi bibit playstore yaitu “com.bibit.bibitid”.

Proses web scraping data \\menghasilkan 500 data ulasan beserta raiting tiap ulasan. Setelah melewati tahap preprocessing, di dapat 464 data bersih yang terbagi menjadi 310 data ber label positif dan 154 data berlabel negatif. Karna pada penelitian ini menggunakan 4 Folds Cross Validation maka dataset akan otomatis terbagi menjadi 75% data testing dan 25% data training.

4.2 Proses Pengujian

Proses pengujian penelitian ini menggunakan metode Naïve Bayes, Support Vector Machine, C4.5 dan K-Nearest Neighbor dengan seleksi fitur Particle Swarm Optimization. Pengujian dilakukan sebanyak 2 kali, untuk mengetahui hasil perbandingan accuracy, recall dan precision setiap metode. pada pengujian pertama tanpa menggunakan

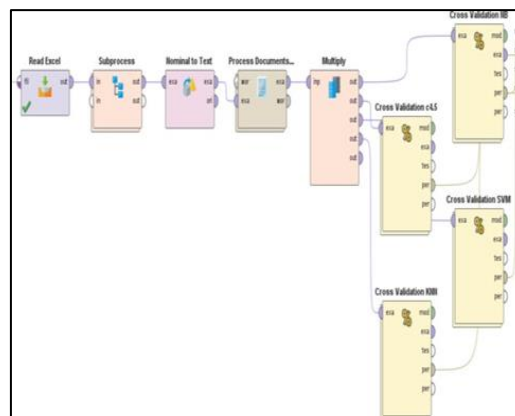
seleksi fitur Particle Swarm Optimization. Lalu pengujian kedua ditambah dengan seleksi fitur Particle Swarm Optimization. Proses pengujian penelitian ini menggunakan RapidMiner.

4.3 Hasil Pengujian

Pengujian dilakukan dengan menggunakan bantuan software RadpidMiner Studio versi 9.10.000. Dibawah ini merupakan hasil pengujian yang telah dilakukan.

4.3.1 Pengujian Pertama

Pada pengujian pertama ke empat metode (naïve bayes, SVM, C4.5 & KNN) di uji tanpa menggunakan seleksi fitur Particle swarm optimization untuk melakukan analisis sentiment terhadap dataset yang sudah di siapkan. Proses pengujian menggunakan software RapidMiner Studio. Berikut adalah desain pada RapidMiner.



Gambar 3. Proses Pengujian Pertama

1.Naïve Bayes

Hasil accuracy alogaritma Naïve Bayes sebesar 84,91%

accuracy : 84.91% +/- 2.68% (micro average: 84.91%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 115 | 31 | 78.77% |
| pred.positif | 39 | 279 | 87.74% |
| class recall | 74.68% | 90.00% | |

2.Support Vector Machine

Hasil accuracy 71,77%, recall 58,12% dan precision 78,88%

accuracy : 71.77% +/- 4.13% (micro average: 71.77%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 27 | 4 | 87.10% |
| pred.positif | 127 | 306 | 70.67% |
| class recall | 17.53% | 98.71% | |

3.C.45

Hasil accuracy 76,94%, recall 66,73% dan precision 80,80%

accuracy : 76.94% +/- 3.02% (micro average: 76.94%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 56 | 9 | 86.15% |
| pred.positif | 98 | 301 | 75.44% |
| class recall | 36.36% | 97.10% | |

4.K-Nearest Neighbor (KNN)

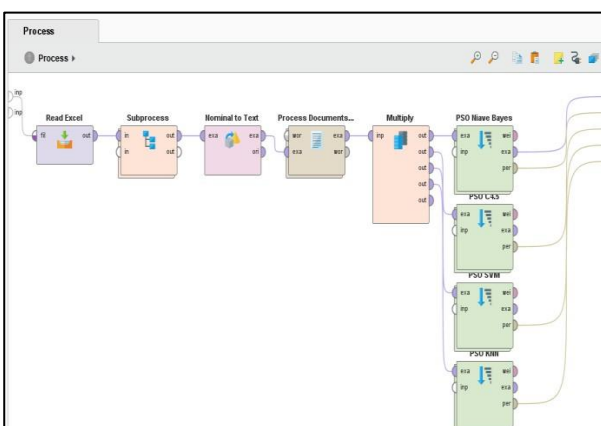
Hasil accuracy 66,81%, recall 50,00% dan precision 33,41%

accuracy : 66.81% +/- 0.50% (micro average: 66.81%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 0 | 0 | 86.15% |
| pred.positif | 154 | 310 | 66.81% |
| class recall | 0.00% | 100.00% | |

4.3.1 Pengujian Kedua

Pada pengujian kedua masih menggunakan dataset yang sama dan masih menggunakan algoritma naiva bayes, support vector machine, C4.5 & K-Nearest Neighbor. Bedanya pada pengujian kedua ditambahkan seleksi fitur Particle Swarm Optimization pada skema desain pengujian. Proses pengujian dilakukan dengan menggunakan software RapidMiner Studio sebagai alat bantu dalam melakukan analisis sentimen. Berikut adalah skema desain pengujian kedua.



Gambar 4. Proses Pengujian Kedua

1.Naïve Bayes

Hasil accuracy 87,28%, recall 84,11% dan precision 86,67%

accuracy : 87.28% +/- 4.30% (micro average: 87.28%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 115 | 20 | 85.19% |
| pred.positif | 39 | 290 | 88.15% |
| class recall | 74.68% | 93.55% | |

2.Support Vector Machine

Hasil accuracy 75,00%, recall 62,50% dan precision 85,19%

accuracy : 75.00% +/- 4.22% (micro average: 75.00%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 39 | 1 | 97.50% |
| pred.positif | 115 | 309 | 72.88% |
| class recall | 25.32% | 99.68% | |

3.C4.5

Hasil accuracy 79,74%, recall 70,62% dan precision 84,12%

accuracy : 79.74% +/- 2.28% (micro average: 79.74%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 67 | 7 | 90.54% |
| pred.positif | 87 | 303 | 77.69% |
| class recall | 43.51% | 97.74% | |

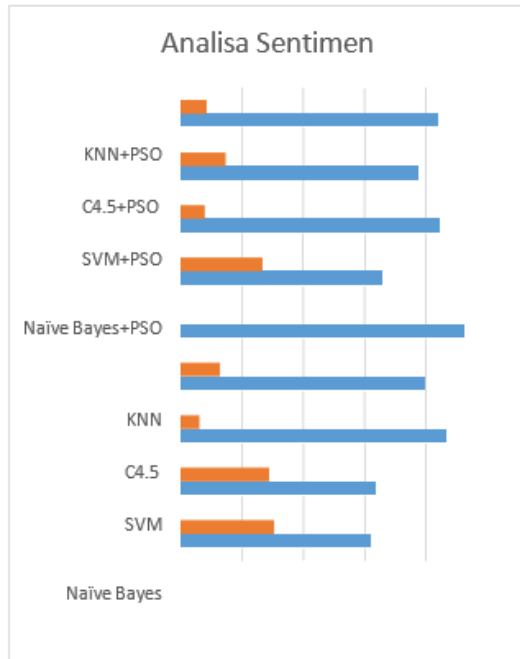
4.K-Nearest Neighbor

Hasil accuracy 74,35%, recall 62,02% dan precision 81,69%

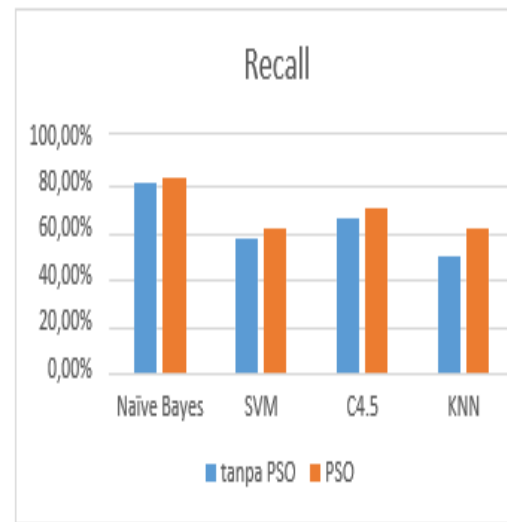
accuracy : 74.35% +/- 4.79% (micro average: 74.35%)

| | true negatif | true positif | class precision |
|--------------|--------------|--------------|-----------------|
| pred.negatif | 39 | 4 | 90.70% |
| pred.positif | 115 | 306 | 72.68% |
| class recall | 25.32% | 98.71% | |

4.3.3 Hasil Klasifikasi Sentimen



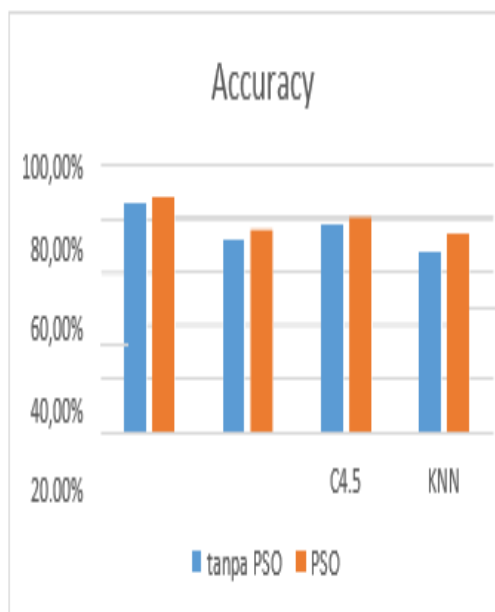
Gambar 5. Grafik Analisis Sentimen



Gambar 7. Hasil Recall Analisis Sentimen

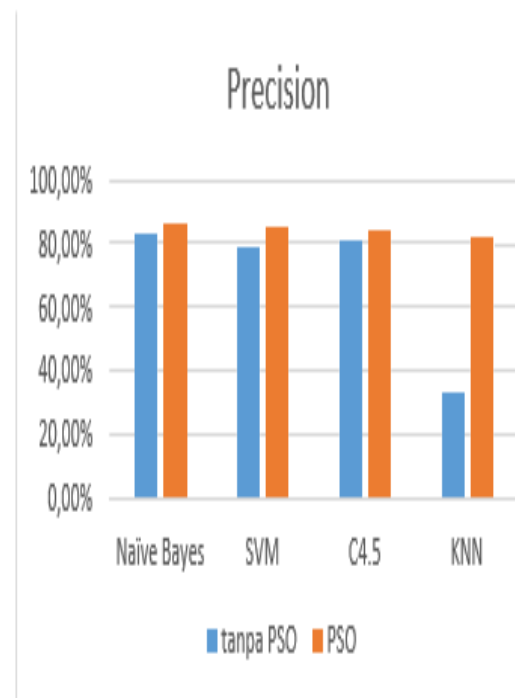
Pada grafik recall diatas metode naïve bayes memiliki nilai recall tertinggi dengan 82,34% dan setelah ditambah seleksi fitur Particle Swarm Optimization menjadi 84,11%.

4.3.3 Hasil Analisis Sentimen



Gambar 6. Hasil Akurasi Analisis Sentimen

Dari grafik diatas dapat dilihat metode naïve bayes memiliki accuracy yang paling tinggi yaitu 84,91% dan setelah ditambah seleksi fitur Particle swarm Optimization menjadi 87,28%.



Gambar 8. Hasil Precision Analisis Sentimen

Dari grafik precicion diatas dapat dilihat metode naïve bayes memiliki nilai tertinggi dengan 83,25% dan setelah ditambah seleksi fitur Particle Swarm Optimization menjadi 86,67%.

Tabel 1. Perbandingan Hasil Metode

| Pengujian | Accuracy | Recall | Precision | Klasifikasi Sentimen | | Keterangan | |
|-------------------|----------|--------|-----------|----------------------|---------|------------|---------|
| | | | | Positif | Negatif | Positif | Negatif |
| Dataset | - | - | - | 310 | 154 | 66,81 | 33,19 |
| Naïve Bayes | 84,91% | 82,34% | 83,25% | 318 | 146 | 68,53 | 31,47 |
| SVM | 71,77% | 58,12% | 78,88% | 433 | 31 | 93,32 | 6,68 |
| C4.5 | 76,94% | 66,73% | 80,80% | 399 | 65 | 85,99 | 14,01 |
| KNN | 66,81% | 50,00% | 33,41% | 464 | 0 | 100,00 | 0,00 |
| Naïve Bayes + PSO | 87,28% | 84,11% | 86,67% | 329 | 135 | 70,91 | 29,09 |
| SVM+ PSO | 75,00% | 62,50% | 85,19% | 424 | 40 | 91,38 | 8,62 |
| C4.5+ PSO | 79,74% | 70,62% | 84,12% | 390 | 74 | 84,05 | 15,95 |
| KNN+ PSO | 74,35% | 62,02% | 81,69% | 421 | 43 | 90,73 | 9,27 |

4.4 Pembahasan

Dari hasil pengujian yang telah dilakukan pada pengujian pertama dan pengujian kedua menunjukkan penggunaan algoritma naïve bayes memiliki performance yang lebih baik dari pada algoritma Support Vector Machine, C4.5 & K-Nearest Nighbor. Dimana pada pengujian pertama algoritma naïve bayes mendapat accuracy sebesar 84,91%, recall 82,34%, dan precision 83,25%, melebihi algoritma lain. Pada pengujian kedua dengan tambahan seleksi fitur Particle swarm Optimization algoritma Naïve Bayes kembali menjadi yang terbaik dengan accuracy 87,28% meningkat 2,37% dari pengujian pertama, recall 84,11% meningkat 1,77% dan precision 86,67% meningkat 3,42%.

Dari penelitian ini dapat disimpulkan bahwa algoritma Naïve Bayes merupakan algoritma yang paling tepat dalam melakukan analisa sentimen pada ulasan aplikasi bibit. Serta penggunaan seleksi fitur Particle Swarm Optimization terbukti dapat meningkatkan accuracy, recall, & precision.

5. Penutup

5.1 Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Penggunaan analisis sentimen dapat mengetahui sentimen positif dan negative yang terkandung pada ulasan aplikasi bibit.
2. Metode naïve bayes, SVM, C4.5 dan KNN dapat diterapkan untuk menganalisis sentimen pada ulasan aplikasi bibit.
3. Berdasarkan perbandingan nilai accuracy, recall, precision, metode naïve bayes lebih tinggi dibandingkan metode SVM, C4.5 dan KNN, sehingga paling tepat dalam melakukan analisis sentimen pada ulasan aplikasi bibit.

5.2 Saran

Adapun beberapa saran masukan dari penulis untuk penelitian selanjutnya, sebagai berikut:

1. Menerapkan metode Naïve bayes, SVM, C4.5 dan KNN untuk menganalisis sentimen pada studi kasus yang berbeda.
2. Masih ada data kotor dalam dataset penelitian ini. Hal ini disebabkan karna pada ulasan aplikasi bibit banyak ditemukan Bahasa gaul, kata singkatan, dan typo. Dimana kosa kata tersebut tidak terkandung pada kamus stopwords dan kamus stemming yang digunakan. Maka dari itu peneliti menyarankan untuk melakukan stopwords dan stemming secara manual pada kosa kata tersebut sehingga di dapat data yang benar-benar bersih.
3. Penerapan metode yang digunakan masih belum optimal karna penelitian ini berfokus pada perbandingan. Maka dari itu diharapkan pada penelitian selanjutnya lebih berfokus pada peningkatan nilai accuracy, recall, dan precision agar di dapat nilai yang optimal.

DAFTAR PUSTAKA

- [1] I. Rozi, S. Pramono and E. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *EECCIS*, vol. 6, no. 1, pp. 37-43, 2013.
- [2] L. G. R. D. M. H. M. & L. Zhang, "Combining Lexiconbased and Learning-based Methods for

- Twitter Sentiment Analysis," Chicago: Hewlett-Packard Development Company, 2011.
- [3] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," PRISMA, Prosiding Seminar Nasional Matematika, 2021.
- [4] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool, 2012.
- [5] D. R. W. G. a. S. M. H. S. A. Saputra, "Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," J. Resti, vol. 3, no. 3, pp. 377-382, 2019.