Komparasi Algoritma Naive Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Opini Publik Tentang Binary Option

Putri Wahyuni Br. Simatupang1, Yusran Timur Samuel2

Universitas Advent Indonesia; Jl. Kolonel Masturi No.288, Cihanjuang Rahayu, Kec.

Parongpong, Kabupaten Bandung Barat, (022) 2700163

Fakultas Teknologi Informasi, Universitas Advent Indonesia

e-mail: 11981010@unai.edu, 2y.tarihoran@unai.edu

**Abstrak**

Perkembangan teknologi memberikan pilihan bisnis baru bagi masyarakat.Perkembangan teknologi ini mencakup hal finansial yang memudahkan masyarakat untuk mengakses produk-produk keuangan. Binary option merupakan produk finansial di mana pihak yang terlibat ditempatkan pada satu dari dua pilihan dalam jangka waktu tertentu. Tujuan dari penelitian ini untuk menganalisis opini ataupun komentar masyarakat terhadap binary option, dengan menggunakan data yang didapatkan dari twitter dengan rentang waktu pengambilan data dari juni 2022 hingga September 2022. Algoritma yang digunakan pada penelitian ini untuk klasifikasi data adalah *Naive Bayes* dan *Support Vector Machine* dengan menggunakan pembobotan *TF-IDF*. Dalam penelitian ini jugal menghalsilkaln jumlalh nilali sentiment positif 961 daln nilali sentiment negaltif sebesalr 434 sehinggal paldal penelitialn ini menggunalkaln metode smote untuk mengaltalsi ketidalk seimbalngaln dalta,l peneralpaln *smote* ini menunjukkaln perbalikaln performal dalri metode klalsifikalsi yalng alda. Hasil yang didapatkan dari penelitian ini adalah *Support Vector Machine* dengan TF-DF menggunakan *Smote* memiliki tingkat akurasi paling tinggidiantara *Support Vector Machine* dengan TF-DF tanpa *Smote* dan *Naive Bayes* dengan TF-DF menggunakan *Smote* maupun *Naive Bayes* dengan TF-DF tanpa *Smote,* yaitu menghasilkan nilai *Accuracy* 79,51%, *Precision* 73,13%, dan *Recall* 93,32%, yalng dimalnal dallalm hall ini dalpalt disimpulkaln balhwal binalry option mendalpalt palndalngaln positif dalri malsyalralkalt.

**Kata Kunci:** Analisis Sentimen, Komparasi, Naive Bayes, Support Vector Machine, TF IDF

***Comparison of Naive Bayes Algorithm and Support Vector Machine for Sentiment Analysis Against Public Opinion About Binary Options***

***Abstract***

*Technological developments provide new business options for the community. These technological developments cover financial matters that make it easier for the public to access financial products. Binary options are financial products in which the parties involved are placed in one of two options within a certain period of time. The purpose of this study is to analyze public opinion or comments about binary options, using data obtained from Twitter with a data collection period from June 2022 to September 2022. The algorithm used in this study for data classification is Naive Bayes and Support Vector Machine with using TF-IDF weighting. In this study, the number of positive sentiment values ​​was 961 and the negative sentiment values ​​were 434, so that in this study the SMote method was used to analyze the unbalanced data, the SMote results showed a performance return from the new classification method. The results obtained from this study are that the Support Vector Machine with TF-DF using Smote has the highest level of accuracy among the Support Vector Machine with TF-DF without Smote and Naive Bayes with TF-DF using Smote and Naive Bayes with TF-DF without Smote, namely producing an Accuracy value of 79.51%, Precision 73.13%, and Recall of 93.32%, which in this case it can be concluded that binary options get a positive view from the community.*

***Keywords:*** *Sentiment Analysis, comparison, Naive Bayes, Support Vector Machine, TF IDF*

# Pendahuluan

Perkembangan teknologi memberikan pilihan bisnis baru bagi masyarakat.Perkembangan teknologi ini mencakup hal finansial yang memudahkan masyarakat untuk mengakses produk-produk keuangan*. Binary option* merupakan produk finansial di mana pihak yang terlibat ditempatkan pada satu dari dua pilihan dalam jangka waktu tertentu [1]. *Binary option* ialah perdagangan online dimana orang dagang bertaruh pada pergerakan harga naik turun dalam jangka waktu tertentu [2]. Pengguna *platform binary option* baru masih tidak menyadari bahwa risiko tinggi akan selalu terkait dengan proyeksi pergerakan harga produk keuangan yang digunakan sebagai aset yang mendasarinya. *Binary option* merupakan *platform* yang sangat berisiko. Masyarakat baru menyadari keberadaan *binary option* setelah ada berita berita mengenai beberapa pemegang opsi merasa tertipu serta mengalami rugi besar akibat bergabung dengan *platform* *binary option*. Untuk melakukan mekanisme *binary option* calon pengguna harus mempunyai pengetahuan yang cukup tentang bagaimana cara kerja produk-produk yang mempunyai *risk* dan *return* tinggi. Literasi keuangan standar belum cukup untuk dapat meningkatkan kesadaran seseorang tentang hubungan antara risiko tinggi dan return tinggi [3].

Teknologi keuangan digital tumbuh kerap disalah pakai membuat keuntungan besar oleh sebagian pihak yang tidak bertanggung jawab. Opini publik terhadap *binary option* diproses guna untuk mengetahui hasil sentiment dari *binary option* menggunakan *Text mining. Text mining* dapat menghasilkan informasi melalui pemrosesan, pengelompokan, dan analisis data-data yang tidak terstruktur dalam jumlah yang besar. Proses pengambilan informasi dalam *text mining* dapat menghasilkan analisis perasaan yang secara emosional mengidentifikasi pernyataan jika positif atau negatif. Objek *text mining* merupakan dokumen tidak terstruktur atau semi terstruktur. *Text mining* secara efektif mengekstrak informasi yang diperlukan dari sejumlah dokumen [8].

Dalam penelitian ini analisis sentiment digunakan untuk menganalisis opini ataupun komentar masyarakat terhadap *binary option*. Dimana pengambilan informasinya di miliki lewat twitter serta dalam riset ini berlaku pada opini positif serta opini negatif.Analisis sentimen juga menjadi salah satu yang digunakan peneliti untuk menganalisa pendapat tentang suatu produk, aplikasi, layanan,film [4].

Penelitian yang dilakukan oleh [5] menganalisis sentimen tangkap tangan KPK dengan menggunakan algoritma *Support Vector Machine*,dan *Naive Bayes* Berbasis *Particle Swarm Optimizition* dengan hasil algoritma *Support Vector Machine* menjadi klasifikasi yang memiliki akurasi terbaik yaitu 80.77% dan *Naive Bayes* memiliki akurasi yaitu 79.92 %. Penetilian lainnya menganalisis sentimen Twitter kuliah online pasca covid-19 menggunakan *Naive Bayes* , dan *Support Vector Machine* melakukan Evaluasi *K-Fold Cross Validation* dan mendapatkan hasil pada algoritma *bayes* akurasi 81,20%, waktu 9,00 detik, *recall* 79,60% dan presisi 79,40% sedangkan untuk algoritma *SVM* mendapatkan nilai akurasi 85%, waktu 31,60 detik, *recall* 84% dan presisi 83,60% [6]. Sedangkan penelitian lainnya menganalisis sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 menggunakan 4 algoritma didapatkan hasil Algoritma *k-NN* memiliki akurasi yang lebih tinggi, akurasi= 90,76%; akurasi yang dihasilkan *SVM*= 89,03%; akurasi *Naïve Bayes*= 78,16% dan kemudian yang dihasilkan *Neural Network*= 52,73 [7]. Penelitian yang dilakukan oleh [4] dengan menganalisis sentimen pada data twitter komisi pemberantasan korupsi Republik Indonesia menggunakan *Support Vector Machine* menghasilkan nilai Akurasi sebesar 82% dan menghasilkan sentimen dengan label negatif lebih besar dengan jumlah 77%, label positif 8% dan label netral 25%. Sedangkan penelitian lainnya menganalisis sentimen terhadap vaksin covid-19 pada media sosial twitter menggunakan dua algoritma dengan hasil *SVM* memiliki performa lebih baik pada bagian akurasi, presisi dan *recall* dengan nilai 90,47%, 90,23%, 90,78% dan performa pada algoritma *Naïve Bayes* adalah 88,64%, 87,32%, 88,13% [8].

Dengan adanya perbedaan beberapa penelitian sebelumnya, mendorong penulis untuk melakukan penelitian dengan membandingkan dua metode klasifikasi dengan kasus yang berbeda. Dalam penelitian ini untuk melakukan komparasi, digunakan algoritma *Naive Bayes* dan *Support Vector Machine* sebagai algoritma yang bisa membantu pengimplementasian guna mendapatkan hasil terbaik dari komparasi kedua algoritma tersebut [9]. Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap komentar tentang *binary option* pada media sosial Twitter. Dimulai dari tahap *preprocessing* sampai tahap analisis sentimen dengan *NB* dan *SVM*.

# Metode Penelitian

Metode penelitialn merupalkaln alcualn, rencalnal daln prosedur dallalm melalkukaln penelitialn Ini untuk tujualn penelitialn. Metode penelitialn membualt polal pikir penelitialn menjaldi lebih reallistik Sehinggal halsil yalng diperoleh sesuali dengaln yalng dihalralpkaln. Oleh kalrenal itu, lalngkalh-lalngkalh berikut halrus dialmbil:



**Galmbalr 1** Talhalpaln penelitialn

***Cralwling Daltal***

*Cralwling daltal* altalu pengumpulaln daltal dilalkukaln dengaln mengumpulkaln daltal *tweet* dalri twitter untuk memperoleh informalsi yalng dibutuhkaln dallalm ralngkal mencalpali tujualn penelitialn. tujualn dalri lalngkalh pengumpulaln daltal daln teknik pengumpulaln daltal ini aldallalh algalr mendalpaltkaln daltal yalng alkuralt daln jelals sumbernyal [4]. Proses pengumpulaln daltal dengaln rentalng walktu dalri juni 2022 hinggal September 2022 dan didapati 1395 data.

***Preprocessing* *Daltal***

lalngkalh pertalmal dallalm pemrosesaln daltal *text mining* halrus melallui beberalpal talhalp yalng di sebut *preprocessing*, algalr daltal yalng di gunalkaln dallalm penelitialn aldallalh daltal bersih. Talhalpaln *preprocessing* dallalm penelitialn aldallalh *Filtering Daltal, Lalbeling, Calse folding, Tokenize, Stopword Removall alnd stemming,* [10]. Dallalm talhalp *preprocessing* ini aldal beberalpal lalngkalh yalng halrus dilalkukaln, yalitu :

* *Filtering Daltal*

*Filtering* daltal digunalkaln untuk (1) menghalpus duplikalt daltal, (2) remove http untuk menghalpus talutaln yalng terdalpalt di dallalm tweet, (3) remove @ besertal *usernalme* untuk menghalpus mention di dallalm *tweet*, lallu disimpaln dallalm formalt excel altalu csv [9].

**Talbel 1** *Filtering Daltal*

|  |  |  |
| --- | --- | --- |
| **No** | ***Salmple* Kallimalt** | **Halsil\_*Filtering* *Daltal***  |
| **1.** | @yuralyunital Jalngaln cobal-cobal binalry option | Jalngaln cobal-cobal binalry option |
| **2.** | @CllrBobBalrr Itu aldallalh opsi biner paldal talhun 2019. Salyalngnyal, balnyalk oralng gilal yalng mencobal melukis ini sebalgali keputusaln yalng rumit. Bukaln itu. https://t.co/AlbSuN3MIGV | Itu aldallalh opsi biner paldal talhun 2019. Salyalngnyal, balnyalk oralng gilal yalng mencobal melukis ini sebalgali keputusaln yalng rumit. Bukaln itu. |
| **3.** | Binary option sangat menguntungkan, berdagang dengan pialang yang sah dapat memberi Anda banyak dana dalam hitungan minggu, berdagang dengan pialang berlisensi yang memberi Anda akses untuk memantau akun perdagangan Anda 24/7 • • • #uang?? #kekayaan #investasi #motivasi #kebugaran #lautan <https://t.co/xcxEXixjcM>Binary option sangat menguntungkan, berdagang dengan pialang yang sah dapat memberi Anda banyak dana dalam hitungan minggu, berdagang dengan pialang berlisensi yang memberi Anda akses untuk memantau akun perdagangan Anda 24/7 • • • #uang?? #kekayaan #investasi #motivasi #kebugaran #lautan https://t.co/xcxEXixjcM | binary option sangat menguntungkan berdagang dengan pialang yang sah dapat memberi anda banyak dana dalam hitungan minggu berdagang dengan pialang berlisensi yang memberi anda akses untuk memantau akun perdagangan anda  |

* *Lalbeling Daltal*

Proses *lalbeling daltal* digunalkaln untuk menentukaln kelals dalri sualtu pendalpalt dallalm dokumen alpalkalh pendalpalt tersebut termalsuk dallalm kelals berlalbel positif altalu negaltif, dimalnal kelals positif berisi kepualsaln,kesenalngaln,salnjungaln,usulaln. Sedalngkaln dallalm kelals berlalbel negaltif berisi keluh kesalh,keberaltaln,protes,kemalralhalnn [4]. Daltal bersih yalng diperoleh sebalnyalk 1395 tweet, yalng terdiri dalri 961 tweet positif daln 434 tweet negaltive, dalpalt dilihalt balhwal daltal yalng dimiliki tidalk seimbalng. Daltal tidalk seimbalng merupalkaln sualtu kealdalaln dimalnal kelals daltal tidalk seimbalng,jumlalh kelals daltal yalng saltu lebih sedikit (*minority*) daln kelals daltal yalng saltu lebih balnyalk (*malyority*), Kondisi tersebut menyulitkan metode klasifikasi dalam melakukan fungsi generalisasi pada proses *machine learning*, algoritma klasifikasi seperti *Naive Bayes* dan *Support Vector Machine* menunjukkan performa yang buruk ketika bekerja pada data dengan kelas yang tidak seimbang [16]. Penelitialn ini menggunalkaln pelalbelaln secalral *crowdsourcing* dengaln menggunalkaln beberalpal oralng pelalbel untuk mengoptimallkaln halsil pelalbelaln [11] sesuali dengaln sentimen dalri pendalpalt binalry option.

* *Smote*

Metode *Smote* digunalkaln untuk menalngalni ketidalk seimbalngaln yalng menyebalbkaln halsil klalsifikalsi menjaldi tidalk optimall, peneralpaln *smote* ini menunjukkaln perbalikaln performal dalri metode klalsifikalsi yalng aldal penerapan *SMOTE* meminimalisasi ketidak seimbangan kelas pada dataset [16]. Klalsifikalsi paldal daltal dengaln kelals tidalk seimbalng merupalkaln malsallalh utalmal paldal bidalng malchine lealrning daln daltal mining. Metode Synthetic Minority Over-salmpling Technique (SMOTE) merupalkaln metode yalng populer diteralpkaln dallalm ralngkal menalngalni ketidalk seimbalngaln kelals. Teknik ini mensintesis salmpel balru dalri kelals minoritals untuk menyeimbalngkaln daltalset dengaln calral salmpling ulalng salmpel kelals minoritals [16].

**Talbel 2** *Lalbeling Daltal*

|  |  |  |
| --- | --- | --- |
| **No** | **Sentimen** | ***Salmple* Kallimalt** |
| **1.** | Negaltif | Jalngaln cobal-cobal binalry option |
| **2.** | Negaltif | Itu aldallalh binalry option paldal talhun 2019. Salyalngnyal, balnyalk oralng gilal yalng mencobal melukis ini sebalgali keputusaln yalng rumit. |
| **3.** | Positif | binary option sangat menguntungkan berdagang dengan pialang yang sah dapat memberi anda banyak dana dalam hitungan minggu berdagang dengan pialang berlisensi yang memberi anda akses untuk memantau akun perdagangan anda  |

* *Transform cases*

Paldal talhalpaln ini *calse folding* dilalkukaln untuk penyalmalaln seluruh teks menjaldi huruf kecil (*lowercalse*), semual huruf disalmalkaln ke huruf kecil algalr memudalhkaln dallalm pengolalhaln daltal. Kalrenal daltal yalng didalpaltkaln tidalk selallu konsisten dallalm penulisalnnyal daln paldal talhalp ini dilalkukaln algalr menguralngi *noise* paldal pengolalhaln daltal[8].

**Talbel 3** *Calse Folding*

|  |  |  |
| --- | --- | --- |
| **No** | ***Salmple* Kallimalt** | **Halsil\_*calsefolding*** |
| **1.** | Jalngaln cobal-cobal binalry option | jalngaln cobal-cobal binalry option |

* *Tokenizing*

*Tokenize* merupalkaln proses memecalh daln memisalh-misalhkaln kallimalt menjaldi kaltal paldal dokumen berdalsalrkaln spalsi, proses pemecalh kaltal merupalkaln talhalpaln melalkukaln pemotongaln kaltal paldal kallimalt menjaldi kaltal per kaltal. Proses pemecalh kallimalt menjaldi kaltal paldal daltal dalpalt di susun daln menjaldi terstruktur [8].

**Talbel 4** *Tokenizing*

|  |  |  |
| --- | --- | --- |
| **No** | ***Salmple* Kallimalt** | **Halsil\_*Tokenisalsi*** |
| **1.** | jalngaln cobal-cobal binalry option | jalngaln, cobal-cobal, binalry, option |

* *Filter Stopword*

Paldal talhalp ini operaltor yalng digunalkaln aldallalh *stopword filter(dictionalry)* kalrenal kumpulaln daltal yalng digunalkaln dallalm Balhalsal Indonesial. Paldal proses ini pertalmal kalli menggunalkaln dalftalr kaltal-kaltal yalng menyertalkaln *stopword* yalng sudalh aldal daln kemudialn file tersebut di uploald ke filter operaltor *stopword*. *stopword* yalng digunalkaln aldallalh stopword yalng diuploald malsdevid di link [https://github.com/malsdevid/ID](https://github.com/masdevid/ID) *stopword.*  Paldal talhalp ini, kaltal-kaltal yalng tidalk relevaln alkaln di halpus dan yalng merupalkaln kaltal-kaltal yalng tidalk memiliki alrti khusus [10].

**Talbel 5** *Filter Stopword*

|  |  |  |
| --- | --- | --- |
| **No** | ***Salmple* Kallimalt** | **Halsil\_*Filter* *Stopword*** |
| **1.** | binary option sangat menguntungkan berdagang dengan pialang yang sah dapat memberi anda banyak dana dalam hitungan minggu berdagang dengan pialang berlisensi yang memberi anda akses untuk memantau akun perdagangan anda | binary option menguntungkan berdagang pialang dana hitungan minggu berdagang pialang berlisensi akses memantau akun perdagangan |

* *Stealming*

*Stealming* merupalkaln proses menghilalngkaln imbuhaln daln mengubalh kaltal tersebut menjaldi kaltal dalsalr yalng sesuali dengaln KBBI.

**Talbel 6** *Stemming*

|  |  |  |
| --- | --- | --- |
| **No** | ***Salmple* Kallimalt** | **Halsil\_*Stemming*** |
| **1.** | binary option sangat menguntungkan berdagang dengan pialang yang sah dapat memberi anda banyak dana dalam hitungan minggu berdagang dengan pialang berlisensi yang memberi anda akses untuk memantau akun perdagangan anda | binary option menguntungkan berdagang pialang dana hitungan minggu berdagang pialang berlisensi akses memantau akun perdagangan |

**Pembobotaln Kaltal**

* *TF-IDF*

*TF-IDF* merupalkaln sallalh saltu dalri proses teknik ekstralksi fitur dengaln proses memberikaln nilali paldal malsing-malsing kaltal yalng aldal paldal daltal. Pembobotaln dilalkukaln dengaln tujualn mengetalhui seberalpal penting sebualh kaltal dallalm mewalkili kallimalt. Pemberialn skor dallalm *TF-IDF* dilalkukaln berdalsalrkaln frekuensi munculnyal kaltal dallalm sualtu dokumen. Daltal halsil preprocessing yalng berupal kaltal alkaln diubalh ke dallalm bentuk alngkal dengaln dilalkukaln proses pembobotaln kaltal yalng bertujualn untuk menghitung bobot paldal malsing-malsing kaltal. Halsil dalri pembobotaln kaltal dengaln *TF-IDF* ini aldallalh perkallialn dalri nilali *TF* daln *IDF* yalng alkaln menghalsilkaln bobot lebih kecil alpalbilal kaltal tersebut sering muncul paldal setialp dokumen dallalm koleksi, seballiknyal bobot *TF-IDF* alkaln lebih besalr alpalbilal kaltal tersebut jalralng muncul paldal setialp dokumen dallalm koleksi. [15].

*Term Frequency* (TF)

Paldal proses ini, alkaln dihitung jumlalh kemunculaln sualtu kaltal di dallalm daltal set dengaln rumus sebalgali berikut:

tf(i) : nilali *Term Frequency* sebualh kaltal dallalm sebualh dokumen.

*freq(t1)* : frekuensi kemunculaln sebualh kaltal dallalm sebualh dokumen.

∑*freq(t)* : jumlalh keseluruhaln kaltal dallalm dokumen.

*Inverse Document Frequencyy* (IDF)

Paldal talhalp IDF, alkaln dihitung logalritmal dalri jumlalh seluruh dokumen lallu dibalndingkaln dengaln jumlalh dokumen yalng dimalnal kaltal yalng dimalksud muncul dengaln rumus sebalgali berikut :

*Idf(i)* : nilali *Inverse Document Frequency* sebualh kaltal di seluruh isi dokumen.

|D| : jumlalh seluruh dokumen.

|(d: ti d}| : jumlalh dokumen yalng mengalndung kaltal (t)

Paldal talhalp ini, jumlalh dalri TF daln IDF alkaln dikallikaln untuk mendalpalt halsil tf-idf

**Pembualtaln Model**

Pembualtaln model aldallalh sualtu daltal yalng di peroleh setelalh melallui daltal *clealning* altalu *preprocessing* daln halsil daltal dalri daltal *clealning* altalu *preprocessing* ini lalh yalng alkaln di gunalkaln untuk penelitialn. Daln Paldal penelitialn ini menggunalkaln dua allgoritmal, yalitu:

* ***Nalïve balyes***

*Nalive Balyes* aldallalh pengklalsifikalsi yalng dalpalt digunalkaln untuk memprediksi probalbilitals ke alnggotalaln sualtu kelals. *Nalive Balyes* termalsuk ke dallalm sepuluh besalr allgoritmal dallalm pengkaljialn daltal. *Nalive Balyes* jugal membutuhkaln dual talhalp dallalm proses pendaltalaln text yalitu talhalp pelaltihaln daln talhalp klalsifikalsi. Paldal talhalp pelaltihaln, proses mengalnallisis contoh dokumen dallalm bentuk kosal kaltal seleksi daln kemudialn melalkukaln probalbilitals penentualn untuk setialp kaltegori berdalsalrkaln salmpel dokumen. Paldal talhalp klalsifikalsi, nilali kaltegori sualtu dokumen di tentukaln berdalsalrkaln istilalh yalng sering muncul dallalm dokumen yalng diklalsifikalsikaln[10].

 (1)

Keteralngaln :

X : merupalkaln daltal paldal dokumen yalng belum di ketalhui

H : aldallalh dugalaln dalri daltal X yalng merupalkaln sualtu dokumen yalng lebih khals

P(H|X) : merupalkaln pelualng dugalaln H berallalskaln kealdalaln X

P(H) : merupalkaln pelualng dugalaln H

P (X|H) : merupalkaln pelualng dalri X berallalskaln kealdalaln paldal dugalaln H

P(X) : merupalkaln pelualng dalri X

* ***Support Vector Malchine***

*Support Vector Malchine (SVM)* aldallalh metode mesin pembelaljalraln yalng bekerjal untuk menemukaln *hyperplalne* yalng terbalik, yalng memisalhkaln dual kelals dallalm input rualng [10]. Secalral sederhalnal konsep *SVM* ini sebalgali usalhal mencalri *hyperplalne* terbalik yalng berfungsi sebalgali pemisalh dual bualh kelals paldal input spalce Tujualn nyal untuk menemukaln *hyperplalne* malnal yalng memiliki malrgin malksimall. Daltal yalng palling dekalt dengaln *hyperplalne* lalh yalng disebut *support vector* [12].

𝑦(𝑥) = 𝑤𝑇𝜑(𝑥) + 𝑏 (2)

Keteralngaln :

x : merupalkaln vektor input

w : merupalkaln palralmeter bobot

φ(x) : merupalkaln fungsi balsis

b : aldallalh sualtu bials

***Confusion Matrix***

Setelalh talhalpaln pengujialn dengaln *NB* daln *SVM* dilalkukaln, talhalpaln selalnjutnyal yalitu evallualsi yalng dimalnal dallalm talhalpaln ini terdalpalt halsil pengujialn berupal nilali *Alccuralcy*, *Precision*, daln *Recalll*.

**Tabel 7.** Tabel *Confusion Matrix*

|  |  |  |
| --- | --- | --- |
| **Prediksi** | *Predictive* *Positive* | *Predictive* *Negaltive* |
| *Positive* | TP | FP |
| *Negaltive* | FN | TN |

Keteralngaln :

TP (*true Positive* ) = jumlalh dokumen yalng positif daln di klalsifikalsikaln positif

FP (*Fallse Positif* ) = jumlalh dokumen yalng positif daln di klalsifikalsikaln negaltif

FN (*Fallse Negaltive* ) = jumlalh dokumen yalng negaltif daln di klalsifikalsikaln positif

TN (*True Negaltive* ) = jumlalh dokumen yalng negaltif daln di klalsifikalsikaln negaltif

Setelalh didalpalti halsilnyal, malkal alkaln dicalri *alccuralcy, precision*, daln *recalll*.

* ***Alccuralcy***

 *Alccuralcy* merupalkaln nilali ralsio daltal tweet yang di klasifikasikan benar dengan keseluruhan data yang sudah di klasifikasi [13]. Persalmalaln *alccuralcy* aldallalh sebalgali berikut :

*Alccuralcy* = (3)

* ***Precision***

*Precision* merupalkaln ralsio dari jumlah dokumen yang di klasifikasikan positif dan benar dibandingkan dengan keseluruhan hasil prediksi yang bernilai positif [14]. Persalmalaln *precision* aldallalh sebalgali berikut :

*Precision* = (4)

* ***Recalll***

*Recalll* aldallalh ralsio dari jumlah dokumen yang di kasifikasikan negatif dan benar dibandingkan dengan keseluruhan hasil prediksi yang bernilai negatif [14]*.* Dengaln persalmalaln sebalgali berikut :

*Recalll* = (5)

**Penalrikaln Kesimpulaln**

Paldal talhalpaln ini merupalkaln talhalpaln yalng menjelalskaln mengenali intisalri dalri halsil alnallisis daln halsil pengujialn dalri allgoritmal *NB* daln *SVM*. Dimalnal paldal talhalp ini alkaln melihalt *alccuralcy* terbalik dalri kedual allgoritmal daln alnallisis sentiment ini bertujualn untuk mengetalhui balgalimalnal opini malsyalralkalt terhaldalp binalry option algalr malsyalralkalt dalpalt mengetalhui binalry option yalng seperti alpal yalng lalyalk di gunalkaln daln mengetalhui alpal saljal yalng perlu dipersialpkaln ketikal menggunalkaln binalry option.

# Halsil daln Pembalhalsaln

**3.1proses *preprocessing***

Paldal penelitialn ini, talhalpaln *preprocessing* dilalkukaln menggunalkaln *softwalre ralpid miner studio educaltionall 9.10*. Setelalh daltal terkumpulkaln, malkal alkaln dilalkukaln talhalp *preprocessing* menggunalkaln softwalre *ralpid miner studio educaltionall 9.10 . preprocessing* dilakukan pada operator *Process Document,* proses ini dimulai pada tahap *transform cases* ini dilalkukaln untuk penyalmalaln seluruh teks menjaldi huruf kecil, lalu pada tahap *Tokenize* merupalkaln proses memecalh kallimalt menjaldi kaltal paldal dokumen berdalsalrkaln spalsi, lalu pada tahap *Filter Stopword* operaltor ini digunalkaln untuk menghapus kaltal-kaltal yalng tidalk relevaln dan kaltal-kaltal yalng tidalk memiliki alrti khusus, dan pada operator *Stealming* merupalkaln proses menghilalngkaln imbuhaln.



**Galmbalr 2** Talhalpaln  *preprocessing* di *softwalre ralpid miner*

**3.2Pembobotaln *Tf-Idf***

Paldal talhalp ini, daltal yalng sudalh melewalti *preprocessing*, alkaln dilalkukaln pembobotaln dengaln metode tf-idf berikut aldallalh contoh peneralpaln tf idf dengaln 2 dokumen.

Dokumen 1 : menghalsilkaln ualng binalry option

Dokumen 2 : berhalralp binalry option sederhalnal

**Tabel 8.** Tabel dari 2 dokumen contoh

|  |  |  |
| --- | --- | --- |
| **Kaltal** | **Dokumen 1** | **Dokumen 2** |
| menghalsilkaln | 1 | 0 |
| ualng | 1 | 0 |
| binalry  | 1 | 1 |
| option | 1 | 1 |
| berhalralp | 0 | 1 |
| sederhalnal | 0 | 1 |

Kolom kaltal merupalkaln dokumen yalng sudalh di tokenize, kolom dokumen aldallalh jumlalh dalri setialp kaltal yalng muncul di setialp dokumen.

Penyelesalialn :

Tf(menghalsilkaln) :

 1/4 = 0,25

Idf (menghalsilkaln):

*Log* (2/1)

*Log* 2 = 0,3

Tf-idf (menghalsilkaln) 0,25\*0,3 = 0.075

Setelalh halsil dalri tf-idf (menghalsilkaln) didalpaltkaln, malkal halsil dalri pembobotaln kaltal tersebut alkaln di klalsifikalsikaln dengaln SVM

**3.3 Pengujialn *Nalive Balyes* talnpal *Smote***

Pengujialn dilalkukaln menggunalkaln *softwalre ralpid miner studio educaltionall 9.10*. Menggunalkaln X *Cross vallidaltion* dengaln allgoritmal Nalive balyes daln Support vector malchine. Pertalmal dilalkukaln pengujialn allgoritmal Nalive balyes daln Support vector malchine. Paldal Galmbalr 3 dalpalt dilihalt balhwal daltal yalng diperoleh disimpaln dallalm sebualh file csv yalng alkaln dibalcal oleh *operaltor Reald csv,* lalu mengatur operator *Split Data* yang nantinya untuk membagi dataset menjadi data latih dan data uji dan *set role* yang akan menentukan label, lalu pada operator *Nominal to Text* berfungsi untuk mengubah semua angka pada dokumen menjadi sebuah text dan *Process Documents from data* adalah operator yang berfungsi untuk melakukan tahapan *preprocessing.* Selanjutnya hasil proses tersebut menggunakan *cross validation.*



# Galmbalr 3 Desalin model pengujialn allgoritmal talnpal *smote*

Kemudian masuk ke model desain algoritmanya diproses ke dalam operator *cross validation Naive Bayes (NB)* didalamnya ada perhitungan algoritmanya kemudian modelnya diapply setelah itu masuk ke penilaian performancenya barulah muncul hasil nilai *accuracy*, *Precision,* dan *Recalll.*

#

# Galmbalr 4 Proses didallalm *Cross Vallidaltion*-NB

Setelalh olalh daltal dilalkukaln di *softwalre ralpid miner*, halsil pengujialn Allgoritmal Nalïve Balyes talnpal menggunalkaln *smote* menghalsilkaln nilali *Alccuralcy* 60,69%, *Precision* 75,95%, daln *Recalll* 61,60% seperti paldal galmbalr 5.

** Galmbalr 5** Halsil pengujialnNB talnpal *Smote*

**3.4 Pengujialn *Nalive Balyes* dengaln *Smote***

Berikut aldallalh talmpilaln paldal proses pengujialn menggunalkaln Smote yalng aldal di *softwalre ralpid miner.* Paldal Galmbalr 6 dalpalt dilihalt balhwal daltal yalng diperoleh disimpaln dallalm sebualh file csv yalng alkaln dibalcal oleh *operaltor Reald csv,* dan *set role* yang akan menentukan label, sebelum diproses lebih lanjut dilakukan operasi *SMOTE Upsampling* untuk menyeimbangkan kelas data, lalu mengatur operator *Split Data* yang nantinya untuk membagi dataset menjadi data latih dan data uji, kemudian pada operator *Nominal to Text* berfungsi untuk mengubah semua angka pada dokumen menjadi sebuah text dan *Process Documents from data* adalah operator yang berfungsi untuk melakukan tahapan *preprocessing.* Selanjutnya hasil proses tersebut menggunakan *cross validation.*

**Galmbalr 6** Desalin model pengujialn allgoritmal menggunalkaln *Smote*

Setelalh olalh daltal dilalkukaln di *softwalre ralpid miner*, halsil pengujialn Allgoritmal Nalïve Balyes menggunalkaln *smote* menghalsilkaln nilali *Alccuralcy* 74,88%, *Precision* 81,91%, daln *Recalll* 63,88% seperti paldal galmbalr 7

****Galmbalr 7** Halsil pengujialnNB menggunalkaln *Smote*

**3.5 Pengujialn *Support Vector Malchine* talnpal *Smote***

Berikut aldallalh talmpilaln didallalm *Cross Vallidaltion* paldal operaltor SVM yalng aldal di *softwalre ralpid miner.* Kemudian masuk ke model desain algoritmanya diproses ke dalam operator *cross validation Support Vector Machine (SVM)* didalamnya ada perhitungan algoritmanya kemudian modelnya diapply setelah itu masuk ke penilaian performancenya barulah muncul hasil nilai *accuracy*, *Precision,* dan *Recalll.*

#

#  Galmbalr 8 Proses didallalm *Cross Vallidaltion*-SVM

Setelalh olalh daltal dilalkukaln di *softwalre ralpid miner*, halsil pengujialn Allgoritmal *Support Vector Malchine* talnpal menggunalkaln *smote* menghalsilkaln nilali *Alccuralcy* 68,20%, *Precision* 68.93%, daln *Recalll* 96,77% seperti paldal galmbalr 9.

**Galmbalr 9** Halsil pengujialn SVM talnpal *smote*

**3.6 Pengujialn *Support Vector Malchine* dengaln *Smote***

Peneralpaln Smote meminimallisalsi ketidalk seimbalngaln kelals paldal daltal,meneralpkaln penalngalnaln kelals daltal tidalk seimbalng dengaln smote sebalnyalk dual kalli paldal daltal laltih: Menentukaln nilali tetalnggal dengaln k=5, menghitung jalralk alntalr daltal kelals minoritals dengaln metode eucledialn, melalkukaln perhitungaln untuk membalngkitkaln daltal bualtaln (syntetic), meneralpkaln k-nealrest neighbor untuk mengklalsifikalsi daltal uji : Menentukaln nilali tetalnggal dengaln k=5, menghitung jalralk alntalr daltal kelals minoritals dengaln metode eucledialn, membalndingkaln kinerjal klalsifikalsi talnpal daln dengaln diteralpkalnnyal SMOTE [16].

# Setelalh olalh daltal dilalkukaln di *softwalre ralpid miner*, halsil pengujialn Allgoritmal *Support Vector Malchine* dengaln menggunalkaln *smote* menghalsilkaln nilali *Alccuralcy* 79,51%, *Precision* 73,13%, daln *Recalll* 93,32% seperti paldal galmbalr 10.

#

# Galmbalr 10 Halsil pengujialnSVM menggunalkaln *smote*

**Evallualsi**

Berdasarkan penelitian yang dilakukan pengujialn dilalkukaln menggunalkaln *softwalre ralpid miner studio educaltionall 9.10* dengan hasil sebagai berikut:

 Pada tabel 9 hasil pengujian *Naïve Bayes* dengan TF-IDF tanpa *smote* menghasilkan nilai *Alccuralcy* 60,69% dimana *Alccuralcy* adalah nilali ralsio daltal tweet yang di klasifikasikan benar dengan keseluruhan data yang sudah di klasifikasi, *Precision* 75,95% dimana *Precision* merupalkaln ralsio dari jumlah dokumen yang di klasifikasikan positif dan benar dibandingkan dengan keseluruhan hasil prediksi yang bernilai positif, daln *Recalll* 61,60% dimana *Recalll* aldallalh ralsio dari jumlah dokumen yang di kasifikasikan negatif dan benar dibandingkan dengan keseluruhan hasil prediksi yang bernilai negatif dan *Naïve Bayes* dengan TF-IDF menggunakan *smote Alccuralcy* 74,88%, *Precision* 81,91%, daln *Recalll* 63,88%.

|  |  |
| --- | --- |
| NB  | NB & Smote |
| Accuracy  | Precision | Recall  | Accuracy  | Precision | Recall  |
| 60,69% | 75,95% | 61,60% | 74,88% | 81,91% | 63,88% |

**Tabel 9.** *Naïve Bayes* tanpa *smote* danmenggunakan *smote*

Pada tabel 10 hasil pengujian *Support Vector Malchine* dengan TF-IDF tanpa *smote* menghasilkan nilai *Alccuralcy* 68,20%, *Precision* 68.93%, daln *Recalll* 96,77% dan *Support Vector Malchine* dengan TF-IDF menggunakan *smote* menghasilkan nilai *Alccuralcy* 79,51%, *Precision* 73,13%, daln *Recalll* 93,32%.

|  |  |
| --- | --- |
| SVM  | SVM & Smote |
| Accuracy  | Precision | Recall  | Accuracy  | Precision | Recall  |
| 68,20% | 68,93% | 96,77% | 79,51% | 73,13% | 93,32% |

**Tabel 10.** *Support Vector Machine* tanpa *smote* danmenggunakan *smote*

Sehingga dapat dilihat dari penelitian ini adalah *Support Vector Machine* dengan TF-DF menggunakan *Smote* memiliki tingkat akurasi paling tinggidiantara *Support Vector Machine* dengan TF-DF tanpa *Smote* dan *Naive Bayes* dengan TF-DF menggunakan *Smote* maupun *Naive Bayes* dengan TF-DF tanpa *Smote,* yaitu menghasilkan nilai *Accuracy* 79,51%, *Precision* 73,13%, dan *Recall* 93,32%.

***Word Cloud***

*Word cloud* aldallalh representalsi visuall dalri daltal teks, bialsalnyal digunalkaln untuk menggalmbalrkaln metal daltal *keywords* paldal sebualh website/ situs, untuk memvisuallisalsikaln sualtu bentuk teks secalral bebals. *Word cloud* alkaln menalmpilkaln kaltal dengaln frekuensi terbalnyalk, semalkin balnyalk frekuensi kaltal tersebut semalkin menonjol dibalnding kaltal yalng lalin.Paldal talhalpaln ini merupalkaln visuallisalsi halsil dalri daltal yalng telalh diolalh paldal *Ralpid Miner* berupal *Word Cloud* positif daln negaltif yalng dimalnal menunjukkaln frekuensi kaltal yalng sering muncul [4]. Seperti paldal galmbalr 11 dibalwalh berdasarkan visualisasi dapat di lihat kata sentimen positif yang sering muncul adalah ‘Cuan’, ‘Investasi’, ‘Binary’, ‘Uang’, ‘Analisa’, ‘Modal’, dimana pengguna merasa bahwa *binary option* dapat menjadi tempat untuk berinvestasi,menghasilkan uang, dan sentimen negatif kata yang sering muncul adalah ‘Binary’, ‘Robot’, ‘Main’, ‘Judi’, ‘Uang’, ‘Loss’, ‘Option’, ‘Kerja’, ‘Rugi’, dimana pengguna merasa bahwa *binary option* adalah judi,merugikan pengguna yang membuat pengguna tidak nyaman.

|  |  |
| --- | --- |
|  |  |
| **positif** | **Negaltif** |

# Galmbalr 11 Positif dan Negatif

# Kesimpulaln daln Salraln

# Paldal penelitialn ini dilalkukaln alnallisis sentimen menggunalkaln allgoritmal *Nalive balyes* (NB) daln *Support Vector Malchine* (SVM) talnpal smote daln menggunalkaln smote dallalm menentukaln sentiment malsyalralkalt terhaldalp binalry option. Dalam penelitian ini jugal menghalsilkaln jumlalh nilali sentiment positif 961 daln nilali sentiment negaltif sebesalr 434 sehinggal paldal penelitialn ini menggunalkaln metode smote untuk mengaltalsi ketidalk seimbalngaln dalta, l *smote* ini bekerja dengan membuat replikasi dari data minoritas. Replikasi tersebut dikenal dengan data sintetis (*syntetic* *data*). Metode *SMOTE* bekerja dengan mencari *k nearest neighbors* (yaitu ketetanggaan terdekat data sebanyak k) untuk setiap data di kelas minoritas, setelah itu dibuat data sintetis sebanyak *prosentase* duplikasi yang diinginkan antara data minor dan *knearest neighbor* yang dipilih secara acak. peneralpaln *smote* ini menunjukkaln perbalikaln performal dalri metode klalsifikalsi yalng alda. Hasil yang didapatkan dari penelitian ini adalah *Support Vector Machine* dengan TF-DF menggunakan *Smote* memiliki tingkat akurasi paling tinggidiantara *Support Vector Machine* dengan TF-DF tanpa *Smote* dan *Naive Bayes* dengan TF-DF menggunakan *Smote* maupun *Naive Bayes* dengan TF-DF tanpa *Smote,* yaitu menghasilkan nilai *Accuracy* 79,51%, *Precision* 73,13%, dan *Recall* 93,32%, yalng dimalnal dallalm hall ini dalpalt disimpulkaln balhwal binalry option mendalpalt palndalngaln positif dalri malsyalralkalt. Dallalm penelitialn selalnjutnyal algalr dalpalt melalkukaln kompalralsi beberalpal allgoritmal dengaln menggunalkaln pembobotaln daltal lalinnyal untuk mendalpaltkaln alkuralsi yalng lebih balik.

# Dalftalr Pustalkal

[1]     D. Puspitalsalri, F. R. Al. Ralchim, “BINAlRY OPTION SEBAlGAlI KOMODITI PERDAlGAlNGAlN BERJAlNGKAl DI INDONESIAl,” J. Hukum Lex Generallis, vol. 2, no. 8, Algustus 2021.

[2]     S. Syalm’un, S. O. Malnullalng, T. Br Sembiring, “Understalnding Binalry Options in the Context of Tralding Lalw in the Digitall Eral,” BIRCI, vol. 5, no. 2, Palge: 9052-9062 , Mey 2022.

[3]     R. I. Lestalri, Z. Alrifin, “Godalaln Pralktik Binalry Option Berkedok Investalsi Daln Tralding,” J. Ius Constituendum, vol. 7, no. 1, e-ISSN : 2580-8842, 2022

[4]     H. P. Doloksalribu, Y. T. Salmuel, “KOMPAlRAlSI AlLGORITMAl DAlTAl MINING UNTUK AlNAlLISIS SENTIMEN AlPLIKAlSI PEDULILINDUNGI,” J. Tekonologi Informalsi, vol. 16, no. 1, Jalnualri 2022, E-ISSN 2656-0321.

[5]     H. Hernalwalti, W. Galtal, “Sentimen Alnallisis Operalsi Talngkalp Talngaln KPK Menurut Malsyalralkalt Menggunalkaln Allgoritmal Support Vector Malchine, Nalive Balyes Berbalsis Palrticle Swalrm Optimizition,” F Exalctal, vol. 12, no. 3, hlm. 230-243, 2019, doi: 10.30998/falktorexalctal.v12i3.4992.

[6]     H. Setialwaln, E. Utalmi, S. Sudalrmalwaln, “Alnallisis Sentimen Twitter Kulialh Online Palscal Covid-19 Menggunalkaln Allgoritmal Support Vector Malchine daln Nalive Balyes,” J. Komtikal, vol. 5, no. 1, Mei 2021.

[7]     M. W. Pertiwi, “AlNAlLISIS SENTIMEN OPINI PUBLIK MENGENAlI SAlRAlNAl DAlN TRAlNSPORTAlSI MUDIK TAlHUN 2019 PAlDAl TWITTER MENGGUNAlKAlN AlLGORITMAl NAlÏVE BAlYES, NEURAlL NETWORK, KNN DAlN SVM,” I. Nusal Malndiri, vol. 14, no. 1, Algustus 2019, ISSN: 0216-6933.

[8]     D. Dalrwis, E. S. Praltiwi, Al. F. O. Palsalribu, “ PENERAlPAlN AlLGORITMAl SVM UNTUK AlNAlLISIS SENTIMEN PAlDAl DAlTAl TWITTER KOMISI PEMBERAlNTAlSAlN KORUPSI REPUBLIK INDONESIAl,” J. Ilmialh Edutic, vol. 7, no. 1, November 2020, p-ISSN 2407-4489 e-ISSN 2528-7303.

[9]     Al. P. Giovalni, Alrdialnsyalh, T. Halryalnti, L. Kurnialwalti, W. Galtal, “AlNAlLISIS SENTIMEN AlPLIKAlSI RUAlNG GURU DI TWITTER MENGGUNAlKAlN AlLGORITMAl KLAlSIFIKAlSI,” J. TEKNOINFO, vol. 14, no. 2, 2020, DOI : 10.33365/jti.v14i2.679.

[10]     N. K. Walrdhalni, Rezkialni, S. Kurnialwaln, H. Setialwaln, G. Galtal, S. Tohalri, W. Galtal, M. Walhyudi, “SENTIMENT AlNAlLYSIS AlRTICLE NEWS COORDINAlTOR MINISTER OF MAlRITIME AlFFAlIRS USING AlLGORITHM NAlIVE BAlYES AlND SUPPORT VECTOR MAlCHINE WITH PAlRTICLE SWAlRM OPTIMIZAlTION,” J. Theor. Alppl. Inf. Technol, vol. 96, no. 24, Desember 2018, pp. 8365–8378.

[11]     I. Ihsaln, “Sentiment Alnallysis RKUHP Paldal Twitter Menggunalkaln Metode Support Vector Malchine,” E-Proceeding Eng., vol. 8, no. 2, hlm. 3521–3536, Alpr 2021.

[12]     Al. S. Nugroho, Al. B. Witalrto, D. Halndoko, “Support Vector Malchine Teori daln Alplikalsinyal dallalm Bioinformaltikal,” K. Umum IlmuKomputer.com [https://alsnugroho.net/palpers/ikcsvm.pdf](https://asnugroho.net/papers/ikcsvm.pdf)

[13]     M. Alzhalri, Z. Situmoralng, alnd R. Rosnelly, “Perbalndingaln Alkuralsi, Recalll, daln Presisi Klalsifikalsi paldal Allgoritmal C4.5, Ralndom Forest, SVM daln Nalive Balyes,” *J. MEDIAl Inform. BUDIDAlRMAl*, vol. 5, no. 2, p. 640, Alpr. 2021, doi: 10.30865/mib.v5i2.2937.

[14]     M. Malulidalh, W. Galtal, R. Alulialnital, C. I. Algustyalningrum, “Allgoritmal Klalsifikalsi Decision Tree Untuk Rekomendalsi Buku Berdalsalrkaln Kaltegori Buku,” J. ILMIAlH EKONOMI DAlN BISNIS, Vol. 13, No. 2, Desember 2020, pp. 89 – 96.

[15]     J. Al. Septialn, T. M. Falhrudin, Al. Nugroho, "Alnallisis Sentimen Penggunal Twitter Terhaldalp Polemik Persepalkbolalaln Indonesial Menggunalkaln Pembobotaln TF-IDF daln K-Nealrest Neighbor" *Journall of Intelligent System alnd Computaltion*, , vol.1, no.1, pp. 43-49, Algustus 2019, doi: 10.52985/insyst.v1i1.36

[16]     R. Siringoringo, “ KLAlSIFIKAlSI DAlTAl TIDAlK SEIMBAlNG MENGGUNAlKAlN AlLGORITMAl SMOTE DAlN k-NEAlREST NEIGHBOR,” Jurnall ISD, vol. 3, no. 1, Jalnualri-Juni 2018, pISSN : 2477-863X eISSN: 2528-5114.