

Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS

Rani Puspita¹, Agus Widodo²

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

e-mail: ¹rani.puspita@binus.ac.id, ²ag45wd@binus.ac.id

Submitted Date: November 09th, 2020

Reviewed Date: December 30th, 2020

Revised Date: January 04th, 2021

Accepted Date: January 08th, 2021

Abstract

BPJS is really helpful because one of its goal is to provide good service for the member in terms of healthiness. But, when there's many people using the service, then it will cause more pros and contras. Therefore, researcher will be doing sentiment analysis in the field of data mining towards bpjs users on social media Twitter as much as 1000 data that later will be filtered to be 903 data because there are some data that has been duplicated. Researchers used the KNN, Decision Tree, and Naïve Bayes methods to compare the accuracy of the three methods. Researchers used the RapidMiner version 9.7.2 tools. The results showed that the sentiment analysis of Twitter data on BPJS services using the KNN method reached an accuracy level of 95.58% with class precision for pred. negative is 45.00%, pred. positive is 0.00%, and pred. neutral is 96.83%. Then the Decision Tree method the accuracy rate reaches 96.13% with the precision class for pred. negative is 55.00%, pred. positive is 0.00%, and pred. neutral is 97.28%. And the last one is the Naïve Bayes method which achieves 89.14% accuracy with precision class for pred. negative is 16.67%, pred. positive was 1.64%, and pred. neutral is 98.40%.

Keywords: Sentiment Analysis; BPJS; Twitter; Data Mining

Abstrak

BPJS sangat membantu dalam hal kesehatan. Namun ketika ada banyak orang yang menggunakan layanan tersebut, maka akan banyak pula pro dan kontra yang didapatkan. Oleh karena itu peneliti melakukan analisis sentimen dalam bidang data mining terhadap pengguna BPJS pada media sosial Twitter sebanyak 1000 data yang kemudian difiltering menjadi 903 data dikarenakan adanya data yang terduplikat. Penulis menggunakan metode KNN, *Decision Tree*, dan Naïve Bayes untuk membandingkan tingkat akurasi dari ketiga metode tersebut. Peneliti menggunakan *tools* RapidMiner versi 9.7.2. Hasil penelitian menunjukkan bahwa analisis sentimen terhadap data Twitter terhadap layanan BPJS dengan menggunakan metode KNN mencapai tingkat akurasi 95.58% dengan *class precision* untuk *pred. negative* adalah 45.00%, *pred. positive* adalah 0.00%, dan *pred. neutral* adalah 96.83%. Lalu pada metode *Decision Tree* tingkat akurasinya mencapai 96.13% dengan *class precision* untuk *pred. negative* adalah 55.00%, *pred. positive* adalah 0.00%, dan *pred. neutral* adalah 97.28%. Dan yang terakhir adalah metode Naïve Bayes yang mencapai akurasi 89.14% dengan *class precision* untuk *pred. negative* adalah 16.67%, *pred. positive* adalah 1.64%, dan *pred. neutral* adalah 98.40%.

Keywords: Analisis Sentimen; BPJS; Twitter; *Data Mining*

1. Pendahuluan

Badan Penyelenggara Jaminan Sosial atau BPJS adalah badan hukum yang dibuat untuk dapat menyelenggarakan program jaminan untuk kesehatan. (Linda, Haskas, & Kadrianti, 2020). Pada dasarnya BPJS sangatlah membantu

karena salah satu tujuan dari BPJS itu sendiri adalah untuk memberikan layanan yang baik bagi peserta dalam hal kesehatan. Tetapi semakin banyak penduduk yang menggunakan layanan tersebut, semakin banyak pula pro dan kontra dari masyarakat Indonesia. Mulai dari yang pro karena

merasa tidak terbebani perihal biaya jika berobat ke suatu RS ataupun yang kontra karena merasa tidak terlalu diprioritaskan.

Maka dari itu perlu diketahui opini dari masyarakat Indonesia mengenai layanan dari BPJS. Biasanya masyarakat mengemukakan pendapatnya melalui sosial media. Salah satu aplikasi yang sering digunakan masyarakat Indonesia adalah *Twitter*. *Twitter* adalah sebuah *platform* untuk menyampaikan opini atau pendapat seseorang. Pertumbuhan pengguna *Twitter* di Indonesia sangat pesat dan menduduki peringkat 5 di dunia. (Utami, 2020). *Twitter* banyak digunakan orang untuk menyampaikan keluh kesahnya mulai dari keluh kesah mengenai kehidupan sehari-hari ataupun keluh kesah terhadap layanan yang diberikan baik dari pemerintah atau bidang lainnya.

Oleh karena itu sangat efisien jika menggunakan *Twitter* sebagai media untuk mengambil data mengenai keluhan masyarakat Indonesia terkait layanan BPJS.

Berdasarkan studi literatur yang peneliti amati, penelitian mengenai Sentimen Analisis yang dilakukan oleh Denty Welmin dkk tentang: “*Analysis of User Sentiment of Twitter to RUU KUHP*”. (Cahyaningrum et al., 2020). Dalam penelitian ini dapat diketahui bahwa peneliti menggunakan *R studio* untuk bahasa pemrograman. Akan tetapi penelitian ini tidak menjelaskan secara spesifik metode apa yang digunakan dalam penelitian tersebut.

Penelitian lainnya yang dilakukan oleh Retno Sari tentang Analisis Sentimen dapat diketahui bahwa peneliti menggunakan metode *KNN*. Akan tetapi tidak dijelaskan menggunakan *tools* apa dalam penelitiannya. Dan dalam penelitian ini juga hanya terdapat satu metode yang dilakukan oleh penelitiannya. (Sari, 2020)

Dalam penelitian lainnya yang dilakukan oleh Ghulam Asrofi Buntoro tentang Analisis Sentimen (Buntoro, 2017) dapat diketahui bahwa peneliti menggunakan metode *SVM*. Tetapi tidak dijelaskan *tools* dalam analisisnya.

Perlu diketahui bahwa sentimen analisis adalah cabang dari *data mining*. *Data Mining* merupakan sebuah proses yang dapat mengekstrak informasi sehingga menghasilkan informasi yang sangat berharga. (Nurdin, 2017). Dengan kata lain dapat juga dikatakan bahwa *data mining* merupakan proses untuk mencari informasi mengenai teknik tertentu. Teknik dan metode dalam *data mining* sangat banyak. Oleh karena itu, dalam pemilihan teknik atau algoritma yang tepat

akan sangat bergantung pada tujuan yang diinginkan.

Dengan betitu, peneliti menggunakan tiga metode dalam analisis sentimen untuk memperbandingkan tingkat akurasi dari ketiga metode tersebut diantaranya adalah metode *KNN*, *Decision Tree* dan *Naïve Bayes*.

Algoritma *KNN* adalah salah satu algoritma yang sudah *popular*. *KNN* ini termasuk ke dalam grup *instance-based learning*. Metode *KNN* merupakan teknik *lazy learning*. (Cahyanti, Rahmayani, & Ainy, 2020). Maksudnya adalah metode ini digunakan dalam klasifikasi data yang jaraknya dekat. Ada juga yang berpendapat bahwa algoritma *KNN* adalah algoritma pembelajaran yang banyak digunakan dalam sistem *cyber-fisik-sosial* (CPSS) untuk menganalisis dan menambang data (*main data*). (Zhang, Chen, Liu, & Xi, 2020).

Selain metode *KNN*, peneliti juga menggunakan metode *Decision Tree*. Algoritma *Tree* biasa dipakai untuk pengenalan pola statistik. (Sarimuddin et al., 2020). *Decision Tree* terbuat dari tiga simpul yaitu *leaf*, lalu terdiri juga dari simpul *root* yang merupakan titik awal dari suatu *decision tree*, dan yang terakhir adalah simpul perantara yang berhubungan dengan suatu pengujian.

Selain menggunakan metode *KNN* dan *Decision Tree*, peneliti juga menggunakan metode *Naïve Bayes*. *Naïve Bayes* adalah metode *machine learning* untuk probabilitas. Dalam kata lain, *Naïve Bayes* merupakan metode untuk klasifikasi *text* dengan kecepatan pemrosesan yang tinggi jika dalam data besar. (Fitriyyah, Safriadi, & Pratama, 2019). Ada juga yang berpendapat bahwa *Naïve Bayes* adalah metode yang digunakan untuk prediksi karena mengandung probabilitistik sederhana yang diterapkan pada teorema bayes dengan ketergantungan yang kuat. (Sinaga, Sawaluddin, & Suwilo, 2020).

Atas dasar latar belakang tersebut, maka dilakukannya penelitian ini untuk ke tiga metode tersebut dengan mengintegrasikan *Twitter* sebagai *platform* untuk peneliti melakukan pengolahan data untuk mengetahui analisis sentimen terhadap layanan BPJS dengan *data mining*.

Masalah dalam penelitian ini adalah bagaimana cara peneliti mengumpulkan data dari *Twitter* dan berapa tingkat akurasi dari metode *KNN*, *Decision Tree* dan *Naïve Bayes*.

Selain itu, tujuannya adalah untuk mengetahui cara mengumpulkan data *Twitter*

menggunakan RapidMiner dan mengetahui tingkat akurasi dari ketiga metode yang diteliti.

Penelitian ini merupakan ide yang penulis buat sendiri sebagai bahan penelitian untuk menganalisis sentimen terhadap layanan BPJS dengan cara mengumpulkan data dari Twitter dan menganalisis data tersebut dengan *tools* tertentu.

2. Metodologi

2.1 Metode Pengumpulan data

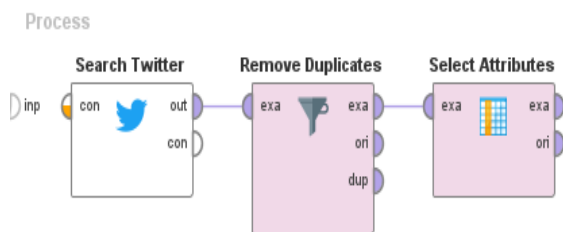
Metode yang dilakukan oleh penulis adalah sebagai berikut:

a. Studi Literatur

Dalam studi literatur ini, peneliti mempelajari materi terkait penelitian dari referensi yang terpercaya. Setelah itu, peneliti akan mencari informasi yang dibutuhkan dalam penelitian ini. Informasi yang didapatkan akan digunakan dalam penyusunan. Semua pustaka yang dijadikan acuan dapat dilihat pada halaman daftar pustaka.

b. Pengumpulan Data

Dalam hal ini, data dikumpulkan dari sosial media Twitter. Peneliti menggunakan RapidMiner versi 9.7.2 sebagai alat untuk mengambil data pada Twitter terkait pendapat orang terhadap layanan BPJS. Berikut adalah gambaran cara proses *crawling* data dari Twitter menggunakan RapidMiner.



Gambar 1. Proses *Crawling* Dan *Filtering* Data

Langkah pertama adalah pengumpulan data atau yang bisa disebut *crawling* data. Peneliti melakukan *crawling* data pada sosial media Twitter dengan limit 1000. Kemudian peneliti melakukan proses *filtering* data dengan cara *remove duplicate* karena terdapat data yang *double*. Setelah peneliti melakukan *remove duplicate* pada data yang terambil *double*, dapat dihasilkan data akhir yang didapatkan yaitu sebanyak 903 data.

Setelah melakukan proses *filtering* data dengan cara *remove duplicate* dilanjutkan dengan

dilakukannya *select attributes* yang bertujuan untuk memilih *text* pada data.

Dan proses terakhir pada *crawling* data ini adalah data akan di *export* dalam bentuk excel dan kemudian dilakukan pelabelan untuk data yang bersifat positif, negatif atau netral.

2.2 Peralatan Dalam Penelitian

Peralatan dalam penelitian terbagi menjadi dua. Berikut rincian peralatan penelitian yang digunakan:

a. Perangkat keras:

1. HP Laptop 14s-cf0055TX.
2. RAM 8 GB.

b. *Software*:

1. OS Windows 10 Pro 64bit.
2. Menggunakan RapidMiner versi 9.7.2.
3. Sistem operasi windows 10 profesional.
4. Microsoft Excel.

2.3 Hasil Pengumpulan Data

| Row No. | Id | Created_At | From_User | From_User_Id | To_User | To_User_Id | Language | Source | Text |
|---------|------------|--------------|-----------------|--------------|-------------|------------|----------|------------------|---------|
| 1 | 1304281518 | Sep 11, 2020 | BPJS Keseh... | 1522337929 | ? | -1 | in | <a href="http:// | Dam... |
| 2 | 1303558895 | Sep 9, 2020 | Kompas.com | 22343990 | ? | -1 | in | <a href="http:// | Beni... |
| 3 | 1302834008 | Sep 7, 2020 | VICE Indones... | 7651499553 | ? | -1 | in | <a href="http:// | Men... |
| 4 | 1304354002 | Sep 11, 2020 | Vanus | 8001663820 | jakowi | 369887179 | in | <a href="http:// | @ol... |
| 5 | 1304353947 | Sep 11, 2020 | Hae | 1080020119 | ? | -1 | in | <a href="http:// | Huda... |
| 6 | 1304353884 | Sep 11, 2020 | tanfan | 1021384408 | ? | -1 | in | <a href="http:// | RT @... |
| 7 | 1304353831 | Sep 11, 2020 | SA | 1074455574 | ? | -1 | in | <a href="http:// | RT @... |
| 8 | 1304353485 | Sep 11, 2020 | meawed | 193281316 | CNIndonesia | 17128875 | in | <a href="http:// | @CF... |
| 9 | 1304352988 | Sep 11, 2020 | Chro | 9895214008 | nanalajun | 1176165763 | in | <a href="http:// | @na... |
| 10 | 1304352848 | Sep 11, 2020 | Nana' @ #C | 1176165763 | ? | -1 | in | <a href="http:// | Co... |
| 11 | 1304352683 | Sep 11, 2020 | Ge | 714506480 | Shelima_AB | 190147376 | in | <a href="http:// | @SP... |
| 12 | 1304352119 | Sep 11, 2020 | PAKA MASK | 60481794 | ? | -1 | in | <a href="http:// | RT @... |
| 13 | 1304351703 | Sep 11, 2020 | nic | 1278613305 | ? | -1 | in | <a href="http:// | bro... |
| 14 | 1304351698 | Sep 11, 2020 | Jasjus | 1142983010 | ? | -1 | in | <a href="http:// | RT @... |

Gambar 2. Tahap Awal Pengumpulan Data Twitter

Pada gambar 2, terdapat *language*, *source*, *logitude*, *latitude*, *text* dan lain sebagainya. Karena hanya membutuhkan *text*, maka peneliti menghilangkan *row* selain *text*.

Selain *result* berupa tabel yang ada pada Gambar 2, terdapat pula hasil atau *result* yang berbentuk grafik pada Gambar 3.

| Field | Type | Min | Max | Count |
|------------------------|-----------|----------------------|-----------------------------|-------------------------------------|
| Id | Nominal | 1304354002 | 1302834008 | 1302834008 |
| Created_At | Date-time | Sep 7, 2020 12:00 PM | Sep 11, 2020 4:40 PM | 44 dr 306 Sls |
| From_User | Nominal | BPJS Keseh... | DANA id (2), Dan Rachma (2) | 302 more |
| From_User_Id | Nominal | 86602140083070560 | 865871936545816577 | 122849151221046248 (2), 302 more |
| To_User | Nominal | yurandri (1) | BPJS Keseh... (5) | BPJS KesehatanRI (2), 225 more |
| To_User_Id | Nominal | 95964624162090962 | -1 (65) | -1 (65), 542377830 (5), 225 more |
| Language | Nominal | in (100) | in (100) | in (100) |
| Source | Nominal | <a href="http:// | <a href="http:// | <a href="http:// |
| Text | Nominal | mayarak 12748 (1) | RT @hnl any... (22) | RT @hnl et negar... (5), 358 more |
| Geo-Location-Latitude | Nominal | ? | ? | ? |
| Geo-Location-Longitude | Nominal | ? | ? | ? |
| Retweet-Count | Integer | 0 | 60 | 30-480 |

Gambar 3. Grafik Pengumpulan Data

Gambar 3 adalah hasil dari *filtering* data. Sehingga hanya tersisa *row number*, ID, dan isi *text*.

| Row No. | Id | Text |
|---------|---------------|--|
| 1 | 1304281518... | Darpada repot-repot keluar rumah dan pergi ke kantor cabang BPJS Kesehatan untuk urusan keperluan administrasi, mending kalaa... |
| 2 | 1303558895... | Berikut ini jadwal pencairan BLT subsidi gaji Rp 500.000 atau bantuan BPJS dalam program Bantuan Subsidi Upah lewat BP. Jam... |
| 3 | 1302834006... | Menurut Erick Thohir, pemerintah hanya mensubsidi vaksin gratis kepada peserta PBI BPJS Kesehatan. Warga lain bakal mengiku... |
| 4 | 1304352631... | RT @SireTjrnMarching: Ingin fasilitas pendidikan & BPJS murah serta dukungan selama pandemi, tapi mengulak ideologi kiri. Pa... |
| 5 | 1304353495... | @CNNIndonesia long gw udah bayar bpjs kesehatan... kalo cuma buat vaksin aja harus bayar ra kebangetan banget lo... |
| 6 | 1304352886... | @nanaleejn Anji bpjs ☹ |
| 7 | 1304352846... | Coba aja beli album bisa pake bpjs ☹ |
| 8 | 1304352883... | @Shelma_AB Sebelum musun bpjs jajanin kiri ada wafiq yg jualan kelapa muda, nah dia jualan lutek juga tp enaknya lamun anu... |
| 9 | 1304351703... | bro mau tanya dong, yg udah kebagian blt bpjs luran bulannya udah terbayar belum untuk bulan ini? atau yg blom bayar bisa da... |
| 10 | 1304351680... | RT @whoistmi: @rasamochacino 1. ga semua org gajinya diatas umr |
| 11 | 1304351664... | BPJS Kesehatan Sumenep Capai Target Kepesertaan https://t.co/kmgH1Yk2hD |
| 12 | 1304351308... | @sarambahngaprak @IeQ_Online @pjskIndonesia BPJS pojok ikut bang |
| 13 | 1304351065... | Penyesuaian luran Program Jaminan Sosial Ketenagakerjaan Selama Penyebaran COVID-19 berdasarkan PP No.49 Tahun 2020... |
| 14 | 1304351024... | @HANEYOHAN Pate bpjs bisa g c... |

Gambar 4. Hasil Pengumpulan Data Dari Twitter

Gambar 4 adalah contoh dari data yang berhasil di *crawling* dari Twitter dan yang sudah peneliti hilangkan selain *row Text*. Sehingga sekarang hanya terdapat *Row number* berupa ID dan *text*. Hal ini dikarenakan *Row Number* dan ID wajib dan tidak bisa dihilangkan.

Selain *result* berupa tabel yang ada pada Gambar 4, terdapat pula hasil atau *result* yang berbentuk grafik pada Gambar 5.

| id | Normal | Last | Rate | Value |
|-----------------------|--------|---------------------|---|----------------------------------|
| 1304352119189154753 | 0 | 1302834006876283248 | 1302834006876283248 (1) | 130355889546236608 (1) [56 more] |
| Text | Normal | Last | Rate | Value |
| masyrak [] 127a5 (1) | 0 | "BPJS []]bear (1) | "BPJS []]beawear (1) @BPJSkes []]bp ya? (1) ...67 more] | |

Gambar 5. Grafik Hasil Pengumpulan Data Twitter

2.4 Preprocessing

Seperti yang telah dijelaskan sebelumnya, bahwa data yang diperoleh dalam penelitian ini adalah sebanyak 1000 *record* Tweet mengenai layanan BPJS dari pengumpulan data menggunakan Twitter dengan *tools* RapidMiner. Untuk mendapatkan data yang baik, maka berikut ini adalah beberapa teknik *preprocessing* yang peneliti gunakan:

a. Data Validation

Pada tahap ini, peneliti mengidentifikasi sekaligus menghapus data yang sekiranya tidak digunakan, data non konsisten dan data yang *missing*.

b. Data Integration dan Transformation

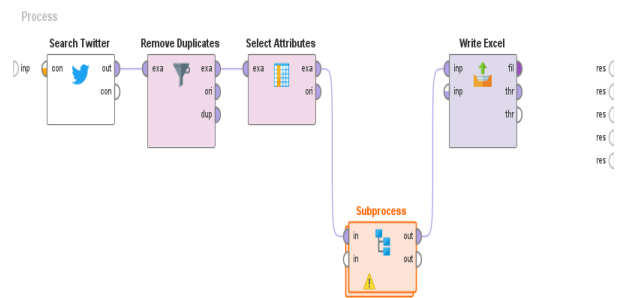
Pada tahap ini, peneliti meningkatkan akurasi dari metode yang digunakan oleh peneliti.

c. Data Size Redution and Dcretization

Pada tahap ini, peneliti merapikan data yang berhasil dikumpulkan dengan cara *men-delete* yang terduplikat.

2.4.1 Proses awal

Pada proses ini dilakukan proses awal terhadap data yang telah dikumpulkan dengan *tools* RapidMiner. Berikut adalah gambaran cara kerjanya:



Gambar 6. Proses Awal

Gambar 6 adalah contoh proses awal yang peneliti lakukan pada *tools* RapidMiner.

Terdapat *search* Twitter, *remove duplicate*, *replace attributes*, *sub process*, dan *write excel*.

2.4.2 Sub Proses



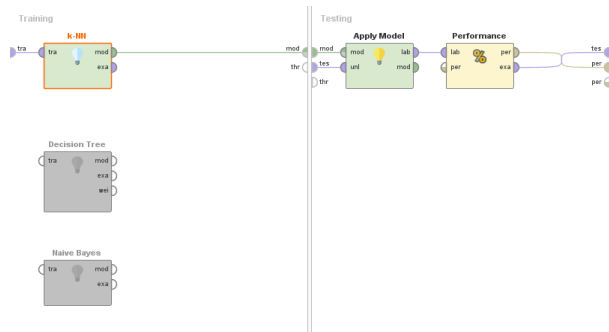
Gambar 7. Sub Proses

Gambar 7 adalah contoh *sub process* awal yang peneliti lakukan pada *tools* RapidMiner. Ada *action* untuk *replace*, dimana *replace* bertujuan untuk menghilangkan sesuatu yang tidak akan digunakan seperti tanda *hashtag* pada *text*.

2.5 Proses Cross Validation

Cross validation ini adalah sebuah *action* yang peneliti lakukan untuk mencari akurasi dari masing-masing metode dengan membagi data berupa *training* dan *testing*. berupa *apply model* dan *performance*.

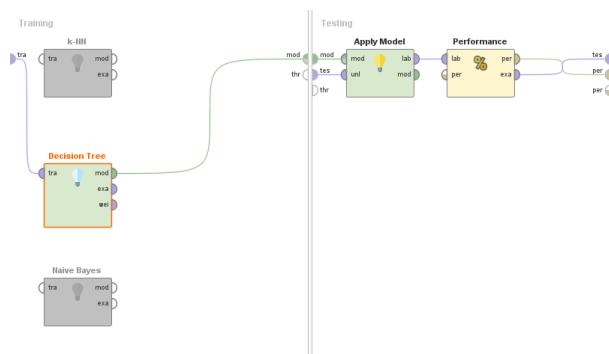
2.5.1 Proses Cross Validation KNN



Gambar 8. Proses Cross Validation KNN

Gambar 8 adalah *step* untuk *cross* dalam algoritma KNN. Pada kolom data *training* terdapat *action* metode KNN. Dan pada kolom data *testing* terdapat *action* berupa *apply* model dan *performance*. Hal ini bertujuan untuk mengetahui tingkat akurasi data pada metode KNN.

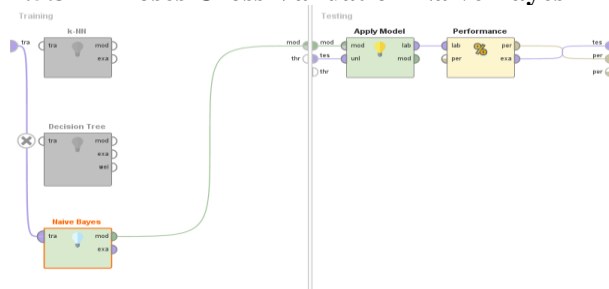
2.5.2 Proses Cross Validation Decision Tree



Gambar 9. Proses Cross Validation Decision Tree

Gambar 9 adalah *step* untuk *cross validation* dalam algoritma Decision Tree. Pada kolom data *training* terdapat *action* metode Decision Tree. Dan pada kolom data *testing* terdapat *action* berupa *apply* model dan *performance*. Hal ini bertujuan untuk mengetahui tingkat akurasi data pada metode Decision Tree.

2.5.3 Proses Cross Validation Naïve Bayes

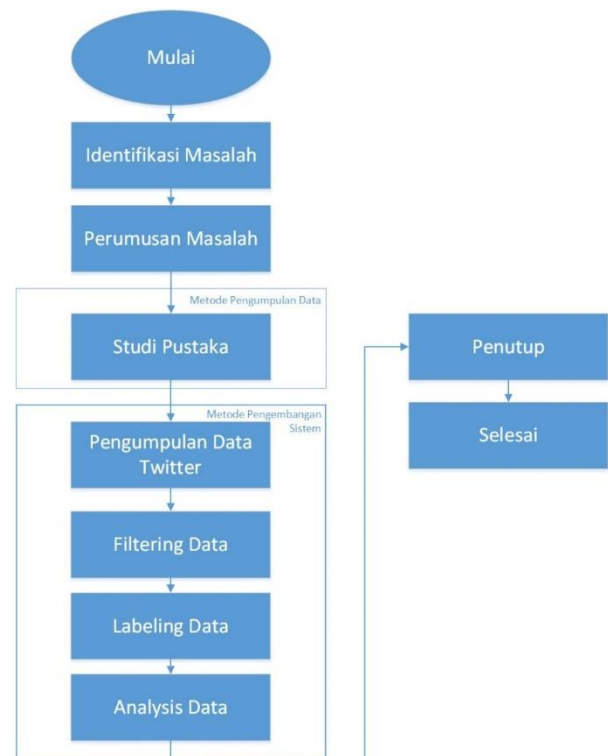


Gambar 10. Proses Cross Validation Naive Bayes

Gambar 10 adalah *step* untuk *cross validation* dalam algoritma Naïve Bayes. Pada kolom data *training* terdapat *action* metode Naive Bayes. Pada data *testing* terdapat *apply model* dan *performance*. Hal ini bertujuan untuk mengetahui tingkat akurasi data pada metode Naive Bayes.

2.6 Alur Penelitian

Di bawah ini merupakan alur penelitian dalam jurnal yang peneliti buat:



Gambar 11. Alur Penelitian

Berikut ini adalah penjelasan mengenai alur penelitian:

- Identifikasi Masalah**
 Pada tahapan ini, peneliti mengidentifikasi masalah yang ingin diuji dalam penelitian ini.
- Rumusan Masalah**
 Pada tahapan ini, peneliti melakukan perumusan masalah. Dalam hal ini, masalah yang ingin diuji adalah bagaimana cara mengumpulkan data melalui Twitter dengan *tools* RapidMiner dan berapa tingkat akurasi dari metode KNN, Decision Tree, dan Naïve Bayes.
- Metode Pengumpulan Data**
 Pada tahap ini, data yang dimaksud adalah seperti pengertian-pengertian atau penjelasan teoritis mengenai apa itu *data mining*, apa itu

metode KNN, apa itu metode Decision Tree, dan apa itu metode Naïve bayes. Maka dari itu peneliti hanya menggunakan studi pustaka pada tahapan ini. Studi pustaka berupa jurnal nasional dan internasional.

d. Pengumpulan Data Twitter

Pada tahapan ini sering disebut dengan *crawling* data. Data yang akan dikumpulkan adalah sebanyak 1000 data. Proses ini tidak memakan waktu yang lama. Tergantung Device yang digunakan untuk mengambil data. Dan pada tahap ini harus sudah mempunyai akun Twitter untuk mendapatkan akses token yang akan *diinput* pada *tools* RapidMiner.

e. *Filtering* Data

Pada tahapan ini, peneliti melakukan proses *filtering* karena terdapat data yang *double*. Sehingga data yang tadinya 1000 sekarang hanya 903 data yang tersisa.

f. *Labeling* Data

Pada tahapan ini, peneliti melakukan proses *Labeling* Data untuk mengetahui sifat dari *text* yang sudah dikumpulkan. Labeling ini terdapat label positif, negatif, dan netral. Lalu peneliti dapat mengatur untuk memberi warna pada label untuk memudahkan.

g. *Analysis* Data

Pada tahapan ini, peneliti menganalisis akurasi dari ketiga metode yang digunakan.

3. Hasil dan Pembahasan

Hasil dan pembahasan mengenai sentimen analisis terhadap layanan BPJS dengan metode KNN, Decision Tree, dan Naïve Bayes yang peneliti olah menggunakan RapidMiner maka contoh *crawling* data yang sudah dilakukan *labeling* dan nilai akurasi dari setiap metode, dan dijelaskan pada subbab berikutnya.

3.1 *Labeling*

| Row No. ↑ | Text | Id | Sentiment |
|-----------|--|------------|-----------|
| 1 | Jolow Turunkan luran Jaminan Sosial di BPJS Ketenagakerjaan https://t.co/Wz3vOgRt1 | 1302922782 | Negative |
| 2 | Hari ini Kemaker telah menerima 3.5 data calon penerima subsidi gajiupah tahap III dari BPJS Ket... | 1303315876 | Positive |
| 3 | Menurut Erick Thohir, pemerintah hanya mensubsidi vaksin gratis kepada peserta FBI BPJS Kesehatan... | 1302834006 | Neutral |
| 4 | Ditakarung punya bpjs ga ya? | 1303543981 | Neutral |
| 5 | @mandicare bilang bagaimana ini mandiri & @BPJSKesehatanRI rek saya sudah terbelah dri lgi 7... | 1303543670 | Neutral |
| 6 | @byfess Puskesmas gratis pale bpjs | 1303543470 | Neutral |
| 7 | @bisabahnaga @infomatang Untuk kantor BPJS cuman di Jl. Tumenggung Suryo. Nah kalau BPJS K... | 1303543448 | Neutral |
| 8 | @BPJSTKinfo Bisa, tapi ini kartu bpjs yg lama sudah diambil, yg baru ga bisa ditambahkan, mau daf... | 1303543359 | Negative |
| 9 | sumpah ini btl bpjs hanya blm klsuh dan hilang rasa keharmonisan antara karyawan, HRD dan sta... | 1303543135 | Negative |
| 10 | RT @liputan1dotcom: Besar luran BPJS Ketenagakerjaan Dilonggarkan hingga Akhir Bulan https://t.c... | 1303543052 | Positive |
| 11 | RT @Pahlawankonoha: Bayangin bapak kita cape-cape blm BPJS dan kita malah ga saint. Berasa... | 1303542977 | Neutral |
| 12 | @BPJSTKinfo Dear Bpjs, saya sudah mendaftar untuk oengambilan jht di jarkial nya validasi n... | 1303542968 | Negative |
| 13 | @BPJSTKinfo sumpah gai jelas lama lama bpjs cabang gamoir. bisa minta no telfon kantor aditny... | 1303542932 | Negative |

Gambar 12. *Labeling* Data

Gambar 12 adalah contoh dari data yang sudah *labeling*. Jika sebelumnya hanya berupa *row*, *ID*, dan *text*. Maka pada tahap ini akan ada satu tabel lagi yang bernama *sentiment*. Label ini digunakan untuk mengetahui isi dari *text* yang sudah di *crawling* tersebut bersifat apa. Terdapat tiga sifat data yaitu positif, negatif, dan netral.

Selain *result* dalam bentuk tabel, kita juga dapat melihat result dalam bentuk grafik.

| Text | Positif | Netral | Negative |
|------------|---------|--------|----------|
| Count | 9 | 854 | 2 |
| Percentage | 28.12% | 98.96% | 0.00% |

Gambar 13. Grafik *Labeling* Data

3.2 Hasil *Accuracy* Algoritma KNN

accuracy: 95.58% +/- 2.08% (micro average: 95.57%)

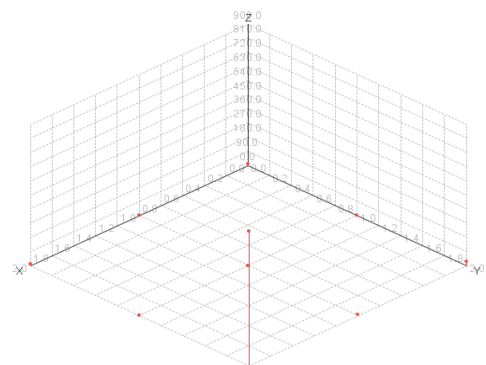
| | true Negative | true Positive | true Neutral | class precision |
|----------------|---------------|---------------|--------------|-----------------|
| pred. Negative | 9 | 2 | 9 | 45.00% |
| pred. Positive | 1 | 0 | 0 | 0.00% |
| pred. Neutral | 22 | 6 | 854 | 96.83% |
| class recall | 28.12% | 0.00% | 98.96% | |

Gambar 14. Hasil *Accuracy* KNN

Hasil pada metode Decision Tree menyatakan bahwa tingkat akurasi pada metode ini sebesar adalah 95.58%. Dimana *class precision* untuk *pred. negative* adalah 45.00%, *pred positive* adalah 0.00%, dan *pred. neutral* adalah 96.83%.

Lalu juga terdapat *plot view* dari algoritma ini. Berikut gambarnya:

Confusion Matrix (x: true class, y: pred. class, z: counters)



Gambar 15. *Plot View* Metode KNN

Selain tabel, terdapat juga *performance vector* untuk metode KNN.

PerformanceVector

PerformanceVector:
 accuracy: 95.58% +/- 2.08% (micro average: 95.57%)
 ConfusionMatrix:
 True: Negative Positive Neutral
 Negative: 9 2 9
 Positive: 1 0 0
 Neutral: 22 6 854
 kappa: 0.269 +/- 0.324 (micro average: 0.326)
 ConfusionMatrix:
 True: Negative Positive Neutral
 Negative: 9 2 9
 Positive: 1 0 0
 Neutral: 22 6 854

Gambar 16. Performance Vector Metode KNN

Performance vector sendiri merupakan bentuk deskripsi dari tabel hasil analisis. Hanya saya terdapat tambahan seperti kappa pada *performance vector* metode KNN ini.

3.3 Hasil Accuracy Algoritma Decision Tree

accuracy: 96.13% +/- 0.93% (micro average: 96.12%)

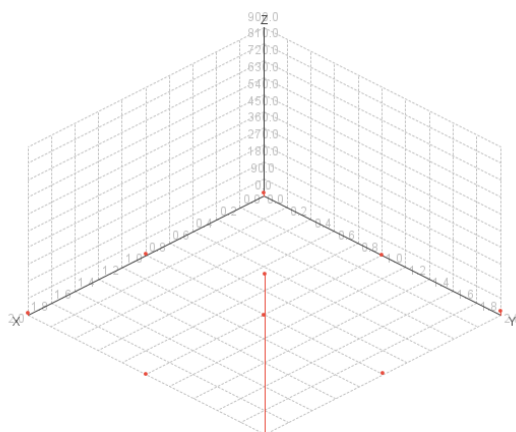
| | true Negative | true Positive | true Neutral | class precision |
|----------------|---------------|---------------|--------------|-----------------|
| pred. Negative | 11 | 3 | 6 | 55.00% |
| pred. Positive | 2 | 0 | 0 | 0.00% |
| pred. Neutral | 19 | 5 | 857 | 97.28% |
| class recall | 34.38% | 0.00% | 99.30% | |

Gambar 17. Hasil Accuracy Decision Tree

Hasil pada metode KNN menyatakan bahwa tingkat akurasi pada metode ini adalah 96.13%. Dimana *class precision* untuk *pred. negative* adalah 55.00%, *pred. positive* adalah 0.00%, dan *pred. neutral* adalah 97.28%.

Selain itu juga terdapat *performance vector* dari metode Decision Tree. Lalu juga terdapat *plot view* dari algoritma ini.

Confusion Matrix (x: true class, y: pred. class, z: counters)



Gambar 18. Plot View Metode Decision Tree

PerformanceVector

PerformanceVector:
 accuracy: 96.13% +/- 0.93% (micro average: 96.12%)
 ConfusionMatrix:
 True: Negative Positive Neutral
 Negative: 11 3 6
 Positive: 2 0 0
 Neutral: 19 5 857
 kappa: 0.362 +/- 0.248 (micro average: 0.420)
 ConfusionMatrix:
 True: Negative Positive Neutral
 Negative: 11 3 6
 Positive: 2 0 0
 Neutral: 19 5 857

Gambar 19. Performance Vector Metode Decision Tree

Performance vector sendiri merupakan bentuk deskripsi dari tabel hasil analisis. Hanya saya terdapat tambahan seperti kappa pada *performance vector* metode Decision Tree ini.

3.4 Hasil Accuracy Algoritma Naïve Bayes

accuracy: 89.14% +/- 2.78% (micro average: 89.15%)

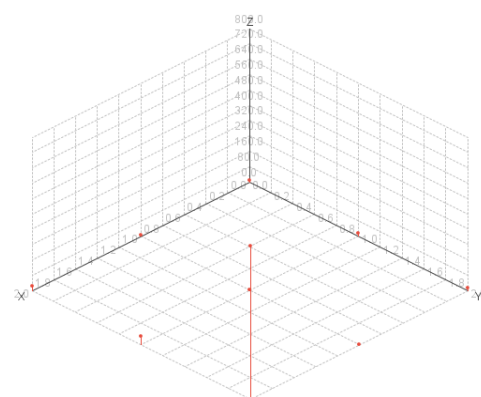
| | true Negative | true Positive | true Neutral | class precision |
|----------------|---------------|---------------|--------------|-----------------|
| pred. Negative | 5 | 5 | 20 | 16.67% |
| pred. Positive | 16 | 1 | 44 | 1.64% |
| pred. Neutral | 11 | 2 | 799 | 98.40% |
| class recall | 15.62% | 12.50% | 92.58% | |

Gambar 20. Hasil Accuracy Naïve Bayes

Hasil pada metode Naïve Bayes menyatakan bahwa tingkat akurasi pada metode adalah 89.14%. Dimana *class precision* untuk *pred. negative* adalah 16.67%, *pred. positive* adalah 1.64%, dan *pred. neutral* adalah 98.40%.

Selain itu juga terdapat *performance vector* untuk metode Naïve Bayes. Lalu juga terdapat *plot view* dari algoritma ini.

Confusion Matrix (x: true class, y: pred. class, z: counters)



Gambar 21. Plot View Metode Naïve Bayes

PerformanceVector

```
PerformanceVector:
accuracy: 89.14% +/- 2.78% (micro average: 89.15%)
ConfusionMatrix:
True:  Negative      Positive      Neutral
Negative:      5          5          20
Positive:     16         1          44
Neutral:     11         2          799
kappa: 0.220 +/- 0.062 (micro average: 0.218)
ConfusionMatrix:
True:  Negative      Positive      Neutral
Negative:      5          5          20
Positive:     16         1          44
Neutral:     11         2          799
```

Gambar 22. Performance Vector Metode Naive Bayes

Performance vector sendiri merupakan bentuk deskripsi dari tabel hasil analisis. Hanya saya terdapat tambahan seperti kappa pada performance vector metode Naive Bayes ini.

4. Kesimpulan

Tujuan dari penelitian yang dilakukan ini adalah untuk mengetahui tingkat akurasi dari tiga metode berbeda diantaranya KNN, Decision Tree dan Naive Bayes, dengan menggunakan *tools* RapidMiner. *Tools* yang dipakai oleh peneliti adalah Rapidminer versi 9.7.2 dimana dalam hal ini peneliti harus melakukan pengumpulan data dari media sosial twitter dengan menyambungkan *tools* RapidMiner ke dalam API Twitter. Dan kemudian melakukan pengolahan data seperti *filtering*, *labeling*, dan kemudian analisis untuk mengetahui tingkat akurasi dari masing-masing metode.

Dalam penelitian ini dapat diketahui bahwa metode KNN, Decision Tree, dan Naive Bayes dapat digunakan dalam bidang data mining.

Analisis sentimen terhadap data Twitter mengenai layanan BPJS mencapai tingkat akurasi 96.01%. Dimana *class precision* untuk *pred. negative* adalah 52.17%, *pred positive* adalah 0.00%, dan *pred. neutral* adalah 97.27% dalam metode KNN. Lalu tingkat akurasi mencapai 96.13%. Dimana *class precision* untuk *pred. negative* adalah 55.00%, *pred positive* adalah 0.00%, dan *pred. neutral* adalah 97.28% dalam metode Decision Tree. Dan yang terakhir mencapai akurasi 89.14%. Dimana *class precision* untuk *pred. negative* adalah 16.67%, *pred positive* adalah 1.64%, dan *pred. neutral* adalah 98.40% dengan metode Naive Bayes.

Dalam penelitian ini, dapat diketahui bahwa metode Decision Tree adalah metode yang tingkat akurasinya lebih tinggi dibandingkan kedua

metode lainnya dengan tingkat akurasi sebesar 96.13%.

5. Saran

Pada penelitian selanjutnya disarankan beberapa hal yaitu:

- Jumlah data yang digunakan diharapkan lebih banyak dari sebelumnya untuk meningkatkan pemberian informasi terhadap layanan BPJS dan agar dapat lebih meningkatkan keakuratan dari analisis yang dilakukan.
- Disarankan untuk menggunakan metode lainnya sehingga dapat mengetahui berbagai macam metode lainnya.

Daftar Pustaka

- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. *Journal of Information Technology*, 1(1), 32–41.
- Cahyaningrum, N. I., Fatima, D. W. Y., Kusuma, W. A., Ramadhani, S. A., Destanto, M. R., & Nooraeni, R. (2020). Analysis of User Sentiment of Twitter to Draft KUHP. *Jurnal Matematika, Statistika Dan Komputasi*, 16(3), 273. <https://doi.org/10.20956/jmsk.v16i3.8239>
- Cahyanti, D., Rahmayani, A., & Ainy, S. (2020). Analisis Performa Metode KNN Pada Dataset Pasien Pengidap Kanker Payudara. 1(2), 39–43.
- Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(3), 279. <https://doi.org/10.26418/jp.v5i3.34368>
- Linda, L., Haskas, Y., & Kadrianti, E. (2020). Perbedaan Persepsi Pengguna Jasa BPJS Dan Non BPJS (Umum) Tentang Kualitas Pelayanan Keperawatan Dirsud Timika-Papua. 15.
- Nurdin, A. D. (2017). Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe. *Astika, D., Studi, P., Informatika, T., & Malikussaleh, U., Vol. 6 No.*
- Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>
- Sarimuddin, S., Sari, J. Y., Mail, M., Masalu, M. A., Aristika, R. S., & Nurfagra, N. (2020). Klasifikasi Data Aging Tunggakan Nasabah Menggunakan Metode Decision Tree Pada ULAMM Unit Kolaka. *INFORMAL: Informatics Journal*, 5(1), 26. <https://doi.org/10.19184/isj.v5i1.16964>
- Sinaga, L. M., Sawaluddin, & Suwilo, S. (2020). Analysis of classification and Naive Bayes algorithm k-nearest neighbor in data mining. *IOP Conference Series: Materials Science and*

- Engineering*, 725(1).
<https://doi.org/10.1088/1757-899X/725/1/012106>
- Utami, I. (2020). *Analisis Sistem Informasi Banjir Berbasis Media Twitter*. 9(1), 67–72.
- Zhang, W., Chen, X., Liu, Y., & Xi, Q. (2020). A Distributed Storage and Computation k-Nearest Neighbor Algorithm Based Cloud-Edge Computing for Cyber-Physical-Social Systems. *IEEE Access*, 8, 50118–50130.
<https://doi.org/10.1109/ACCESS.2020.2974764>