**KLASIFIKASI TWITTER TENTANG PENDAPAT MASYARAKAT INDONESIA TERHADAP LAYANAN MASYARAKAT BPJS KESEHATAN DENGAN MENGGUNAKAN NAÏVE BAYES CLASSIFIACTION (NBC)**

**Revikhasah Alfian Kamal dan Muhammd Ridwan Pratama**

FakultasaIlmuaKomputer, UniversitasaDian Nuswantoro  
Jl. Imam Bonjol No. 205-207, Pendrikan Kidul, Semarang 50131, Indonesia  
e-mail: alfian.pbun@gmail.com, [muh.ridwan.777@gmail.com](mailto:muh.ridwan.777@gmail.com)

**ABSTRAK**

BPJS kesehatan adalah badan hukum publik yang di bentuk untuk menyelenggarakan asuransi perawatan kesehatan untuk semua orang Indonesia. BPJS kesehatan bekerjasama dengan fasilitas kesehatan milik pemerintah atau swasta untuk menyelenggarakan pelayanannya. Melalui program ini pemerintah berupaya agar masyarakat bisa mendapatkan pelayanan kesehatan dengan baik. Namun pada penerapannya masih belum sesuai dengan tujuan dan manfaat dari BPJS kesehatan sehingga menimbulkan opini masyarakat tentang pemanfaatan layanan BPJS kesehatan. Tujuan dari penelitian yang dilakukan adalah menganalisis opini publik terkait dengan pelayanan BPJS Kesehatan. Pada penelitian ini penulis mengklasifikasikan sebanyak 1500 tweet yang didapatkan dari crawling data mengginakan Twitter API. Anilisis dan pengolahan data dilakukan dengan menggunakan metode naive bayes classification.

***Kata Kunci****: klasifikasi, Information Retrievel, Naïve Bayessian Classification*

**I PENDAHULUAN**

1. **Latar Belakang**

Jaminan sosial merupakan hak setiap orang yang pemenuhannya dijamin oleh konstitusi dan peraturan perundang-undangan yang berlaku. Pelaksanaannya terutama dikaitkan dengan upaya pemerintah dalam pengentasan kemiskinan yang dilakukan secara bertahap sesuai dengan kemampuan negara, swasta, dan masyarakat dalam pembiayaannya. Salah satu metode pembiayaan jaminan sosial adalah dengan melibatkan peserta sendiri, melalui kewajiban pembayaran iuran yang disebut asuransi sosial. Jenis program jaminan yang paling tua dan penting adalah jaminan sosial kesehatan.

Tenaga kerja merupakan faktor strategis dalam upaya mewujudkan pembangunan nasional Indonesia. Peran negara dalam mewujudkan pembangunan nasional adalah dengan menjamin dan mewujudkan kesejahteraan tenaga kerja. Salah satu bentuk terwujudnya kesejahteraan tenaga kerja adalah terpenuhinya jaminan kesehatan. Pembangunan kesehatan merupakan bagian yang tak terpisahkandari pembangunan nasional sehingga harus mempunyai acuan yang jelas tentang arah pembangunan kesehatan yang dapat dipedomani oleh seluruh komponen pelaku pembangunan. (Dr. Ekowati Retnaningsih 2013)

Menurut data pemakaian fasilitas Kesehatan BPJS tahun 2016, faskes peserta JKN-KIS meningkat setiap tahunnya hal ini menunjukan bahwa masyarakat yang menggunakan layanan BPJS semakin bertambah. Penyelenggaraan asuransi kesehatan bertujuan untuk dimanfaatkan oleh peserta sebagai bagian dari upaya pemerintah untuk mencegah dan mengentaskan kemiskinan. Melalui program ini pemerintah berupaya untuk memelihara akses penduduk miskin terhadap pelayanan kesehatan yang masih rendah dancenderung menurun. Selain itu, dengan program ini diharapkan akan lebih tepat sasaran mengingat pada program sebelumnya pernah terjadi penyalahgunaan program kesehatan oleh masyarakat yang mampu. Pada penelitian yang pernah dilakukan sebelumnya ditemukan bahwa persepsi masyarakat terhadap pelayanan petugas kesehatan dirumah sakit perlu ditingkatkan. Karena pelayanan yang diberikan oleh petugas kesehatan di rumah sakit yang kurang baik maka menimbulkan banyak opini masyarakat tentang pelayanan bpjs kesehatan dirumah sakit. Analisis sentimen adalah riset komputasional dari opini, sentimen, dan emosi yang diekspresikan secara tekstual. Tujuan dari analisis sentimen adalah untuk mengelompokan polaritas dari teks atau pendapat yang ada dalam dokumen, apakah pendapat yang ditemukan bersifat positif, negatif. (Rudelvi 2018).

Penelitian ini bertujuan untuk menganalisis sentimen publik terkait dengan manfaat pelayanan Badan penyelenggara Jaminan Sosial Kesehatan pada selluruh masyarakat Indonesia metode yang digunakan untuk klasifikasi dokumen teks adalah model naïve bayes.

1. **Rumusan Masalah**

Berdasarkan pendahuluan diatas dapat disimpulkan rumusan masalah sebagai berikut:

1. Banyaknya pendapat dan komentar yang diuraikan oleh masyarakat membuat penelitian terhadap kinerja BPJS Kesehatan yang berupa tweet positive, negatif, dan keluhan.
2. Bagaimana cara kinerja algoritma Naïve Bayes Classifiaction (NBC) untuk mengklasifikasikan Twitter tentang opini masyarakat tentang BPJS Kesehatan di Indonesia?
3. **Batasan Masalah**

Ada beberapa batasan masalah yang dapat ditangkap pada penelitian kali ini, sebagai berikut :

1. Crawling data dengan menggunakan Twitter API dengan bahasa Indonesia.
2. Menentukan kata dari kamus sendiri untuk melihat cuitan positif, negatif, dan keluhan berdasarkan analisis setiap postingan.
3. Pemodelan terhadap data yang telah diambil menggunakan metode Naive Bayes Classifiaction(NBC) dan data yang diambil selama tahun 2019.
4. **Tujuan Masalah**

Beberapa tujuan masalah sebagai berikut:

1. Menggunakan text mining yang bekerja untuk membedakan mana komentar/cuitan yang berupa positif, negatif ataupun keluhan bertujuan untuk kedepan sebagai penilaian terhadap pelayanan masyarakat BPJS Kesehatan.
2. Penerapan algoritma Naïve Bayes Classification dalam melakukan penilaian terhadap opini masyarakat tentang pelayanan masyarakat BPJS Kesehatan.
3. **Manfaat**

Ada 3 target yang terkena dalam manfaat penelitian ini sebagai berikut :

1. Bagi Peneliti

* Dapat mengerti bagaimana cara mengklasifikasikan sebuah data.
* Dapat mengetahui lebih dominan mana pendapat masyarakat terhadap jasa pelayanan masyarakat BPJS Kesehatan.

1. Bagi BPJS

* Mengetahui opini masyarakat apakah kebanyakan opini tersebut positif, negatif atau keluhan terhadap pelayanan yang diberikan.
* Dapat mengevaluasi apa yang harus diperbaiki dalam kinerja BPJS terhadap masyarakat.

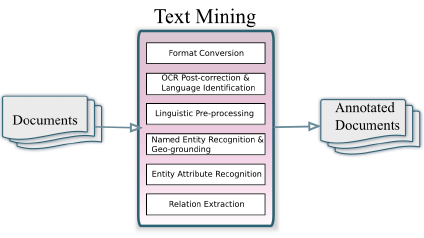
1. Bagi Masyarakat

* Mendapatkan pelayanan dan kinerja yang lebih baik dari BPJS

**II LANDASAN TEORI**

1. **Tinjauan Pustaka**
2. Text Mining

Text mining adalah proses ekstraksi pola berupa informasi dan pengetahuan yang sebelumnya tidak diketahui pada sejumlah besar sumber data yang berupa teks (chiwara, dkk, 2016). Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Proses yang umum dilakukan oleh penambangan teks di antaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks, dll (Turban, dkk, 2015)



1. Preprocessing data

Dokumen teks dari hasil pencarian yang telah dikumpulkan kemudian dilakukan pemrosesan data. Pada tahapan ini dilakukan beberapa hal, yaitu tokenized, transform cases, filter tokens, filter stopword dan Stemming.

1. Tokenized memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata. Tokenisasi menghilangkan delimiter seperti tanda titik (.), koma (,), spasi, dan karakter angka yang ada pada kata tersebut**.**
2. Transform cases Merupakan proses untuk merubah bentuk kata-kata, pada proses ini karakter dijadikan menjadi huruf kecil atau lower case semua
3. Filtering Proses menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi klasifikasi teks. Pada proses ini kata yang termasuk adalah seperti penunjuk waktu, kata tanya dan kata sambung.
4. Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasar. Tujuan dari proses stemming adalah menghilangkan imbuhan imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata.
5. Naive Bayes

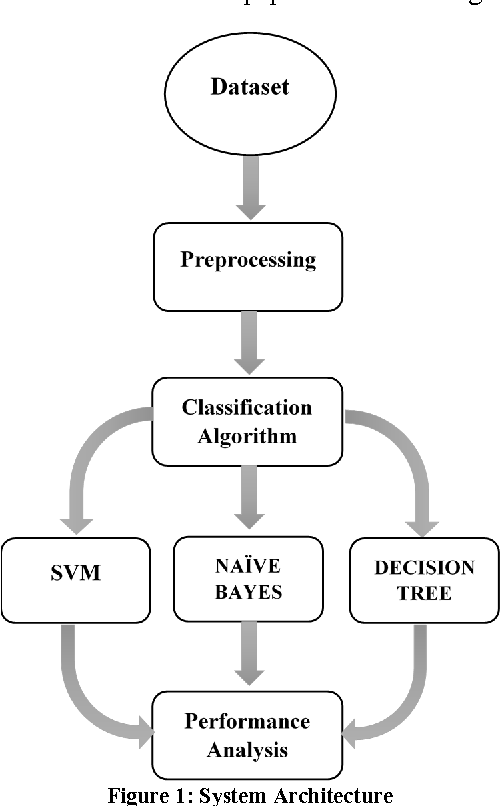
Twitter API (Application Programming Interface) merupakan sejumlah fungsi yang dapat digunakan pengembang perangkat lunak untuk mengolah data saat membangun perangkat lunak. Twitter API menyediakan beberapa fungsi untuk melakukan suatu tugas tertentu, sehingga pengembang perangkat lunak hanya memanggil fungsi tersebut di dalam perangkat lunak yang dibangun. Twitter API menggunakan arsitektur REST (Representational State Transfer) sehingga Twitter API dapat digunakan pada format data yang beragam seperti XML maupun JSON.

1. Xampp

Xampp XAMPP adalah sebuah software web server apache yang didalamnya sudah tersedia database server MySQL dan dapat mendukung pemrograman PHP.

1. Naïve Bayes

Dalam penelitian ini yang menjadi data uji adalah dokumen opini. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya atau disebut data training. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya atau disebut data testing. Dalam algoritma Naive Bayes Classifier setiap dokumen direpresentasikan dengan pasangan atribut “x1, x2, x3, ...xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya.



Ada perumusan untuk naive bayes yakni sebagai berikut :



Dimana P(H|X) adalah probabilitas posterior H berdasarkan syarat X yang diperoleh dari probabilitas hipotesis H (P(H)) dikali dengan probabilitas X berdasarkan kondisi hipotesis H (P(X|H)) dan dibagi dengan probabilitas X (P(X)).

1. **Kerangka Pemikiran**

Penelitian ini menggunakan text mining yang bekerja untuk membedakan mana komentar/cuitan yang berupa positif, negatif ataupun keluhan bertujuan untuk kedepan sebagai penilaian terhadap pelayanan masyarakat BPJS Kesehatan.

**III METODE PENELITIAN**

**3.1 Intrumen Penelitian**

Beberapa ini merupakan komponen yang dibutuhkan untuk melakukan penelitian ini sebagai berikut :

1. **Kebutuhan Perangkat Lunak (Software)**

Perangkat lunak yang digunakan untuk membangun system ini antara lain :

1. Sistem Operasi : W indows 10
2. Bahasa pemrograman PHP untuk crawling Twitter API
3. Google Chrome untuk search engine
4. Text editor Visual Studio Code
5. Server WEB menggunakan XAMPP
6. CorelDRAW 2019 untuk editing gambar
7. **Kebutuhan Perangkat Keras (Hardware)**

Perangkat keras yang digunakan untuk membuat system ini antara lain :

1. Prosessor intel core i5-8250U (1.6 GHz – 3,4 GHz)
2. Grafis AMD Radeon 520
3. RAM sebesar 4 GB DDR4

**3.2 Prosedure Pengambilan dan Pengumpulkan Data**

Pengambilan data dalam penelitian ini yaitu data dari Tweet dan melalui Twitter API yang akan di implementasikan ke dalam program berbahasa pemrograman PHP menggunakan text editor Visual Studio Code dan untuk mengambil data melalui Teknik crawling. Data kunci yang akan dijadikan patokan untuk menentukan dalam kategori positif, negatif atau netral mengacu pada kata-kata yang berada di kbbi.web.id.

**3.3 Teknik Analisis Data**

Dalam penelitian ini diperlukan Analisa data dengan langkah-langkah sebagai berikut:

1. Mengcrawling data postingan yang ada di BPJS.
2. Datanya akan dibagi menjadi 2 yaitu 80% data training dan 20% data testing.
3. Data training memberikan label pada setiap tweet dibedakan menjadi 2 kategori ialah data positif merupakan tweet yang tidak ada ujaran kebencian, negatif, dan keluhan merupakan data yang ada unsur kebencian.
4. Data testing akan melalui proses preprocessing untuk mendapatkan data yang akan diklasifikasikan.
5. Melakukan proses klasifikasi dengan menggunakan algoritma Naïve Bayes Cllassifier agar mendapatkan klasifikasi berupa positif, negatif, dan keluhan.
   1. **Model / Metode yang Diusulkan**

Penelitian ini akan menggunakan metode machine learning, dengan teknik pembelajaran terbimbing (supervised learning). Metode yang digunakan untuk mengolah data text yang sudah terkumpul adalah Naive Bayes Classifier. Pada penelitian proses utama dibagi ke dalam 2 proses, proses yang pertama adalah proses pelatihan dan proses yang kedua merupakan proses pengujian.

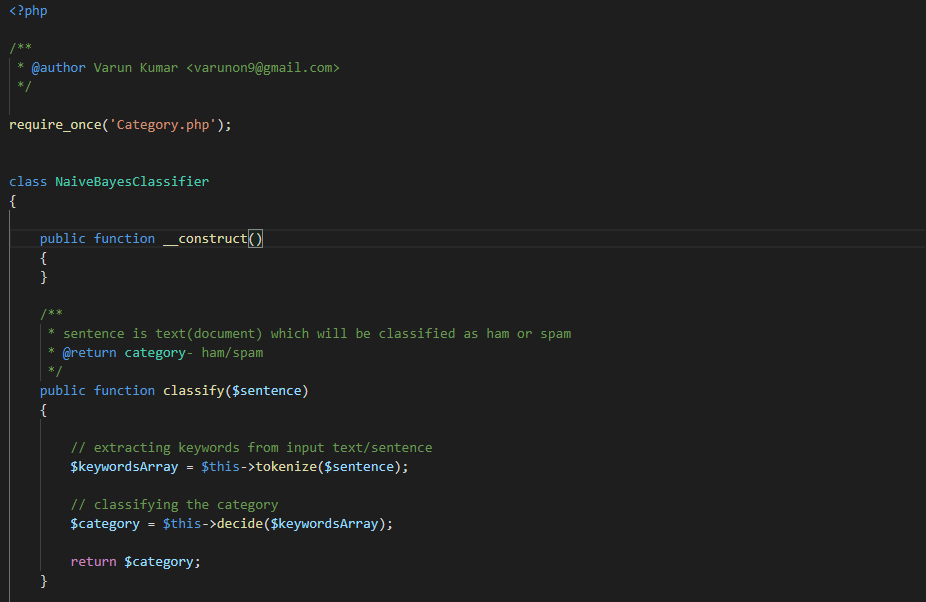
**IV. METODE PENGUMPULAN DATA**

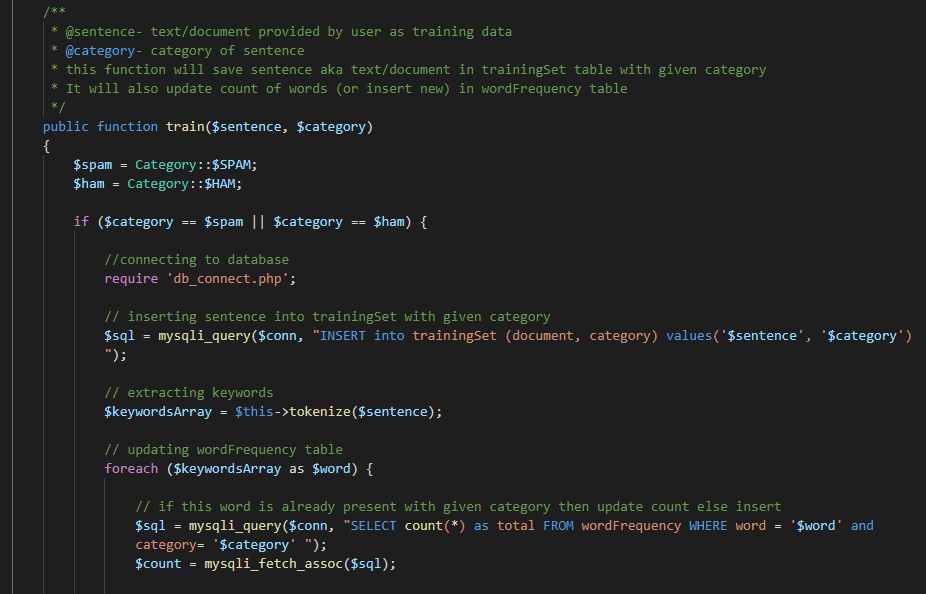
Untuk proses pengambilan data tweet (crawling) menggunakan python, langkah pertama yaitu install python di laptop, selanjutnya mendaftar twitter api. Setelah semua terinstal buka Command Prompt, dengan menggunakan bahasa pemrograman pyhton, untuk melakukan crawling topik apa yang akan diambil dengan memasukkan keyword “bpjs, bpjskesehatan”. Dengan hasil pencarian awal mencapai 100 data kemudian dilakukan normalisasi data secara manual dan setelah itu dilakukan pelabelan untuk 2 kelas yaitu negatif dan positif.

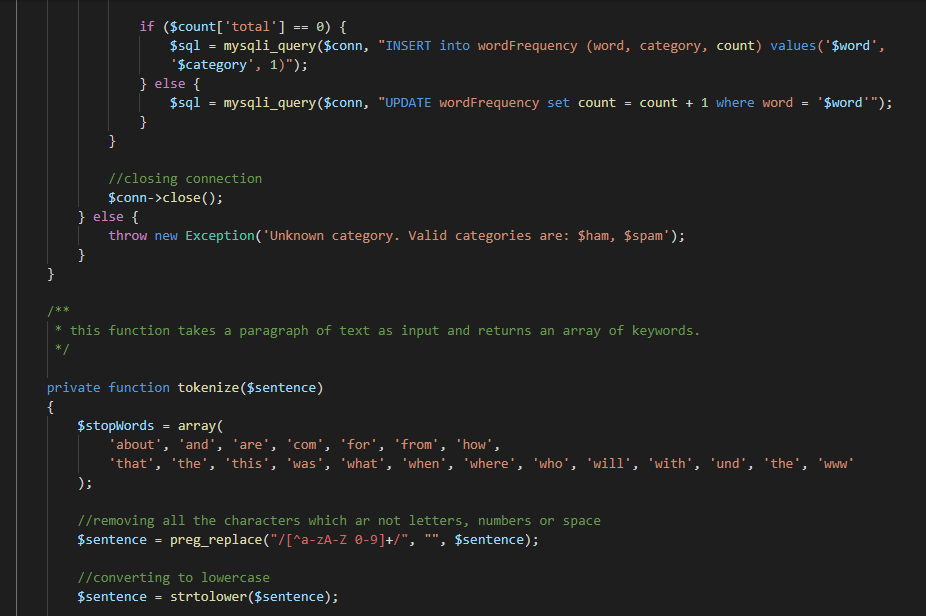
**4.1 Hasil Preprocessing**

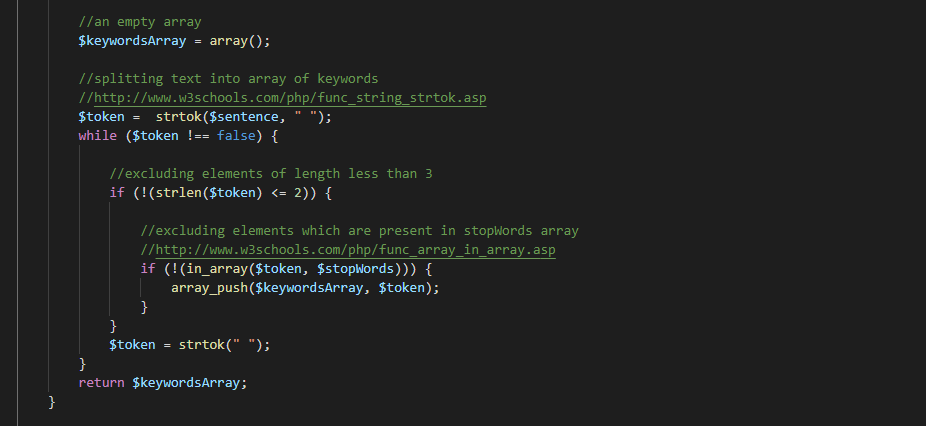
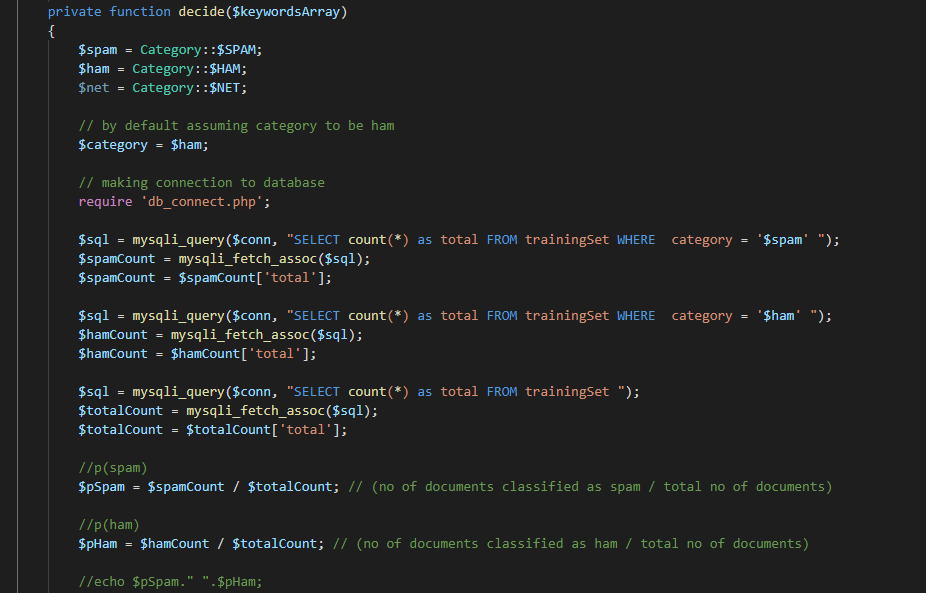
Proses pre processing pada penelitian ini diawali dengan proses data cleaning, case volding, tokenizing, stopword removal dan stemmin*g.* Pada proses ini atribut dari data dapat dikurangkan sehingga akan lebih mempermudah dalam proses komputasi.

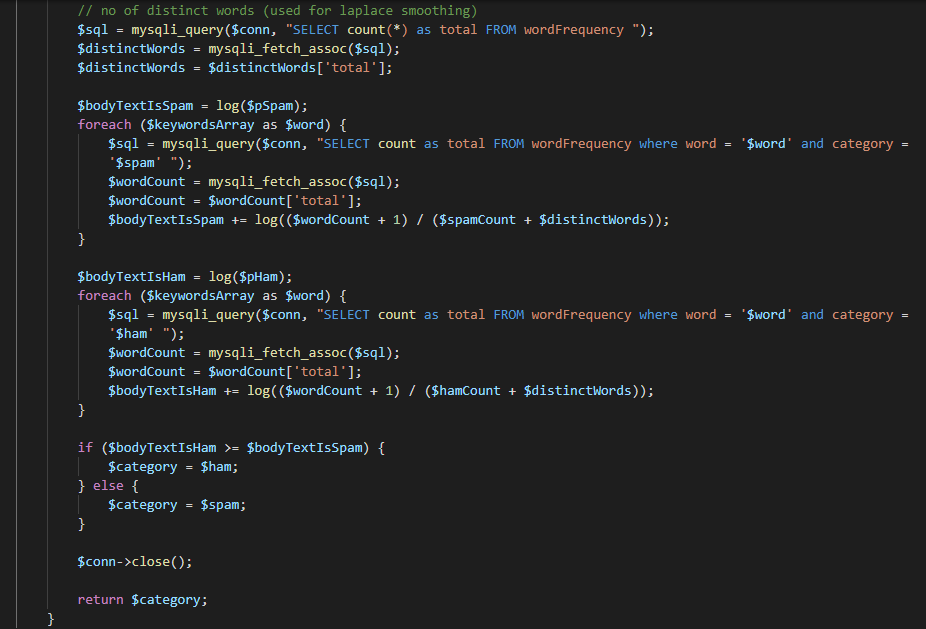
**4.2. Implementasi Algoritma Naive Bayes**





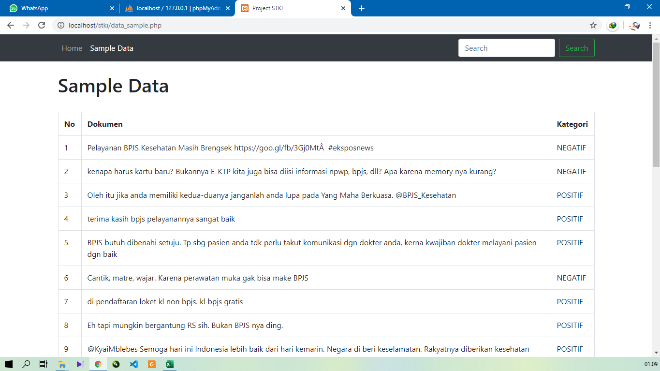


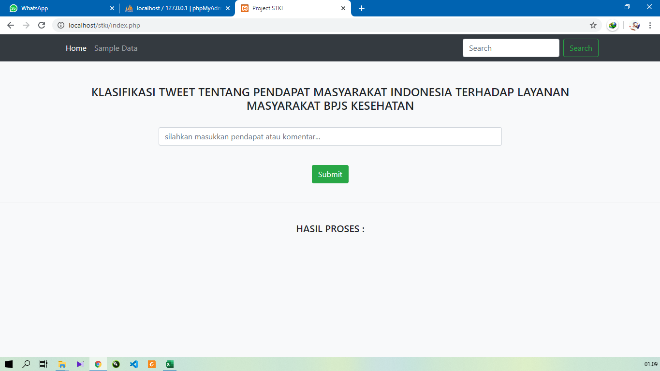




**4.2. Interface Program**

Halaman utama dari program sistem klasifikasi tweet tentang kinerja layanan masyarakat BPJS Kesehatan. menampilkan data yang berupa data tweet. Data tweet tersebut disajikan dalam bentuk tabel. Dibawah tabel data tweet akan terdapat sebuah kolom untuk memasukkan dokumen jika ingin mengklasifikasikan data yang baru. Saat pengguna menekan masukan data, maka algoritma NBC akan bekerja mengklasifikasikan dokumen tersebut.





**4.4. Pengujian**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | True Class | |
|  | Positif | Negatif |
| Predicted Class | Positif | 9(True Positif) | 2(False Positif) |
| False | 2 (False Negatif) | 7(True Negatif) |

Accuracy =

=

=

= 80%

**V. KESIMPULAN DAN SARAN**

**5.1. Kesimpulan**

Berdasarkan pada penelitian ini, data berasal dari keyword **“bpjs, bpjskeseehatan”** Dari 100 *tweet* dijadikan sebagai data *training* dan 20 % dari data training akan menjadi data testing. Proses dimulai dengan *cleaning, case volding, tokenizing, stopword removal dan stemming*. Proses selanjutnya merupakan proses pelabelan manual dengan memberi label negatif dan positif. Hingga mendapatkan tingkat akurasi sebesar 85%.

**5.1. Saran**

Saran yang dapat diberikan guna perkembangan klasifikasi pendapat masyarakat mengenai kinerja layanan masyarakat BPJS Kesehatan adalah memperbanyak kamus selain menggunakan libraby sastrawi. Karena tingkat keakuratan bergantung pada jumlah kamus.

**DAFTAR PUSTAKA**

1. Andrianti, S. D. (2018). *SENTIMENT ANALYSIS ON TWITTER ACCOUNT USING NAIVE BAYES CLASSIFIER ALGORITHM Case Study : Indonesia Healthcare and Social Security Agency ( BPJS Kesehatan )*. *2*(2), 23–28.
2. Ipmawati, J., Kusrini, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, *6*(1), 28–36.
3. Nurulbaiti, F., & Subekti, R. (2018). Analisis Sentimen Terhadap Data Tweet Untuk Badan Penyelenggara Jaminan Sosial ( Bpjs ) Menggunakan Program R. *Jurnal Pendidikan Matematika Dan Sains UNY*, 1–9.
4. Yanis, R. Y. (2018). Sentiment Analysis of Bpjs Kesehatan Services To Smk Eklesia and Bina Insani Jailolo Teachers. *Jurnal Terapan Teknologi Informasi*, *2*(2), 25–34. https://doi.org/10.21460/jutei.2018.22.105
5. Zakkiyah, T. H. (2016). *Perbandingan 3 Metode Data Mining Dalam Pencarian Pengetahuan Kualitas Pelayanan Kesehatan Bagi Pasien BPJS*.
6. Aditya, R. B. (2015) Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining
7. Fauzi, A. M. (2018) Klasifikasi Berita pada Twitter Menggunakan Metode Naive Bayes dan Query Expansion Hipernim-Hiponim http://j-ptiik.ub.ac.id
8. Rustiana, D. (2017) ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES