

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Tinjauan pertama penelitian yang dilakukan oleh Septian Nersa Putra (2018). Pada penelitian ini, sistem akan melakukan klasifikasi terhadap berita yang diupload pada akun twitter Divis Humas Polri ke dalam tiga kategori yaitu berita kegiatan polisi, komentar masyarakat dan layanan masyarakat selama empat tahun ke belakang. Setelah berita diklasifikasi kemudian dicari sentimen dari setiap topik. Metode yang di gunakan adalah *Naïve Bayes Classifier*.

Tinjauan yang ke-dua ditulis oleh Yuna Sophia Dewi Febriant pada tahun (2017). Objek yang diteliti adalah *tweets* yang ada di akun twitter STMIK Akakom yang kemudian diklasifikasikan dalam *tweets* positif atau negatif dengan metode *Naive Bayes*.

Tinjauan ke-tiga ditulis oleh Dimas Bagus C. W. (2017). Peneliti mengembangkan aplikasi dengan metode metode *text mining* untuk melakukan analisis sentimen untuk menemukan topik topik yang terdapat pada data tweets dan melihat keterkaitan antar kata pada khusus pilkada DKI Putaran 2.

Tinjauan ke-empat oleh Widya Sulistyani (2020). Pada penelitian ini, peneliti membangun sistem yang dapat menemukan topik tersembunyi di dalam

tweets menggunakan pemodelan topik Latent Semantic Analysis, study kasus Gubenur Jawa Tengah.

Tinjauan ke-lima oleh Ghulam Asrofi Buntoro (2017). Penelitian ini, data yang diambil hanya tweet dalam bahasa Indonesia, yaitu 100 tweet dengan kata kunci AHY, 100 tweet dengan kata kunci Ahok, dan 100 tweet dengan kata kunci Anies. Data diambil secara acak baik dari user biasa ataupun media online di Twitter dan dianalisis sentimen menggunakan *Naïve Bayes* dan *Support Vector Machine*.

Pada penelitian ini akan menganalisis sentimen negatif,positiv dan netral dari *tweets* nasabah pada akun twitter BNICustomerCare pada pelayanan Internet Banking menggunakan metode *naïve bayes*. Dan *tweets* yang digunakan untuk analisis adalah data *tweets* tahun 2020

Perbandingan dari penelitian – penelitian di atas dengan yang akan dibuat dapat dilihat pada tabel 2.1

Tabel 2. 1 Tinjauan Pustaka

Penulis	Objek	Metode	Hasil
Septian Narsa Putra(2018)	Berita dan twet Divisi Humas Polri	<i>Naive Bayes Classifier</i>	Mengklasifikasikan berita berdasarkan topik kemudian dicari sentimen dari setiap topik.

Yuna Sophia Dewi Febriant(2017)	<i>Tweets</i> pada akun Twitter STMIK Akakom	<i>Naive Bayes Classifier</i>	Menganalisa dan mengklasifikasikan sentimen yang ada pada twitter STMIK Akakom.
Dimas Bagus C. W. (2017)	<i>Tweets</i> terkait pilkada DKI putaran ke-2	Supervised Learning : Naive Bayes dan Unsupervised Learning : Association Rule, K Means & Topic Modeling	Membangun aplikasi dengan menggunakan metode-metode <i>text mining</i> untuk melakukan analisis sentiment dan menemukan topik topik yang terdapat pada data tweets.
Widya Sulistyani (2020)	Data <i>tweets</i> Gubernur Jawa Tengah.	Latent Semantic Analysis (LSA)	Sistem yang dapat menemukan topik tersembunyi di dalam tweets menggunakan pemodelan topik Latent Semantic Analysis.
Ghulam Asrofi Buntoro(2017)	Data <i>Tweets</i> pemilihan calon gubernur DKI 2017	<i>Naive Bayes Classifier dan Support Vector Machine</i>	Menganalisis sentimen yang berkaitan dengan <i>tweets</i> calon Gubernur DKI 2017 (AHY,Ahok dan Anies)
Awaluddin(2020)	<i>Tweets</i> Twitter BNI Cutomer Care	<i>Naive Bayes Classifier</i>	Mencari sentiment dari setiap <i>tweets</i> pada akun twitter BNI Customer Care.

2.2 Dasar Teori

2.2.1 Twitter

Twitter merupakan sebuah sosial media atau layanan jejaring sosial dan mikroblogging yang di gunakan untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan *tweets*. Twitter berdiri pada bulan maret 2006 oleh Jack Dorsey dan diluncurkan pada bulan juli. Pengguna yang ingin menggunakan twitter harus melakukan pendaftaran (*sig-in*) terlebih dahulu untuk mendapatkan akaun twitter supaya bisa menulis *tweets* dan sebagainya. Sedangkan pengguna yang tidak terdaftar hanya bisa membaca *tweets* saja.

Pengguna bisa menulis pesan berdasarkan topik dengan tanda #(tagar). Sedangkan untuk menyebut atau membalas pesan dari pengguna lain bisa menggunakan tanda @(diikuti nama akun yang akan dibalas).

Twint merupakan tools scrapping untuk twitter yang dibuat secara khusus dalam bahasa python, tools ini dapat digunakan tanpa API dari Twitter dengan keterbatasan hanya dapat melakukan scrapping data 3200 twit. Tools ini juga dapat melakukan scrapping user, followers, retweet dsb.

2.2.2 *Text mining*

Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Masukan untuk penambangan teks adalah data yang tidak (atau kurang) terstruktur, sedangkan masukan untuk penambangan data adalah data yang terstruktur (Ronen Feldman, 2006). *Text mining* dapat diterapkan untuk pelacakan topik, kategorisasi dan analisis sentimen. Contoh teks mining pada bidang marketing untuk meningkatkan *cross-selling* dan *up-selling* dengan menganalisa data tak-terstruktur yang dihasilkan oleh *call centers*. Text yang dihasilkan oleh catatan di call-centers dan juga transkrip percakapannya dengan pelanggan dianalisa dengan menggunakan algoritma-algoritma *text mining* untuk mengekstrak informasi yang baru dan bisa ditindaklanjuti tentang persepsi pelanggan terhadap produk dan layanan pelanggan. Sehingga perusahaan tersebut bisa mengambil keputusan untuk strategi marketing yang akan diambil kedepannya.

Dalam proses penambangan teks terdapat proses perubahan data menjadi data yang terstruktur sesuai dengan kebutuhan, proses ini disebut *Text Preprocessing*, dalam proses ini terdapat beberapa tahap diantaranya sebagai berikut :

1. *Case Folding* adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima dan karakter selain huruf dihilangkan.
2. Tahapan *Tokenizing* adalah proses pemotongan kalimat (*string*) berdasarkan kata penyusunnya. Kemudian tahap *filtering*, merupakan tahapan mengambil kata - kata penting dari hasil token dan membuang sebagian kata tertentu.
3. Tahapan *Stemming* merupakan tahapan pemetaan dan penuraian suatu kata menjadi kata dasarnya.
4. Tahapan *Tagging* merupakan tahap memberikan label.

2.2.3 Naïve Bayes Classifier

Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Bayes merupakan teknik prediksi berbasis *probabilistic* sederhana yang berdasar pada penerapan teorema bayes dengan asumsi independensi yang kuat (Eko Prastyo, 2012). *Naive Bayes* merupakan salah satu contoh metode *supervised document classification* yang berarti membutuhkan data latih dalam melakukan klasifikasi.

Dalam algoritma Naïve Bayes Classifier setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kalimat pertama, x_2 adalah kalimat kedua dan seterusnya. Sedangkan V adalah himpunan kalimat. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua dokumen yang diujikan (*VMAP*), dimana persamaannya adalah sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n | V_j)} \dots \dots \dots (1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kalimat (V_j) sehingga persamaan dapat ditulis sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \dots \dots \dots (2)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \dots \dots \dots (3)$$

Keterangan :

V_j = Document kalimat

J = 1,2,3...n. Dimana penelitian ini

j_1 = sentimen negatif,

j_2 = sentimen positif,

j_3 = sentimen netral.

$P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs_j|}{|contoh|} \times 100\% \dots \dots \dots (4)$$

Keterangan :

|docs j| = dokumen uji/train_set

|contoh| = jumlah semua kalimat document

Pada pengklasifikasian menggunakan Naïve Bayes dibagi kedalam 2 proses, yaitu proses *training* dan *testing*. Proses *training* digunakan untuk menghasilkan model analisis yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan data mentah yang baru.

2.2.4 Tools

Jupyter Notebook merupakan tool yang populer untuk mengolah data di python. Jupyter Notebook memungkinkan untuk mengintegrasikan antara kode dengan output di dalam satu dokumen secara interaktif.

Python adalah bahasa pemrograman interpreter/penerjemah multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas Pustaka standart yang besar serta komprehensif. Salah satu fitur yang tersedia pada python adalah sebagai bahasa pemrograman yang dinamis yang dilengkapi dengan manajemen memori otomatis.

Visual Studio Code adalah perangkat lunak penyunting kode-sumber buatan Microsoft untuk Linux, macOS, dan Windows. Visual Studio Code menyediakan fitur seperti penyorotan sintaksis, penyelesaian kode, kutipan kode, merefaktor kode, pengawakutuan, dan Git.

Framework Flask adalah kerangka kerja aplikasi web bersifat kerangka kerja mikro yang ditulis dalam bahasa pemrograman Python dan menggunakan dependensi Werkzeug dan Jinja2. Flask disebut kerangka kerja mikro karena tidak membutuhkan alat-alat tertentu atau pustaka. Flask mendukung ekstensi yang dapat menambahkan fitur aplikasi seolah-olah mereka diimplementasikan dalam Flask itu sendiri. Ekstensi yang ada seperti pemetaan objek-relasional, validasi form, penanganan unggahan, berbagai teknologi otentikasi terbuka, lapisan abstraksi basisdata, validasi form, atau komponen lain.