

# KLASIFIKASI KEPERIBADIAN BERDASARKAN POSTINGAN TWITTER DENGAN ALGORITME NAÏVE BAYES STUDI KASUS : CV. WILIS ELEKTRONIK

Adi Setiawan<sup>1)</sup>, Harunur Rosyid<sup>2)</sup>

<sup>1)</sup> CV. Wilis Elektronik

<sup>2)</sup> Teknik Informatika Universitas Muhammadiyah Gresik

Jl. Sumatra 101 Gresik Kota Baru (GKB), Randuagung, 61121 Telp.(031) 3951414 , Faks. (0561) 740186  
e-mail: adisett57@gmail.com<sup>1)</sup>

(Naskah masuk : 28 Juli 2022 Diterima untuk diterbitkan : 30 November 2022)

## ABSTRAK

*Interaksi seseorang terhadap orang lain dan lingkungan sekitarnya dapat dipengaruhi oleh kepribadian orang tersebut. Sehingga dalam dunia kerja, kepribadian adalah salah satu faktor pendukung keberhasilan kinerja. Salah satu inventori kepribadian adalah Big Five Personality yang di kemukakan oleh Lewis Golberg pada tahun 1992. Goldberg, merumuskan pengelompokan ciri-ciri sifat seseorang yang membedakannya dengan orang lain menjadi 5 himpunan kepribadian. yaitu Extraversion, Agreeableness, Conscientiousness, Neuroticism, dan Openness. Kepribadian seseorang dapat diklasifikasi berdasarkan teks yang diunggahnya di media sosial. Penentuan kata-kata yang digunakan berulang-ulang dapat menggambarkan kepribadian orang tersebut. Metode Naive Bayes dipilih dalam penelitian ini untuk mengklasifikasi kepribadian pegawai di CV. Wilis Elektronik. Adapun data yang dipakai adalah data masukan yang berasal dari hasil crawling akun twitter pegawai, data talih berasal dari Open Psycometric Project berupa 1000 jawaban kuesioner yang dikumpulkan secara onilne, dan data keluaran yang merupakan hasil dari proses klasifikasi. Serta dilakukan preprocessing dan pembobotan menggunakan algoritme TF-IDF pada data masukan dan data latih. Dari hasil analisis dan implementasi diperoleh kesimpulan bahwa sistem dapat melakukan prediksi atau klasifikasi kepribadian. Secara umum didapatkan hasil yang sesuai seperti yang diharapkan, dari proses crawling data sampai proses klasifikasi. Sistem Klasifikasi Kepribadian Berdasarkan Postingan Twitter ini dapat menyederhanakan tes kepribadian yang ada di CV. Wilis Elektronik. Sehingga HRD tidak perlu lagi mengadakan sesi wawancara dan deskripsi diri untuk mengetahui kepribadian pegawainya.*

**Kata Kunci:** Klasifikasi, Big Five Personality, Naïve Bayes, Twitter.

## ABSTRACT

*A person's interactions with other people and the surrounding environment can be influenced by that person's personality. So that in the world of work, personality is one of the factors supporting the success of performance. One of the personality inventories is the Big Five Personality which was put forward by Lewis Golberg in 1992. Goldberg, formulated the grouping of a person's characteristics that distinguish him from other people into 5 personality sets. namely Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. A person's personality can be classified based on the text he uploads on social media. Determination of words used repeatedly can describe the person's personality. The Naive Bayes method was chosen in this study to classify the personality of employees at CV. Electronics Wilis. The data used are input data derived from the results of crawling employee Twitter accounts, data from the Open Psychometric Project in the form of 1000 questionnaire answers collected online, and output data which is the result of the classification process. As well as preprocessing and weighting using the TF-IDF algorithm on the input data and training data. From the analysis and implementation results, it can be concluded that the system can predict or classify personality. In general, the results are as expected, from the data crawling process to the classification process. The Personality Classification System Based on Twitter Posts can simplify the personality tests in the CV. Electronics Wilis. So that HRD no longer needs to hold interview sessions and self-descriptions to find out the personality of its employees.*

**Keywords:** Classification, Naïve Bayes, Big Five Personality, Twitter.

## I. PENDAHULUAN

Interaksi seseorang terhadap orang lain dan lingkungan sekitarnya dapat dipengaruhi oleh kepribadian, sehingga pada dunia kerja kepribadian merupakan salah satu faktor pendukung keberhasilan kinerja. Salah satu inventori kepribadian adalah *Big Five Personality* yang dikemukakan oleh Lewis Goldberg pada tahun 1992. Goldberg, merumuskan pengelompokan ciri-ciri sifat seseorang yang membedakannya dengan orang lain menjadi 5 himpunan kepribadian. Yaitu Extraversion, Agreeableness, Conscientiousness, Neuroticism, dan Openness. Inventori kepribadian Big Five Personality telah digunakan dalam penelitian sebelumnya di Indonesia untuk mengukur keterkaitan antara dimensi kepribadian dengan performa kerja, komitmen organisasional, keberhasilan kewirausahaan dan perilaku konsumen (Mar'i, F., dkk, 2018).

CV. Wilis Elektronik adalah salah satu badan usaha yang menawarkan jasa perbaikan peralatan elektronik rumah tangga seperti televisi, sound system, kipas angin, kulkas, rice cooker, setrika listrik dsb. yang ada di Jl. Pahlawan Mubin Wonokoyo, Kecamatan Menganti, Kabupaten Gresik. Sebagai perusahaan yang bergerak dibidang jasa, CV. Wilis Elektronik mengharuskan seluruh pegawainya terutama bagian customer service untuk bersikap ramah ketika menyambut pelanggan. Namun pada kenyataannya masih banyak pegawai yang kurang ramah ketika melayani pelanggan. Hal tersebut dibuktikan dari banyaknya respon pelanggan yang mengeluhkan sikap pegawai yang kurang ramah melalui kotak saran yang disediakan. Kepribadian yang tidak cocok dengan jenis pekerjaan merupakan salah satu faktor yang menyebabkan hal tersebut.

Selama ini, untuk mengetahui kepribadian pegawainya, CV. Wilis Elektronik harus mengadakan tes kepribadian seperti wawancara dan deskripsi diri. Tes kepribadian tersebut dilakukan dengan cara mewawancarai para pegawai dan meminta para pegawai untuk mendeskripsikan diri mereka melalui selembar kertas yang sudah disediakan. CV. Wilis Elektronik masih belum menggunakan sebuah sistem atau aplikasi sebagai media untuk memprediksi kepribadian pegawainya. Hal ini dapat dilihat dari pelaksanaan tes kepribadian yang masih menggunakan media kertas dan wawancara dengan para pegawai. Hal ini membuat para pegawai harus datang ke lokasi tes, Serta menuliskan tentang deskripsi diri mereka. Selain itu tim penyelenggara tes juga harus memeriksa hasil tes tersebut satu persatu.

Pengembangan sistem dapat menjadi solusi dari proses penentuan kepribadian pegawai di CV. Wilis Elektronik. . Kepribadian seseorang dapat diklasifikasi berdasarkan teks yang diunggahnya di media sosial. Penentuan kata-kata yang digunakan berulang-ulang dapat menggambarkan kepribadian orang tersebut (Alvini Fikriani, 2019). Dari banyaknya platform media sosial yang beredar, Twitter adalah salah satu media sosial yang cukup populer, dalam waktu singkat data yang dihasilkan oleh Twitter terus bertumbuh dan semakin banyak. Oleh karena itu, CV. Wilis Elektronik membutuhkan pengembangan sistem yang terintegrasi dengan media sosial twitter untuk mendapatkan informasi dari para pegawai. Aplikasi ini nantinya akan mengambil beberapa postingan berupa teks dari akun twitter pegawai tersebut. Adapun tweet tersebut akan dilakukan preprocessing menggunakan NLTK dan TF-IDF (Indri Nurandini, 2019). Setelah itu hasil dari preprocessing akan dilakukan klasifikasi teks untuk menentukan kepribadian pemilik tweet kedalam suatu kelas atau characteristics primary yang

dibuat oleh open psychometrics project (Hanif Akhtar, 2018) menggunakan metode supervised learning Naive Bayes dan SVM (support vector machine).

Penggunaan kedua algoritme yaitu Naïve Bayes dan SVM akan menunjukkan manakah algoritme yang terbaik untuk melakukan klasifikasi kepribadian berdasarkan postingan di twitter. Pemilihan kedua algoritme tersebut didasarkan pada keberhasilan tingkat akurasi terbaik jika dibandingkan dengan algoritme klasifikasi lainnya dalam mengklasifikasikan dokumen text (Meilana, 2021). Keunggulan algoritme Naïve Bayes adalah pengklasifikasiannya dapat dilakukan secara independen pada masing-masing fitur objek yang diklasifikasikan (Kepin Sihotang, 2019). Selain itu keunggulan algoritme SVM adalah mempunyai konsep yang matang dan jelas secara matematis serta mampu menangani masalah klasifikasi serta regresi secara linear atau non linear (Kristian Indradiarta Gunawan, 2021). Sehingga pada penelitian ini akan digunakan kedua algoritme tersebut untuk klasifikasi kepribadian pegawai pada CV. Wilis elektronik. Selain itu, akan dibuktikan juga manakah dari kedua algoritme tersebut yang memiliki performa terbaik.

## II. LANDASAN TEORI

### 2.1 Big Five

Big Five merupakan klasifikasi ilmiah dari sifat kepribadian, sistem yang memetakan sifat kepribadian yang sesuai berdasarkan penilaian dalam faktor lima besar. Sifat-sifat kepribadian yang digolongkan menjadi lima faktor berasal dari studi statistik tentang tanggapan terhadap item kepribadian. Lima faktor dari sifat-sifat kepribadian itu adalah Extraversion, Agreeableness, Conscientiousness, Emotional Stability dan Intellect. Extraversion mencakup sikap mudah bergaul, banyak bicara, tegas dan aktif. Agreeableness mencakup sikap kooperatif, fleksibel, toleran dan pemaaf. Conscientiousness mencakup sifat dapat diandalkan, berjuang untuk meraih prestasi, pekerja keras, tekun dan tertib. Emotional Stability mencakup sifat tenang, percaya diri, dan ulet. Intellect mencakup rasa ingin tahu, berpikiran luas, cerdas dan berbudaya (Rizka Shinta Wulandari, 2021). Lima faktor Big Five mempengaruhi performa pekerjaan yang spesifik berdasarkan studi yang menunjukkan bahwa aspek kepribadian merupakan prediktor yang valid dalam mengukur performa akademik atau pekerjaan [Purba Daru Kusuma, 2021]. Keberhasilan karier seorang individu dapat dikaitkan dengan Extraversion, Emotional Stability, dan Agreeableness. Seorang individu yang mempunyai Extraversion tinggi, Emotional Stability tinggi, dan Agreeableness yang rendah mengalami tingkat kepuasan karier yang lebih tinggi dibandingkan mereka yang mempunyai Extraversion rendah, Emotional Stability yang rendah serta Agreeableness yang tinggi. Extraversion mempunyai hubungan yang positif dalam kesuksesan karier, seperti hubungan dengan gaji, promosi, dan kepuasan karier (Casi Setianingsih, 2021).

### 2.2 Data Mining

Menurut (Y. Mardi, 2019) data mining adalah proses menemukan informasi menarik yang ada pada sekumpulan data dengan metode tertentu. Teknik pada data mining sangat bervariasi. selain itu, dibutuhkan pemilihan algoritme yang tepat karena akan bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan.

### 2.3 Text Mining

Text Mining adalah bentuk penerapan teknik data mining yang bertujuan untuk mencari pola pada text sehingga didapatkan sebuah informasi tertentu (Kepin Sihontang, 2019). Proses text mining dibagi kedalam beberapa tahap yaitu, Tokenisasi atau pemotongan kalimat berdasarkan kata yang menyusunnya serta menghapus semua karakter tanda baca. Case Folding atau mengkonversi seluruh kata menjadi huruf kecil, Filtering atau proses pengambilan kata-kata inti dengan cara menghapus kata-kata yang tidak deskriptif atau kurang penting, Stemming

atau proses pencarian kata dasar dari hasil filtering ditahap stemming terdapat proses pengambilan berbagai bentuk kata kedalam representasi serupa.

## 2.4 Pembobotan TF-IDF

Pembobotan TF-IDF atau *Term Frequency Inverse Document Frequency* merupakan sebuah proses transformasi data dari data tekstual kedalam bentuk numerik dan dilakukan pembobotan pada setiap kata (Jeremy Andre Septian, 2020). TF-IDF dapat dijadikan sebagai *size statistik* yang digunakan untuk mengevaluasi seberapa penting sebuah kata didalam suatu dokumen. TF merupakan frekuensi kemunculan fitur atau kata pada tiap data yang merepresentasikan seberapa penting kata itu didalam suatu dokumen. DF merupakan frekuensi dokumen yang mengandung fitur tersebut sehingga diketahui seberapa umum kata tersebut. IDF merupakan inverse dari hasil DF. Hasil dari proses ini merupakan perkalian dari TF dengan IDF. Sehingga dapat diketahui jika kata yang memiliki bobot besar adalah kata yang sering muncul dalam sebuah dokumen dan begitu juga sebaliknya.

## 2.5 Algoritme Naive Bayes Classifier

Klasifikasi algoritme Naive Bayes Classifier adalah salah satu algoritme yang diasumsikan bahwa ada atau tidak ada ciri-ciri tertentu dari sebuah kelas, maka tidak ada hubungannya dengan ciri-ciri dari kelas yang lainnya (Meilena 2021). Berikut adalah bentuk yang umum dari teorema bayes ditunjukkan pada persamaan 2.1 berikut:

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} \quad (2.1)$$

Sedangkan formulasi Naive Bayes Classifier untuk klasifikasi ditunjukkan pada persamaan 2.2 berikut:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2.2)$$

## 2.6 Algoritme Support Vector Machine (SVM)

Algoritme Support Vector Machine atau SVM dalam pembelajaran mesin adalah model pembelajaran supervised dengan mempelajari algoritme yang terkait untuk menguji data dan mengidentifikasi pola, yang mana digunakan untuk regresi dan analisis klasifikasi [Alvini Fikriani 2019]. Teori yang menjadi dasar SVM sudah berkembang sejak 1960-an, namun baru diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992 dan sejak itu SVM berkembang dengan pesat (Ibnu Asror 2019). Berikut adalah bentuk umum dari algoritme SVM ditunjukkan pada persamaan 2.3.

$$LD = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n a_i . a_j . y_i . y_j . x_i . x_j \quad (2.3)$$

## III. ANALISA DAN PERANCANGAN SISTEM

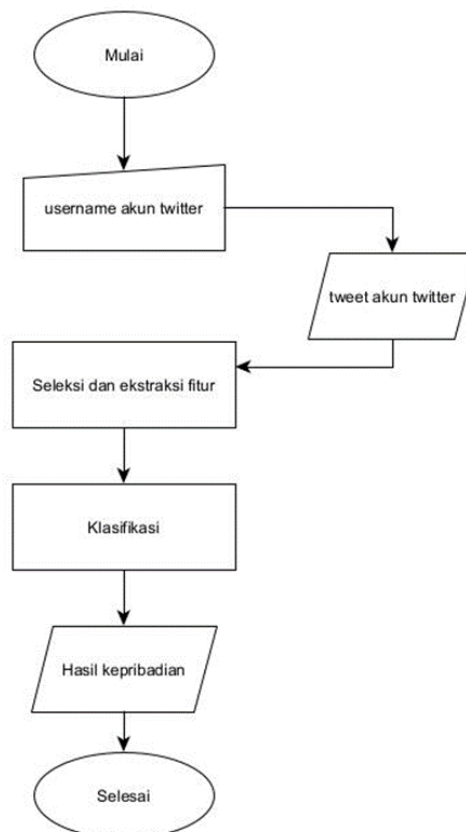
### 3.1 Hasil Analisis

Hasil dari analisis yang terkumpul dari proses pengamatan dan penelitian yang dilakukan dapat diketahui adanya proses prediksi atau klasifikasi kepribadian yang masih dilakukan melalui tes kepribadian dengan cara mewawancarai calon pegawai dan meminta calon pegawai untuk mendeskripsikan diri mereka melalui selembar kertas. Pengembangan sistem dapat menjadi solusi dari proses penentuan kepribadian calon pegawai di CV.Wilis Elektronik. Ciri kepribadian seseorang dapat diperoleh dari teks yang diunggahnya di media sosial Twitter.

Dibuatnya sistem yang baru ini diharapkan dapat membantu mempersingkat proses prediksi atau klasifikasi kepribadian. Sehingga tidak perlu lagi mengadakan wawancara dan serangkaian tes kepribadian seperti mendeskripsikan diri. Adanya implementasi sistem ini akan membuat proses klasifikasi kepribadian semakin ringkas. Untuk dapat melakukan proses klasifikasi kepribadian seseorang berdasarkan teks yang diunggahnya pada Twitter menggunakan metode Naive Bayes maka perlu dilakukan beberapa proses diantaranya :

1. Pengambilan tweet atau Crawling dari akun
2. media sosial Twitter
3. Ekstraksi fitur dan seleksi fitur
4. Stemming
5. Pembobotan TF-IDF
6. Klasifikasi teks menggunakan metode Naïve Bayes

Pada prosedur yang ada, bisa digambarkan sebuah flowchart sebagai visualisasi dari proses yang ada pada sistem klasifikasi kepribadian. Berikut adalah gambar alur diagram atau flowchart seperti pada Gambar 3.1



**Gambar 3.1** Flowchart sistem klasifikasi kepribadian

Gambar 3.1 menggambarkan sebuah flowchart dimulai dari HRD memasukkan username akun twitter melalui form inputan yang ada pada sistem klasifikasi kepribadian. Selanjutnya sistem akan melakukan crawling atau pengambilan data. Data tersebut akan disimpan dalam bentuk dokumen text atau string panjang untuk dilakukan seleksi dan ekstraksi fitur seperti tokenisasi, stemming, penghilangan stop words dan pembobotan TF-IDF. Setelah proses seleksi dan ekstraksi fitur selesai dilakukan. Sistem akan melakukan perhitungan klasifikasi menggunakan metode Naïve Bayes.

### 3.2 Representasi Model

Data yang digunakan didalam sistem dibagi menjadi beberapa bagian yaitu data masukan, data train dan data keluaran. Data masukan adalah data pertama yang akan diproses oleh sistem, data masukan berupa tweet dalam bentuk text yang diambil dari seorang pengguna twitter. Contoh data masukan dapat dilihat pada Gambar 3.2



**Gambar 3.2** Contoh Data Masukan berupa Tweet dari salah satu pengguna

Data train adalah data yang digunakan dalam proses klasifikasi pada sistem. Contoh data train dapat dilihat pada tabel 3.1

**Tabel 3.1** Data train pada sistem

Nama Data	Keterangan
Wordlist_id	Kumpulan kata yang merupakan kata dasar untuk dijadikan fitur perhitungan klasifikasi
Ina_stopwords	Kumpulan kata stop words bahasa Indonesia
Slangword_id	Kumpulan kata non formal atau tidak baku yang sering digunakan oleh sekelompok orang tertentu
User_freq_words	Kumpulan kata-kata yang sering digunakan pengguna
Test_data	Teks dari kumpulan tweet yang digunakan sebagai data percobaan
MNB_classifiers	Kumpulan classifiers untuk metode Naïve Bayes
MNB_predict	Hasil prediksi label metode Naïve Bayes

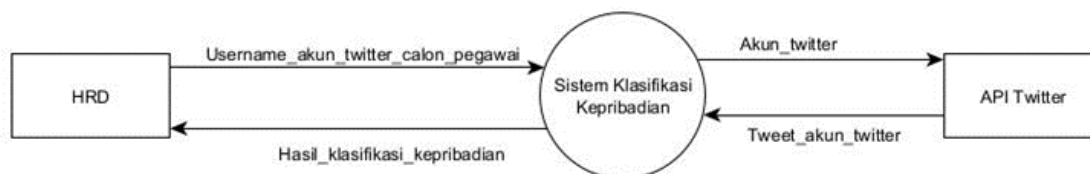
Data keluaran adalah data hasil klasifikasi kepribadian dari sistem. Data keluaran berupa nilai Ya atau Tidak dari setiap label kepribadian dari pengguna yang diklasifikasi. Terdapat juga nilai Karakteristik Utama yang diperoleh dari pengertian awal label tersebut. Sebagai contoh, pada label Neuroticism, nilai “Ya” berarti ciri pengguna tersebut adalah “mudah terpengaruhi oleh emosi negatif”. Begitu juga sebaliknya, jika nilai “Tidak” maka pegawai tersebut “tidak mudah terpengaruhi oleh emosi. Selain itu terdapat nilai Karakteristik Berdasarkan Gabungan antara dua kepribadian yang didapat dari kombinasi antara 2 label kepribadian. Sebagai contoh, seorang pengguna mendapat nilai “Ya” pada label Neuroticism dan “Tidak” pada label Agreeableness, maka orang tersebut cenderung memiliki karakteristik “Temperamental, Pemarah dan Tidak Sabar” [Rizka Shinta Wulandari, 2021]. Tabel dimensi karakteristik utama kepribadian dan tabel kombinasi antara 2 label kepribadian dapat dilihat pada lampiran A.1 dan lampiran A.2. Tabel 3.2 adalah contoh data keluaran dari sistem.

**Tabel 3.2** Contoh Data Keluaran Sistem Klasifikasi Kepribadian

Nama Data	Nilai
Username	Adi_sett
Agreeableness	Ya
Neuroticism	Tidak
Extraversion	Ya
Openness	Ya
Conscientiousness	Tidak
Karakteristik Utama	Tidak mudah terpengaruhi oleh emosi negatif, tidak mudah marah, sabar dst.
Karakteristik Berdasar- kan Gabungan antara dua kepribadian	Mengikuti naluri, Tenang, Sabar, Tidak mudah marah, Individualistis, Eksentrik

### 3.3 Diagram Konteks

Diagram konteks merupakan diagram yang menunjukkan sebuah proses tunggal dalam sistem yang berhubungan langsung dengan semua stakholder yang terlibat dalam sistem seperti pada Gambar 3.3



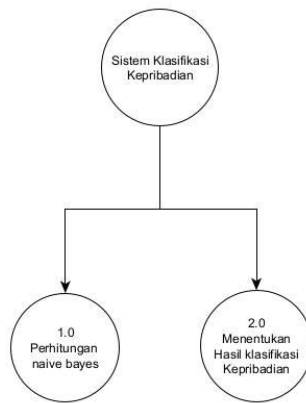
**Gambar 3.3** Diagram Konteks Sistem Klasifikasi Kepribadian.

Keterangan :

1. HRD menginputkan username atau nama pengguna akun twitter dari pegawai yang akan diklasifikasi kedalam sistem klasifikasi kepribadian.
2. Sistem klasifikasi kepribadian meneruskan akun twitter calon pegawai ke API twitter
3. API twitter mengembalikan data tweet berupa teks sesuai dengan akun yang diterima.
4. Sistem klasifikasi kepribadian melakukan perhitungan klasifikasi kepribadian dan mengirimkan hasilnya ke HRD

### 3.4 Diagram Berjenjang

Diagram berjenjang sangat diperlukan dalam perancangan semua proses dan merupakan awal dalam menggambarkan Data Flow Diagram ke level yang lebih dalam.



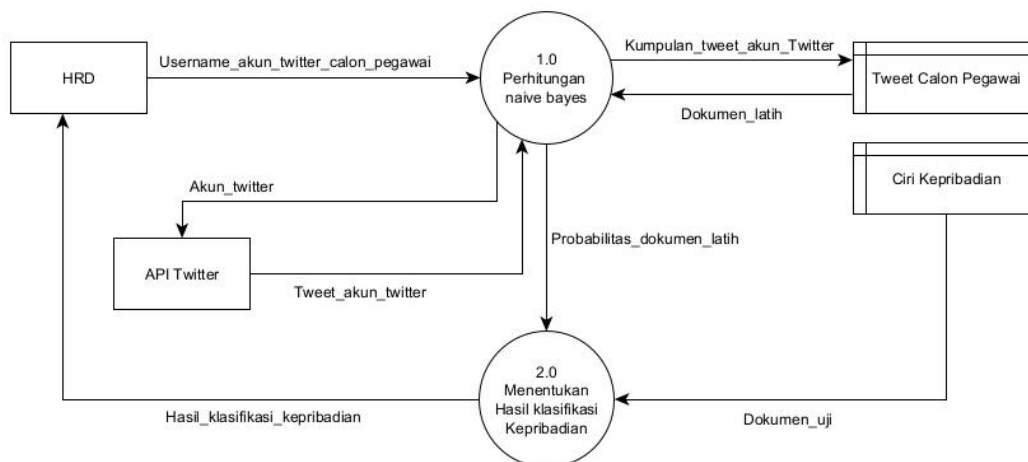
**Gambar 3.4** Diagram Berjenjang Sistem Klasifikasi Kepribadian.

Gambar 3.3 menggambarkan bahwa sistem yang dibuat terdiri dari 2 level, yaitu:

1. Top Level : Sistem Klasifikasi Kepribadian
2. Level 0 : Merupakan hasil uraian dari proses keseluruhan pada sistem klasifikasi kepribadian yaitu:
  - a. Perhitungan Naïve Bayes
  - b. Menentukan hasil klasifikasi kepribadian

### 3.5 Data Flow Diagram Level 0

Data Flow Diagram atau DFD memberikan tampilan secara visual tentang aliran data dan informasi dari sistem. Visual dari DFD ini menggambarkan siapa saja yang terlibat pada sistem dari awal sampai akhir seperti pada Gambar 3.4



**Gambar 3.5** Data Flow Diagram Level 0 Sistem Klasifikasi Kepribadian

Gambar 3.5 menjelaskan bahwa Data Flow Diagram Level 0 yang ditunjukkan pada gambar diatas menjelaskan beberapa proses yang terjadi diantaranya :

1. HRD menginputkan data berupa username akun twitter calon pegawai ke dalam sistem klasifikasi, data tersebut diteruskan ke API Twitter. API Twitter mengirimkan respon berupa tweet akun twitter calon pegawai ke sistem klasifikasi kepribadian, sistem klasifikasi kepribadian menyimpan data tersebut ke dalam database tweet calon pegawai. Sistem mengambil data kepribadian dari database Ciri kepribadian untuk dikirimkan ke proses selanjutnya.
2. Dari data tweet calon pegawai dan data ciri kepribadian, terbentuklah dokumen uji dan dokumen latih yang akan digunakan untuk perhitungan menggunakan metode Naïve Bayes.



3. Hasil perhitungan akan dikirimkan kembali ke HRD berupa laporan hasil klasifikasi kepribadian.

#### IV. IMPLEMENTASI DAN PENGUJIAN SISTEM

##### 4.1 Implementasi Sistem

Sistem klasifikasi kepribadian berdasarkan postingan twitter ini berfokus pada proses prediksi kepribadian berdasarkan postingan twitter menggunakan inventori kepribadian Big Five Personality. Didalam sistem ini terdapat dua algoritme yang digunakan dalam proses klasifikasi, yaitu algoritme Naïve Bayes dan Support Vector Machine atau SVM. Dari kedua algoritme tersebut akan dilakukan perbandingan untuk menentukan manakah dari kedua algoritme tersebut yang memiliki performa terbaik dalam melakukan klasifikasi kepribadian. Didalam Sistem klasifikasi kepribadian berdasarkan postingan twitter ini proses input username akun twitter dilakukan oleh HRD atau panitia penyelenggara tes, username akun twitter yang diinputkan adalah username akun twitter milik pegawai di CV. Wilis Elektronik yang ingin diprediksi kepribadiannya.

Implementasi sistem adalah tahap untuk menentukan sistem supaya siap untuk dioperasikan. Implementasi sistem bertujuan untuk mengkonfirmasi modul perancangan, sehingga sistem ini dapat digunakan. Selanjutnya untuk melihat sistem yang telah dibangun apakah sudah sesuai dengan yang diharapkan, maka perlu pengujian yang mewakili sistem secara keseluruhan. Pada tahap implementasi, sistem klasifikasi kepribadian berdasarkan postingan twitter ini dirancang untuk dapat digunakan dalam keadaan yang sebenarnya.

Implementasi sistem ini membutuhkan software pendukung dalam pembuatannya seperti Visual Studio Code, Lampp server, dan Mozilla firefox. Serta membutuhkan library pendukung diantaranya scikit-learn, pandas, tweepy, sastrawi, matplotlib, skmultilearn, nltk, seaborn, dan xlrd. Aplikasi ini dibangun menggunakan sebuah framework bernama streamlit dengan versi terbaru saat ini yaitu versi 1.9.0.

##### 4.1.1 Pengujian Proses Klasifikasi

HRD atau penyelenggara tes menginputkan data berupa username akun twitter dari pegawai yang ingin diprediksi pada sidebar sebelah kiri halaman dashboard. Setelah username berhasil diinputkan selanjutnya HRD atau penyelenggara tes menekan tombol klasifikasi yang berada dibawah field inputan username. Secara default jika username belum diinputkan sistem akan memberi notifikasi yang mengintruksikan untuk menginputkan username terlebih dahulu. Tampilan dashboard dapat dilihat pada Gambar 4.1



**Gambar 4.1** Tampilan Halaman Input Data Sistem Klasifikasi Kepribadian

Setelah HRD atau penyelenggara tes menekan tombol klasifikasi sistem akan memproses data dan melakukan klasifikasi atau prediksi kepribadian pegawai berdasarkan username akun twitter yang sudah diinputkan. Setelah semua proses selesai dilakukan, sistem klasifikasi kepribadian akan menampilkan hasil klasifikasi di halaman yang sama yaitu halaman dashboard. Hasil klasifikasi meliputi tabel klasifikasi kepribadian, karakteristik utama, dan karakteristik berdasarkan gabungan antara dua kepribadian. Tampilan hasil prediksi atau klasifikasi dapat dilihat pada Gambar 4.2

**Sistem Klasifikasi Kepribadian Berdasarkan Postingan di Twitter**

**User Lookup**

Profil Pengguna

Created : 2022-04-26 05:03:43+08:00

Username : @sudutpelita

Lima Tweet Terbaru :

Hasil Klasifikasi Kepribadian :

extraversion	Neuroticism	Agreeableness	Conscientiousness	Openness
Ya	Tidak	Ya	Tidak	Ya

**Karakteristik Utama :**

senang berhubungan dan bersosialisasi dengan orang lain , tidak mudah terpengaruh oleh emosi negatif , senang bekerja sama dengan orang lain , Santai dan kurang termotivasi untuk sukses , Menyukai perubahan dan mempunyai keingintahuan tinggi terhadap hal-hal baru.

**Karakteristik Berdasarkan Gabungan antara dua kepribadian :**

Ceroboh, Susah diatur, Sangat tegang, Emosional, mudah tertipu, Sederhana, Bergantung pada orang lain, Menyukai keributan, Nakal, Toleran, Penyayang, Penurut, Genit, Emosi meledak-ledak, Detail, Sentimental, Penyayang, Sensitif, Mengikuti naluri, Bertindak diluar kebiasaan, Cemas berlebihan, Ingin menjadi pusat perhatian.

**Gambar 4.2** Tampilan Halaman Hasil Prediksi Sistem Klasifikasi Kepribadian

#### 4.1.2 Pengujian Evaluasi dan Perbandingan Model

Setelah sistem berhasil melakukan prediksi atau klasifikasi kepribadian, selanjutnya akan dilakukan evaluasi dan juga perbandingan model. Model berisi 1000 kelas yang masing-masing memiliki lima label. Label dimodelkan berdasarkan Big Five Personality yaitu EXT atau Extraversion, NEU atau Neuroticism, ARG atau Agreeableness, CON atau Conscientiousness, dan OPN atau Openness. Selain itu model juga dibagi menjadi 4 bagian yaitu train\_data dan train\_label yang akan digunakan sebagai dokumen latih, test\_data dan test\_label yang akan digunakan sebagai dokumen uji. Dokumen latih berjumlah 800 data sedangkan dokumen uji berjumlah 200 data. Dari proses evaluasi dan perbandingan model, sistem akan menghasilkan output berupa classifications report dan tabel confusion matrix dari kedua algoritme yang digunakan, yaitu Naïve Bayes dan SVM. Tampilan classifications report dapat dilihat pada Gambar 4.3 dan Gambar 4.4. sedangkan untuk tampilan tabel confusion matrix dapat dilihat pada Gambar 4.5 dan 4.6.

Nilai Akurasi Pengujian & Pelatihan Dataset :

CLASSIFICATION REPORT NAIVE BAYES

Accuracy score : 60.0

	precision	recall	f1-score	support		
Extraversion			0.64	0.47	0.54	134
Neuroticism			0.60	0.59	0.59	116
Agreeableness			0.65	0.62	0.64	113
Conscientiousness			0.82	0.59	0.68	143
Openness			0.51	0.71	0.59	73
micro avg			0.64	0.58	0.61	579
macro avg			0.64	0.60	0.61	579
weighted avg			0.66	0.58	0.61	579
samples avg			0.67	0.58	0.59	579

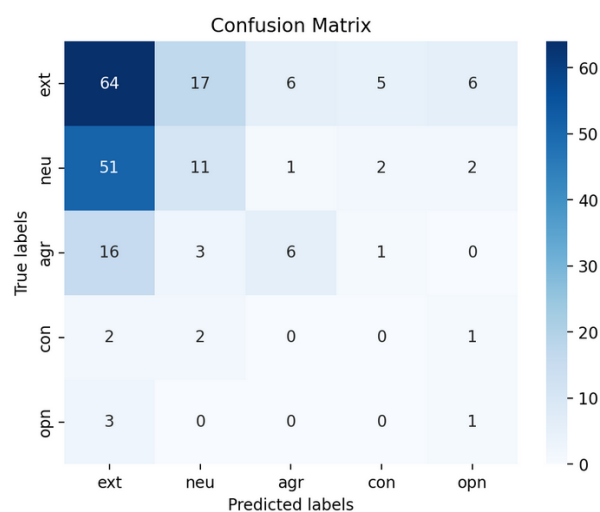
Gambar 4.3 Tampilan Hasil Classifications Report Algoritme Naïve Bayes

CLASSIFICATION REPORT SVM

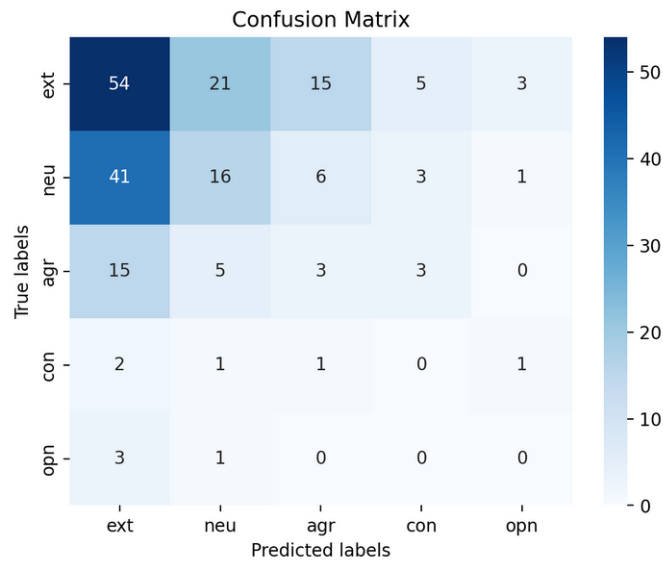
Accuracy score : 49.99999999999999

	precision	recall	f1-score	support		
Extraversion			0.53	0.46	0.49	113
Neuroticism			0.51	0.59	0.55	98
Agreeableness			0.57	0.58	0.57	106
Conscientiousness			0.60	0.51	0.55	121
Openness			0.58	0.56	0.57	106
micro avg			0.56	0.54	0.55	544
macro avg			0.56	0.54	0.55	544
weighted avg			0.56	0.54	0.55	544
samples avg			0.56	0.53	0.51	544

Gambar 4.4 Tampilan Hasil Classifications Report Algoritme SVM



Gambar 4.5 Tampilan Tabel Confusion Matrix Algoritme Naïve Bayes



**Gambar 4.6** Tampilan Tabel Confusion Matrix Algoritme SVM

## 4.2. Analisa Hasil Pengujian Sistem

Hasil pengujian terhadap sistem klasifikasi kepribadian berdasarkan postingan di twitter telah dilakukan dan didapatkan kesimpulan bahwa sistem dapat melakukan klasifikasi kepribadian berdasarkan postingan di twitter. Secara umum diperoleh hasil sesuai dengan yang diharapkan. Mulai dari proses klasifikasi kepribadian sampai dengan evaluasi dan perbandingan model. Berdasarkan proses evaluasi dan perbandingan model, dapat disimpulkan bahwa metode atau algoritme yang memiliki performa terbaik dalam melakukan klasifikasi text adalah algoritme Naïve Bayes dengan akurasi skor sebesar 60% sedangkan algoritme SVM hanya mendapatkan akurasi skor sebesar 49.9%. selain melalui akurasi skor hasil perbandingan juga didapat melalui tabel confusion matrix dimana algoritme Naïve bayes cenderung dapat memprediksi label dengan benar dan hanya sedikit label yang diprediksi salah jika dibandingkan dengan algoritme SVM. Oleh sebab itu kedepannya, hasil klasifikasi yang akan dijadikan sebagai acuan adalah klasifikasi menggunakan algoritme Naïve bayes. Sistem klasifikasi kepribadian ini dapat membantu mengklasifikasikan kepribadian pegawai di CV. Wilis Elektronik berdasarkan postingannya di twitter. Sehingga tidak perlu lagi diadakan tes kepribadian dengan cara mendeskripsikan diri dan mengadakan wawancara.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi yang dilakukan pada Sistem Klasifikasi Kepribadian Berdasarkan Postingan Twitter, dapat diambil kesimpulan bahwa sistem dapat melakukan prediksi atau klasifikasi kepribadian. Selain itu dapat disimpulkan juga bahwa algoritme Naïve Bayes memiliki tingkat akurasi yang lebih baik dibandingkan dengan algoritme Support Vector Machine atau SVM dalam mengklasifikasikan kepribadian berdasarkan postingan di twitter. Secara umum diperoleh hasil yang sesuai dengan yang diharapkan, mulai dari proses crawling data hingga proses klasifikasi. Sistem Klasifikasi Kepribadian Berdasarkan Postingan Twitter ini dapat menyederhanakan tes kepribadian yang ada di CV. Wilis Elektronik. Sehingga HRD tidak perlu lagi mengadakan sesi wawancara dan deskripsi diri untuk mengetahui kepribadian pegawainya

## 5.2 Saran

Adapun saran yang dapat diberikan dari pembuatan sistem klasifikasi kepribadian ini adalah diharapkan pada pengembangannya Sistem Klasifikasi Kepribadian ini tidak hanya berbasis website tetapi juga dapat dikembangkan ke aplikasi mobile. Selain itu proses translate pada data ciri kepribadian dapat dilakukan secara spesifik bersama seseorang yang ahli dibidangnya untuk meningkatkan akurasi dan mendapatkan hasil yang lebih optimal dalam mengklasifikasikan kepribadian berdasarkan postingan di twitter.

### DAFTAR PUSTAKA

- [1] Arifin, N., Enri, U., & Sulistiyowati, N. (Vol. 6 No. 2 Desember 2021). Penerapan Algoritme Support Vector Machine (Svm) Dengan Tf-Idf N-Gram Untuk Text Classification. *String (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 2549 - 2837.
- [2] Deolika, A., Kusriani, & Luthfi, E. T. (Vol.3, No.2. Desember 2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*
- [3] Fikriani, A., Asror, I., & Murti, Y. R. (Vol.6, No.3 Desember 2019). Klasifikasi Kepribadian Berdasarkan Data Twitter Dengan Menggunakan Metode Support Vector Machine. *E-Proceeding Of Engineering*, Page 10436.
- [4] Gunawan, K. I., & Santoso, J. (Volume 03, Nomor 01, Maret 2021). Multilabel Text Classification Menggunakan Svm Dan Doc2vec Classification Pada Dokumen Berita Bahasa Indonesia. *Journal Of Information System, Graphics, Hospitality And Technology*.
- [5] Handayani, F., & Pribadi, F. S. (Januari - Juni 2015). Implementasi Algoritme Naive Bayes Classifier Dalam Pengklasifikasian Teks Otomatis Pengaduan Dan Pelaporan Masyarakat Melalui Layanan Call Center 110. *Jurnal Teknik Elektro Vol. 7 No. 1*.
- [6] Mar'i, F., Mahmudy, W. F., & Yusainy, C. (Oktober 2019). Sistem Rekomendasi Profesi Berdasarkan Dimensi Big Five Personality Menggunakan Fuzzy Inference System Tsukamoto. *Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik) Vol. 6, No. 5, 457-466*.
- [7] Meilana, Astuti, Y., Wulandari, I. R., Sulistiyowati, I., & M, B. A. (Volume 10, Nomor 2, Mei 2021). Algoritme Naive Bayes Untuk Mengklasifikasikan Kepribadian Siswa Smp Berdasarkan Tipologi Hippocrates-Galenus. *Sistemasi: Jurnal Sistem Informasi*, 480-489.
- [8] Pamungkasa, F. S., & Kharisudin, I. (2021). Analisis Sentimen Dengan Svm, Naive Bayes Dan Knn Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter. *Prisma, Prosiding Seminar Nasional Matematika*, 628-634.
- [9] Rahman, M., Darmawidjadja, M., & Alamsah, D. (Jan 2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian
- [10] Regularization Neural Network (Rbnn). *Jurnal Informatika Vol.11.No1*.
- [11] Saputro, I. W., & Sari, B. W. (Vol. 6, No. 1, Januari 2019). Uji Performa Algoritme Naive Bayes Untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*
- [12] Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitt terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan Tf Idf Dan K Nearest Neighbor. *Journal Of Intelligent Systems And Computation*
- [13] Sihotang, K., & Ghaniy, R. (Volume 9 Nomor 1 Bulan Mei 2019). Penerapan Metode Naive Bayes Classifier Untuk Penentuan Topik Tugas Akhir Pada Website Perpustakaan Stikom Binaniaga. *Jurnal Ilmiah Teknologi - Informasi Dan Sains (Teknois)*, 63-72.
- [14] Wulandari, R. S., Kusuma, P. D., & Setianingsih, C. (Vol.8, No.6 Desember 2021). Sistem Pemetaan Faktor Kepribadian Big Five Sebagai Rekomendasi Pemilihan Pekerjaan Dengan Algoritme C4.5. *E-Proceeding Of Engineering*, 12178.