

Prediksi Kepribadian Berdasarkan Status Sosial Media *Facebook* Menggunakan Metode Naïve Bayes Dan K-Nearest Neighbor

Andi Oktafiqurahman¹, Kusri², Asro Nasiri³

^{1,2,3}MTI Universitas AMIKOM Yogyakarta

andi.oktafiqurahman@students.amikom.ac.id¹, kusri@amikom.ac.id², asro.nasiri@amikom.ac.id³

ABSTRACT

Social media is now almost a part of life for people of almost all ages with diverse personalities. This study was conducted to predict a person's personality based on Facebook social media status using the Naïve Bayes algorithm and K-NN personalities based on the big five. The data taken from *MyPersonality* which is divided into 40% training data, and 60% test data produces an accuracy rate of 100%, 100% precision, 100% recall can be concluded that the prediction variables are (significant) with the Naive Bayes method. While predictions using the KKN method showed an average personality prediction value with an accuracy of 58.96% (Insignificant), precision 99.12% (Significant), and recall 2.34% (Insignificant).

Keywords: Naïve Bayes, K-NN, Prediction, Personality, Facebook

I. PENDAHULUAN

Sosial media sekarang sangat berkembang dan penggunaanya pun sangat beragam, hampir semua golongan mempunyai sosial media baik pria maupun wanita. Sosial media ini memang diciptakan untuk memudahkan masyarakat dalam melaksanakan kegiatan-kegiatan sulit yang dilakukan manusia[1].

Media sosial menciptakan interaksi dengan sesama pengguna terjadi begitu saja dengan orang yang kita kenal maupun dengan orang baru yang mana tidak saling mengenal sebelumnya. Media sosial adalah sebuah bentuk baru dari sebuah teknologi informatika yang sangat berkembang pesat dan penggunaanya selalu bertambah dengan cepat setiap harinya. Dengan perkembangan teknologi yang sangat pesat sekarang media sosial selalu memiliki bentuk dan jenis yang berubah tetapi saat ini media sosial memiliki beberapa bentuk dan jenis seperti *weblogs* ("blogs"), *microblogs*, situs jejaring sosial, dan *media-sharing*, RSS dan *feed* sindikasi web lain, *wiki*, sosial *bookmark*, *mashup*, *widget*, *microblogging*, dan lain sebagainya[2].

Di dalam dunia psikologi terdapat beberapa kepribadian yang dimiliki oleh manusia itu sendiri, diantaranya *sanguine*, *koleris*, *melankolis*, *plegmatis*. Dimana orang yang bertipe *sanguine* ini memiliki sifat yang lebih terbuka dan mudah bersosialisasi, sedangkan *koleris* memiliki sifat yang lebih menggunakan logika, dan *melankolis* cenderung memiliki kepribadian perfeksionis, sensitif, berempati, dan cenderung pendiam, untuk *plegmatis* sendiri memiliki sifat yang lebih tenang dan memiliki empati yang tinggi.

Terdapat sekali kata-kata yang dapat mendeskripsikan kepribadian seseorang. Hampir setiap dari kita menggunakan saat berinteraksi atau berkomunikasi dengan orang lain sebagai makhluk sosial. Kita sering mengatakan bahwa si A seperti ini dan si B seperti itu atau mungkin si A memiliki "beberapa kepribadian" dan si B "tidak memiliki kepribadian sama sekali". Kita menggunakan kata kepribadian itu sebenarnya kita sendiri. Sering kali kita sebenarnya hanya menggambarkan watak atau beberapa kepribadian tertentu. Pemikiran bahwa kita mempunyai beberapa kepribadian yang berbeda ternyata kurang tepat. Namun ini sangat menarik. Dalam berkomunikasi kita di sosial media tanpa harus melihat orang secara langsung kita seringkali menetapkan kepribadian seseorang sama atau berbeda dengan diri kita dimana pada dasarnya atau lebih baik kita katakan setiap orang memiliki pola kepribadian masing-masing.

Penelitian ini bertujuan untuk melakukan prediksi kepribadian apa yang dimiliki oleh *user* dari media sosial *facebook* menggunakan metode *naïve bayes* dan *k-nearest neighbor*. dengan cara melakukan prediksi berdasarkan status media sosial *facebook*. Dengan mengetahui kepribadian seseorang maka kita akan lebih mudah untuk melakukan interaksi atau juga memudahkan perusahaan yang ingin mengetahui kepribadian calon karyawan dari media sosial apakah kepribadiannya berjalan sejajar dengan visi dan misi perusahaan tersebut. Seperti yang kita ketahui di dunia pekerjaan kepribadian seseorang bisa berdampak dengan hasil kerjanya. Algoritma *naïve bayes* dan *k-nearest neighbor* dapat

digunakan untuk melakukan prediksi tersebut. Algoritma *naïve bayes* memang sangat populer dalam penelitian yang dilakukan untuk melakukan prediksi. Dan kelebihan dari algoritma ini yaitu tidak memerlukan begitu banyak data pelatihan (*training data*) untuk menentukan estimasi parameter yang digunakan dalam proses pengklasifikasian. begitu juga dengan *k-nearest neighbor* algoritma ini cukup populer karena algoritma ini dapat digunakan dalam proses klasifikasi maupun regresi. Dapat digunakan dalam jumlah kelas yang beragam. Sangat mudah jika terdapat penambahan data karena data baru dapat langsung dikelaskan dan tidak memerlukan proses training dan proses pembangunan model, dan lain sebagainya.

Tujuan dalam penulisan penelitian ini adalah mengetahui kepribadian seseorang dari status *facebook* menggunakan algoritma *naïve bayes* dan *k-nearest neighbor*.

Batasan masalah akan dirumuskan sebagai berikut:

- Penelitian dilakukan berdasarkan status pengguna sosial media facebook
- Membuat sistem prediksi kepribadian berdasarkan status sosial media menggunakan model pendekatan *big five personality traits*.
- Menggunakan algoritma *naïve bayes* dan *k-nearest neighbor* untuk membuat sistem prediksi

II. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Mohammad Ichsanudin A dengan judul Prediksi kepribadian berdasarkan media sosial twitter menggunakan metode *naïve bayes classifier* dengan tujuan menghasilkan prediksi kepribadian dengan status sosial media twitter dengan model prediksi yang digunakan yaitu *big five*. Menghasilkan kesimpulan dari penelitian ini bahwa akurasi yang dihasilkan cukup rendah dengan akurasi hanya sebesar 42%. Yang diperoleh dari hasil prediksi *label* kepribadian tertinggi terdapat pada *label Agreeableness* dengan jumlah 167. Sedangkan, nilai terendah diperoleh oleh *label Extraversion* dengan jumlah 14 tweet. Namun pada penelitian terdapat kelemahan yaitu dengan jumlah *dataset* yang digunakan tidak banyak dikarenakan tidak semua akun tidak memiliki ciutan atau tweet lebih dari 100 tweet [3].

Penelitian yang dilakukan oleh Sandhya katiyar dengan judul *personality prediction from stack overflow by using naïve bayes*

theorem in data mining. Penelitian ini bertujuan untuk menganalisis kepribadian seseorang di forum *stack overflow* untuk membangun sebuah tim dalam suatu project menggunakan algoritma *naïve bayes*. Hasil dari penelitian ini menghasilkan kesimpulan dari 16 jenis kepribadian berdasarkan MBTI di dapatkan hasil yaitu ENTP yang mendefinisikan karakteristik seperti jujur, tegas dan memiliki kualitas kepemimpinan. Penelitian memiliki kesamaan dengan penulis yaitu melakukan prediksi kepribadian dan menggunakan algoritma *naïve bayes* namun sosial media yang digunakan berbeda dan jumlah kepribadian yang digunakanpun berbeda[4].

Penelitian yang dilakukan oleh Siska Febriani dengan judul Analisis data hasil diagnosa untuk klasifikasi gangguan kepribadian menggunakan algoritma C4.5. dengan tujuan untuk melakukan prediksi terhadap gangguan kepribadian untuk mendapatkan informasi yang dibutuhkan oleh masyarakat. Dalam penelitian ini hanya menggunakan 3 gangguan psikologis yaitu gangguan kecemasan (GAD), gangguan mood (gangguan depresi mayor), dan gangguan somatoform (gangguan konversi) dari penelitian ini menghasilkan tingkat akurasi sebesar 72,67%. Kesamaan penelitian ini dan penulis adalah melakukan prediksi kepribadian hanya saja penelitian melakukan prediksi terhadap gangguan kepribadian dan juga menggunakan algoritma C4.5, sedangkan penulis menggunakan *naïve bayes* dan *k-nearest neighbor* [5].

Penelitian yang dilakukan oleh Irina Deeva dengan judul "*Computational Personality Prediction Based on Digital Footprint of A Social Media User*" pada penelitian ini menentukan apakah psikometri dapat berkolerasi sama baiknya di lingkungan media yang berbeda dengan tujuan tersebut matriks kolerasi dibangun antara fitur dan psikometrik. Dengan data yang dibagi menjadi 30% data uji dan 70% data latih untuk menghindari pelatihan yang berulang oleh model. Dengan itu metode *train-test-split* dengan *hyperparameter shuffle* = 42. Hasil regresi unvariat disajikan untuk facebook. Sedangkan untuk penelitian yang akan dilakukan adalah untuk melakukan prediksi dimana membandingkan data yang telah dilakukan prediksi dengan model pendekatan *deep learning* untuk permodelan kepribadian *big five personality traits* dengan model pendekatan *machine learning* [6].

Penelitian yang dilakukan oleh Hans Christian dengan judul "*Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging*" pada penelitian ini menunjukkan perbandingan metode ekstraksi fitur yang berbeda dengan pendekatan algoritma yang berbeda dalam membangun sistem kepribadian untuk sumber data media sosial facebook. Penelitian ini melakukan pendekatan *deep learning* arsitektur dengan model bahasa BERT, RoBERTa, dan XLNet. fitur statistik dan rata-rata model mengungguli sebagian besar model kepribadian dengan menghasilkan akurasi tertinggi 86,17% dan skor pengukuran f1 0,912 pada dataset Facebook dan akurasi 88,49% dan skor pengukuran f1 0,882 pada dataset Twitter. Sedangkan untuk penelitian yang akan dilakukan menggunakan metode machine learning dengan menggunakan algoritma *naïve bayes* dan *k-nearest neighbor*[7].

III. METODE PENELITIAN

3.1 Big Five Personality

Extraversion, *Agreeableness*, *Conscientiousness*, *Neuroticism* dan *Openness* adalah ciri-ciri dari kepribadian tingkat tinggi yang mencakup teori *The Big Five*. Ciri-ciri dari *big five personality* adalah orthogonal dan bipolar, dimana dari profil individu ke individu di *big five* dapat menjadi unik. Dengan pengecualian keterbukaan terhadap pengalaman yang muncul kemudian di masa remaja, beberapa peneliti lebih lanjut berpendapat bahwa *big five* adalah sifat bawaan disposisi genetic yang tidak berubah dari waktu ke waktu. Namun, kemungkinan besar ada hubungan transaksional antara kepribadian dan factor kontekstual sosial. Sifat kepribadian dari *big five* dibagi menjadi 5 dengan dimana:

1. *Extraversion* (Ekstraversi)

Dimensi kepribadian ini berkaitan dengan tingkat kenyamanan seseorang dalam berinteraksi dengan orang lain. Karakteristik positif individu *Extraversion* adalah senang bergaul, mudah bersosialisasi, hidup berkelompok dan tegas.

2. *Agreeableness* (Mudah Akur atau Mudah Bersepakat)

Individu yang ini cenderung patuh dengan individu lainnya dan memiliki kepribadian yang ingin menghindari konfli. Karakteristik positif dari dimensi individu ini adalah kooperatif (dapat bekerjasama),

penuh kepercayaan, bersifat baik, hangat dan lembut dan juga suka membantu.

3. *Conscientiousness* (Berhati-hati)

Individu di dimensi ini memiliki sifat yang lebih berhati-hati dalam melakukan suatu Tindakan atau sangat penuh pertimbangan dalam mengambil sebuah keputusan, dimensi kepribadian ini juga memiliki sifat yang disiplin diri yang tinggi dan dapat dipercaya. Karakteristik positif pada dimensi ini adalah dapat diandalkan, bertanggung jawab, tekun, dan berorientasi pada pencapaian.

4. *Neuroticism* (Neurotisme)

Dimensi kepribadian ini memiliki kemampuan seseorang dalam menahan tekanan atau stress. Karakteristik positif dari dimensi ini disebut dengan *Emotional Stability* atau emosi yang stabil. Individu ini dengan emosional yang stabil cenderung tenang saat menghadapi masalah, percaya diri, memiliki pendirian yang teguh.

5. *Openness* (Terbuka)

Dimensi kepribadian ini mengelompokkan individu berdasarkan ketertarikan terhadap hal-hal yang baru dan berkeinginan untuk mengetahui serta mempelajari sesuatu yang baru. Karakteristik positif dari dimensi kepribadian ini adalah lebih kreatif, imajinatif, intelektual, penasaran dan berpikiran luas.

3.2 Natural Language Processing (NLP)

Natural Language Processing adalah salah satu bagian dari *artificial intelligence* (AI) yang memberikan computer dapat berinteraksi seperti membaca, memahami, dan menafsirkan maksud dari manusia menggunakan bahasa alami. NLP membantu mengukur sentiment dan menentukan bagian mana yang penting dari bahasa manusia. Untuk computer ini adalah hal yang susah untuk dilakukan karena banyaknya data yang tidak terstruktur, tidak memiliki aturan formal, dan tidak ada kolerasinya di dunia nyata. Beberapa tugas umum yang dilakukan NLP adalah ekstraksi informasi, kategorisasi dokumen, pencocokan teks semantik.

a. *Kepribadian*

Dalam dunia psikologi terdapat banyak kepribadian yang dimiliki oleh manusia. Banyak kata-kata yang dapat mendeskripsikan kepribadian atau watak seseorang. Setiap hari kita hampir menggunakan saat berinteraksi

dengan orang lain. Kita terkadang mengatakan bahwa si A seperti ini dan si B seperti ini atau si A mempunyai “dua kepribadian” atau si B “tidak memiliki kepribadian sama sekali” atau bahkan memiliki banyak kepribadian. Sering kali kita menggunakan kata kepribadian itu seenaknya. Namun, sebenarnya kita hanya menggambarkan watak atau sejumlah kepribadian tertentu.

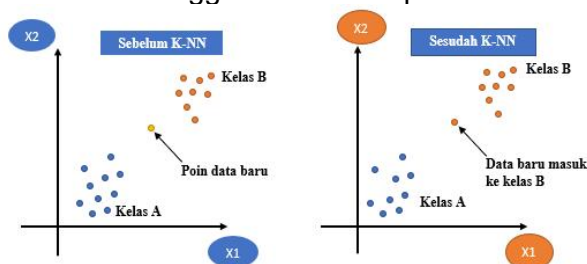
b. Naïve Bayes

Naïve bayes adalah suatu algoritma yang biasa digunakan untuk melakukan klasifikasi biner dan multikelas. Algoritma ini dapat menilai model dengan sangat cepat dan skala secara linear dalam jumlah prediksi dan baris. Algoritma ini juga dapat melakukan klasifikasi yang mempresentasikan setiap kelas object berdasarkan kesimpulan atau rekapitulasi probabilistic dan menemukan kemungkinan besar kelas yang sesuai dengan tiap objek yang akan ditentukan kelasnya dari data uji yang ada berdasarkan atribut atau variable yang telah diketahui nilainya. Untuk persamaan dari naïve bayes dapat dilihat dibawah ini:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad \dots\dots\dots(1)$$

c. K-Nearest Neighbor (KNN)

Algoritma KNN atau K-nearest Neighbor adalah sebuah algoritma yang digunakan untuk mengelompokkan suatu data berdasarkan data training (data latih) yang berasal dari k tetangga terdekat (*nearest neighbor*). Dimana k adalah jumlah tetangga terdekat. Konsep cara kerja dari algoritma ini adalah yaitu mendapatkan jarak terpendek antara data yang akan dilakukan evaluasi dan k tetangga dalam data latih. Tujuan algoritma *K-Nearest Neighbor* adalah untuk mengklasifikasikan objek baru yang didasarkan pada atribut dan sample data training. Untuk proses perhitungan jarak dilakukan menggunakan konsep *Euclidean*.



Gambar 1 K-NN

IV. Implementasi Sistem dan Hasil

1. Preprocessing Data

Dataset yang digunakan dalam penelitian Prediksi Kepribadian Berdasarkan Status Sosial Media Facebook Menggunakan Metode Naïve Bayes Dan Knn terdiri dari 9.990 Data Status Facebook dari penelitian sebelumnya dari *Kaggle Machine Learning* yang sudah di uji menggunakan *Big Five Personality Trait* dengan lima kriteria variabel yaitu sEXT, sNEU, sAGR, sCON, sOPN.

2. Implementasi Metode Naïve Bayes

Langkah selanjutnya setelah *Dataset* sudah siap kemudian dilakukan penerapan metode *Naïve Bayes* menggunakan *Software RapidMiner Studio*. Dilakukan implementasi metode Naïve Bayes untuk memprediksi kepribadian berdasarkan status sosial media Facebook. Model Naïve Bayes telah dilatih menggunakan data latih yang telah melalui tahap *preprocessing*, pada penelitian ini memprediksi kriteria variabel yaitu sEXT,sNEU,sAGRsCON,sOPN dengan hasil implementasi *naïve bayes* dapat dilihat pada Tabel 1.

Tabel 1 Hasil metode Naïve Bayes

Variabel	Accuracy	Precision	Recall
sEXT	100%	100%	100%
sNEU	99,95%	99,92%	100%
sAGR	99,89%	99,80%	100%
sCON	99,84%	99,89%	99,77%
sOPN	99,95%	99,80%	100%

3. Implementasi Metode KNN

Dilakukan implementasi metode *KNN* untuk memprediksi kepribadian berdasarkan status sosial media Facebook. Model *KNN* telah dilatih menggunakan data latih yang telah melalui tahap *preprocessing*, pada penelitian ini memprediksi kriteria variabel yaitu sEXT,sNEU,sAGRsCON,sOPN dengan hasil implementasi *KNN* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Metode KNN

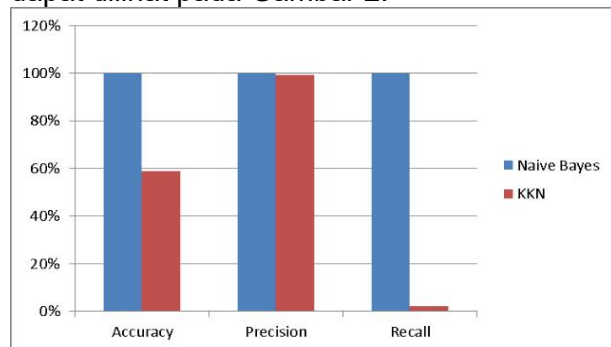
Variabel	Accuracy	Precision	Recall
sEXT	57,11%	100%	0,49%
sNEU	38,46%	100%	2,18%
sAGR	47,04%	100%	0,40%
sCON	57,16%	95,59%	7,45%
sOPN	74,54%	100%	1,03%

4. Perbandingan Hasil Antara Metode Naïve Bayes dan K-NN

Naive Bayes: Metode Naive Bayes didasarkan pada Teorema Bayes, yang menggunakan probabilitas kondisional untuk melakukan klasifikasi. Metode ini mengasumsikan bahwa semua fitur atau variabel adalah independen secara kondisional.

K-Nearest Neighbors (KNN): Metode KNN berdasarkan pada prinsip bahwa objek yang mirip cenderung berada dalam kategori yang sama. Algoritma ini mencari K tetangga terdekat dari objek yang akan diklasifikasikan dan menggunakan mayoritas kategori tetangga tersebut untuk mengklasifikasikan objek tersebut.

Berdasarkan implementasi Metode Naïve Bayes dan KNN yang telah dilakukan terdapat perbandingan hasil yang cukup signifikan, dapat dilihat pada Gambar 2.



Gambar 2 Perbandingan Hasil

Implementasi metode *Naive Bayes* sebagai metode prediksi kepribadian berdasarkan status sosial media *facebook* memperoleh nilai rata-rata cukup signifikan dengan nilai *accuracy* 100%, *Precision* 100%, dan nilai *Recall* 100%. Hal ini berbanding terbalik dengan hasil implementasi metode *KNN* dimana diperoleh rata-rata hasil *accuracy* sebesar 58,69%, *precision* 99,12% dan nilai *Recall* 2,34.

V. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dengan judul "Prediksi Kepribadian Berdasarkan Status Facebook menggunakan Metode *Naive Bayes* dan *K-Nearest Neighbors (KNN)*", dapat disimpulkan bahwa kedua metode *Naive Bayes* dan *KNN* dapat digunakan untuk memprediksi kepribadian berdasarkan status Facebook.

Tabel 3 Kesimpulan

Metode	Accuracy	Precision	Recall
Naive Bayes	100%	100%	100%
KNN	58,96%	99,12%	2,34%

Prediksi kepribadian menggunakan metode *Naive Bayes* menunjukkan rata-rata nilai prediksi kepribadian dengan akurasi 100%, *precision* 100%, *Recall* 100% dapat disimpulkan bahwa variabel prediksi yang (Signifikan) dengan metode *Naive Bayes*. Sedangkan prediksi menggunakan metode *KNN* menunjukkan rata-rata nilai prediksi kepribadian dengan akurasi 58,96% (Tidak signifikan), *precision* 99,12% (Signifikan), dan *recall* 2,34% (Tidak signifikan).

Daftar Pustaka

- [1] M. Astuti, & N. Matondang, "Manajemen Pemasaran UMKM dan Digital Sosial Media". Yogyakarta: DEEPUBLISH, 2020.
- [2] W. Yasya, P. Muljono, & Hardinsyah. "Laporan pengguna media sosial media facebook. pengaruh penggunaan media sosial facebook dan dukungan", Tesis, Sekolah Pascasarjana, Institut Pertanian Bogor, 2019.
- [3] M. Ichsanudin A, A. S. Yuda irawan, & A. Solehudin, "Prediksi Kepribadian Berdasarkan Media Sosial twitter menggunakan Metode Naive bayes Classifier", *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 988-996, 2021.
- [4] S. Katiyar*, S. Kumar, and H. Walia, "Personality prediction from stack overflow by using naïve Bayes theorem in Data Mining," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 1555–1559, 2020.
- [5] S. Febriani, H. Sulistiani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5", *Jurnal Teknologi dan Sistem Informasi*, Vol. 2, No. 4, 2021.
- [6] I. Deeva, "Computational personality prediction based on digital footprint of a social media user," *Procedia Computer Science*, vol. 156, pp. 185–193, 2019
- [7] H. Christian, D. Suhartono, A. Chowanda, and K. Z. Zamli, "Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging," *Journal of Big Data*, vol. 8, no. 1, 2021.