
PENERAPAN TEXT MINING DALAM MENGANALISIS KEPRIBADIAN PENGGUNA MEDIA SOSIAL

M Pramadani Riyanis Putra¹, Kiki Rizky Nova Wardani²

^{1,2}Program Studi Teknik Informatika, Universitas Bina Darma

e-mail: *¹mpramadaniriyanisputra@gmail.com, ²kikirizkynovawardani@binadarma.ac.id.

Abstrak

Facebook merupakan aplikasi jejaring sosial dimana para pengguna mengungkapkan banyak tentang diri mereka sendiri melalui laman postingan mereka. Sehingga penulis ingin mengetahui informasi apa saja yang dapat diambil tentang kepribadian pengguna. Data *mining* memainkan peran penting yang bertujuan untuk mengubah data mentah menjadi suatu struktur yang dapat dimengerti untuk dapat digunakan lebih lanjut. *Text mining* mengacu pada proses mengambil informasi berkualitas tinggi dari teks, salah satu metode klasifikasi yang dapat digunakan ialah algoritma *K-Nearest Neighbor*. Berdasarkan teori *big five personality* hasil penelitian menyimpulkan bahwa tingkat akurasi yang diperoleh yaitu 92.92%, dari 550 data dengan nilai karakter kepribadian *openness* yang paling tinggi yaitu 239, *Conscientiousness* sebanyak 16 data, *Extraversion* sebanyak 173 data, *Agreeableness* sebanyak 50 data, *Neuroticism* sebanyak 33 data dan 39 data yang tidak dapat diklasifikasikan.

Kata kunci : *Text Mining, Big Five Personality, K Nearest Neighbor, Facebook*

Abstract

Facebook is a social networking application where users reveal a lot about themselves through their posting pages. So the writer wants to know what information can be taken about the user's personality. Data mining plays an important role which aims to transform raw data into a structure that can be understood for further use. Text mining refers to the process of retrieving high quality information from text, one of the classification methods that can be used is the K-Nearest Neighbor algorithm. Based on the theory of big five personality the results of the study concluded that the accuracy rate obtained was 92.92%, from 550 data with the highest openness personality character value of 239, Conscientiousness of 16 data, Extraversion of 173 data, Agreeableness of 50 data, Neuroticism of 33 data and 39 data that cannot be classified.

Keywords : *Text Mining, Big Five Personality, K Nearest Neighbor, Facebook*

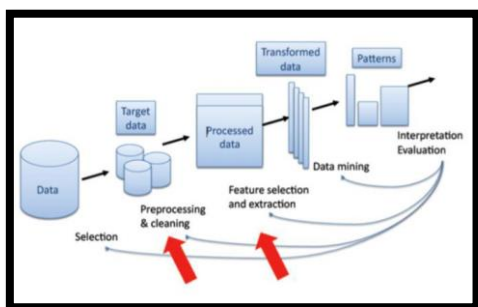
I. PENDAHULUAN

Perkembangan penggunaan media sosial di Indonesia sangat berkembang pesat. Bila dilihat dari jumlah pengguna internet, maka bisa dibilang seluruh pengguna media sosial di Indonesia. *We Are Social* mengatakan 132,7 juta pengguna internet, 130 juta diantaranya pengguna aktif di media sosial. Media sosial adalah sebuah media *online*, dengan para penggunanya dengan mudah berpartisipasi, berbagi. Blog dan jejaring social merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Seorang pasti memiliki macam-macam motivasi dalam menggunakan medsos. Untuk mencari kabar perkembangan sesuatu, untuk berbagi informasi maupun untuk mengikuti salah satu yang menjadi *trend* saat ini yaitu menggunakan medsos sebagai untuk apresiasi diri. *Facebook* salah satu medsos yang sangat populer dan banyak digemari orang di semua dunia. Dari anak orangtua sudah menggunakan *facebook*. Kebanyakan dari mereka memakai *facebook* untuk menjalin hubungan pertemanan dan berkomunikasi dengan banyak orang, yaitu sahabat bahkan teman jauh. Pengertian *facebook* secara lengkap ialah sebuah situs internet sosial yang memungkinkan pengguna dapat saling berinteraksi dengan pengguna lain di seluruh dunia maya. Mark Zuckerberg mendirikan *facebook* saat ia belajar di fakultas psikologi di *Harvard University*. Pada Februari tahun 2004, Mark menegakkan *Facebook*, dimana nama itu berasal dari lembaran *profil* yang dibagikannya kepada mahasiswa baru di *Harvard*. Pada awalnya, *facebook* dikenalkan hanya untuk kalangan mahasiswa. Hal itu digunakan sebagai media perkenalan bagi para mahasiswa *Harvard*. Seseorang yang menjadikan *Facebook* sebagai media sosial favorit

mereka adalah orang yang suka menjalin persahabatan dan juga berbagi. Hal ini dilihat dari orang yang suka menggunakan *Facebook* akan menunjukkan segala kegiatannya di *wall* mereka. Masyarakat dalam mengakses media sosial tentu bukan lagi menyangkut persoalan pribadi saja, melainkan sudah berkembang pada *aksesibilitas* ranah publik yang tinggi. Batas privasi seakan menjadi sampingan manakala menu perhatian ruang publik semakin meningkat. Derasnya perhatian ruang publik yang dikonsumsi menandakan demokrasi semakin menunjukkan tarungnya. Eksistensi dan perhatian masyarakat terhadap apa yang menjadi kehidupan ruang publik sangat diperhatikan. Hal itu ditandai dengan munculnya kepribadian yang berbeda-beda dimana perkataan yang mereka tulis diberanda media social ada yang menghujat, memberitakan berita yang tidak nyata (*hoax*) dan lain sebagainya. Algoritma *Nearest Neighbor* (*k-nearest neighbor*) adalah algoritma untuk melakukan *klasifikasi* terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Kasus khusus di mana *klasifikasi* diprediksikan berdasarkan data pembelajaran yang paling dekat disebut algoritma *k-nearest neighbor*. Dalam metode KNN penggunaan atribut kata yang hadir menjadi dasar klasifikasi untuk penggolongan atau klasifikasi teks. Dari penjelasan tersebut peneliti ingin melakukan pengklasifikasian teks dengan menggunakan metode KNN untuk mendapatkan hasil. Untuk selanjutnya di golongkan mana yang akan masuk dalam kepribadian *Big Five Personality* yaitu *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness* dan *Neuroticism*.

II. TINJAUAN PUSTAKA

Text mining adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terkandung dalam sebuah dokumen. Prosedur utama dalam metode ini terkait dengan menemukan kata-kata yang dapat mewakili isi dari dokumen untuk selanjutnya dilakukan analisis keterhubungan antar dokumen dengan menggunakan metode statistik tertentu seperti analisis kelompok, klasifikasi dan asosiasi. Tahapan dalam *text mining* secara umum dapat dilihat pada gambar di bawah ini.



Gambar 1. Tahapan *Text Mining*

Tahapan dalam *text mining* dalam gambar 1 dapat dijelaskan sebagai berikut :

1. Pada tahap pertama : Melakukan persiapan/pengumpulan data yang dibutuhkan atau disebut dengan data awal. Data ini akan di seleksi kemudian masuk kedalam tahap kedua.
2. Pada tahap kedua : Melakukan *Pre-processing* dan *cleansing*. Adapun tahapan proses *pre processing* meliputi *tokenizing*, *filtering*, *stemming* dan *Tagging*. setelah data dilakukan *pre processing* maka data selanjutnya akan di *cleaning*, proses pembersihan data ini dilakukan untuk menghilangkan *noise* yang ditemukan berupa *missing value*. Setelah dua proses ini dilakukan maka data siap diolah ke proses berikutnya.

3. Pada tahap ketiga : Melakukan pengolahan data dengan menggunakan algoritma data *mining* yang terdiri dari akurasi atau disebut sebagai proses *feature selection* dan *feature extraction* dengan algoritma data *mining*. Pada tahap ini laksanakan terlebih dahulu *feature selection* sebelum *feature extraction*. Proses *feature selection* adalah teknik terpenting dalam *pre-processing*. Teknik ini mengurangi jumlah fitur yang terlibat dalam menentukan suatu nilai kelas target proses pengambilan data ini adalah data yang sudah melewati proses *text mining* atau data yang sudah terstruktur. Kemudian data yang didapatkan dan dilakukan *selection* maka data tersebut di *extraction*. Proses ini ialah salah satu cara untuk mengenali suatu objek tersebut. Tahap inilah yang akan mencocokkan kepribadian dengan data yang didapat.
4. Setelah tahap ketiga selesai maka pada tahap keempat: Menentukan kamus / *dictionary manual* terkait dengan pemilihan status dan di masukan ke dalam *big five personality*.
5. Pada tahap kelima : Melakukan evaluasi komparasi dari hasil untuk mengukur tingkat akurasi kerja proses *text mining* dengan KNN terhadap proses manualnya.

Algoritma *K-Nearest Neighbor* (K-NN) adalah teknik klasifikasi yang sangat populer yang diperkenalkan oleh Fix dan Hodges, yang telah terbukti menjadi algoritma sederhana yang baik. KNN merupakan salah satu metode yang digunakan dalam pengklasifikasian dengan menggunakan algoritma supervised. Ketepatan algoritma KNN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan

relevansinya terhadap klasifikasi. Ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate*. [1]

Algoritma *K-Nearest Neighbor* menggunakan klasifikasi ketetangga (Neighbor) sebagai nilai prediksi dari query instance yang baru. Algoritma ini sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan ketetanggaannya. [2]

- a. Menentukan parameter K
- b. Menghitung jarak antara data yang akan di evaluasi dengan semua pelatihan
- c. Mengurutkan jarak yang terbentuk
- d. Menentukan jarak terdekat sampai urutan k
- e. Memasangkan kelas yang bersesuaian
- f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

III. METODELOGI PENELITIAN

3.1 Analisa Text Mining

1. Proses data awal

Pada penelitian ini data awal yang digunakan dalam text mining adalah data yang di dapatkan dari media sosial facebook yang berupa status penggunaannya dan disebut sebagai data awal. Data yang diambil dengan menggunakan teknik web scrapping, kemudian lakukan transformasi dari bentuk yang tidak terstruktur ke dalam format data terstruktur. Pada gambar 2 data awal yang didapatkan sebanyak 550 data status pengguna media sosial.

No	Status	Sumber	Kode Status
1	bagaimana cara memenangkan, hati ini yang lagi sakit?	Facebook	Fb_01
2	Pasti aku juara kelas tahun ini.	Facebook	Fb_02
3	moga tahun ini kita menjadi pribadi yang lebih baik lagi, dan apa yang dicita-citakan tercapai amin	Facebook	Fb_03
4	Kasihannya pengemis di jalan yang berjuang hidup	Facebook	Fb_04
5	Semoga mereka kembali ke jalan Allah	Facebook	Fb_05
6	Orang dibwah pasti miskin	Facebook	Fb_06
7	ayo... ngumpul... lama tidak bertemu	Facebook	Fb_07
8	Yang gakak idup kepo itu cuka betanye dulu tope kejadiannya jgn asal menuduh sekali ika jett wang aku lebih dari jett	Facebook	Fb_08
9	Jangan selesaikan masalah dengan mengeluh	Facebook	Fb_09
10	Nha Kenangan 7 tahun yg lalu Rinduuuuuu masa putih abu2, masa2 paling indah di sekolah	Facebook	Fb_10
11	Tetap bersyukur dalam keadaan apapun, karena kebahagiaan berawal dari rasa syukur yang selalu kita panjatkan	Facebook	Fb_11
12	Ayo Perangi Narkoba, Kurangi main game, jauhi maksiat....	Facebook	Fb_12
13	Coba jadi aku sebentar saja	Facebook	Fb_13
14	Walaupun belum hari ini, mungkin esok atau lusa	Facebook	Fb_14
15	indah pada waktunya	Facebook	Fb_15
16	Semoga lelah hari ini menjadi menjadi berkah hari tua	Facebook	Fb_16
17	Alhamdulillah selalu bertemu Dan di kelilingi oleh orang2 yang baik hati...	Facebook	Fb_17

Gambar 2. Data awal

2. Proses pre-processing dan cleansing

Pada tahap ini *pre-processing text* akan mengubah data yang tadinya belum terstruktur menjadi terstruktur sehingga memudahkan dalam melakukan penelitian ini. Adapun tahapan pre-processing :

1. *Tokenizing*, Tahap ini ialah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya untuk menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks.
2. *Filtering*, Pada tahap ini mengambil kata-kata penting disini yang telah dipotong kata per kata dilakukan filter. Pada tahap ini dilakukan pembuangan kata-kata yang di anggap tidak penting seperti tanda hubung dan lainnya.
3. *Steaming*, Pada tahap ini mencari root kata dari tiap hasil *filtering*.
4. *Tagging*, Pada tahap ini mencari bentuk awal/root dari tiap kata lampau atau hasil *Stemming*.

data yang telah di *pre-processing* data tersebut akan dilakukan data *cleaning*, data *cleaning* adalah sebuah fase dimana data-data tidak lengkap, mengandung eror, dan tidak konsisten dibuang dari koleksi data.

3. Proses feature selection dan feature extraction.

Proses Pada tahap *Feature Selection* adalah teknik terpenting dalam *pre-processing*. Teknik ini mengurangi jumlah fitur yang terlibat dalam menentukan suatu nilai kelas target. proses pengambilan data ini adalah data yang sudah melewati proses *text mining* atau data yang sudah terstruktur

No	Status	Sumber	Kode Status
1	cara tenang hati lagi sakit	Facebook	Fb_01
2	Pasti aku juara kelas tahun	Facebook	Fb_02
3	tahun kita menjadi pribadi lebih baik lagi apa cita-cita tercapai amin	Facebook	Fb_03
4	Kasih pengemis jalan berjuang hidup	Facebook	Fb_04
5	Semoga mereka kembali jalan Allah	Facebook	Fb_05
6	Orang pasti miskin	Facebook	Fb_06
7	sabtu	Facebook	Fb_07
8	Itu dulu asal menuduh sekali lebih dari	Facebook	Fb_08
9	Jangan selesaikan masalah dengan mengeluh	Facebook	Fb_09
10	Kenangan 7 tahun lalu Rindu putih paling indah sekolah	Facebook	Fb_10
11	Tetap bersyukur dalam keadaan apapun karena kebahagiaan berawal dari rasa syukur yang selalu kita panjatkan	Facebook	Fb_11
12	Ayo Perangi Narkoba Kurangi main game jauh maksiat	Facebook	Fb_12
13	Coba jadi aku sebentar saja	Facebook	Fb_13
14	Walaupun belum hari ini, mungkin esok atau lusa	Facebook	Fb_14
15	Indah pada waktu	Facebook	Fb_15
16	Semoga lelah hari ini menjadi menjadi berkah hari tua	Facebook	Fb_16
17	Allhamdulillah selalu bertemu dan di kelilingi yang baik hati...	Facebook	Fb_17

Gambar 3. data yang diselection

Setelah data di selecteion, maka data tersebut siap dilakukan extraction, Pada tahap *Feature extraction* ialah salah satu cara untuk mengenali suatu objek tersebut. Tahap inilah yang akan mencocokkan kepribadian dengan data yang didapat

kepribadian	Kata – kata	status	keterangan
Openness	Baru, tenang, apa, sangat, ingin, sekali	Cara tenang hati lagi sakit	Openness
Conscientiousness	Hati, awas, jatuh, serius,		
Extraversion	Hai, halo, kabar, iya,		
Agreeableness	Janji, benar, akan, baik		
Neuroticism	Marah, tidak, jangan, gugup		

Gambar 4. tabel kecocokan kepribadian

Data gambar 3 yang telah dicocokkan pada gambar 4 maka data tersebut sudah di extraction, hasil yang didapatkan seperti gambar 5

No	Status	Kepribadian	Sumber	Kode Status
1	cara tenang hati lagi sakit	Openness	Facebook	Fb_01
2	Pasti aku juara kelas tahun	Openness	Facebook	Fb_02
3	tahun kita menjadi pribadi lebih baik lagi apa cita-cita tercapai amin	Agreeableness	Facebook	Fb_03
4	Kasih pengemis jalan berjuang hidup	Extraversion	Facebook	Fb_04
5	Semoga mereka kembali jalan Allah	Extraversion	Facebook	Fb_05
6	Orang pasti miskin	Extraversion	Facebook	Fb_06
7	sabtu	Agreeableness	Facebook	Fb_07
8	Itu dulu asal menuduh sekali lebih dari	Conscientiousness	Facebook	Fb_08
9	Jangan selesaikan masalah dengan mengeluh	Openness	Facebook	Fb_09
10	Kenangan 7 tahun lalu Rindu putih paling indah sekolah	Extraversion	Facebook	Fb_10
11	Tetap bersyukur dalam keadaan apapun karena kebahagiaan berawal dari rasa syukur yang selalu kita panjatkan	Extraversion	Facebook	Fb_11
12	Ayo Perangi narkoba kurang: main game jauh maksiat	Openness	Facebook	Fb_12
13	Coba jadi aku sebentar saja	Openness	Facebook	Fb_13
14	Walaupun belum hari ini mungkin esok atau lusa	Openness	Facebook	Fb_14
15	Indah pada waktunya	Openness	Facebook	Fb_15
16	Semoga lelah hari ini menjadi menjadi berkah hari tua	Extraversion	Facebook	Fb_16
17	Allhamdulillah selalu bertemu dan di kelilingi yang baik hati	Agreeableness	Facebook	Fb_17

Gambar 5. Data yang di extraction.

IV. HASIL DAN PEMBAHASAN

4.1 Hasil

Adapun proses data mining tersebut menggunakan aplikasi Rapidminer. Proses data mining dilakukan setelah tahapan proses text ming selesai dilaksanakan. Proses pemilihan karakter kepribadian berdasar kan status pengguna media sosial.

1. Implementasi Rapidminer.

Inputkan data yang sudah di extraction kedalam rapid miner.

Row No.	kepribadian	jumlah
1	Agreeableness	62
2	Conscientiousness	18
3	Extraversion	185
4	Neuroticism	33
5	Openness	252
6	Grand Total	550

Gambar 6. Hasil input data di rapid miner

Data yang telah berhasil diinputkan makan tambahkan replace missing value, proses ini bertujuan untuk melihat apakah ada data yang hilang untuk diganti dengan nilai minimum, maksimum atau rata-rata dari atribut itu.

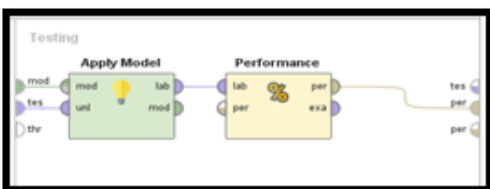
Gambar 7. Hasil dari replace missing value

Gambar 7 menunjukkan bahwa tidak ada data yang missing atau data yang hilang. setelah menambahkan replace missing value maka tambahkan select attributer, cross validation dan terakhir menambahkan algoritma KNN didalam subtraining. Fungsi penambahan algoritma ini sebagai model yang akan dilakukan dalam penelitian ini.



Gambar 8. Data telah di apply model

Maka tahapan terakhir adalah menapply model tersebut untuk dilihat performance nya.



Gambar 9. Data yang telah di performance

Setelah data di performance maka hasil perhitungannya untuk mendapatkan hasil nilai akurasi dan performa adalah

accuracy: 92.92% +/- 2.49% (micro average: 92.91%)						
	true Openess	true Agreeablene...	true Extraversion	true Conscientio...	true Neuroticism	class precision
pred. Openess	239	6	4	1	0	95.60%
pred. Agreeablene...	1	50	0	0	0	98.04%
pred. Extraversion	5	2	173	0	0	95.11%
pred. Conscientio...	1	0	2	16	0	84.21%
pred. Neuroticism	6	4	6	1	33	66.00%
class recall	94.84%	80.65%	93.51%	88.89%	100.00%	

Gambar 10. Hasil perhitungan akurasi

Pada gambar 10 merupakan hasil akurasi yang di dapat dari perhitungan menggunakan tools *Rapidminer*. Dengan hasil yang didapatkan akurasi sebesar 92.92% +/- 2.49.

4.2 Pembahasan

Berdasarkan hasil yang didapat model KNN menunjukkan bahwa tingkat akurasi yang di hasilkan oleh algoritma *K-Nearest Neighbor* memiliki kekuatan yang sangat tinggi. Hal ini dibuktikan dengan hasil akurasi perhitungan mencapai 92.92% dari data sebanyak 550 data, nilai tersebut membuktikan bahwa model yang dibangun dapat digunakan untuk melakukan klasifikasi teks dan ketepatan dalam mengklompokan kepribadian dengan *big five personality*. Pemahaman terhadap macam-macam karakter dasar dapat mendorong setiap individu untuk mengenal dirinya dari dalam, yaitu kekuatan dan kelemahan karakter dasar yang dimiliki.

Diharapkan setelah seseorang mengenal karakter dan memahami jiwanya sendiri, orang tersebut dapat meningkatkan kekuatan kepribadian yang dimilikinya dan belajar menyesuaikan diri dengan orang lain serta bukan berusaha untuk menirukan orang lain. Dengan pemahaman tersebut dapat diketahui bahwa seseorang harus memahami kepribadiannya sendiri dan

berusaha untuk menimbulkan kekuatan atau sisi kepribadian positif yang dimilikinya, terutama bila kepribadian yang dimilikinya merupakan faktor yang penting dan mempunyai pengaruh yang besar dalam menunjang pekerjaan yang sedang digeluti oleh orang tersebut.

Sebagai contoh seorang bekerja, Public Relation maka kepribadian yang dimiliki akan mempengaruhi citra perusahaan karena bertanggung jawab untuk berhadapan langsung dengan publik perusahaan di antaranya media. Seperti yang diungkapkan di atas bahwa kepribadian merupakan perangsang bagi orang lain. Jadi bagaimana orang lain itu bereaksi terhadap seseorang, itulah kepribadian seseorang tersebut. [3]

Kalau orang telah memilih istilah kepribadian, selanjutnya masih perlu ditunjukkan, bahwa istilah psikologi kepribadian lebih tepat dari pada istilah teori kepribadian. Kendatipun sudah jelas, bahwa kepribadian itu mempunyai arti deskripsi, namun masih ada kemungkinan pembuatan deskripsi itu dilakukan dari berbagai sudut pandang ilmu pengetahuan ataupun sudut pandang yang lain. Karena itu supaya lebih jelas mengenai sarannya, yaitu bahwa orang mempersoalkan kepribadian itu dalam arti psikologi, jadi dari sudut pandang psikologi, baiklah secara eksplisit digunakan istilah Psikologi kepribadian. Dimensi *Big Five* pertama kali diperkenalkan oleh Goldberg pada tahun 1981.

Dimensi ini tidak mencerminkan perspektif teori tertentu, tetapi merupakan hasil dari analisis bahasa alami manusia dalam menjelaskan dirinya sendiri dan orang lain. Taksonomi *Big Five* bukan bertujuan untuk mengganti sistem yang terdahulu melainkan sebagai penyatu karena dapat memberikan penjelasan sistem kepribadian secara umum. *Big Five Personality* merupakan pendekatan

dalam psikologi kepribadian yang mengelompokkan trait kepribadian dengan analisis vektor. Bahwa sifat dari kepribadian adalah bipolar dan mengikuti distribusi lonceng yaitu kebanyakan orang akan memiliki skor yang berada dekat dengan titik tengah dari setiap sifat dan hanya sedikit orang yang memiliki skor pada titik ekstrem. [4] Penelitian yang meneliti unit dasar dari kepribadian dengan melakukan pengamatan dan menganalisis kata-kata orang dalam kehidupan sehari-hari untuk menggambarkan kepribadian seseorang. Pada akhirnya dihasilkan *five factors* dari kepribadian yaitu : *Openness* (O), *Conscientiousness* (C), *Extraversion* (E), *Agreeableness* (A) dan *Neuroticism*(N).

1. ***Openness*** (Terbuka terhadap hal-hal baru), Pribadi yang secara konsisten mencari pengalaman-pengalaman yang berbeda dan beragam akan memperoleh skor tinggi.
2. ***Conscientiousness*** (**Sifat berhati-hati**), Orang dengan *conscientiousness* yang tinggi adalah seseorang yang bertanggung jawab, terorganisir, dapat diandalkan dan orang yang gigih. Orang yang mempunyai skor rendah adalah orang yang mudah bimbang, tidak terorganisir dan tidak dapat diandalkan.
3. ***Extraversion*** (**Mudah bersosialisasi**), Dimensi dari *extraversion* berhubungan dengan tingkat kenyamanan dalam sebuah hubungan. Seseorang dengan kepribadian *ekstrovert* cenderung suka berteman, tegas dan ramah, sedangkan *introvert* cenderung pendiam, pemalu dan tenang.
4. ***Agreeableness*** (**Mudah akur atau mudah bersepakat**), Dimensi dari *agreeableness* mengacu pada kecenderungan untuk tunduk kepada orang lain. Orang dengan tingkat *agreeableness* yang tinggi, adalah

orang yang kooperatif, hangat dan penuh kepercayaan, sedangkan seseorang yang mendapat skor rendah, adalah orang yang dingin, tidak mudah patuh dan *antagonistic*.

5. **Neuroticism (Stabilitas emosional)**, Pribadi yang tinggi skor *neuroticism* cenderung mudah menjadi cemas, temperamental, mengasihani diri, emosional, dan rapuh terhadap gangguan yang berkaitan dengan stress. Pribadi yang skornya rendah biasanya tenang, bertemperamen lembut, dan puas diri.

Dibawah ini adalah Nilai akurasi dari kriteria-kriteria kepribadian berdasarkan *big five personality* :

Tabel 1. Tabel Penjelasan Nilai karakter

Kepribadian	Kata – Kata	Nilai akurasi	Jumlah Kata	Nilai class recall
<i>Openness</i>	Baru, apa itu, sangat, ingin, sekali	92,92%	239	94,84%
<i>Conscientiousness</i>	Hati, awas, jatuh, serius,		50	80,65%
<i>Extraversion</i>	Hai, halo, kabar, lagi, iya,		173	93,51%
<i>Agreeableness</i>	Janji, bener, akan, baik,		16	88,89%
<i>Neuroticism</i>	Marah, tidak, jangan, gugup		33	100%

V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dibahas pada bab sebelumnya, maka penulis mencoba menarik kesimpulan dan saran yang mungkin dapat dijadikan pertimbangan bagi pengguna sebagai berikut:

1. Berdasarkan perhitungan data mining menggunakan algoritma K-Nearest Neighbor, dapat ditarik kesimpulan bahwa tingkat akurasi yang diperoleh yaitu 92.92%, dengan nilai karakter *openness* yang paling tinggi yaitu 239 dari 550 data. Dan untuk nilai karakter yang paling rendah yaitu *conscientiousness* dengan nilai 16 dari 550 data.
2. Dari hasil observasi terhadap data status social media *facebook* melalui proses pengklasifikasian *KNN* didapatkan jumlah nilai karakter yang paling sering dibuat oleh masyarakat yaitu *openness*
3. Penentuan data awal dapat mempengaruhi hasil pengujian. Karena data yang diambil harus banyak sehingga dapat mengetahui hasil kepribadian *big five personality* dari tools *RapidMiner*

VI. SARAN

Berdasarkan kesimpulan yang telah diuraikan diatas, maka penulis menyampaikan beberapa saran untuk pengembangan dari Penerapan Text Mining dalam menganalisis kepribadian pengguna media sosial dengan K-Nearest Neighbor (Studi Kasus : Facebook) dibuat guna kebaikan dimasa yang akan datang. Adapun saran yang diberikan penulis sebagai berikut :

1. Masih memiliki kelemahan dalam mencocokkan kepribadian *big five personality*.
2. Dengan adanya penelitian ini diharapkan dapat dimanfaatkan untuk

melihat kepribadian pengguna media sosial berdasarkan *big five personality*.

VII. DAFTAR PUSTAKA

- [1] T. Rismawan, “Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Web,” 2016.
- [2] A. Maulidi, “KESANTUNAN BERBAHASA PADA MEDIA JEJARING SOSIAL FACEBOOK,” hlm. 8.
- [3] A. Aprilia, J. Wijaya, dan Y. Wiyanto, “ANALISA PENGARUH TIPE KEPERIBADIAN DAN GAYA KOMUNIKASI PUBLIC RELATIONS MANAGER HOTEL ”X” SURABAYA DALAM MEMBANGUN HUBUNGAN BAIK DENGAN MEDIA DAN MENINGKATKAN PUBLISITAS,” hlm. 10.
- [4] R. Hardiyanti, “Burnout Ditinjau Dari Big Five Factors Personality Pada Karyawan Kantor Pos Pusat Malang,” vol. 01, hlm. 18, 2013.