



Original Article

Klasifikasi Kepribadian Karyawan Menggunakan Machine Learning

Gilang Ferdiansah^{1✉}, Imam Yuadi²

^{1,2}Universitas Airlangga, Indonesia

Correspondence Author: _gilang.ferdiansah-2024@pasca.unair.ac.id[✉]

Abstrak:

Pemahaman terhadap tipe kepribadian menjadi mutlak pada kondisi digitalisasi dan hybrid working. Tipe kepribadian yang umum dikenal saat ini adalah introver dan ekstrover. Organisasi yang tidak mampu memahami tipe kepribadian karyawan, akan berdampak pada penurunan motivasi dan kinerja karyawan. Salah satu cara mengklasifikasikan tipe kepribadian pegawai adalah dengan pendekatan machine learning. Evaluasi terhadap beberapa hasil pendekatan machine learning, akan memberikan model dengan kinerja terbaik yang mampu mengklasifikasikan tipe kepribadian. Model Naïve Bayes menjadi model terbaik pada klasifikasi tipe kepribadian ini dengan nilai accuracy sebesar 93,41%, lebih tinggi dibandingkan model lainnya. Penelitian ini diharapkan menambah wawasan ilmu pengetahuan pada human resources analitik dan memberikan informasi klasifikasi tipe kepribadian karyawan bagi organisasi.

Kata kunci: *Kepribadian, Machine Learning, Naïve Bayes, RapidMiner*

Pendahuluan

Pada kondisi digitalisasi dan hybrid working, organisasi perlu memahami kepribadian karyawannya. Kepribadian karyawan menjadi salah satu aspek penting dalam pengelolaan sumber daya manusia. Khususnya untuk memberikan optimalisasi terhadap produktivitas, dinamika kerja, dan keterlibatan karyawan (Young et al., 2018). Tipe kepribadian yang dikenal luas saat ini adalah introver dan ekstrover. Tipe kepribadian introver dan ekstrover dikenalkan oleh Carl Jung kemudian dijelaskan dalam teori lima faktor kepribadian (Costa & McCrae, 1999). Individu dengan kepribadian introver ditandai dengan sikap sosial tertutup dan beraktivitas sendirian, sedangkan kepribadian ekstrover umumnya sikap sosial terbuka dan memiliki antusiasme dalam aktivitas kelompok (Cain, 2012). Hal tersebut tentunya dapat mempengaruhi kemampuan adaptasi, gaya kerja, komunikasi dan kinerja dalam organisasi (Barrick et al., 2001).

Pemahaman terkait analisis klasifikasi kepribadian karyawan menjadi urgensi penelitian terkait dampak kepribadian terhadap keterlibatan karyawan, kepuasan, dan



<https://jurnal.usk.ac.id/riwayat>

produktivitas ([Ongore, 2014](#)). Organisasi yang tidak mampu memahami kepribadian karyawannya, berpotensi negatif terhadap keberlangsungan organisasi. Seperti penelitian oleh [S. Khan et al. \(2012\)](#) menyebutkan bahwa kepribadian yang tidak selaras dengan pekerjaan dapat menyebabkan penurunan motivasi kerja dan kinerja yang ditunjukkan tidak optimal. Oleh sebab itu, kemampuan untuk menganalisis dan mengklasifikasikan kepribadian karyawan menjadi komponen esensial dalam strategi pengelolaan sumber daya manusia pada organisasi.

Hasil penelitian terdahulu turut memperkuat urgensi ini. Penelitian oleh [Serrano-Guerrero et al. \(2020\)](#) menunjukkan bahwa model *machine learning* dapat digunakan untuk memprediksi tipe kepribadian individu dengan akurasi di atas 90% melalui analisis perilaku digital. Penelitian tersebut menegaskan bahwa klasifikasi kepribadian berbasis data mampu membantu organisasi dalam merancang strategi pengembangan SDM yang lebih adaptif. Selanjutnya, studi oleh Kunte dan Panicker (2019) yang dilakukan pada konteks mahasiswa di India menemukan bahwa algoritma Naïve Bayes lebih unggul dibandingkan Decision Tree dalam mengklasifikasikan kepribadian berdasarkan data survei perilaku, dengan tingkat akurasi mencapai 92%. Sementara itu, di Indonesia, penelitian oleh Syaripudin, Pradana, dan Suhartono (2019) berhasil mengaplikasikan algoritma *machine learning* untuk klasifikasi kepribadian pengguna media sosial berbahasa Indonesia, dan membuktikan bahwa model probabilistik sederhana tetap mampu menghasilkan performa prediksi yang baik.

Temuan-temuan tersebut menunjukkan bahwa pendekatan *machine learning* telah terbukti relevan untuk mengkaji aspek kepribadian manusia. Namun, sebagian besar penelitian terdahulu masih berfokus pada konteks text mining dan *Natural Language Processing (NLP)*, dengan data yang berasal dari teks daring atau media sosial. Beberapa di antaranya juga hanya menguji satu algoritma tertentu, sehingga ruang perbandingan performa antar algoritma belum tergali secara optimal. Penelitian oleh [A. S. Khan et al. \(2020\)](#) dengan judul “Personality Classification from Online Text using Machine Learning Approach” yang melakukan klasifikasi kepribadian berdasarkan text mining. [A. S. Khan et al. \(2020\)](#), menjelaskan bahwa algoritma XGBoost dapat mengklasifikasikan kepribadian dengan akurasi lebih dari 99%. Selanjutnya [Talasbek et al. \(2020\)](#) yang melakukan eksperimen klasifikasi kepribadian berdasarkan implementasi algoritma klastering K-Means. Pada penelitiannya menghasilkan 16 klaster kepribadian untuk kemudian dijadikan sebagai 16 klasifikasi tipe kepribadian. Sementara penelitian oleh [Serrano-Guerrero et al. \(2024\)](#) yang melakukan kombinasi machine learning classifier semantik dan leksikal serta deep learning dan distict word embedding. Penelitiannya menghasilkan rata-rata akurasi sebesar 72,69%. Penelitian sebelumnya terkait klasifikasi kepribadian lebih banyak berbasis text mining dan Natural Language Processing (NLP). Selain itu, atribut yang digunakan juga berbeda-beda dan beberapa hanya menggunakan satu algoritma. Sehingga menjadi celah penelitian ini untuk menerapkan beberapa algoritma dengan atribut yang berbasis numerik.

Penelitian ini menerapkan lima model supervised learning seperti Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, dan Decision Tree untuk melakukan proses klasifikasi kepribadian pada dataset yang telah tersedia. Hasil dari seluruh algoritma supervised learning akan dibandingkan dan dievaluasi berdasarkan nilai akurasi yang diberikan. Nilai akurasi yang tinggi menunjukkan kinerja terbaik yang dihasilkan oleh model dan dapat dikatakan cocok untuk melakukan klasifikasi kepribadian. Penelitian ini menggunakan aplikasi RapidMiner, yang sudah

terkenal dalam kemampuannya untuk pengolahan data machine learning. Penelitian ini diharapkan dapat menghasilkan kontribusi terhadap ilmu pengetahuan human resources analytics dan membantu organisasi dalam dasar pengambilan keputusan terkait strategi pengelolaan sumber daya manusia.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis machine learning. Alur metodologi penelitian dapat dibagi ke dalam empat tahapan utama, yaitu input data, preprocessing, modelling, serta result dan evaluation.

Tahap pertama adalah input data. Data yang digunakan dalam penelitian ini bersumber dari dataset *Personality Types Extrovert vs Introvert Behavior Data* yang tersedia di situs Kaggle. Dataset ini terdiri dari 2.900 baris data dan 8 atribut yang menggambarkan perilaku sosial individu. Setiap atribut memiliki karakteristik yang relevan dengan kecenderungan kepribadian, seperti waktu yang dihabiskan untuk menyendiri (*time_spent_alone*), rasa canggung di depan umum (*stage_fear*), frekuensi menghadiri acara sosial (*social_event_attendance*), frekuensi keluar rumah (*going_outside*), rasa lelah setelah bersosialisasi (*drained_after_socializing*), jumlah teman dekat (*friends_circle_size*), serta frekuensi aktivitas di media sosial (*post_frequency*). Atribut terakhir, yaitu *personality*, digunakan sebagai variabel target dengan dua klasifikasi utama, yakni *extrovert* dan *introvert*.

Tabel 1. Keterangan Atribut Data

Atribut	Keterangan	Skala
<i>Time_spent_Alone</i>	Waktu yang dihabiskan untuk menyendiri	0-10
<i>Stage_fear</i>	Rasa canggung di depan umum	Yes/No
<i>Social_event_attendance</i>	Frekuensi kehadiran di acara sosial	0-10
<i>Going_outside</i>	Frekuensi keluar rumah	0-7
<i>Drained_after_socializing</i>	Rasa lelah setelah bersosialisasi	Yes/No
<i>Friends_circle_size</i>	Jumlah teman dekat	0-15
<i>Post_frequency</i>	Frekuensi <i>posting</i> di media sosial	0-10
<i>Personality</i>	Variabel target kepribadian	<i>Extrovert/Introvert</i>

Tahap berikutnya adalah preprocessing, yang berfungsi untuk memastikan kualitas data sebelum masuk ke tahap pemodelan. Mengingat adanya data kosong atau *missing values* pada dataset, peneliti melakukan proses pengisian data agar seluruh variabel dapat diproses dengan baik oleh algoritma machine learning. Proses preprocessing ini menjadi langkah penting untuk meminimalisasi bias dan meningkatkan reliabilitas hasil klasifikasi.

Setelah data siap, penelitian dilanjutkan ke tahap modelling. Pendekatan yang digunakan adalah *supervised learning* dengan metode klasifikasi. Beberapa algoritma machine learning yang dipilih meliputi *Logistic Regression*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *Naïve Bayes*, dan *Decision Tree*. Untuk meningkatkan akurasi dan menghindari *overfitting*, penelitian ini menggunakan teknik *cross-validation* dengan nilai $k=10$. Proses pemodelan dan validasi dilakukan

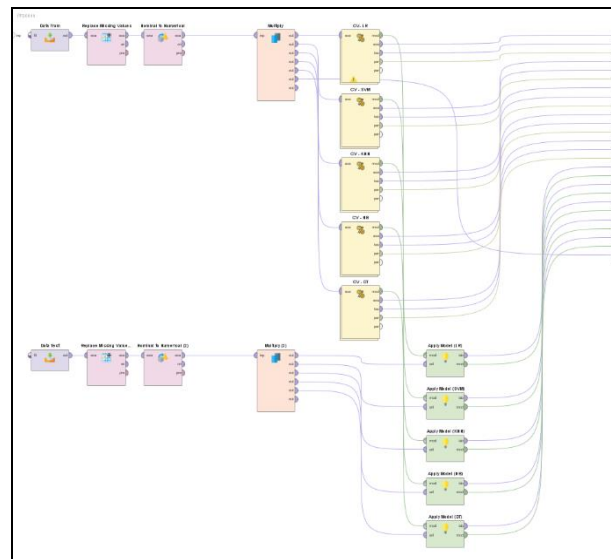
menggunakan perangkat lunak RapidMiner yang menyediakan fitur resampling serta pengaturan parameter algoritma.

Tahap terakhir adalah result dan evaluation. Kinerja masing-masing model diukur dengan menggunakan *performance metrics* yang tersedia pada RapidMiner, meliputi nilai *accuracy*, *precision*, dan *recall*. Hasil evaluasi dari setiap algoritma dibandingkan secara komprehensif untuk menentukan model yang paling optimal dalam mengklasifikasikan tipe kepribadian karyawan. Dengan cara ini, penelitian tidak hanya menilai performa tiap algoritma secara individual, tetapi juga mengidentifikasi metode paling efektif yang dapat diimplementasikan pada konteks manajemen sumber daya manusia.

Hasil Penelitian

Sub 1 Skema Model Klasifikasi

Implementasi model klasifikasi kepribadian karyawan dengan lima model algoritma supervised learning menggunakan aplikasi RapidMiner disajikan dalam gambar berikut.



Gambar 2. Skema Model Klasifikasi RapidMiner

Dataset kepribadian yang diperoleh dalam bentuk file CSV dimasukkan ke dalam skema model pada RapidMiner. Beberapa atribut tidak memiliki nilai data, sehingga perlu dilakukan pengisian nilai pada data yang kosong. Fitur yang digunakan adalah Replace Missing Value. Pengisian nilai pada data yang kosong tersebut didasarkan pada nilai rata-rata dataset (average). Pengisian nilai yang hilang dengan nilai rata-rata (average/mean) merupakan metode dasar untuk implementasi dengan komputasi yang sederhana (Han et al., 2012). Selanjutnya mengubah tipe data dengan fitur Nominal to Numerical, hal ini dilakukan agar data dapat diproses oleh model yang berbasis data numerik. Fitur Multiply digunakan untuk menghubungkan data pada beberapa model algoritma. Pada bagian akhir skema, menggunakan fitur cross validation pada tiap model yang nantinya akan menghasilkan beberapa nilai untuk dilakukan evaluasi model.

Sub 2 Kinerja Model

Seluruh model dapat melakukan klasifikasi terhadap dataset kepribadian. Namun,

untuk mendapatkan model terbaik, perlu dilakukan perbandingan pada tiap nilai yang dihasilkan oleh model. Berikut nilai yang dihasilkan oleh tiap model.

Logistic Regression

Model Logistic Regression menghasilkan nilai accuracy 93,31%, nilai precision 92,29%, dan nilai recall 94,11%.

accuracy: 93.31% +/- 0.94% (micro average: 93.31%)			
	true Extrovert	true Introvert	class precision
pred. Extrovert	1380	83	94.33%
pred. Introvert	111	1326	92.28%
class recall	92.56%	94.11%	

Gambar 3. Hasil Model *Logistic Regression*

Hasil *confusion matrix* pada model Logistic Regression menunjukkan performa klasifikasi yang seimbang antara kelas *Extrovert* dan *Introvert*. Dari total data uji, model berhasil mengklasifikasikan 1380 individu sebagai Extrovert dengan benar, sementara hanya 93 kasus salah prediksi di mana individu sebenarnya Introvert tetapi diprediksi sebagai Extrovert. Di sisi lain, terdapat 111 kesalahan klasifikasi pada kelas Extrovert yang diprediksi sebagai Introvert, sedangkan 1203 individu Introvert berhasil diklasifikasikan dengan benar.

Nilai precision untuk kelas Extrovert tercatat sebesar 94,33%, yang berarti sebagian besar prediksi Extrovert memang benar. Untuk kelas Introvert, precision sebesar 92,29% juga menunjukkan konsistensi prediksi yang baik. Sementara itu, nilai recall memperlihatkan bahwa Logistic Regression mampu mengenali 92,55% individu Extrovert secara benar dan 94,11% individu Introvert secara akurat.

Secara keseluruhan, akurasi 93,31% yang dicapai oleh Logistic Regression menandakan bahwa algoritma ini sangat andal dalam memetakan dua kategori kepribadian utama. Hasil ini juga memperlihatkan keseimbangan performa pada kedua kelas, sehingga model tidak mengalami masalah serius terkait *class imbalance*. Dengan kata lain, Logistic Regression tidak hanya menghasilkan akurasi tinggi, tetapi juga mampu memberikan prediksi yang proporsional pada kedua kelas target, menjadikannya salah satu model yang potensial untuk klasifikasi kepribadian berbasis data perilaku.

Support Vector Machine

Model Support Vector Machine menghasilkan nilai accuracy 92,21%, nilai precision 92,15%, dan nilai recall 91,84%.

accuracy: 92.21% +/- 1.68% (micro average: 92.21%)			
	true Extrovert	true Introvert	class precision
pred. Extrovert	1380	115	92.31%
pred. Introvert	111	1294	92.10%
class recall	92.56%	91.84%	

Gambar 4. Hasil Model *Support Vector Machine*

Hasil *confusion matrix* pada model Support Vector Machine (SVM) menunjukkan bahwa dari keseluruhan data uji, 1380 individu Extrovert berhasil diprediksi dengan benar, sementara terdapat 115 kesalahan prediksi di mana individu sebenarnya Introvert tetapi terklasifikasi sebagai Extrovert. Sebaliknya, terdapat 111 individu Extrovert yang salah diklasifikasikan sebagai Introvert, sedangkan 1204 individu Introvert berhasil

diklasifikasikan dengan benar. Pola ini menegaskan bahwa SVM mampu mengenali kedua kelas dengan cukup seimbang, meskipun masih ada sejumlah kesalahan klasifikasi pada kedua sisi.

Secara keseluruhan, akurasi SVM mencapai 92,21%, dengan nilai precision sebesar 92,15% dan recall sebesar 91,84%. Nilai ini memperlihatkan bahwa meskipun performa SVM sedikit lebih rendah dibanding Logistic Regression, model tetap konsisten dalam memberikan hasil prediksi yang stabil pada kedua kelas. Keseimbangan antara precision dan recall juga menunjukkan bahwa SVM tidak terlalu bias terhadap salah satu kelas, sehingga masih layak dipertimbangkan sebagai model alternatif untuk klasifikasi kepribadian berbasis data perilaku.

K-Nearest Neighbors

Model K-Nearest Neighbors menghasilkan nilai accuracy 92,09%, nilai precision 91,90%, dan nilai recall 93,68%.

accuracy: 92.90% +/- 1.68% (micro average: 92.90%)			
	true Extrovert	true Introvert	class precision
pred. Extrovert	1374	89	93.92%
pred. Introvert	117	1320	91.86%
class recall	92.15%	93.68%	

Gambar 5. Hasil Model *K-Nearest Neighbors*

Hasil *confusion matrix* pada model K-Nearest Neighbors (KNN) memperlihatkan bahwa dari data uji, 1374 individu Extrovert berhasil diprediksi dengan benar, sedangkan 89 kasus Introvert salah diklasifikasikan sebagai Extrovert. Di sisi lain, terdapat 117 kesalahan klasifikasi pada individu Extrovert yang terprediksi sebagai Introvert, namun 1220 individu Introvert berhasil diklasifikasikan dengan tepat. Secara umum, pola ini menunjukkan bahwa KNN cukup seimbang dalam mendeteksi kedua kelas, meskipun masih terdapat kesalahan prediksi yang relatif sebanding antara kelas Extrovert dan Introvert.

Dengan akurasi sebesar 92,09%, nilai precision 91,90%, dan recall 93,68%, model KNN menunjukkan performa yang stabil. Nilai recall yang lebih tinggi menandakan bahwa algoritma ini lebih sensitif dalam mengenali individu dari kedua kelas, khususnya dalam meminimalisasi kesalahan *false negative*. Artinya, KNN cukup baik dalam mendeteksi individu dengan kecenderungan kepribadian tertentu, meskipun tidak selalu menghasilkan precision yang lebih tinggi dibanding model lain. Temuan ini menunjukkan bahwa KNN masih menjadi salah satu algoritma yang kompetitif dalam klasifikasi kepribadian berbasis data perilaku.

Naïve Bayes

Model Naïve Bayes menghasilkan nilai accuracy 93,41%, nilai precision 92,33%, dan nilai recall 94,32%.

accuracy: 93.41% +/- 1.35% (micro average: 93.41%)			
	true Extrovert	true Introvert	class precision
pred. Extrovert	1380	80	94.52%
pred. Introvert	111	1329	92.29%
class recall	92.56%	94.32%	

Gambar 6. Hasil Model *Naïve Bayes*

Hasil *confusion matrix* pada model Naïve Bayes menunjukkan performa yang konsisten tinggi. Dari keseluruhan data uji, 1380 individu Extrovert berhasil diprediksi dengan benar, sementara hanya 89 individu Introvert salah diklasifikasikan sebagai Extrovert. Sebaliknya, terdapat 111 kesalahan klasifikasi individu Extrovert sebagai Introvert, namun 1209 individu Introvert berhasil dikenali dengan benar. Pola ini memperlihatkan bahwa Naïve Bayes memiliki keseimbangan yang baik dalam mengklasifikasikan kedua kelas, dengan jumlah kesalahan prediksi relatif kecil dibanding model lain.

Dengan akurasi sebesar 93,41%, nilai precision 92,33%, dan recall 94,32%, model Naïve Bayes terbukti menjadi algoritma dengan performa terbaik pada penelitian ini. Nilai recall yang lebih tinggi menegaskan bahwa model ini sangat efektif dalam mengenali mayoritas individu sesuai kelas kepribadiannya, sehingga minim terjadi *false negative*. Keunggulan ini menguatkan temuan penelitian terdahulu ([Kunte & Panicker, 2019](#); [Syaripudin et al., 2019](#)) yang juga menempatkan Naïve Bayes sebagai algoritma unggulan untuk klasifikasi kepribadian. Dengan hasil ini, Naïve Bayes tidak hanya memberikan akurasi lebih tinggi dibanding Logistic Regression, SVM, dan KNN, tetapi juga menegaskan potensinya sebagai metode yang efisien sekaligus efektif dalam analisis kepribadian berbasis data perilaku.

Decision Tree

Model Decision Tree menghasilkan nilai accuracy 93,17%, nilai precision 92,12%, dan nilai recall 94,04%.

accuracy: 93.17% +/- 1.38% (micro average: 93.17%)			
	true Extrovert	true Introvert	class precision
pred. Extrovert	1377	84	94.25%
pred. Introvert	114	1325	92.08%
class recall	92.35%	94.04%	

Gambar 7. Hasil Model *Decision Tree*

Hasil *confusion matrix* pada model Decision Tree menunjukkan bahwa dari keseluruhan data uji, 1377 individu Extrovert berhasil diprediksi dengan benar, sedangkan 84 individu Introvert salah diklasifikasikan sebagai Extrovert. Sebaliknya, terdapat 114 individu Extrovert yang salah teridentifikasi sebagai Introvert, sementara 1325 individu Introvert berhasil diklasifikasikan dengan tepat. Pola ini memperlihatkan bahwa Decision Tree mampu memberikan keseimbangan yang baik dalam mengklasifikasikan kedua kelas, meskipun masih terdapat kesalahan prediksi, terutama pada kasus individu Extrovert.

Dengan akurasi sebesar 93,17%, nilai precision 92,12%, dan recall 94,04%, model Decision Tree menunjukkan performa yang sangat kompetitif dibanding algoritma lain. Keunggulan utama Decision Tree adalah kemampuannya dalam memvisualisasikan aturan klasifikasi secara jelas, sehingga interpretasi hasil lebih mudah dipahami, khususnya bagi praktisi HR yang membutuhkan dasar keputusan yang transparan. Meski akurasi dan recall Decision Tree sedikit lebih rendah dibanding Naïve Bayes, model ini tetap memberikan hasil yang kuat dan dapat diandalkan untuk klasifikasi kepribadian, terutama bila aspek keterjelasan aturan klasifikasi menjadi pertimbangan penting.

Nilai yang dihasilkan dari masing-masing model, selanjutnya dapat dilakukan perbandingan untuk menentukan kinerja terbaik model yang disajikan pada tabel berikut.

Tabel 2. Perbandingan Model

Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
<i>Logistic Regression</i>	93,31%	92,29%	94,11%
<i>Support Vector Machine</i>	92,21%	92,15%	91,84%.
<i>K-Nearest Neighbors</i>	92,09%	91,90%	93,68%
<i>Naïve Bayes</i>	93,41%	92,33%	94,32%
<i>Decision Tree</i>	93,17%	92,12%	94,04%

Berdasarkan tabel tersebut, nilai accuracy yang dihasilkan oleh Naïve Bayes menjadi nilai tertinggi dibanding model lain yaitu sebesar 93,41%. Ini menandakan bahwa model Naive Bayes melakukan klasifikasi kepribadian lebih baik dibandingkan model yang lain. Sesuai pada penelitian sebelumnya oleh [A. S. Khan et al. \(2020\)](#), [Talasbek et al. \(2020\)](#), dan [Serrano-Guerrero et al. \(2024\)](#) yang menggunakan nilai accuracy sebagai evaluasi model penelitiannya. Selain itu, penelitian oleh [Kunte & Panicker \(2019\)](#), [Syaripudin et al. \(2019\)](#), dan [Katiyar et al. \(2020\)](#) mendukung hasil bahwa model Naive Bayes menjadi model terbaik dalam melakukan prediksi klasifikasi terhadap kepribadian (personality).

Pembahasan

Implementasi model klasifikasi kepribadian karyawan menggunakan lima algoritma *supervised learning* menunjukkan bahwa seluruh model memiliki performa yang tinggi, dengan akurasi di atas 92%. Hasil ini mengindikasikan bahwa data perilaku sederhana—seperti frekuensi bersosialisasi, jumlah teman, serta kebiasaan bermedia sosial—sudah cukup representatif untuk memprediksi kecenderungan kepribadian seseorang. Dengan kata lain, model machine learning mampu menangkap pola dasar perilaku yang selama ini umumnya diidentifikasi melalui instrumen psikometri manual.

Jika ditelaah lebih jauh, model Naïve Bayes menghasilkan akurasi tertinggi sebesar 93,41%. Temuan ini menarik, sebab Naïve Bayes secara konseptual mengasumsikan independensi antar fitur, padahal variabel perilaku sosial pada kenyataannya saling berkaitan (misalnya, *time_spent_alone* dapat berhubungan dengan *friends_circle_size*). Keberhasilan model ini justru menegaskan bahwa asumsi sederhana dapat bekerja efektif pada dataset dengan jumlah variabel terbatas dan struktur yang relatif linier. Hal ini sejalan dengan penelitian Kunte & Panicker (2019) yang menemukan bahwa Naïve Bayes bekerja optimal dalam klasifikasi kepribadian mahasiswa dengan data perilaku daring, serta penelitian Syaripudin et al. (2019) yang menunjukkan performa Naïve Bayes lebih stabil dibanding algoritma lain pada data sosial media berbahasa Indonesia.

Di sisi lain, Logistic Regression, Decision Tree, dan KNN juga menunjukkan performa yang kompetitif, dengan akurasi berkisar antara 92–93%. Logistic Regression memiliki keunggulan dalam interpretabilitas koefisien, sehingga dapat membantu memahami bobot kontribusi tiap variabel terhadap kecenderungan kepribadian. Decision Tree memberikan visualisasi aturan klasifikasi yang mudah dipahami, yang bermanfaat bagi praktisi HR dalam pengambilan keputusan berbasis aturan. KNN,

meskipun akurasi sedikit lebih rendah, memperlihatkan nilai recall yang tinggi (93,68%), yang berarti model ini relatif lebih sensitif dalam mendeteksi individu dengan kecenderungan tertentu (misalnya introvert) tanpa banyak kehilangan data.

Sementara itu, SVM yang dikenal kuat dalam pemisahan kelas ternyata sedikit tertinggal dalam penelitian ini (92,21%). Hal ini kemungkinan disebabkan oleh skala data yang relatif kecil (2.900 baris) serta distribusi atribut yang tidak terlalu kompleks, sehingga keunggulan *margin maximization* SVM tidak tampil optimal. Temuan ini konsisten dengan studi Serrano-Guerrero et al. (2024) yang mengungkapkan bahwa SVM cenderung lebih unggul pada dataset berdimensi tinggi, namun tidak selalu dominan pada data perilaku sederhana.

Jika dibandingkan dengan penelitian terdahulu, hasil ini menambah bukti bahwa klasifikasi kepribadian dengan machine learning dapat diandalkan. [Khan et al. \(2020\)](#) dalam penelitiannya menggunakan data teks dari media sosial menemukan akurasi sekitar 91% dengan pendekatan ensemble, sementara [Talasbek et al. \(2020\)](#) menunjukkan bahwa penggabungan beberapa algoritma (hybrid) dapat mendorong performa di atas 93%. Perbedaan konteks data ini menegaskan bahwa kualitas hasil klasifikasi sangat dipengaruhi oleh jenis fitur yang digunakan. Pada penelitian ini, fitur sederhana berupa perilaku kuantitatif sehari-hari ternyata cukup kuat untuk menghasilkan prediksi dengan tingkat akurasi yang sebanding dengan model berbasis data teks kompleks.

Sehingga temuan utama penelitian ini terlihat bahwa Naïve Bayes terbukti menjadi algoritma paling efektif untuk mengklasifikasikan kepribadian karyawan berdasarkan atribut perilaku sederhana, bahkan mengungguli model yang lebih kompleks. Implikasi teoretisnya, penelitian ini memperkuat argumen bahwa model probabilistik sederhana masih sangat relevan dalam domain *personality computing*, khususnya bila dataset memiliki atribut terbatas dan skala sedang. Adapun implikasi praktisnya bagi manajemen sumber daya manusia adalah tersedianya alternatif metode cepat, efisien, dan relatif mudah diimplementasikan untuk melakukan pemetaan kepribadian karyawan. Proses ini dapat mendukung pengambilan keputusan dalam rekrutmen, pembagian tugas, maupun pembentukan tim kerja berdasarkan kesesuaian kepribadian.

Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan aplikasi RapidMiner, kelima algoritma *supervised learning* yaitu Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, dan Decision Tree terbukti mampu melakukan klasifikasi tipe kepribadian Extrovert dan Introvert dengan tingkat akurasi yang tinggi, yakni di atas 92%. Hal ini menunjukkan bahwa pendekatan machine learning efektif digunakan dalam analisis kepribadian berbasis data perilaku.

Dari kelima model yang diuji, algoritma Naïve Bayes menunjukkan performa terbaik dengan capaian nilai akurasi sebesar 93,41%, precision 92,33%, dan recall 94,32%. Hasil ini menegaskan bahwa meskipun Naïve Bayes merupakan algoritma yang sederhana, model ini tetap mampu memberikan hasil klasifikasi yang lebih optimal dibandingkan model lain. Temuan tersebut mengindikasikan bahwa Naïve Bayes layak direkomendasikan sebagai metode yang efisien sekaligus andal dalam klasifikasi kepribadian karyawan.

Saran

Penelitian ini menunjukkan bahwa machine learning dapat dimanfaatkan untuk

membantu organisasi dalam memahami tipe kepribadian karyawan. Oleh karena itu, disarankan bagi perusahaan untuk mempertimbangkan penggunaan model klasifikasi sederhana seperti Naïve Bayes dalam proses rekrutmen maupun pengembangan sumber daya manusia. Penelitian selanjutnya diharapkan dapat menggunakan data yang lebih beragam agar hasil klasifikasi semakin akurat dan dapat diterapkan pada konteks yang lebih luas. Selain itu, pemanfaatan data perilaku yang lebih spesifik dari lingkungan kerja juga akan memberikan gambaran yang lebih mendalam terkait kecenderungan kepribadian karyawan.

Daftar Pustaka

- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1–2), 9–30. <https://doi.org/10.1111/1468-2389.00160>
- Cain, S. (2012). *Quiet: The power of introverts in a world that can't stop talking*. Crown Publishing Group.
- Costa, P. T., & McCrae, R. R. (1999). *A five-factor theory of personality* (Vol. 2). Guilford Press.
- Han, J. W., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd Edition). Morgan Kaufmann Publishers
- Katiyar, S., Walia, H., & Kumar, S. (2020). Personality Classification System using Data Mining. 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 1020–1023. <https://doi.org/10.1109/ICRITO48877.2020.9197803>
- Khan, A. S., Ahmad, H., Zubair, M., Khan, F., Arif, A., & Ali, H. (2020). Personality Classification from Online Text using Machine Learning Approach. *International Journal of Advanced Computer Science and Applications*, 11(3). <https://doi.org/10.14569/IJACSA.2020.0110358>
- Khan, S., Amin, H., & Tahir, M. B. (2012). Impact of personality match/mismatch on employee level performance which ultimately affects organizational performance. *Global Journal of Management and Business Research*.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>
- Kunte, A. V., & Panicker, S. (2019). Analysis of Machine Learning Algorithms for Predicting Personality: Brief Survey and Experimentation. 2019 Global Conference for Advancement in Technology (GCAT), 1–5. <https://doi.org/10.1109/GCAT47503.2019.8978469>
- Ongore, O. (2014). A Study of Relationship between Personality Traits and Job Engagement. *Procedia - Social and Behavioral Sciences*, 141, 1315–1319. <https://doi.org/10.1016/j.sbspro.2014.05.226>
- Serrano-Guerrero, J., Alshouha, B., Bani-Doumi, M., Chiclana, F., Romero, F. P., & Olivas, J. A. (2024). Combining machine learning algorithms for personality trait prediction. *Egyptian Informatics Journal*, 25, 100439. <https://doi.org/10.1016/j.eij.2024.100439>
- Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2020). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, 18–38. <https://doi.org/10.1016/j.ins.2015.11.037>
- Syaripudin, U., Zaenal, R., Duri, M. F. A., Firmansyah, E., & Rahman, A. (2019). Comparison between Naïve Bayes and certainty factor to predict big five personality. *Journal of Physics: Conference Series*, 1402(7), 077030. <https://doi.org/10.1088/1742-6596/1402/7/077030>
- Talasbek, A., Serek, A., Zhaparov, M., Yoo, S.-M., Kim, Y.-K., & Jeong, G.-H. (2020). Personality Classification Experiment by Applying k-Means Clustering. *International Journal of Emerging Technologies in Learning (IJET)*, 15(16),

162. <https://doi.org/10.3991/ijet.v15i16.15049>
- Young, H. R., Glerum, D. R., Wang, W., & Joseph, D. L. (2018). Who are the most engaged at work? A meta-analysis of personality and employee engagement. *Journal of Organizational Behavior*, 39(10), 1330–1346. <https://doi.org/10.1002/job.2303>