

UTS 1
KECERDASAN BISNIS B
Analisis Prediktif pada Data HR Competency Scores for Screening

DOSEN PROGRAM STUDI

Ir. Paulina Heruningsih Prima Rosa, S.Si., M.Sc.



DISUSUN OLEH

- | | |
|--------------------------------|--------------------|
| 1. Resiana Kinanti Jati | (215314159) |
| 2. Loadtriani Oktavia S | (215314172) |

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS SANATA DHARMA
YOGYAKARTA

2024

A. LATAR BELAKANG

Rekrutmen dan seleksi kandidat calon pekerja memainkan peran penting dalam menjamin keberhasilan suatu organisasi atau perusahaan. Proses penyaringan awal yang efektif dapat secara signifikan meningkatkan kualitas kumpulan kandidat dan meningkatkan peluang untuk menemukan kandidat yang tepat (Zaky, 2021). Pada praktiknya, proses rekrutmen membutuhkan kemampuan yang jeli dalam mengidentifikasi dan memilih kandidat dengan kompetensi yang sesuai dengan kebutuhan posisi. Salah satu masalah dalam mengidentifikasi kandidat adalah efektifitas penilaian yang digunakan untuk memperoleh skor kompetensi yang dilakukan oleh departemen *Human Resources* (HR).

Untuk menganalisis data penilaian apa saja yang diperlukan dalam proses seleksi, usaha yang dapat dilakukan yaitu dengan observasi pada beberapa perusahaan terkhusus departemen HR. Namun, kami sebagai penyusun menemukan sebuah *dataset* pada laman *platform* berbagi data yaitu Kaggle yang dapat digunakan untuk analisis pada makalah ini.

Masalah efektifitas penggunaan data skor kompetensi ini tidak hanya berdampak pada departemen HR, tetapi juga berpengaruh pada keseluruhan struktur organisasi perusahaan. Pemimpin atau manajer dari berbagai departemen turut mengandalkan karyawan baru yang memenuhi kualifikasi untuk posisi atau kebutuhan yang diharapkan.

Dalam dunia *data science*, ilmu kecerdasan bisnis (BI) dapat membantu analisis permasalahan ini, yang mana memanfaatkan instrumen analisis prediktif sehingga perusahaan dapat mengembangkan model prediktif untuk memproyeksikan keberhasilan kandidat berdasarkan skor kompetensi mereka (Bogen & Rieke, 2018). Ini juga membantu dalam mengidentifikasi data yang paling berpengaruh dalam penilaian.

B. RUMUSAN MASALAH

Berdasarkan latar belakang di atas, dirumuskan permasalahan yang akan diselesaikan dalam makalah ini yaitu:

1. Apa faktor-faktor yang paling signifikan dalam menentukan kelayakan kandidat untuk wawancara?
2. Apa metode/ algoritma yang menghasilkan akurasi terbaik untuk memprediksi dan menjelaskan kemungkinan kandidat dipanggil untuk wawancara?

C. TUJUAN

Untuk menjawab rumusan masalah di atas, maka dapat dijabarkan tujuan sebagai berikut:

1. Mengetahui faktor-faktor yang paling signifikan dalam menentukan kelayakan kandidat untuk wawancara.
2. Mengetahui algoritma yang tepat untuk memprediksi dan menjelaskan kemungkinan kandidat dipanggil untuk wawancara.

D. PEMBAHASAN

Pada data yang telah didapatkan, kami menggunakan konsep model prediktif. Model Prediktif adalah representasi matematis atau statistik yang membantu kita memprediksi nilai yang mungkin dari suatu data berdasarkan informasi yang kita miliki. Ini membantu kita dalam memahami hubungan antara berbagai faktor (yang disebut fitur) dan hasil yang kita inginkan (yang disebut target). Berikut adalah beberapa metode penelitian yang umumnya digunakan dalam pengembangan model prediktif :

1. Pencarian Data : Data di peroleh dari Kaggle dengan judul HR Competency Scores for Screening
2. Pemahaman Data : memahami data yaitu atribut-atribut beserta maknanya. Data yang diperoleh terdiri dari 10 atribut dan dijelaskan dalam tabel berikut di bawah ini.

No	Atribut	Penjelasan
1	years_of_experience	Jumlah tahun pengalaman yang dimiliki oleh kandidat di bidangnya
2	functional_competency_score	Skor yang mewakili kompetensi fungsional kandidat berdasarkan tes
3	top1_skills_score	Skor keterampilan paling berharga yang dimiliki kandidat
4	top2_skills_score	Skor keterampilan berharga kedua yang dimiliki kandidat
5	top3_skills_score	Skor keterampilan paling berharga ketiga yang dimiliki kandidat
6	behavior_competency_score	Skor yang mewakili kompetensi perilaku kandidat yang dimiliki kandidat
7	top1_behavior_skill_score	Skor keterampilan perilaku paling berharga yang dimiliki kandidat
8	top2_behavior_skill_score	Skor keterampilan perilaku paling berharga kedua yang dimiliki kandidat
9	top3_behavior_skill_score	Skor keterampilan perilaku paling berharga ketiga yang dimiliki kandidat

10	call_for_interview	Variabel biner yang menunjukkan apakah kandidat untuk wawancara atau tidak
----	--------------------	--

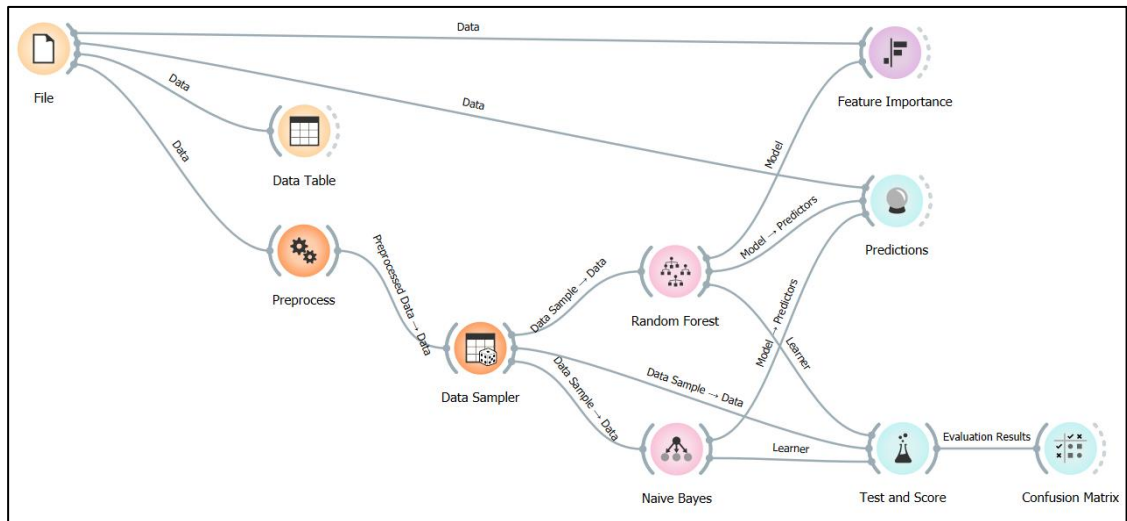
3. Preprocessing Data : Setelah mendapatkan data, langkah selanjutnya adalah melakukan preprocessing untuk menangani nilai yang hilang dan data yang duplikat. Langkah ini penting untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang unggul dan siap untuk digunakan dalam pembuatan model prediktif.
4. Pelatihan Model ; Setelah dilakukan preprocessing pada data, selanjutnya dilakukan pelatihan model yang akan digunakan. Pada data ini, kami menggunakan 2 metode klasifikasi yaitu Random Forest dan Naive Bayes. Random Forest pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001, Random Forest adalah teknik yang banyak digunakan dalam klasifikasi. Dalam teknik ini, klasifikasinya berasal dari beberapa pohon klasifikasi yang digabungkan. Metode ini memiliki reputasi yang bagus di kalangan ahli statistik dan pembelajaran mesin karena kemampuannya menghasilkan hasil klasifikasi yang akurat untuk berbagai jenis data. Random forest terutama cocok digunakan dalam situasi di mana terdapat banyak fitur dalam data, yang semakin sering terjadi karena pengumpulan dan penyimpanan data yang semakin mudah (Amaratunga et al., 2008). Sedangkan Naive Bayes adalah metode yang paling sederhana di antara model-model lainnya. Ia menganggap bahwa semua atribut contoh saling bebas satu sama lain jika diberikan kelasnya. Meskipun asumsi ini tidak sepenuhnya akurat dalam kehidupan nyata, Naive Bayes sering memberikan hasil klasifikasi yang baik (Mccallum & Nigam, 1998) Kedua metode ini akan digunakan dalam pelatihan model untuk melakukan klasifikasi pada data yang telah dipersiapkan sebelumnya.
5. Evaluasi Model : Setelah melatih model-model Random Forest dan Naive Bayes, langkah selanjutnya adalah mengevaluasi kinerjanya. Evaluasi dilakukan dengan menghitung sejumlah metrik evaluasi, seperti *Area Under ROC Curve* (AUC), Klasifikasi Akurasi, *F1 Score*, *Presisi*, *Recall*, dan Koefisien Korelasi Matthews untuk kedua metode. Matrik-matrik ini memberikan gambaran menyeluruh tentang seberapa baik model-model tersebut dapat melakukan prediksi dan klasifikasi data yang diberikan. Dengan demikian, evaluasi ini memungkinkan perbandingan antara kinerja model-

model tersebut dan pemilihan model terbaik untuk analisis data yang sedang dilakukan.

Sebagaimana telah dijabarkan pada tabel penjelasan atribut di atas, data yang diperlukan dalam analisis ini adalah data yang berisikan atribut-atribut untuk penilaian kompetensi beserta skornya. Selanjutnya, akan disajikan contoh data yang diperoleh dari laman *platform* Kaggle dengan link *url* <https://www.kaggle.com/datasets/muhammadjawwadismail/hr-competency-scores-for-screening> dalam tabel berikut.

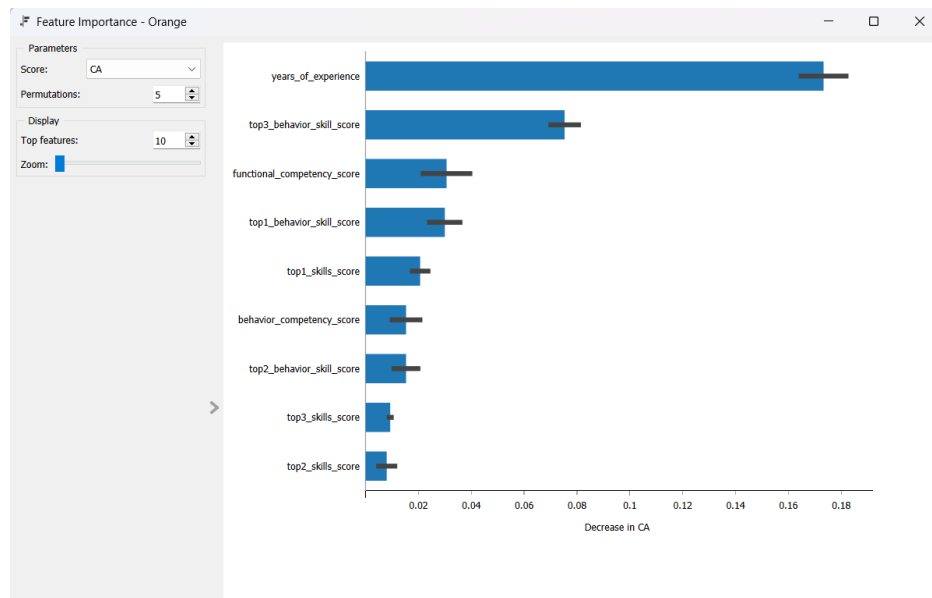
No	Years of experience	Functional competency score	top1 skills score	top2 skills score	top3 skills score	Behavior competency score	top1 behavior skill score	top2 behavior skill score	top3 behavior skill score	Call for interview
1	0.333333	0.749498	0.777778	0.7	0.8	0.769231	0.931689	0.662042	0.752463	1
2	0.133333	0.490638	0.777778	0.3	0.4	0.153846	0.352093	0.132231	0.316905	0
3	0	0.460256	0.444444	0.3	0.3	0.051282	0.052347	0.089765	0.254859	0
4	0	0.507347	0.555556	0.4	0.5	0.384615	0.309913	0.289758	0.440784	0
5	0.4	0.66202	0.666667	0.7	0.5	0.461538	0.497929	0.397544	0.39276	1
6	0.4	0.66202	0.666666667	0.7	0.5	0.461538462	0.497928738	0.39754422	0.392759525	1
7	0.8	0.585157778	1	0.8	0.8	0.871794872	0.759200003	0.821151425	0.721994076	1
8	0.333333333	0.691946667	0.888888889	0.6	0.5	0.461538462	0.495587553	0.584719615	0.440508602	1
9	0	0.477886667	0.333333333	0.5	0.5	0.512820513	0.365087223	0.48169207	0.330779931	0
10	0.266666667	0.662735556	0.888888889	0.7	0.6	0.538461538	0.482059551	0.601124041	0.647886671	1

Setelah menerapkan langkah metodologis, selanjutnya disajikan penyajian hasil luaran dari olah data dengan mempresentasikan hasil prediktif yang dihasilkan dengan menggunakan software Orange.

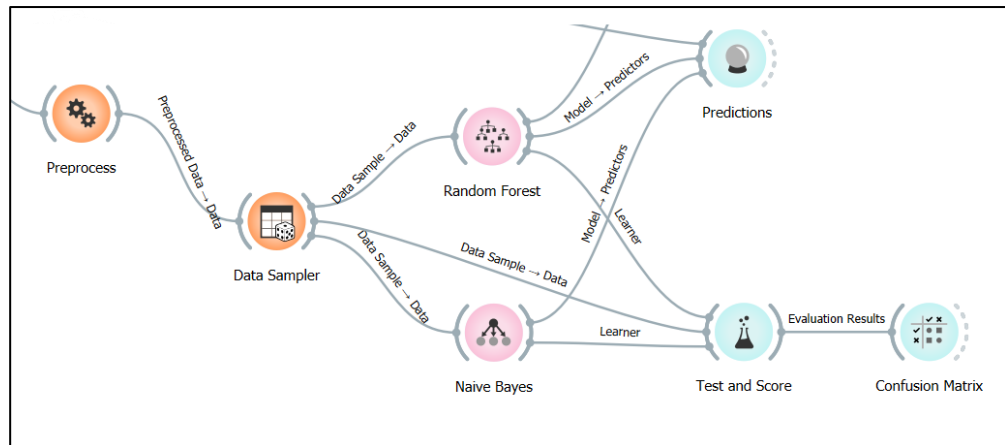


Gambar di atas merupakan bagan alur proses dalam software Orange dan berhasil dijalankan. Hasil analisis prediktif akan dipaparkan dalam poin-poin berikut:

- Mencari faktor-faktor apa saja yang paling signifikan dalam penentuan kandidat dengan menggunakan widget Feature Importance (atribut fitur yang penting).



- Memperoleh algoritma yang tepat untuk memprediksi dan menjelaskan kemungkinan kandidat dipanggil untuk wawancara.



- a. Test and Score untuk mendapatkan algoritma pemodelan yang menghasilkan akurasi terbaik sebelum proses prediksi.

Test and Score - Orange

Cross validation
 Number of folds: 5
☒ Stratified
☐ Cross validation by feature
☐ Random sampling
 Repeat train/test: 10
 Training set size: 70 %
☒ Stratified
☐ Leave one out

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.959	0.921	0.921	0.925	0.921	0.846
Naive Bayes	0.963	0.892	0.891	0.912	0.892	0.804

Confusion Matrix untuk masing-masing algoritma:

Confusion Matrix - Orange

Learners
 Random Forest
 Naive Bayes

Output
☒ Predictions
☐ Probabilities
☒ Apply Automatically

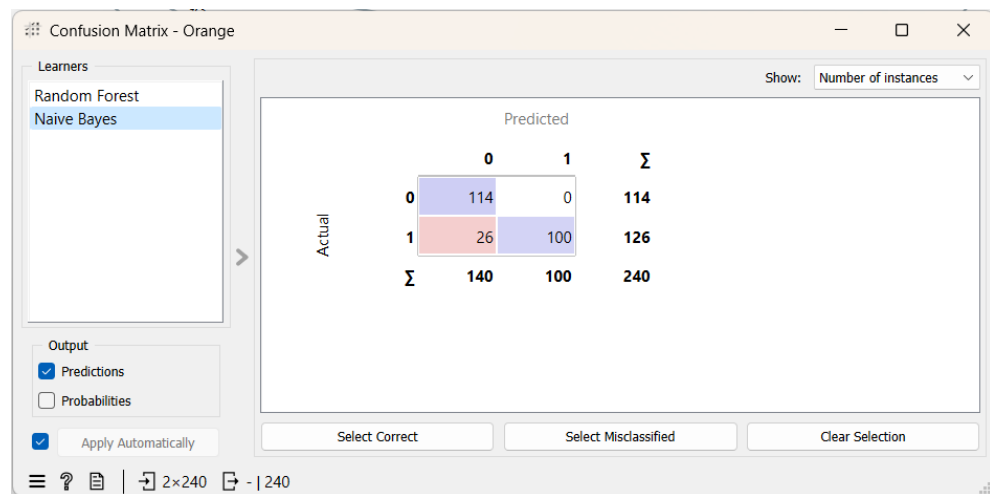
Show: Number of instances

		Predicted		Σ
		0	1	
Actual	0	110	4	114
	1	15	111	126
Σ		125	115	240

Select Correct Select Misclassified Clear Selection

- True Negatives (TN): 110 contoh diprediksi dengan benar sebagai kelas 0.
- False Negatives (FN): 15 contoh salah diprediksi sebagai kelas 0 padahal sebenarnya kelas 1.
- False Positives (FP): 4 contoh salah diprediksi sebagai kelas 1 padahal sebenarnya kelas 0.

- True Positives (TP): 111 contoh diprediksi dengan benar sebagai kelas 1.



- True Negatives (TN): 114 contoh diprediksi dengan benar sebagai kelas 0.
- False Negatives (FN): 26 contoh yang diprediksi secara salah sebagai kelas 0 padahal sebenarnya kelas 1.
- False Positives (FP): 0 contoh salah diprediksi sebagai kelas 1 padahal sebenarnya kelas 0.
- True Positives (TP): 100 contoh diprediksi dengan benar sebagai kelas 1.

b. Prediction

Predictions - Orange

Show probabilities for: Classes in data ☒ Show classification errors

Restore Original Order

	Random Forest	error	Naive Bayes	error	call_for_interview	ears_of_experience	onal_competency	top1_skills_score	top2_skills
1	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.333333	0.749498	0.777778	0.7
2	0.71 : 0.29 → 0	0.290	1.00 : 0.00 → 0	0.000	0	0.133333	0.490638	0.777778	0.3
3	0.80 : 0.20 → 0	0.200	1.00 : 0.00 → 0	0.000	0	0	0.460256	0.444444	0.3
4	0.73 : 0.27 → 0	0.267	1.00 : 0.00 → 0	0.002	0	0	0.507347	0.555556	0.4
5	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.4	0.66202	0.666667	0.7
6	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.8	0.585158	1	0.8
7	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.333333	0.691947	0.888889	0.6
8	0.88 : 0.12 → 0	0.120	1.00 : 0.00 → 0	0.000	0	0	0.477887	0.333333	0.5
9	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.266667	0.662736	0.888889	0.7
10	0.93 : 0.07 → 0	0.070	1.00 : 0.00 → 0	0.000	0	0.0666667	0.492876	0.888889	0.3
11	0.95 : 0.05 → 0	0.050	1.00 : 0.00 → 0	0.000	0	0.0666667	0.546536	0.222222	0.1
12	1.00 : 0.00 → 0	0.000	1.00 : 0.00 → 0	0.000	0	0	0.501478	0.333333	0.4
13	0.95 : 0.05 → 0	0.050	1.00 : 0.00 → 0	0.000	0	0	0.591209	0.666667	0.5
14	0.54 : 0.46 → 0	0.544	1.00 : 0.00 → 0	1.000	1	0.2	0.457838	0.666667	0.5
15	0.77 : 0.23 → 0	0.230	1.00 : 0.00 → 0	0.000	0	0.2	0.345198	0.666667	0.4
16	0.55 : 0.45 → 0	0.550	1.00 : 0.00 → 0	1.000	1	0	0.577047	0.333333	0.5
17	1.00 : 0.00 → 0	0.000	1.00 : 0.00 → 0	0.000	0	0	0.479309	0.666667	0.3
18	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.466667	0.9071	1	0.7
19	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.8	0.726576	1	0.8
20	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.866667	0.695636	1	1.0
21	0.02 : 0.98 → 1	0.020	0.61 : 0.39 → 0	0.610	1	0.4	0.222222	0.888889	0.5
22	0.48 : 0.52 → 1	0.516	1.00 : 0.00 → 0	0.000	0	0.2	0.41524	0.666667	0.3
23	1.00 : 0.00 → 0	0.000	1.00 : 0.00 → 0	0.000	0	0	0.472764	0.444444	0.4
24	0.98 : 0.02 → 0	0.017	1.00 : 0.00 → 0	0.000	0	0.133333	0.420658	0.666667	0.6
25	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.001	1	0.266667	0.66708	0.888889	0.5
26	0.00 : 1.00 → 1	0.000	0.01 : 0.99 → 1	0.007	1	0.466667	0	1	0.7
27	0.00 : 1.00 → 1	0.000	0.00 : 1.00 → 1	0.000	1	0.266667	0.72954	0.888889	0.7

Hasil performa dari alur prediksi adalah sebagai berikut:

Show performance scores						
Target class: (Average over classes)						
Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.991	0.977	0.977	0.977	0.977	0.953
Naive Bayes	0.968	0.907	0.906	0.920	0.907	0.827

Hasil luaran ini bermanfaat bagi beberapa pihak antara lain staf pekerja departemen HR yang akan melakukan rekrutmen dan pihak lain dalam suatu perusahaan yang terlibat untuk mengoptimalkan atau upaya efisiensi proses seleksi sehingga data kompetensi yang digunakan tepat sasaran. Para calon kandidat pekerja (pelamar) turut mendapat manfaat dari hasil luaran ini untuk memperoleh informasi kompetensi apa saja yang perlu dipersiapkan secara maksimal agar dapat memiliki peluang diterima dalam proses rekrutmen yang besar.

E. KESIMPULAN DAN SARAN

- Faktor yang paling signifikan mempengaruhi proses rekrutmen dan seleksi apabila diranking dalam 3 urutan teratas yaitu : years of experience, top3 behaviour skill score, behaviour competency score. Hal ini dapat ditarik kesimpulan bahwa para calon kandidat pekerja perlu mempersiapkan pengalaman kerja, keterampilan (*skill*), dan wawasan perilaku (*behaviour*) agar nantinya dapat memperoleh skor kompetensi yang maksimal dalam 3 faktor tersebut.
- Hasil akurasi pemodelan untuk dataset HR Competency Scores for Screening dengan Random Forest adalah 92% sementara dengan Naive Bayes adalah 89%.
- Model Naive Bayes memiliki skor kinerja lebih rendah dibandingkan dengan model Random Forest. Maka algoritma Random Forest merupakan algoritma yang tepat untuk analisis prediktif dataset yang kami gunakan.

Dalam penelitian atau analisa selanjutnya yang mungkin juga terkait dengan analisis prediktif HR Competency Scores for Screening, penggunaan *hyperparameter* dapat dilakukan untuk meningkatkan akurasi pemodelan data. Keterbatasan pada *software* Orange untuk melakukan *hyperparameter* menjadi kendala bagi kami. Oleh karena itu, eksplorasi dengan menggunakan *software* pemrograman maupun olah data lainnya tentu dapat dilakukan untuk memenuhi kebutuhan penelitian.

F. REFERENSI

Amaratunga, D., Cabrera, J., & Lee, Y. S. (2008). Enriched random forests. *Bioinformatics*, 24(18), 2010–2014.
<https://doi.org/10.1093/bioinformatics/btn356>

Bogen, M., & Rieke, A. (2018). *An Examination of Hiring Algorithms, Equity, and Bias HELP WANTED*.

Mccallum, A., & Nigam, K. (1998). *A Comparison of Event Models for Naïve Bayes Text Classification*.

Zaky, M. (2021). Penerapan Teknologi Informasi dalam Proses Rekrutmen dan Seleksi Sumber Daya Manusia. *KOMITMEN: Jurnal Ilmiah Manajemen*, 2(2).