

## Prediksi Pengunduran Diri Karyawan Perusahaan “Y” Menggunakan Random Forest

Daniel Dwi Eryanto Manurung<sup>1</sup>, Fachran Sandi<sup>2</sup>, Fajar Akbardipura<sup>3</sup>, Hashfi Ashfahan<sup>4</sup>,  
Desta Sandya Prasvita<sup>5</sup>  
Informatika

Universitas Pembangunan Nasional Veteran Jakarta  
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok,  
Daerah Khusus Ibukota Jakarta 12450  
danielDEM@upnvj.ac.id<sup>1</sup>, fachrans@upnvj.ac.id<sup>2</sup>, fajarakbardipura@upnvj.ac.id<sup>3</sup>,  
hashfiashfahan@upnvj.ac.id<sup>4</sup>, desta.sandya@upnvj.ac.id<sup>5</sup>

**Abstrak.** Kemajuan teknologi mempengaruhi beberapa bidang, terutama dalam bidang manajemen SDM. Teknologi tersebut merubah cara departemen SDM beroperasi, sehingga dapat mengarahkan pada tingkat efisiensi yang lebih tinggi dan juga hasil yang lebih baik. Jumlah karyawan mengundurkan diri yang tinggi pada perusahaan dapat menimbulkan kemunduran bisnis, oleh karena itu penting untuk perusahaan mengetahui alasan karyawan resign dari perusahaan. Dalam penelitian ini, perusahaan dapat memprediksi pengunduran diri karyawan. Hal ini penting dilakukan bagi HRD serta manajer untuk memahami apa yang biasanya menjadi faktor *resign* seorang karyawan dan melakukan upaya pencegahan berdasarkan alasan tersebut. Pada penelitian ini menggunakan data karyawan fiktif yang dimiliki oleh perusahaan. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, pengolahan menggunakan algoritma penelitian, analisis data untuk mencari korelasi antar atribut, dan interpretasi LIME yang dapat membantu untuk memahami secara individual penyebab karyawan mengundurkan diri. Pada penelitian ini menggunakan metode *Random Forest*, hasil akurasi yang didapatkan sebesar 87% dan *error* sebesar 13%.

**Kata Kunci:** SDM, Random Forest, Pengunduran Diri.

### 1 Pendahuluan

Kemajuan teknologi mempengaruhi beberapa bidang, terutama dalam bidang manajemen SDM. Teknologi tersebut merubah cara departemen SDM beroperasi, sehingga dapat mengarahkan pada tingkat efisiensi yang lebih tinggi dan juga hasil yang lebih baik. Bidang manajemen SDM harus dapat menyesuaikan kemajuan teknologi tersebut supaya tidak tertinggal, hal ini dapat dilakukan dengan mengoptimalkan penggunaan teknologi dalam pengelolaan SDM beserta alat-alat lainnya yang dapat membantu untuk menganalisis permasalahan yang dihadapi oleh bidang manajemen SDM [5]. Keputusan dalam penggunaan teknologi data mining atau big data dapat dijadikan suatu pertimbangan dalam bidang manajemen SDM. Menggabungkan data mining dan teknik analisis prediktif yang canggih dipastikan akan sangat meningkatkan kinerja dari manajemen SDM itu sendiri. Analisis yang dilakukan pada data SDM dapat berupa deskriptif analitik, prediktif dan korelasi yang digabungkan dengan teknik machine learning untuk pengambilan sebuah keputusan [4].

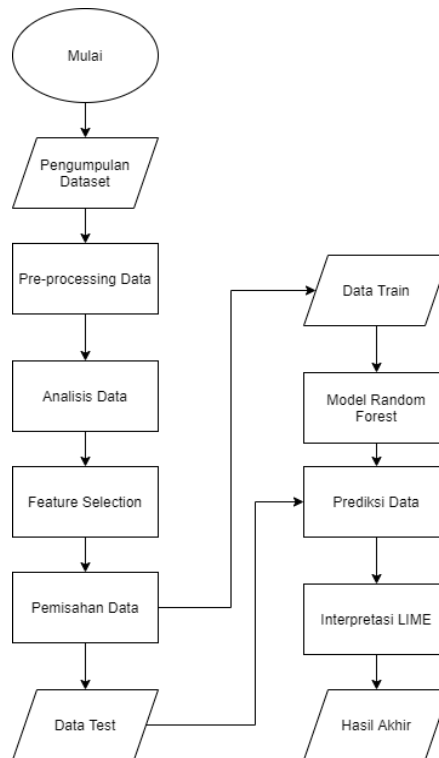
Jumlah karyawan mengundurkan diri yang tinggi pada perusahaan dapat menimbulkan kemunduran bisnis, oleh karena itu penting untuk perusahaan mengetahui alasan karyawan resign dari perusahaan. Dalam penelitian ini, perusahaan dapat memprediksi pengunduran diri karyawan. Hal ini penting dilakukan bagi HRD serta manajer untuk memahami apa yang biasanya menjadi faktor resign seorang karyawan dan melakukan upaya improvement berdasarkan alasan tersebut [7]. Dengan mengetahui faktor-faktor pengunduran diri karyawan, perusahaan dapat mencegah terjadinya pengunduran diri karyawannya, sehingga perusahaan dapat meminimalisir pengeluaran UP (Uang Pesangon) yang diberikan dan juga tidak perlu kesulitan dalam mencari karyawan baru yang harus dilatih ulang dari awal.

Beberapa penelitian lain yang telah dilakukan, seperti penelitian tentang analisis kemungkinan pengunduran diri calon mahasiswa baru. Penelitian tersebut menggunakan *decision tree* sebagai metode untuk prediksi dan hasil dari penelitian tersebut didapatkan akurasi 94,29%, kemudian menggunakan aplikasi rapid miner dengan akurasi 86.70%, selanjutnya membandingkan *decision tree* dan ID3 dengan akurasi ID3 lebih unggul yaitu sebesar 84.5% [1]. Sedangkan penelitian lainnya yang dilakukan tentang prediksi pengunduran diri mahasiswa Universitas

Amikom Yogyakarta, penelitian tersebut menggunakan metode *Naïve Bayes* dan hasil dari penelitian tersebut memiliki akurasi sebesar 77,78% dengan tingkat error sebesar 22,22% [3].

## 2 Metode Penelitian

Dalam penelitian ini dilakukan eksperimen dengan menggunakan data karyawan fiktif yang dimiliki oleh perusahaan. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, pengolahan menggunakan algoritma penelitian, dan analisis data untuk mencari korelasi antar atribut untuk menentukan atribut yang paling berpengaruh dalam penelitian ini. Berikut adalah diagram alir dari penelitian yang disajikan pada Gambar 1.



**Gambar. 1.** Flowchart Metode Penelitian

### 2.1 Pengumpulan Data

Pada tahapan ini, kami mengumpulkan data yang dibutuhkan untuk menjalankan penelitian ini. Data tersebut kami peroleh dari situs Kaggle. Data yang digunakan merupakan data fiktif, tetapi data tersebut memiliki kemiripan dengan data yang ada pada bagian SDM di perusahaan-perusahaan.

### 2.2 Pra-proses Data

Sebelum data kami gunakan, kami melakukan pra-proses untuk menghilangkan noise pada dataset dan juga melakukan penambahan fitur. Dalam data yang kami gunakan terdapat noise seperti duplikat data, tipe data yang tidak sesuai dan *missing value*. Sehingga kami perlu melakukan pre-proses dengan melakukan penghapusan data yang duplikat, mengubah tipe data, menghapus data yang bernilai *NaN* dan menambahkan 2 fitur.

### 2.3 Analisis Data

Langkah ini merupakan langkah yang cukup penting dalam melakukan penelitian yang berkaitan dengan *machine learning* atau *big data*. Pada tahapan ini mengeksplor data lebih dalam untuk melihat apakah terdapat *outlier* dan korelasi antara variabel tertentu, kemudian memvisualisasikan data tersebut.

## 2.4 Feature Selection

Pada tahapan ini, kami melakukan *feature selection*, pada tahapan ini kami menggunakan metode PCA di mana PCA ini berfungsi untuk mengelompokkan data berdasarkan pengamatan antar variabel yang berkorelasi. Dalam penelitian ini PCA membantu untuk mengurangi dimensi untuk meringkas variabel *numerik*.

## 2.5 Pemisahan Data

Data yang sudah di bersihkan atau di pra-proses lalu dilakukan *feature selection* guna mencari variabel yang digunakan. Dibagi menjadi 2 data yaitu data *training* dan data *testing* sebesar 80% dan 20% dari 3310 data. Data tersebut tidak langsung digunakan dalam pemodelan, kita lihat terlebih dahulu apakah data *train* yang akan digunakan dalam pemodelan sudah *balance* atau belum, jika belum maka perlu dilakukan *upsampling* atau *downsampling* [6]. Untuk pemodelan ini menggunakan *downsampling* untuk menyeimbangkan data *train* tersebut.

## 2.6 Model Random Forest

Algoritma *random forest* adalah salah satu algoritma *supervised learning*. Ini adalah teknik pembelajaran *ensemble* yang berdasarkan pada algoritma *decision tree*. Teknik yang terintegrasi ini menggabungkan prediksi beberapa estimator dasar yang dibangun menggunakan algoritma *decision tree* untuk meningkatkan efisiensi estimator tunggal [2]. *Random forest* menumbuhkan banyak pohon klasifikasi, yang disebut *Forest*(hutan). Jika kita ingin mengklasifikasikan data baru, setiap pohon akan memiliki prediksi kelasnya sebagai pemungutan suara [2]. *Random forest* ini akan memilih kategori dengan suara terbanyak. Secara umum, semakin banyak pohon di hutan acak, semakin tinggi hasil akurasi.

## 2.7 Prediksi Data

Pada tahap sebelumnya model *random forest* sudah dilatih terlebih dahulu sehingga selanjutnya kita langsung masuk ke dalam implementasi model menggunakan data *test*. Hasil yang keluar berupa prediksi dari model yang sudah dilatih. Kemudian menggunakan *confusion matrix* untuk mengevaluasi model tersebut.

## 2.8 Interpretasi LIME

Pada tahapan ini kita menggunakan LIME yang merupakan singkatan dari *Local Interpretable Model-Agnostic Explanations*. Biasanya, komputer tidak menjelaskan prediksi mereka dan hanya menampilkan angka dari prediksi tersebut contohnya seperti akurasi dan *recall*. LIME membantu kita untuk memahami secara individual penyebab karyawan mengundurkan diri dan juga menjelaskan model *machine learning* yang digunakan [8]. Sehingga, pengguna yang menggunakan model tersebut dapat mempercayai hasil dari model yang telah diimplementasikan.

# 3 Hasil dan Pembahasan

## 3.1 Pengumpulan Data

Penelitian ini menggunakan data *Human Resources Data set* dari situs Kaggle. Dari data tersebut terdapat 35 atribut dan total sebanyak 3310 data. Data yang diunduh dari situs tersebut berformat *.txt* yang harus kita lihat terlebih dahulu isi dari file data tersebut sehingga kita bisa melihat bagaimana data tersebut terstruktur. Kemudian kita dapat memutuskan bagaimana cara untuk mengimport data tersebut.

### 3.2 Pra-proses Data

Setelah memperoleh data yang ingin kita gunakan untuk penelitian ini, jika dilihat lebih lanjut dalam dataset tersebut terlihat ada beberapa tahapan yang harus kita lakukan terlebih dahulu sebelum memasukkan data tersebut ke dalam model, terdapat beberapa proses, yaitu :

1. Cara umum dalam transformasi data / *data wrangling* adalah dengan merubah tipe data dan menghapus variabel yang berlebihan. Variabel yang berlebihan disini salah satunya adalah *Employee\_Name* dan *EmpID* yang memberikan informasi yang sama. Untuk kasus ini, penggunaan *Employee\_Name* lebih efisien, sehingga memutuskan untuk menghapus *EmpID*. Sebelumnya, dilakukan pemeriksaan apakah terdapat data yang duplikat atau tidak dalam *Employee\_Name*. Data yang duplikat dapat dilihat pada tabel 1.

Tabel 1. Data Duplikat

Employee Name <chr>	EmpID <dbl>	MarriedID <int>	MaritalStatusID <int>
Young, Darien	9071302177	0	0
Yound, Darien	9049431318	0	2
Warner, Larissa	9024100552	1	1
Warner, Larissa	9001310400	0	0

2. Menghapus data yang berlebihan. Dalam dataset tersebut terdapat atribut kolom deskripsi dan kolom id, seperti *MarriedID* dan *MaritalDesc* atau *Employee\_Name* dan *EmpID*. Jika seperti itu, maka simpan yang kolom deskripsi dan hapus kolom id.
3. Mengubah tipe data. Proses ini berlaku untuk kolom *DOB*, *DateofHire*, *DateofTermination*, *LastPerformanceReview\_Date*, *FromDiversityJobFairID*, dan *Termd*. Untuk *FromDiversityJobFairID* dan *Termd* sebelumnya berformat numerik sehingga diubah menjadi tipe factor karena variabel tersebut hanya berisi 1 dan 0. Untuk variabel lain yang diubah berkaitan dengan tanggal, sebelumnya bertipe *character* yang tidak bisa digunakan untuk pengolahan lebih lanjut sehingga diubah menjadi tipe *date*.
4. Menghapus *missing value*. Dataset ini memiliki beberapa *missing value* yang dapat dilihat pada tabel 2. Untuk variabel *DateofTermination* terdapat *missing value* yang cukup banyak, hal ini karena memang banyak pegawai yang belum berhenti dari pekerjaannya sehingga tidak ada tanggal pemberhentian dan variabel *LastPerformanceReview\_Date* juga memiliki *missing value* yang banyak juga karena mungkin saja ada beberapa pegawai yang belum pernah di ulas performanya. Sehingga *missing value* tersebut diganti dengan "01-01-2020". Sisa *missing value* pada variabel *EngagementSurvey*, *EmpSatisfaction*, dan *SpecialProjectsCount* dapat dikatakan cukup kecil dari total keseluruhan data, sehingga memutuskan untuk dihapus saja *missing value* tersebut.

Tabel 2. Missing Value

Department	0	ManagerName	0
RecruitmentSource	0	PerformanceScore	0
EngagementSurvey	78	EmpSatisfaction	78
SpecialProjectsCount	78	LastPerformanceReview_Date	3103

### 3.3 Feature Engineering

Tahapan ini merupakan proses penambahan fitur ke dalam dataset. Sebelumnya, *DOB* dan atribut lainnya yang berkaitan dengan tanggal telah diubah tipenya menjadi *date* dan juga pengisian *missing value* dengan tanggal "01-01-2020" semua proses itu mengarah ke proses penambahan fitur ini. Kita dapat membuat suatu fitur yang berguna untuk pemodelan nanti. Fitur yang akan dibuat adalah *WorkingDay* dan *DayAfterReview*. Atribut *WorkingDay* menunjukkan berapa lama pegawai telah bekerja. Jika pegawai masih aktif maka tanggal "recent" yaitu "01-01-2020" dikurangi dengan *DateofHire*, dan jika pegawai tersebut telah berhenti bekerja maka "recent" dikurangi dengan *DateofTermination*. Untuk atribut *DayAfterReview*, jika "recent" dikurangi *LastPerformanceReview\_Date* maka akan menghasilkan "0" yang berarti memang pegawai tersebut belum pernah di ulas atau *review*. Penambahan fitur tersebut dapat dilihat pada tabel 3.

Tabel 3. Feature Engineering

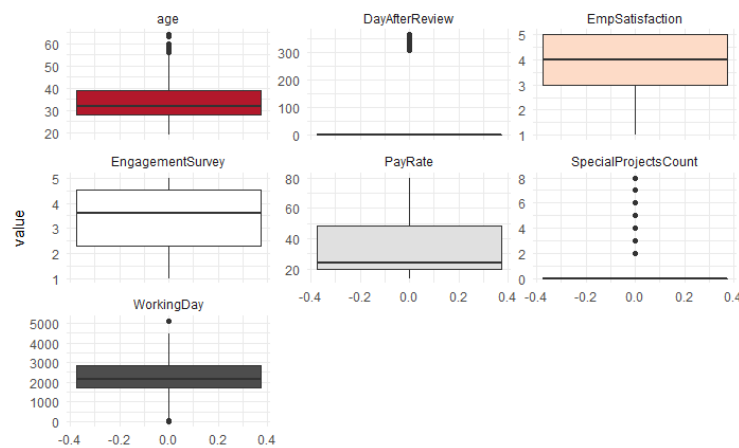
LastPerformanceReview_Date	Age	DayAfterReview	WorkingDay
28/01/2019	34	338	1822
14/01/2019	36	352	2368
01/01/2019	47	0	1083
28/01/2019	39	338	1920
22/02/2019	48	313	2284
04/01/2019	38	362	5105

### 3.4 Analisis Data

Tahapan ini lebih banyak mengeksplorasi data untuk melihat adanya *outlier*, korelasi dan mendapatkan hipotesis yang nantinya akan membantu dalam prediksi. Beberapa proses eksplorasi dilakukan, yaitu :

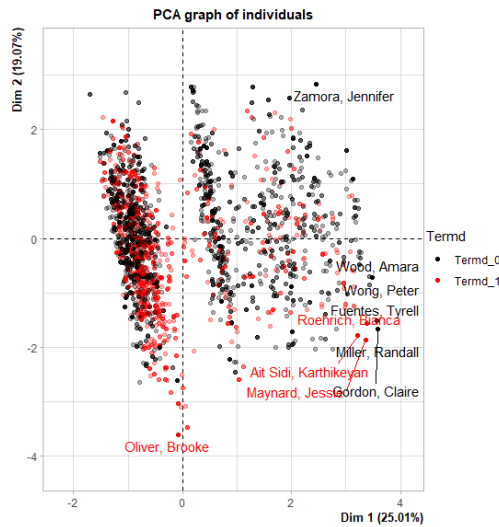
1. Melihat *Data Outlier*.

*Data outlier* ini divisualisasikan menggunakan *boxplot* dengan setiap variabel *numerik* dalam dataset. *Boxplot* dapat dilihat pada Gambar 2 dibawah ini.

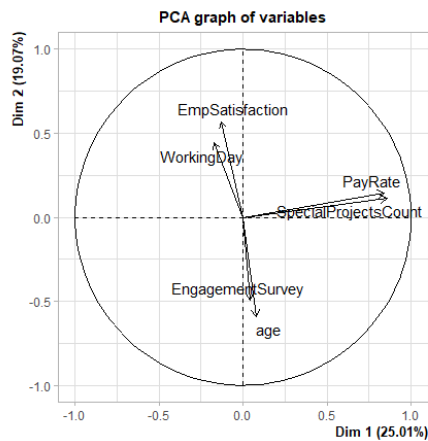


**Gambar. 2.** Boxplot

Dari gambar *boxplot* diatas terdapat 4 variabel yang memiliki *outlier*, yaitu *age*, *DayAfterReview*, *SpecialProjectsCount*, dan *WorkingDay*. Setelah itu kita memvisualisasikan dengan *biplo*t menjadi dua dimensi, dengan membuat *PCA* terlebih dahulu untuk mengurangi dimensi dengan meringkas variabel *numerik*. Berikut gambar *biplo*t dapat dilihat pada Gambar 3. Kemudian lihat bagaimana variabel berkontribusi pada setiap *PC* yang dapat dilihat pada Gambar 4.



Gambar. 3. PCA Variabel



Gambar. 4. Biplot

- Melihat Korelasi antara Manager dan PerformanceScore.  
Untuk melihat korelasi tersebut dapat menggunakan *chi-square* untuk melihat apakah dua data tersebut saling ketergantungan dengan menggunakan fungsi `chisq.test()` pada bahasa R. Hasil uji dapat dilihat pada tabel 4.

Tabel 4. Chi-Square Test

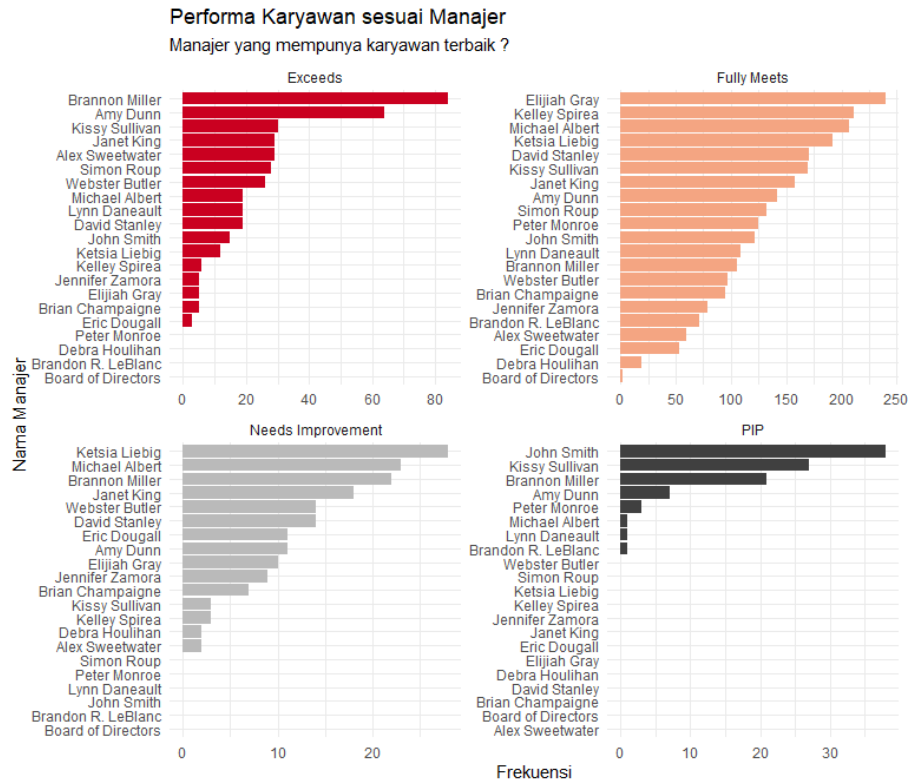
X-squared	826.79
df	60
p-value	< 2.2e-16

Untuk melihat output lebih jelas, harus ditentukan terlebih dahulu hipotesis H0 dan H1.

**H0 : Tidak ada hubungan antara Manager dan PerformanceScore**

**H1 : Ada hubungan antara Manager dan PerformanceScore**

dapat diketahui bahwa  $p\text{-value} < \alpha$ , kita asumsikan bahwa  $\alpha$  yang digunakan adalah 0,95. Sehingga dapat disimpulkan bahwa  $H_1$  yang berarti ada hubungan antara *Manager* dan *PerformanceScore*. Kemudian dapat kita visualisasikan performa pegawai sesuai manager seperti pada Gambar 10 dibawah ini.



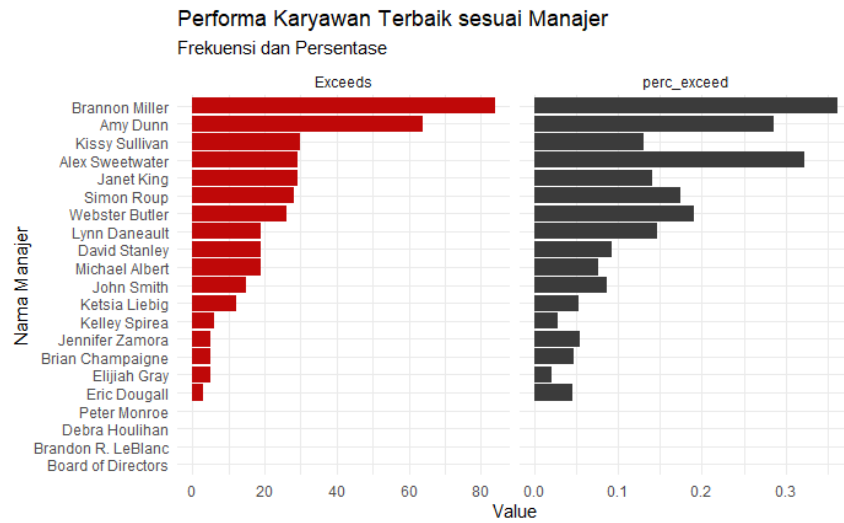
**Gambar. 5.** Performa Pegawai

Dari Gambar 5 dapat dilihat bahwa “*Exceeds*” merupakan performa kinerja terbaik dan “*PIP*” merupakan yang terendah, jika dilihat gambar tersebut memberi tahu bahwa Brannon Miller merupakan manajer terbaik, tetapi gambar tersebut belum cukup memberi tahu tentang yang terjadi antara manajer dan pegawai mereka terkait *PerformanceScore*, hal tersebut mungkin saja menjadikan Brannon Miller manajer terbaik jika dihitung berdasarkan frekuensi. Kemudian lihat banyaknya pegawai Brannon Miller yang terdapat pada tabel 4.

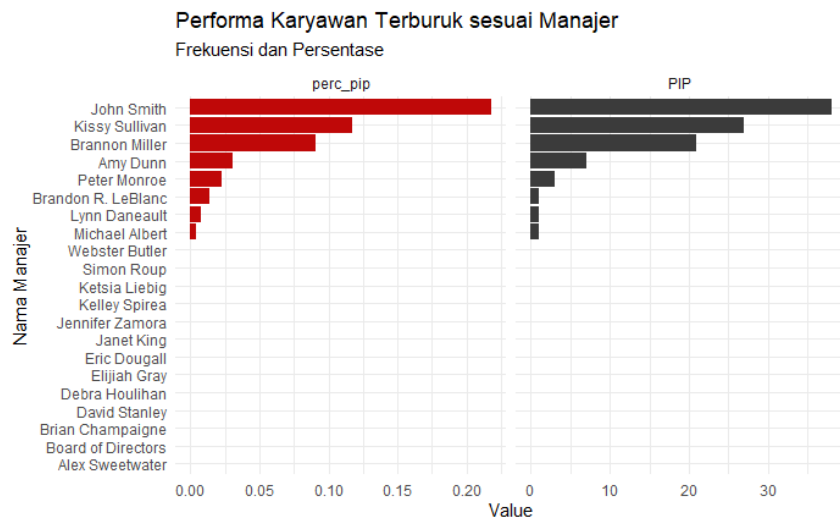
**Tabel 5.** Frekuensi Pegawai

ManagerName	n
Elijah Gray	255
Michael Albert	250
Brannon Miller	232
Ketsia Liebig	232

Terlihat jelas bahwa Brannon Miller memang memiliki pegawai yang banyak. Kemudian proses selanjutnya dengan membuat perbandingan data untuk melihat persentase pegawai terbaik dengan total banyaknya pegawai dari masing-masing manajer. Hasil visualisasi *frekuensi* dan *persentase* pegawai dapat dilihat pada Gambar 6 dan Gambar 7.



**Gambar. 6.** Performa Karyawan Terbaik



**Gambar. 7.** Performa Karyawan Terburuk

Dari Gambar 6 terbukti bahwa Brannon Miller merupakan manajer dengan performa pegawai terbaik, tetapi dapat dilihat juga bahwa Amy Dunn yang terbaik kedua berdasarkan *frekuensi* bukan merupakan yang terbaik kedua berdasarkan persentase, melainkan Alex Sweetwater yang memiliki persentase pegawai terbaik dari Amy Dunn.

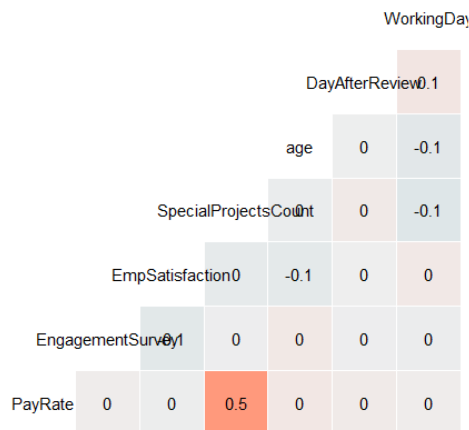
- Melihat perbandingan PayRate antara departemen.  
Perbandingan rasio upah untuk setiap departemen dapat dilihat pada tabel 6.

**Tabel 6.** PayRate Department



Department	avg PayRate	median PayRate	min PayRate	max PayRate
Sales	54.99021	55.00	25.00	65
Software Engineering	51.66522	52.25	27.00	62
Executive Office	51.85417	47.00	27.00	80
IT/IS	46.38903	45.00	21.00	65
Admin Offices	36.01150	29.00	16.56	62
Production	23.61290	21.00	14.00	80

Dari tabel 6, terlihat bahwa Production memiliki avg dan median terendah dibandingkan departemen lain. Hal tersebut mungkin saja terjadi karena setiap departemen mempunyai peran pekerjaan yang berbeda dan ada faktor lain yang mempengaruhi upah atau gaji. Oleh karena itu, tidak dapat langsung disimpulkan bahwa terdapat kesetaraan *PayRate* setiap departemen. Selanjutnya akan dilihat lebih dalam lagi tentang analisis *PayRate* yang terdapat pada Gambar 8.



**Gambar. 8.** Korelasi Variabel PayRate

Dari Gambar 8, dapat dilihat hanya antara variabel *PayRate* dan *SpecialProjectsCount* yang memiliki nilai korelasi yang tinggi dibandingkan *PayRate* dengan variabel lainnya. Hal tersebut dapat diartikan bahwa semakin tinggi *SpecialProjectsCount* maka semakin tinggi juga *PayRate*. Kemudian akan dihitung menggunakan korelasi pearson. Hasil hitung dari korelasi pearson dapat dilihat pada tabel 7.

**Tabel 7.** Korelasi Pearson

t	31.591
df	3230
p-value	< 2.2e-16

Korelasi pearson disini akan membuktikan apakah variabel yang bersangkutan saling mempengaruhi satu sama lain dan dapat disimpulkan bahwa  $p\text{-value} < \alpha$  ( 0,05 ) dengan hipotesis :

**H0 : Tidak ada hubungan antara *PayRate* dan *SpecialProjectsCount***

**H1 : Ada hubungan antara *PayRate* dan *SpecialProjectsCount***

Sama seperti perhitungan chi-square sebelumnya, dapat disimpulkan bahwa H1, yaitu terdapat hubungan antara *PayRate* dan *SpecialProjectsCount*. Kemudian, pada tabel 8 dapat dilihat departemen apa yang memiliki *SpecialProjects* yang tinggi.

**Tabel 8. Special Projects Count**

Department	Total	Avg
IT/IS	2917	5.55619048
Software Engineering	479	4.16521739
Executive Office	87	3.62500000
Admin Offices	253	3.16250000
Sales	65	0.19578313
Production	95	0.04406308

Dapat dilihat dari tabel 8, bahwa departemen IT yang memiliki *SpecialProjectsCount* yang tinggi dengan rata-rata setiap pegawai IT memiliki 5 project. Sehingga dapat disimpulkan bahwa pegawai IT memiliki *PayRate* yang tinggi karena memiliki lebih banyak *SpecialProjectsCount* dibandingkan pegawai pada departemen *Sales* dan *Production*.

### 3.5 Pemisahan Data

Sebelum masuk ke dalam pemodelan, data harus dipisah terlebih dahulu menjadi data train dan data test. Variabel yang menjadi data train dan data test adalah *Termd*. Setelah data dipisah, maka pada penelitian ini dilihat terlebih dahulu apakah data train yang digunakan sudah seimbang (*balanced*) atau belum. Dalam data train yang akan digunakan tersebut ternyata belum seimbang, dapat dilihat pada tabel 9.

**Tabel 9. Imbalanced Data Train**

0	1
0.6702744	0.329725
	6

Terlihat bahwa data tersebut belum seimbang, maka dalam penelitian ini menggunakan *downsampling* untuk mengatasi ketidak seimbangan data train tersebut. Penggunaan *downsampling* disini akan menghapus nilai mayoritas. Berikut hasil setelah *downsampling* yang dapat dilihat pada tabel 10.

**Tabel 10. Balanced Data Train**

0	1
0.5	0.5

### 3.6 Modelling Random Forest

Sekarang kita masuk ke dalam model *random forest*. Model yang sudah kita buat dapat kita ukur kinerjanya menggunakan *confusion matrix*.

**Tabel 11. Hasil Model Random Forest**

Accuracy	0.8775
Sensitivity	0.8545
Specifity	0.9245

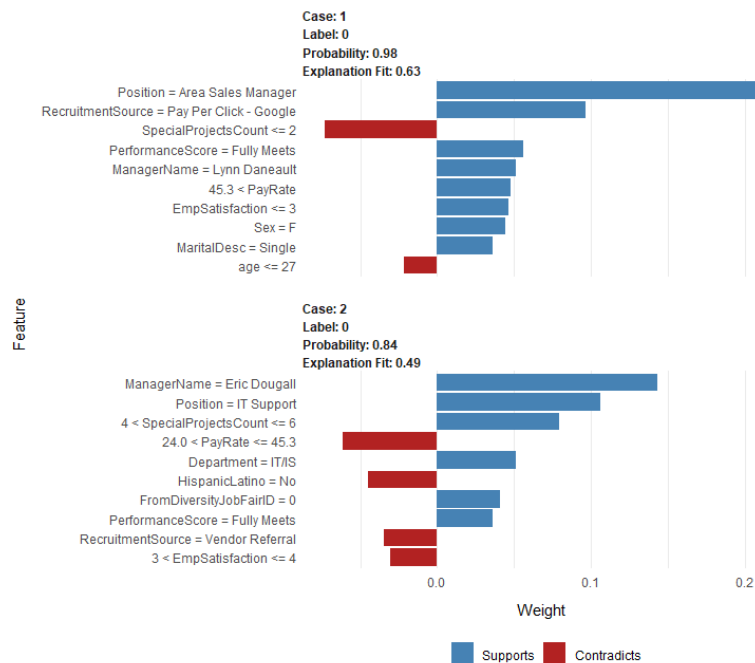
**Tabel 12. Confusion Matrix**

Confusion Matrix		Reference	
		0	1
Prediction	0	370	16
	1	63	196

Dari tabel 11, dapat dilihat hasil akurasi yang didapatkan dari model yang sudah dibuat sebesar 0,8755 atau sebesar 87%. Dari tabel 12 kita dapat melihat juga kinerja dari model yang sudah kita buat menggunakan confusion matrix. Dari table confusion matrix kita dapat menyimpulkan bahwa model yang sudah kita buat sudah cukup baik dan didukung dengan akurasi yang cukup tinggi yaitu sebesar 87%.

### 3.7 Interpretasi LIME

Setelah proses implementasi model *random forest*, kesulitan yang dialami adalah cara menginterpretasikan model *random forest* yang telah diimplementasikan. Hasil dari model tersebut perlu ditafsirkan supaya lebih mudah dalam mempresentasikan kepada pihak atas atau manajer. Oleh karena itu, perlu alat tambahan yang bernama LIME (*Local Interpretable-Model Agnostic Explanations*). Hasil visualisasi dari LIME dapat dilihat pada Gambar 9.



**Gambar. 9.** LIME

Dapat dilihat dari Gambar 9, hanya menggunakan 2 contoh saja dan dari plot tersebut menunjukkan bagaimana setiap variabel berpengaruh dalam proses prediksi yang telah dibuat sebelum dengan model *random forest*. Keterangan “Label:0” menunjukkan nilai dari target variabel apa yang sedang di prediksi, dalam hal ini 0 = tidak berhenti kerja. Keterangan “Probability” menunjukkan probabilitas dari pengamatan “Label:0”. Warna dari setiap batang plot mewakili apakah variabel tersebut mendukung atau bertentangan. Pada bagian “Explanation Fit” menjelaskan seberapa baik LIME menjelaskan model yang dibuat. Untuk observasi pertama terlihat “Explanation Fit” bernilai 63% dan pada observasi kedua bernilai 49%, sehingga dapat diartikan bahwa LIME hanya menjelaskan sedikit dari model yang dibuat.

## 4 Kesimpulan

Penelitian ini mengimplementasikan metode *Random Forest* untuk memprediksi pengunduran diri karyawan perusahaan “Y” menggunakan dataset dari situs website *Kaggle*. Dataset tersebut memiliki 35 atribut dan 3310 *record data*. Dalam penelitian ini *Random Forest* dapat memprediksi pengunduran diri karyawan dengan akurasi sebesar 87% dan *error* sebesar 13%.

## Referensi

- [1] Andie. (2016). Penerapan Decision Tree Untuk Menganalisis Mahasiswa Baru. *Technologia*, 7(1), 8–14.
- [2] Fauzi, A. M. (2018). Random Forest Approach for Sentiment Analysis in Indonesia Language. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), 46-50.
- [3] Mahangara, A., & Laksito, A. D. (2019). Prediksi Pengunduran Diri Mahasiswa Universitas Amikom Yogyakarta Menggunakan Metode Naive Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(1), 273–280. <https://doi.org/10.24176/simet.v10i1.2967>
- [4] Mishra, S. N., & Lama, D. R. (2016). A Decision Making Model for Human Resource Management in Organizations using Data Mining and Predictive Analytics. *International Journal of Computer Science and Information Security (IJCSIS)*, 14(5), 217–221.
- [5] Nasril, F., Indiyati, D., & Ramantoko, G. (2021). Talent Performance Analysis Using People Analytics Approach. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(1), 216–230. <https://doi.org/10.33258/birci.v4i1.1585>
- [6] Provost, F. (2000). Machine learning from imbalanced data sets 101. *Proceedings of the AAAI'2000 Workshop on ...*, 3. <https://www.aaai.org/Papers/Workshops/2000/WS-00-05/WS00-05-001.pdf%5Cnpapers://1c40c143-2a6e-4e94-8c17-c5bc9ae73d7e/Paper/p11435>
- [7] Setianto, K. S., & Jatikusumo, D. (2020). Employee Turnover Analysis Using Comparison of Decision Tree and Naïve Bayes Prediction Algorithms on K-Means Clustering Algorithms at PT. AT. *Jurnal Mantik*, 4(3), 1573-1581.
- [8] Visani, G., Bagli, E., & Chesani, F. (2020). OptiLIME: Optimized lime explanations for diagnostic computer algorithms. *CEUR Workshop Proceedings*, 2699.