

## KLASIFIKASI DONATUR PADA YAYASAN JAJAN PAHALA MENGUNAKAN ALGORITMA *RANDOM FOREST*

Nusa Seldi Wibisono<sup>1</sup>, Satrio Agung Wicaksono<sup>2</sup>, Nanang Yudi Setiawan<sup>3</sup>

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>nusaseldi@student.ub.ac.id, <sup>2</sup>satrio@ub.ac.id, <sup>3</sup>nanang@ub.ac.id

### Abstrak

Yayasan Jajan Pahala adalah salah satu yayasan yang bergerak pada bidang sosial, keagamaan, dan kemanusiaan yang sudah ada sejak tahun 2015 dan resmi menjadi yayasan pada tahun 2019. Pada pelaksanaan kegiatan yang dilakukan, Yayasan Jajan Pahala mendapatkan data mengenai donasi yang dilakukan dan data informasi donatur yang berdonasi. Data tersebut hanya disimpan dan tidak ada analisis lebih jauh dari data yang diperoleh. Padahal, dari data yang didapatkan itu bisa diolah dan dianalisis menjadi informasi yang berguna untuk mendukung keputusan berdasarkan data. Salah satu bentuk analisis data adalah klasifikasi untuk mengelompokkan kelas dari data tersebut. Penulis bertujuan untuk menggunakan model klasifikasi pada data donatur untuk mengklasifikasikan apakah donatur akan rutin berdonasi atau hanya sekali berdonasi. Dari penerapan model klasifikasi tersebut akan dilihat bagaimana hasil evaluasinya. Algoritma yang digunakan pada klasifikasi adalah random forest dengan kombinasi cross-validation dan tuning hyperparameter. Data donatur yang digunakan sebanyak 227 data yang memuat mengenai informasi donatur. Dari penerapan model klasifikasi random forest didapatkan hasil akurasi pada V-fold Cross Validation sebesar 0,729 dengan parameter yang paling optimal pada mtry 13 dan min\_n 20. Hasil evaluasi model dengan akurasi, presisi, recall, dan f-measure mendapatkan nilai secara berurutan sebesar 78%, 67%, 57%, dan 61%. Hasil ini mendeskripsikan bahwa performa model cukup baik.

**Kata kunci:** *Yayasan, Donatur, Data analisis, Klasifikasi, Random forest, Dashboard*

### Abstract

*Yayasan Jajan Pahala is one of the foundations engaged in the social, religious and humanitarian fields that has existed since 2015 and officially became a foundation in 2019. In the implementation of the activities carried out, the Jajan Pahala Foundation gets data about donations made and donor information that donates. The data is only stored and there is no further analysis of the data obtained. In fact, the data obtained can be processed and analyzed into useful information to support data-based decisions. One form of data analysis is classification to classify the class of the data. The author aims to use a classification model on donor data to classify whether donors will regularly donate or only donate once. From the application of the classification model, it will be seen how the evaluation results are. The algorithm used in the classification is random forest with a combination of cross-validation and hyperparameter tuning. The donor data used is 227 data containing donor information. From the application of the random forest classification model, the accuracy result on V-fold Cross Validation is 0.729 with the most optimal parameters at mtry 13 and min\_n 20. The model evaluation results with accuracy, precision, recall, and f-measure get sequential values of 78%, 67%, 57%, and 61%. These results describe that the model's performance is quite good.*

**Keywords:** *Foundation, Donors, Data analytic, Classification, Random forest, Dashboard*

## 1. PENDAHULUAN

Jajan Pahala merupakan salah satu yayasan yang bergerak di bidang sosial, kemanusiaan, dan keagamaan dengan lokasi kantor pusat di Bekasi. Sebelum menjadi Yayasan di tahun

2019, Jajan Pahala sudah ada sejak tahun 2015 sebagai sebuah komunitas anak muda penggerak “Sedekah is fun” dengan tujuan utamanya yaitu menumbuhkan rasa ingin berbagi sekaligus memfasilitasi teman-teman yang ingin bersedekah. Pada mulanya Jajan Pahala

berkegiatan di Bandung, namun seiring berjalannya waktu, Jajan Pahala juga ada di beberapa kota, yaitu Jakarta, Bekasi, Bogor, Cilegon, dan Yogyakarta. Visi dari Jajan Pahala sendiri adalah menebar kebaikan di lingkungan sekitar dengan misinya yaitu saling mengingatkan dalam kebaikan, menumbuhkan rasa kepedulian terhadap sekitar, tanggap terhadap isu sosial dan lingkungan sekitar, menjalankan kegiatan berbasis kekeluargaan, dan melakukan aktivitas sosial secara profesional, tepat sasaran, dan transparan. Untuk mencapai visi dan misi tersebut, jajan pahala melakukan kegiatan yang terbagi dua yaitu kegiatan rutin dan kegiatan eventual. Contoh kegiatan rutin dari jajan pahala adalah berbagi 1000 sembako, berbagi ifthar, dan bagi-bagi nasi. Untuk kegiatan eventual jajan pahala seperti JP peduli bencana, JP membangun desa, dan JP peduli covid.

Selama ini keberlangsungan kegiatan Yayasan Jajan Pahala bersumber dari donasi yang masuk melalui rekening satu-satunya Jajan Pahala yaitu Bank Syariah Indonesia (BSI) atas nama Yayasan Jajan Pahala. Dari donasi yang masuk tersebut, akan didapatkan data donasi yang masuk ke rekening Jajan Pahala. Data riwayat donasi yang masuk selama ini hanya disimpan oleh Yayasan Jajan Pahala dan belum ada pengolahan data lebih jauh. Padahal dari riwayat data donasi tersebut dapat dilakukan analisis historical data untuk membantu Yayasan Jajan Pahala mengetahui gambaran umum donasi yang terjadi dan informasi donatur dalam berdonasi. Dari hasil tersebut dapat dimanfaatkan untuk memecahkan masalah dan menjadi dasar dalam pengambilan keputusan yang berdasarkan data.

Salah satu cara mengolah data riwayat donasi adalah dengan data analisis. Pengertian dari data analisis secara sederhana adalah proses pemeriksaan, transformasi, dan pemodelan data untuk mencari informasi yang berguna dan mendukung pengambilan keputusan. Metode data analisis yang pertama digunakan adalah analisis deskriptif untuk melihat gambaran apa yang telah terjadi pada donasi. Perhitungan yang digunakan dalam analisis deskriptif yaitu distribusi frekuensi untuk melihat persebaran data berdasarkan bank pengirim, kode unik donasi, dan hari. Selain itu, data analisis yang dilakukan selanjutnya adalah klasifikasi untuk mengelompokkan donatur dengan melihat informasi dari donatur tersebut. Klasifikasi merupakan suatu bentuk analisis data yang

membuat model untuk menggambarkan dan mengkategorikan data berdasarkan dari label kelasnya (Han, Kamber, dan Pei, 2012). Pada penerapannya, donatur akan diklasifikasikan dengan melihat potensi dalam frekuensi berdonasi yaitu rutin atau hanya sekali saja donasi. Variabel yang digunakan untuk mengklasifikasi kelas donatur adalah yang berhubungan dengan demografi donatur seperti jenis kelamin, tingkat pendidikan, dan pekerjaan. Selain demografi, rerata donasi dan donasi pertama donatur juga digunakan dalam analisis model. Algoritma *random forest* akan dipakai dalam pembuatan model klasifikasi. *Random forest* adalah *ensemble classifier* yang menggunakan gabungan dari pohon keputusan untuk membuat prediksi (Breiman, 2001). Kelebihan algoritma *random forest* dibanding algoritma klasifikasi lainnya adalah kemampuan untuk menangani dataset dengan jumlah variabel prediktor yang besar dan secara konsisten mendapat akurasi tertinggi dibandingkan dengan model lainnya dalam menjalankan klasifikasi (Fernández-Delgado, et al., 2014).

Berdasarkan atas penelitian Baskoro (2021) yang membandingkan tiga algoritma yaitu *random forest*, *naïve bayes*, dan *k-nn* untuk prediksi penerima beasiswa di Universitas Muhammadiyah Pringsewu. Dari ketiga algoritma yang diuji, *random forest* memiliki performa akurasi yang lebih baik dibanding dua algoritma lainnya dengan hasil sebesar 98,08%.

Penelitian lainnya yang juga menggunakan *random forest* adalah penelitian yang dijalankan oleh Nugraha Listiana Hanun pada tahun 2020. Objek penelitian mengenai kelayakan pemberian kredit di Koperasi Mitra Sejahtera. Data yang dipakai pada penelitian adalah data kreditur dari tahun 2015 sampai 2016 yaitu sebanyak 259 kreditur. Hasil akurasi dari analisis untuk menentukan kredit yang tidak bermasalah dan bermasalah yakni sebesar 87,88%.

Berdasarkan uraian tersebut, peneliti bermaksud melakukan penelitian dengan judul “Klasifikasi Donatur Pada Yayasan Jajan Pahala Menggunakan Algoritma *Random Forest*”. Diharapkan dari penelitian ini dapat mengetahui hasil dari analisis terhadap donasi yang masuk selama ini dan dapat menganalisis kualitas dari model yang dibuat dalam mengklasifikasikan donatur pada Yayasan Jajan Pahala

## 2. LANDASAN PUSTAKA

## 2.1. Analisa Data

Analisa data merupakan proses pemeriksaan, pembersihan, transformasi, dan pemodelan data yang bertujuan untuk menemukan informasi yang berguna, menunjukkan kesimpulan, dan mendukung pengambilan keputusan (Kudyba, 2014). Analisis data adalah proses untuk menggunakan data mentah dan mengubahnya menjadi informasi yang berguna untuk pengambilan keputusan oleh pengguna. John Tukey seorang Ahli Statistik, mendefinisikan analisis data pada tahun 1962, sebagai prosedur untuk menganalisis data, teknik untuk menafsirkan hasil prosedur tersebut, cara merencanakan pengumpulan data untuk membuatnya lebih mudah, lebih tepat atau lebih akurat.

## 2.2. Data Preprocessing

Data preprocessing adalah rangkaian proses yang dilakukan untuk mentransformasikan data mentah menjadi format yang tepat dan sesuai untuk diproses. Data preprocessing dilakukan dengan tujuan untuk mengurangi kesalahan dan bias pada data sebelum analisis dilakukan (Tong et al., 2011). Data preprocessing sendiri terdiri dari beberapa proses seperti penggabungan data, pembersihan data, dan penyeleksian data.

Pembersihan data melibatkan siklus berulang dari penyaringan, diagnosis, perawatan dan dokumentasi dari proses ini. Ketika pola kesalahan diidentifikasi, pengumpulan data dan prosedur entri harus disesuaikan untuk memperbaiki pola tersebut dan mengurangi kesalahan di masa depan.

Van den Broeck et al., 2005 menggambarkan pembersihan data dalam empat siklus. Pertama adalah penyaringan, yaitu melibatkan pencarian fitur yang dicurigai secara sistematis dalam kuesioner penilaian, database, atau kumpulan data analisis. Siklus kedua adalah diagnosis, yaitu mengidentifikasi sifat data yang rusak atau hilang. Siklus ketiga adalah perawatan, yaitu menghapus, mengedit, atau membiarkan data yang rusak tersebut. Siklus keempat adalah dokumentasi, yaitu mendokumentasikan perubahan yang terdeteksi.

## 2.3 Descriptive Analytics

*Descriptive analytics* adalah salah satu bentuk data analisis yang menggunakan historical data untuk mengidentifikasi tren dan hubungan antar variabel kemudian disajikan dengan cara yang mudah dipahami. *Descriptive*

*analytics* fokus pada apa yang telah terjadi. Dari historical data tersebut nantinya akan diolah dan dianalisis untuk menghasilkan insight. Selanjutnya, insight yang didapat akan disajikan dengan bentuk yang mudah dipahami dengan cara visualisasi data menggunakan grafik. Sebagian besar organisasi menggunakan *descriptive analytics* dan *Key Performance Indicators (KPI) dashboard* pada departemen sales dan marketing (Bughin 2017, Delen and Ram 2018, Mintz et al. 2019).

## 2.4. Klasifikasi

Klasifikasi menurut Han, Kamber, dan Pei, (2012) adalah bentuk dari analisis data yang mencari model (atau fungsi) untuk menggambarkan dan membedakan kelas data. Pada klasifikasi, label kelas telah ditentukan terlebih dahulu sehingga klasifikasi masuk ke dalam *supervised learning*. Secara umum, pendekatan klasifikasi terdiri dari dua langkah proses, yang pertama adalah *learning step* yaitu proses klasifikasi model dibentuk dan yang kedua adalah *classification step* yaitu dari model yang telah dibuat digunakan untuk memprediksi label kelas dari data yang ada. Beberapa contoh teknik klasifikasi yaitu *Decision Tree*, *Naive Bayes*, *Support Vector Machine*, dan *Random Forest*.

## 2.5. Random Forest

Algoritma *Random Forest* adalah salah satu algoritma machine learning yang dapat digunakan untuk klasifikasi dan regresi. *Random Forest* merupakan kumpulan dari banyak pohon keputusan yang bertindak sebagai *base classifier*. Cara kerja dari algoritma ini adalah dengan membangun banyak pohon keputusan yang setiap pohon memberikan prediksi kelas. *Random Forest* menggabungkan prediksi ini dengan memilih kelas yang paling sering diprediksi (voting mayoritas). Dalam pengaturan *random forest*, pohon dibangun dengan menggunakan set data latih yang dipilih secara acak dan subset acak dari variabel prediktor untuk memodelkan hasil. Dari hasil setiap pohon digabungkan untuk memberikan prediksi untuk setiap pengamatan. Oleh karena itu, *random forest* sering kali memberikan akurasi yang lebih tinggi dibandingkan dengan model pohon keputusan tunggal dengan tetap mempertahankan beberapa kualitas yang bermanfaat dari model pohon (misalnya kemampuan untuk menafsirkan hubungan antara

prediktor dan hasil) (Speiser, Durkalski, & Lee, 2015). Menurut Breiman (2001), *Random Forest* dapat berjalan secara efisien pada database yang besar, dapat menangani ribuan variabel tanpa harus dilakukan penghapusan variabel, dan mempunyai metode efektif untuk menjaga akurasi ketika terdapat data yang hilang.

**2.6. V-Fold Cross Validation**

*Cross Validation* adalah adalah *resampling method* yang sudah mapan. Meskipun terdapat sejumlah variasi, metode *cross validation* yang paling umum adalah *V-Fold Cross Validation*. Pada *V-Fold Cross Validation*, data dipartisi secara acak ke dalam V set dengan ukuran yang kurang lebih sama yang disebut *folds* (Kuhn & Silge, 2022). Masing-masing dari set penilaian V berisi 1/V dari training set dan masing-masing mengecualikan titik data yang berbeda. Misalnya V = 10, maka ada sepuluh versi berbeda dari 90% data dan juga sepuluh versi dari 10% sisanya untuk setiap *resample* (Kuhn & Johnson, 2019).

**2.7. Tuning Hyperparameter**

*Tuning* adalah proses pencarian *hyperparameter* yang optimal dari model algoritma sesuai dengan dataset yang digunakan. Ada dua *hyperparameter* dasar yang menonjol pada random forest yaitu *mtry* dan *min\_n*. Dari kedua parameter tersebut, *mtry* memiliki dampak paling besar (Probst, Wright, & Boulesteix, 2019; van Rijn & Hutter, 2018). *Mtry* mengacu pada jumlah variabel yang dipilih secara acak dan dipertimbangkan untuk pemisahan di setiap node selama pembuatan pohon keputusan. Ini adalah parameter penting yang mempengaruhi kinerja dan akurasi model *Random Forest*. Nilai dari parameter *mtry* biasanya adalah akar kuadrat dari jumlah variabel input (Gislason et al., 2006). *Min\_n* adalah ukuran minimum jumlah sampel yang harus ada di setiap node terminal (*leaf*) dari pohon keputusan. Jika jumlah sampel dalam suatu node kurang dari nilai ini, maka node tersebut tidak akan dibagi lagi (Kuhn dan Johnson, 2013).

**2.8. Evaluasi Model Klasifikasi**

Setelah model klasifikasi berhasil dibangun, dibutuhkan alat yang dapat menentukan kualitas dari model klasifikasi tersebut. Dalam pemodelan klasifikasi, ada beberapa evaluasi model yang bisa digunakan

salah satunya adalah *confusion matrix*. Pada *confusion matrix* akan terlihat jumlah data yang tepat diklasifikasi dan jumlah data yang tidak tepat diklasifikasi (Rahman, Darmawidjadja, & Alamsah, 2017). Tampilan dari *confusion matrix* terdapat pada tabel 2.1.

Tabel 1. Confusion Matrix

		Kelas Prediksi	
		Positif	Negatif
Kelas Sebenarnya	Positif	True Positives (TP)	False Negatives (FN)
	Negatif	False Positives (FP)	True Negatives (TN)

Keterangan :

1. *True Positives* (TP) adalah jumlah data positif yang benar diklasifikasikan sebagai kelas positif.
2. *True Negatives* (TN) adalah jumlah data negatif yang benar diklasifikasikan sebagai kelas negatif.
3. *False Positives* (FP) adalah jumlah data negatif yang salah diklasifikasikan sebagai kelas positif.
4. *False Negatives* (FN) adalah jumlah data positif yang salah diklasifikasikan sebagai kelas negatif.

Dari *confusion matrix* tersebut, bisa mendapatkan perhitungan akurasi, presisi, *recall*, *F-Measure*. Akurasi adalah ukuran seberapa banyak prediksi benar yang dibuat oleh model. Presisi adalah ukuran seberapa banyak kasus yang diprediksi dengan benar ternyata positif. *Recall* memberitahu seberapa banyak positif sebenarnya yang dapat diprediksi dengan benar. *F-Measure* adalah rata-rata harmonik dari presisi dan *recall* yang memberikan ide gabungan tentang kedua metrik (Karimi, 2021).

$$Akurasi: \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Presisi: \frac{TP}{TP + FP} \tag{2}$$

$$Recall: \frac{TP}{TP + FN} \tag{3}$$

$$F - measure: \frac{2(Presisi * Recall)}{(Presisi + Recall)} \tag{4}$$

**2.9. Dashboard**

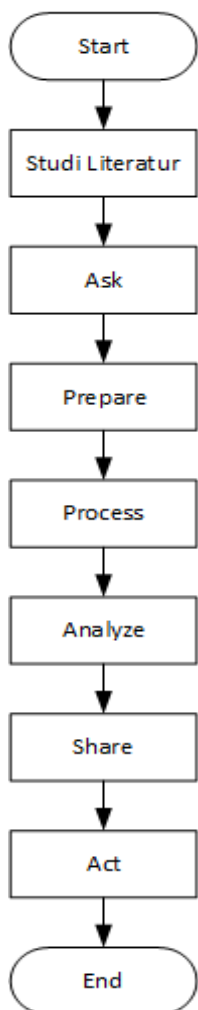
Pengertian *Dashboard* adalah tampilan visual dari data yang digunakan untuk memantau suatu kondisi dan memfasilitasi pemahaman yang dapat mencakup elemen infografis atau



visualisasi naratif (Wexler et al, 2017). Menurut Sarikaya et al (2019) tujuan penggunaan *dashboard* mendorong pemilihan dalam sisi desain visual dan kemampuan fungsional. Dua faktor berikut menangkap peran dari setiap *dashboard* dalam proses analisis dan komunikasi.

**3. METODOLOGI PENELITIAN**

Alur penelitian klasifikasi dimulai dengan studi literatur. Dilanjut dengan fase *Ask* atau identifikasi masalah yang akan diselesaikan. Berikutnya adalah *prepare* yang mana merupakan fase pengumpulan data. Selanjutnya fase *process* yaitu pembersihan data yang telah dikumpulkan. Kemudian fase *analyze* yaitu melakukan analisis data dengan analisis deskriptif dan klasifikasi untuk menyelesaikan masalah. Setelah itu, masuk ke fase *share* yang mana data yang telah diproses pada tahap sebelumnya divisualisasikan dengan grafik. Terakhir yaitu fase *Act* yang mana akan diambil kesimpulan dari proses yang telah dijalankan. Alur penelitian ditunjukkan pada gambar 2.



Gambar 1. Tahapan Penelitian

**3.1 Studi Literatur**

Tahap pertama adalah studi literatur yang diperlukan untuk mempelajari informasi dan pengetahuan mengenai tahapan dan proses data analisis. Studi literatur dilaksanakan dengan tujuan sebagai landasan teori dalam penelitian dan juga sebagai dasar dalam implementasi analisis data.

**3.2 Ask**

Tahap kedua adalah *Ask* yang mana pada tahap ini akan dilakukan identifikasi masalah yang akan diselesaikan pada penelitian dan juga memastikan ekspektasi dari stakeholder. Dalam penelitian ini, identifikasi masalah yang terjadi adalah belum adanya pengolahan pada data yang dimiliki oleh yayasan. Padahal dari data yang terkumpul, bisa dimanfaatkan untuk menjadi informasi yang berguna bagi yayasan jika data tersebut dapat diolah. Salah satu cara untuk memanfaatkan data yang ada yaitu dengan melakukan analisis pada data riwayat donasi menggunakan metode klasifikasi random forest untuk mengklasifikasikan donatur. Dari hasil analisis data tersebut, diharapkan bisa membantu Yayasan dalam mengategorikan donatur yang selama ini berdonasi.

**3.3 Prepare**

Tahap ketiga adalah mempersiapkan data apa saja yang dibutuhkan untuk menyelesaikan masalah. Mulai dari mengenali perbedaan data format, tipe, dan struktur sampai mengumpulkan dan menyimpan data untuk analisis. Dalam penelitian ini, data yang dibutuhkan adalah data mengenai donatur yang berdonasi dan juga data transaksi yang dilakukan oleh donatur. Untuk proses pengumpulan data didapatkan secara langsung dari data internal Yayasan Jajan Pahala. Data yang terkumpul adalah data riwayat donasi selama enam bulan dari Agustus 2021 sampai Januari 2022 dan data terkait informasi donatur. Dari data riwayat donasi, diperoleh data transaksi sebanyak 887 records dan atribut yang berjumlah 11. Untuk data mengenai informasi donatur, didapat sebanyak 243 records dan 5 atribut.

**3.4 Process**

Tahap keempat adalah pembersihan data dan transformasi data untuk memastikan data

yang digunakan relevan dan konsisten. Setelah membersihkan data yang error, tidak akurat, dan tidak konsisten selanjutnya akan dilakukan penyeleksian data. Pada tahap seleksi data, atribut yang tidak dibutuhkan pada penelitian akan dihapus dan hanya diambil atribut yang dibutuhkan saja. Dari data yang didapat terdapat beberapa data yang tidak lengkap dan juga data yang tidak konsisten. Untuk mengatasi data yang tidak lengkap, diputuskan untuk menghapus baris data yang tidak lengkap. Sedangkan untuk data yang tidak konsisten, akan diperbaiki format data sehingga data tersebut konsisten. Data yang telah dibersihkan akan digunakan pada tahap selanjutnya yaitu Analyze.

### 3.5 Analyze

Tahap kelima adalah eksplorasi dan analisis data untuk menemukan pola dan mengambil kesimpulan dengan menggunakan data *analysis tools*. Dari data yang telah dibersihkan di tahap sebelumnya akan digunakan untuk tahap selanjutnya yaitu analisis data. Tipe analisis yang dipakai adalah *descriptive analytics* yaitu salah satu bentuk analisis yang menggunakan *historical data* untuk mengidentifikasi tren dan hubungan antar variabel. Setelah analisis deskriptif dilakukan, akan dilanjut dengan metode klasifikasi *random forest* untuk mengelompokkan donatur berdasarkan label yang telah dibuat sebelumnya. Hasil dari analisis berupa tren dari data dan klasifikasi dari donatur yang akan divisualisasikan pada tahap berikutnya.

### 3.6 Share

Tahap keenam adalah menafsirkan hasil analisis dengan membuat visualisasi dan menyampaikan hasilnya kepada stakeholder. Setelah tahap analisis selesai, selanjutnya akan disampaikan hasil dari analisis dengan visualisasi yang jelas dan menarik melalui grafik atau *dashboard*. Dari hasil analisis yang telah dijalankan ditahap sebelumnya, data hasil analisis akan divisualisasikan dengan cara membuat *dashboard*.

### 3.7 Act

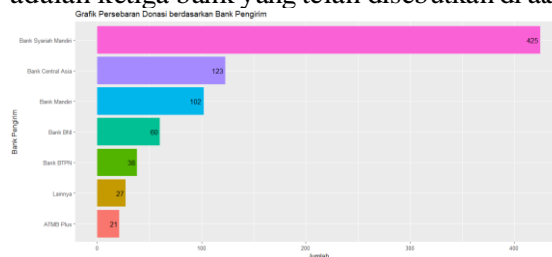
Tahap ketujuh adalah saatnya mengaplikasikan *insight* yang didapat untuk menyelesaikan masalah. Dari hasil analisis yang telah dilakukan sebelumnya, peneliti akan memberikan rekomendasi kepada *stakeholder* sesuai dengan temuan yang ada sehingga

*stakeholder* dapat membuat keputusan yang berdasarkan data.

## 4. HASIL DAN PEMBAHASAN

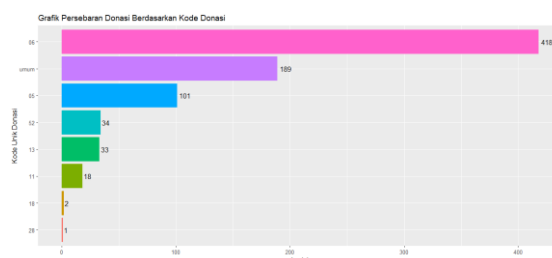
### 4.1 Hasil Descriptive Analytics

Analisis deskriptif berfokus untuk menjawab apa yang telah terjadi. Dengan analisis deskriptif ini akan dapat terlihat gambaran data yang telah didapat secara umum. Analisis yang pertama adalah melihat persebaran dari bank pengirim yang melakukan donasi yang dapat dilihat pada gambar 2. Dari 796 transaksi yang terkumpul, sebanyak 53,4% transaksi dilakukan melalui Bank Syariah Mandiri yang sekarang telah bersatu dengan bank syariah lainnya menjadi Bank Syariah Indonesia (BSI). Urutan kedua ditempati Bank Central Asia (BCA) dengan 15,4% dan di urutan ketiga ada Bank Mandiri dengan 12,8%. Jika diperhatikan, nilai transaksi yang mencapai lebih dari 10% adalah ketiga bank yang telah disebutkan di atas.



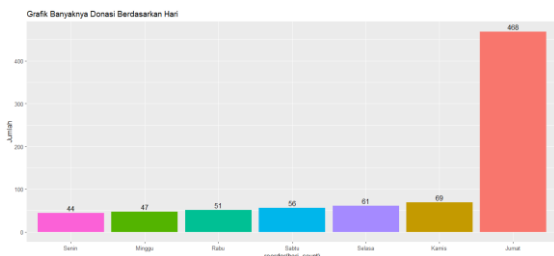
Gambar 2. Persebaran berdasarkan bank pengirim

Setelah melihat persebaran bank pengirim, analisis selanjutnya akan mengamati persebaran donasi berdasarkan kode donasi. Dari hasil analisis didapatkan kode donasi dengan jumlah terbanyak adalah 06 sebesar 418 untuk kegiatan bagi-bagi nasi. Berikutnya adalah donasi umum sebanyak 189 diikuti dengan donasi 05 untuk bencana semeru sebanyak 101. Selanjutnya ada kode donasi 52 untuk Employee Giveback Pertamina Trans Kontinental yang berjumlah 34 dan kode donasi 13 untuk pejuang kanker sebanyak 33. Pemetaan dari kode donasi ini dapat melihat kegiatan apa yang banyak didukung oleh donatur. Gambar persebaran kode unik donasi dapat dilihat pada gambar 3.



Gambar 3. Persebaran berdasarkan kode donasi

Analisis ketiga adalah melihat banyak donasi berdasarkan hari sesuai pada gambar 4. Ada yang menarik dari hasil yang didapat yaitu jumlah donasi mencapai puncak pada Hari Jumat dengan total mencapai 58,8% dari keseluruhan donasi. Sedangkan pada hari lainnya, jumlah donasi hanya berkisar antara 5,5% sampai 8,7%.



Gambar 4. Persebaran berdasarkan hari

### 4.2 Hasil Pengujian Akurasi Random Forest

Hasil akurasi dari model klasifikasi didapat menggunakan fungsi accuracy() dari paket yardstick yang akan membandingkan data dari kelas sebenarnya dengan data kelas prediksi. Dari hasil pengujian, didapatkan akurasi pada data uji dari setiap rasio data latih yang dapat dilihat di tabel di bawah. Akurasi tertinggi didapat pada rasio 90% data latih yaitu sebesar 0,78. Untuk perhitungan selanjutnya akan memakai persentase akurasi terbaik yaitu 90:10.

Tabel 2. Hasil akurasi pada data uji

Persentase data latih & data uji	Akurasi
50 : 50	0,65
60 : 40	0,76
70 : 30	0,76
80 : 20	0,76
90 : 10	0,78

### 4.3 Hasil Pengujian V-Fold Cross Validation

Pada perhitungan Cross Validation, V yang diterapkan adalah sepuluh yang berarti ada 10 Fold pembagian data. Dari 10 Fold tersebut, dikombinasikan lagi dengan tuning hyperparameter via grid search. Terdapat 10 fold ditambah dengan 10 kombinasi mtry dan min\_n dari grid search. Pada tabel di bawah ini adalah rata-rata dari nilai kesepuluh fold yang telah dikombinasikan dengan hyperparameter grid search. Dapat dilihat hasil terbaik dari parameter pada mtry bernilai 13 dan min\_n bernilai 20. Kombinasi tersebut menghasilkan rata-rata sebesar 0,7297. Untuk lebih jelasnya ditunjukkan pada gambar 5.

No	mtry	min_n	.metric	.estimator	mean	n	std_err
1	3	8	accuracy	binary	0.6767749	10	0.02492960
2	7	28	accuracy	binary	0.7004113	10	0.03098913
3	15	32	accuracy	binary	0.7242641	10	0.03400125
4	6	6	accuracy	binary	0.7049351	10	0.02651373
5	11	39	accuracy	binary	0.7051732	10	0.03012031
6	9	14	accuracy	binary	0.7294805	10	0.02639284
7	2	11	accuracy	binary	0.6027273	10	0.03975728
8	13	20	accuracy	binary	0.7297186	10	0.02887305
9	5	31	accuracy	binary	0.6908658	10	0.02499256
10	11	24	accuracy	binary	0.7097186	10	0.03125455

Gambar 5. Rata-rata hasil V-Fold Cross Validation

### 4.4 Confusion Matrix

Evaluasi lebih lanjut dari model klasifikasi menggunakan confusion matrix yang dapat dilihat pada tabel di bawah. Dari confusion matrix, didapat perhitungan untuk nilai akurasi, presisi, recall, F-Measure. Nilai presisi dapat dihitung dengan menerapkan persamaan 2, Recall didapatkan nilainya dengan mengaplikasikan persamaan 3, dan nilai F-Measure didapat dengan cara menggunakan persamaan 4. Data latih yang dipakai adalah 90% dikarenakan pada rasio tersebut mendapat hasil akurasi yang tertinggi.

Tabel 3. Confusion matrix hasil pengujian data uji

Data latih 90% & data uji 10%		Aktual	
		Tidak Rutin	Rutin
Prediksi	Tidak Rutin	14	3
	Rutin	2	4

$$Akurasi : \frac{TP+TN}{TP+TN+FP+FN} = \frac{4+14}{4+14+2+3} = 0,78$$

$$Presisi : \frac{TP}{TP+FP} = \frac{4}{4+2} = 0,67$$

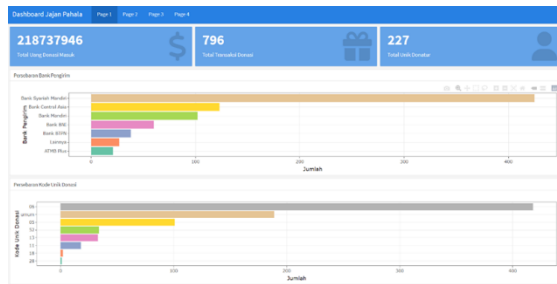
$$Recall : \frac{TP}{TP+FN} = \frac{4}{4+3} = 0,57$$

$$F\text{-measure} : \frac{2(Presisi * Recall)}{(Presisi + Recall)} = \frac{2(0,67 * 0,57)}{(0,67 + 0,57)} = 0,61$$

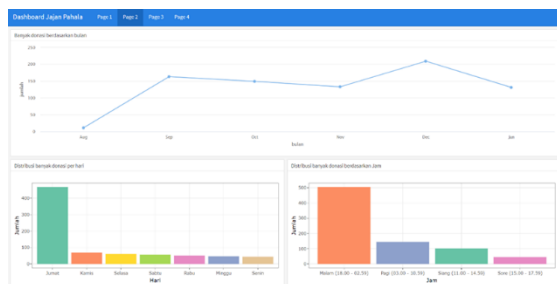
### 4.5 Hasil Visualisasi Dashboard

Pembuatan dashboard dilakukan untuk melihat hasil dari analisis deskriptif pada riwayat donasi dan juga menampilkan hasil dari klasifikasi donatur dengan random forest. Dashboard yang dibuat bernama Dashboard Jajan Pahala dan terdiri dari empat halaman. Halaman satu dan dua menampilkan visual dari hasil analisis riwayat donasi, halaman ketiga menampilkan data terkait donatur dan halaman empat memperlihatkan hasil dari klasifikasi

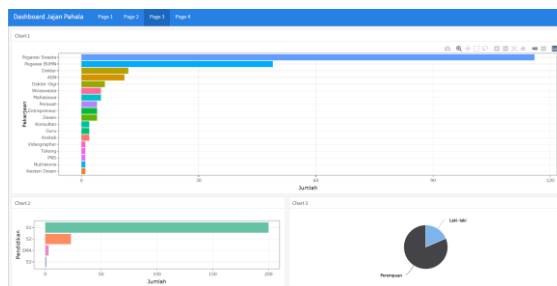
Random Forest. Hasil visualisasi dapat dilihat pada gambar 6, 7, 8, dan 9.



Gambar 6. Halaman pertama dashboard



Gambar 7. Halaman kedua dashboard



Gambar 8. Halaman ketiga dashboard

nama_donatur	jenis_kelamin	Pekerjaan	Pendidikan	Agama	gaji_Rutin	gaji_Tidak_Rutin	donasi	gaji_Utan
1	Male	Persewaan	SD	Islam	0,20	0,00	0,00	0,00
2	Male	Persewaan	SD	Islam	0,40	0,00	0,00	0,00
3	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
4	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
5	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
6	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
7	Male	Persewaan	SD	Islam	0,40	0,00	0,00	0,00
8	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
9	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
10	Male	Persewaan	SD	Islam	0,40	0,00	0,00	0,00
11	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
12	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
13	Male	Persewaan	SD	Islam	0,40	0,00	0,00	0,00
14	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
15	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00
16	Male	Persewaan	SD	Islam	0,30	0,00	0,00	0,00

Gambar 9. Halaman keempat dashboard

5. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan terhadap pertanyaan penelitian, maka dapat ditarik Kesimpulan pada analisis deskriptif yang telah dilakukan, didapatkan bahwa dari aspek bank yang digunakan donatur untuk mengirim donasi selain BSI adalah Bank BCA sebesar 15,4% dan Bank Mandiri sebesar 12,8%. Hal ini dapat dijadikan acuan jika Yayasan Jajan

Pahala ingin membuka rekening baru. Dari aspek selanjutnya yaitu kode unik donasi diperoleh kode 06 (Bagi-bagi Nasi) menjadi kegiatan yang paling banyak transaksi donasinya yaitu sebesar 52,5%. Aspek terakhir yang dilihat adalah persebaran donasi berdasarkan hari yang ditemukan bahwa Hari Jumat menjadi yang terbanyak dengan jumlah 58,8% sedangkan hari lainnya hanya 5,5% sampai 8,7%.

Hasil evaluasi model yang dipakai untuk klasifikasi kelas donatur pada penelitian ini cukup baik. Hal ini dapat dilihat pada pengujian cross validation model random forest menghasilkan akurasi sebesar 72,9% dengan parameter optimal pada mtry 13 dan min\_n 20. Untuk pengukuran evaluasi menggunakan nilai akurasi diperoleh nilai sebesar 78% yang berarti model membuat prediksi benar untuk 78% dari donatur rutin dan tidak rutin. Presisi (kelas “rutin”) didapatkan nilai sebesar 67% yang berarti dari semua donatur yang diprediksi sebagai donatur rutin, 67% benar donatur rutin. Nilai recall (kelas “rutin”) bernilai 57% yang berarti dari semua donatur rutin, model ini dapat mendeteksi 57%-nya. Nilai dari keseimbangan antara presisi dan recall yaitu f-measure didapatkan sebesar 61%.

Berdasarkan kesimpulan di atas, terdapat beberapa saran untuk diperbaiki dan dilanjutkan pada penelitian selanjutnya yaitu dapat mempertimbangkan untuk menambah jumlah data yang digunakan dengan tetap memperhatikan distribusi dari jumlah kelas. Hal ini dilakukan untuk meningkatkan jumlah data training pada model dan juga menghindari terjadinya imbalanced data dari kelas yang ada. Kemudian penelitian selanjutnya dapat menambah atribut dari informasi donatur. Hal ini dilakukan untuk mendapatkan hasil akurasi model yang lebih baik lagi. Sebagai contoh atribut yang dapat ditambah yaitu usia donatur dan range penghasilan donatur.

6. DAFTAR PUSTAKA

Amusa, L., North, D., & Zewotir, T. (2021). Optimal hyperparameter tuning of random forests for estimating causal treatment effects. *Songklanakarin Journal of Science & Technology*, 43(4).

Baskoro, B., Sriyanto, S., & Rini, L. S. (2021). Prediksi penerima beasiswa dengan menggunakan teknik data mining di Universitas Muhammadiyah Pringsewu. *In Prosiding Seminar Nasional Darmajaya*



- (Vol. 1, pp. 87-94).
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114, 24-31.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Duxbury Press.
- Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.
- Den Broeck, V., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *Africa Centre for Health and Population Studies*.
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1), 3133-3181.
- Friedman, V. (2008). Data visualization and infographics. Tersedia di <https://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/> [Diakses 7 Februari 2024].
- Gerson, R. F. (2001). *Marketing management*. Prentice Hall.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Isabona, J., Imoize, A. L., & Kim, Y. (2022). Machine learning-based boosted regression ensemble combined with hyperparameter tuning for optimal adaptive learning. *Sensors* (Basel, Switzerland), 22(10), 3776. <https://doi.org/10.3390/s22103776>.
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Kuhn, M., & Silge, J. (2022). *Tidy modeling with R*. O'Reilly Media.
- Kuhn, Max & Johnson, Kjell. (2013). *Applied Predictive Modeling*. 10.1007/978-1-4614-6849-3.
- Kudyba, S. (2014). *Big data, mining, and analytics*. Auerbach Publications.
- Mu'alim, F., & Hidayati, R. (2022). Implementasi metode random forest untuk penjurusan siswa di Madrasah Aliyah Negeri Sintang. *JUPITER (Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 14(1), 116-125. <https://doi.org/10.5281/4588/5.jupiter.2022.04>.
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi untuk diagnosa diabetes menggunakan metode Bayesian regularization neural network (RBNN). *Jurnal Informatika*, 11(1), 36.
- Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 682-692. <https://doi.org/10.1109/TVCG.2018.2864903>.
- Setyohadi, D., Kristiawan, F., & Ernawati, E. (2017). Perbaikan performansi klasifikasi dengan preprocessing iterative partitioning filter algorithm. *Telematika: Jurnal Informatika dan Teknologi Informasi*, 14(1), 12-20. <https://doi.org/10.31315/telematika.v14i01.1960>.
- Speiser, J. L., Durkalski, V. L., & Lee, W. M. (2015). Random forest classification of etiologies for an orphan disease. *Statistics in medicine*, 34(5), 887-899.
- Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*.
- Viegas, F., & Wattenberg, M. (2011). How to make data look sexy. Tersedia di [https://web.archive.org/web/20110506065701/http://articles.cnn.com/2011-04-19/opinion/sexy.data\\_1\\_visualization-21st-century-engagement?\\_s=PM%3AOPINION](https://web.archive.org/web/20110506065701/http://articles.cnn.com/2011-04-19/opinion/sexy.data_1_visualization-21st-century-engagement?_s=PM%3AOPINION) [Diakses 7 Februari 2024].
- Zailani, A. U., & Hanun, N. L. (2020). Penerapan algoritma klasifikasi random forest untuk penentuan kelayakan pemberian kredit di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, 6(1), 7-14.