

Perbandingan *Naive Bayes* dan *Random Forest* untuk Prediksi Perilaku Peserta Program Rujuk Balik

Widdi Djatmiko¹, Kusrini², Hanafi³

^{1,2,3}Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

¹widdi.dj@students.amikom.ac.id, ²kusrini@amikom.ac.id, ³hanafi@amikom.ac.id

Abstract

Program Rujuk Balik (PRB) is a service provided to participants of BPJS Kesehatan who have a history of chronic diseases, including diabetes mellitus, hypertension, heart disease, asthma, Chronic Obstructive Pulmonary Disease (COPD), epilepsy, stroke, schizophrenia, systemic lupus erythematosus. Participants who participate in the Program Rujuk Balik (PRB) already have stable/controlled health conditions but still require treatment. Participants registered in the Program Rujuk Balik (PRB) are required to make regular (active) monthly visits to the designated Fasilitas Kesehatan Tingkat Pertama (FKTP) as chosen by the Members. However, it has been observed that some Participants still exhibit a passive behavior by not being willing to make regular visits because they feel healthy and not currently ill. The objective of this study is to predict the behavior of Participants in the Program Rujuk Balik (PRB) and compare the performance of the Naive Bayes and Random Forest algorithms. The dataset used is private data obtained from BPJS Kesehatan Branch Office in Balikpapan. In this study, various data splits were performed, including 70:30, 80:20, and 90:10. The final results of this study indicate that the Random Forest algorithm outperforms Naive Bayes, with the highest values in the 70:30 data split technique, with an accuracy of 97.52%, precision of 96.56%, and recall of 98.71%.

Keywords: predictions, PRB, health, Naive Bayes, Random Forest

Abstrak

Program Rujuk Balik (PRB) merupakan pelayanan yang diberikan kepada Peserta BPJS Kesehatan yang memiliki riwayat penyakit kronis antara lain penyakit diabetes melitus, hipertensi, jantung, asma, Penyakit Paru Obstruktif Kronis (PPOK), *epilepsy*, *stroke*, *schizophrenia*, *systemic lupus erythematosus*. Peserta yang mengikuti Program Rujuk Balik (PRB) telah memiliki kondisi kesehatan yang sudah stabil/terkontrol namun masih memerlukan pengobatan. Peserta yang terdaftar dalam Program Rujuk Balik (PRB) diwajibkan untuk melakukan kunjungan rutin (aktif) setiap bulan ke Fasilitas Kesehatan Tingkat Pertama (FKTP) sesuai dengan yang dipilih oleh Peserta. Namun masih didapati sebagian Peserta masih memiliki perilaku tidak bersedia melakukan kunjungan rutin (pasif) dikarenakan Peserta masih merasa dalam kondisi sehat dan tidak dalam keadaan sedang sakit. Tujuan dalam penelitian ini akan melakukan prediksi perilaku Peserta Program Rujuk Balik (PRB) dan akan membandingkan performa dari algoritma *Naive Bayes* dan *Random Forest*. Dataset yang digunakan bersifat *private data* yang diambil dari BPJS Kesehatan Kantor Cabang Balikpapan. Pada penelitian ini dilakukan variasi *split* data antara lain 70:30, 80:20, 90:10. Hasil akhir dari penelitian ini yaitu algoritma *Random Forest* memiliki yang lebih baik dibandingkan dengan *Naive Bayes* dengan nilai tertinggi pada teknik *split* data 70:30 pada nilai *accuracy* sebesar 97,52%, *precision* sebesar 96,56%, *recall* sebesar 98,71%.

Kata kunci: Prediksi, PRB, Kesehatan, *Naive Bayes*, *Random Forest*

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Program Jaminan Kesehatan Nasional (JKN) merupakan program yang dapat memberikan jaminan perlindungan kesehatan bagi seluruh masyarakat Indonesia sehingga masyarakat berhak mendapatkan manfaat yang perlindungan kesehatan [1]. BPJS Kesehatan adalah Badan Hukum Publik yang ditunjuk oleh Pemerintah sebagai organisasi yang menyelenggarakan program Jaminan Kesehatan Nasional (JKN) bagi seluruh rakyat Indonesia [2]. Masyarakat yang telah terdaftar sebagai Peserta dan telah membayar iuran atau iurannya dibayar oleh pemerintah maka berhak mendapatkan kebutuhan dasar kesehatan yang dapat terpenuhi secara layak. Program manfaat yang didapatkan oleh masyarakat yang telah terdaftar sebagai BPJS Kesehatan yaitu pelayanan promotif dan preventif. Salah satu upaya program unggulan yang dijalankan untuk memenuhi kewajiban

pelaksanaan program promotif dan preventif tersebut yaitu Program Rujuk Balik (PRB) [3].

Program Rujuk Balik (PRB) merupakan pelayanan yang diberikan kepada Peserta BPJS Kesehatan yang memiliki riwayat penyakit kronis antara lain penyakit diabetes melitus, hipertensi, jantung, asma, Penyakit Paru Obstruktif Kronis (PPOK), *epilepsy*, *stroke*, *schizophrenia*, *systemic lupus erythematosus*. Peserta yang mengikuti Program Rujuk Balik (PRB) telah memiliki kondisi kesehatan yang sudah stabil/terkontrol namun masih memerlukan pengobatan untuk jangka panjang. Pemberian pelayanan Program Rujuk Balik (PRB) dilakukan di Fasilitas Kesehatan Tingkat Pertama (FKTP) rekomendasi/rujukan balik dari dokter spesialis/sub spesialis dokter di Fasilitas Kesehatan Rujukan Tingkat Lanjut (FKRTL) telah mendaftarkan Peserta untuk menjadi peserta rujuk balik [4].

Proses penyembuhan penderita penyakit kronis dipengaruhi oleh perilaku Peserta menjalani terapi pengobatan seperti rutin pemeriksaan kondisi kesehatan dan memastikan secara rutin konsumsi obat sesuai dengan resep yang diterbitkan oleh dokter. Perilaku Peserta yang tidak patuh dalam pemeriksaan kondisi kesehatan di fasilitas kesehatan dan pengambilan obat yang sudah diresepkan tentu saja akan berdampak buruk dan menjadi penghambat kesembuhan dalam proses terapi pengobatan [5].

Peserta yang terdaftar dalam Program Rujuk Balik (PRB) diwajibkan untuk melakukan kunjungan rutin (aktif) setiap bulan ke FKTP sesuai dengan yang dipilih oleh Peserta [6]. Hal ini dilakukan agar Peserta dapat memeriksakan kondisi kesehatan dan melakukan pengambilan obat sesuai dengan dosis dan jadwal yang sudah ditetapkan oleh dokter spesialis di FKRTL. Sebagian Peserta masih memiliki perilaku tidak bersedia melakukan kunjungan rutin (pasif) dikarenakan Peserta masih merasa dalam kondisi sehat dan tidak dalam keadaan sedang sakit [7].

Kondisi kesehatan Peserta yang tidak rutin berkunjung (pasif) akan semakin memburuk dan tidak dapat tertangani dengan baik. Perilaku Peserta PRB yang cenderung pasif dalam pemeriksaan kesehatan dan pengambilan obat akan menyebabkan proses pengobatan menjadi terhambat dan memiliki dampak risiko peningkatan jumlah penyakit, pasien rawat inap, dan tingginya angka kematian [8]. Berdasarkan adanya risiko dampak buruk yang bisa terjadi dan dapat merugikan bagi Peserta akibat adanya pola perilaku kunjungan peserta yang terdaftar dalam PRB maka penelitian ini bertujuan untuk prediksi pola perilaku Peserta PRB menggunakan pendekatan algoritma *machine learning*.

Dalam penelitian ini, penulis akan melakukan eksperimen berupa aktifitas prediksi perilaku Peserta Program Rujuk Balik (PRB) dan akan membandingkan performa dari algoritma *Naive Bayes* dan *Random Forest* dengan menggunakan dataset yang tersedia di BPJS Kesehatan Kantor Cabang Balikpapan. Pada beberapa penelitian sebelumnya telah dibahas penggunaan algoritma *Naive Bayes* dan *Random Forest* dalam upaya memberikan pengambahan ilmu pengetahuan baru berupa hasil dan kesimpulan atas berbagai permasalahan yang muncul.

Pada penelitian yang dilakukan oleh Efayanti, telah melakukan pengukuran kepatuhan dan keaktifan Peserta PRB dengan menggunakan 82 sampel responden untuk mendapatkan informasi. Berdasarkan studi tersebut didapatkan hasil yang menunjukkan tingkat kepatuhan Peserta Program PRB yang aktif melakukan kunjungan pemeriksaan kesehatan dan pengambilan obat rujuk balik yaitu sebesar 62,20% Peserta. Sedangkan masih terdapat 37,80% Peserta yang dinyatakan pasif terhadap kunjungan pemeriksaan kesehatan dan pengambilan obat rujuk balik [9].

Selanjutnya penelitian yang dilakukan oleh Maulida, dkk melakukan pengambilan sampel terhadap 53 Peserta yang terdaftar dalam Program PRB untuk mengetahui tingkat keaktifan pola perilaku Peserta Program PRB. Berdasarkan studi tersebut didapatkan hasil tingkat keaktifan Peserta Program PRB yaitu mencapai 34%. Sedangkan masih terdapat 66% Peserta yang dinyatakan pasif terhadap kunjungan pemeriksaan kesehatan dan pengambilan obat rujuk balik [10].

Pola perilaku Peserta yang aktif dan pasif berkunjung di Fasilitas Kesehatan dipengaruhi oleh faktor demografis dan riwayat medis Peserta Program PRB [11]. Pada beberapa studi telah dilakukan penelitian tentang prediksi kunjungan pasien di fasilitas kesehatan dengan menggunakan pendekatan model *machine learning*. Penelitian yang dilakukan oleh Fan, dkk melakukan prediksi tingkat keaktifan atau kehadiran pasien yang berkunjung di fasilitas kesehatan untuk mendapatkan pelayanan kesehatan menggunakan *machine learning* antara lain *logistic regression*, *k-nearest neighbor (KNN)*, *boosting*, *decision tree*, *random forest*, dan *bagging*. Hasil pengujian didapatkan akurasi dari setiap model tersebut yaitu *logistic regression*=68.79%, *k-nearest neighbor (KNN)*=90.14%, *boosting*=99.44%, *decision tree*=80.06%, *random forest*=96.92%, dan *bagging*=98.50% [12].

Selanjutnya pada penelitian Alaidah, dkk melakukan penelitian terkait dengan prediksi kunjungan medis dan ketidakhadiran pasien di fasilitas kesehatan menggunakan pendekatan *machine learning* antara lain menggunakan metode *Logistic Regression*, *Support Vector Machines*, *Naïve Bayes*, *Decision Tree* dan *Random Forest*. Data yang digunakan untuk pemodelan sistem yaitu berupa dataset demografi pasien, riwayat medis pasien, lokasi wilayah tempat tinggal, dan informasi kehadiran kunjungan pasien. Hasil dari penelitian ini menunjukkan tingkat performa dari metode *Random Forest* memiliki *recall* tertinggi diantara metode lainnya yaitu mencapai nilai 100%. Metode *Random Forest* dan *Naïve Bayes* umumnya digunakan untuk pemecahan permasalahan dengan menggunakan pendekatan *machine learning* [13].

Pada penelitian yang dilakukan oleh Tugiman, dkk melakukan penelitian terkait prediksi penggunaan obat yang dikonsumsi oleh pasien menggunakan metode *Naïve Bayes*. Penelitian bertujuan untuk peningkatan jumlah kunjungan pasien ke Fasilitas Kesehatan untuk mendapatkan pelayanan kesehatan. Para peneliti berupaya untuk prediksi ketersediaan obat yang membagi dalam 2 klas yaitu cukup dan tidak cukup. Hasil dari penelitian ini memiliki tingkat performa akurasi sebesar 50%, presisi sebesar 50% dan *recall* sebesar 100% [14].

Metode *Random Forest* memiliki keunggulan yang dapat menunjukkan hasil yang lebih baik pada dataset asli dan sebagian besar dataset seimbang [15]. Selain itu, metode *Random Forest* dapat melakukan beberapa

pemilihan fitur secara acak selanjutnya mengurutkan setiap node sehingga dapat menghasilkan nilai tingkat error atau kesalahan yang relatif lebih rendah [16]. Sementara *Naïve Bayes* memiliki kecepatan dalam proses perhitungan karena pemodelan mudah diimplementasikan dalam pendekatan *machine learning* [17]. Selain itu, metode *Naïve Bayes* tidak terlalu banyak membutuhkan data *training* untuk dapat menentukan class pada tahapan klasifikasi [18].

Dari beberapa uraian di atas, penulis tertarik mengajukan penelitian menggunakan algoritma *Naive Bayes* dan *Random Forest* pada Peserta PRB sebagai objek penelitian. Hasil penelitian ini diharapkan memberikan gambaran kepada BPJS Kesehatan untuk mengetahui pola perilaku Peserta PRB sehingga dapat membangun pengetahuan dan memberikan penilaian akurasi performa kedua model algoritma tersebut.

2. Metode Penelitian

Naive Bayes dapat digunakan untuk berbagai macam kebutuhan antara lain untuk klasifikasi, prediksi, pemodelan probabilitas, pengenalan pola dan beberapa permasalahan yang terkait dengan klasifikasi lainnya [19]. Sementara algoritma *Random Forest* merupakan model Algoritma *decision tree* yang dikembangkan dengan menggunakan *decision tree* yang tersusun lebih banyak, dimana pada *decision tree* sebelumnya dilakukan proses training dengan sampel individu dan setiap atribut dibagi pada *tree* yang dipilih antara atribut yang bersifat acak [20].

2.1. Naïve Bayes

Algoritma *Naive Bayes* merupakan sebuah proses klasifikasi dengan metode pemeriksaan pada tingkat kemunculan data yang sering ditemukan dalam data pelatihan yang menghitung sekumpulan kemungkinan untuk mencari peluang terbesar dari probabilitas klasifikasi [21]. Algoritma *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan data yang besar [22]. Model persamaan *Naive Bayes* yaitu sebagai berikut:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (1)$$

Keterangan:

Y : Data dengan kelas yang belum diketahui

X : Hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$: Probabilitas hipotesis x berdasarkan kondisi y (posteriori probability)

$P(x)$: Probabilitas hipotesis x (prior probability)

$P(y|x)$: Probabilitas y berdasarkan kondisi pada hipotesis x

$p(y)$: Probabilitas dari y

2.2. Random Forest

Algoritma *Random Forest* merupakan pengembangan dari *decision tree* atau berbasis pohon keputusan.

Metode *Random Forest* akan melakukan aktifitas pembuatan pohon keputusan dalam jumlah yang cukup besar berdasarkan data yang sudah dilakukan proses training [23]. *Random Forest* merupakan model klasifikasi yang menggunakan pendekatan pohon terstruktur. Pentahapan penggunaan metode *Random Forest* dapat dijelaskan dengan langkah sebagai berikut:

- Pemilihan Dataset: tentukan dataset yang akan digunakan dalam proses klasifikasi.
- Bootstrap samples*: dataset yang telah dikumpulkan akan dibentuk dalam beberapa sample acak dengan variasi penggantian.
- Pohon keputusan: pada tahap ini, sample yang telah dilakukan pengolahan akan dibangun pohon keputusan dengan opsi pemilihan variabel yang terbaik berdasarkan kriteria yang telah ditentukan.
- Penggabungan pohon keputusan: pohon keputusan yang telah melewati proses pengolahan selanjutnya akan digabungkan untuk mengambil hasil mayoritas sebagai keputusan akhir [24]

2.3. Dataset

Dataset yang digunakan pada penelitian ini bersumber dari lembaga BPJS Kesehatan Kantor Cabang Balikpapan sesuai dengan topik yang telah ditentukan yaitu riwayat data Peserta yang terdaftar dalam Program Rujuk Balik (PRB). Selain melakukan pengambilan dataset Peserta PRB, penulis melakukan metode wawancara yang digunakan untuk memperoleh keterangan, informasi dan penjelasan lebih mendalam terkait dengan permasalahan yang akan dibahas dalam penelitian ini sehingga diperoleh data yang akurat dan terpercaya yang bersumber dari BPJS Kesehatan Kantor Cabang Balikpapan. Jumlah dataset yang terkumpul yaitu sebanyak 8.575 Data yang terdiri dari 11 atribut.

Tabel 1. Atribut Dataset Peserta Program PRB

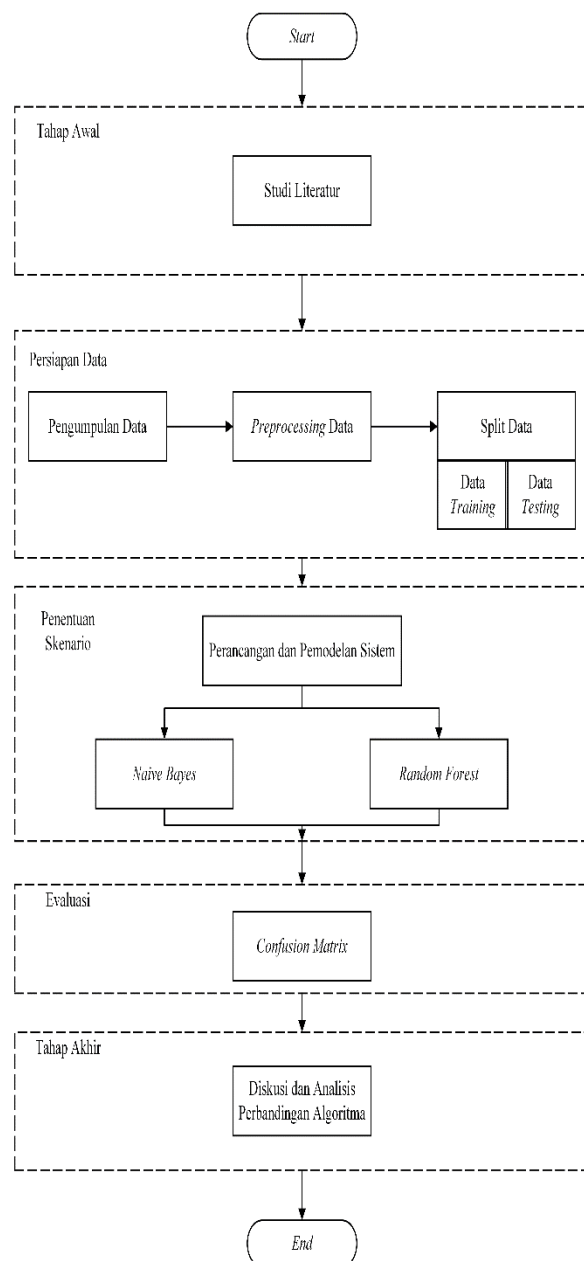
No	Atribut	Deskripsi	Value
1	Dati2	Lokasi Kabupaten Kota Peserta PRB	Kota Balikpapan Kab. Pasir Utara Kab. Pasir Kab. Berau
2	Usia Peserta	Usia Peserta PRB	8 – 99 tahun
3	Jenis FKTP	Jenis FKTP yang dikunjungi oleh Peserta PRB	Dokter Umum Klinik Pratama Puskesmas
4	Kepemilikan FKTP	Pengelola FKTP yang dikunjungi oleh Peserta PRB	BUMN PemKab POLRI Swasta TNI

No	Atribut	Deskripsi	Value
5	Diagnosa	Jenis penyakit Peserta PRB	Asthma Diabetes Mellitus Epilepsi Hypertensi Penyakit Jantung PPOK (Penyakit Paru Obstruktif Kronik) Schizophrenia Stroke Systemic Lupus Erythematosus
6	HubKel	Hubungan anggota keluarga Peserta PRB	P (Peserta) I (Istri) S (Suami) A (Anak) T (Tambahan)
7	Progprolanis	Peserta terdaftar sebagai Prolanis (Program Pengelolaan Penyakit Kronis)	Ya Tidak
7	Kelas	Jenis kelas rawat Peserta PRB	1 2 3
8	Jenis Kelamin	Jenis kelamin Peserta PRB	Laki Laki Perempuan
10	Jenis Peserta	Segmen Peserta PRB terdaftar	PBI Non PBI
11	Status PRB	Status perilaku Peserta PRB	Aktif Pasif

Data yang telah berhasil dikumpulkan selanjutnya akan ditentukan kelas atau label dari dataset Peserta PRB yaitu atribut atau variabel yang menjelaskan status PRB. Label status PRB akan memberikan informasi apakah Peserta rutin berkunjung ke Fasilitas Kesehatan akan dilabeli dengan atribut “Aktif” dan Peserta yang tidak rutin berkunjung ke Fasilitas Kesehatan akan dilabeli dengan atribut “Pasif”. Dataset Peserta PRB yang digunakan berdasarkan atribut yang tersedia akan menjadi dasar faktor penentu perilaku Peserta yang berkunjung di Fasilitas Kesehatan.

2.4. Alur Penelitian

Pada bagian ini akan dijelaskan mengenai alur penelitian yang disusun secara lengkap dan sistematis sehingga penulisan penelitian dapat dilakukan dengan baik dan benar sesuai penjelasan pada Gambar 1.



Gambar 1. Alur Penelitian

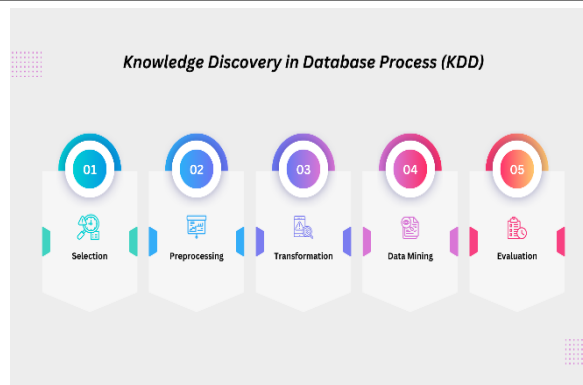
1. Tahap awal akan dilakukan proses penelusuran informasi secara detail dan mendalam melalui studi literatur dengan upaya pencarian referensi dari berbagai sumber antara lain jurnal, buku ilmiah, tesis yang relevan berdasarkan permasalahan yang akan diteliti sebagai bahan rujukan dalam menentukan metode yang tepat dan sesuai untuk penyelesaian permasalahan.
2. Persiapan data akan dilakukan proses tahapan aktifitas yang akan memudahkan penulis dalam melakukan penelitian.
 - a. Proses pengumpulan dataset yang bersumber dari lembaga BPJS Kesehatan Kantor Cabang Balikpapan sesuai dengan topik yang telah ditentukan yaitu riwayat data Peserta yang terdaftar dalam Program Rujuk Balik (PRB). Data yang berhasil terkumpul yaitu sebanyak

8.575 dataset Peserta yang terdaftar dalam Program Rujuk Balik (PRB). Data Peserta yang berhasil terkumpul akan dibangun model untuk mengklasifikasikan Peserta PRB menjadi 2 *class* yaitu Peserta Aktif atau Pasif. Berdasarkan dari *class* yang telah ditentukan tersebut akan diperoleh informasi Peserta melakukan upaya pemeriksaan kondisi kesehatan dan pengambilan obat secara rutin. Proses labeling akan dilakukan memberikan label berdasarkan data demografi dan kondisi diagnosa penyakit medis dari Peserta PRB.

- b. Tahap *preprocessing* data untuk dilakukan pemeriksaan dan perbaikan yang berpotensi tidak sesuai atau tidak dibutuhkan (data *cleansing*) pada data yang akan digunakan pada penelitian.
 - c. Pembagian data akan dilakukan menjadi beberapa eksperimen untuk mengetahui tingkat performa terbaik antara lain dengan rasio 70% data testing dan 30% data training, 80% data testing dan 20% data training, 90% data testing dan 10% data training.
3. Pada tahap penentuan skenario, penulis akan melakukan aktifitas klasifikasi terhadap dataset yang telah dikumpulkan dengan melakukan pemodelan dan membanding 2 metode menggunakan algoritma *Naive Bayes* dan *Random Forest*.
 4. Pada tahap evaluasi akan dilakukan proses pengujian kinerja pada setiap algoritma terhadap dataset dan *split* data yang telah dilakukan eksperimen dengan nilai luaran berupa *confusion matrix* antara lain *accuracy*, *precision*, dan *recall* sebagai perbandingan tingkat performa untuk kedua algoritma.
 5. Pada tahap akhir akan dilakukan diskusi dan analisa perbandingan kinerja untuk kedua algoritma. Pada proses ini akan disampaikan metode yang memiliki tingkat performa terbaik prediksi pola perilaku Peserta Program Rujuk Balik (PRB) BPJS Kesehatan.

3. Hasil dan Pembahasan

Dalam proses *data mining* dibutuhkan konsep metode pengolahan data dalam bentuk pentahapan pemecahan masalah yang biasa disebut *Knowledge Discovery in Database Process (KDD)*. *Knowledge Discovery in Database Process (KDD)* merupakan suatu rangkaian aktifitas yang dilakukan oleh peneliti untuk mendapatkan informasi dengan yang dibutuhkan dengan menggunakan pendekatan algoritma tertentu berdasarkan dataset yang sudah didapatkan [25]. Dalam proses *KDD*, terdapat beberapa tahapan penting untuk melakukan *data mining* antara lain *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, *Evaluation*.



Gambar 2. Knowledge Discovery in Database Process (KDD)

3.1. Selection

Dataset yang dibutuhkan dalam penelitian ini yaitu sekumpulan data yang diperoleh dari BPJS Kesehatan Kantor Cabang Balikpapan yang memuat informasi berupa riwayat catatan peserta yang terdaftar dalam program PRB. Terdapat 11 atribut dalam dataset yang akan digunakan dalam penelitian ini. Penggunaan atribut yang relevan dalam penelitian akan bermanfaat dan memberikan kontribusi dalam analisis data. Seluruh atribut memiliki hubungan keterkaitan yang akan menentukan kondisi perilaku peserta PRB dalam berkunjung di Fasilitas Kesehatan.

3.2. Preprocessing

Pada tahapan *Preprocessing* akan dilakukan proses analisa jumlah data yang tidak tersedia pada setiap atribut sehingga dapat dilakukan penanganan data yang hilang (*missing values*). Data yang hilang (*missing values*) akan mengakibatkan data tidak dapat diproses karena tidak memiliki nilai yang dapat dilakukan proses perhitungan. Penanganan terhadap data yang kosong dapat dilakukan dengan menghapus baris tidak relevan dan data yang dianggap anomali sehingga menyisakan data yang valid untuk dijadikan sebagai data penentu yang akan digunakan dalam perhitungan. Hasil proses *missing values* sebagaimana tercantum pada tabel 2.

Tabel 2. Hasil proses *missing values*

No	Atribut	Missing Values
1	Dati2	0%
2	Usia Peserta	0%
3	Jenis FKTP	0%
4	Kepemilikan FKTP	0%
5	Diagnosa	0%
6	PISAT	0%
7	Progprolanis	0%
8	Kelas	0%
9	Jenis Kelamin	0%
10	Jenis Peserta	0%
11	Status PRB	0%

3.3. Transformation

Selanjutnya dilakukan proses transformasi data untuk mengubah data menjadi format yang sesuai untuk analisis. Transformasi data bertujuan untuk menjadikan data dapat lebih berkualitas dan memudahkan dalam proses perhitungan. Setelah seluruh tahapan *transformation* telah dilakukan, maka akan didapatkan data yang layak untuk digunakan sebagai variabel dalam penelitian ini. Contoh dataset yang telah berhasil dilakukan dikumpulkan dan transformasi data yaitu sebagaimana yang tercantum dalam tabel 3.

Tabel 3. Pembagian data *training* dan data *testing*

No	Atribut	P1	P2	...	P8575
1	Dati2	Kab. Berau	Kab. Pasir	...	Kota Balikpapan
2	Usia Peserta	57	58	...	31
3	Jenis FKTP	Klinik Pratama	Klinik Pratama	...	Klinik Pratama
4	Kepemilikan FKTP	Pemkab	Swasta	...	Swasta
5	Diagnosa	Penyakit Jantung	Diabetes Melitus	...	Diabetes Melitus
6	PISAT	P	P	...	T
7	Prognosis	Tidak	Ya	...	Ya
8	Kelas	1	1	...	2
9	Jenis Kelamin	Perempuan	Laki Laki	...	Perempuan
10	Jenis Peserta	Non PBI	Non PBI	...	Non PBI
11	Status PRB	Pasif	Pasif	...	Aktif

Berdasarkan data pada Tabel 3 di atas, diketahui terdapat data sebanyak 8.575 yang berhasil dikumpulkan dari BPJS Kesehatan Kantor Cabang Balikpapan. Dataset tersebut mencakup 11 atribut yang terdiri dari data Peserta yang terdaftar dalam Program Rujuk Balik (PRB) yang memuat informasi kondisi wilayah (dati2), demografi, kelas rawat, diagnosa penyakit, dan lain sebagainya. Atribut Status PRB yang memuat informasi aktif dan pasif akan ditentukan sebagai *class* pada proses prediksi perilaku Peserta PRB. Dalam melakukan proses klasifikasi perilaku Peserta PRB, dibutuhkan informasi yang cukup memadai dan mendukung dalam konteks pembuatan model pembelajaran mesin. Pemilihan atribut di atas merupakan informasi data Peserta yang dapat bermanfaat untuk membantu memahami interaksi dan variasi antara atribut dalam menentukan label dari dataset dengan jumlah besar yang kompleks dan beragam.

3.4. Data Mining (Tahap Pemodelan)

Pada penelitian ini akan dilakukan proses *split* data menjadi 2 bagian yaitu data *training* dan data *testing*. Variasi pembagian data akan dilakukan menjadi beberapa bagian antara lain dengan rasio 70% data

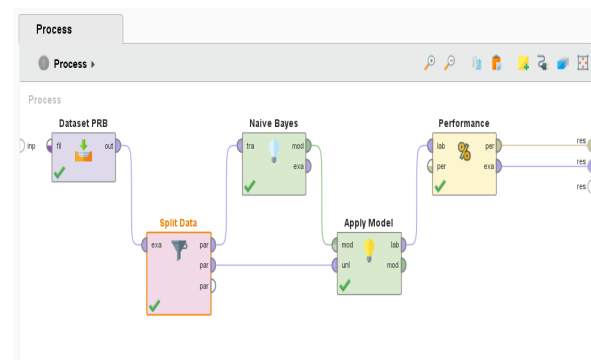
training dan 30% data *testing*, 80% data *training* dan 20% data *testing*, 90% data *training* dan 10% data *testing*. Variasi pembagian data ini merupakan upaya eksperimen untuk mengetahui teknik *split* data yang memiliki performa terbaik. Jumlah komposisi pembagian data sebagaimana tercantum pada tabel 4 di bawah ini.

Tabel 4. Pembagian data *training* dan data *testing*

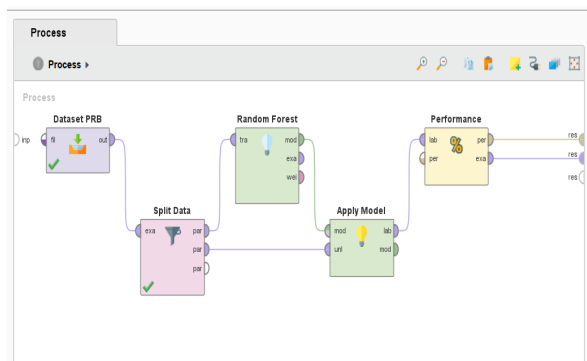
No	% Training	% Testing	Jumlah Data Training	Jumlah Data Testing
1	70%	30%	6.002	2.573
2	80%	20%	6.860	1.715
3	90%	10%	7.718	857

Setelah ditentukan data yang telah dilakukan pembagian (*split* data), tahapan selanjutnya yaitu masuk dalam tahap proses pemodelan data dengan menggunakan algoritma yang sudah dipilih antara lain *Naive Bayes* dan *Random Forest*. Pemodelan data menggunakan *tools Rapid Miner* untuk melakukan proses seluruh rangkaian dalam penelitian ini. *Rapid Miner* merupakan alat bantu (*tools*) dalam melakukan proses analisa untuk data *mining*, *machine learning*, dan *predictive analytics* [26]. Pengguna *Rapid Miner* dapat memanfaatkan banyak fitur *machine learning* dengan metode data *mining* yang cukup variatif.

Proses pemodelan mining menggunakan *Naive Bayes* dan *Random Forest* yaitu melalui tahapan *entry dataset* melalui *tools Rapid Miner*. Dataset tersebut meliputi seluruh atribut yang memiliki hubungan keterkaitan dengan label yang telah ditentukan. Pada penelitian ini label yang akan ditentukan yaitu pada atribut Status PRB. Label tersebut akan menjadi hasil prediksi yang akan menentukan apakah Peserta yang terdaftar dalam program PRB memiliki perilaku aktif atau pasif dalam berkunjung secara rutin ke Fasilitas Kesehatan untuk memeriksa kondisi kesehatan dan melakukan pengambilan obat yang telah diresepkan. Pemodelan data pada algoritma *Naive Bayes* dan *Random Forest* dapat dilihat pada gambar 3 dan gambar 4 dibawah ini.



Gambar 3. Pemodelan *Naive Bayes*



Gambar 4. Pemodelan *Random Forest*

3.5. Evaluation

Hasil pengujian performa terhadap algoritma *Naive Bayes* dan *Random Forest* menggunakan pendekatan *confusion matrix*. *Confusion matrix* umumnya banyak digunakan untuk melakukan menguji performa suatu model dengan cara membandingkan antara data aktual dengan data prediksi [27]. Pengukuran kinerja suatu model menggunakan *Confusion Matrix* dapat dimanfaatkan untuk menentukan kinerja nilai pemodelan algoritma klasifikasi berdasarkan hasil perhitungan nilai akurasi yang diperoleh dari jumlah prediksi kasus yang benar dan jumlah prediksi yang salah. Pada penelitian ini akan dilakukan skenario pembagian data (*split data*) sebanyak 3 variasi data training (70%, 80%, 90%) untuk mengetahui performa terbaik pada setiap model yang akan dilakukan pengujian.

3.5.1 *Naive Bayes* split data 70:30

Pemodelan data menggunakan *Naive Bayes* pada split data 70:30 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 5 di bawah ini.

accuracy: 87.84%			
	true PASIF	true AKTIF	class precision
pred. PASIF	1055	125	89.41%
pred. AKTIF	188	1205	86.50%
class recall	84.88%	90.60%	

precision: 86.50% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	1055	125	89.41%
pred. AKTIF	188	1205	86.50%
class recall	84.88%	90.60%	

recall: 90.60% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	1055	125	89.41%
pred. AKTIF	188	1205	86.50%
class recall	84.88%	90.60%	

Gambar 5. *Confusion Matrix Naive Bayes* split data 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 5, performa model *Naive Bayes* pada split data 70:30 menghasilkan nilai *Accuracy* = 87.84%, *Precision* = 86.50%, *Recall* = 90.60%.

3.5.2 *Naive Bayes* split data 80:20

Pemodelan data menggunakan *Naive Bayes* pada split data 80:20 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 6 di bawah ini.

accuracy: 88.34%			
	true PASIF	true AKTIF	class precision
pred. PASIF	699	70	90.90%
pred. AKTIF	130	816	86.26%
class recall	84.32%	92.10%	

precision: 86.26% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	699	70	90.90%
pred. AKTIF	130	816	86.26%
class recall	84.32%	92.10%	

recall: 92.10% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	699	70	90.90%
pred. AKTIF	130	816	86.26%
class recall	84.32%	92.10%	

Gambar 6. *Confusion Matrix Naive Bayes* split data 80:20

Berdasarkan hasil pengujian yang tercantum pada gambar 6, performa model *Naive Bayes* pada split data 80:20 menghasilkan nilai *Accuracy* = 88.34%, *Precision* = 86.26%, *Recall* = 92.10%.

3.5.3 *Naive Bayes* split data 90:10

Pemodelan data menggunakan *Naive Bayes* pada split data 90:10 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 7 di bawah ini.

accuracy: 87.86%			
	true PASIF	true AKTIF	class precision
pred. PASIF	348	38	90.16%
pred. AKTIF	66	405	85.99%
class recall	84.06%	91.42%	

precision: 85.99% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	348	38	90.16%
pred. AKTIF	66	405	85.99%
class recall	84.06%	91.42%	

recall: 91.42% (positive class: AKTIF)			
	true PASIF	true AKTIF	class precision
pred. PASIF	348	38	90.16%
pred. AKTIF	66	405	85.99%
class recall	84.06%	91.42%	

Gambar 7. *Confusion Matrix Naive Bayes* split data 90:10

Berdasarkan hasil pengujian yang tercantum pada gambar 7, performa model *Naive Bayes* pada split data 90:10 menghasilkan nilai *Accuracy* = 87.86%, *Precision* = 85.99%, *Recall* = 91.42%.

3.5.4 *Random Forest* split data 70:30

Pemodelan data menggunakan *Random Forest* pada split data 70:30 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 8 di bawah ini.

accuracy: 97.52%

	true PASIF	true AKTIF	class precision
pred. PASIF	2791	40	98.59%
pred. AKTIF	109	3062	96.56%
class recall	96.24%	98.71%	

precision: 96.56% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	2791	40	98.59%
pred. AKTIF	109	3062	96.56%
class recall	96.24%	98.71%	

recall: 98.71% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	2791	40	98.59%
pred. AKTIF	109	3062	96.56%
class recall	96.24%	98.71%	

Gambar 8. Confusion Matrix Random Forest split data 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 8, performa model *Random Forest* pada split data 70:30 menghasilkan nilai *Accuracy* = 97.52%, *Precision* = 96.56%, *Recall* = 98.71%.

3.5.5 Random Forest split data 80:20

Pemodelan data menggunakan *Random Forest* pada split data 80:20 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 9 di bawah ini.

accuracy: 96.55%

	true PASIF	true AKTIF	class precision
pred. PASIF	3182	105	96.81%
pred. AKTIF	132	3441	96.31%
class recall	96.02%	97.04%	

precision: 96.31% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	3182	105	96.81%
pred. AKTIF	132	3441	96.31%
class recall	96.02%	97.04%	

recall: 97.04% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	3182	105	96.81%
pred. AKTIF	132	3441	96.31%
class recall	96.02%	97.04%	

Gambar 9. Confusion Matrix Random Forest split data 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 9, performa model *Random Forest* pada split data 80:20 menghasilkan nilai *Accuracy* = 96.55%, *Precision* = 96.31%, *Recall* = 97.04%.

3.5.5 Random Forest split data 90:10

Pemodelan data menggunakan *Random Forest* pada split data 90:10 menghasilkan *Confusion Matrix* yang terbagi menjadi 3 pengujian (*Accuracy*, *Precision*, *Recall*) seperti terlihat pada gambar 10 di bawah ini.

accuracy: 94.79%

	true PASIF	true AKTIF	class precision
pred. PASIF	3523	196	94.73%
pred. AKTIF	206	3793	94.85%
class recall	94.48%	95.09%	

precision: 94.85% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	3523	196	94.73%
pred. AKTIF	206	3793	94.85%
class recall	94.48%	95.09%	

recall: 95.09% (positive class: AKTIF)

	true PASIF	true AKTIF	class precision
pred. PASIF	3523	196	94.73%
pred. AKTIF	206	3793	94.85%
class recall	94.48%	95.09%	

Gambar 10. Confusion Matrix Random Forest split data 70:30

Berdasarkan hasil pengujian yang tercantum pada gambar 10, performa model *Random Forest* pada split data 90:10 menghasilkan nilai *Accuracy* = 94.79%, *Precision* = 94.85%, *Recall* = 95.09%.

Berikut hasil rekap keseluruhan pengujian performa *confusion matrix* dari model *Naive Bayes* dan *Random Forest* menggunakan *Rapid Miner* sebagaimana tercantum pada tabel 5 di bawah ini.

Tabel 5. Hasil perhitungan *Confusion Matrix*

Model	Split Data	Accuracy	Precision	Recall
Naive Bayes	70:30	87,84%	86,50%	90,60%
Naive Bayes	80:20	88,34%	86,26%	92,10%
Naive Bayes	90:10	87,86%	85,99%	91,42%
Random Forest	70:30	97,52%	96,56%	98,71%
Random Forest	80:20	96,55%	96,31%	97,04%
Random Forest	90:10	94,79%	94,85%	95,09%

Berdasarkan proses pengujian menggunakan *confusion matrix*, performa *Random Forest* memiliki keunggulan dalam pemodelan data dibandingkan dengan *Naive Bayes*. Hasil performa terbaik yaitu dengan menggunakan algoritma *Random Forest* pada pembagian data menggunakan 70% data *training* dan 30% data *testing*.

4. Kesimpulan

Setelah melakukan seluruh tahapan penelitian ini dengan menggunakan 2 *class* yang telah ditentukan yaitu Aktif atau Pasif menggunakan algoritma *Naive Bayes* dan *Random Forest*, maka dapat diberikan kesimpulan bahwa untuk melakukan prediksi pola perilaku Peserta Program Rujuk Balik (PRB) dapat menggunakan pendekatan data mining yang dikelola dengan baik. Penerapan metode *Random Forest* terbukti memiliki tingkat performa yang lebih baik apabila dibandingkan dengan metode *Naive Bayes*. *Random Forest* memiliki performa terbaik pada split

data 70:30 dengan nilai *Accuracy* = 97.52%, *Precision* = 96.56%, *Recall* = 98.71%.

Perilaku pasien yang terdaftar dalam Peserta Program Rujuk Balik (PRB) dapat diprediksi dengan pendekatan menggunakan *machine learning* yang didukung dengan dataset yang memadai dan informatif berupa analisis data historis dan pengolahan data demografis. Pemanfaatan *machine learning* dapat membangun model prediktif yang memanfaatkan berbagai variabel untuk meramalkan tingkat kepatuhan pasien terhadap perawatan bagi Peserta yang terdaftar dalam Program PRB.

Prediksi pola perilaku Peserta Program PRB merupakan salah satu upaya yang dapat direkomendasikan kepada BPJS Kesehatan dan tim medis di FKTP dan FKRTL dalam menyusun strategi yang efektif dengan upaya pendekatan komunikasi lebih mendalam untuk meningkatkan kepatuhan pasien khususnya bagi Peserta yang memiliki perilaku Pasif. Hal ini dapat berdampak buruk bagi kesehatan Peserta yang Pasif dan perlu mendapatkan perhatian serius terhadap perawatan dan tindakan medis yang direkomendasikan. Pola perilaku Peserta PRB yang telah diketahui lebih awal menggunakan *machine learning* dapat membantu pasien dan tenaga medis untuk merencanakan perawatan yang lebih personal dan sesuai dengan kebutuhan individu serta dapat mengambil tindakan pencegahan atau pengelolaan medis yang sesuai.

Berdasarkan hasil dari penelitian ini, maka dapat diharapkan untuk penelitian berikutnya dapat melakukan perluasan dengan menggunakan metode algoritma data mining lainnya dan pengembangan studi kasus yang memiliki jenis atribut yang lebih banyak.

Daftar Rujukan

- [1] Republik Indonesia. 2013. *Peraturan Menteri Kesehatan Republik Indonesia Nomor 71 Tahun 2013 Tentang Pelayanan Kesehatan pada Jaminan Kesehatan Nasional*. Kementerian Kesehatan. Jakarta
- [2] Pemerintah Indonesia. 2011. *Undang-Undang Republik Indonesia Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial*. Lembaran Negara Republik Indonesia Tahun 2011 Nomor 116. Sekretariat Negara. Jakarta.
- [3] Pemerintah Indonesia. 2018. *Perpres Nomor 82 Tahun 2018 tentang Jaminan Kesehatan*. Lembaran Negara Republik Indonesia Tahun 2018 Nomor 165. Sekretariat Negara. Jakarta.
- [4] BPJS Kesehatan. 2014. *Peraturan Badan Penyelenggara Jaminan Sosial Kesehatan Nomor 1 Tahun 2014 Tentang Penyelenggaraan Jaminan Kesehatan*. BPJS Kesehatan. Jakarta.
- [5] Dolovich L, Nair K, Sellors C, Lohfeld L, Lee A, Levine M. *Do patients' expectations influence their use of medications? Qualitative study*. Can Fam Physician. 2008 Mar;54(3):384-93. PMID: 18337533; PMCID: PMC2278356.
- [6] Republik Indonesia. 2013. *Peraturan Menteri Kesehatan Republik Indonesia Nomor 28 Tahun 2014 tentang Pedoman Pelaksanaan Program Jaminan Kesehatan Nasional*. Kementerian Kesehatan. Jakarta.
- [7] J. Keperawatan Abdurrahman, D. Sukma Dewi Arimbi, R. Lisa Indra, and P. Studi Keperawatan STIKes Hang Tuah Pekanbaru, *Pengaruh Pendidikan Kesehatan Terhadap Motivasi Mengontrol Kadar Gula Darah Pada Pasien DM Tipe II*. vol. 4, no. 1, pp. 2579–8723, 2020.
- [8] M. A. Fischer et al. *Primary medication non-adherence: Analysis of 195,930 electronic prescriptions*. J Gen Intern Med, vol. 25, no. 4, pp. 284–290, Apr. 2010, doi: 10.1007/s11606-010-1253-9.
- [9] Dwi Efayanti, "Analisis Faktor Yang Mempengaruhi Kepatuhan Pengambilan Obat Peserta Program Rujuk Balik Di Bandar Lampung," Jurnal Farmasi Indonesia, vol. 9, no. 1, 2017.
- [10] Maulida G, Asrinawaty, & Aquarista M F, "Faktor Yang Berhubungan dengan Ketidakepatuhan Peserta Program Rujuk Balik (PRB) di Wilayah Puskesmas Pelambuan Banjarmasin Tahun 2020," Jurnal Kesehatan Masyarakat, 2020.
- [11] Fitriyani N. E, Dewi I R, and Mersi Nawangsari, "Tingkat Kepatuhan Penggunaan Obat pada Pasien Diabetes Melitus Program Rujuk Balik Apotek Kimia Farma 437 di Kota Purwokerto Tipe 2," Jurnal Penelitian Sains dan Kesehatan Avicenna, vol. 2, no. 2, May 2023.
- [12] G. Fan, Z. Deng, Q. Ye, and B. Wang, "Machine learning-based prediction models for patients no-show in online outpatient appointments," Data Science and Management, vol. 2, pp. 45–52, Jun. 2021, doi: 10.1016/j.dsm.2021.06.002.
- [13] Albtoul Alaidah, Eman Alamoudi, Dauaa Shalabi, Malak AlQahtani, Hajar Alnamshan, and Nirase Fathima Abubacker, "Mining and Predicting No-Show Medical Appointments: Using Hybrid Sampling Technique," Communication and Intelligent Systems, vol. 204, 2021.
- [14] Tugiman, Lily Damayanti, Alexius Hendra Gunawan, and Samuel Ryon Elkana, "Prediksi Penggunaan Obat Peserta Jaminan Kesehatan Nasional Menggunakan Algoritma Naïve Bayes Classifier," JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST), vol. 3, no. 1, 2022.
- [15] R. Couronné, P. Probst, and A. L. Boulesteix, "Random forest versus logistic regression: A large-scale benchmark experiment," BMC Bioinformatics, vol. 19, no. 1, Jul. 2018, doi: 10.1186/s12859-018-2264-5.
- [16] Rusydi Umar, Imam Riadi, and Purwono, "Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial," JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 4, no. 2, 2020.
- [17] Liyana Febriyanti and Hadi Zakaria, "Implementasi Data Mining Untuk Memprediksi Produktivitas Pada Tanaman Kacang Tanah Menggunakan Metode Naive Bayes," LOGIC: Jurnal Ilmu Komputer dan Pendidikan, vol. 1, no. 2, pp. 105–118, 2023.
- [18] Hakam Febtadianrano Putro, Retno Tri Vulandar, and Wawan Laksito Yuly Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," Jurnal TIKomSiN, vol. 8, no. 2, Oct. 2020.
- [19] Rizky I I K M, Irianto S Y, Sriyanto. *Perbandingan Kinerja Algoritma Naive Bayes, Support Vector Machine dan Random forest untuk Prediksi Penyakit Ginjal Kronis*. Agustus 2023. Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat.
- [20] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 10, no. 3, pp. 432–437, Dec. 2021, doi: 10.32736/sisfokom.v10i3.1302.
- [21] R. Leonardo and J. Pratama, "Address: Universitas Prima Indonesia, Teknik Informatika," Jl. Sekip Sei Kambing Medan, no. 123, 2011, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Bank+Ma>
- [22] Kusri and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: CV. Andi Offset, 2009.
- [23] Q. Raihananda et al., "Application of Random Forest Method Classification to Predict BPJS Kesehatan Card Users Who Receive Contribution Assistance in Karangasem District, Bali Province 2017," Jurnal Matematika, Statistika dan Komputasi, vol. 17, no. 2, pp. 178–188, Dec. 2020, doi: 10.20956/jmsk.v17i2.11710.

- [24] A. Sandi, "Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest," *Jurnal Informatika dan Teknologi*, vol. 6, no. 2, p. 238, 2023, doi: 10.29408/jit.v6i2.10104.
- [25] M. H. Dunham, *Data Mining*. Upper Saddle River: Prentice Hall, 2003.
- [26] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.
- [27] J. Joseph, S. Senith, A. Alfred Kirubaraj, and S. R. Jino Ramson, "Machine Learning for Prediction of Clinical Appointment No-Shows," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 7, no. 4, pp. 558–574, Aug. 2022, doi: 10.33889/IJMEMS.2022.7.4.036.