

# Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode *Random Forest* dan *Support Vector Machine*

Qonita Iman<sup>a1</sup>, Arie Wahyu Wijayanto<sup>a2</sup>

<sup>a</sup>Program Studi Statistika, Politeknik Statistika STIS

Jl. Otto Iskandardinata No.64C, Jatinegara, Jakarta Timur, Daerah Khusus Ibukota Jakarta 13330

<sup>1</sup>211709949@stis.ac.id

<sup>3</sup>ariewahyu@pertama.edu

## Abstrak

Program Beras Miskin (Raskin) atau Beras Sejahtera (Rastra) merupakan program subsidi pemerintah dalam bentuk non-tunai yakni beras yang bertujuan menanggulangi kemiskinan dan memberikan perlindungan sosial di bidang pangan. Sayangnya, program ini masih belum berjalan efektif sepenuhnya. Permasalahan yang masih terjadi salah satunya mengenai rumah tangga sasaran. Penelitian ini bertujuan untuk melakukan klasifikasi/pengelompokkan rumah tangga penerima Raskin/Rastra menggunakan pendekatan *data mining* serta melakukan perbandingan metode yakni *Random Forest* dengan SVM (*Support Vector Machine*). Data yang digunakan merupakan data mikro Susenas Provinsi Jawa Barat tahun 2017. Variabel yang digunakan adalah penerima bantuan Raskin/Rastra sebagai target/kelas dan variabel penjelas (atribut) yang terdiri dari pekerjaan KRT, luas lantai, jenis dinding, jenis lantai, sumber air, sumber penerangan, pengeluaran per kapita. Program yang digunakan untuk pengolahan data adalah SPSS dan R. Hasil penelitian menunjukkan metode *random forest* dan SVM memiliki ketelitian yang cukup baik dalam melakukan klasifikasi, yakni sebesar 71–72%. Secara keseluruhan, selisih kinerja klasifikasi kedua metode *random forest* dan SVM memiliki kinerja yang sama baiknya dalam melakukan klasifikasi.

**Kata kunci:** Beras miskin, klasifikasi, *random forest*, SVM

# Classification of Poor Rice (Raskin)/Prosperous Rice (Rastra) Recipient Households in West Java Province in 2017 using the Random Forest Method and Support Vector Machine

## Abstract

Beras Miskin (Raskin) or Beras Sejahtera (Rastra) Program is a government subsidized program in the form of non-cash, that is rice, which aims to alleviate poverty and provide social protection in the food sector. Unfortunately, this program is still not fully effective. One of the problems that still occur is the targeted household. The purpose of this study is to classify/group Raskin/Rastra targeted households using a data mining approach and to compare methods, which are Random Forest and SVM (Support Vector Machine). The data used Susenas micro data for West Java Province in 2017. The variables are recipients of Raskin/Rastra assistance as a target/class and explanatory variables (attributes) consisting of head of household's work, floor area, wall type, floor type, water source, lighting source, per capita expenditure. The programs used for data processing were SPSS and R. The results showed that the random forest and SVM methods had a fairly good accuracy in classification, which is 71–72%. Overall, both the random forest and SVM methods performed equally well in classifying.

**Keywords:** Poor rice, classification, random forest, SVM

## I. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan ekonomi makro yang masih terus terjadi hingga saat ini, tak terkecuali di Indonesia. Kemiskinan adalah suatu kondisi di masa seseorang atau sekelompok orang tidak mampu secara ekonomi untuk memenuhi kebutuhan dasar hidupnya[1]. Penanggulangan masalah kemiskinan

merupakan poin utama yang menjadi fokus dalam mewujudkan kesejahteraan masyarakat[2].

Salah satu kebutuhan dasar manusia yang harus terpenuhi adalah pangan. Kemiskinan sendiri memiliki kaitan erat terhadap ketahanan pangan. Hal ini dikarenakan penduduk miskin umumnya lebih terfokus pada pemenuhan kebutuhan pangan dibandingkan kebutuhan non-pangan[3]. Keterbatasan akses terhadap pangan yang

terjadi pada masyarakat mengindikasikan adanya gangguan pada ketahanan pangan.

Indonesia sebagai negara agraris yang memiliki sektor utama pertanian juga perlu mengedepankan pertahanan pangan. Ketahanan pangan di Indonesia dapat dikatakan dalam kondisi yang cukup baik. Daftar pustaka [4] mengungkapkan terjadi peningkatan produksi sejumlah tanaman utama di Indonesia, termasuk padi, lebih tinggi dibandingkan pertumbuhan penduduknya. Dalam pelaksanaan ketahanan pangan nasional, pemerintah memiliki tanggung jawab untuk memastikan terpenuhinya kebutuhan pangan masyarakat. Berdasarkan hal tersebut, Pemerintah Indonesia melakukan program subsidi pangan dalam rangka mengatasi (atau setidaknya mengurangi) permasalahan kemiskinan. Oleh karena itu, pemerintah membuat suatu program yang disebut sebagai Program Beras Miskin (Raskin) atau Beras Sejahtera (Rastra).

Program Beras Miskin (Raskin) atau yang dikenal juga dengan Beras Sejahtera (Rastra) merupakan program pemerintah dalam bentuk non-tunai yakni beras yang diperuntukkan bagi rumah tangga berpendapatan rendah (rumah tangga miskin dan rentan). Program ini diselenggarakan oleh Pemerintah Pusat dengan tujuan menanggulangi kemiskinan dan memberikan perlindungan sosial di bidang pangan[5].

Kementerian Perdagangan Dalam Negeri (Kemendagri) mengungkapkan program Raskin/Rastra sendiri pelaksanaannya diatur dalam UU No. 18 Tahun 2012 tentang Pangan, PP No. 17 Tahun 2015 tentang Ketahanan Pangan dan Gizi, dan beberapa peraturan lainnya. Banyaknya penerima bantuan ini dihitung berdasarkan rumah tangga yang menerima Kartu Perlindungan Sosial (KPS) atau Kartu Keluarga Sejahtera (KKS) atau Surat Keterangan Rumah Tangga Miskin (SKRTM).

Terdapat enam indikator yang digunakan untuk mengukur keberhasilan Program Raskin/Rastra, yang disebut dengan 6T, yakni: 1) tepat sasaran, 2) tepat waktu, 3) tepat jumlah, 4) tepat harga, 5) tepat kualitas, dan 6) tepat administrasi. Dalam menanggulangi kemiskinan, efektivitas program Raskin sangat tergantung pada ketepatan sasaran kelompok rumah tangga miskin ataupun rentan[6].

Sayangnya, program ini masih belum berjalan efektif sepenuhnya. Di beberapa wilayah di Indonesia masih ditemukan adanya permasalahan pada keberhasilan program ini. Permasalahan yang masih terjadi sebagian besar adalah mengenai kualitas beras, jumlah beras yang diterima, terutama rumah tangga sasaran[7]–[11]. Pada tahun 2014, 17,5% rumah tangga di Indonesia memenuhi syarat pada tiga program bantuan sosial. Sayangnya, rumah tangga tersebut tidak menerima satupun program bantuan sosial, termasuk Raskin/Rastra[12].

BPK mengungkapkan data penerima bantuan raskin belum mutakhir dan kualitas beras meragukan[13]. Hal ini menyebabkan BPK melakukan pemeriksaan kinerja atas pengelolaan raskin yang dilaksanakan di 10 provinsi Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Lampung, Yogyakarta, Nusa Tenggara Timur, Sulawesi Selatan, Kalimantan Timur, dan Papua Barat. Masih kurangnya efektivitas pemberian Beras Miskin (Raskin) yang terjadi

di beberapa wilayah di Indonesia menunjukkan perlunya teknis yang lebih baik dan lebih terarah dalam menentukan suatu rumah tangga berhak menjadi penerima program Raskin/Rastra atau tidak.

Kemajuan teknologi hingga saat ini mengantarkan manusia untuk lebih efisien dan lebih teliti dalam melakukan suatu pekerjaan. Belum tepatnya sasaran penerima bantuan Raskin/Rastra memerlukan metode analisis yang lebih kompleks. Keberadaan *data mining* sebagai kemajuan statistik berbasis teknologi membuat pengambilan keputusan berdasarkan analisis statistik yang lebih kompleks menjadi lebih efisien dalam pelaksanaannya. Dengan adanya metode *data mining*, pengelompokan keluarga yang berhak menerima bantuan Raskin/Rastra diharapkan lebih tepat sasaran dan pemanfaatannya dapat semakin dirasakan masyarakat. Selain itu juga pengelompokan dengan *data mining* meningkatkan efisiensi secara statistik dalam pengambilan kebijakan mengenai RTS-PM (Rumah Tangga Sasaran Penerima Manfaat).

Berdasarkan penjelasan latar belakang di atas, peneliti tertarik meneliti lebih jauh mengenai klasifikasi penerima program bantuan Raskin/Rastra menggunakan pendekatan *data mining*. Penelitian ini bertujuan untuk:

- Mengetahui klasifikasi/pengelompokan rumah tangga penerima Raskin/Rastra menggunakan pendekatan *data mining*
- Melakukan perbandingan metode antara *Random Forest* dengan SVM (*Support Vector Machine*)

## II. METODOLOGI

### A. Cara Sumber Data dan Program Pengolahan

Data yang digunakan bersumber dari data mikro Susenas Provinsi Jawa Barat tahun 2017. Data berjumlah 23.756 yang merupakan seluruh rumah tangga di Jawa Barat. Tabel 1 menunjukkan variabel-variabel yang digunakan pada penelitian ini. Program yang digunakan untuk pengolahan data adalah SPSS dan R. Metode yang digunakan adalah *random forest* dan *support vector machine* (SVM).

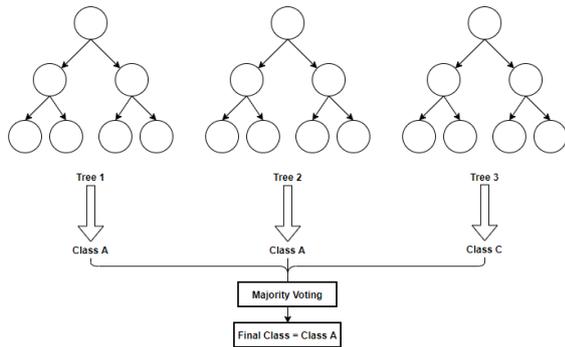
TABEL I  
DAFTAR VARIABEL DAN KODE PERTANYAAN

Deskripsi	Kode	Variabel
Penerima bantuan Raskin/Rastra	R2001	Dependen
Kedudukan pekerjaan KRT	R805	Independen
Luas lantai	R1604	Independen
Jenis dinding terluas	R1608	Independen
Jenis lantai terluas	R1609	Independen
Sumber air	R1611A	Independen
Sumber penerangan	R1618	Independen
Pengeluaran per kapita	EXP_CAP	Independen

### B. Random Forest

*Random Forest* merupakan salah satu metode CART (Classification and Regression Tree) dalam *data mining* dan tidak memerlukan asumsi apapun. Metode ini menggunakan konsep pohon keputusan (*decision tree*). Model ini dibentuk dari banyak pohon sehingga membentuk sebuah kumpulan pohon seperti hutan (*forest*) dengan menerapkan metode *bootstrap aggregating*

(*bagging*) dan *random feature selection* [14], [15]. Ilustrasi dari pohon keputusan dan pengambilan keputusan dengan *random forest* terdapat pada gambar 1.



Gambar 1. Ilustrasi metode *random forest*  
sumber: *medium.com*

Kelebihan *random forest*:

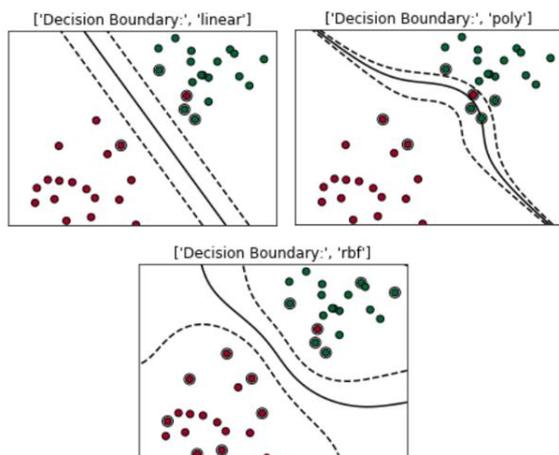
- Akurasi sangat baik dibandingkan metode klasifikasi lainnya
- Mampu menangani data *imbalanced*
- Dapat menangani data dengan sampel besar
- Dapat mengatasi data dengan *noise* dan *missing data*
- *Error* yang dihasilkan relatif rendah

Kekurangan *random forest*:

- Akurasi yang dihasilkan cenderung tidak stabil
- Waktu pemrosesan yang lama
- Membutuhkan *tuning* model yang tepat untuk data

C. *Support Vector Machine (SVM)*

*Support Vector Machine (SVM)* merupakan salah satu metode klasifikasi dan prediksi yang digunakan untuk mengatasi permasalahan yang ada pada metode statistika klasik. SVM dalam praktiknya menggunakan *hyperplane*, di mana pemisah dari kelompok kelas dibentuk melalui suatu dimensi vektor berukuran *n*. *Hyperplane* yang digunakan seoptimal mungkin dengan melihat jarak maksimum dari pemisah kelas atau yang disebut dengan *margin*[16]. Ilustrasi dari klasifikasi dengan SVM menggunakan beberapa jenis *hyperplane* terdapat pada gambar 2.



Gambar 2. Ilustrasi metode SVM berdasarkan jenis *hyperplane*  
sumber: *towardsdatascience.com*

Kelebihan SVM:

- Efektif pada data dengan dimensi tinggi
- Implementasi metode relatif mudah
- Bekerja dengan baik jika pemisah antarkelas jelas
- Mampu mengatasi *generalized error*

Kekurangan SVM:

- Sulit digunakan pada data dengan sampel besar
- Kurang cocok digunakan jika jumlah atribut lebih banyak dari *record*
- Hanya dapat digunakan untuk mengelompokkan 2 kelas

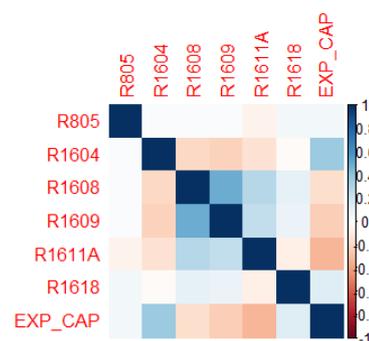
D. *Preprocessing*

Dalam melakukan pengolahan dan pemodelan, data harus terlebih dahulu “dibersihkan” melalui serangkaian proses yang disebut dengan *preprocessing*. Hal ini bertujuan agar data yang digunakan pada pemodelan selanjutnya telah bebas dari *noise* sehingga model yang dihasilkan kualitasnya semakin baik.

*Preprocessing* yang akan dilakukan terdiri dari imputasi *missing data*, pengecekan redundansi, dan standardisasi.

1) *Missing Data*: Data yang dikumpulkan masih memiliki *missing data*, hal ini diakibatkan pada kode pertanyaan R805, R806, R1611A, dan R1618 ada yang memang tidak diisi karena alur pertanyaan. Hal ini diatasi dengan memberi kategori baru pada masing-masing variabel dikarenakan tidak memiliki syarat untuk mengisi jawaban tersebut sehingga kurang baik apabila diatasi dengan ukuran pemusatan data seperti median ataupun modus. Proses pemberian kategori baru ini dilakukan pada program SPSS pada menu “*Record into Same Variable*”.

2) *Redundansi*: Pada gambar 3, terlihat bahwa plot korelasi tidak terdapat warna yang pekat atau nilai korelasi (*r*) lebih dari 0,7. Hal ini menunjukkan masing-masing variabel penjelas tidak memiliki korelasi yang tinggi (*r* < 0,7) sehingga tidak terdapat redundansi pada data dan semua atribut dapat digunakan untuk melakukan klasifikasi.



Gambar 3. Plot korelasi variabel penjelas

3) *Standardisasi*: Setelah *missing value* dan redundansi data dicek, data perlu distandardisasi dikarenakan terdapat variabel numerik yang memiliki satuan dan rentang yang berbeda. Variabel-variabel tersebut adalah R1604 (luas lantai) dan EXP\_CAP (pengeluaran per kapita). Hal ini agar variabel-variabel

numerik yang ada memiliki rentang yang sama. Metode standardisasi yang digunakan pada penelitian ini adalah normalisasi dengan *z-score*.

E. Pengujian Model

Pada tahap ini data yang digunakan untuk melakukan pemodelan adalah *training set*. Pemodelan data dilakukan untuk melihat hasil kinerja dari metode/model yang dipilih, yakni seberapa cocok model yang terbentuk dari *training set* yang sudah ditentukan. Proporsi yang ditentukan pada *training set* dan *test set* ini adalah 70% dan 30%.

F. Evaluasi dan Validasi Hasil

Setelah dilakukan pengujian model, evaluasi hasil diperlukan guna melihat dan menilai kinerja dari model yang berdasarkan data yang digunakan (di mana pada tahap ini data yang digunakan adalah *validation set* dan/atau *testing set*). Selanjutnya adalah validasi hasil yang bertujuan menilai hasil prediksi dari model. Evaluasi dan validasi hasil yang digunakan adalah dengan *confusion matrix*, *cross validation*, dan *ROC curve*.

1) *Confusion Matrix*: Metode ini menampilkan tabel yang berisi kelas yang diprediksi dan kelas yang sebenarnya, di dalamnya terdapat jumlah yang tepat dan tidak tepat dikelompokkan[17]. Baris pada tabel menunjukkan *true class*, sedangkan kolom pada tabel menunjukkan *predicted class*. Banyaknya amatan (*record*) yang diprediksi dengan tepat dapat dilihat melalui jumlah pada diagonal table 2.

TABEL II  
CONFUSION MATRIX PADA KLASIFIKASI 2 KELAS

		Kelas Prediksi	
		Ya	Tidak
Kelas Sebenarnya	Ya	TP	FN
	Tidak	FP	TN

*Confusion matrix* yang terbentuk dapat menunjukkan banyak ukuran evaluasi model. Di antaranya adalah[18], [19]:

- Akurasi (Accuracy)

Nilai akurasi menunjukkan ketepatan prediksi dari model yang dibentuk, baik kelas positif (ya) ataupun kelas negatif (tidak). Rumus akurasi sebagai berikut:

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

- Sensitivitas (Sensitivity)

Sensivitas mengukur proporsi dari nilai positif (ya) yang tepat diprediksi terhadap seluruh nilai positif (ya) yang sebenarnya. Ukuran ini juga disebut dengan *recall* atau *true positive rate* (TPR). Rumus sensitivitas sebagai berikut:

$$\text{Sensitivitas} = \frac{TP}{TP+FN} \tag{2}$$

- Spesifisitas (Specificity)

Sensivitas mengukur proporsi dari nilai negatif (tidak) yang tepat diprediksi terhadap seluruh nilai negatif (tidak) yang sebenarnya. Ukuran ini juga disebut dengan *true positive rate* (TPR). Rumus sensitivitas sebagai berikut:

$$\text{Spesifisitas} = \frac{TN}{TN+FP} \tag{3}$$

- Positive Predicated Value (PPV)

Nilai PPV menunjukkan proporsi kasus yang diprediksi masuk ke dalam kelas positif (ya) dan diprediksi secara benar. Rumus PPV sebagai berikut:

$$\text{PPV} = \frac{TP}{TP+FP} \tag{4}$$

- Negative Predicated Value (NPV)

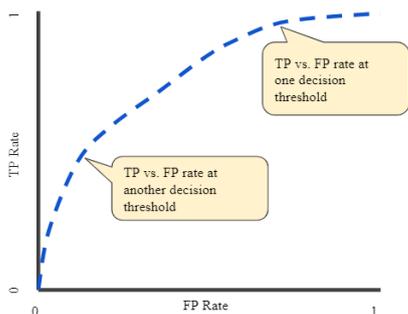
Nilai NPV menunjukkan proporsi kasus yang diprediksi masuk ke dalam kelas negatif (tidak) dan diprediksi secara benar. Rumus NPV sebagai berikut:

$$\text{NPV} = \frac{TN}{TN+FN} \tag{5}$$

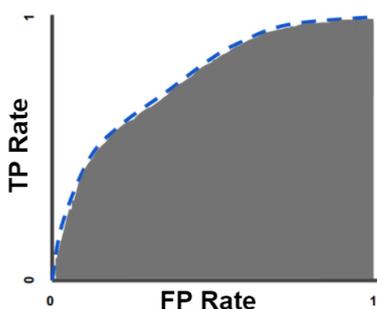
2) *Cross Validation*: Metode yang juga disebut sebagai estimasi rotasi (*rotation estimation*) ini digunakan untuk melihat performa dari klasifikasi model ataupun membandingkan dua atau beberapa metode klasifikasi. Metode ini sering digunakan dalam evaluasi model karena mudah dipahami dan diimplementasikan, bias yang dihasilkan juga lebih rendah disbanding metode lainnya. *Cross validation* dilakukan dengan membagi data menjadi *k* bagian data yang sama besar. Pembagian data bersifat acak (random) di mana data terdiri dari *training set* sebanyak (*k*-1) bagian dan 1 bagian sisa yang menjadi *testing set* yang saling independen (*mutually exclusive*). Pengujian prediksi masing-masing data amatan dilakukan hanya satu kali[20], [21].

Proses validasi dan testing dilakukan sebanyak *k* kali dan akurasi keseluruhan dari metode ini dilihat dari rata-rata akurasi setiap prosesnya. Banyaknya *folds* yang umum digunakan adalah 10 dengan beberapa variasi. Daftar pustaka [22] menyatakan bahwa semakin kecil iterasi, hasilnya akan semakin reliabel. Oleh karena itu, banyaknya bagian data (*k*) yang digunakan pada penelitian ini adalah 10-*folds* dengan 1 kali iterasi.

3) *Kurva AUC-ROC (AUC-ROC Curve)*: Metode ini menggunakan kurva dalam visualisasinya, kurva ini menggambarkan seberapa mampu model klasifikasi memisahkan kelas pada data. ROC menunjukkan kurva probabilitas dan AUC (*Area Under Curve*) menunjukkan derajat keterpisahan. Kurva ROC diplot dengan *Sensitivity/FPR (False Positive Rate)* pada sumbu horizontal dan *Specificity/TPR (True Positive Rate)* pada sumbu vertikal. Kurva AUC-ROC hanya dapat digunakan pada klasifikasi 2 kelas, tidak dapat dilakukan pada kelas berganda (*multiclass*)[23]. Kurva ROC pada data kontinu berbentuk garis kontinu sehingga membentuk kurva kontinu, pada data diskrit kurva akan berbentuk garis patahan. Gambar 4 dan gambar 5 menunjukkan kurva ROC dan AUC pada data kontinu.



Gambar 4. Kurva ROC pada data kontinu



Gambar 5. AUC pada data kontinu  
sumber: wemp.app

Penilaian model dilihat dari besarnya nilai AUC yang terbentuk dari area di bawah kurva ROC, nilai AUC berada pada rentang 0 hingga 1. Semakin jauh kurva ROC dari garis diagonal ( $x = y$ ), maka nilai AUC akan semakin besar (mendekati 1) yang menunjukkan akurasi data yang diuji semakin baik. Sebaliknya, semakin mendekati garis diagonal ( $x = y$ ), maka nilai AUC akan semakin rendah dan menunjukkan akurasi data yang diuji semakin rendah/buruk[24].

III. HASIL DAN PEMBAHASAN



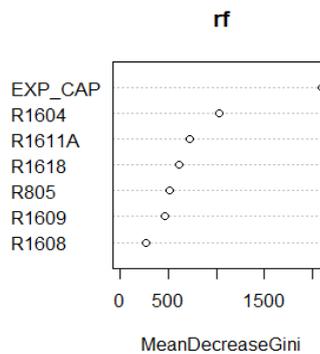
Gambar 6. Proporsi penerima bantuan Raskin/Rastra

Pada gambar 6 terlihat bahwa proporsi penerima bantuan Raskin/Rastra adalah sebanyak 10.082 rumah tangga atau sebesar 42 persen. Berdasarkan proporsi penerima bantuan ini, rumah tangga yang termasuk dalam kategori keluarga miskin dan berhak mendapatkan bantuan Raskin/Rastra ini masih sangat banyak, yakni hampir dari setengah dari total rumah tangga di Provinsi Jawa Barat tahun 2017.

Daftar pustaka [25] mengungkapkan dalam melakukan analisis klasifikasi, asumsi yang digunakan ialah data yang

digunakan seimbang. Ketidakseimbangan kelas atau *imbalanced class* merupakan kondisi di mana satu kelas memiliki jumlah sampel yang besar dan kelas lainnya hanya memiliki jumlah sampel yang sedikit. Berdasarkan hal tersebut, proporsi kelas yang ditunjukkan pada gambar 6 mengindikasikan tidak adanya *imbalanced class* pada variabel dependen (kelas).

Model *random forest* yang terbentuk menggunakan pohon sebanyak 500 pohon keputusan, dan model SVM yang terbentuk menggunakan *hyperplane* model linier (*linear*) dengan nilai cost sebesar 0,1.



Gambar 7. Plot variable important

Berdasarkan gambar 7, variabel yang paling berpengaruh dalam pembentukan pohon keputusan adalah pengeluaran per kapita (EXP\_CAP), kemudian disusul dengan luas lantai (R1604), sumber air (R1611A), sumber penerangan (R1618), pekerjaan kepala rumah tangga (R805), lantai terluas (R1609), dan terakhir adalah dinding terluas (R1608).

TABEL III  
PERBANDINGAN EVALUASI MODEL DENGAN SPLIT VALIDATION

	Random Forest	SVM
Akurasi	0,7208	0,7239
Sensivicity	0,7030	0,7094
Specificity	0,7336	0,7343
Positive Pred. Value	0,6556	0,6583
Negative Pred. Value	0,7740	0,7779

Dari tabel 3, terlihat rata-rata akurasi yang dihasilkan metode SVM lebih baik yakni sebesar 72,39%. Nilai *sensitivity*, *specificity*, *positive predicted value*, dan *negative predicted value* juga lebih baik pada metode SVM. Namun secara keseluruhan, perbedaan nilai kedua metode tersebut sangat kecil. Seperti pada nilai akurasi yang dihasilkan, perbedaan dari kedua metode hanya sebesar 0,31%. Hal ini menunjukkan, dengan evaluasi model menggunakan *split validation*, ketelitian kedua metode dapat dikatakan sama dan sudah baik dalam melakukan pengelompokan/klasifikasi.

TABEL IV  
CONFUSION MATRIX MODEL RANDOM FOREST DENGAN CROSS VALIDATION

	Penerima	Bukan Penerima
Penerima	1780	853
Bukan Penerima	1207	3287

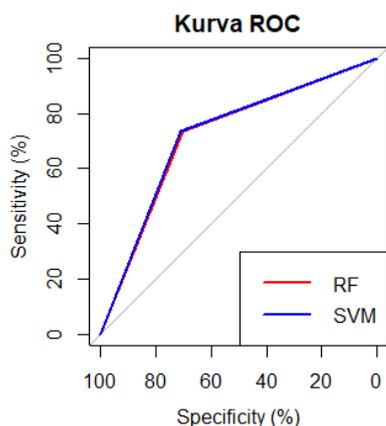
TABEL V  
CONFUSION MATRIX MODEL SVM DENGAN CROSS VALIDATION

	Penerima	Bukan Penerima
Penerima	2122	1107
Bukan Penerima	865	3033

TABEL VI  
PERBANDINGAN EVALUASI MODEL DENGAN CROSS VALIDATION

	Random Forest	SVM
Akurasi	0,7110	0,7233
Sensitivity	0,5959	0,7104
Specificity	0,7940	0,7326
Positive Pred. Value	0,6760	0,6572
Negative Pred. Value	0,7314	0,7781

Pada tabel 6, dapat dilihat rata-rata akurasi yang dihasilkan metode SVM lebih baik yakni sebesar 72,33%. Nilai *sensitivity*, *specificity*, dan *negative predicted value* juga lebih baik, hanya saja pada *positive predicted value* nilainya sedikit lebih kecil dari metode *random forest*. Perbedaan nilai pada kedua metode dengan evaluasi model *cross validation* sedikit terlihat jika dibandingkan dengan evaluasi model *split validation*. Kedua metode juga sudah baik dalam melakukan pengelompokan/klasifikasi. Metode SVM dapat dikatakan sedikit lebih baik dibandingkan dengan *random forest* dari segi evaluasi modelnya.



Gambar 8. Kurva ROC

Dapat dilihat kurva ROC pada gambar 8, garis ROC pada metode *random forest* dan SVM berimpit. Metode *random forest* memiliki nilai AUC sebesar 71,8% dan metode SVM memiliki nilai AUC sebesar 72,2%. Dapat dikatakan berdasarkan kedua garis di atas, ketelitian kedua metode hampir sama baiknya.

Berdasarkan seluruh evaluasi model yang dilakukan pada kedua metode, *random forest* dan SVM, ketelitian

yang dihasilkan hampir sama. Selisih kinerja klasifikasi yang dihasilkan kedua metode *random forest* dan SVM tidak signifikan. Oleh karena itu, dapat dikatakan bahwa kedua metode klasifikasi memiliki kinerja yang setara dan sama baiknya dalam melakukan klasifikasi

#### IV. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan yang telah dilakukan sebelumnya, dapat diambil kesimpulan bahwa:

- Rata-rata akurasi yang dihasilkan dengan metode *random forest* adalah sebesar 71–72% dan metode SVM adalah sebesar 72%. Hal ini menunjukkan model yang dibentuk dari kedua metode sudah baik dalam melakukan pengelompokan (klasifikasi) rumah tangga penerima program bantuan Raskin/Rastra.
- Secara keseluruhan, selisih kinerja klasifikasi kedua metode *random forest* dan SVM tidak signifikan. Maka, kedua metode klasifikasi memiliki kinerja yang setara dan sama baiknya dalam melakukan klasifikasi.

Saran yang dapat diberikan untuk pemerintah sebagai pembuat kebijakan adalah melakukan koordinasi lanjutan dengan pakar *data mining* dalam mengambil keputusan rumah tangga sasaran penerima bantuan Raskin/Rastra ke depannya. Berikutnya, saran yang dapat diberikan guna menunjang penelitian selanjutnya adalah menggunakan data mikro yang lebih mutakhir dan melakukan perbandingan keberhasilan program di beberapa wilayah.

#### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, “Konsep Kemiskinan dan Ketimpangan.” <https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html>.
- [2] A. V. K. Fetiningrum, “Inovasi Program Bantuan Pangan Non Tunai (BPNT) Sebagai Upaya Pengentasan Kemiskinan,” Universitas Brawijaya, 2017.
- [3] N. Ainistikmalia, “Analisis Kemiskinan Multidimensi dan Ketahanan Pangan Provinsi Kalimantan Utara,” Universitas Padjajaran, 2020.
- [4] Badan Pusat Statistik, *Potensi Pertanian Indonesia-Analisis Hasil Pencacahan Lengkap ST2013*. Jakarta, 2014.
- [5] Kementerian Dalam Negeri, “Tanya Jawab Program Raskin.” <https://raskin.bangda.kemendagri.go.id/tj-raskin.html>.
- [6] Rubiah, “Analisis Efektivitas dan Efisiensi Distribusi Raskin di Kelurahan Tempuling Kecamatan Tempuling Kabupaten Indragiri Hilir,” *J. Agribisnis*, vol. 6, no. 2, hal. 49–62, 2017.
- [7] A. Solikhin, “Analisis Pelaksanaan Penyaluran Beras untuk Rumah Tangga Miskin (Raskin) di Desa Tanjung Bakau Kecamatan Rangsang Kabupaten Kepulauan Meranti,” Universitas Islam Negeri Sultan Syarif Kasim Riau, 2016.
- [8] S. Khomsatun, “Analisis Penerima Raskin (Beras untuk Rumah Tangga Miskin) di Provinsi Jawa Tengah,” *Econ. Dev. Anal. J.*, vol. 6, no. 3, hal. 283–305, 2017, doi: 10.15294/edaj.v6i3.22272.
- [9] M. Y. A. Pulungan, “Efektivitas Program Beras Miskin (Raskin) di Kelurahan Limau Sundai Kecamatan Binjai Barat Kota Binjai Tahun 2015,” Universitas Sumatera Utara, 2017.
- [10] F. Mutia, “Analisis Efektivitas dan Dampak Pelaksanaan Program Beras untuk Keluarga Miskin di Kenagarian Biaro Gadang Kecamatan Ampek Angkek Tahun 2017,” Universitas Andalas, 2018.
- [11] Neli, L. O. Turi, dan Rizal, “Pendistribusian Beras Subsidi (Raskin) pada Masyarakat Wakatobi,” *J. Online Progr. Stud. Pendidik. Ekon.*, vol. 4, no. 4, hal. 99–106, 2019.
- [12] A. Tohari, C. Parsons, dan A. Rammohan, “Targeting poverty under complementarities: Evidence from Indonesia’s unified targeting system,” *J. Dev. Econ.*, vol. 140, no. April 2018, hal. 127–144, 2019, doi: 10.1016/j.jdeveco.2019.06.002.
- [13] CNN Indonesia, “BPK: Ada Risiko Penyimpangan Subsidi Raskin

- di Bulog,” Apr 29, 2015. <https://www.cnnindonesia.com/ekonomi/20150429111318-78-49982/bpk-ada-risiko-penyimpangan-subsidi-raskin-di-bulog>.
- [14] T. Purwa, “Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced,” *J. Mat. Stat. dan Komputasi*, vol. 16, no. 1, hal. 58–73, 2019, doi: 10.20956/jmsk.v16i1.6494.
- [15] I. L. Mulyahati, “Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python,” Universitas Islam Indonesia, 2020.
- [16] D. A. Otcherea, T. O. A. Ganat, R. Gholami, dan S. Ridhaa, “Application of Supervised Machine Learning Paradigms in The Prediction of Petroleum Reservoir Properties: Comparative Analysis of ANN and SVM Models,” *Epilepsy Res.*, hal. 108182, 2020, doi: 10.1016/j.petro.2020.108182.
- [17] A. Tharwat, “Classification Assessment Methods,” *Appl. Comput. Informatics*, 2018, doi: 10.1016/j.aci.2018.08.003.
- [18] M. Sultana, A. Haider, dan M. S. Uddin, “Analysis of Data Mining Techniques for Heart Disease Prediction,” *2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEICT 2016*, 2016, doi: 10.1109/CEEICT.2016.7873142.
- [19] I. D. Dinov, “Model Performance Assessment,” in *Data Science and Predictive Analytics: Biomedical and Health Applications using R*, 2018, hal. 475–496.
- [20] L. Xu *et al.*, “Representative Splitting Cross Validation,” *Chemom. Intell. Lab. Syst.*, vol. 183, hal. 29–35, 2018, doi: 10.1016/j.chemolab.2018.10.008.
- [21] T. H. Kerbaa, A. Mezache, dan H. Oudira, “ScienceDirect ScienceDirect Model Selection of Sea Clutter Using Cross Validation Model Selection of Sea Clutter Using Cross Validation Method a , b Method,” *Procedia Comput. Sci.*, vol. 158, hal. 394–400, 2019, doi: 10.1016/j.procs.2019.09.067.
- [22] J. G. Moreno-Torres, J. A. Saez, dan F. Herrera, “Study on The Impact of Partition-Induced Dataset Shift on k-Fold Cross-Validation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 8, hal. 1304–1312, 2012, doi: 10.1109/TNNLS.2012.2199516.
- [23] Sarang Narkhede, “Understanding AUC - ROC Curve,” *Towards Data Science*, 2018. <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.
- [24] A. King dan R. Eckersley, “Descriptive Statistics III: ROC Analysis,” in *Statistics for Biomedical Engineers and Scientists*, Academic Press, 2019, hal. 57–69.
- [25] Y. Sun, M. S. Kamel, A. K. C. Wong, dan Y. Wang, “Cost-Sensitive Boosting for Classification of Imbalanced Data,” *Pattern Recognit.*, vol. 40, no. 12, hal. 3358–3378, 2007, doi: 10.1016/j.patcog.2007.04.009.