J M S K

E-ISSN: 2614-8811

JURNAL MATEMATIKA. STATISTIKA & KOMPUTASI

https://journal.unhas.ac.id/index.php/jmsk/index

Vol. 17, No. 2, 240-251, January, 2021

DOI: 10.20956/jmsk.v17i2.11680

Workforce Classification in West Java 2018 With Random Forest

Klasifikasi Angkatan Kerja di Jawa barat Tahun 2018 dengan Metode *Random Forest*

Ahmad Safrian F[1], M. Rizki R[2], Rendy R.R[3], Ria Nurul A[4], Tika Novitasari[5], Rani Nooraeni[6]

Abstract

Unemployment in Indonesia is a serious problem. The high unemployment rate in Indonesia is due to the number of available jobs is not proportional to the increasing number of the labor force. Based on data from BPS, West Java is the largest contributor to a number of the unemployment in Indonesia 2018, which has an open unemployment rate of 8.52 percent. The purpose of this study is to classify the labor age population between 15 – 64 years into two groups (classes), namely unemployed or non-unemployed groups in West Java Province 2018 and to analyze the variables that have a major contribution in the classification process. The classification is carried out using a random forest model to improve classification accuracy. The random forest model that is formed uses the random over-under sampling (ROUS) method because it produces the best performance compared without ROUS in dealing with imbalanced data. The classification results of the model after resamling shows the level of accuracy, sensitivity, specificity, and AUC, which are 73.89 percent, 80.64 percent, 67.40 percent, and 75.73 percent. It shows that the model after resampling with random over-under sampling is better in the classification of the labor status of the working age population which includes the workforce in West Java in 2018. In addition, from the results of the analysis, the variable marital status is variable that has a major contribution in determining unemployment status.

Keyword : unemployment, random forest, random over-under sampling

Abstrak

Pengangguran di Indonesia merupakan masalah yang serius. Tingginya angka pengangguran tersebut dikarenakan jumlah lapangan kerja yang tersedia tidak sebanding dengan jumlah angkatan kerja yang terus meningkat. Berdasarkan data BPS, Jawa Barat sebagai penyumbang terbesar jumlah pengangguran di indonesia tahun 2018, dengan tingkat pengangguran terbuka sebesar 8,52 persen. Tujuan penelitian ini untuk melakukan klasifikasi penduduk usia kerja antara 15 – 64 tahun yang termasuk angkatan kerja kedalam dua kelompok (kelas), yaitu berstatus pengangguran atau bukan pengangguran (bekerja) di Provinsi Jawa Barat tahun 2018 dan menganalisis variabel yang paling berkontribusi besar dalam proses pengklasifikasian. Pengklasifikasian tersebut dilakukan dengan model *random forest* karena dapat meningkatkan ketepatan klasifikasi. Model *random forest* yang terbentuk menggunakan metode *random over-under sampling* (ROUS) karena menghasilkan performa terbaik dibandingkan tanpa ROUS dalam mengatasi data *imbalance*. Hasil pengklasifikasian dari model tersebut setelah resamling menunjukkan tingkat akurasi, *sensitivity*, *spesifity*, dan AUC, yaitu sebesar 73,89 persen, 80,64 persen, 67,4 persen,

**Program Studi Diploma IV Statistika, Jurusan Statistika Ekonomi Politeknik Statitika STIS*

***E-Mail**: 211709510@stis.ac.id[1], 211709838@stis.ac.id[2], 211709966@stis.ac.id[3], 211709974@stis.ac.id[4], 211710033@stis.ac.id[5], raninoor@stis.ac.id[6]*

dan 75,73 persen. Selain itu, dari hasil analisis variabel status perkawinan merupakan variabel yang memiliki kontribusi besar dalam menentukan status angkatan kerja.

Kata Kunci : pengangguran, *random forest*, *random over-under sampling*

1. PENDAHULUAN

Pembangunan ekonomi merupakan suatu proses yang berkesinambungan antara sumber daya manusia, sumber daya alam, modal, teknologi dan lain-lain dengan tujuan untuk meningkatkan kesejahteraan ekonomi masyarakat. Keberhasilan pembangunan ekonomi tersebut

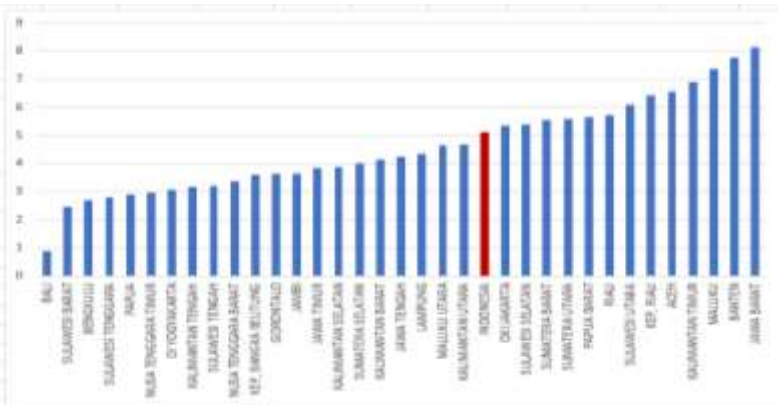
dilihat dari laju pertumbuhan ekonomi yang diiringi dengan pengurangan kemiskinan dan ketimpangan pendapatan. Selain itu dalam pembangunan ekonomi tersebut pasti akan dihadapkan dengan berbagai macam permasalahan, salah satunya masalah pengangguran.

Pengangguran adalah seseorang yang berusia 15 tahun ke atas yang tidak bekerja dan sedang mencari pekerjaan [15]. Pengangguran adalah masalah yang sangat buruk efeknya kepada perekonomian dan masyarakat [25]. Pengangguran tersebut merupakan suatu permasalahan kompleks yang harus dihadapi oleh negara maju maupun negara berkembang karena secara langsung maupun tidak langsung akan berdampak pada perekonomian, individu dan masyarakat, seperti kemiskian, kesejahteraan masyarakat, kriminalitas, dan masalah sosial ekonomi lainnya. Masalah pengangguran menjadi poin penting yang dimasukkan ke dalam tujuan pembangunan berkelanjutan (*Sustainable Development Goals*/SDGs) yang kedelapan, yaitu "Pada tahun 2030, mencapai ketenagakerjaan secara penuh dan produktif dan pekerjaan yang layak bagi seluruh perempuan dan laki-laki, termasuk untuk kaum muda dan orang dengan disabilitas, juga kesetaraan upah bagi pekerjaan yang mempunyai nilai yang sama".

Permasalahan pengangguran di Indonesia merupakan masalah yang serius dan menjadi perhatian khusus oleh pemerintah. Tingginya angka pengangguran di Indonesia tersebut dikarenakan jumlah lapangan kerja yang tersedia tidak sebanding dengan jumlah angkatan kerja yang terus meningkat. Upaya pemerintah untuk mengurangi pengangguran ditunjukkan dengan penetapan target angka tingkat pengangguran dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2015 – 2019. Selain itu, Menurut Menteri Ketenagakerjaan, Muhammad Hanif Dhakiri, sepanjang tahun 2015 – 2018, pemerintah telah berhasil membuka 10,34 juta lapangan kerja, kata dia, rata-rata setiap tahun telah tercipta 2,58 juta lapangan kerja baru bagi masyarakat untuk setiap tahunnya. Angka tingkat pengangguran terbuka (TPT) Indonesia pada tahun 2018 (bulan Agustus) sebesar 5,34 persen [2]. Angka TPT tersebut mengalami penurunan sebesar 0,16 persen dari tahun 2017. Hal ini dapat dikatakan sejalan dengan kebijakan pembukaan lapangan pekerjaan oleh pemerintah. Akan tetapi jumlah lapangan pekerjaan tersebut belum dapat memenuhi permintaan lapangan pekerjaan dan belum dapat menyerap angkatan kerja secara maksimal. Hal tersebut ditunjukkan dengan hasil evaluasi dari RPJMN 2015 – 2019, bahwa realisasi pengurangan pengangguran di Indonesia belum dapat mencapai target karena target yang ditetapkan berkisar antara empat sampai lima persen dan TPT Indonesia tahun 2018 masih di atas angka lima persen.

Apabila masalah pengangguran di Indonesia tidak segera diatasi akan berdampak pada terganggunya proses pembangunan ekonomi Indonesia dalam jangka panjang. Secara langsung, pengangguran dapat mengakibatkan menurunnya tingkat kemakmuran dan kesejahteraan ekonomi karena pendapatan riil yang diperoleh masyarakat menjadi lebih rendah. Selain itu, pengangguran juga dapat menghambat laju pertumbuhan ekonomi dan meningkatkan kemiskinan Indonesia. Menurut [4], risiko menjadi pengangguran akan meningkat seiring dengan semakin rendahnya pendidikan pada studi kasus Rusia. Seiring dengan temuan tersebut,

di Latvia, [14] mendapati bahwa status pendidikan mempengaruhi posisi tawar seseorang dalam mencari pekerjaan dan dapat mempengaruhi besaran penghasilannya.



Sumber : BPS (diolah)

Gambar 1.1. Tingkat Pengangguran Tebuka (TPT) di Indonesia Tahun 2018

Dari 34 provinsi di Indonesia penyumbang terbesar untuk jumlah pengangguran, yaitu Provinsi Jawa Barat, Banten, dan Maluku dengan masing-masing angka tingkat penganggurannya sebesar 8,52 persen, 8,17 persen, dan 7,27 persen [2]. Angka TPT ketiga provinsi tersebut jauh di atas angka TPT nasional, yaitu sebesar 5,34 persen. Dari Gambar 1. dapat dilihat bahwa Provinsi Jawa Barat merupakan provinsi dengan permasalahan tingkat pengangguran tertinggi di Indonesia. Dalam periode 2011 – 2018, Jawa Barat ini memiliki angka TPT di atas 8 persen. Jika dibandingkan dari tahun sebelumnya angka TPT Jawa Barat tahun 2018 mengalami penurunan sebesar 0,05 persen dari tahun sebelumnya. Hal tersebut masih jauh dari target Rancangan Pembangunan Jangka Menengah Daerah (RPJMD) Jawa Barat karena target penurunan tingkat pengangguran sebesar 0,5 persen setiap tahunnya.

Berdasarkan uraian permasalahan di atas, dalam penelitian ini akan dilakukan pengklasifikasian angkatan kerja berdasarkan faktor yang mempengaruhi tingkat pengangguran di Jawa Barat pada tahun 2018. Tingkat pengangguran terbuka ini mampu menggambarkan tingginya jumlah pengangguran terhadap banyaknya angkatan kerja berusia diatas 15 tahun yang terdapat di wilayah tertentu. Telah banyak berbagai literatur yang meneliti variabel – variabel yang diduga memiliki pengaruh terhadap keputusan seseorang untuk menganggur. Jenis kelamin berpengaruh positif terhadap keputusan seseorang untuk menganggur dimana perempuan memiliki kecenderungan menganggur lebih tinggi [1]. Hal ini terjadi juga pada usia seseorang yang berpengaruh negatif terhadap keputusan seseorang untuk menganggur. Selanjutnya, status perkawinan dimana kecenderungan seseorang yang belum kawin memiliki peluang menganggur lebih besar daripada yang sudah kawin. Tingkat pendidikan secara signifikan berpengaruh terhadap status kerja seseorang [11]. Selain itu, klasifikasi wilayah tempat tinggal penduduk berpengaruh terhadap kondisi status pekerjaan seeorang dimana seseorang yang tinggal di wilayah pedesaan memiliki kecenderungan untuk meganggur lebih besar karena ketersediaan lapangan pekerjaan di pedesaan sedikit.

Penelitian ini menggunakan metode *random forest* untuk mengetahui seberapa penting variabel-variabel di atas dalam mengklasifikasikan penduduk angkatan kerja kedalam kelompok berstatus pengangguran atau bukan pengangguran. Metode *random forest* merupakan teknik terbaik dalam melakukan pengklasifikasian karena menghasilkan *error term* paling minimum dibandingkan dengan metode lainnya. Selain itu, Metode tersebut menjadi pilihan alternatif untuk metode pengolahan data dengan dimensi data yang besar karena sangat sulit apabila dilakukan pengolahan data dengan metode konvensional. Dengan demikian, akan dilakukan

penelitan dengan judul "Analisis pengangguran di Jawa Barat Tahun 2018 dengan pendekatan *Random Forest".* Penelitian ini dilakukan dengan tujuan untuk melakukan klasifikasi penduduk angkatan kerja ke dalam kelompok berstatus pengangguran atau bukan pengangguran dengan pendekatan *random forest* dan mengetahui variabel mana yang paling berperan penting dalam melakukan pengklasifikasian tersebut.

2. TINJAUAN PUSTAKA

2.1. Angkatan kerja

Menurut Badan Pusat Statistik (BPS), penduduk yang termasuk angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) yang bekerja, atau punya pekerjaan namun sementara tidak bekerja dan pengangguran. Bekerja didefinisikan sebagai kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu. Kegiatan tersebut termasuk pola kegiatan pekerja tak dibayar yang membantu dalam suatu usaha/kegiatan ekonomi.

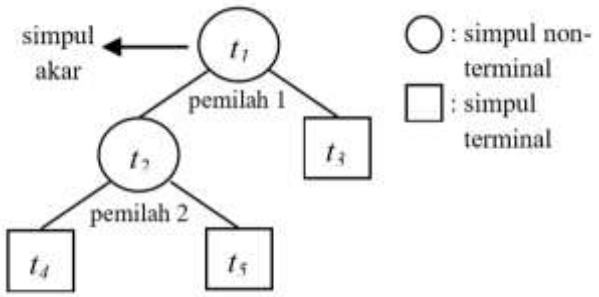
Pengangguran adalah suatu keadaan dimana seseorang yang tergolong dalam angkatan kerja yang ingin mendapatkan kerja tetapi mereka belum dapat memperoleh pekerjaan tersebut [25]. Pengangguran merupakan suatu ukuran yang dilakukan jika seseorang tidak memiliki pekerjaan tetapi mereka sedang melakukan usaha secara aktif dalam empat minggu terakhir untuk mencari pekerjaan [13].

2.2. *Classification and Regression Tree* (CART)

CART merupakan metode eksplorasi data yang didasarkan pada teknik pohon keputusan. Pohon klasifikasi dihasilkan saat variabel respons berupa data kategorik, sedangkan pohon regresi dihasilkan saat variabel respons berupa data numerik [6]. Pohon terbentuk dari proses pemilahan rekursif biner pada suatu gugus data sehingga nilai peubah respons pada setiap gugus data hasil pemilahan akan lebih homogen [6,21].

Pembentukan pohon klasifikasi CART meliputi tiga hal [6], yaitu sebagai berikut.

1. Pemilihan pemilah (*split*)
2. Penentuan simpul terminal
3. Penandaan label kelas



Gambar 2.1. Struktur pohon metode CART

Pada Gambar 1. Menunjukkan pohon disusun oleh simpul $t_1, t_2, \ldots, t_5$. Setiap pemilah (*split*) memilah simpul non-terminal menjadi dua simpul yang saling lepas. Hasil prediksi respons suatu amatan terdapat pada simpul terminal.

2.3. *Random Forest*

Metode *random forest* merupakan pengembangan dari metode CART, yaitu klasifikasi berbasis *decision tree*/pohon keputusan [20]. Sehingga *random forest* dapat disebut juga dengan *random decision forest.* Metode *random forest* akan menghasilkan banyak *tree* yang terbentuk

dari *training set*. *Decision* mayoritas dari seluruh *tree* yang terbentuk adalah hasil akhir dari klasifikasinya. Metode *random forest* secara umum memiliki nilai akurasi yang lebih tinggi dibanding metode lain. Selain itu, *random forest* lebih cocok digunakan untuk *training set* yang berukuran besar dan tetap dapat digunakan saat terdapat *missing data*. Akan tetapi akan membutuhkan waktu yang lebih lama dalam pembuatan modelnya saat *decision tree* yang dihasilkan semakin banyak.

2.4. Metode *Random Over-Under Sampling.*

Random over-under sampling (ROUS) merupaka salah satu fitur yang terdapat pada metode resampling Random Over Sampling Example (ROSE) yang diperkenalkan Giovanna Menardi dan Nicola Torelli dalam mengatasi data *imbalance* yang berpengaruh pada estimasi model dan ukuran evaluasi. Metode ROSE menambah data pada suatu kelas dengan data buatan berdasarkan pendekatan *smoothed bootstrap* [18]. Sedangkan ROUS merupakan kombinasi dari teknik undersampling dan oversampling. Kelas minoritas dilakukan oversampling dengan pengembalian dan kelas mayoritas dilakukan undersampling tanpa pengembalian[16].

2.5. Ukuran Evaluasi

Dalam metode klasifikasi, digunakan *test set* untuk menilai seberapa baik klasifikasi yang dihasilkan. Jika label suatu baris dalam *test set* sama dengan hasil klasifikasi yang dihasilkan oleh model, maka disebut sebagai *correct classification*. Sedangkan apabila label suatu baris dalam *test set* berbeda dengan hasil klasifikasi yang dihasilkan oleh model, maka disebut sebagai *missclassification*. Sehingga semakin banyak jumlah *correct classification*, maka akan semakin tinggi akurasi dari sebuah *classifier* dan sebaliknya semakin banyak jumlah *missclassification*, maka akan semakin rendah akurasi dari *classifier* [20].

Salah satu alat bantu yang digunakan untuk menilai seberapa baik *classifier* adalah *confusion matrix* yang dihasilkan dari aplikasi model pada *test set*.

Tabel 2.1. *Confusion Matrix*

		Predict Class		
		C	-C	
Actual Class	C	True Positive (TP)	False Negative (FN)	P
	-C	False Positive (FP)	True Negative (TN)	N
		P’	N’	All

Dari *confusion matrix* dapat diturunkan menjadi berbagai ukuran evaluasi, diantaranya:

1. Akurasi adalah perentase baris *test set* yang terklasifikasi dengan benar. $\mathbf{Akurasi} = \frac{\mathbf{(TP+TN)}}{\mathbf{All}}$
2. Sensitivity adalah True Positive recognition rate. $\mathbf{Sensitivity} = \frac{\mathbf{TP}}{\mathbf{P}}$
3. Specifity adalah True Negatif recognition rate. $\mathbf{Specifity} = \frac{\mathbf{TN}}{\mathbf{N}}$

Selain itu, terdapat ukuran evaluasi lain yang sering digunakan dalam mengukur performa dari model klasifikasi, yaitu AUC (*Area Under Curve*) [21]. Kurva yang dimaksud disini adalah kurva ROC (*Receiver Operating Characteristics*) yang menunjukkan hubungan antara *false positive rate* (1 – *Specifity*) dan *true positive rate* (*Sensitivity*). Luas AUC berada pada rentang 0 sampai 1, sehingga semakin luas (mendekati 1) maka semakin baik pula performa klasifikasinya.

2.6.*Mean Decrease Gini*

Mean Decrease Gini merupakan salah satu ukuran kriteria pemisahan yang digunakan dalam *random forest* yang juga digunakan dalam CART. Untuk menghitung *Mean Decrease Gini* digunakan rumus sebagai berikut:

$$\boldsymbol{Gini\,(K) = 1 - \sum_{i=1}^{n} P_i^2}$$

Pada setiap pemisahan, salah satu kelas digunakan untuk membentuk pemisahan dan terjadi penurunan gini. Jumlah semua penurunan pada *forest* karena variabel tertentu, dinormalisasi dengan jumlah *tree*, membentuk ukuran Gini [5].Angka *Mean Decrease Gini* merupakan angka yang menjelaskan seberapa penting suatu variabel independen dalam kontribusinya dalam model. Semakin tinggi angka *Mean Decrease Gini*, maka semakin tinggi pula kontribusinya dalam model tersebut.

3. METODOLOGI

3.1. Ruang Lingkup Penelitian

Menurut data yang dihimpun dari BPS, provinsi dengan predikat TPT tertinggi dimiliki oleh Provinsi Jawa Barat. TPT Provinsi Jawa Barat pada 2018, berada di level 8,52 persen, berada di atas TPT Nasional sebesar 5,34 persen. Selain itu, TPT Jawa Barat selama delapan tahun berturut-turut berada di atas delapan persen. Sehingga berdasarkan pertimbangan tersebut, lokus yang dipilih dalam penelitian ini adalah Provinsi Jawa Barat.

Tabel 3.1. Variabel-variabel yang digunakan pada penelitian

No	Jenis Variabel	Keterangan	Tipe Data
(1)	(2)	(3)	(4)
1	Klasifikasi wilayah	0 =Perkotaan; 1 = Perdesaan	Nominal
2	Jenis Kelamin	0 = Laki-laki; 1 = Perempuan	Nominal
3	Umur	0 = 15 – 28 tahun; 1 = ≥ 29 tahun	Nominal
4	Status Perkawinan	0 = Belum kawin, Cerai Hidup dan Cerai mati; 1 = Kawin	Nominal
5	Tingkat Pendidikan	0 = Tidak ada, Paket A; SDLB; SD; Paket B; SMPLB; SMP; Paket C; Paket C; SMA; dan SMK; 1 = D I/ D II; D III; D IV / S 1; S 2 dan S 3	Nominal
6	Pelatihan	0 = Ya; 1= Tidak	Nominal
7	Pengalaman Kerja	0 = Ya ; 1 = Tidak	Nominal
8	Status Pengangguran (Y)	0 = Tidak ; 1 = Ya	Nominal

3.2. Sumber Data

Sumber data penelitian ini menggunakan *raw data* hasil Survei Angkatan Kerja Nasional (Sakernas) tahun 2018. Sampel yang diambil merupakan individu yang termasuk sebagai angkatan kerja berusia 15 - 64 tahun. Jumlah total angkatan kerja dalam sampel sejumlah 20.074 individu.

3.3. Tahapan Analisis

Penelitian ini melakukan klasifikasian dengan metode *Random Forest* menggunakan aplikasi R. Pengklasifikasian ini terdiri dari dua kelas yaitu pengagguran dan bukan

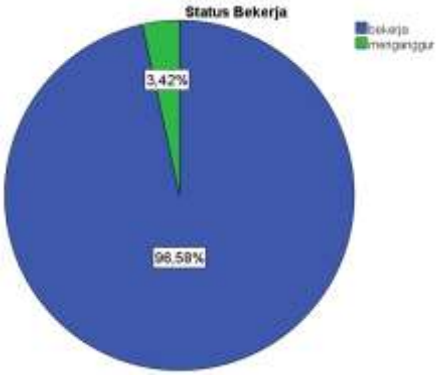
pengangguran (bekerja) berdasarkan hasil Sakernas tahun 2018. Berikut beberapa tahapan analisis data dengan metode *random forest* :

1. Menentukan dan mempersiapkan seluruh variabel independen dan dependen.
2. Melakukan pengkodean terhadap variabel-variabel bersifat kategorik.
3. Mengubah tipe data variabel-variabel kategorik dengan fungsi *as.factor().*
4. Menyeimbangkan data dengan metode *random over-under sampling* untuk mengatasi data timpang.
5. Memisahkan data menjadi data *training* dan data *testing*, dengan proporsi masing-masing sebesar 80 persen dan 20 persen.
6. Membuat model *random forest* berdasarkan data *training.*
7. Menguji model dengan data *testing*.
8. Membuat *confussion matrix*.
9. Menghitung ukuran evaluasi/performa model dengan melihat tebakan benar.
10. Mentukan *Mean Decrease Gini* untuk menjelaskan seberapa penting suatu variabel independen dalam model.

4. HASIL DAN PEMBAHASAN

4.1. Gambaran Angkatan Kerja

Berdasarkan data Sakernas 2018 berikut bahwa jumlah orang yang memiliki pekerjaan lebih besar daripada orang yang tidak memiliki pekerjaan (menganggur) untuk Provinsi Jawa Barat. Hal ini dilihat dari persentase penduduk bekerja sebesar 96,58 persen sedangkan persentase penduduk yang mengaggur sebesar 3,42 persen dari total penduduk angkatan kerja.



Sumber : BPS (diolah)

Gambar 4.1. Persentase Jumlah Angkatan kerja berdasarkan Status Bekerja

Adanya perbedaan yang besar ini menununjukkan terdapat data yang timpang dan model klasifikasi yang dibangun akan tidak efektif. Sehingga penelitian ini menggunakan metode *random forest* dengan *random over-under sampling* untuk mengatasi data yang *unbalanced* (timpang).

Tabel 4.1. Jumlah unit sebelum dan sesudah proses *random over-under sampling*

	Data	
Status	Sebelum disetimbangkan	Hasil *over-undersampling*
Bekerja	19388 (96,58%)	9877 (49,20%)

Menganggur	686 (3,42%)	10197 (51,79%)

Berdasarkan tabel diatas terdepatat perbedaan yang sangat jauh sebelum dan sesudah dilakukan penyeimbangan data dengan metode *random over-under sampling*. Sebelum dilakukan penyeimbangan data, persentase penduduk yang bekerja dan yang menganggur di Jawa Barat tahun 2018 masing-masing sebesar 96,58 persen dan 3,42 persen. Hal tersebut menunjukkan bahwa perbedaan proporsi penduduk yang bekerja dan menganggur sangat berbeda. Namun, setelah dilakukan penyeimbangan data dengan metode *random over-under sampling* jumlah penduduk yang berstatus bekerja dan menganggur jauh lebih proporsional dan hampir sama, dengan persentase penduduk yang bekerja dan yang menganggur di Jawa Barat tahun 2018 masing-masing sebesar 49,20 persen dan 51,79 persen.

4.2. Klasifikasi Angkatan Kerja dengan *Random Forest*

Penelitian ini melakukan pengklasifikasian penduduk angkatan kerja kedalam kelompok menganggur dan bekerja berdasarkan data Sakernas 2018 di Jawa Barat dengan variabel – variabel yang telah ditentukan menggunakan metode *random forest* dengan 80 persen dari data total atau sebanyak 16.059 data untuk data *training* dan 20 persen dari data total atau sebanyak 4.015 data untuk data *testing*. Untuk memeriksa dan menemukan model terbaik, peneliti melakukan perbandingan data *confusion matrix* sebelum dan sesudah *resampling* dengan *random over-under sampling* dengan hasil sebagai berikut :

Tabel 4.2. *Confusion matrix* (tanpa *resampling*)

		Nilai Sebenarnya	
		Bekerja	Pengangguran
Nilai Observasi	Bekerja	3877	147
	Pengangguran	0	0

Tabel 4.3. *Confusion matrix* (dengan *resampling*)

		Nilai Sebenarnya	
		Bekerja	Pengangguran
Nilai Observasi	Bekerja	1587	667
	Pengangguran	381	1379

Berdasarkan tabel 4.2, dapat diketahui bahwa model *random forest* pada data sebelum *resampling* (data *imbalanced*) dengan *random over-under sampling* mengalami *overfitting*. *Overfitting* dapat terjadi karena model cenderung mengarah pada salah satu kategori. Sehingga hasil prediksi menjadi bias dan akurasi yang menyesatkan [19]. Salah satu indikasi terjadi *overfitting* adalah dari hasil nilai prediksi dan nilai sebenarnya, model *random forest* tidak mampu memprediksi pengangguran berdasarkan variabel independen yang ada.

Sedangkan pada tabel 4.3,data hasil *resampling* menggunakan *random over-under sampling* mampu membuat model *random forest* memprediksi pengangguran berdasarkan

varaibel independen yang ada. Karena resampling dapat mengatasi masalah pada data *imbalanced* (tidak seimbang), maka model *random forest* dengan *random over-under sampling* lebih cocok dalam mengklasifikasikan angkatan kerja di Jawa Barat tahun 2018.

Tabel 4.4. Ukuran performa sebelum dan sesudah *resampling*

	Ukuran Performa (Persen)			
Kondisi	Akurasi	Sensitivity	Specificity	AUC
Sebelum	96,59	1	0	0,50
Sesudah	73,89	80,64	67,40	75,73

Berdasarkan ukuran performa model pada tabel 4.4, untuk model *random forest* tanpa *resampling* menghasilkan akurasi dan *sensitivity* yang cukup tinggi sebesar 96,56 persen dan 100 persen. Namun, untuk *specificity* sangat rendah sebesar 0 persen dan AUC sebesar 50 persen. Angka *specificity* yang sangat rendah karena ketidakmampuan model dalam memprediksi pengangguran akibat data yang *overfitting*. Sehingga model *random forest* tanpa *resampling* tidak tepat dipakai untuk melakukan klasifikasi meskipun akurasi dan *sensitivity* sangat tinggi.

Sedangkan untuk hasil model random forest setelah *resampling* diperoleh tingkat akurasi model random forest sebesar 73,89 persen Artinya, ada sekitar 26,11 persen data yang misklasifikasi dalam memprediksi data tersebut masuk dalam kategori yang benar dibandingkan dengan data aslinya. Untuk performa sensitifity berada di nilai 80,64 persen. Artinya, sebanyak 80,64 persen data diprediksi benar bahwa penduduk angkatan kerja tersebut diklasifikasikan memiliki pekerjaan dibandingkan dengan total penduduk angkatan kerja memiliki pekerjaan menurut data aslinya. Sedangkan 19,36 persen tersebut salah mengklasifikasikan yang sebenarnya memiliki pekerjaan namun diklasifikasikan menganggur. Sedangkan specifity menjelaskan bahwa terdapat 67.40 persen data yang diklasifikasikan benar bahwa penduduk tersebut menganggur dibandingkan total penduduk yang sebenarnya menganggur. Sedangkan 33.60 persen menjelaskan banyaknya penduduk yang salah diklasifikasikan, yang sebenarnya memiliki pekerjaan dimasukkan ke kelas menganggur. Untuk nilai AUC berada di angka 75.73 persen karena semakin mendekati 100 persen, maka suatu model dikatakan bagus dalam mengklasifikasikan data. Nilai AUC yang berada di 75.4 persen memberikan gambaran bahwa model setelah *resampling* dengan *random over-under sampling* lebih baik dalam klasifikasi status bekerja individu pada data Sakernas 2018 di Jawa Barat.

Tabel 4.5. Tabel *Mean Decrease Gini* dengan *Random Forest*

Variabel	Nilai *Mean Decrease Gini*
Status Perkawinan	1665.70
Jenis Kelamin	234.77
Pengalaman Kerja	171.63
Pengalaman Pelatihan Bersertifikat	98.83
Pendidikan Terakhir	94.14

Klasifikasi Wilayah	85.70
Umur	81.52
Pengalaman Pelatihan Tak Bersertifikat	16.15

Berdasarkan tabel 4.5 di atas dapat diketahui bahwa variabel status perkawinan memiliki angka *Mean Decrease Gini* terbesar dibandingkan variabel lainnya. Sehingga, variabel status perkawinan memiliki kontribusi yang besar dalam mengklasifikasikan penduduk yang memiliki pekerjaan dan yang menganggur berdasarkan hasil Sakernas 2018 di Jawa Barat. Sesuai dengan hasil penelitian [12] bahwa status perkawinan menjadi faktor yang krusial dalam menentukan posisi seseorang untuk berkeinginan mencari pekerjaan. Kemudian variabel selanjutnya adalah variabel jenis kelamin yang memiliki kontribusi terbesar kedua dalam menetukan penduduk yang memiliki pekerjaan dan yang menganggur.

5. KESIMPULAN

Berdasarkan penelitian yang dilakukan, model *random forest* dengan metode *resampling random over-under sampling* untuk mengatasi data *imbalanced* menghasilkan performa pengklasifikasian lebih baik dibandingkan tanpa *resampling*. Selain itu, variabel yang berperan penting dalam mengklasfikasian penduduk angkatan kerja adalah variabel status perkawinan dan jenis kelamin. Dengan demikian dapat disimpulkan bahwa penerapan model *random forest* dengan metode *random over-under sampling* lebih baik dalam melakukan pengklasfikasian pada data penduduk angkatan kerja di Jawa Barat tahun 2018 yang memiliki karakteristik *imbalanced* (rasio data tidak berimbang).

DAFTAR PUSTAKA

[1] Aryati, F. & Sunaryanto, H., 2014. Analisis Pengangguran Terdidik di Provinsi Bengkulu. *Jurnal Ekonomi dan Perencaan Pembangunan (JEPP)*, 05(04), 70 - 79.

[2] Badan Pusat Statistik. 2020. *Tingkat Pengangguran Terbuka Menurut Provinsi Tahun 1986 – 2020*. Badan Pusat Statistik

[3] Bappenas. 2017. *Evalusai Paruh Waktu RPJMN 2015 – 2019*. Bappenas.

[4] Blinova, T., Bylina, S., & Rusanovskiy, V., 2015. Vocational Education in the System of Determinants of Reducing Youth Unemployment: Interregional Comparisons. *Procedia - Social and Behavioral Sciences* 214, 526 – 534

[5] Breiman, L., & Cutler, A. (2011). Manual–setting up, using, and understanding random forests V4. 0. 2003. *URL https://www. stat. berkeley. edu/~ breiman/Using_random_forests_v4. 0. pdf*. [20 November 2020]

[6] Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J., 1984. *Classification and Regression Trees.* Chapman & Hall.

[7] Breiman, L., 2001. *Random Forests. Machine Learning*, 45:5-32.

[8] Cahyono, & Kahar, S., 2019. *Menguak Fakta Tingginya Pengangguran di Jawa Barat*. Koran Perdjoeangan, Jakarta. https://www.koranperdjoeangan.com/menguak-fakta-tingginya-pengangguran-di-jawa-barat-2/ . [11 September 2020]

[9] Dinas Tenaga Kerja Kabupaten Buleleng. 2019. *Banyaknya Pengangguran karana kurangnya Pelatihan keterampilan kerja*. Website Pemerintah Kabupaten Buleleng, Buleleng. https://www.bulelengkab.go.id/detail/artikel/banyaknya-pengangguran-karana-kurangnya-pelatihan-keterampilan-kerja-11 . [10 September 2020]

[10] Dewi, N.K., Syafitri, U.D., & Mulyadi, S.Y., 2011. Penerapan Metode Random Forest dalam Driver Analysis (The Application of Random Forest in Driver Analysis). *Forum Statistika dan Komputasi*, 16 (1), 35-43

[11] Hasby, M., 2019. *Faktor- Faktor yang Memengaruhi Pengangguran Pemuda di Provinsi Banten tahun 2018 (Skripsi).* Politeknik Statistika STIS.

[12] Jacob, M., & Kleinert, C. (2014). Marriage, Gender, and Class: The Effects of Partner Resources on Unemployment Exit in Germany. *Social Forces, 92*(3), 839-871.

[13] Kaufman, B.E., & Hotchkiss, J.L., 1999. *The Economic Labor Markets*. Georgia State University.

[14] Lavrinovicha, I., Lavrinenko, O., & Treinovskis, J.T., 2015. Influence of Education on Unemployment Rate and Incomes of Residents. *Procedia - Social and Behavioral Sciences,* 174, 3824 – 3831

[15] Lipsey, R.G., Purvis, D.D., Courant, P.N., & Steiner, P.O., 1997. *Pengantar Makroekonomi. Jilid kedua. Agus Maulana [penerjemah]*. Binarupa Aksara.

[16] Lunardon, N., Menardi, G., & Torelli, N. 2014. ROSE: A Package for Binary Imbalanced Learning. *The R Journal*, 6(1), 79.

[17] Maulida, I., & Nooraeni, R., 2020. Penerapan Metode Random Forest Untuk Klasifikasi Wanita Usia Subur di Perdesaan Dalam Menggunakan Internet (SDKI 2017). *Jurnal Matematika Dan Statistika Serta Aplikasinya*, 8(1), 72-76.

[18] Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. Data Mining and Knowledge Discovery, 28(1), 92-122.

[19] Mishra, S. 2017. Handling imbalanced data: SMOTE vs. random undersampling. *International Research Journal of Engineering and Technology*, 4(8), 317-320.

[20] Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R., 2018. *Data Mining dengan R Konsep Serta Implementasi*. InMedia.

[21] Purwa, T., 2019. Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017). *Jurnal Matematika, Statistika Dan Komputasi*, 16(1), 58.

[22] Putra, D.A., 2018. *Pemerintah Jokowi Ciptakan 10,34 Juta Lapangan Kerja Baru Sejak 2015 Hingga 2018*. Merdeka.com, Jakarta. https://www.merdeka.com/uang/pemerintah-jokowi-ciptakan-1034-juta-lapangan-kerja-baru-sejak-2015-hingga-2018.html . [10 September 2020]

[23] Sartono B, Syafitri UD. 2010. *Ensemble Tree: an Alternative toward Simple Classification & Regression Tree*. *Forum Statistika dan Komputasi*. 15(1):1-7.

[24] Sukirno, S., 2000. *Makro Ekonomi Modern*. Raja Grafindo Persada.

[25] Sukirno, S., 2006. *Makroekonomi Teori Pengantar*. Raja Grafindo Persada.