

Perbandingan *Hot-deck*, *Support Vector Machine*, dan *Random Forest* dalam Mengidentifikasi Industri Mikro dan Kecil Terdampak Covid-19 Tahun 2020

Penerapan Pada Data Survei Industri Mikro dan Kecil Tahunan 2020

(*Comparison of Hot-deck, SVM, and Random Forest in Identifying Micro and Small Industry Affected by Covid-19 in 2020*)

Iman Jihad Fadillah^{1*}, Lalu Moh. Arsal Fadila², Lalu Muhamad Winadi Darundiye²

¹Badan Pusat Statistik

²Politeknik Statistika STIS

DKI Jakarta, Indonesia

Email: jihadiman22@gmail.com

ABSTRAK

Penyebaran Covid-19 telah dinyatakan sebagai pandemi sejak Maret 2020. Pandemi yang dibarengi kebijakan oleh pemerintah berakibat pada penurunan di sektor ekonomi, khususnya di industri mikro dan kecil (IMK). Mengidentifikasi IMK terdampak pandemi Covid-19 menjadi salah satu langkah yang penting. Ada dua jenis metode untuk mengidentifikasi yang umum digunakan, yaitu metode berbasis statistika dan metode berbasis *machine learning*. Setiap metode memiliki hasil pengukuran yang berbeda-beda. Oleh karena itu, diperlukan suatu metode yang tepat untuk mengidentifikasi IMK yang terdampak pandemi Covid-19. Penelitian ini bertujuan untuk membandingkan metode *hot-deck*, SVM dan *random forest*, sehingga diperoleh metode yang terbaik untuk mengidentifikasi IMK terdampak Covid-19. Hasil yang didapatkan bahwa metode *random forest* merupakan metode yang terbaik dalam mengidentifikasi IMK terdampak Covid-19.

Kata kunci: IMK, *Hot-deck*, SVM, *Random Forest*, Covid-19

ABSTRACT

The spread of Covid-19 has been declared a pandemic since March 2020. The pandemic coupled with policies by the government resulted in a decline in the economic sector, especially in micro and small industries (IMK). Identifying IMK affected by the Covid-19 pandemic is an important step. There are two types of identification methods that are commonly used, namely statistical-based methods and machine learning-based methods. Each method has different measurement results. Therefore, an appropriate method is needed to identify IMKs affected by the Covid-19 pandemic. This study aims to compare the hot-deck, SVM and random forest methods, in order to obtain the best method to identify IMK affected by Covid-19. The results obtained are that the random forest method is the best method in identifying IMK affected by Covid-19.

Keywords: IMK, *Hot-deck*, SVM, *Random Forest*, Covid-19

PENDAHULUAN

Penyebaran Covid-19 atau *Corona Virus Disease 2019* telah dinyatakan sebagai pandemi global oleh *World Health Organization* (WHO) sejak Maret 2020. Pemerintah Indonesia juga menetapkan Covid-19 sebagai kedaruratan kesehatan masyarakat melalui Keputusan Presiden Nomor 11 Tahun 2020 tentang Penetapan Kedaruratan Kesehatan Masyarakat *Corona Virus Disease 2019* (COVID-19) serta bencana nonalam berdasarkan Keputusan Presiden Nomor 12 Tahun 2020 tentang Penetapan Bencana Nonalam. Sebagai tindak lanjut penetapan tersebut, pemerintah melakukan berbagai kebijakan untuk mencegah penyebaran Covid-19 diantaranya melakukan pembatasan jarak sosial serta karantina wilayah.

Pandemi yang dibarengi kebijakan pembatasan oleh pemerintah berakibat pada penurunan di berbagai sektor di Indonesia, utamanya adalah sektor ekonomi. Hal itu dapat terlihat dari pertumbuhan ekonomi Indonesia berkontraksi sebesar 2,07 persen pada tahun 2020. Sektor industri pengolahan merupakan sektor yang paling terdampak dari krisis ekonomi tersebut. Pertumbuhan sektor Industri pengolahan mengalami kontraksi lebih dalam, yaitu sebesar - 2,93 persen. Selain itu, BPS (2021) menunjukkan jumlah pekerja pada sektor industri pengolahan terjadi penurunan akibat efek pandemi sebesar 8,93 persen. Sebagian pabrik mulai tutup secara permanen atau tutup sementara. Sebagian pabrik juga secara bertahap mulai mengurangi jumlah pekerja dan produksinya akibat dari situasi dan permintaan yang menurun.

Industri mikro dan kecil (IMK) yang merupakan bagian dari industri pengolahan juga mengalami dampak dari pandemi ini. Menurut BPS (2021), angka pertumbuhan produksi usaha IMK terendah terjadi pada triwulan II-2020 yaitu sebesar -21,31 persen. Produksi usaha IMK yang mengalami kontraksi berakibat pada terjadi penutupan usaha. Sampai dengan akhir tahun 2020, sekitar 11,25 persen tutup sementara, bahkan 7,06 persen usaha IMK tutup. Selain itu, Menurut BPS (2022), usaha IMK mengaku mengalami banyak kesulitan pada kegiatan pemasaran, permodalan serta bahan baku. Produksi usaha IMK yang menurun dan perubahan jumlah usaha IMK yang fluktuatif serta kendala yang dialami menunjukkan bahwa usaha IMK masih sangat tergantung pada kondisi pandemi dan kebijakan dari pemerintah. Ketika kondisi pandemi Covid-19 memburuk, yang ditindaklanjuti dengan kebijakan pengetatan aktivitas ekonomi, maka usaha IMK langsung terdampak.

Berdasarkan uraian tersebut, mengidentifikasi industri mikro dan kecil yang terdampak pandemi Covid-19 menjadi salah satu langkah yang penting, sehingga diperlukan suatu metode yang tepat untuk mengidentifikasi masalah ini. Kemajuan di bidang statistik berbasis teknologi dapat menjadi alat untuk mengetahui permasalahan tersebut. Ada dua jenis metode klasifikasi utama yang umum digunakan, yaitu metode berbasis statistika dan metode berbasis *machine learning* (He et al., 2018). Salah satu metode klasifikasi berbasis statistika adalah metode *hot-deck*. Kelebihan dari metode *hot-deck* yaitu metode ini digunakan pada banyak jenis data dengan penggunaannya relatif sederhana dan tidak diperlukannya asumsi yang rumit dibanding metode lainnya (Fadillah dan Muchlisoh, 2020). Fadillah dan Puspita (2022) memanfaatkan metode *hot-deck* untuk mengidentifikasi permasalahan pengklasifikasian Klasifikasi Baku Lapangan Usaha Indonesia (KBLI). Penelitian tersebut menyatakan bahwa *hot-deck* dapat menghasilkan klasifikasi yang tinggi dan *precise* pada identifikasi KBLI. Selain itu, Raudhatunnisa dan Wilantika (2022) melakukan perbandingan antara metode *hot-deck*, *K-nearest neighbor* (KNN), dan *predictive mean matching*. Penelitian tersebut menyimpulkan bahwa metode *hot-deck* merupakan metode terbaik dibandingkan dua metode lainnya.

Metode klasifikasi berikutnya adalah berbasis *machine learning*. Ada berbagai macam metode klasifikasi berbasis *machine learning*. Contohnya adalah metode *random forest* dan *support vector machine* (SVM) yang diteliti oleh Iman dan Wijayanto (2021). Penelitian tersebut menyimpulkan bahwa kedua metode klasifikasi tersebut memiliki kinerja yang setara dan sama baiknya dalam melakukan klasifikasi rumah tangga penerima beras miskin (raskin)/beras sejahtera (rastra). Sementara itu, Sihombing dan Yulianti (2021) melakukan perbandingan antara metode *classification and regression tree* (CART), *Naïve Bayes*, *random forest* dan SVM. Hasil dari penelitian tersebut menyatakan bahwa *random forest* adalah metode yang memiliki kinerja terbaik dibandingkan metode klasifikasi lainnya. Penelitian lain yang dilakukan oleh Geraldly dan Maghfiroh (2021) menyatakan bahwa SVM adalah metode yang memiliki kinerja terbaik dibandingkan metode *naïve-Bayes*, KNN, *random forest*, dan *gradient boosting*. Menurut Iman dan Wijayanto (2021), *random forest* memiliki kelebihan diantaranya akurasi sangat baik dibandingkan metode klasifikasi lainnya, dapat menangani data dengan sampel besar, mengatasi data dengan *noise* dan *missing data*, serta *error* yang dihasilkan relatif rendah. Sementara itu metode SVM efektif pada data dengan dimensi tinggi serta Implementasi metode ini relatif mudah.

Beberapa penelitian terdahulu tersebut telah mencoba membandingkan banyak metode klasifikasi dan menyatakan bahwa metode *hot-deck*, SVM dan *random forest* merupakan metode klasifikasi terbaik. Namun, hasil klasifikasi suatu metode dapat berbeda jika diterapkan pada dataset yang berbeda pula. Selain itu, masih jarang ditemukan penelitian terdahulu yang membandingkan ketiga metode tersebut. Oleh karena itu, pada penelitian ini, peneliti bertujuan untuk membandingkan beberapa metode klasifikasi yaitu, metode *hot-deck*, SVM dan *random forest*, sehingga diperoleh metode yang terbaik untuk mengidentifikasi IMK terdampak Covid-19.

METODE

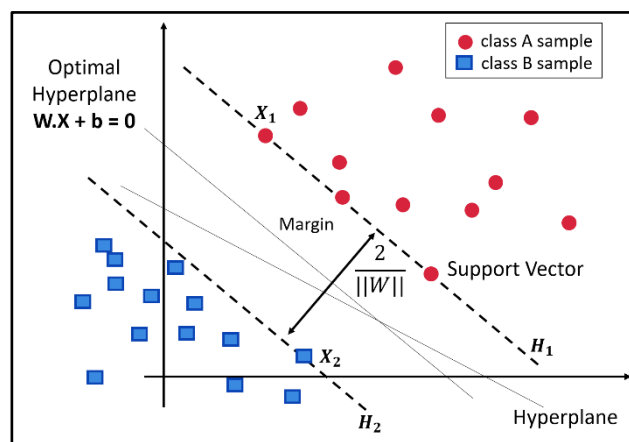
Hot-deck

Hot-deck merupakan suatu metode yang algoritmanya memanfaatkan konsep *similarity* (Fadillah dan Muchlisoh, 2020). *Hot-deck* dapat diterapkan dalam dua tahap (Batista dan Monard, 2002), tahap pertama membentuk *cluster-cluster*, dan tahap kedua melakukan donor dari nilai yang dianggap memiliki unsur *similarity*. Pembentukan *cluster* dilakukan dengan mengurutkan data menggunakan variabel prediktor. Pengurutan data menggunakan variabel prediktor dipilih berdasarkan hubungannya dengan variabel yang menjadi amatan (Grau et al., 2004). Fadillah (2019) memberikan contoh, jika suatu data yang sudah diurutkan memiliki jumlah observasi sebanyak 6. Kemudian 3 diantaranya adalah data *testing*, dimana y_1 , y_4 dan y_5 adalah data *train*, sedangkan y_2 , y_3 , dan y_6 adalah data *testing*. Maka, y_2 , y_3 diprediksi oleh y_1 , dan y_6 diprediksi oleh y_5 . Jika data *testing* digunakan pada y_1 maka beberapa nilai awal kemungkinan diperlukan, dapat berasal dari catatan survei sebelumnya atau menggunakan nilai data setelahnya yang tersedia yaitu y_2 .

Hot-deck dapat diterapkan pada data bertipe numerik, kategorik, maupun *mixed data*. Akan tetapi, dalam prakteknya, *hot-deck* masih jarang digunakan dalam proses klasifikasi. Salah satu penelitian yang memanfaatkan metode *hot-deck* untuk mengklasifikasikan KBLI adalah penelitian yang dilakukan oleh Fadillah dan Puspita (2022). Penelitian tersebut menyatakan bahwa *hot-deck* dapat menghasilkan klasifikasi yang tinggi dan *precise* pada KBLI 2 digit dan 5 digit.

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan suatu algoritma pembelajaran mesin yang digunakan dalam masalah klasifikasi dua kelas (Vapnik, 1995). Magnin et.al (2009) menyatakan bahwa SVM mengimplementasikan vektor dipetakan secara nonlinier ke ruang fitur berdimensi sangat tinggi. Pada fitur ini, permukaan pemisahan linier dibuat untuk memisahkan data *training* dengan meminimalkan *margin* antara vektor dari dua kelas. Ketika melakukan suatu pengklasifikasian, jarak antar kelas pada SVM dilakukan dengan mencari *hyperplane* terbaik (Franklin, 2005). Untuk mendapatkan *hyperplane* terbaik yang dimana memiliki kesalahan klasifikasi minimum, maka dicari *hyperplane* marginal maksimum atau yang disebut dengan *maximum marginal hyperplane* (MMH) (Arista, 2021). Berikut merupakan ilustrasi dari *hyperplane* yang terdapat pada Gambar 1.



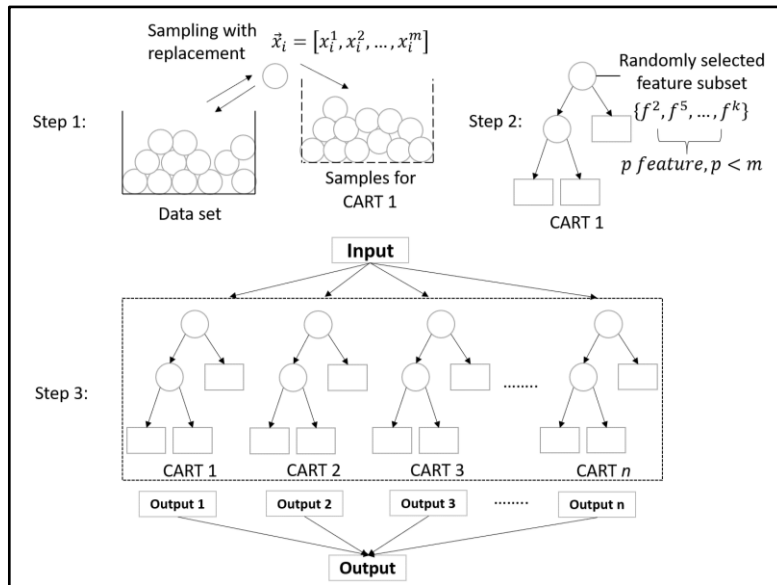
Sumber: researchget.net

Gambar 1. SVM dengan *Maximum Marginal Hyperlane* (MMH)

Arista (2021) menyatakan bahwa SVM memiliki kelebihan dalam melakukan klasifikasi, yakni memiliki banyak opsi kernel. Tujuan dari penggunaan kernel yaitu mengimplementasikan suatu model pada ruangan yang lebih tinggi (*feature space*) tanpa harus mendefinisikan fungsi pemetaan dari ruang input ke *feature space*, sehingga untuk kasus nonlinier dapat menjadi linear pada *feature space*. Ada berbagai macam fungsi kernel yaitu linear, polynomial, radial basis function, dan sigmoid.

Random Forest

Random forest merupakan suatu algoritma dalam melakukan klasifikasi yang termasuk *ensemble learning* (Mukhlis, 2016). Pada algoritma ini terdapat *k* pohon dengan vektor random yang independen dengan vektor-vektor random sebelumnya, tetapi memiliki distribusi yang identik (Breiman, 2001). Breiman (2001) juga menyatakan bahwa metode ini memanfaatkan algoritma *decision tree* dalam melakukan klasifikasi. Kemudian dibentuk sebuah model dengan menerapkan metode *bootstrap aggregating* ketika membentuk sebuah sampel *training set*, dan setiap *tree* yang dibentuk menggunakan metode yang sama untuk membangun CART (*Classification and Regression Tree*). Berikut adalah proses dari algoritma *random forest* yang terlihat pada Gambar 2.



Sumber : Lin, et. al (2016)

Gambar 2. Proses Algoritma *Random Forest*

Dalam penggunaan metode *random forest*, diperlukan variabel prediktor yang digunakan sebagai prediksi dalam pembentukan *decision tree* pada model *random forest*. Pentingnya suatu variabel prediktor dilihat dari nilai skor *important* (Schratz et al., 2022). Semakin tinggi skornya, maka semakin penting variabel prediktor tersebut. Kemudian untuk variabel prediktor yang digunakan merupakan variabel yang berhubungan dengan variabel yang akan diprediksi (Hendrawati, 2015).

Parameter Pengukuran Evaluasi Model

Confussion matrix merupakan tabel yang digunakan untuk menggambarkan kinerja pengklasifikasian. Pada tabel tersebut memberikan informasi mengenai kelas yang diprediksi dan kelas yang sebenarnya yang didalamnya terdapat jumlah yang benar dan tidak benar untuk dikelompokkan (Arista, 2021). Beberapa pengukuran yang dihasilkan dari *confussion matrix* adalah *accuracy*, *recall*, *specificity*, *positive predicted value* (PPV), dan *negative predicted value* (NPV) (Marjia Sultana; Afrin Hider; Mohammad Shorif Uddin, 2016; I.D. Dinov, 2018).

Tabel 1. Confussion Matrix

| | | Kelas Sebenarnya | |
|----------------|----------|---------------------|---------------------|
| | | Positive | Negative |
| Kelas Prediksi | Positive | True Positive (TP) | False Positive (FP) |
| | Negative | False Negative (FN) | True Negative (TN) |

Secara matematik *accuracy*, *recall*, *specificity*, PPV, dan NPV dirumuskan sebagai berikut:

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

$$recall = \frac{TP}{TP+FN} \tag{2}$$

$$specificity = \frac{TN}{FP+TN} \tag{3}$$

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

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

Nilai *accuracy* menunjukkan hasil ketepatan dari prediksi model yang dibentuk, dimana hasil prediksi yang benar dibandingkan keseluruhan total data. Kemudian nilai *recall* mengukur rasio dari hasil kelas yang benar diprediksi *positive* dengan jumlah dari kelas *positive*. Ukuran ini bisa disebut dengan *false positive rate* (FPR). Sementara itu, *specificity* menghitung rasio dari hasil kelas yang benar diprediksi *negative* dengan jumlah dari kelas *negative*. Ukuran ini bisa disebut dengan *true positive rate* (TPR). Selanjutnya nilai PPV menunjukkan rasio dari kelas yang benar diprediksi *positive* dan diprediksi secara benar. Nilai NPV menunjukkan rasio dari kelas yang benar diprediksi *negative* dan diprediksi secara benar.

Analisis Data

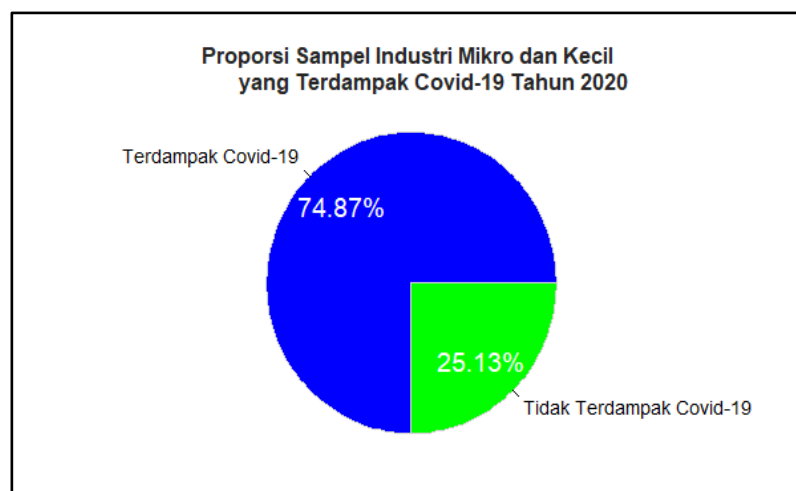
Analisis data pada penelitian ini memanfaatkan R-Studio, secara berurutan, metode *hot-deck*, SVM, dan *random forest*, menggunakan *package VIM, e1071, dan randomForest* untuk melakukan proses klasifikasi pada ketiga metode tersebut. Sebelum melakukan klasifikasi, ditentukan variabel *important* melalui perhitungan skor *important* pada variabel-variabel prediktor. Kemudian, data *training* dan data *testing* ditentukan terlebih dahulu yakni sebesar 75 persen dan 25 persen. Metode *hot-deck* mengacu pada *sequential hot-deck, random forest* dibentuk 500 *tree* untuk keperluan beban komputasi, dan SVM menggunakan *hyperplane* model fungsi kernel linear dengan *cost* sebesar 1. Selanjutnya adalah membandingkan hasil klasifikasi berdasarkan ukuran nilai dari *confusion matrix*. Sementara Sun et.al (2007) menyatakan bahwa asumsi dalam melakukan klasifikasi adalah data yang digunakan tidak mengandung *imbalanced class* yang dimana satu kelas memiliki jumlah sampel yang besar dan kelas lainnya hanya memiliki jumlah sampel yang sedikit.

Data dan Sumber Data

Data yang digunakan pada penelitian ini bersumber pada Survei IMK Tahunan 2020 BPS. Unit observasi pada survei ini adalah usaha/perusahaan IMK. Jumlah data yang digunakan sebanyak 88.883 *record*. Adapun variabel yang digunakan adalah industri yang terdampak Covid-19, pendapatan usaha/perusahaan IMK, klasifikasi industri, rata-rata hari kerja per bulan, KBLI 2-digit, dan bahan baku. Variabel industri yang terdampak Covid-19 digunakan sebagai variabel klasifikasi, sedangkan yang lainnya merupakan variabel prediktor.

HASIL DAN PEMBAHASAN

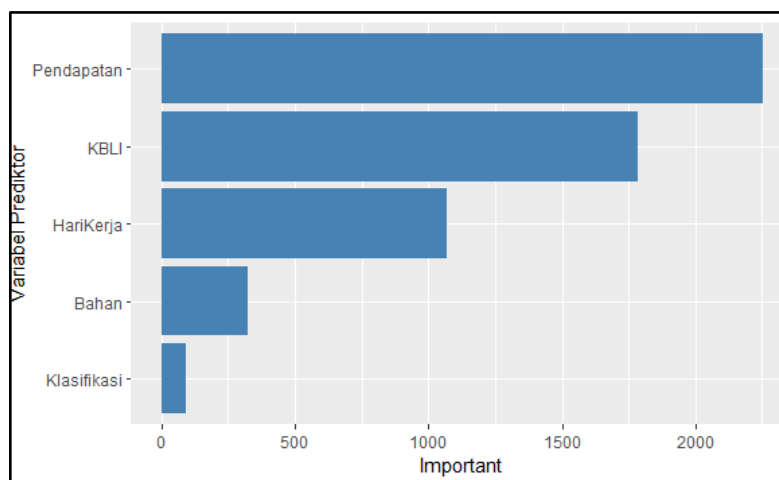
Analisis data pada penelitian ini menggunakan data IMK terdampak Covid-19. Dari jumlah data sebanyak 88.883 usaha/perusahaan IMK, 74,87 persen terdampak Covid-19. Sedangkan sisanya, 25,13 persen tidak terdampak Covid-19. Hal ini menunjukkan bahwa hampir tiga dari empat usaha IMK terdampak Covid-19 pada tahun 2020. Sebagaimana digambarkan pada Gambar 5, berdasarkan hasil tersebut, proporsi data IMK terdampak Covid-19 sebagai variabel klasifikasi dianggap tidak memiliki *imbalanced class* karena sampel yang digunakan sangat besar.



Gambar 3. Proporsi Sampel Industri Mikro dan Kecil yang Terdampak Covid-19 Tahun 2020

Kemudian, untuk mengetahui variabel prediktor yang dapat digunakan untuk pembentukan *decision tree* pada model *random forest*, maka ditentukan melalui skor *important* dengan terbentuk 500 *tree*. Berdasarkan

Gambar 4 diperoleh bahwa urutan variabel prediktor yang paling dapat digunakan adalah pendapatan usaha, kemudian KBLI 2-digit, rata-rata hari kerja per bulan, bahan baku, dan terakhir yaitu klasifikasi industri.



Gambar 4. Grafik Skor *Important* Variabel Prediktor

Berdasarkan hasil perhitungan yang telah dilakukan, didapatkan *confussion matrix* untuk ketiga metode klasifikasi yang masing-masing dijelaskan pada Tabel 2, Tabel 3, dan Tabel 4. Dari ketiga tabel tersebut, metode *random forest* menghasilkan *true positive* terbanyak, sedangkan metode *hot-deck* yang menghasilkan *true negative* paling banyak.

Tabel 2. Confussion Matrix Hot-Deck

| | | Kelas Sebenarnya | |
|----------------|-----------------|------------------|-----------------|
| | | Terdampak | Tidak Terdampak |
| Kelas Prediksi | Terdampak | 12805 | 3774 |
| | Tidak Terdampak | 3995 | 1646 |

Tabel 3. Confussion Matrix SVM

| | | Kelas Sebenarnya | |
|----------------|-----------------|------------------|-----------------|
| | | Terdampak | Tidak Terdampak |
| Kelas Prediksi | Terdampak | 16570 | 5097 |
| | Tidak Terdampak | 230 | 323 |

Tabel 4. Confussion Matrix Random Forest

| | | Kelas Sebenarnya | |
|----------------|-----------------|------------------|-----------------|
| | | Terdampak | Tidak Terdampak |
| Kelas Prediksi | Terdampak | 16665 | 4677 |
| | Tidak Terdampak | 135 | 743 |

Melalui hasil perhitungan dari ketiga *confussion matrix* tersebut, dapat dilakukan perhitungan untuk masing-masing nilai pengukuran. Adapun hasil *accuracy*, *recall*, *specificity*, *positive predicted values*, dan

negative predicted values untuk masing-masing metode, yakni *hot-deck* SVM, dan *random forest*, pada pengklasifikasian IMK yang terdampak Covid-19 di Indonesia pada tahun 2020 ditampilkan pada Tabel 5.

Tabel 5. Perbandingan Parameter Pengukuran Evaluasi Model

| Parameter | <i>hot-deck</i> | SVM | <i>Random Forest</i> |
|--------------------|-----------------|--------|----------------------|
| <i>Accuracy</i> | 0,6504 | 0,7603 | 0,7834 |
| <i>Recall</i> | 0,7622 | 0,9863 | 0,9920 |
| <i>Specificity</i> | 0,3037 | 0,0596 | 0,1371 |
| <i>PPV</i> | 0,7724 | 0,7648 | 0,7809 |
| <i>NPV</i> | 0,2918 | 0,5841 | 0,8462 |

Pada Tabel 5 diketahui bahwa nilai *accuracy* dalam pengklasifikasian IMK terdampak Covid-19 di Indonesia tahun 2020 yang dihasilkan oleh *random forest* lebih baik daripada *hot-deck* dan SVM. Dimana tingkat *accuracy* pada *random forest* sebesar 0,7834 lebih tinggi dibandingkan dengan nilai *accuracy* dari *hot-deck* dan SVM yang sebesar 0,6504 dan 0,7603. Hasil ini sejalan dengan penelitian Sihombing dan Yuliaty (2021) terkait kinerja dari metode *random forest*. Selain itu, menunjukkan adanya perbedaan kinerja antara metode SVM dan *random forest* pada dataset penelitian ini dibandingkan penelitian oleh Iman dan Wijayanto (2021) yang menggunakan dataset yang berbeda. Jika dilihat dari nilai *recall*, metode *random forest* menghasilkan hasil yang lebih baik dibandingkan kedua metode, yakni sebesar 0,9920. Akan tetapi tidak jauh berbeda dengan nilai *recall* SVM yang sebesar 0,9863. Sedangkan Metode *hot-deck* menghasilkan nilai *recall* yang jauh lebih rendah yakni 0,7622. Sementara itu, jika dilihat dari pengukuran lainnya, metode *hot-deck* menghasilkan nilai *specificity* yang lebih tinggi daripada kedua metode lainnya, yakni sebesar 0,3037. Serupa dengan hasil sebelumnya jika dilihat dari nilai PPV dan NPV, metode *random forest* menghasilkan nilai *predicted values* yang sedikit lebih baik dibandingkan kedua metode, yakni PPV sebesar 0,7809 dan NPV sebesar 0,8462. Dengan demikian, secara keseluruhan, dalam mengidentifikasi IMK terdampak Covid-19 tahun 2020, metode *random forest* merupakan metode yang terbaik dibandingkan metode *hot-deck* dan SVM.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, perbandingan metode *hot-deck*, SVM, dan *random rorest* dalam mengidentifikasi IMK terdampak Covid-19 tahun 2020 menunjukkan bahwa, metode *random forest* menghasilkan hasil klasifikasi yang secara konsisten lebih baik dibandingkan metode *hot-deck* dan SVM dan baik dari sisi *accuracy*, *positif rate (recall dan specificity)*, dan *predicted values* (PPV dan NPV). Adapun hasil klasifikasi dari metode SVM memberikan hasil yang cukup baik, jika dibandingkan dengan metode *hot-deck*. Kemudian metode *hot-deck* menunjukkan hasil klasifikasi yang tidak lebih baik dibandingkan kedua metode lainnya.

Metode *random forest* memberikan hasil yang konsisten dan prediksi pengklasifikasian yang akurat dilihat dari nilai *accuracy*, *positif rate (recall dan specificity)*, dan *predicted values* (PPV dan NPV) yang tinggi. Hal ini menunjukkan bahwa penggunaan metode *random forest* dapat menjadi salah satu metode yang digunakan sebagai solusi untuk menangani permasalahan dalam mengidentifikasi IMK yang terdampak Covid-19.

DAFTAR PUSTAKA

- Batista, G. E. A. P. A., & Monard, M. C. (2002). A study of k-nearest neighbour as an imputation method. *Frontiers in Artificial Intelligence and Applications*, 87(May 2014), 251–260.
- BPS. (2021). *Industri Mikro dan Kecil di Masa Pandemi Covid-19 2020*.
- BPS. (2022). *Profil Industri Mikro dan Kecil 2020*. Badan Pusat Statistik.
- Breiman, L. (2001). Random Forest. *Machine Learning*, 45, 5–32.
- Fadillah, I. J. (2019). Perbandingan Metode Hot-Deck Imputation Dan Metode Knni Dalam Mengatasi Missing

Values. *Skripsi, June*.

- Fadillah, I. J., & Muchlisoh, S. (2020). Perbandingan Metode Hot-Deck Imputation Dan Metode Knni Dalam Mengatasi Missing Values. *Seminar Nasional Official Statistics*, 2019(1), 275–285. <https://doi.org/10.34123/semnasoffstat.v2019i1.101>
- Fadillah, I. J., & Puspita, C. D. (2022). Application of The Sequential Hot-deck Imputation Method for Identification of Indonesian Standard Classification of Business Fields (KBLI). *Proceedings of The International Conference on Data Science and Official Statistics*, 2022(1), 734–741. <https://doi.org/10.34123/icdsos.v2021i1.70>
- Franklin, J. (2005). The Elements of Statistical Learning : Data Mining , Inference and Prediction Probability Theory : The Logic of Science The Fundamentals of Risk Measurement Mathematicians , pure and applied , think there is something weirdly different about. *The Mathematical Intelligencer*, 27(2), 83–85. <http://link.springer.com/article/10.1007/BF02985802?LI=true#>
- Gerald, H., & Rahmatuti Maghfiroh, L. (2021). Penerapan Klasifikasi Kueri untuk Meningkatkan Efektivitas Mesin Pencari (Implementation of Query Classification to Improve Effectiveness of Search Engine). *Seminar Nasional Official Statistics*, 1012–1018. www.bps.go.id
- Grau, E. A., Frechtel, P., Odom, D. M., & Painter, D. (2004). a Simple Evaluation of the Imputation Procedures Used in Nsduh. *Proceedings of the Survey Research Methods Section, American Statistical Association, January*, 3588–3595.
- He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, 98, 105–117. <https://doi.org/10.1016/j.eswa.2018.01.012>
- Hendrawati, T. (2015). Kajian Metode Imputasi Dalam Menangani Missing Data. *Prosiding Seminar Nasional Statistika | Dapertemen Statistika FMIPA Universitas Padjadjaran*, 637–642.
- I.D. Dinov. (2018). Model Performance Assessment. *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*, 475–496.
- Iman, Q., & Wijayanto, A. W. (2021). Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 9(2), 178. <https://doi.org/10.26418/justin.v9i2.44137>
- Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélérini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., & Benali, H. (2009). Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2), 73–83. <https://doi.org/10.1007/s00234-008-0463-x>
- Marjia Sultana; Afrin Hider; Mohammad Shorif Uddin. (2016). Analysis of Data Mining Techniques for Disease Prediction. *Conf. Electr. Eng. Inf. Commun. Technol.* <https://doi.org/0.1109/CEEICT.2016.7873142>.
- Mukhlis. (2016). Kajian Metode Klasifikasi Support Vector Machine Dengan Velocity-based Fruit Fly Optimization Algorithm dan Random Forest Pada Data Berdimensi Tinggi. *Sekolah Tinggi Ilmu Statistik*.
- Namira Arista. (2021). Pendeteksian Kecenderungan Depresi Pada Pengguna Twitter Menggunakan Support Vector Machine (SVM). *Sekolah Tinggi Ilmu Statistik*.
- Raudhatunnisa, T., & Wilantika, N. (2022). Performance Comparison of Hot-Deck Imputation, K-Nearest Neighbor Imputation, and Predictive Mean Matching in Missing Value Handling, Case Study: March 2019 SUSENAS Kor Dataset. *Proceedings of The International Conference on Data Science and Official Statistics*, 2021(1), 753–770. <https://doi.org/10.34123/icdsos.v2021i1.93>
- Schratz, P., Lang, M., Bischl, B., & Binder, M. (2022). Filter Based Feature Selection for “mlr3.” *CRAN*.
- Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 417–426. <https://doi.org/10.30812/matrik.v20i2.1174>
- Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12), 3358–3378. <https://doi.org/10.1016/j.patcog.2007.04.009>
- Vapnik, C. C. V. (1995). Support Vector Network. *Machine Learning*, 20, 273–292.