IJSA (2024), Vol. 8, No. 01, 37-46
DOI: https://doi.org/10.29244/ijsa.v8i1p37-46

Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

Application of Random Forest Method to Identify Food and Beverage Industries Experiencing Raw Material Difficulties

Penerapan Metode Random Forest untuk Mengidentifikasi Industri Makanan dan Minuman yang Mengalami Kesulitan Bahan Baku

Iman Jihad Fadillah[1*], Indah Noor Safrida[2], Rima Kusumaningtyas[3]

[1,2,3]BPS-Statistics Indonesia, Indonesia
[*]corresponding author: jihadiman22@gmail.com

Abstract

The food and beverage industry experienced a significant increase after the pandemic. However, challenges continue to hit this industry, especially for micro and small-scale businesses. To overcome this problem, the right approach is needed. One of the first steps is to provide quality data as a basis for decision-making and problem-solving. However, statistical activities such as censuses and surveys often need help in the form of missing values. One effective method for dealing with this is using the random forest method. This research aims to use a machine learning-based imputation method, namely the random forest method, to identify micro and small-scale food and beverage industries that are experiencing raw material difficulties. This research uses data from the Micro and Small Industry Survey 2020, with a total of 26,058 records observed. The variables used include the food and beverage industry experiencing raw material difficulties, business/company income, 5-digit KBLI, average working days in a month, industry classification, and industries affected by COVID-19. The accuracy resulting from the random forest method in handling the classification of this data is 80.37%. Meanwhile, the imputation time required for classification is 281.62 seconds. The research results show that the random forest method provides accurate and consistent predictions in identifying food and beverage industries experiencing raw material difficulties. However, it is also necessary to consider the relatively long computing time required to implement this method.

Keywords: food and beverage industry, missing value, random forest.

[*] Received: Okt 2023; Reviewed: Mar 2024; Published: Juni 2024

1. Pendahuluan

Industri makanan dan minuman terus mengalami peningkatan paskapandemi. Badan Pusat Statistik (2022) mencatat industri makanan dan minuman di Indonesia tumbuh sebesar 2,54 persen pada tahun 2021 dan meningkat menjadi 4,90 persen dari tahun 2021 ke 2022. Industri ini masih menjadi kontributor terbesar pada nilai PDB subsektor industri pengolahan. Pada triwulan I-2023, industri makanan dan minuman menyumbang 6,47 persen pada perekonomian nasional (Badan Pusat Statistik, 2023a). Dikarenakan secara konsisten memberikan kontribusi besar pada kinerja industri pengolahan, industri makanan dan minuman dipilih sebagai industri prioritas dalam Rencana Induk Pembangunan Industri Nasional (RIPIN) 2015-2035 (Kemenperin, 2015).

Meskipun demikian, industri makanan dan minuman sering kali mengalami kesulitan dalam keberlangsungannya, terutama untuk skala mikro dan kecil. Salah satu kendala yang dialami adalah ketersediaan bahan baku. Data BPS tahun 2020 menunjukkan bahwa sekitar 20,55 persen industri makanan dan minuman skala mikro dan kecil mengalami kesulitan bahan baku (Badan Pusat Statistik, 2023b). Industri makanan dan minuman sangat erat kaitannya dengan sektor pertanian, dimana bahan-bahan utamanya disokong oleh produk dari pertanian, utamanya subsektor pertanian dan perikanan. Adanya perlambatan pertumbuhan pada sektor pertanian pada tahun 2020 hingga 2021 memberikan dampak terhadap pertumbuhan industri makanan dan minuman. Permasalahan ini tentunya memerlukan penanganan sehingga dapat menanggulangi permasalahan ketersediaan bahan baku untuk keberlangsungan industri makanan dan minuman.

Penanganan awal yang dapat dilakukan pada masalah tersebut adalah dengan menyediakan data yang berkualitas untuk dasar pengambilan keputusan dan penyelesaian masalah. Han et al. (2012) menjelaskan bahwa salah satu ciri data yang berkualitas adalah kelengkapan/*completeness*. Namun permasalahan yang sering ditemui pada kegiatan statistik baik sensus maupun survei adalah adanya *missing values* (data hilang/tidak lengkap). Biemer & Lyberg (2003) menyatakan bahwa *missing values* ditemukan hampir di semua usaha pengumpulan data berskala besar dan dapat menjadi masalah dalam pelaksanaan survei atau sensus. Banyak alasan terjadinya *missing values* pada data survei. Menurut Kaiser (2014), *missing values* dapat terjadi karena responden tidak menjawab semua pertanyaan pada kuesioner, kesalahan eksperimen, data yang disensor atau tidak dikenal, dan lainnya.

Salah satu metode yang dapat digunakan untuk mengatasi *missing values* adalah dengan imputasi, yaitu mengisi atau mengganti nilai yang hilang pada suatu dataset tanpa mengurangi jumlah unit data yang diobservasi. Menurut Jerez et al. (2010), metode imputasi dapat digolongkan menjadi dua, yaitu metode imputasi berbasis statistik dan *machine learning*. Metode imputasi berbasis statistik melakukan proses imputasi menggunakan kaidah-kaidah analisis statistik, seperti *mean imputation*, *hot-deck imputation*, dan *multiple imputation*. Sementara metode imputasi berbasis *machine learning* melakukan proses imputasi dengan memanfaatkan pembelajaran/*learning machine* untuk memprediksi nilai yang hilang. Beberapa

contoh metode imputasi berbasis *machine learning* yaitu *K-Nearest Neighbour Imputation* (KNNI), *Self-organisation maps* (SOM), *Support Vector Machine* (SVM), CART*, Naïve Bayes*, dan *random forest.*

Beberapa penelitian terdahulu mengenai *machine learning* sudah banyak dilakukan. Fadillah & Puspita (2021), Iman & Wijayanto (2021), dan Sihombing & Yuliati (2021) membahas perbandingkan antara beberapa metode imputasi berbasis *machine learning* dan menyatakan bahwa *random forest* adalah metode yang memberikan kinerja terbaik. Selain itu, Fadillah & Puspita (2021) melakukan identifikasi menggunakan metode imputasi berbasis statisitik yaitu *Hot-deck Imputation* pada data Industri Mikro dan Kecil, dan memberikan hasil identifikasi dengan akurasi yang cukup baik. Namun, pada penelitian Fadillah et al. (2022) menjelaskan bahwa akurasi yang dihasilkan oleh metode *random forest* secara konsisten lebih baik dibandingkan dengan metode *Hot-deck Imputation*.

Beberapa penelitian terdahulu sudah membahas mengenai perbandingan dari beberapa metode imputasi berbasis *machine learning* dan menyatakan metode *random forest* adalah metode yang terbaik. Penelitian lain juga sudah meneliti terkait identifikasi, namun menggunakan metode imputasi berbasis statistik. Oleh karena itu penelitian ini bertujuan untuk mengidentifikasi industri makanan dan minuman pada skala mikro dan kecil yang mengalami kesulitan bahan baku menggunakan metode metode imputasi berbasis *machine learning* yaitu *random forest*. Secara khusus, tujuan dari penelitian ini adalah untuk membandingkan akurasi dan waktu komputasi yang dihasilkan oleh metode *random forest* dalam mengidentifikasi industri makanan dan minuman pada skala mikro dan kecil yang mengalami kesulitan bahan baku.

2. Metodologi

2.1 Bahan dan Data

Penelitian ini menggunakan data yang berasal dari Survei Industri Mikro dan Kecil Tahunan Tahun 2020 Badan Pusat Statistik. Data yang digunakan berfokus pada usaha/perusahaan di sektor Industri Makanan dan Minuman skala mikro dan kecil. Variabel yang digunakan meliputi Industri Makanan dan Minuman yang mengalami kesulitan bahan baku, pendapatan usaha/perusahaan, Klasifikasi Baku Lapangan Usaha Indonesia 5-digit (KBLI 5-digit), rata-rata hari kerja dalam sebulan, klasifikasi industri, dan industri yang terkena dampak Covid-19. Variabel yang digunakan berdasarkan desain sampling yang digunakan pada Survei Industri Mikro dan Kecil Tahunan Tahun 2020 yaitu terkait KBLI dan juga klasifikasi industri mikro dan kecil, kemudian variabel yang lain berdasarkan penelitian terdahulu oleh Fadillah et al. (2022) yang mengkaji terkait Survei Industri Mikro dan Kecil Tahunan Tahun 2020. Variabel Industri Makanan dan Minuman yang mengalami kesulitan bahan baku, akan digunakan sebagai variabel klasifikasi, sedangkan variabel lainnya, yaitu pendapatan usaha/perusahaan, Klasifikasi Baku Lapangan Usaha Indonesia 5-digit (KBLI 5-digit), rata-rata hari kerja dalam sebulan, klasifikasi industri, dan industri yang terkena dampak Covid-19, akan digunakan sebagai variabel prediktor.

2.2 Metode Penelitian

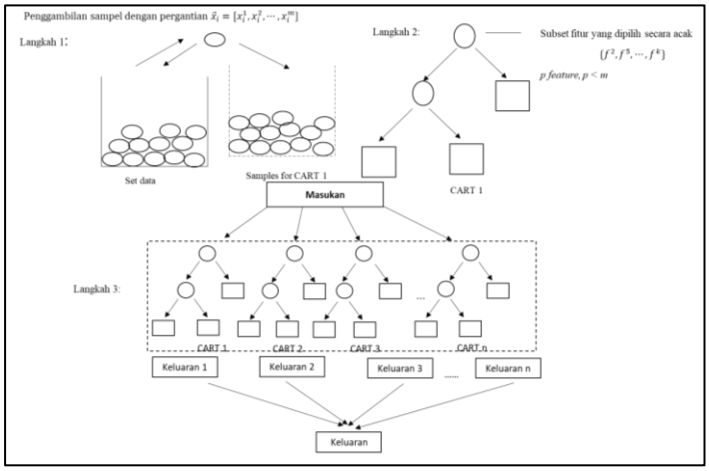
Missing Values

Ada banyak hal yang menyebabkan terjadinya *missing values*, diantaranya penolakan dari responden untuk menjawab beberapa pertanyaan karena pertanyaannya sangat pribadi/rahasia, kesalahan pada saat pengumpulan data misalnya pertanyaan terlewat sehingga tidak memperoleh jawaban, selain *itu missing values* juga diakibatkan kesalahan pada entri data (Fadillah & Muchlisoh, 2019). Berdasarkan penelitian Irawan et al. (2017), *missing values* disebabkan karena informasi tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Hal ini akan menyebabkan menurunnya keakuratan dan kualitas data yang diolah.

Menurut Fadillah & Muchlisoh (2019), *missing values* yang terjadi dikarenakan penghapusan item pertanyaan secara keseluruhan pada unit observasi yang membuat hilangnya informasi yang sudah dikumpulkan dan membuat pendugaan parameter menjadi tidak efisien.

Random Forest

Random forest adalah salah satu metode berbasis klasifikasi dan regresi dimana terdapat proses agregasi pohon Keputusan. Primajaya & Sari (2018) serta Breiman (2001) menyatakan pada algoritma *random forest* terdapat k pohon dengan vektor random yang independen dengan vektor-vektor random sebelumnya, tetapi memiliki distribusi yang identik. Metode ini memanfaatkan algoritma *decision tree* dalam melakukan klasifikasi, sehingga terbentuklah metode *boostrap aggregating* ketika membentuk sebuah sampel *training set*, dan setiap *tree* yang dibentuk menggunakan metode sama dalam membangun CART (*Classification and Regression Tree*). CART merupakan metode eksplorasi data yang didasarkan pada teknik *decision tree*. *Decision tree* dihasilkan saat peubah respons berupa data kategorik, sedangkan *regression tree* dihasilkan saat peubah respons berupa data numerik. Berikut adalah proses algoritma *random forest.*



Sumber: (Lin et al., 2017)

Gambar 1: Proses Algoritma *Random Forest*

Penggunaan metode *random forest* memerlukan variabel prediktor untuk membentuk *decision tree* pada modelnya. Sebelum melakukan klasifikasi, ditentukan variabel *important* melalui perhitungan skor *important* pada variabel-variabel prediktor. Pentingnya suatu variabel prediktor dilihat dari nilai skor *important* (Schratz et al., 2022). Skor *important* adalah ukuran yang menggambarkan seberapa pentingnya setiap fitur dalam model *random forest* untuk melakukan prediksi yang akurat. Semakin tinggi skor *important,* maka semakin penting variabel prediktor tersebut. Sebaliknya, semakin rendah skor *important* maka semakin kurang penting variabel prediktor tersebut untuk melakukan proses identifikasi.

Analisis Data dan Evaluasi Model

Analisis data pada penelitian ini memanfaatkan Software R, dimana metode *random forest* menggunakan *package randomForest* pada R-Studio untuk melakukan proses identifikasi. Pada penelitian ini, analisis akan dibagi menjadi dua bagian (simulasi I dan simulasi II). Simulasi I menggunakan keseluruhan data sebagai variabel prediktor, sedangkan simulasi II akan menggunakan sebagian variabel prediktor yang memiliki nilai skor *important* tinggi (variabel *important*). Analisis dibedakan menjadi dua simulasi dengan tujuan untuk membandingkan imputasi menggunakan keseluruhan variabel pada dataset dan hanya menggunakan sebagian variabel saja. Hal ini dikarenakan pada beberapa kondisi, tidak semua variabel dapat digunakan untuk klasifikasi. Analisis akan dilakukan dengan membagi terlebih dahulu *data training* dan *data testing* masing-masing sebesar 75 persen dan 25 persen. Selanjutnya akan dilakukan evaluasi model.

Pada penelitian ini, evaluasi model dari hasil identifikasi menggunakan metode *random forest* akan diukur menggunakan tabel *confusion matrix*. Menurut Arista (2021), *confusion matrix* adalah suatu tabel yang digunakan untuk menggambarkan performa pengklasifikasian. Tabel ini memberikan informasi mengenai kelas yang diprediksi dan kelas yang sebenarnya, serta jumlah observasi yang diklasifikasikan dengan benar dan tidak benar. Ada banyak ukuran yang dapat dihasilkan menggunakan *confusion matrix*. Salah satu ukuran yang umum digunakan adalah *accuracy*. Adapun bentuk tabel *confusion matrix* adalah sebagai berikut:

Tabel 1: *Confusion matrix*

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

Secara matematik, ukuran *accuracy* dirumuskan sebagai berikut:

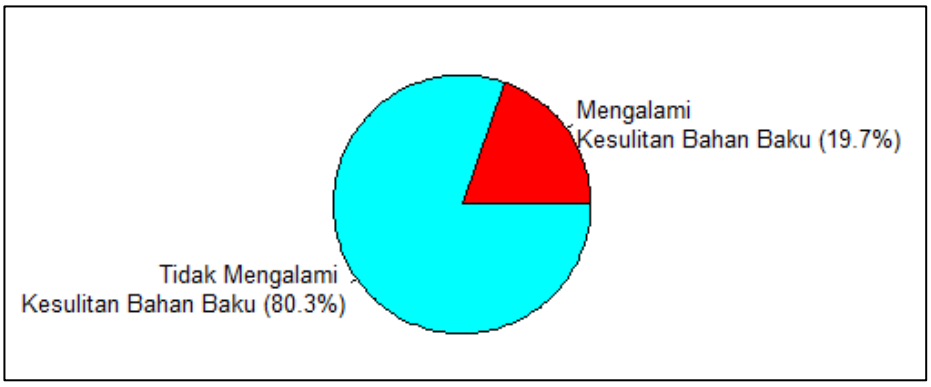
$$accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

Nilai akurasi membandingkan antara jumlah prediksi yang benar dibandingkan dengan total data yang diprediksi. Han et al. (2012) menjelaskan bahwa setiap pengguna data mungkin memiliki penilaian yang berbeda terhadap kualitas akurasi, salah satunya dijelaskan bahwa ketika akurasi mencapai 80 persen, maka database

tersebut dapat dikatakan akurat. Selain itu, akan diukur juga waktu komputasi yang diperlukan dalam proses identifikasi industri makanan dan minuman yang mengalami kesulitan bahan baku.

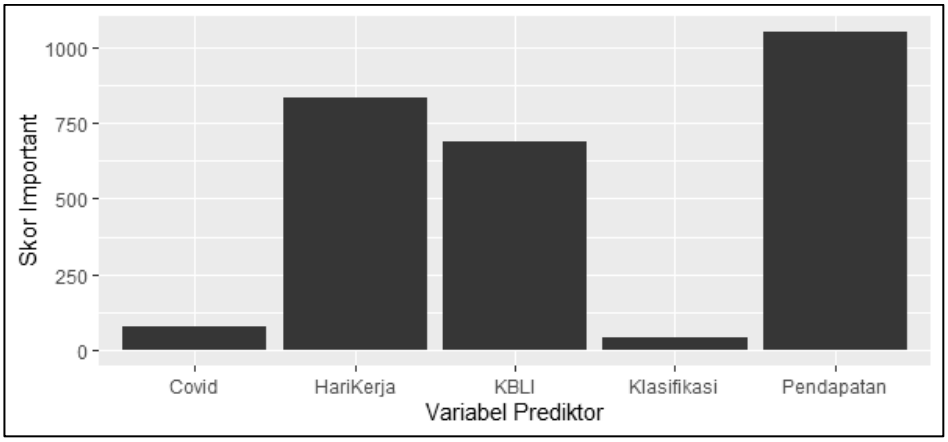
3. Hasil dan Pembahasan

Analisis data pada penelitian ini menggunakan data industri makanan dan minuman yang mengalami kesulitan bahan baku. Jumlah data yang digunakan sebanyak 26.058 *record*. Gambar 2 menjelaskan sebaran data untuk usaha-usaha yang mengalami kesulitan bahan baku dan juga usaha yang tidak mengalami kesulitan bahan baku. Sebanyak 19,67 persen usaha adalah usaha yang mengalami kesulitan bahan baku, dan sisanya sebanyak 80,33 persen adalah usaha yang tidak mengalami kesulitan bahan baku. Sebagaimana digambarkan pada Gambar 2, berdasarkan hasil tersebut dengan jumlah data yang besar, maka variabel klasifikasi dianggap tidak memilki *imbalanced class.*



Gambar 2: Proporsi Kesulitan Bahan Baku

Kemudian dilakukan perhitungan skor *important* pada model *random forest* untuk menentukan variabel yang akan digunakan pada simulasi II. Berdasarkan Gambar 3 diperoleh skor *important* untuk masing-masing variabel adalah sebagai berikut:



Gambar 3: Grafik Skor *Important* Variabel Prediktor

Berdasarkan hasil skor *important* di atas, dapat dilihat bahwa dari kelima variabel prediktor yang digunakan, tiga variabel prediktor memiliki skor *important* yang tinggi sedangkan dua variabel lainnya memiliki skor *important* yang rendah. Berdasarkan hasil skor *important* tersebut, ketiga variabel tersebut akan digunakan sebagai variabel *important* pada simulasi ke II. Ketiga variabel tersebut antara lain, variabel rata-rata hari kerja dalam sebulan, KBLI 5-digit, dan pendapatan usaha/perusahaan.

Berdasarkan hasil perhitungan yang dilakukan, adapun *confusion matrix* untuk masing-masing analisis adalah sebagai berikut:

Tabel 2: *Confusion matrix* simulasi I

		Kelas Sebenarnya	
		Positive	*Negative*
Kelas Prediksi	*Positive*	*53*	51
	Negative	*1228*	5182

Tabel 3: *Confusion matrix* simulasi II

		Kelas Sebenarnya	
		Positive	*Negative*
Kelas Prediksi	*Positive*	80	148
	Negative	1201	5085

Berdasarkan *confusion matrix* pada Tabel 2 dan Tabel 3, diperoleh hasil perhitungan *accuracy* untuk masing-masing simulasi. Selain itu, diperoleh juga waktu komputasi yang diperlukan oleh tiap-tiap simulasi dalam melakukan proses identifikasi data. Hasil perhitungan dari kedua simulasi secara lebih rinci dapat ditemukan dalam Tabel 4 di bawah ini.

Tabel 4: Perbandingan *accuracy* dan waktu komputasi

Parameter	Simulasi	
	I	II
Accuracy (%)	80,37%	79,29%
Waktu komputasi (detik)	281,62	260,75

Berdasarkan hasil Tabel 4 tersebut, diketahui nilai akurasi yang dihasilkan oleh metode *random forest* dalam mengidentifikasi industri makanan dan minuman yang mengalami kesulitan bahan baku akurat. Hal ini terlihat dari akurasi yang tinggi sebesar 80,37 persen pada simulasi I. Namun, ketika mengurangi sebagian variabel prediktor pada simulasi II, nilai akurasi yang dihasilkan menjadi kurang akurat. Dari hasil tersebut juga, terlihat bahwa akurasi dari simulasi I lebih tinggi dibandingkan dengan simulasi II dimana terdapat selisih sekitar 1,08 persen. Hal ini menandakan pada kasus ini, menggunakan seluruh variabel prediktor memberikan hasil yang lebih baik dibandingkan dengan menggunakan hanya sebagian variabel prediktor.

Namun, jika melihat dari sisi waktu komputasi, kedua metode terbilang cukup lama dalam membentuk model imputasi. Hal ini ditandai dengan waktu komputasi yang lebih dari empat menit. Walaupun demikian, proses imputasi pada kedua simulasi terbilang cepat karena hanya pada tahap pembentukan model saja yang

memakan waktu cukup lama. Jika dibandingkan antara simulasi I dan simulasi II terlihat ada selisih sebesar 20,85 detik (7,41 persen). Jika menimbang antara nilai akurasi yang hanya selisih 1,08 persen, perbedaan waktu komputasi sebesar 7,41 persen dapat menjadi pertimbangan ketika hendak melakukan identifikasi industri makanan dan minuman yang mengalami kesulitan bahan baku. Tentu saja dengan tetap mempertimbangan bahwa hasil identifikasi yang dihasilkan akurat. Secara keseluruhan terlihat penerapan metode random forest sebagai metode imputasi berbasis machine learning sejalan dengan sebagian besar penelitian terdahulu, yang memberikan hasil yang akurat. Dengan demikian, secara keseluruhan dalam melakukan identifikasi industri makanan dan minuman yang mengalami kesulitan bahan baku, metode random forest merupakan metode yang secara signifikan menghasilkan hasil yang akurat pada satu kondisi (simulasi I) dan bisa memberikan hasil yang kurang akurat pada kondisi lainnya (simulasi II).

4. Simpulan dan Saran

Berdasarkan hasil penjabaran dan analisis yang telah dilakukan sebelumnya, penerapan metode *random forest* sebagai metode imputasi berbasis *machine learning* untuk mengidentifikasi industri makanan dan minuman yang mengalami kesulitan bahan baku menunjukkan bahwa metode *random forest* menghasilkan hasil klasifikasi yang akurat pada saat keseluruhan informasi variabel prediktor dimanfaatkan. Hal ini ditandai dengan nilai akurasi yang dihasilkan oleh metode tersebut pada simulasi I. Selain itu, pemilihan variabel prediktor yang efektif, akan meningkatkan tingkat akurasi dari metode ini. Walaupun demikian, waktu komputasi yang relatif cukup lama, bisa menjadi kekurangan pada metode ini dalam melakukan klasifikasi. Hal ini ditandai dengan waktu komputasi pada kedua simulasi yang telah dilakukan.

Metode *random forest* memberikan hasil yang secara tepat dan konsisten dalam memprediksi industri makanan dan minuman yang mengalami kesulitan bahan baku. Hal ini ditandai dengan nilai *accuracy* yang tinggi pada simulasi I. Hal ini menunjukkan bahwa penggunaan metode ini dapat menjadi salah satu solusi untuk mengidentifikasikan industri makanan dan minuman yang mengalami kesulitan bahan baku menggunakan metode imputasi berbasis *machine learning*. Namun, jika proses pengidentifikasian hanya memiliki sedikit waktu dan memiliki jumlah data yang banyak, penggunaan metode ini juga perlu untuk diperhitungkan kembali dan mungkin dapat menggunakan alternatif metode imputasi berbasis statistik maupun metode imputasi berbasis *machine learning* lainnya. Sebaliknya jika tidak mempermasalahkan waktu, maka metode dapat menjadi solusi yang sangat baik.

Daftar Pustaka

Arista, N. (2021). *Pendeteksian Kecenderungan Depresi pada Pengguna Twitter Menggunakan Support Vector Machine (SVM)*. Politeknik Statistika STIS.

Badan Pusat Statistik. (2022). *Berita Resmi Statistik Pertumbuhan Ekonomi Indonesia Triwulan IV 2021*.

Badan Pusat Statistik. (2023a). Berita Resmi Statistik Pertumbuhan Ekonomi Indonesia Triwulan IV 2022. *Www.Bps.Go.Id*. Retrieved from https://www.bps.go.id/pressrelease/2020/02/05/1755/ekonomi-indonesia-2019-tumbuh-5-02-persen.html

Badan Pusat Statistik. (2023b). Profil Industri Mikro dan Kecil 2022. *Badan Pusat Statistik*, *13*: 1–239.

Biemer, P. P., & Lyberg, L. E. (2003). Introduction to Survey Quality. In *Analytical Biochemistry*. Canada: John Wiley & Sons.

Breiman, L. (2001). Random Forests. *Machine Learning*, (45): 5–32. https://doi.org/10.1109/ICCECE51280.2021.9342376

Fadillah, I. J., Fadila, L. M. A., & Darundiye, L. muhamad W. (2022). 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. *Seminar Nasional Official Statistics*, *2022*(1): 147–154. https://doi.org/10.34123/semnasoffstat.v2022i1.1235

Fadillah, I. J., & Muchlisoh, S. (2019). 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. (2021). 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*, *2021*(1): 734–741. https://doi.org/10.34123/icdsos.v2021i1.70

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining. Concepts and Techniques* (3rd ed.). https://doi.org/10.3726/978-3-653-01927-8/2

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

Irawan, N. D., Wijono, W., & Setyawati, O. (2017). Perbaikan Missing value Menggunakan Pendekatan Korelasi Pada Metode K-Nearest Neighbor. *Jurnal Infotel*, *9*(3). https://doi.org/10.20895/infotel.v9i3.286

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*(2): 105–115. https://doi.org/10.1016/j.artmed.2010.05.002

Kaiser, J. (2014). Dealing with Missing Values in Data. *Journal of Systems Integration*, 42–51. https://doi.org/10.20470/jsi.v5i1.178

Kemenperin. (2015). Rencana Induk Pembangunan Industri Nasional 2015 - 2035. *Rencana Induk Pembangunan Industri Nasional 2015-2035*, 1–98.

Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, *1*(1): 27. https://doi.org/10.24014/ijaidm.v1i1.4903

Schratz, P., Lang, M., Bischl, B., Binder, M., & Zobolas, J. (2022). *Package 'mlr3filters.'* Retrieved from github.com/mlr-org/mlr3filters/issues

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