

PENERAPAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE) PADA KASUS DAMPAK COVID-19 TERHADAP PENDUDUK USIA KERJA DI KALIMANTAN TIMUR

(SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE APPROACH IN CASE OF THE IMPACT OF COVID-19 ON POPULATION WORKING AGE IN EAST KALIMANTAN)

Ika Ayuningtyas, Ely Uswatun Kasanah

BPS Provinsi Kalimantan Timur^{1, 2}

E-mail: ika.ayu@bps.go.id

ABSTRAK

Pandemi Covid-19 memberikan dampak terhadap semua aspek kehidupan termasuk kondisi ketenagakerjaan di Indonesia, termasuk di Kalimantan Timur. Penelitian ini bertujuan untuk melihat faktor apa saja yang menjadi penyebab terjadinya perubahan status ketenagakerjaan yang terjadi akibat pandemi Covid-19 di Kalimantan Timur. Metode yang digunakan yakni analisis regresi logistik dengan menerapkan *Synthetic Minority Oversampling Technique*. Penggunaan metode ini akibat data yang tidak seimbang pada data jumlah penduduk usia kerja yang terdampak Covid-19. Berdasarkan penelitian yang telah dilakukan, SMOTE terbukti dapat meningkatkan akurasi model yang digunakan, sehingga model regresi logistik mampu memprediksi kejadian penduduk usia kerja yang terkena dampak Covid-19 berdasarkan variabel bebas yang digunakan. Variabel yang secara signifikan memengaruhi penduduk usia kerja Kalimantan Timur terkena dampak Covid-19, yaitu umur, pendidikan, kategori lapangan usaha, status pekerjaan kategori pegawai/buruh dan klasifikasi daerah tempat tinggal. Klasifikasi daerah tempat tinggal merupakan variabel yang rasio kecenderungannya paling besar, dimana penduduk usia kerja di wilayah perkotaan memiliki peluang terdampak Covid-19 lebih besar 3,031 kali dibandingkan penduduk usia kerja di wilayah perdesaan.

Kata kunci: dampak covid-19 pada ketenagakerjaan, karakteristik terdampak covid-19, data tidak seimbang, *overfitting*

ABSTRACT

The Covid-19 pandemic has an impact on all aspects of life, including labor conditions in Indonesia, including in East Kalimantan. This study aims to see what factors are the causes of changes in employment status that occur due to the Covid-19 pandemic in East Kalimantan. The method used is logistic regression analysis by applying the Synthetic Minority Oversampling Technique. This method is due to unbalanced data on the number of working age population affected by Covid-19. Based on the research, SMOTE has been proven to increase the accuracy of the model used. So that the logistic regression model can predict the incidence of the working-age population affected by Covid-19 based on the independent variables used. Variables that significantly affect the working-age population of East Kalimantan affected by Covid-19 are age, education, business field category, employment status in the employee category, and the residential area classification. The residential area classification is the variable with the greatest trend ratio. The working-age population in urban areas has a 3.031 times greater chance of being affected by Covid-19 than the working-age population in rural areas.

Keywords: *the impact of covid-19 on employment, characteristics affected by covid-19, imbalanced data, overfitting*

PENDAHULUAN

Pandemi Covid-19 memberikan dampak terhadap semua aspek kehidupan, baik pada masalah kesehatan maupun masalah ekonomi. Pandemi ini memaksa pemerintah untuk melakukan pembatasan aktivitas masyarakat. Pembatasan aktivitas ini turut berdampak pada dinamika ketenagakerjaan di Indonesia. Beberapa penelitian menyatakan bahwa pandemi Covid-19 ini memengaruhi kondisi tenaga kerja di Indonesia, seperti hilangnya mata pencaharian sebagian masyarakat, terjadinya PHK serta meningkatnya angka pengangguran (Fahri *et al.*, 2020; Indayani

& Hartono, 2020). Dampak pandemi Covid-19 juga dirasakan penduduk usia kerja di Kalimantan Timur. Badan Pusat Statistik (BPS) Kalimantan Timur mencatat sebanyak 14,82 persen dari penduduk usia kerja di Kalimantan Timur merasakan dampak pandemi Covid-19 (BPS Provinsi Kalimantan Timur, 2020). Dampak yang dirasakan oleh penduduk usia kerja seperti menjadi pengangguran, berhenti bekerja, sementara tidak bekerja atau dirumahkan, atau mengalami pengurangan jam kerja. Perubahan status ketenagakerjaan akibat adanya pandemi Covid-19 ini menyebabkan pemerintah harus mengatur strategi kebijakan agar permasalahan ini tidak menyebabkan dampak lain yang lebih luas, seperti kemiskinan.

Penelitian ini bertujuan untuk melihat bagaimana karakteristik penduduk usia kerja di Kalimantan Timur yang mengalami perubahan status ketenagakerjaan akibat Covid-19. Metode yang akan digunakan yakni regresi logistik. Regresi logistik merupakan salah satu alat analisis statistik yang dapat melihat hubungan antara variabel terikat berupa data kategorik dengan variabel bebas (Hosmer & Lemeshow, 2000). Akan tetapi penggunaan metode ini kurang tepat digunakan pada kondisi data yang tidak seimbang (*imbalanced data*), karena model klasifikasi ini akan cenderung menihilkan peluang dari kelompok minoritas sehingga klasifikasi dari hasil prediksi akan cenderung memasukkan kategori mayoritas (King & Zeng, 2003). Salah satu alternatif untuk meningkatkan akurasi model pada regresi logistik adalah melakukan *Synthetic Minority Oversampling Technique* (SMOTE). Dengan penggunaan teknik akurasi model regresi logistik yang digunakan akan semakin meningkat, sehingga dapat memprediksi respon dengan lebih tepat (Hanifah *et al.*, 2015; Hazwani *et al.*, 2019).

Berdasarkan latar belakang di atas, metode regresi logistik dengan menerapkan SMOTE akan digunakan untuk melihat faktor apa saja yang menyebabkan terjadinya perubahan status ketenagakerjaan yang diakibatkan pandemi Covid-19 di Kalimantan Timur. Metode ini digunakan karena kondisi data jumlah penduduk usia kerja yang mengalami dampak Covid-19 di Kalimantan Timur memiliki proporsi yang tidak seimbang.

METODE

Regresi Logistik

Regresi logistik adalah model regresi yang menunjukkan pengaruh variabel bebas, baik berupa data kontinyu maupun kategorik, terhadap variabel respon berupa data kategorik (Hosmer & Lemeshow, 2000). Model regresi logistik dapat dituliskan sebagai berikut:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \dots\dots\dots(1)$$

Karena model di atas tidak linier pada parameter maka dilakukan transformasi logit untuk mempermudah proses estimasi parameter, yakni

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \dots\dots\dots(2)$$

Metode yang digunakan adalah metode *Maximum Likelihood Estimation* (MLE).

Pengujian parameter secara serentak digunakan untuk mengetahui apakah variabel bebas dalam model secara serentak berpengaruh secara signifikan terhadap variabel respon. Hipotesis yang digunakan dalam uji ini, yakni

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

$$H_1: \text{minimal ada dari } \beta_j \neq 0, \text{ dimana } j = 1, 2, \dots, p$$

dengan statistik uji:

$$G = -2 \log\left(\frac{l_0}{l_1}\right) = -2[\log(l_0) - \log(l_1)] \dots\dots\dots(3)$$

dimana l_0 merupakan nilai maksimum fungsi probabilitas di bawah H_0 , sedangkan l_1 merupakan nilai maksimum fungsi probabilitas di bawah H_1 . Statistik uji G mengikuti distribusi *chi-square* dengan derajat bebas sebanyak p variabel bebas. Keputusan tolak H_0 jika nilai statistik uji $G > \chi_p$ atau minimal ada satu variabel bebas yang berpengaruh signifikan terhadap variabel respon.

Pengujian parameter secara parsial untuk mengetahui apakah suatu variabel bebas berpengaruh signifikan terhadap variabel respon. Hipotesis yang digunakan dalam pengujian ini, yakni

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, \text{ dimana } j = 1, 2, \dots, p$$

dengan statistik uji Wald:

$$W = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \dots\dots\dots(4)$$

Keputusan tolak H_0 jika nilai statistik uji $W > Z_{\alpha/2}$ atau variabel bebas ke- j berpengaruh signifikan terhadap variabel respon.

Synthetic Minority Oversampling Technique (SMOTE)

Metode SMOTE merupakan salah satu solusi dalam mengatasi data yang tidak seimbang (Chawla *et al.*, 2002). Metode ini dilakukan dengan cara menambah jumlah kelas minoritas agar setara dengan kelas mayoritas dengan cara membangkitkan data buatan. Data buatan atau sintesis dibuat berdasarkan *k -nearest neighbor*. Jumlah *k -nearest neighbor* ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya (Barro *et al.*, 2013). Pembangkitan data buatan yang berskala numerik berbeda dengan kategorik. Data numerik diukur jarak kedekatannya dengan jarak *Euclidean* sedangkan data kategorik lebih sederhana yaitu dengan nilai modus.

Pengukuran Performa Model Klasifikasi

Pengukuran performa model klasifikasi dengan dua kategori dapat dilihat dari *confusion matrix* di bawah ini:

Tabel 1. Kesesuaian Klasifikasi

| Aktual | Prediksi Model | |
|--------|-------------------------------|-------------------------------|
| | 0 | 1 |
| 0 | Benar <i>(Specificity)</i> | Salah |
| 1 | Salah | Benar <i>(Sensitivity)</i> |

Indikator *sensitivity*; persentase observasi yang diklasifikasikan 1 dari total yang sebenarnya 1 dan indikator *specificity*; persentase observasi yang diklasifikasikan 0 dari total yang sebenarnya 0.

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data mikro yang berasal dari Survei Angkatan Kerja Nasional (Sakernas) Provinsi Kalimantan Timur. Sampel yang digunakan sebanyak 75 persen sampel panel pada bulan Agustus 2019 dan Agustus 2020. Penduduk usia kerja terdampak Covid-19 dibatasi hanya pada pekerja usia kerja yang mengalami pengangguran karena Covid-19, Bukan Angkatan Kerja karena Covid-19, dan sementara tidak bekerja karena Covid-19. Jumlah sampel sebanyak 7.933 unit sampel. Dimana sebanyak 7.783 sampel atau sebesar 98,1 persen sampel penduduk usia kerja tidak terkena dampak Covid-19, sedangkan hanya sebanyak 1,9 persen atau 150 sampel penduduk usia kerja yang terkena dampak Covid-19. Dengan demikian, data pekerja usia kerja yang terdampak Covid-19 merupakan data yang tidak seimbang (*imbalanced data*). Adapun variabel respon dan variabel bebas yang digunakan sebagai berikut:

Tabel 2. Variabel yang digunakan

| Variabel | Nama | Keterangan |
|----------|--|---|
| y | Dampak Covid-19 terhadap Penduduk Usia Kerja | 0: Tidak terkena Dampak Covid-19 1: Terkena Dampak Covid-19 |
| Umur | Usia Penduduk Usia Kerja | Satuan (Tahun) |
| JK | Jenis Kelamin | 0: Laki-Laki 1: Perempuan |
| Kateg | Kategori Lapangan Usaha | 0: Pertanian 1: Pertambangan 2: Industri Pengolahan 3: Perdagangan 4: Lainnya |
| Stat | Status Pekerjaan | 0: Berusaha (sendiri/dibantu) 1: Buruh/pegawai 2: Pekerja bebas/tidak dibayar |
| Pend | Pendidikan yang ditamatkan | 0: SMP ke bawah 1: SMA ke atas |
| Klas | Klasifikasi Wilayah | 0: Desa 1: Kota |

HASIL DAN PEMBAHASAN

Perbandingan antara jumlah penduduk usia kerja yang terdampak Covid-19 dan yang tidak, berdasarkan Sakernas Agustus 2020 memiliki perbedaan yang cukup besar. Data seperti ini apabila dilakukan pemodelan menggunakan analisis regresi logistik tentu akan memberikan hasil estimasi yang kurang akurat. Oleh karena itu perlu dilakukan *oversampling* agar data lebih proporsional, sehingga model estimasi yang diperoleh akan lebih tepat.

Tabel 3. Perbandingan Jumlah Data Penduduk Usia Kerja Sebelum dan Setelah menggunakan SMOTE

| Penduduk Usia Kerja | Data | |
|-------------------------------|---------------|-------------------|
| | Sebelum SMOTE | Menggunakan SMOTE |
| Tidak terkena dampak Covid-19 | 7.783 | 7.783 |
| Terkena dampak Covid-19 | 150 | 7.650 |

Sumber: Sakernas Agustus 2019 dan 2020 (data diolah).

Dari tabel 3 terlihat data sebelum dan setelah menggunakan SMOTE memiliki perbedaan. Data sebelum menggunakan SMOTE tidak seimbang, dimana penduduk usia kerja yang tidak terkena dampak Covid-19 jauh lebih banyak dibandingkan dengan yang terkena dampak Covid-19. Namun, setelah dilakukan metode SMOTE, perbandingan data kejadian penduduk usia kerja yang mengalami dampak Covid-19 menjadi lebih seimbang.

Berdasarkan tabel 4, dapat diketahui bahwa pada model regresi logistik pada data yang tidak seimbang terjadi *overfitting*. Salah satu indikasi terjadi *overfitting* adalah hasil dari tabel prediksi dan klasifikasi pada model regresi logistik tidak mampu memprediksi kejadian penduduk usia kerja yang terkena dampak Covid-19 berdasarkan variabel bebas yang ada.

Tabel 4. Tabel Klasifikasi dan Prediksi (tanpa SMOTE)

| | Penduduk Usia Kerja | Prediksi | | Persentase |
|-------------------|-------------------------------|-------------------------------|-------------------------|------------|
| | | Tidak terkena dampak Covid-19 | Terkena dampak Covid-19 | |
| Observasi | Tidak terkena dampak Covid-19 | 7 783 | 0 | 100,0 |
| | Terkena dampak Covid-19 | 150 | 0 | 0,0 |
| Persentase | | | | 98,1 |

The cut value is .500

Sumber: Sakernas Agustus 2019 dan Agustus 2020 (diolah).

Sedangkan pada tabel 5, diketahui dengan menggunakan data hasil metode SMOTE, model regresi logistik mampu memprediksi kejadian penduduk usia kerja yang terkena dampak Covid-19 berdasarkan variabel bebas yang digunakan. Persentase hasil prediksinya sebesar 68,9 persen. Dikarenakan penggunaan SMOTE dapat mengatasi terjadinya *overfitting* pada data yang tidak seimbang, maka model regresi logistik dengan SMOTE lebih sesuai digunakan untuk memodelkan kejadian dampak Covid-19 terhadap penduduk usia kerja di Kalimantan Timur.

Tabel 5. Tabel Klasifikasi dan Prediksi (Menggunakan SMOTE)

| | Penduduk Usia Kerja | Prediksi | | Persentase |
|-------------------|-------------------------------|-------------------------------|-------------------------|------------|
| | | Tidak terkena dampak Covid-19 | Terkena dampak Covid-19 | |
| Observasi | Tidak terkena dampak Covid-19 | 4 631 | 3 152 | 59,5 |
| | Terkena dampak Covid-19 | 1 643 | 6 007 | 78,5 |
| Persentase | | | | 68,9 |

The cut value is .500

Sumber: Sakernas Agustus 2019 dan Agustus 2020 (diolah).

Pemodelan dengan Regresi Logistik

Pemodelan untuk penelitian dampak covid-19 terhadap penduduk usia kerja Provinsi Kalimantan Timur dilakukan menggunakan analisis regresi logistik berganda dengan data yang sudah dilakukan SMOTE. Hasil pengujian secara simultan dengan menggunakan data hasil SMOTE sebagai berikut:

Tabel 6. Uji Simultan (Omnibus Tests of Model Coefficients)

| | Chi-square | Df | p-value |
|--------------|-------------------|-----------|----------------|
| Model | 2835.339 | 10 | .000 |

Catatan: *) signifikan pada taraf signifikansi 5%

Sumber: Sakernas Agustus 2019 dan Agustus 2020 (diolah).

Berdasarkan hasil uji simultan di atas diperoleh *p-value* 0,000. Nilai tersebut lebih kecil dari lima persen, dapat disimpulkan bahwa minimal terdapat satu variabel bebas yang berpengaruh signifikan secara statistik terhadap penduduk usia kerja yang terdampak Covid-19 atau dapat dikatakan model dinyatakan fit. Selanjutnya uji parameter secara parsial untuk masing-masing variabel, sehingga diperoleh hasil sebagai berikut:

Tabel 7. Uji Parsial Determinan yang Memengaruhi Penduduk Usia Kerja terdampak Covid-19 di Provinsi Kalimantan Timur

| Variabel | Nama Variabel | Koefisien | Standard Error | Sig. | Odds Ratio |
|---|---------------|-----------|----------------|-------|------------|
| Umur | Umur | 0.009 | 0.004 | 0.000 | 1.009 |
| Pendidikan yang Ditamatkan | Pend | -0.642 | -0.615 | 0.000 | 0.526 |
| Jenis Kelamin | JK | 0.215 | 0.337 | 0.000 | 1.240 |
| Kategori Lapangan Usaha: Pertambangan | Kateg1 | -1.850 | -1.295 | 0.000 | 0.157 |
| Kategori Lapangan Usaha: Industri Pengolahan | Kateg2 | -0.150 | 0.061 | 0.033 | 0.860 |
| Kategori Lapangan Usaha: Perdagangan | Kateg3 | -0.447 | -0.107 | 0.000 | 0.639 |
| Kategori Lapangan Usaha: Lainnya | Kateg4 | -0.605 | -1.051 | 0.000 | 0.546 |
| Status Pekerjaan: Buruh/Pegawai | Stat1 | 0.497 | 0.125 | 0.000 | 1.643 |
| Status Pekerjaan: Pekerja Bebas/tidak dibayar | Stat2 | -0.001 | -0.621 | 0.987 | 0.999 |
| Klasifikasi Daerah Tempat Tinggal | Klas | 1.109 | 0.935 | 0.000 | 3.031 |
| Konstan | | -0.793 | -4.010 | 0.000 | 0.453 |

Catatan: *) signifikan pada taraf signifikansi 5%
 Sumber: Sakernas Agustus 2019 dan Agustus 2020 (diolah).

Berdasarkan hasil pengujian parsial pada masing-masing variabel, diperoleh salah satu kategori dari status pekerjaan tidak signifikan karena nilai *p-value*nya di atas 5 persen, yaitu kategori pekerja bebas/tidak dibayar. Dengan demikian dapat disimpulkan, pada tingkat kepercayaan 5 persen, secara statistik terdapat cukup bukti untuk menyatakan bahwa umur, pendidikan, kategori lapangan usaha, status pekerjaan kategori pegawai/buruh dan klasifikasi daerah tempat tinggal berpengaruh terhadap peluang penduduk usia kerja terkena dampak Covid-19.

Dari Tabel 7, variabel yang tidak signifikan secara statistik kemudian dikeluarkan dari model, sehingga model yang terbentuk untuk kejadian penduduk usia kerja terdampak covid-19 adalah sebagai berikut:

$$g(x) = -0,793 + 0,009umur_i - 0,642pend_i + 0,215JK_i - 1,850kateg1_i - 0,150kateg2_i - 0,447kateg3_i - 0,605kateg4_i + 0,497Stat1_i + 1,109Klas_i$$

Interpretasi hasil estimasi model regresi logistik yakni dengan melihat rasio kecenderungan (*odds ratio*) pada masing-masing variabel. Nilai rasio kecenderungan untuk setiap variabel dinyatakan oleh $\exp(\hat{\beta})$. Interpretasi model logistik terhadap pekerja usia kerja terdampak Covid-19 dapat dituliskan sebagai berikut:

- Pertambahan umur akan memperbesar probabilitas seorang penduduk usia kerja terkena dampak covid-19. Setiap bertambahnya usia pekerja sebesar 1 tahun, maka kecenderungan penduduk usia kerja terkena dampak Covid-19 adalah sebesar 1,009 kali lebih besar.
- Peluang penduduk usia kerja yang berpendidikan SMA ke atas terdampak Covid-19 lebih kecil sebesar 0,526 kali dibandingkan yang berpendidikan SMP ke bawah, atau peluang penduduk usia kerja yang berpendidikan SMP ke bawah terdampak Covid-19 lebih besar 1,901 kali dibandingkan yang berpendidikan SMA ke atas.
- Sedangkan untuk penduduk usia kerja yang berjenis kelamin perempuan memiliki peluang terdampak Covid-19 yang lebih besar 1,240 kali dibandingkan penduduk laki-laki.
- Peluang penduduk usia kerja bekerja di pertanian terdampak Covid-19 lebih besar 6,369 kali dibandingkan yang bekerja di pertambangan. Sedangkan penduduk yang bekerja di sektor pertanian terdampak Covid-19 memiliki peluang lebih besar 1,163 kali dibandingkan yang bekerja di sektor industri pengolahan, lebih besar 1,565 kali dibandingkan pada sektor perdagangan, dan lebih besar 1,832 kali dibandingkan sektor lainnya. Interpretasi ini

memiliki kecenderungan yang cukup berbeda dibandingkan kajian lainnya, yang disebutkan bahwa sektor pertanian lebih tahan terhadap guncangan pandemi Covid-19 (Purba *et al.*, 2020; Robins *et al.*, 2020). Sehingga perlu dilakukan peninjauan kembali terkait variabel lapangan usaha pada penduduk usia kerja ini.

- e. Peluang penduduk yang bekerja sebagai buruh/karyawan terdampak Covid-19 lebih besar 1,643 kali jika dibandingkan penduduk yang berusaha, baik yang berusaha .
- f. Penduduk usia kerja di wilayah perkotaan memiliki peluang terdampak Covid lebih besar 3,031 kali dibandingkan penduduk usia kerja di wilayah perdesaan.

KESIMPULAN

Proporsi penduduk usia kerja yang terdampak Covid-19 di Kalimantan Timur tidak seimbang, karena jumlah kejadiannya sangat kecil. Penelitian ini menggunakan metode SMOTE untuk menangani ketidakseimbangan tersebut. Berdasarkan hasil penelitian di atas diketahui bahwa penggunaan metode SMOTE pada model regresi logistik dapat menghindari terjadinya *overfitting* pada data tidak seimbang dan meningkatkan tingkat akurasi pengklasifikasian untuk data minoritas.

Variabel umur, jenis kelamin, pendidikan yang ditamatkan, lapangan usaha, status pekerjaan kategori pegawai/buruh, serta klasifikasi wilayah tempat tinggal berpengaruh terhadap peluang penduduk usia kerja terdampak Covid-19, seperti menjadi pengangguran, menjadi bukan angkatan kerja (tidak bekerja lagi), atau dirumahkan (memiliki pekerjaan namun sementara tidak bekerja).

DAFTAR PUSTAKA

- Barro, R. A., Sulvianti, I. D., & Afendi, F. M. (2013). Penerapan Synthetic Minority Oversampling Technique (SMOTE) terhadap Data Tidak Seimbang pada Pembuatan Model Komposisi Jamu. *Xplore*, 1(1).
- BPS Provinsi Kalimantan Timur. (2020). Berita Resmi Statistik. In *Keadaan Ketenagakerjaan Kalimantan Timur Agustus 2020* (Vol. 19, Issue 27).
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16(Sept. 28), 321–357. <https://arxiv.org/pdf/1106.1813.pdf><http://www.snopes.com/horrors/insects/telamonia.asp>
- Fahri, Abd. Jalil, & Kasnelly, S. (2020). Meningkatnya Angka Pengangguran Di Tengah Pandemi (Covid-19). *Jurnal Ekonomi Syariah*, 2(pengangguran akibat covid 19), 45–60.
- Hanifah, F. S., Wijayanto, H., & Kurnia, A. (2015). SMOTE bagging algorithm for imbalanced dataset in logistic regression analysis (case: Credit of bank X). *Applied Mathematical Sciences*, 9(137–140), 6857–6865. <https://doi.org/10.12988/ams.2015.58562>
- Hazwani, A., Rahim, A., Rashid, N. A., Nayan, A., & Ahmad, A. (2019). SMOTE Approach to Imbalanced Dataset in Logistic Regression Analysis. In *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (ICMS2017)*. Springer Singapore. <https://doi.org/10.1007/978-981-13-7279-7>
- Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regressin. In *John Wiley & Sons, Ic* (2nd edition).
- Indayani, S., & Hartono, B. (2020). Analisis Pengangguran dan Pertumbuhan Ekonomi sebagai Akibat Pandemi Covid-19. *Jurnal Perspektif*, 18(2), 201–208.
- King, G., & Zeng, L. (2003). Logistic regression in rare events data. *Journal of Statistical Software*, 8, 137–163. <https://doi.org/10.18637/jss.v008.i02>
- Purba, H. J., Yusuf, E. S., & Erwidodo. (2020). Dampak Pandemi Covid-19 terhadap Pertumbuhan Ekonomi dan Sektor Pertanian. In *Dampak Pandemi Covid-19: Perspektif Adaptasi dan Resiliensi Sosial Ekonomi Pertanian* (pp. 23–46).
- Robins, L., Crimp, S., van Wensveen, M., Alders, R. ., Bourke, R. M., Butler, J., Cosijn, M., Davila, F., Lal, A., McCarthy, J. ., McWilliam, A., Palo, A. S. M., Thomson, N., Warr, P., & Webb, M. (2020). *COVID-19 and food systems in Indo-Pacific: An Assessment of vulnerabilities, impacts and opportunities for action* (Issue 96).