

Klasifikasi Multi Class Pada Metode Kerja Jarak Jauh Menggunakan Algoritma Decision Tree dan Imbalance Data

Jefina Tri Kumalasari¹, Agustiena Merdekawati²,
Apriliani Hidayati³

Program Studi Sistem Informasi^{1,2,3},
Fakultas Teknologi Informasi^{1,2,3},
Universitas Bina Sarana Informatika^{1,2,3}

e-mail:jefina.jtk@bsi.ac.id¹,
agustiena.atd@bsi.ac.id²,19211131@bsi.ac.id³

Received: November 20, 2023. **Revised:** December 27, 2024. **Accepted:** January 30, 2024. **Issue Period:** Vol.8 No.1 (2024), Pp.109-117

Abstrak – Pandemi Covid-19 baru saja berlalu di seluruh dunia walaupun ada beberapa negara mulai mengalami pandemic lagi. Pandemi ini telah merubah kebiasaan dalam keseharian masyarakat khususnya dalam bekerja. Dimana sebelumnya bekerja dari rumah sebagai upaya mencegah peyebaran virus. Namun hingga pandemi usai, berdasarkan survey yang telah dilakukan terhadap 100 pekerja didapati bahwa bekerja dari rumah (*Work From Home*) atau bekerja secara *remote* masih menjadi pilihan yang banyak diminati oleh para pekerja karena dianggap memberikan fleksibilitas dan menghemat lebih banyak waktu. Sebagian dari mereka yang lebih memilih bekerja dari kantor (*Work From Office*) karena dinilai lebih mudah menjaga fokus dan dapat meningkatkan produktivitas dan ada pula yang lebih minat mode kerja campuran (*a mixed-mode of working*). Sehingga dibutuhkan analisis dan perbandingan untuk mengetahui lokasi kerja mana yang lebih banyak diminati pekerja. Salah satu solusi untuk mengatasi permasalahan ini yaitu diperlukan metode klasifikasi multi class untuk memproses faktor yang mempengaruhi pilihan lokasi kerja tersebut. Metode klasifikasi yang digunakan sebagai pengolah data adalah metode *Decision Tree*. Metode pada permasalahan ketidakseimbangan kelas menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Pengujian yang dilakukan menggunakan *Decision Tree* dan SMOTE split data yang mendapatkan akurasi hingga 83,08% pada rasio 0,5 (5:5). Pada penelitian ini didapati hasil sebanyak 13% pekerja lebih memilih bekerja dari rumah, 25% memilih bekerja dari kantor dan 27% memilih mode kerja campuran.

Kata Kunci: *Decision Tree*, SMOTE, WFH dan WFO

Abstract – The COVID-19 pandemic that has hit the world has forced workers to work from home to prevent the spread of the virus. However, until the pandemic is over, based on survey conducted on 100 employee, found that working from home (*Work from Home*) is still a choice that many workers are interested in because it considered to provide flexibility and save more times. But some of them prefer to work from the office because it is considered easier than to focus and can increase



DOI: 10.52362/jisamar.v8i1.1350

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

productivity and more interested mixed mode of working. The analysis and comparison determined to find out about which work locations are more popular with workers. One solution to overcome this problem is that a classification method is needed to group the factors that influence the choice of work location. The classification method used for data processing is the Decision Tree method. The method for class imbalance problems uses the Synthetic Minority Over-sampling Technique (SMOTE) method. Tests were executed using Decision Tree and SMOTE split data which obtained an accuracy of up to 83.08% at a ratio of 0.5 (5:5). In this research, it was found that 13% of workers preferred to work from home, 25% chose to work from an office, and 27% chose mixed work models.

Keywords: Decision Tree, SMOTE, WFH and WFO

I. PENDAHULUAN

Pandemi COVID-19 telah menimbulkan tantangan yang belum pernah terjadi sebelumnya terhadap sistem kesehatan masyarakat, perekonomian global [1] Pandemi Covid 19 telah menyebabkan perubahan drastis pada organisasi sosial dan profesi di berbagai negara di dunia, termasuk dalam memenuhi kebutuhan yaitu metode dalam bekerja. Salah satu penyebabnya karena pemerintah berbagai negara telah menerapkan pembatasan sosial melalui pengendalian pada banyak aspek masyarakat, termasuk mobilitas. Kehidupan sosial dan cara kerja kita telah berubah, dan banyak negara menciptakan kebijakan yang mempertimbangkan lockdown (pemerintah memaksa orang untuk tinggal di rumah) atau penjarakan sosial (orang menjaga jarak tertentu satu sama lain) untuk terus berlanjut di masa depan [2]

Efek dari adanya pandemi ini juga dialami oleh pekerja, pergeseran metode kerja dari bekerja dari kantor menjadi bekerja dari rumah demi mencegah penyebaran COVID-19 [3] Namun setelah pandemi usai 2 tahun yang lalu, ditemukan bahwa metode bekerja dari rumah masih banyak diminati dikarenakan sebagian pekerja merasa lebih fleksibel dan menghemat banyak waktu. Namun sebagian lainnya lebih menyukai bekerja dari kantor karena dianggap membuat pekerja lebih fokus dan meningkatkan produktifitas. Selain karena merasa lebih nyaman percepatan kemajuan teknologi komunikasi merubah cara kerja saat ini dan di masa depan. Kemampuan *soft skill* dalam teknologi akan mendukung karyawan dalam berkarya baik secara *hybrid*, *remote* ataupun bekerja di kantor.

Data diperoleh berdasarkan faktor maupun alasan pekerja dimana ada yang lebih menyukai salah satu metode kerja di kantor, di rumah, dan gabungan dari keduanya. Mengetahui apa yang diperlukan atau yang diinginkan karyawan diharapkan akan meningkatkan produktifitas dan efisiensi. Hal ini bermanfaat diperlukan model atau metode yang mengolah data menggunakan metode *Decision Tree* guna mempermudah mengetahui yang diinginkan karyawan dalam bekerja.

Klasifikasi sebagai *supervised learning* memprediksi data masukan yang diberikan. Menggunakan label dimana model akan dilatih menggunakan *train* data kemudian guna mengevaluasi test data sebelum digunakan memprediksi data yang baru. Multiclass classification digunakan pada kasus ini, dimana memiliki lebih dari dua label yang berbeda. Klasifikasi mengklasifikasikan setiap item pada suatu dataset ke dalam kelas-kelas atau kelompok yang sudah didefinisikan sebelumnya [4]

Algoritma pohon keputusan (*decision tree*) sudah tidak asing lagi digunakan dalam penelitian dimana sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan gunakan masing-masing rangkaian pembagi, anggota himpunan hasil menjadi mirip satu dengan yang lainnya. Data dalam Pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record [5]



Pada penelitian terdahulu terkait *Decision Tree*, penggunaan algoritma *Decision Tree* pernah digunakan untuk judul penelitian “Rekomendasi Merk Mobil Untuk Calon Pembeli Menggunakan Algoritma *Decision Tree* [6]. Seringkali pembeli merasa kecewa ketika membeli mobil karena tidak seperti yang diharapkan. Walaupun di zaman sekarang sudah ada mesin pencari, namun banyaknya informasi mengharuskan calon pembeli untuk menyimpulkan secara mandiri apakah hal tersebut sesuai atau tidak. Metode yang diusulkan ialah *Decision Tree*. Dengan hasil akurasi tertinggi yaitu pada kedalaman tree sebanyak 7 dengan akurasi 85% dan rata-rata precision 87% dan dari berbagai batasan kedalaman tree yang digunakan, dihasilkan nilai recall, precision, f1 measure, dan accuracy di atas 70%.

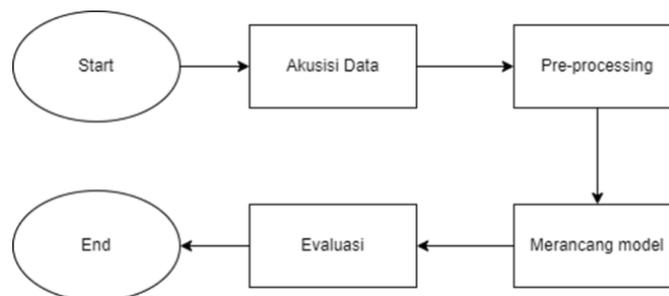
Synthetic Minority Over-sampling Technique (SMOTE) merupakan pendekatan untuk menyeimbangkan data sampel pada kelas yang memiliki ketidakseimbangan berlebihan (mayoritas) dengan fokus terhadap kelas minoritas, dengan tujuan meningkatkan kinerja dari metode klasifikasi. Pada SMOTE kemungkinan terjadi overfitting yaitu data pada kelas minoritas yang terduplikasi (Franseda et al., 2020).

Penelitian terkait lainnya berjudul “Integrasi Metode *Decision Tree* dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu lintas” [7]). Dalam penelitian ini digunakan SMOTE untuk meningkatkan evaluasi pada penelitian ini menggunakan proses *Knowledge Discovery in Database* (KDD). Pengujian dilakukan dengan tiga desain model yaitu *Split Validation Decision Tree* dan SMOTE diperoleh akurasi 69.23%. Pengujian menggunakan *Cross Validation Decision Tree* dan SMOTE diperoleh akurasi 63.56%. Pengujian menggunakan *Decision Tree* dan SMOTE Split Data diperoleh akurasi 71.12% dengan perbandingan 1:9. Sehingga, setelah ketiga desain model tersebut dibandingkan, maka *Decision Tree* dan SMOTE *Split Data* mendapatkan akurasi yang paling baik. Selain itu diperoleh pula presisi 89.71% (3:7) dan area under curve (AUC) sebesar 0.773 (1:9).

Penelitian terkait lainnya berjudul “*Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE*” [8]. Dalam penelitian ini selain menggunakan *Decision Tree* dan SMOTE, digunakan pula metode *Bagging*, *SVM* (*Support Vector Machine*), *MLP* (*Multi-Layer-Perceptron*) dan *Simple Logistic*. Penelitian yang bertujuan untuk menentukan pengklasifikasi terbaik untuk kumpulan data seimbang untuk memprediksi diabetes melitus tipe 2 dan menghasilkan akurasi hingga 94,7013%, dan kurva karakteristik operator penerima (ROC) 0,953 dengan pengklasifikasi pohon keputusan (*decision tree*). Validasi dicapai melalui validasi silang 10 kali lipat dengan percobaan yang dilakukan pada catatan klinis 734 pasien.

2. Metode Penelitian

Tahapan-tahapan yang akan dilakukan untuk mengimplementasi metode *Decision Tree* untuk klasifikasi data hasil survey jenis pekerjaan yang dipilih oleh pekerja ditujukan pada gambar 2.1



Gambar 1 Alur Penelitian

Sumber: Penulis

Dataset yang akan diuji dan dilatih sebagai tahapan pertama penelitian didapatkan dari data public pada situs kaggle. Tahapan penelitian yang akan dilalui yaitu akusisi data, pre-processing, merancang model dan evaluasi. Metode utama yang penulis gunakan ialah *Decision Tree* sebagai pembuat pohon keputusan yang kinerja modelnya ditingkatkan menggunakan algoritma SMOTE Upsampling sebagai penyeimbang data dari kelas minoritas sehingga dapat menghasilkan persentasi akurasi yang lebih tinggi[4]. Berikut tahapan berdasarkan gambar dari alur penelitian:



DOI: 10.52362/jisamar.v8i1.1350

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

1. Akuisi Data

Tahap yang awal dalam memulai sebuah penelitian, operasi yang dilakukan yaitu mengunduh/men-download dataset dari berbagai situs penyedia dataset secara gratis, dalam hal ini penulis mengakuisi data dari dataset public dari situs kaggle. Dalam penelitian ini dataset yang akan dipilih adalah data sekunder yang telah disediakan oleh situs penyedia repository.

2. Pre-processing

Pada tahap pre-processing ini, dari operasi yang digunakan untuk mendeteksi dan melakukan koreksi data-data yang kurang tepat, menghapus data berlebih, dan juga menambahkan atau menggantikan data yang terhapus. Selain itu proses labeling juga dilakukan pada tahap ini [9] Sehingga data yang sudah dibersihkan akan kemungkinan memberikan lebih banyak keberhasilan dalam data mining [10]. Pada tahap ini penggunaan operator Remove missing value digunakan guna menghilangkan nilai yang kosong.

3. Merancang Model

Langkah ini merupakan inti dari proses pengolahan data karna pada langkah ini dataset survey pekerja akan diekstrak. Algoritma klasifikasi dimana yang dipilih dalam penelitian ini ialah Decision Tree yang merupakan sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan dan setiap simpul daun pada decision tree menandai label kelas [11]

Bagian dari decision tree yaitu (leaf), batang/cabang dan akar yang bekerja sebagai titik awal dari suatu Decision Tree, batang/cabang sebagai perantara antara akar dan daun[12]. Decision tree akan menemukan solusi penyelesaian masalah dengan menjadikan standard note yang saling berkaitan untuk membentuk struktur pohon. Untuk menuju cabang berikutnya, sebuah cabang akan mewakili setiap atribut. Sehingga, setiap pohon tentunya memiliki cabang hingga berakhir di daun. Algoritma iterative dichotomizer (ID3) yaitu pembaruan dari Decision Tree C.45[13]

4. Evaluasi

Confusion matriks digunakan untuk mengetahui performa klasifikasi dimana terdapat sebuah pengukuran performa yang sering digunakan pada masalah klasifikasi dimana output dapat terdiri dari dua kelas atau lebih. confusion matrix merupakan teknik menganalisis seberapa baik kinerja dari pengklasifikasi dapat mengenali tupel dari kelas yang berbeda [14] Terdapat empat atribut yang merupakan kombinasi dari nilai yang diprediksi (predicted) dan nilai yang sebenarnya (actual)[15]

Berikut Persamaan model confusion matrix:

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

$$\text{Sensitive} = \text{Recall} = \text{TP}_{\text{rate}} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \text{TN}_{\text{rate}} = \frac{TN}{TN+FP}$$

$$\text{FP}_{\text{rate}} = \frac{FP}{FP+TN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

Dimana :

TP = True Positif , Memprediksi nilai sebenarnya atau kelas prediksi sesuai dengan kebenaran.

TN = True Negatif, nilai negative, dan model diprediksi negative



DOI: 10.52362/jisamar.v8i1.1350

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

FP = False positif, memprediksi positif, dan itu salah

FN = Nilai sebenarnya salah, dan model yang diprediksi salah.

Precision: menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model

Recall atau *sensitivity*: Nilai recall yang tinggi menunjukkan kecilnya nilai False Negative (terprediksi negative sebenarnya positive) [9] merupakan salah satu penggambaran keberhasilan model dalam menemukan kembali sebuah informasi

Berikut tahapan terakhir dari sebuah proses pengolahan data dimana data akan diuji secara akurasi dan visualisasi data yang sebelumnya telah diolah oleh algoritma Decision Tree dan SMOTE. Dimana teknik SMOTE bermanfaat untuk mengatasi masalah class imbalance problem (CIP). Untuk meningkatkan metode klasifikasi SMOTE memodifikasi dataset yang tidak sebanding untuk menciptakan data sintetik yang baru dari kelas minoritas. Dimana kemungkinan data kelas minoritas terduplikasi atau *overfitting* [16]. Penulis menggunakan tools rapidminer v.10.2 untuk mengolah data, visualisasi data dan menguji akurasi sehingga dapat memaksimalkan hasil penelitian

3. Hasil dan Pembahasan

1. Akusisi Data

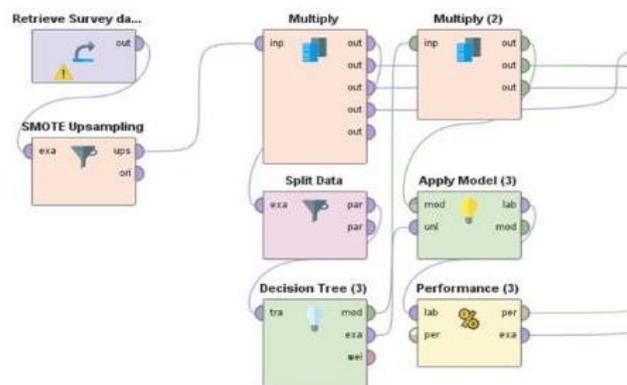
Tahap ini merupakan tahap yang pertama dalam penelitian. Data yang diperoleh merupakan data sekunder hasil survey yang terdapat pada situs kaggle yang di upload pada pertengahan tahun 2023. Dan didapatkan hasil survey dari 100 sample pekerja yang bertujuan untuk mengetahui persentase jumlah pekerja yang memilih bekerja dari rumah (WFH) atau berkerja dari kantor (WFO) maupun keduanya.

2. Pre-processing

Pada tahap pre-processing, atribut yang berupa pertanyaan survey kemudian disesuaikan dengan jenis datanya. Terdapat 11 atribut yaitu. Setelah itu penulis membagi data untuk kemudian diuji dengan perbandingan 5:5, dimana 50% data untuk training dan 50% data untuk testing sehingga menciptakan output penelitian yakni berupa sistem kerja yang dipilih oleh para pekerja terhadap pandemi covid-19 dalam bentuk pohon keputusan (decision tree).

3. Merancang Model

Agar dapat menampilkan visualisasi model serta menampilkan akurasi sebagai langkah pertama untuk merancang model adalah dengan membuat desain model. Berikut desain model yang akan ditampilkan pada gambar 3.1 yaitu menganalisa peningkatan kinerja pada model Decision Tree.



Gambar 2 Model Proses Pengujian

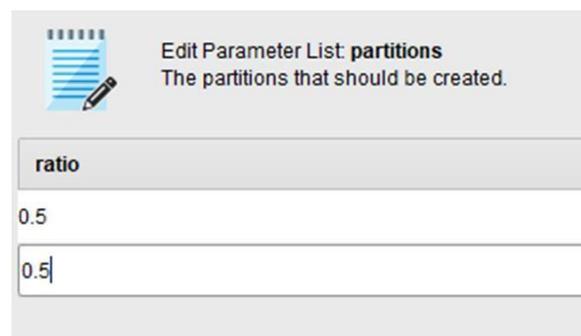
Sumber: penulis

Pada gambar 3.1 dataset yang telah diunduh dimasukkan kedalam repository yang ada didalam rapidminer kemudian sesuaikan atribut dengan tipe datanya. Fenomena imbalance class terjadi pada kasus ini dimana pola yang memilih bekerja secara remote, berkerja secara hybrid dan berkerja dikantor memiliki kemiringan data. Dimana kelas mayoritas mengabaikan kelas minoritas. Yaitu pada atribut “Which work type do you prefer the most?” sebagai label. Guna mengatasi hal tersebut digunakan Syntetic Minority Over-sampling Technique atau operasi SMOTE Upsampling. Terakhir untuk penentuan akurasi penulis menggunakan operasi performance.

4. Evaluasi

Penelitian ini bertujuan untuk mengetahui persentase jumlah pekerja yang memilih bekerja dari rumah (WFH) atau berkerja dari kantor (WFO) maupun keduanya menggunakan metode klasifikasi Decision Tree, operator performance guna mengetahui hasil dari confusion matriks dan teknik penyeimbang data yakni SMOTE. Berdasarkan faktor internal maupun eksternal pasca pandemi yang membuat sebagian pekerja lebih memilih salah satu jenis kerja. Output dari evaluasi ini berupa akurasi perhitungan data dan hasilnya berupa visualisasi data.

Pada pengujian dengan menggunakan model Decision Tree dan SMOTE Split Data dilakukan split yaitu membagi data menjadi 2 subset atau partisi.



Gambar 3 Split Data

Sumber: Penulis

Split data salah satu metode untuk membagi data dari aspek yang ada sebagian mempengaruhi model klasifikasi pada algoritma pembelajaran mesin [17]. Split data adalah metode yang akan membagi data menjadi 2 yaitu data latih dan data uji. Pada gambar rasio yang diuji antara 0.1 (1:9) hingga 0.9 (9:1).

Tabel 1. Perbandingan Hasil Pengujian

Ratio	Akurasi
0,1 (1:9)	69,23%
0,2 (2:8)	76,92%
0,3 (3:7)	71,79%
0,4 (4:6)	76,92%
0,5 (5:5)	83,08%
0,6 (6:4)	82,05%
0,7 (7:3)	76,92%



DOI: 10.52362/jisamar.v8i1.1350

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

0,8 (8:2)	75,00%
0,9 (9:1)	76,07%

Sumber: Penulis

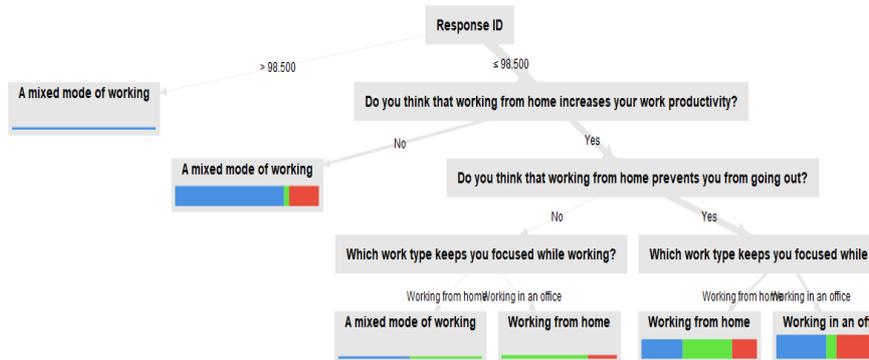
Berdasarkan tabel 3.1 dan gambar 3.3. Pengujian yang dilakukan berdasarkan jumlah data pada tiap-tiap rasio. Hasil terbaik dari pengujian ini terletak pada ratio 0,5 (5:5) atau 50% sebagai data training dan 50% sisanya sebagai data testing yang memiliki akurasi 83,8%. Dari table di bawah tampak yang dapat bekerja secara hybrid di prediksi 77,78% tetapi setelah dilakukan recall meningkat 84% atau benar dapat bekerja secara hybrid lebih besar. Presentase orang yang bekerja di rumah atau secara remote adalah 84% namun setelah dilakukan recall tetap 84% dan terakhir orang diprediksi senang bekerja di kantor 92,31% namun setelah dilakukan proses recall menghasilkan 80% terjadi penurunan, hal ini perlu lebih diperhatikan kinerja pada saat di kantor agar karyawan lebih efisien ketika bekerja.

accuracy: 83.08%

	true A mixed mode of working	true Working from home	true Working in an office	class precision
pred. A mixed mode of working	21	4	2	77.78%
pred. Working from home	3	21	1	84.00%
pred. Working in an office	1	0	12	92.31%
class recall	84.00%	84.00%	80.00%	

Gambar 4 Tabel Confusion Matriks

Sumber: Penulis

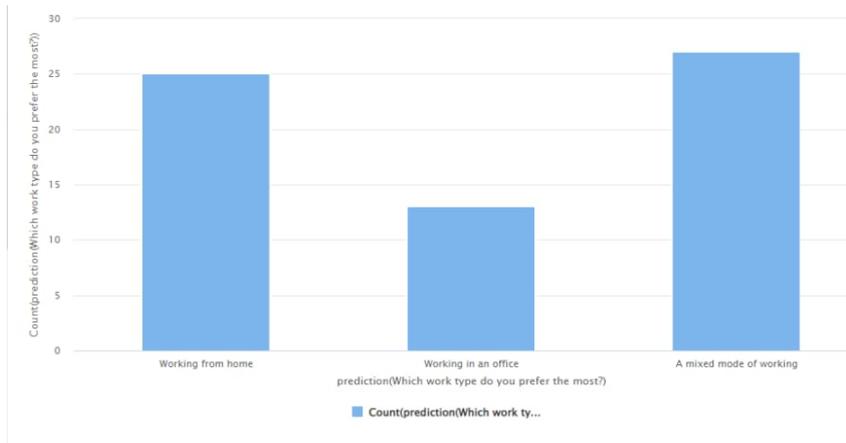


Gambar 5 Hasil Analisa Decision Three

Dari gambar terlihat bahwa beberapa orang walaupun memilih bekerja di rumah tetapi lebih sesuai untuk bekerja di kantor, begitupun sebaliknya.

Berikut merupakan perbandingan data yang memilih untuk bekerja dari rumah, bekerja di kantor dan hibrid bekerja dirumah dan dikantor.





Gambar 6 Perbandingan Jumlah Prediksi

4. Kesimpulan

Dari hasil penelitian dapat disimpulkan bahwa, sebanyak 13% pekerja lebih memilih bekerja dari rumah, 25% memilih bekerja dari kantor dan 27% memilih mode kerja campuran. Serta, pengujian yang dilakukan Decision Tree dan SMOTE split data menghasilkan akurasi tertinggi hingga 83,08% yang terdapat pada ratio 0,5 (5:5) sehingga hasil akurasi ini masuk kedalam kriteria tepat dalam melakukan prediksi.

Daftar Pustaka

- [1] W. Mckibbin and R. Fernando, "Crawford School of Public Policy CAMA Centre for Applied Macroeconomic Analysis The Brookings Institution Centre of Excellence in Population Ageing Research The Global Macroeconomic Impacts of COVID-19: Seven Scenarios *," 2020.
- [2] D. Rianto Rahadi, "DILEMA WORK FROM HOME DIMASA PANDEMI STUDI KAWASAN INDUSTRI BEKASI," 2021. [Online]. Available: <http://ejournal.unsri.ac.id/index.php/jmbs>
- [3] D. Mustajab *et al.*, "THE INTERNATIONAL JOURNAL OF APPLIED BUSINESS TIJAB Fenomena Bekerja dari Rumah sebagai Upaya Mencegah Serangan COVID-19 dan Dampaknya terhadap Produktifitas Kerja Working from Home Phenomenon as an Effort to Prevent COVID-19 Attacks and Its Impacts on Work Productivity".
- [4] E. Sutoyo, M. Asri Fadlurrahman, J. Telekomunikasi Jl Terusan Buah Batu, K. Dayeuhkolot, K. Bandung, and J. Barat, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *Jurnal Edukasi dan Penelitian Informatika*, vol. 6, no. 3, pp. 379–385, 2020.
- [5] N. Indriani, E. Rainarli, K. Evita Dewi, T. dan Ilmu Komputer, and J. Dipati Ukur, "JURNAL INFOTEL Informatika-Telekomunikasi-Elektronika Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *Jurnal Infotel*, vol. 9, no. 4, 2017, doi: 10.20895/infotel.v9i4.
- [6] B. H. Mawaridi and M. Faisal, "Rekomendasi Merk Mobil Untuk Calon Pembeli Menggunakan Algoritma Decision Tree," *Jurnal Informatika*, vol. 10, no. 2, pp. 157–162, Oct. 2023, doi: 10.31294/inf.v10i2.16000.



DOI: 10.52362/jisamar.v8i1.1350

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- [7] A. Franseda, W. Kurniawan, S. Anggraeni, and W. Gata, “Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 282, Jul. 2020, doi: 10.26418/justin.v8i3.40982.
- [8] M. Shuja, S. Mittal, and M. Zaman, “Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE,” 2020, pp. 195–211. doi: 10.1007/978-981-15-0222-4_17.
- [9] Ni'ma Kholila, “Merdeka belajar twitter,” *Jurnal Ilmiah Teknik Informatika*, vol. 15, no. 2, pp. 252–261, 2021, Accessed: Oct. 11, 2022. [Online]. Available: <https://doi.org/10.35457/antivirus.v15i2.1866>
- [10] P. Ristoski and H. Paulheim, “Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey.” [Online]. Available: <https://ssrn.com/abstract=3199217>
- [11] A. Muzakir and R. A. Wulandari, “Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree,” *Scientific Journal of Informatics*, vol. 3, no. 1, 2016, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [12] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 646, Dec. 2021, doi: 10.32493/informatika.v5i4.7622.
- [13] E. Priyanti, “Penerapan Decision Tree Pada Penentuan Lokasi Waralaba,” *JURNAL SWABUMI*, vol. 11, no. 1, pp. 8–12, 2023.
- [14] J. Han *et al.*, “Designing Data-Intensive Web Applications,” 2012.
- [15] T. Iskandar Zulkarnain Maulana Putra, A. Farhan Bukhori, dan Ilmu Pengetahuan Alam, and U. Gadjah Mada, “Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long Short-Term Memory for Indonesian-language Tweets),” *Jurnal Ilmu Siber dan Teknologi Digital (JISTED)*, vol. 1, no. 1, pp. 1–28, 2022, doi: 10.35912/jisted.v1i1.1509.
- [16] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, “Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE,” *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.
- [17] A. Nurhopipah and U. Hasanah, “Dataset Splitting Techniques Comparison For Face Classification on CCTV Images,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 4, p. 341, Oct. 2020, doi: 10.22146/ijccs.58092.

