p-ISSN: 2963-7856| e-ISSN: 2961-9890
Available online at jerkin.org/index.php/jerkin
Jurnal Pengabdian Masyarakat dan Riset Pendidikan
Volume 4 No. 3 Januari-Maret 2026, pp 17639-17644

Implementasi Algoritma C4.5 untuk Klasifikasi Status Tingkat Pengangguran di Indonesia Berdasarkan Jenjang Pendidikan

Fadel Tri Ismar[1*], Muhammad Reza[2], Muhammad Faizal Afif[3], Muhammad Rayyan Saputra[4], Ammar[5]

[1,2,3,4,5]Sistem Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta
E-mail: fadeltriismar01@gmail.com
* Corresponding Author

https://doi.org/10.31004/jerkin.v4i3.5076

ARTICLE INFO

Article history
Received: 25 Dec 2025
Revised: 05 Jan 2026
Accepted: 16 Jan 2026

Kata Kunci:
Pengangguran, Jenjang Pendidikan, Algoritma C4.5, Data Mining, *Decision Tree*

Keywords:
Unemployment, Education Level, C4.5 Algorithm, Data Mining, Decision Tree

ABSTRACT

Pengangguran merupakan salah satu masalah ekonomi utama di Indonesia yang fluktuatif setiap tahunnya. Data Badan Pusat Statistik (BPS) menunjukkan bahwa tingkat pendidikan tidak selalu menjamin ketersapan tenaga kerja, dimana fenomena pengangguran terdidik (lulusan SMK/Universitas) sering terjadi. Penelitian ini bertujuan untuk mengklasifikasikan status tingkat pengangguran nasional (kategori "Tinggi" atau "Rendah") berdasarkan variabel jenjang pendidikan (SD, SMP, SMA, SMK, Diploma, Universitas). Metode yang digunakan adalah Algoritma Data Mining C4.5 karena kemampuan algoritma ini dalam membentuk pohon keputusan (Decision Tree) yang mudah diinterpretasikan. Proses pengolahan data dilakukan menggunakan tools RapidMiner dengan membagi data menjadi data latih dan data uji. Hasil penelitian ini berupa aturan (rules) yang dapat digunakan pemerintah atau pemangku kebijakan untuk mengetahui jenjang pendidikan mana yang paling berkontribusi terhadap tingginya angka pengangguran nasional.

Unemployment is one of the main economic problems in Indonesia that fluctuates every year. Data from the Central Statistics Agency (BPS) shows that the level of education does not always guarantee the absorption of the workforce, where the phenomenon of educated unemployment (vocational high school/university graduates) often occurs. This study aims to classify the status of the national unemployment rate (category "High" or "Low") based on the variable of education level (elementary school, junior high school, senior high school, vocational high school, diploma, university). The method used is the C4.5 Data Mining Algorithm because of its ability to form a decision tree (Decision Tree) that is easy to interpret. Data processing is carried out using the RapidMiner tool by dividing the data into training data and test data. The results of this study are in the form of rules that can be used by the government or policy makers to determine which level of education contributes most to the high national unemployment rate.

How to Cite: Fadel Tri Ismar, et al (2026). Implementasi Algoritma C4.5 untuk Klasifikasi Status Tingkat Pengangguran di Indonesia Berdasarkan Jenjang Pendidikan, 4(3) 17639-17644. https://doi.org/10.31004/jerkin.v4i3.5076

PENDAHULUAN

Masalah pengangguran dan ketenagakerjaan merupakan isu krusial yang dihadapi oleh hampir semua negara di dunia, terutama negara-negara berkembang. Kedua isu ini saling terkait dan menciptakan sebuah dualisme: di satu sisi, kelebihan tenaga kerja dapat menjadi beban yang menghambat pertumbuhan ekonomi jika tidak dikelola dengan baik. Di sisi lain, jika pemerintah mampu memanfaatkannya secara efektif, surplus tenaga kerja ini justru dapat menjadi aset berharga yang mengakselerasi pembangunan nasional. Dengan demikian, kemampuan pemerintah dalam mengelola sumber daya manusia menjadi faktor penentu apakah ketenagakerjaan akan menjadi peluang atau justru

ancaman bagi stabilitas ekonomi (Alisjahbana, Armida.2008). Indonesia memiliki posisi unik di kancah global dengan angkatan kerja yang tergolong salah satu yang terbesar di dunia (World Bank, 2013). Seiring dengan pertambahan penduduk, jumlah angkatan kerja nasional terus menunjukkan tren peningkatan, seperti yang terekam pada periode 2013-2014 dimana terjadi penambahan sekitar 2,5 juta jiwa (BPS, 2014).

Kondisi ini menempatkan pemerintah pada posisi krusial untuk memastikan pemanfaatan sumber daya manusia secara optimal. Kegagalan dalam menyerap angkatan kerja yang terus bertambah ini akan mengakibatkan peningkatan angka pengangguran, yang pada akhirnya menjadi penghambat utama bagi perekonomian dan pembangunan berkelanjutan (Makna & Indonesia, 2015).

Mengingat pengangguran merupakan isu krusial yang berdampak langsung pada stabilitas ekonomi, analisis mendalam terhadap faktor-faktor penentunya menjadi sangat penting. Salah satu atribut yang paling fundamental dalam menentukan status pekerjaan seseorang adalah jenjang pendidikan. Namun, memetakan pola hubungan antara tingkat pendidikan dengan status pengangguran secara manual pada kumpulan data yang besar merupakan tugas yang kompleks dan rentan terhadap kesalahan interpretasi.

Untuk mengatasi tantangan analisis tersebut, diperlukan sebuah pendekatan komputasi yang sistematis. Oleh karena itu, penelitian ini mengusulkan penerapan data mining dengan menggunakan Algoritma C4.5 untuk membangun model klasifikasi yang dapat memprediksi status pengangguran. Algoritma C4.5 dipilih karena kemampuannya yang andal dalam membentuk pohon keputusan (decision tree), yaitu sebuah model yang mentransformasikan data kompleks menjadi aturan-aturan yang mudah dipahami.

Dengan menggunakan data Badan Pusat Statistik di Indonesia, penelitian ini bertujuan untuk mengimplementasikan Algoritma C4.5 guna mengklasifikasikan status tingkat pengangguran berdasarkan jenjang pendidikan. Hasil dari penelitian ini diharapkan dapat menghasilkan sebuah model klasifikasi yang akurat serta memberikan wawasan berbasis data mengenai bagaimana tingkat pendidikan berkorelasi dengan status pekerjaan, yang pada akhirnya dapat menjadi masukan berharga bagi perumusan kebijakan ketenagakerjaan nasional (M. Taufany Firmansyah,2016)

METODE

Jenis dan Pendekatan Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif dengan metode Data Mining (Penambangan Data). Fokus utama penelitian adalah pada teknik klasifikasi, yaitu implementasi Algoritma C4.5 untuk membangun model Pohon Keputusan (Decision Tree). Pendekatan ini dipilih karena Algoritma C4.5 mampu mentransformasikan data kompleks menjadi serangkaian aturan yang mudah diinterpretasikan, yang sangat relevan untuk memberikan wawasan berbasis data kepada pemangku kebijakan.

Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai tingkat pengangguran di Indonesia. Sumber data utama adalah data resmi yang dipublikasikan oleh Badan Pusat Statistik (BPS).

Data tersebut terdiri dari:

1. Variabel Prediktor (Atribut): Jenjang pendidikan, yang meliputi kategori SD, SMP, SMA, SMK, Diploma, dan Universitas.
2. Variabel Target (Kelas): Status tingkat pengangguran nasional, yang diklasifikasikan menjadi dua kategori, yaitu "Tinggi" atau "Rendah".

Pengumpulan dan Pra-pemrosesan Data

Data tingkat pengangguran dari BPS dikumpulkan dan dilanjutkan dengan tahap pra-pemrosesan. Tahap ini mencakup pembersihan data, penanganan nilai yang hilang (jika ada), dan transformasi data agar sesuai dengan format input yang dibutuhkan oleh Algoritma C4.5.

Pembagian Data

Data yang telah diproses dibagi menjadi dua set utama:

1. Data Latih (Training Data): Digunakan oleh Algoritma C4.5 untuk mempelajari pola dan membangun model Pohon Keputusan.

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

2. Data Uji (Testing Data): Digunakan untuk menguji dan memvalidasi performa model yang telah dibangun.

Implementasi Algoritma C4.5

Algoritma C4.5 diimplementasikan untuk menghasilkan Pohon Keputusan. Proses ini melibatkan perhitungan nilai Entropy dan Gain untuk menentukan atribut (jenjang pendidikan) mana yang paling informatif dan efektif dalam membagi data, sehingga menghasilkan aturan klasifikasi yang optimal.

Evaluasi Model

Model klasifikasi yang dihasilkan dievaluasi untuk mengukur tingkat keakuratan dan kemampuan generalisasinya. Teknik Cross Validation digunakan untuk memastikan estimasi performa model yang objektif. Metrik evaluasi yang digunakan meliputi:

1. Akurasi (Accuracy): Proporsi prediksi yang benar secara keseluruhan.
2. Sensitivitas (Recall): Kemampuan model untuk mengidentifikasi semua kasus positif dengan benar.
3. Presisi (Precision): Proporsi kasus positif yang diprediksi benar dari semua kasus yang diprediksi positif. Perhitungan metrik ini didasarkan pada hasil dari Confusion Matrix.

Alat Bantu Penelitian

Seluruh proses pengolahan dan analisis data, mulai dari pra-pemrosesan, implementasi Algoritma C4.5, hingga evaluasi model, dilakukan menggunakan perangkat lunak RapidMiner. RapidMiner berfungsi sebagai alat bantu komputasi yang memfasilitasi penerapan teknik data mining secara terstruktur dan efisien.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai tingkat pengangguran di Indonesia. Sumber data utama adalah data resmi yang dipublikasikan oleh Badan Pusat Statistik (BPS).

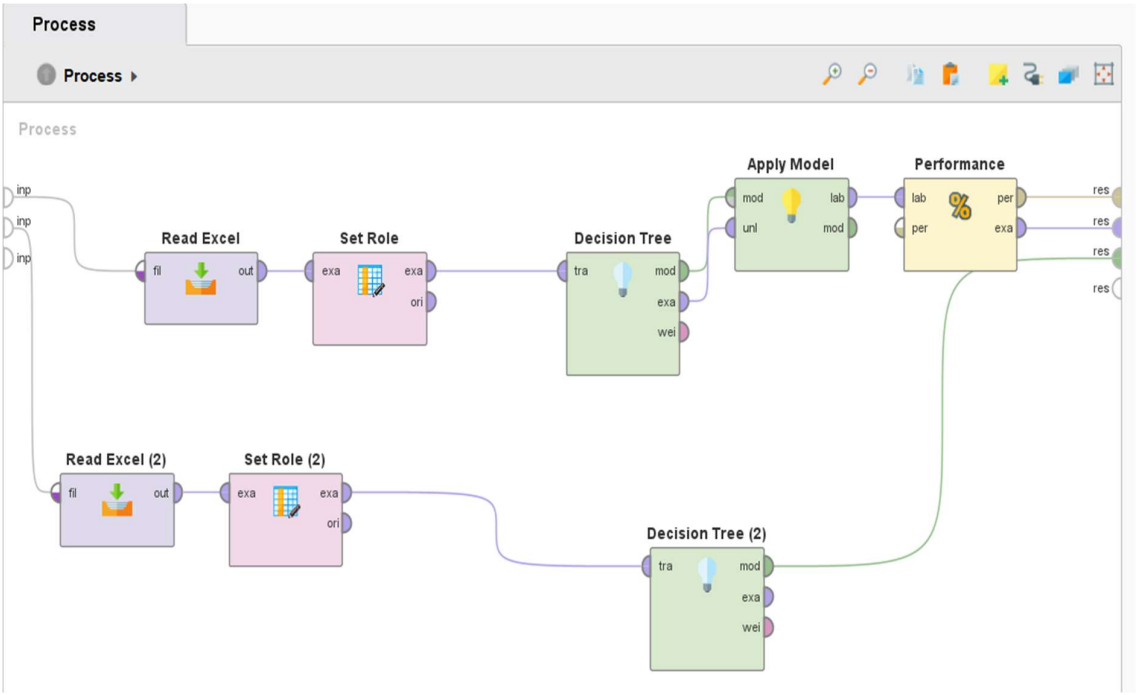
	A	B	E	F	G	H	I	J	K	L	M	N
1	Periode	Bulan	SD	SLTP	SLTA Umu	SLTA Kejur	Akademi/D	Universita	Total	Status_Pengangguran		
2	2006	Februari	2675459	2860007	2842876	1204140	297185	375601	11104693	Tinggi		
3	2006	Agustus	2589699	2730045	2851518	1305190	278074	395554	10932000	Tinggi		
4	2007	Februari	2753548	2643062	2630360	1114675	330316	409890	10547917	Tinggi		
5	2007	Agustus	2179792	2264198	2532204	1538349	397191	566588	10011142	Tinggi		
6	2008	Februari	2216748	2166619	2204377	1165582	519867	626202	9427590	Tinggi		
7	2008	Agustus	2058913	2083656	2241315	1336146	515865	695724	9390881	Tinggi		
8	2009	Februari	2035930	2073998	2156637	1268383	630440	665551	9252567	Tinggi		
9	2009	Agustus	1894982	2160173	2225471	1427131	540127	624231	9268181	Tinggi		
10	2010	Februari	1750531	1948514	1956551	1202862	604473	760822	8591625	Tinggi		
11	2010	Agustus	1663271	1904797	1966547	1335759	380112	736025	8319535	Tinggi		
12	2011	Februari	1632742	1870631	1732644	1202287	419515	479603	7713465	Rendah		
13	2011	Agustus	1637770	1780447	1955364	1276226	303530	419692	7700566	Rendah		
14	2012	Februari	1654877	1756543	1717361	1143826	357777	636928	7610719	Rendah		
15	2012	Agustus	1563228	1620025	1887858	1143890	305586	396037	7235280	Rendah		
16	2013	Februari	1542157	1652436	1631777	1048633	313491	688569	7170541	Rendah		
17	2013	Agustus	1629851	1626013	1893509	847365	195258	398298	7406321	Rendah		
18	2014	Februari	1502479	1552523	1893509	847365	195258	398298	7147069	Rendah		
19	2014	Agustus	1229652	1566838	1962786	1332521	193517	495143	7244905	Rendah		
20	2015	Februari	1320392	1650387	1762411	1174366	254312	565402	7454767	Rendah		
21	2015	Agustus	1004961	1373919	2280029	1569690	251541	653586	7560822	Rendah		
22	2016	Februari	1218954	1313815	1546699	1348327	249362	695304	7024172	Rendah		
23	2016	Agustus	1035731	1294483	1805527	1702738	214051	535594	7031539	Rendah		
24	2017	Februari	1153835	1340156	1545620	1383022	249876	605553	7010000	Rendah		
25	2017	Agustus	1022209	1336467	1779872	1879502	249219	618758	7040000	Rendah		
26	2018	Februari	1093122	1273926	1468205	1427192	234141	677321	6870000	Rendah		
27	2018	Agustus	1003738	1372793	1920700	1730018	241036	737497	7070000	Rendah		
28	2019	Februari	1067750	1306352	1543888	1451240	227361	725803	6820000	Rendah		
29	2019	Agustus	1035417	1331613	1956550	1838634	233777	719212	7050000	Rendah		

Pengolahan Data

Alur kerja ini secara keseluruhan menggambarkan proses evaluasi model klasifikasi Decision Tree dengan skema hold-out(data latih dan data uji terpisah).

Fokus utama:

1. Melatih model decision Tree menggunakan set data pertama.
2. Menerapkan model yang telah (Apply Model) pada set data kedua(Data Uji).
3. Mengevaluasi kinerja model (Peformance) berdasarakan hasil prediksi pada Data uji.



Pengukuran Penelitian

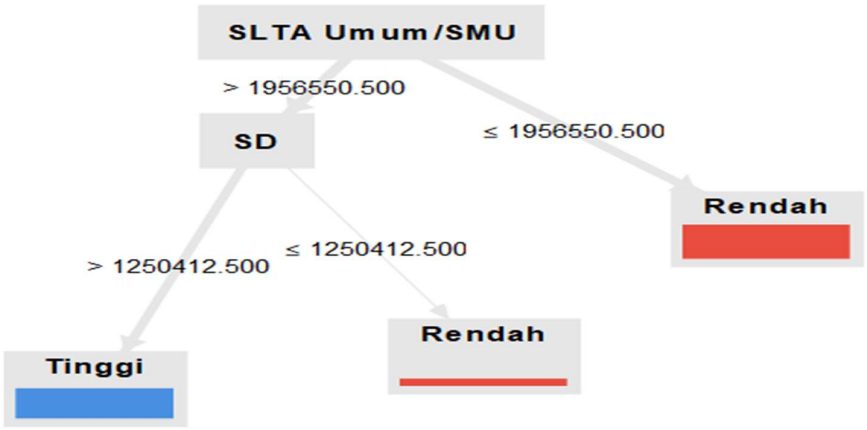
Berdasarkan hasil olah data melalui aplikasi RapidMiner, maka diperoleh Confusion Matrix guna mengukur tingkat akurasinya dari implementasi Algoritma C4.5. Hasil evaluasi menunjukkan performa yang sempurna, yaitu sebesar 100.00% untuk Akurasi keseluruhan. Selain itu, model juga mencapai 100.00% untuk Precision (baik untuk kelas Tinggi maupun Rendah) dan 100.00% untuk Sensitivity atau Recall (baik untuk kelas Tinggi maupun Rendah), seperti yang dapat kita lihat pada tabel hasil evaluasi.

accuracy: 100.00%

	true Tinggi	true Rendah	class precision
pred. Tinggi	15	0	100.00%
pred. Rendah	0	20	100.00%
class recall	100.00%	100.00%	

Berdasarkan data yang disajikan pada tabel di atas, akan menghasilkan sebuah model pohon keputusan seperti berikut :

ISSN 2963-7856 (print), ISSN 2961-9890 (online)



Analisa Hasil Penelitian Data

Melalui data dari pohon Pohon Keputusan yang dihasilkan oleh Algoritma C4.5 ini merupakan model klasifikasi yang sangat ringkas, hanya terdiri dari dua tingkat keputusan (kedalaman 2), yang bertujuan untuk memprediksi status tingkat pengangguran (Tinggi atau Rendah). Variabel SLTA Umum/SMU teridentifikasi sebagai akar (root node), menjadikannya faktor penentu utama dalam klasifikasi.

Berikut adalah rule yang berhasil di dapatkan dari pohon keputusan :

Tree

```
Periode = [-∞ - 2007.7]: Tinggi {Tinggi=4, Rendah=0}
Periode = [2007.7 - 2009.4]: Tinggi {Tinggi=4, Rendah=0}
Periode = [2009.4 - 2011.1]
|   Tidak/belum tamat SD = [-∞ - 274597.1]: Rendah {Tinggi=0, Rendah=1}
|   Tidak/belum tamat SD = [274597.1 - 312415.2]: Tinggi {Tinggi=2, Rendah=0}
|   Tidak/belum tamat SD = [312415.2 - 350233.3]: Rendah {Tinggi=0, Rendah=1}
Periode = [2011.1 - 2012.8]: Rendah {Tinggi=0, Rendah=2}
Periode = [2012.8 - 2014.5]: Rendah {Tinggi=0, Rendah=4}
Periode = [2014.5 - 2016.2]: Rendah {Tinggi=0, Rendah=4}
Periode = [2016.2 - 2017.9]: Rendah {Tinggi=0, Rendah=2}
Periode = [2017.9 - 2019.6]: Rendah {Tinggi=0, Rendah=4}
Periode = [2019.6 - 2021.3]
|   Tidak/belum pernah sekolah = [-∞ - 52192.5]: Rendah {Tinggi=0, Rendah=1}
|   Tidak/belum pernah sekolah = [52192.5 - 72445.0]: Tinggi {Tinggi=1, Rendah=0}
|   Tidak/belum pernah sekolah = [72445.0 - 92697.5]: Tinggi {Tinggi=2, Rendah=0}
Periode = [2021.3 - ∞]
|   Tidak/belum tamat SD = [350233.3 - 388051.4]: Rendah {Tinggi=0, Rendah=1}
|   Tidak/belum tamat SD = [388051.4 - 425869.5]: Tinggi {Tinggi=2, Rendah=0}
```

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma C4.5 dalam mengklasifikasikan status tingkat pengangguran di Indonesia berdasarkan jenjang pendidikan dapat berjalan dengan baik dan menghasilkan model yang sangat mudah diinterpretasikan. Algoritma ini mampu membentuk struktur pohon keputusan (decision tree) yang merepresentasikan hubungan antar variabel pendidikan terhadap status pengangguran dalam bentuk aturan-aturan logis yang sederhana dan sistematis.

Hasil pengujian model menggunakan perangkat lunak RapidMiner menunjukkan performa klasifikasi yang sangat tinggi, dengan nilai akurasi, precision, dan recall mencapai 100%. Hal ini mengindikasikan bahwa model mampu mengklasifikasikan data dengan sangat baik pada dataset yang digunakan. Selain itu, variabel SLTA Umum/SMU teridentifikasi sebagai atribut paling dominan dalam menentukan status tingkat pengangguran, yang menunjukkan bahwa jenjang pendidikan menengah umum memiliki kontribusi signifikan dalam membedakan kondisi pengangguran nasional.

Namun demikian, hasil evaluasi yang sempurna juga mengindikasikan adanya potensi overfitting, terutama mengingat ukuran dataset yang relatif terbatas serta struktur pohon keputusan yang sederhana. Model cenderung sangat sesuai dengan data yang digunakan, tetapi belum tentu memiliki kemampuan generalisasi yang sama baiknya ketika diterapkan pada data dengan karakteristik yang berbeda. Oleh karena itu, meskipun model yang dihasilkan valid secara teknis dan metodologis, interpretasi hasil tetap perlu dilakukan secara kritis dan proporsional.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma C4.5 efektif digunakan sebagai alat bantu analisis dan pengambilan keputusan berbasis data, khususnya dalam memahami pola hubungan antara jenjang pendidikan dan status pengangguran. Model yang dihasilkan dapat menjadi referensi awal bagi pemerintah atau pemangku kebijakan dalam merumuskan strategi peningkatan kualitas pendidikan dan penyerapan tenaga kerja.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

Ba, I., Althubiti, T., Alharbi, A., Alfarsi, K., & Rasheed, S. (2021). A Short Review of Classification Algorithms Accuracy for Data Prediction in Data Mining Applications. 162–174. https://doi.org/10.4236/jdaip.2021.93011

Faid, M. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. 8. https://doi.org/10.34148/teknika.v8i1.95

Klasifikasi, C. U. (2016). Implementasi Metode Decision Tree Dan Algoritma.Landgrebe, D. (1990). iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii /-'// -_.

Makna, R., & Indonesia, M. (2015). Homo Cooperativus : November.

Menerima, D., & Sosial, B. (2022). , Christian Dwi Sofyan. 1(2), 39–47.

Nasional, J., Informasi, S., Riandaru, V., Lazuardi, H., Adhi, A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier. 01, 8–17.

Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi data mining dengan algoritma c4 . 5 untuk penjurusan siswa (studi kasus : sma negeri 1 pontianak). 04(3).

Pertiwi, R. N., Azizah, D. F., & Kurniawan, B. C. (2014). View metadata, citation and similar papers at core.ac.uk. Analisis Efektivitas Pemungutan Pajak Bumi Dan Bangunan (Studi Pada Dinas Pendapatan, Pengelolaan Keuangan Dan Aset Kota Probolinggo), 3(1), 1–7.

Tinggi, S. S., & Utama, K. W. (2017). Penelitian komparasi algoritma klasifikasi. 1(1), 1–12.

Firdaus, N. D., Indriana, M. R., Muizzah, U., & ... (2023). Strategi Harmoni Hak dan Kewajiban Bela Negara Melalui Pajak. Jurnal Pendidikan Transformatif (JPT), 02(06), 24–34. https://jupetra.org/index.php/jpt/article/view/1053%0Ahttps://jupetra.org/index.php/jpt/article/download/1053/355

Tiraada, T. A. M. (2013). Kesadaran Perpajakan, Sanksi Pajak, Sikap Fiskus Terhadap Kepatuhan Wpop Di Kabupaten Minahasa Selatan. Jurnal Emba, 1(3), 999–1008.

Pertiwi, R. N., Azizah, D. F., & Kurniawan, B. C. (2014). View metadata, citation and similar papers at

ISSN 2963-7856 (print), ISSN 2961-9890 (online)