



IMPLEMENTASI ALGORITMA J48 DENGAN TEKNIK BAGGING UNTUK PREDIKSI KIPI PESERTA VAKSINASI COVID-19

Eka Rahmawati¹, Candra Agustina²

¹Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika,

²Program Studi Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika.

eka.eat@bsi.ac.id , candra.caa@bsi.ac.id

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10420

Keywords:

*Covid-19 Vaccine,
J14, AEFI, Bagging.*

Abstract

The Covid 19 vaccination is considered to be the most effective way to prevent the spread of the Corona Virus, in addition to a clean lifestyle such as washing hands, wearing masks, and keeping a distance from other people. Several large vaccine manufacturing companies in the world have issued a product in the form of a Covid-19 vaccine with various levels of effectiveness. The vaccine is still being distributed throughout the world, including Indonesia. The vaccine obtained an emergency distribution permit from the authorized institution and was administered to community groups that meet the requirements. However, during the implementation of the vaccine, many AEFIs (Post Immunization Adverse Events) were found, such as dizziness, fever, headaches, and some even fainted. Although not dangerous but quite disturbing for people with solid activities. Therefore, it is necessary to predict whether participants will get AEFI or not. The data consists of 8 Attributes, after being processed using the J48 Algorithm, the results show that the attributes that have a strong influence are 7 Attributes, while the rest have no major effect. The accuracy level of the prediction model obtained is 91,23% with this level of accuracy, precision 0,918, recall 0,912, F-Measure 0,913 and ROC Area 0,966. It means that the model can be utilized by the parties concerned to then be able to anticipate.

Kata Kunci:

*Vaksin Covid-19,
J48, KIPI, Bagging.*

Abstrak

Vaksinasi Covid 19 dinilai menjadi cara paling efektif untuk mencegah penyebaran Virus Corona, selain gaya hidup bersih seperti mencuci tangan, memakai masker dan menjaga jarak dengan orang lain. Beberapa perusahaan produsen vaksin besar di dunia mengeluarkan produk berupa vaksin Covid-19 dengan berbagai tingkat efektivitasnya. Vaksin tersebut secara masih disebarkan keseluruh penjuru Dunia tak terkecuali Indonesia. Vaksin tersebut mendapatkan ijin edar darurat dari lembaga yang berwenang dan dilakukan pada kelompok masyarakat yang memenuhi persyaratan. Akan tetapi pada pelaksanaan vaksin banyak ditemukan KIPI (Kejadian Ikutan Pasca Imunisasi), seperti pusing, demam, sakit kepala bahkan ada yang sampai pingsan. Meskipun tidak berbahaya namun cukup mengganggu untuk orang-orang dengan aktivitas padat. Oleh karena itu perlu dilakukan penelitian untuk memprediksi apakah peserta akan mengalami KIPI atau tidak. Data terdiri dari 8 Atribut, setelah diolah menggunakan Algoritma J48 mendapatkan hasil bahwa atribut yang punya pengaruh kuat sebanyak 7 Atribut, sedangkan sisanya tidak berpengaruh besar. Tingkat akurasi yang model prediksi yang didapatkan sebesar 91,22% dengan *precision* 0.918, *recall* 0.912, F-Measure 0.913 dan ROC Area 0.966, artinya model dapat

dimanfaatkan pihak-pihak yang terkait untuk kemudian dapat melakukan antisipasi.

Pendahuluan

Munculnya virus Corona pada tahun 2019 di Wuhan telah memberikan dampak keseluruhan belahan dunia. *World Health Organization* (WHO) menyatakan virus ini sebagai pandemi pada Maret 2020[12]. Artinya wabah sudah menyebar secara global diseluruh penjuru Dunia. Tentu saja WHO dan masyarakat didunia tidak tinggal diam, penelitian terus dilakukan untuk menghentikan laju penyebaran virus. Selain langkah-langkah seperti memakai masker, mencuci tangan dan menjaga jarak perusahaan vaksin terkemuka dunia juga berlomba untuk menemukan vaksin yang cocok untuk menaggulangi virus tersebut. Di Indonesia sendiri pemberian vaksin dimulai tanggal 13 Januari 2021. Pemberian vaksin pertama kali dilakukan oleh Presiden RI yaitu Bapak Ir. Joko Widodo. Program awal pemerintah menggunakan vaksin Sinovac yang berasal dari China dan mendapatkan ijin edar darurat dari BPOM. Dalam perkembangannya pemerintah juga menggunakan jenis vaksin lain seperti Astra Zeneca, Sinoparm, Pfizer dan Moderna. Vaksin-vaksin tersebut dalam uji klinis terbukti menunjukkan imunogenisitas yang bagus dengan berbagai tingkat efektifitas[1].

Pada November 2021 capaian vaksinasi di Indonesia mencapai 38,03%.



Gambar 1. Infografis Capaian Vaksinasi Tinggi dan Lonjakan Kasus Covid-19 [2]

Akan tetapi program ini bukan tanpa masalah, karena terjadinya KIPI membuat masyarakat banyak yang ragu bahkan takut untuk melakukan vaksinasi. KIPI yang sering dialami adalah demam, mual muntah, menggigil, kelelahan, nyeri kepala, nyeri sendi dan nyeri pada area yang disuntik. KIPI bersifat sementara dan akan hilang dengan sendirinya tanpa pengobatan dan perawatan khusus[3].

Menurut Survey yang dilakukan oleh Kementerian kesehatan, masyarakat yang menolak program vaksinasi karena takut efek samping sebanyak 12%[4]. Hal ini menunjukkan bahwa adanya KIPI ini menjadi faktor penting yang harus dicari solusinya. Jika sebelum vaksin seseorang bisa diketahui akan mengalami KIPI maka dapat diantisipasi sebelumnya, sehingga resiko dapat diminimalisasi.

KIPI (Kejadian Ikutan Pasca Imunisasi) menurut WHO merupakan kejadian medis yang tidak diinginkan setelah seseorang melakukan imunisasi. Dimana efek tersebut tidak selalu memiliki hubungan kausal dengan penggunaan vaksin[5].

Untuk melakukan prediksi seseorang akan mengalami KIPI atau tidak dapat dilakukan dengan menerapkan algoritma matematika. Terdapat beberapa algoritma yang dapat digunakan, namun algoritma dengan akurasi tertinggi yang digunakan untuk mengolah data. Algoritma yang digunakan diantaranya algoritma J48, dimana algoritma tersebut memiliki tingkat akurasi yang tinggi[6].

Landasan Teori

1. Data Mining

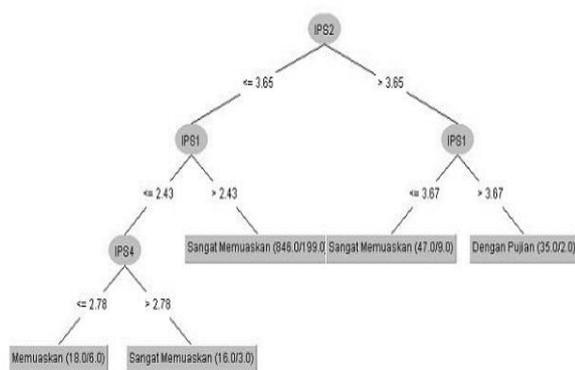
Data mining adalah suatu proses dari penemuan pengetahuan data. Proses dalam data mining dilakukan untuk menemukan informasi yang berarti dari suatu repositori. Dengan suatu proses dalam data mining, maka data dapat diolah untuk memperoleh berbagai informasi yang belum pernah ditemukan sebelumnya[6].

Berbagai penelitianpun telah dilakukan dengan menerapkan algoritma data mining. Terdapat

banyak sekali algoritma dalam Data Mining yang dapat diimplementasikan untuk pengolahan data. Pada penelitian kali ini, algoritma yang akan digunakan adalah *Decision Ttree* J48.

2. Algoritma J48

Decision Ttree menjadi salah satu algoritma yang digunakan dalam data mining. Algoritma J48 adalah implementasi Algoritma C4.5 pada software Weka. Algoritma tersebut dapat menghasilkan pohon keputusan.



Gambar 2. Contoh Pohon Keputusan [7]

Struktur pohon keputusan memiliki simpul akar, cabang dan daun. Simpul cabang tersebut mewakili kondisi atribut masukan, sementara cabang menentukan hasil kondisi, sedangkan daun memiliki label. Ketika digunakan untuk memprediksi kelulusan, algoritma J48 memiliki akurasi yang lebih baik dibandingkan dengan Logistic Model Tree [7].

Dalam sebuah penelitian, algoritma J48 terbukti memiliki akurasi yang tinggi untuk mendeteksi penyakit tiroid yaitu sebesar 99,64% [8]. Berdasarkan penelitian sebelumnya, maka penelitian ini menerapkan algoritma *Decision Tree* J48.

3. Bagging

Bagging merupakan metode ensemble umum yang dilakukan dengan cara memanipulasi data training. Data training diduplikasi sebanyak d kali dengan pengembalian yang akna menghasilkan sebanyak d data training yang baru. Kemudian dari data training akan dibentuk classifier yaitu bagged clasifier[9].

Teknik *bagging* merupakan metode untuk meningkatkan akurasi sebuah algoritma. Menurut Mordelet and Vert dalam kurniawan dan prihandono mengatakan kunci *bagging* dalam menaikkan akurasi adalah sampling with *replacement* karena dapat memperkecil varian dari dataset. Dalam penelitian ini *bagging* terbukti dapat meningkatkan akurasi algoritma naïve bayes, dari semula 77.00% menjadi 80,00% [10].

Metode

Data penelitian diperoleh dengan menyebarkan kueisoner menggunakan media online. Atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Atribut Penelitian

No	Atribut	Keterangan	Nilai
1	No_vaksin	Urutan vaksin	1. Vaksin I 2. Vaksin II
2	Jk	Jenis Kelamin	1. Laki-Laki 2. Perempuan
3	Usia	Kategori Kelompok Umur	1. Remaja 2. Dewasa 3. Lansia
4	TD	Tekanan Darah	1. Hipotensi 2. Normal 3. Hipertensi
5	Diabetes	Apakah responden memiliki riwayat diabetes	1. Ya 2. Tidak
6	Hipertensi	Apakah responden memiliki riwayat darah tinggi	1. Ya 2. Tidak
7	Jenisvaksin	Jenis Vaksin yang diterima Responden	1. Sinovac 2. Astra Zenneca 3. Moderna 4. Pfizer
8	KIPI	Apakah Mengalami KIPI (Label)	1. Ya 2. Tidak

Data yang masuk berjumlah 59, disajikan dalam bentuk CSV agar dapat diolah menggunakan aplikasi Weka. Model penelitian yang digunakan mengadopsi dari model CRIDPS-DM.



Gambar 3. Model Penelitian

Adapun detail dari model penelitian adalah sebagai berikut:

1. *Business Understanding*

Tahapan kegiatan untuk memahami tujuan dari penelitian data mining yang akan dilakukan.

2. *Data Understanding*

Merupakan tahapan untuk mengumpulkan data, memahami data yang akan digunakan dalam penelitian.

3. *Data Preparation*

Merupakan tahap persiapan data untuk dapat diproses lebih lanjut.

4. *Modelling*

Merupakan tahapan untuk menentukan teknik data mining yang digunakan, tools, algoritma yang digunakan serta parameternya.

5. *Evaluation*

Merupakan tahap interpretasi terhadap hasil data.[11].

Hasil dan Pembahasan

Implementasi algoritma dilakukan dengan menggunakan *Tools Weka 3.8.5*. Adapun proses dalam implementasinya adalah sebagai berikut:

Pengujian Awal

Pengujian dilakukan dengan menggunakan tools WEKA. Pengujian awal dilakukan dengan mengujikan data secara langsung dengan 8 atribut. Ketika seluruh atribut diujikan, hasil akurasi menunjukkan 43,86%. Pengujian selanjutnya dilakukan dengan menghapus atribut *No_vaksin* untuk berfokus pada tingkat akurasi yang lebih baik tanpa memperhatikan vaksin pertama atau kedua.

Algoritma J48 diimplementasikan pada data dengan 7 atribut yang disertakan yaitu jenis

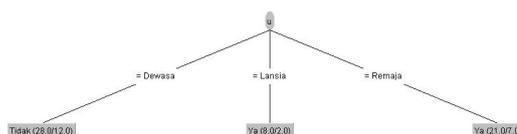
kelamin, usia, tekanan darah, diabetes, hipertensi, jenis vaksin dan KIPI. Hasil dari implementasi menunjukkan tingkat akurasi 75,44%.

Adapun confusion matrix pengujian awal terdapat pada Tabel 2.

Tabel 2. Confussion Matrix Pengujian Awal

		True Values	
		TRUE	FALSE
Prediction	TRUE	16	9
	FALSE	5	27

Pohon keputusan yang diperoleh dari pengujian awal terdapat pada gambar 3.



Gambar 4. Pohon Keputusan Pengujian Awal

Pengujian awal memberkan hasil yang kurang memuaskan. Oleh karena itu, dilakukan pengujian selanjutnya dengan menerapkan Teknik *resample*.

Pengujian Teknik *resample* dilakukan untuk melakukan perhitungan secara acak dan membagi data dalam training dan test set dengan proporsi tertentu. Setelah *resample* diterapkan, hasil dari implementasinya menunjukkan tingkat akurasi 85,96%.

Adapun confusion matrix pengujian dengan *resample* terdapat pada Tabel 3.

Tabel 3. Confussion Matrix Pengujian *Resample+J48*

		True Values	
		TRUE	FALSE
Prediction	TRUE	24	1
	FALSE	7	25

Tingkat akurasi dengan Teknik *resample* lebih baik jika dibandingkan dengan pengujian awal.

Teknik ensemble yang digunakan adalah *Bagging*. Implementasi dari teknik ini menunjukkan tingkat akurasi sebesar 91,22%. Adapun confusion matrix pengujian dengan teknik ensemble terdapat pada Tabel 4.

Tabel 4. Confusion Matrix Pengujian *Resample+J48+Bagging*

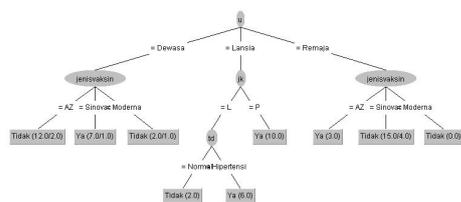
		True Values	
		TRUE	FALSE
Prediction	TRUE	24	1
	FALSE	4	28

Pengujian yang dilakukan menunjukkan adanya perbedaan akurasi. Akurasi dapat meningkat setelah diterapkan *resample* dan *bagging*. Evaluasi Hasil ditunjukkan dengan tabel 5.

Tabel 5. Evaluasi Hasil

	Precision	Recall	F-Measure	ROC Area	Accuracy
J48	0,755	0,754	0,751	0,776	75,44%
J48+Resample	0,879	0,862	0,86	0,912	85,96%
J48+Resample+Bagging	0,918	0,912	0,913	0,966	91,23%

Adapun hasil tree dari implementasi teknik tersebut dengan algoritma J48 terdapat pada Gambar 4.



Gambar 5. Pohon Keputusan Pengujian Akhir

Melihat hasil pengujian, tingkat *precision*, *recall*, *f-Measure*, *ROC Area* dan *Accuracy* paling tinggi Ketika pengujian menerapkan teknik *Resample+Bagging*.

Kesimpulan dan Saran

Penelitian dilakukan untuk memaksimalkan kinerja dari algoritma J48 yang diimplementasikan pada data KIPi Vaksinasi Covid-19. Penelitian dilakukan dengan melakukan seleksi atribut, *pre-processing* data (*resample*), dan penerapan teknik *ensemble*. Dari setiap langkah yang dilakukan dalam penelitian, dapat meningkatkan akurasi hingga 91,22%.

Penelitian selanjutnya dapat dilakukan dengan penggunaan jumlah data yang lebih banyak dan menggunakan teknik lain untuk peningkatan akurasi dari algoritma yang digunakan.

Referensi

- [1] I. N. H. Setiyo Adi Nugroho, "Efektivitas Dan Keamanan Vaksin Covid-19," J. Keperawatan, Vol. 9, p. 47, 2021.
- [2] D. Aghata, "Capaian Vaksinasi Covid-19 Ri Per 20 November 2021, Bertambah 1.423.305," Liputan6.Com, Jakarta, Nov. 2021.
- [3] Unicef, "Vaksin Covid-19 & Kipi," Unicef, 2021.
- [4] K. Ri, Itagi, Who, And Unicef, "Survei Penerimaan Vaksin Covid-19 Di Indonesia," Satuan Gugus Tugas Penanganan Covid-19, No. November, Pp. 1-26, 2020.
- [5] H. Hafizzanovian, D. Oktariana, M. A. Apriansyah, And Y. Yuniza, "Peluang Terjadinya Immunization Stress-Related Response (Isrr) Selama Program Vaksinasi Covid-19," J. Kedokt. Dan Kesehat. Publ. Ilm. Fak. Kedokt. Univ. Sriwij., Vol. 8, No. 3, Pp. 211-222, 2021, Doi: 10.32539/Jkk.v8i3.13807.
- [6] S. Fitri, "Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian , Lazy-Ibk , Zero-R , Dan Decision Tree- J48," Dasi, Vol. 15, No. 1, Pp. 33-37, 2014.
- [7] M. Defriani And I. Jaelani, "Algoritma J48 Dan Logistic Model Tree Untuk Memprediksi Predikat Kelulusan

Mahasiswa: Studi Kasus Stt Xyz,” Intecom
J. Inf. ..., Vol. 3, 2020.

- [8] S. Agustiani, A. Mustopa, A. Saryoko, W. Gata, And S. K. Wildah, “Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid,” *Paradig. - J. Komput. Dan Inform.*, Vol. 22, No. 2, Pp. 153-160, 2020, Doi: 10.31294/p.v22i2.8174.
- [9] N. Altman And M. Krzywinski, “Points Of Significance: Ensemble Methods: Bagging And Random Forests,” *Nat. Methods*, Vol. 14, No. 10, Pp. 933-934, 2017, Doi: 10.1038/Nmeth.4438.
- [10] A. Kurniawan And A. Prihandono, “Penerapan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Pada Algoritma Naive Bayes Dalam,” Vol. 1, Pp. 34-40, 2020.
- [11] D. Feblian And D. U. Daihani, “Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X,” *J. Tek. Ind.*, Vol. 6, No. 1, Pp. 1-12, 2017, Doi: 10.25105/Jti.v6i1.1526.
- [12] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>