

Penerapan Algoritma C4.5 Untuk Prediksi Efektivitas Pembelajaran Daring Selama Masa Pandemi Covid-19 Pada Siswa Sma Negeri Kota Bekasi

Bari Anandanu Wibowo¹, Iin Ernawati², Nurul Chamidah³

Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta^{1,2,3}

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, Jakarta, 12450, Indonesia

E-mail : bari.anandanu99@gmail.com, iinernawati@upnvj.ac.id, nurul.chamidah@upnvj.ac.id

Abstrak -- Masa pandemi covid-19 mengubah segala aspek kehidupan, salah satunya terhadap kegiatan belajar mengajar. Perubahan dalam kegiatan belajar mengajar juga merubah standar efektivitas dalam pembelajaran daring. Efektivitas dalam pembelajaran daring mempunyai beberapa variabel yang berbeda dengan pembelajaran tatap muka. Pada penelitian ini yang akan diteliti adalah memprediksikan efektivitas pembelajaran daring berdasarkan variabel yang tersedia dengan menggunakan metode klasifikasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah metode klasifikasi pohon keputusan algoritma C4.5. Pengumpulan data dilakukan dengan menyebarkan kuesioner kepada siswa sekolah menengah keatas di Kota Bekasi. Data yang sudah dikumpulkan akan dilakukan pra proses data dengan pembersihan dan seleksi data. Data kemudian diterapkan proses algoritma C4.5 dengan melatih dan menguji data. Hasil yang didapatkan dalam penelitian ini adalah dalam penggunaan algoritma C4.5 ditemukan bahwa semua variabel mempengaruhi efektivitas pembelajaran daring dan performa algoritma C4.5 mendapatkan rata-rata akurasi sebesar 69% setelah melakukan evaluasi dengan metode k-fold cross validation dengan nilai k sebesar 4.

Kata Kunci: Efektivitas, Algoritma C4.5, Pembelajaran Daring

Abstract -- The COVID-19 pandemic has changed all aspects of life, one of which is teaching and learning activities. Changes in teaching and learning activities also change the standard of effectiveness in online learning. Effectiveness in online learning has several different variables from face-to-face learning. In this study, what will be investigated is predicting the effectiveness of online learning based on available variables using the classification method. The classification method used in this study is the C4.5 algorithm decision tree classification method. Data was collected by distributing questionnaires to high school students and seniors in Bekasi City. The data that has been collected will be pre-processed with data cleaning and selection. The data is then applied to the C4.5 algorithm process by training and testing the data. The results obtained in this study are in the use of the C4.5 algorithm, it was found that all variables affect the effectiveness of online learning and the performance of the C4.5 algorithm, getting an average accuracy of 69% after evaluating with the k-fold cross validation method with a k value of 4.

Keywords: Effectiveness, C4.5 Algorithm, Online Learning

I. PENDAHULUAN

Pandemi Covid-19 mengubah segala aspek kehidupan, salah satunya adalah kegiatan belajar para siswa. Dengan diadakannya pembatasan sosial berskala besar maka ketentuan pemerintah dalam melaksanakan kegiatan belajar mengajar menumbuhkan sebutan baru yakni pembelajaran jarak jauh. Pembelajaran jarak jauh ini merupakan kegiatan belajar mengajar antara siswa dan guru yang dilakukan bukan dengan tatap muka melainkan secara daring, baik menggunakan fasilitas daring berupa media sosial maupun aplikasi *e-learning*.

Suatu pembelajaran daring berbeda dengan pembelajaran tatap muka, dalam satu sisi tenaga pengajar tidak akan mengetahui secara langsung tentang pemahaman seorang pelajar terkait materi yang diberikannya dan di sisi lain seorang pelajar jika ingin bertanya lebih lanjut terkait materi pelajaran yang telah diterimanya. Dengan hilangnya salah satu fungsi antara pengajar dan pelajar tersebut maka pelaksanaan pembelajaran yang akan terjadi bisa berubah dalam segi efektivitas. Efektivitas merupakan hasil dari suatu usaha yang telah dilakukan.

Algoritma C4.5 merupakan salah satu algoritma dari metode klasifikasi pohon keputusan. Algoritma ini menggunakan model pembelajaran berupa pohon keputusan yang bisa aplikasikan untuk memprediksi data yang akan diketahui jenis klasifikasinya. Algoritma ini juga membutuhkan beberapa variabel yang mendukung untuk bisa dilakukan prediksi. Dengan perubahan kegiatan belajar mengajar dari tatap muka menjadi secara daring menyebabkan perubahan dari segi efektivitasnya, maka dibutuhkan cara untuk mengetahui hal-hal yang berpengaruh dalam efektivitas untuk pembelajaran daring di masa pandemi Covid-19. Oleh karena itu penerapan

algoritma C4.5 merupakan salah satu solusi untuk memprediksikan efektivitas pembelajaran daring yang berguna untuk menentukan hal-hal yang berpengaruh tersebut.

Dalam jurnal [1] menganalisis efektivitas berdasarkan beberapa hal, salah satunya yakni kesiapan pelajar dalam melakukan kegiatan belajar mengajar, penggunaan media pembelajaran daring yang bervariasi, dan kemudahan akses media. Dalam jurnal hasil penelitian ini menunjukkan bahwa adanya dukungan positif untuk dilaksanakannya pembelajaran daring dimasa yang akan datang.

Dalam jurnal [2] menjelaskan bahwa hasil analisa pengisian angket pelaksanaan BDR kelas VI SD di Muhammadiyah 18 Surabaya menjelaskan beberapa indikator efektivitas pembelajaran daring memiliki hasil rata-rata dari perhitungan angket yang telah dilakukan. Beberapa indikator tersebut adalah efektivitas, akses, pendampingan, reaksi tugas, kesulitan materi, materi, media, evaluasi, dan kejelasan instruksi. Masing masing indikator memiliki kualifikasi efektivitasnya masing masing, yang diperoleh dari analisis hasil pengisian angket sebelumnya. Setelah melakukan penelitian tersebut terdapat sebuah kesimpulan berupa pelaksanaan belajar dari rumah untuk siswa kelas VI SD Muhammadiyah 18 Surabaya berlangsung cukup efektif.

Dalam jurnal [3] tertulis bahwa ada beberapa indikator efektivitas pembelajaran jarak jauh yakni konten(isi) pembelajaran, sumber belajar berbasis teknologi, adanya media untuk melakukan pembelajaran praktikum, penggunaan teknologi informasi dan komunikasi, dan memudahkan untuk melakukan evaluasi belajar secara mandiri. Ada juga indikator efektivitas interaksi dalam pembelajaran jarak jauh yakni perilaku dosen saat mengajar, interaksi, antara dosen dan taruna, dan durasi pembelajaran. Serta ada indikator efektivitas pemahaman taruna dalam pembelajaran jarak jauh yakni kemandirian belajar taruna, motivasi belajar taruna, dan keaktifan taruna dalam pembelajaran seperti menyimak, melihat, mendengar, membaca, bertanya, menyimpulkan, dan mengkomunikasikan. Semua indikator diatas didapatkan berdasarkan gabungan dari pendapat Ma'mur dan disesuaikan dengan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 109 Tahun 2013 tentang Penyelenggaraan Pendidikan Jarak Jauh Pada Pendidikan Tinggi. Hasil yang didapatkan dari jurnal ini adalah setelah dilakukan analisis pembelajaran jarak jauh yang dilakukan di politeknik penerbangan Surabaya bisa dikatakan sebagai pembelajaran yang efektif.

Dalam jurnal [4] menghasilkan akurasi sebesar 78,75% dengan menggunakan algoritma C4.5 dalam memprediksikan kompetensi karyawan. Penulis mengembangkan metode algoritma c4.5 dengan framework RapidMiner. Kemudian untuk perhitungan akurasi data training dan testing menggunakan Confusion Matrix dan ROC/AUC (Area Under Cover). Evaluasi yang didapatkan jika menggunakan Confusion Matrix akurasi dari data training sebesar 78.64% dan data testing sebesar 56.00% sedangkan menggunakan ROC/AUC akurasi dari data training sebesar 0.907 dan data testing sebesar 0.783. Penulis juga menyimpulkan dengan menerapkan metode pohon keputusan dapat mempercepat dan tingkat keakuratan dalam pengambilan keputusan.

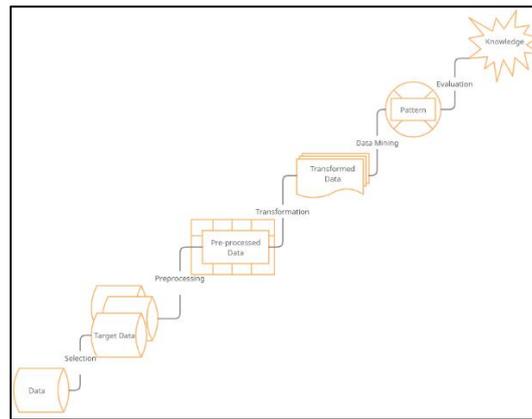
Berdasarkan beberapa penelitian tersebut bisa diambil beberapa indikator yang bisa digunakan untuk efektivitas pembelajaran daring yakni variabel media online dari [1], variabel pemahaman materi, kondisi materi, reaksi tugas, pengumpulan tugas, kejelasan tugas, dan pendamping [2], dan variabel ekspresi guru, keaktifan belajar, belajar mandiri, interaksi, durasi pembelajaran, pengerjaan materi, dan waktu pengerjaan dari [3]. Serta menggunakan algoritma C4.5 sebagai algoritma pohon keputusan dalam proses klasifikasi data karena penggunaan nya dalam [4] menghasilkan akurasi sebesar 78,75%.

II.METODOLOGI PENELITIAN

Berdasarkan [5] pembelajaran daring didefinisikan sebagai proses pembelajaran yang memudahkan pelajar dalam memanfaatkan internet yang terhubung kepada perangkat komputer. Sehingga pelajar dapat mempelajari ilmu baru melalui visualisasi yang merupakan salah satu hal yang penting dalam pembelajaran. Berdasarkan penjelasan diatas bisa ditarik kesimpulan bahwa pembelajaran daring merupakan suatu kegiatan belajar mengajar antara pelajar dan pengajar dengan menggunakan perangkat komputer yang dapat terhubung menggunakan internet.

Berdasarkan [6] efektivitas pembelajaran bisa ditentukan oleh empat indikator yaitu kesesuaian tingkat pembelajaran, kualitas pembelajaran, waktu, dan insentif. Kesesuaian tingkat pembelajaran adalah sejauh mana pengajar memastikan tingkat kesiapan pelajar dalam mempelajari materi baru. Kualitas pembelajaran adalah seberapa besar informasi yang disajikan dengan mudah dapat dipelajari oleh siswa atau meminimalisir tingkat kesalahannya. Semakin kecil tingkat kesalahan yang dilakukan oleh siswa maka semakin efektif pembelajaran yang telah dilakukan. Waktu adalah lamanya waktu yang ditentukan untuk siswa dalam mempelajari materi yang telah diberikan. Pembelajaran dikatakan efektif jika siswa dapat menyelesaikan tugas yang diberikan sesuai dengan waktu yang ditentukan. Insentif adalah seberapa besar kemampuan pengajar untuk memotivasi pelajar dalam mengerjakan tugas-tugas serta mempelajari materi yang telah diberikan. Semakin besar motivasi yang diberikan maka akan semakin besar keaktifan pelajar.

Berdasarkan [7] *data mining* didefinisikan sebagai suatu proses penemuan pola atau pengetahuan bermanfaat dengan sendirinya atau otomatis yang didapatkan dari sekumpulan data yang memiliki jumlah yang besar. *Data mining* bisa dianggap bagian dari *Knowledge Discovery in Database (KDD)*.



Gambar 1. Knowledge Data Discovery (KDD)

Berdasarkan Gambar 1 bisa disimpulkan bahwa *data mining* merupakan suatu teknik penggalian informasi berharga yang tersembunyi dalam koleksi data yang berukuran besar sehingga ditemukan pola atau aturan menarik yang belum diketahui dengan tahapan seleksi data, pembersihan data, transformasi data, dan evaluasi pola.

Berdasarkan [8] algoritma C4.5 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini berupa dapat mengolah data diskrit dan numerik, menghasilkan aturan - aturan yang mudah diinterpretasikan, dapat menangani nilai atribut yang hilang dan tercepat diantara algoritma-algoritma yang lain. Algoritma C4.5 merupakan salah satu algoritma pohon keputusan atau *decision tree*. Algoritma ini juga merupakan versi perbaikan dari algoritma ID3, perbaikan yang telah dilakukan berupa dapat menghilangkan missing data, mengatasi pruning, serta dapat mengolah data yang bertipe kategorikal dan numerik.

Berdasarkan [9] *k-fold cross validation* merupakan salah satu teknik validasi pada dataset untuk menemukan nilai akurasi yang baik. Kemudian berdasarkan [10], metode *cross validation* ini memastikan bahwa setiap data memiliki kesempatan yang sama untuk menjadi data latih dan data uji. Dengan menggunakan nilai $k=10$ data dibagi menjadi berbagai partisi yang dinamakan sebagai *fold*. Dari masing *fold* memiliki jumlah data yang sama atau mendekati sama. Selama k iterasi, dipilih salah satu *fold* sebagai data uji, sedangkan sisanya digunakan sebagai data latih.

Berdasarkan [11], *confusion matrix* didefinisikan sebagai alat yang berguna untuk menganalisis seberapa bagus algoritma klasifikasi dalam mengenali atribut dari kelas yang berbeda. Kemudian berdasarkan [12] menjelaskan bahwa evaluasi dilakukan dengan pengukuran akurasi, presisi, dan *recall* yang kemudian dimasukkan kedalam tabel klasifikasi bersifat prediktif atau disebut juga dengan *confusion matrix*.

Tabel 1. Confusion Matrix

Confusion Matrix		Aktual	
		Efektif	Tidak Efektif
Prediksi	Efektif	TP	FP
	Tidak Efektif	FN	TN

Dilihat dari Tabel 1 terdapat TP (*True positive*) yang berarti jumlah data dengan kelas “Efektif” diprediksikan sebagai kelas “Efektif”, FN (*False Negative*) yang berarti jumlah data dengan kelas “Efektif” diprediksikan sebagai kelas “Tidak Efektif”, FP (*False Positive*) yang berarti jumlah data dengan kelas “Tidak Efektif” diprediksikan sebagai kelas “Efektif”, dan TN (*True Negative*) yang berarti jumlah data dengan kelas “Tidak Efektif” diprediksikan sebagai kelas “Tidak Efektif”.

Pertanyaan-pertanyaan kuesioner yang digunakan dalam penelitian ini mengambil referensi dari beberapa jurnal penelitian yang membuat pertanyaan terkait efektivitas pembelajaran daring. Penyusunan pertanyaan tersebut didasarkan oleh pertanyaan yang diajukan dari jurnal efektivitas pembelajaran daring pada penelitian sebelumnya [1][2][3]. Sehingga variabel yang digunakan untuk penyusunan pertanyaan adalah media online, ekspresi guru, kejelasan tugas, reaksi tugas, keaktifan belajar, belajar mandiri, interaksi, pendamping, pengumpulan tugas, durasi pembelajaran, pemahaman materi, kondisi materi, pengerjaan materi, waktu pengerjaan, dan efektivitas. Dengan pertanyaan sebagai berikut:

Kelas: 10 / 11

Media Online

1. Bagaimana penggunaan media online yang anda gunakan selama masa pembelajaran daring?
 - a. Bervariasi (menggunakan beberapa media)
 - b. Terbatas (hanya menggunakan salah satu)

Ekspresi Guru

2. Bagaimana ekspresi guru saat mengajar selama masa pembelajaran daring?
 - a. Merasa menyenangkan

- b. Merasa terbebani

Kejelasan Tugas

- 3. Bagaimana kejelasan tugas yang telah diberikan selama masa pembelajaran daring?
 - a. Sangat jelas
 - b. Kurang jelas

Pengerjaan Tugas

- 4. Bagaimana reaksi anda terhadap tugas yang telah diberikan selama masa pembelajaran daring?
 - a. Langsung mengerjakan
 - b. Menunda mengerjakan

Keaktifan Belajar

- 5. Bagaimana keaktifan anda dalam menyimak, melihat, mendengar, membaca, bertanya, menyimpulkan dan mengkomunikasikan selama masa pembelajaran daring?
 - a. Selalu aktif
 - b. Sering aktif
 - c. Jarang aktif
 - d. Tidak aktif

Belajar Mandiri

- 6. Bagaimana motivasi belajar mandiri anda selama masa pembelajaran daring?
 - a. Merasa meningkat
 - b. Merasa terbebani

Interaksi Guru

- 7. Bagaimana komunikasi antara guru dengan anda selama masa pembelajaran daring?
 - a. Sering bertanya
 - b. Sering menyimak
 - c. Jarang bertanya atau menyimak

Pendamping

- 8. Selama masa pembelajaran anda sering didampingi oleh siapa?
 - a. Keluarga
 - b. Sendiri
 - c. Teman (melalui media sosial whatsapp/ line)

Pengumpulan Tugas

- 9. Bagaimana kondisi pengumpulan tugas yang anda kerjakan?
 - a. Sebelum *deadline*
 - b. Mendekati *deadline*
 - c. Melewati *deadline*

Durasi Pembelajaran

- 10. Menurut anda, bagaimana durasi pembelajaran yang terjadi selama masa pembelajaran daring?
 - a. Lama, sehingga saya tidak punya waktu sendiri
 - b. Cukup, sehingga saya mampu menyeimbangkan antara belajar dan istirahat
 - c. Cepat, sehingga saya punya banyak waktu sendiri

Pemahaman Materi

- 11. Bagaimana pemahaman materi yang diberikan kepada anda selama masa pembelajaran?
 - a. Mudah dipahami
 - b. Sulit dipahami

Kondisi Materi

- 12. Bagaimana kondisi materi yang diberikan kepada anda selama masa pembelajaran?
 - a. Sesuai harapan
 - b. Tidak sesuai harapan

Pengerjaan Materi

- 13. Bagaimana proses pemahaman terhadap materi yang diberikan selama masa pembelajaran daring?
 - a. Bisa secara mandiri
 - b. Butuh bantuan orang lain

Waktu Pengerjaan

- 14. Bagaimana waktu yang diberikan untuk mengerjakan tugas?
 - a. Cukup untuk mengerjakan
 - b. Kurang untuk mengerjakan

Efektivitas

- 15. Menurut anda apakah pembelajaran daring yang telah dilaksanakan efektif?
 - a. Efektif
 - b. Tidak Efektif

Tabel 2. Nama Variabel

Nama Variabel	Variabel Pengganti
Media Online	V1
Ekspresi Guru	V2
Kejelasan Tugas	V3
Reaksi Tugas	V4
Keaktifan Belajar	V5
Belajar Mandiri	V6
Interaksi	V7
Pendamping	V8
Pengumpulan Tugas	V9
Durasi Pembelajaran	V10
Pemahaman Materi	V11
Kondisi Materi	V12
Pengerjaan Materi	V13
Waktu Pengerjaan	V14
Efektivitas	Label

Data pada penelitian ini diperoleh dengan menyebarkan pertanyaan kuesioner kepada siswa SMA Negeri 10 Kota Bekasi. Pertanyaan yang diajukan berjumlah 16 pertanyaan dengan 2 pertanyaan (kelas dan jenis sekolah) dan 14 pertanyaan yang mewakili setiap variabel dari Tabel 2. Data yang sudah dikumpulkan berjumlah 159 responden dengan keterangan 61 Responden dari kelas 11 dan 98 responden dari kelas 10. Dari total 159 data terdapat 90 data yang berlabel sebagai “Efektif” dan 69 data yang berlabel sebagai “Tidak Efektif”.

Seleksi variabel dilakukan dengan menghapus variabel Timestamps, Kelas dan Jenis Sekolah, penghapusan variabel Timestamps dikarenakan variabel tersebut hanya menunjukkan keterangan waktu data diperoleh, penghapusan variabel Kelas dan Jenis sekolah dikarenakan variabel tersebut merupakan tanda bahwa data yang diperoleh berasal dari siswa dengan jenis SMA Negeri dan dengan kelas 10 atau 11 yang merupakan target responden penelitian ini dan pemilihan variabel yang digunakan dalam pembuatan model. hasil seleksi variabel dari data yang digunakan dengan menghasilkan 14 variabel independen dan 1 variabel dependen.

Untuk pelatihan model dalam penelitian ini digunakan algoritma C4.5. Pada dasarnya metode algoritma C4.5 memiliki 4 tahapan untuk membangun suatu pohon keputusan atau *decision tree* yang dijelaskan sebagai berikut:

1. Tahapan pertama dilakukan dengan memilih variabel yang akan menjadi akar.
2. Tahapan kedua dilakukan dengan membuat cabang untuk setiap nilai nilai dari variabel yang terpilih menjadi akar.
3. Tahapan ketiga dilakukan dengan membagi kasus keputusan kedalam cabang pohon keputusan.
4. Tahapan keempat dilakukan dengan mengulangi tahapan pertama sampai ketiga untuk setiap cabang keputusan sampai semua kasus keputusan didalam cabang pohon keputusan memiliki kelas yang sama.

Untuk tahapan pertama dalam memulai pembuatan pohon keputusan yang menggunakan algoritma C4.5 diperlukan node akar yang dapat diperoleh dengan mencari nilai gain ratio tertinggi dari setiap variabel. Untuk memperoleh nilai gain ratio ini dibutuhkan perhitungan *entropy*, *gain*, *split info*, dan *gain ratio* dalam menentukannya. Berikut rumus *entropy*, *gain*, *split info*, dan *gain ratio* yang digunakan:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

$$Split Info(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

$$Gain Ratio(A) = \frac{Gain(A)}{Split Info(A)} \quad (4)$$

Keterangan:

S: kelompok kasus

A: variabel

n: jumlah partisi variabel A

pi: proporsi dari Si terhadap S

Si: jumlah kasus pada partisi ke -i

Gain(A): Nilai Gain pada variabel A

Split Info(A): Nilai Split Info pada variabel A

III. HASIL DAN PEMBAHASAN

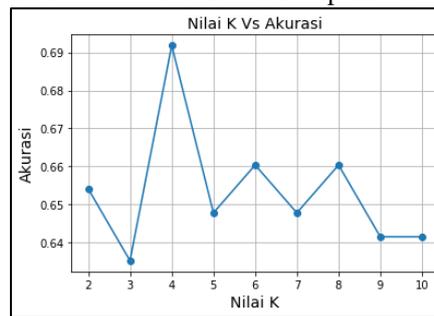
Maka hasil perhitungan yang dilakukan kepada data menggunakan persamaan (1),(2),(3),(4) untuk pencarian node akar adalah sebagai berikut:

Tabel 3. Hasil Perhitungan Algoritma C4.5

Nama Variabel	Nilai Variabel	Efektif	Tidak Efektif	Entropy	Gain	Split Info	Gain Ratio
V1	Bervariasi	63	80	0,989	0,001	0,470	0,002
	Terbatas	6	10	0,954			
V2	Merasa senang	57	53	0,999	0,048	0,891	0,054
	Merasa terbebani	12	37	0,803			
V3	Kurang jelas	25	51	0,913	0,029	0,998	0,029
	Sangat jelas	44	39	0,997			
V4	Langsung mengerjakan	53	42	0,990	0,069	0,972	0,071
	Menunda mengerjakan	16	48	0,811			
V5	Selalu aktif	11	11	1	0,033	1,505	0,022
	Sering aktif	40	39	0,999			
	Jarang aktif	18	38	0,905			
	Tidak aktif	0	2	0			
V6	Merasa meningkat	40	20	0,918	0,097	0,956	0,102
	Merasa terbebani	29	70	0,872			
V7	Sering bertanya	8	6	0,985	0,017	1,075	0,016
	Sering menyimak	53	64	0,993			
	Jarang bertanya atau menyimak	8	20	0,863			
V8	Keluarga	6	9	0,971	0,0006	1,35	0,0005
	Sendiri	31	38	0,992			
	Teman	32	43	0,984			
V9	Sebelum deadline	36	41	0,996	0,018	1,390	0,013
	Mendekati deadline	29	35	0,993			
	Melewati deadline	4	14	0,764			
V10	Cepat	8	9	0,997	0,002	0,905	0,002
	Cukup	56	72	0,988			
	Lama	5	9	0,940			
V11	Mudah dipahami	31	12	0,854	0,090	0,842	0,107
	Sulit dipahami	38	78	0,912			
V12	Sesuai harapan	45	29	0,966	0,078	0,996	0,079
	Tidak sesuai harapan	24	61	0,858			
V13	Mandiri	33	24	0,981	0,034	0,941	0,036
	Bantuan	36	66	0,936			
V14	Cukup	57	64	0,997	0,013	0,793	0,016
	Kurang	12	26	0,899			

Berdasarkan Tabel 3 diketahui bahwa variabel dengan nilai *Gain Ratio* tertinggi adalah variabel V11 (Pemahaman Materi) dengan nilai sebesar 0.107701. Hal tersebut membuat variabel V11 digunakan sebagai node akar. Nilai variabel dari V11 ada 2 yakni Mudah dan Sulit. Dari kedua nilai variabel tersebut nilai variabel Mudah perlu dilakukan perhitungan lebih lanjut karena masih memiliki 2 nilai kelas, dan untuk nilai variabel Sulit masih perlu memiliki perhitungan lebih lanjut juga karena memiliki 2 nilai kelas. Perhitungan lebih lanjut menunjukkan bahwa pohon keputusan yang terbentuk menghasilkan 95 rules keputusan antara kelas “Efektif” dan “Tidak Efektif”. Dengan salah satu contoh rules yang terbentuk adalah “IF (Pemahaman Materi = Mudah Dipahami) && (Durasi Pembelajaran = Cukup) && (Pengerjaan Tugas = Menunda Mengerjakan) && (Kejelasan Tugas = Kurang Jelas) THEN (Label = Tidak Efektif)”. Hasil pohon keputusan tersebut telah terbentuk dengan bantuan library *Chefboost* dengan konfigurasi algoritma C4.5 pada pemrograman *python* menunjukkan bahwa variabel yang berpengaruh dalam pembentukan pohon keputusan adalah semua variabel.

Evaluasi model dengan pemrograman *python* menggunakan *library Chefboost* dan melakukan *K-fold Cross Validation* dengan nilai $k=4$. Penggunaan nilai $k=4$ didapat dari nilai rata-rata akurasi tertinggi dari percobaan nilai $k=2$ sampai dengan $k=10$. Berikut hasil rata-rata akurasi dari setiap nilai k :



Gambar 2 Hasil Nilai K vs Akurasi

Berdasarkan Gambar 2 ditunjukkan bahwa rata-rata akurasi pada nilai $k=4$ adalah nilai rata-rata tertinggi dari hasil percobaan sebesar 0,6925, alasan mengapa nilai $k=4$ tertinggi dengan nilai k yang lain adalah karena dengan nilai k yang lain data latihan yang digunakan terlalu bervariasi dengan sedikit jumlah kasus data yang sama sehingga saat pengujian data nilai akurasi yang didapat bisa sangat tinggi disatu bagian dan sangat rendah di bagian lainnya, maka penggunaan nilai $k=4$ digunakan pada evaluasi model. Kemudian didapatkan nilai Confusion Matrix dari 4-fold cross validation sebagai berikut:

Tabel 4. Hasil Confusion Matrix

	<i>Fold ke-1</i>	<i>Fold ke-2</i>	<i>Fold ke-3</i>	<i>Fold ke-4</i>	Rata-Rata
Akurasi	0,62	0,70	0,60	0,85	0,6925
Presisi	0,59	0,69	0,53	0,82	0,6575
Recall	0,56	0,53	0,53	0,82	0,61

Dilihat dari nilai rata-rata akurasi, presisi dan *recall* bisa disimpulkan bahwa penggunaan algoritma C4.5 dan evaluasi yang menggunakan metode *k-fold cross validation* dengan nilai $k=4$ terhadap dataset ini mempunyai hasil yang bervariasi, hal ini didasarkan dengan melihat salah satu cabang pohon keputusan yang terbentuk hanya mendukung sedikit kasus data. Sehingga dalam pengujian model didapatkan akurasi yang bervariasi, sesuai dengan Tabel 4 dengan akurasi terendah pada *fold ke-3* sebesar 0,60 atau sebesar 60% dan akurasi tertinggi pada *fold ke-4* sebesar 0,85 atau sebesar 85%. Walau menggunakan semua variabel dapat dilihat bahwa nilai rata-rata akurasi masih sebesar 69%, itu karena dengan penggunaan banyak variabel dan setiap kasus data yang ada hanya memiliki sedikit kasus data memungkinkan pelabelan data banyak yang tidak sesuai.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka bisa disimpulkan beberapa hal sebagai berikut:

1. Penerapan algoritma C4.5 untuk memprediksikan efektivitas pembelajaran daring siswa SMA Negeri 10 di Kota Bekasi menggunakan variabel media online, ekspresi guru, kejelasan tugas, reaksi tugas, keaktifan belajar, belajar mandiri, interaksi, pendamping, pengumpulan tugas, durasi pembelajaran, pemahaman materi, kondisi materi, pengerjaan materi, waktu pengerjaan.
2. Berdasarkan hasil penerapan algoritma C4.5 terhadap data klasifikasi efektivitas pembelajaran daring siswa SMA Negeri di Kota Bekasi ditemukan bahwa variabel yang berpengaruh dalam efektivitas pembelajaran daring adalah seluruh variabel.
3. Berdasarkan penggunaan algoritma C4.5 dalam klasifikasi dataset pohon keputusan yang terbentuk menghasilkan total 95 rules.
4. Performa model prediksi efektivitas pembelajaran daring dengan menggunakan metode *k-fold cross validation* dengan nilai k sebesar 4 menghasilkan nilai rata-rata akurasi sebesar 0,6925 atau 69%, nilai rata-rata presisi sebesar 0,6575 atau 66%, dan nilai rata-rata *recall* sebesar 0,61 atau 61% yang menunjukkan bahwa penggunaan algoritma C4.5 dalam klasifikasi data efektivitas pembelajaran daring siswa kurang akurat.

Berdasarkan penelitian yang telah dilakukan maka penulis menyarankan beberapa hal sebagai berikut:

1. Dengan menambah variabel efektivitas pembelajaran daring lain karena pohon keputusan yang terbentuk diketahui menggunakan semua variabel dataset yang menunjukkan pengaruh dalam prediksi efektivitas pembelajaran daring.
2. Menggunakan algoritma klasifikasi lain seperti *ID3*, *CART*, atau *Naïve Bayes Classifier* dalam klasifikasi dataset karena nilai akurasi, *recall*, dan presisi yang sudah didapat kurang bagus.

V. DAFTAR PUSTAKA

- [1] Simatupang, Nova Irawati, Sri Rejeki Ignasia Sitohang, Angelia Patricia Situmorang, & Ismenny Martalena Simatupang. 2020. *Efektivitas Pelaksanaan Pengajaran Online Pada Masa Pandemi Covid-19 Dengan Metode Survey Sederhana*, Jurnal Dinamika Pendidikan. Vol.13, No.2.
- [2] Sari, Dewi Ratna, & Fairuza Amrozi. 2020. *Analisis Efektivitas Pembelajaran Jarak Jauh (Pjj) Di Politeknik Penerbangan Surabaya*. Jurnal Penelitian Politeknik Penerbangan Surabaya, Edisi XXVIII, Vol 5, No 2.
- [3] Kurniasari, Asrilia, Fitroh Setyo Putro Pribowo, & Deni Adi Putra. 2020. *Analisis Efektivitas Pelaksanaan Belajar Dari Rumah (Bdr) Selama Pandemi Covid-19*. Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan dan Hasil Penelitian, Vol 6, No 3.
- [4] Sunge, Aswan Supriyadi. 2018. *Prediksi Kompetensi Karyawan Menggunakan Algoritma C4.5 (Studi Kasus: Pt Hankook Tire Indonesia)*. Seminar Nasional Teknologi Informasi dan Komunikasi 2018
- [5] Abdullah, Walib. 2018. *Model Blended Learning Dalam Meningkatkan Efektifitas Pembelajaran*. IAI Al - Khairat Pamekasan: Jurnal Pendidikan dan Manajemen Islam, Vol. 7. No.1.
- [6] Slavin, E. Robert. 2009. *Cooperative Learning: Teori, Riset, Dan Praktik*. Alih Bahasa: Nurulita, Nusa Media, Bandung.
- [7] Rahmayuni, Indri. 2014. *Perbandingan Performansi Algoritma C4.5 dan Cart dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang*. Jurnal TEKNOIF, Vol.2 No.1.
- [8] Widodo, Prabowo Pudjo., Handayanto, Rahmadya Trias., & Herlawati. 2013. *Penerapan Data Mining Dengan Matlab*. Penerbit Rekayasa Sains, Bandung.
- [9] Haltuf, M.. 2011. *University of Economics in Prague*. Faculty of Business Administration.
- [10] Brownlee, J.. 2016. *Master Machine Learning Algorithms: Discover How They Work And Implement Them From Scratch*. Jason Brownlee.
- [11] Han, J., M. Kamber, & J. Pei. 2011. *Data Mining: Concepts And Techniques*. Elsevier.
- [12] Xhemali, D., Hinde, C. J., & Stone, R. G.. 2009. *Naive Bayes Vs. Decision Trees Vs. Neural Networks In The Classification Of Training Web Pages*. IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1.