




INNOVATIVE: Journal Of Social Science Research

Volume 3 Nomor 2 Tahun 2023 Page 13880-13892

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma

Eri Mardiani¹, Nur Rahmansyah², Sari Ningsih³, Dhieka Avrilia Lantana⁴, Adisti Suryaningtyas Putri Wirawan⁵, Sifonne Adi Wijaya⁶, Dinda Nurkhaliza Putri⁷

(1) Program Studi Informatika, FTKE, Universitas Nasional, Indonesia

(2) Program Studi Animasi Politeknik Negeri Media Kreatif, Indonesia

(3) Program Studi Sistem Informasi, FTKE, Universitas Nasional, Indonesia,

(4) Program Studi Bisnis Digital, FEB, Universitas Nasional, Indonesia

(5) (6)(7)Program Studi Akuntansi, FEB, UPN Veteran Jakarta, Indonesia

Email : erimardiani1@gmail.com¹

Abstrak

Perkembangan teknologi informasi semakin pesat hingga saat ini terus dikembangkan teknologi-teknologi terbaru, penggunaan data sebagai pengolahan data kini semakin banyak diterapkan pada berbagai bidang, dengan penggunaan data untuk mengetahui kondisi suatu bagian atau institusi sangat berpengaruh terhadap kinerja dan kualitas dari suatu bagian atau institusi tersebut, dengan penggunaan data dapat dilakukan perkembangan teknologi yang dapat meningkatkan kualitas dan efisiensi yang dilakukan terhadap perusahaan, pengolahan data dapat diterapkan di semua bidang dan salah satunya adalah di bidang pendidikan yang cocok menerapkan teknik data mining untuk mengolah maupun menganalisis data-data pendidikan. Tujuan dari riset ini dengan menggunakan tools aplikasi Orange Data Mining untuk membantu orang tua dan pelajar untuk mengetahui kinerja dan nilai siswa, untuk mengetahui performa pelajar SMA mulai dari nilai dan informasi demografis mereka. Data dikumpulkan dari tiga SMA dan juga untuk mengembangkan model pembelajaran mesin untuk memprediksi kinerja siswa berdasarkan demografi dan faktor lainnya, sehingga kita dapat mengetahui kondisi pelajar saat ini..

Kata Kunci: *Data mining, Orange, pendidikan, pelajar SMA.*

Abstract

The development of information technology is increasing rapidly until now the latest technologies are being developed, the use of data as data processing is now increasingly being applied to various fields, with the use of data to determine the condition of a section or institution which greatly influences the performance and quality of a section or institution. , with the use of data technological developments can be carried out which can improve the quality and efficiency carried out by companies, data processing can be applied in all fields and one of them is in the field of education which is suitable for applying data mining techniques to process and analyze educational data. The purpose of this research is to use the Orange Data Mining application tools to help parents and students to find out student performance and grades, to find out the performance of high school students starting from their grades and demographic information. The data was collected from three high schools and also to develop a machine learning model to predict student performance based on demographics and other factors, so that we can know the current state of students

Keyword: *data mining, orange, education, high school students*

PENDAHULUAN

Perkembangan teknologi informasi sudah sedemikian pesat salah satunya adalah Internet. Teknologi Internet menghubungkan ribuan jaringan komputer individual dan organisasi di seluruh dunia.

Ada banyak kemajuan teknologi yang tersedia dengan komputer. Komputer tidak hanya membantu kita dalam menyelesaikan pekerjaan, tetapi juga dapat menjadi alat yang menyenangkan dan berguna untuk digunakan. Untuk lebih memaksimalkan pekerjaan teknologi berbasis internet sangat mendukung pekerjaan secara online. Selain itu, dengan teknologi ineternet pekerjaan menjadi jauh lebih cepat dan lebih mudah selesai daripada penggunaan komputer tradisional. Perkembangan teknologi informasi sudah sedemikian pesat. Teknologi Internet menghubungkan ribuan jaringan komputer individual dan organisasi di seluruh dunia.

Saat ini pendidikan merupakan modal utama bagi suatu bangsa dalam upaya meningkatkan kualitas sumber daya manusia yang dimilikinya. Sumber daya manusia yang berkualitas akan mampu mengelola sumber daya alam dan memeberi layanan secara efektif dan efisien untuk meningkatkan kesejahteraan masyarakat. Oleh karena itu, hampir semua bangsa berusaha untuk meningkatkan kualitas pendidikan yang dimilikinya, termasuk indonesia.

Setidaknya ada enam alasan mengapa teknologi Internet begitu populer. Keenam alasan tersebut adalah Internet memiliki konektivitas dan jangkauan yang luas; dapat mengurangi biaya komunikasi; biaya transaksi yang lebih rendah; dapat mengurangi biaya *agency*; interaktif, fleksibel, dan mudah; serta memiliki kemampuan untuk mendistribusikan pengetahuan secara cepat.

Setiap pelajar memiliki sikap, kemampuan, dan latar belakang yang berbeda-beda. Perbedaan-perbedaan ini yang mempengaruhi performa dari masing-masing pelajar, yang signifikan antara pelajar satu dengan yang lainnya. Dataset Student Performance Prediction berisi informasi performa dari pelajar Sekolah Menengah Atas, terdapat nilai matematika, membaca, menulis dan informasi umum, seperti jenis kelamin (*Gender*), ras atau etnik (*Race/ethnicity*), tingkat pendidikan orang tua (*Parental level of education*), pemilihan makan siang (*Lunch*), dan kursus persiapan ujian (*Test preparation course*). Sample dari data ini berasal dari pelajar-pelajar 3 Sekolah Menengah Atas.

Dataset digunakan untuk melihat dampak dari *data categorical* (*Gender, Race/ethnicity, Parental level of education, Lunch, and Test preparation course*) terhadap *data numeric* (nilai matematika, membaca, dan menulis). Pada project ini data yang dipakai merupakan data kursus persiapan ujian (*Test preparation course*) terhadap data nilai matematika, membaca, dan menulis. Bertujuan untuk mengetahui bagaimana aspek kursus persiapan ujian (*Test preparation course*) berpengaruh terhadap nilai-nilai pelajar yang mempengaruhi performa para pelajar. Kursus persiapan ujian (*Test preparation course*) terbagi menjadi dua kategori yaitu "Completed" yang berarti pelajar telah menyelesaikan atau mengikuti kursus persiapan ujian dan "None" bagi pelajar yang tidak. Hasil prediksi akan melihat jika pelajar mengikuti kursus persiapan ujian apakah performa pelajar akan baik atau sebaliknya.

METODE PENELITIAN

Data Mining menggunakan Teknik Klasifikasi dengan 5 Model Algoritma

1. *K-Nearest Neighbor* (k-NN)

Algoritma *K-Nearest Neighbor* (k-NN) merupakan sebuah model algoritma yang digunakan untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Ataupun dapat dipahami juga bahwa *k-nearest neighbor* adalah salah satu algoritma yang paling sederhana dan banyak digunakan. Titik data akan diklasifikasikan berdasarkan kesamaan kelompok tertentu dari titik data lain yang berdekatan. Sehingga, algoritma ini akan memberikan hasil yang kompetitif.

2. *Naïve Bayes*

Salah satu metode data mining ialah klasifikasi *Naive Bayes*. *Naive Bayes Classifier* adalah metode klasifikasi yang berakar pada teorema Bayes. Naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasi probabilitas. Tujuan dari metode *Naive Bayes* adalah untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. Hasil dari perhitungan data mining menggunakan metode klasifikasi *Naive Bayes* akan makin berguna jika penyajiannya menarik dan dapat dipahami dengan baik oleh penerima data.

3. *Decision Tree*

Decision tree adalah algoritma *machine learning* yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan. Klasifikasi ini menggunakan observasi pada *node* untuk menemukan target pada *leaves*. *Decision Tree* merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia dengan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple.[6]

4. *Ensemble Method*

Ensemble Method adalah algoritma dalam pembelajaran mesin (*machine learning*) dimana algoritma ini sebagai pencarian solusi prediksi terbaik dibandingkan dengan algoritma yang lain karena metode ensemble ini menggunakan beberapa algoritma pembelajaran untuk pencapaian solusi prediksi yang lebih baik daripada algoritma yang bisa diperoleh dari salah satu pembelajaran algoritma konstituen saja. Tidak seperti ensemble statistika dalam mekanika statistika biasanya selalu tak terbatas. Ensemble Pembelajaran hanya terdiri dari seperangkat model alternatif yang bersifat terbatas, namun biasanya memungkinkan untuk menjadi lebih banyak lagi struktur fleksibel yang ada diantara alternatif model itu sendiri. Evaluasi prediksi dari ensemble biasanya memerlukan banyak komputasi daripada evaluasi prediksi model tunggal (*single model*).

5. *Linear Regression*

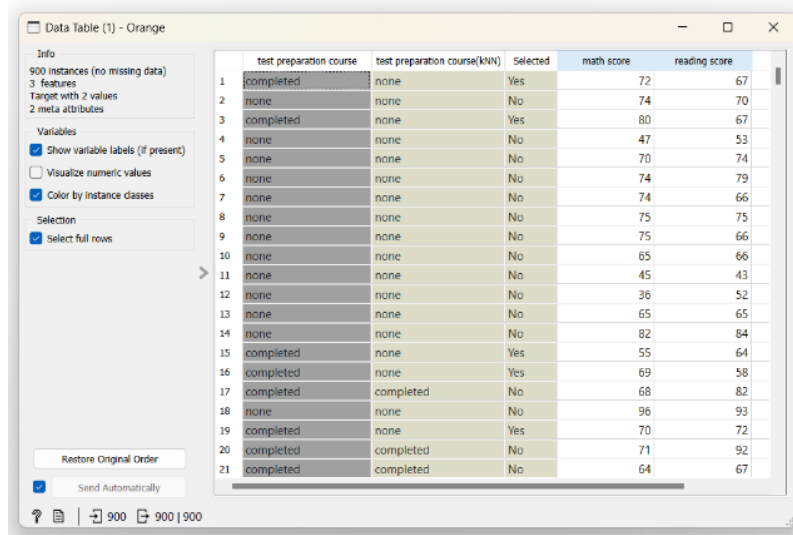
Linear regression merupakan salah satu algoritma yang memodelkan suatu persamaan untuk menghitung estimasi. Pada metode ini bertujuan untuk mencari pola pada nilai numerik, sehingga data yang dibutuhkan berupa data numerik agar dapat diolah dengan model algoritma ini. *Linear regression* memprediksi nilai data yang tidak diketahui dengan

menggunakan nilai data lain yang terkait dan diketahui. Secara matematis memodelkan variabel yang tidak diketahui atau tergantung dan variabel yang dikenal atau independen sebagai persamaan linier..

HASIL DAN PEMBAHASAN

HASIL ANALISIS

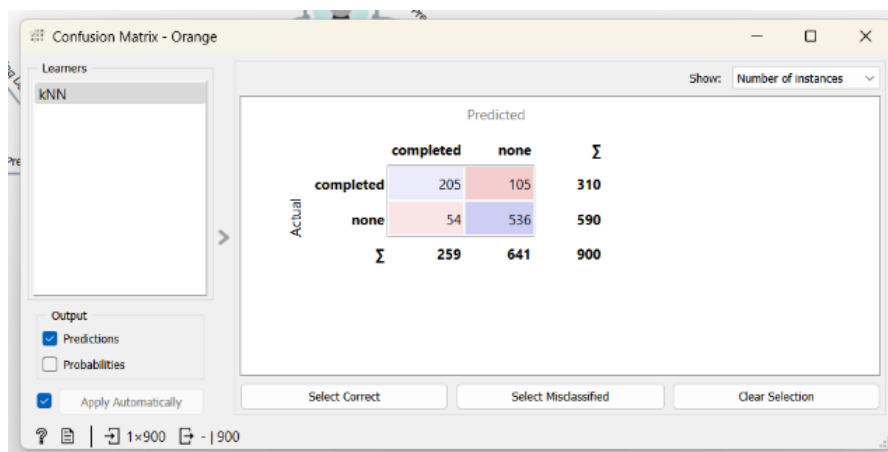
1. K-NEAREST NEIGHBOR (KNN)



	test preparation course	test preparation course(kNN)	Selected	math score	reading score
1	completed	none	Yes	72	67
2	none	none	No	74	70
3	completed	none	Yes	80	67
4	none	none	No	47	53
5	none	none	No	70	74
6	none	none	No	74	79
7	none	none	No	74	66
8	none	none	No	75	75
9	none	none	No	75	66
10	none	none	No	65	66
11	none	none	No	45	43
12	none	none	No	36	52
13	none	none	No	65	65
14	none	none	No	82	84
15	completed	none	Yes	55	64
16	completed	none	Yes	69	58
17	completed	completed	No	68	82
18	none	none	No	96	93
19	completed	none	Yes	70	72
20	completed	completed	No	71	92
21	completed	completed	No	64	67

Gambar 1 Data Tabel KNN

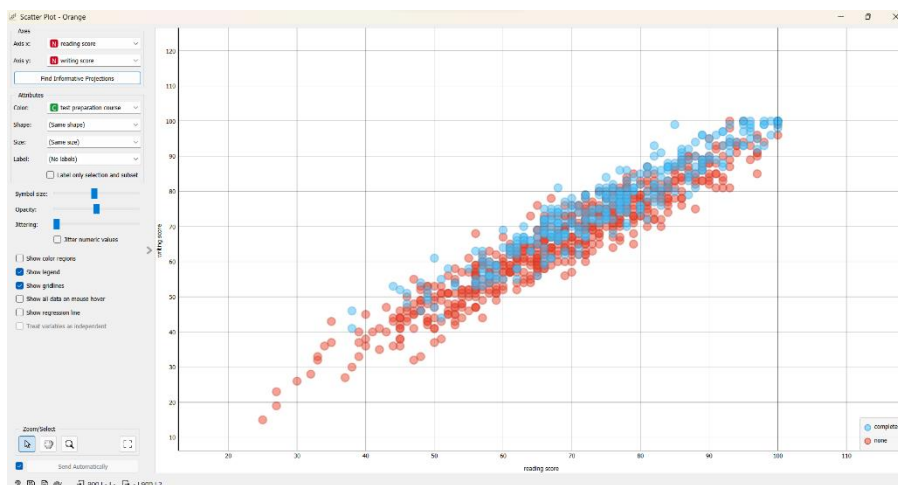
Data table pada dasarnya merupakan tabel yang menunjukkan hasil data mining terkait data yang diinput secara lebih rinci. Pada algoritma K-Nearest Neighbor (k-NN) data table menunjukkan bagaimana algoritma ini mengolah data Test Preparation Course baik secara prediksi ataupun aktual disertai dengan hasil nilai yang diperoleh murid-murid pada masing-masing score. Pada data table ini data tersebut ditunjukkan secara terperinci setiap muridnya per satu data.



		Predicted		Σ
		completed	none	
Actual	completed	205	105	310
	none	54	536	590
Σ		259	641	900

Gambar 2 Confusion KNN

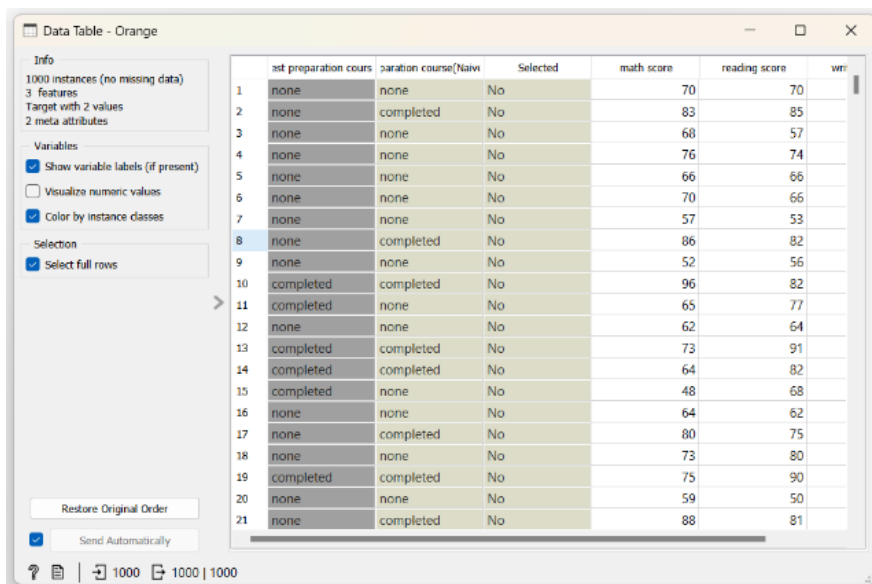
Hasil data mining pada confusion matrix di atas menunjukkan terdapat 205 data yang benar mengikuti Test Preparation Course baik pada data prediksi ataupun aktual, dimana pada data prediksi dan aktual keduanya pada kategori positive yaitu "completed". Dengan ini maka angka 205 menunjukkan true positive. Lalu terdapat 536 data yang benar bahwa mereka tidak mengikuti Test Preparation Course, karena baik data prediksi ataupun aktual masuk kategori negative "none". Sehingga angka 536 menunjukkan true negative. Angka 54 menunjukkan false negative, karena data prediksi pada kategori positive yaitu "completed", sedangkan data ground truth-nya (yang sebenarnya) adalah "none" maka prediksinya false atau salah dan berkategori negative. Dan 105 menunjukkan disebut false positive karena prediksinya kategori negative "none" tetapi ground truth-nya yaitu "completed" maka prediksinya false atau salah dan berkategori positive.



Gambar 3 scatterplot

Pada metode KNN pola titik-titik pada scatterplot tidak menyebar dan memiliki pola yang jelas. Titik-titik scatterplot pada metode ini berada hanya di atas angka 0 (di atas angka 10) pada sumbu Y, hal ini menandakan hasil dari scatterplot ini terdapat gejala heteroskedastisitas (adanya ketidaksamaan varian dari residual untuk semua pengamatan pada model regresi). Syarat untuk terbebas dari gejala heteroskedastisitas merupakan pola titik-titik pada scatterplot menyebar dengan pola tidak jelas dan titik-titik pada scatterplot harus berada pada di atas dan di bawah angka 0 pada sumbu Y.

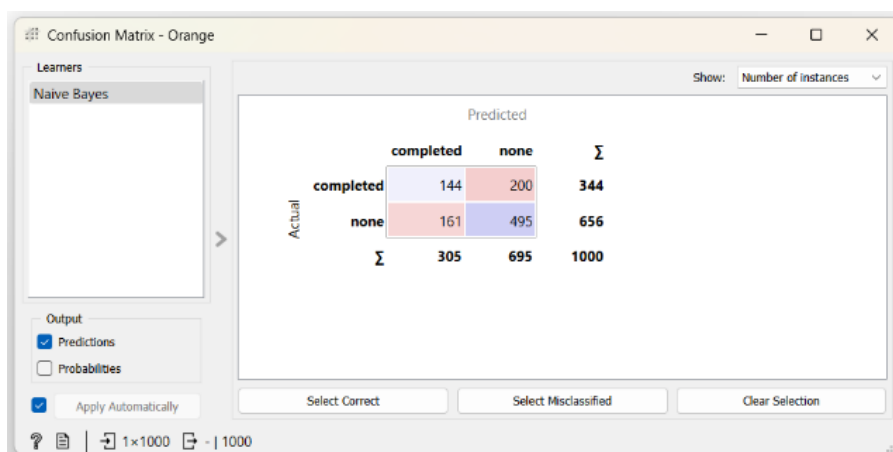
2. Naive Bayes



	test preparation cours	paration course(Naiv	Selected	math score	reading score	wrn
1	none	none	No	70	70	
2	none	completed	No	83	85	
3	none	none	No	68	57	
4	none	none	No	76	74	
5	none	none	No	66	66	
6	none	none	No	70	66	
7	none	none	No	57	53	
8	none	completed	No	86	82	
9	none	none	No	52	56	
10	completed	completed	No	96	82	
11	completed	none	No	65	77	
12	none	none	No	62	64	
13	completed	completed	No	73	91	
14	completed	completed	No	64	82	
15	completed	none	No	48	68	
16	none	none	No	64	62	
17	none	completed	No	80	75	
18	none	none	No	73	80	
19	completed	completed	No	75	90	
20	none	none	No	59	50	
21	none	completed	No	88	81	

Gambar 4 Data Tabel Naive Bayes

Data table pada algoritma *Naive Bayes* pun menunjukkan hasil data mining secara terperinci sesuai dengan dataset yang diinput. Maka dari itu pada data table ini ditunjukkannya data aktual dan prediksi dari algoritma ini sendiri terkait Test Preparation Course dan dilengkapi dengan data terkait nilai-nilai para murid. Semua data pada data table ini ditunjukkan secara per satu data, yang berarti jika pada dataset terdapat 1000 data maka 1000 data ditunjukkan secara tersusun di data table ini.



		Predicted		Σ
		completed	none	
Actual	completed	144	200	344
	none	161	495	656
Σ		305	695	1000

Gambar 5 Confusion Naive Bayes

Pada confusion matrix di atas angka 144 merupakan true positive karena prediksinya sesuai dengan ground truth-nya (yang sebenarnya), serta masuk pada kategori positive, yaitu "completed" yang berarti mereka benar mengikuti Test Preparation Course. Pada angka 495 disebut true negative karena prediksinya sesuai dengan ground truth-nya (yang sebenarnya), tetapi kategorinya negative, yaitu "none" yang berarti mereka benar tidak mengikuti Test

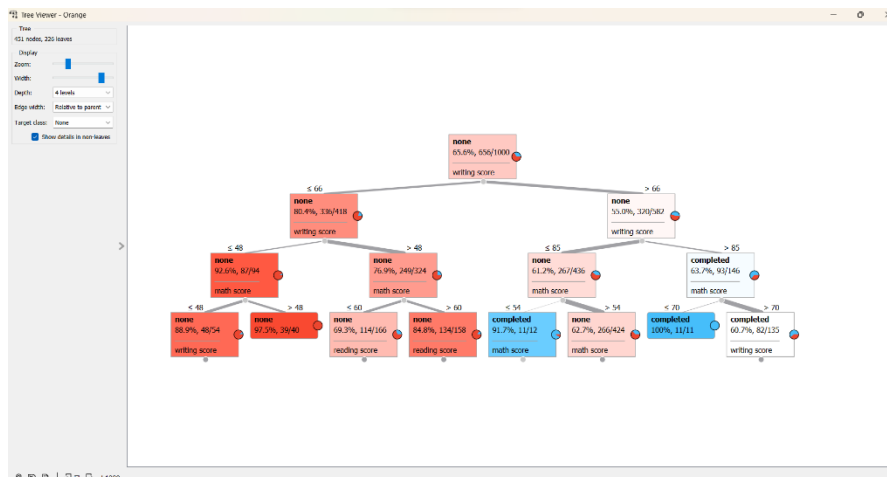
Preparation Course. Pada angka 200 disebut false positive karena prediksinya kategori negative "none" tetapi ground truth-nya yaitu "completed" maka prediksinya false atau salah dan berkategori positive. Pada angka 161 disebut false negative karena prediksinya kategori positive "completed" tetapi ground truth-nya (yang sebenarnya) adalah "none" maka prediksinya false atau salah dan berkategori negative.



Gambar 6 Scatter Plot Naïve Bayes

Pada metode *Naïve Bayes* pola titik-titik pada scatterplot tidak menyebar dan memiliki pola yang jelas. Titik-titik scatterplot pada metode ini berada hanya di atas angka 0 (di atas angka 20) pada sumbu Y, hal ini menandakan hasil dari scatterplot ini terdapat gejala heteroskedastisitas (adanya ketidaksamaan varian dari residual untuk semua pengamatan pada model regresi).

3. DECISION TREE

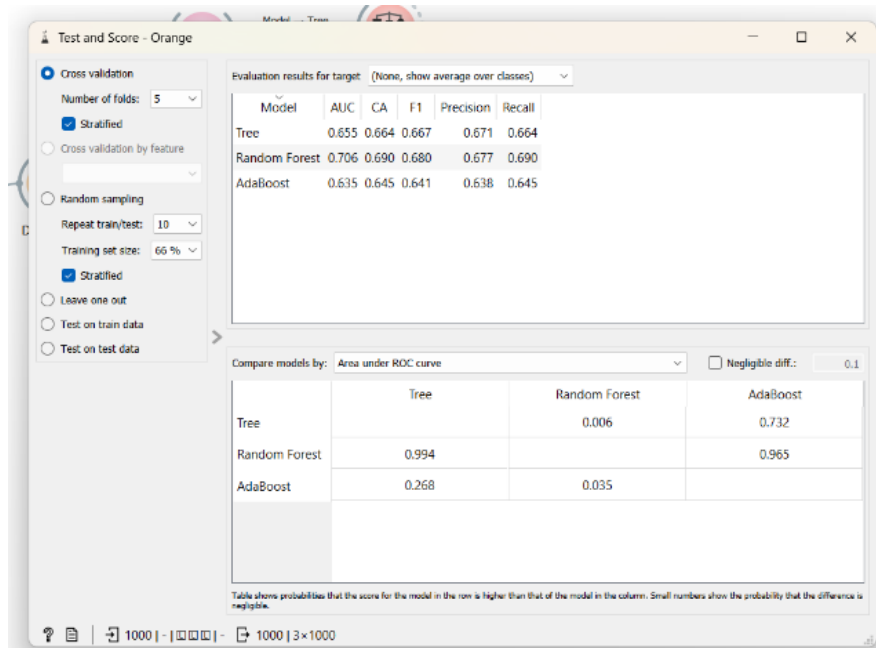


Gambar 7 Hasil Data mining

Hasil data mining dengan algoritma decision tree ini menunjukkan dengan hasil dimana pada kotak paling atas menunjukkan data dengan "none" sebesar 65,6%, yang berarti 656 dari 1000 data menunjukkan bahwa tidak mengikuti Test preparation course. Lalu pada akar pohon

pertama menunjukkan jika nilai writing score kurang atau sama dengan 66 menunjukkan bahwa murid yang memiliki nilai tersebut tidak memiliki perubahan atau pengaruh apapun pada nilai nya karena tidak mengikuti Test preparation course. Hal tersebut didukung dengan tiga akar pohon selanjutnya pada bagian kiri yang menunjukkan pada writing score maupun math score mulai dari 48 hingga 66, semua data masuk pada kategori "none". Adapun pada tiga akar pohon bagian kanan menunjukkan adanya data yang masuk kategori "completed", dimana berarti murid mengikuti Test preparation course. Berdasarkan gambar hasil tree viewer pada akar pohon di bagian kanan tersebut dapat dilihat mayoritas bahwa murid yang memiliki nilai pada writing score atau math score yang dimulai dari 67 hingga lebih atau diatas 85 mengikuti Test preparation course. Hal tersebut menunjukkan bahwa murid yang memiliki nilai pada writing score atau math score yang dimulai dari 67 hingga lebih atau diatas 85 mendapatkan pengaruh pada nilai nya berdasarkan keikutsertaannya pada Test preparation course.

4. ENSEMBLE METHOD



Gambar 8 Test And Score

Dari hasil test and score kita dapat melihat bahwa nilai AUC yang paling besar adalah Random forest yaitu sebesar 0.706, karena Tree hanya 0.655 dan AdaBoost hanya 0.635. Begitupun dengan nilai CA dan F1 Random Forest yaitu 0.690 dan 0.680 yang lebih besar dibandingkan nilai CA dan F1 Tree yaitu 0.664 dan 0.667, serta nilai CA dan F1 AdaBoost yaitu 0.645 dan 0.641. Adapun nilai precision dan recall Random Forest, yaitu 0.677 dan 0.690 yang mana lebih tinggi dibandingkan nilai precision dan recall Tree yaitu 0.671 dan 0.664, serta nilai

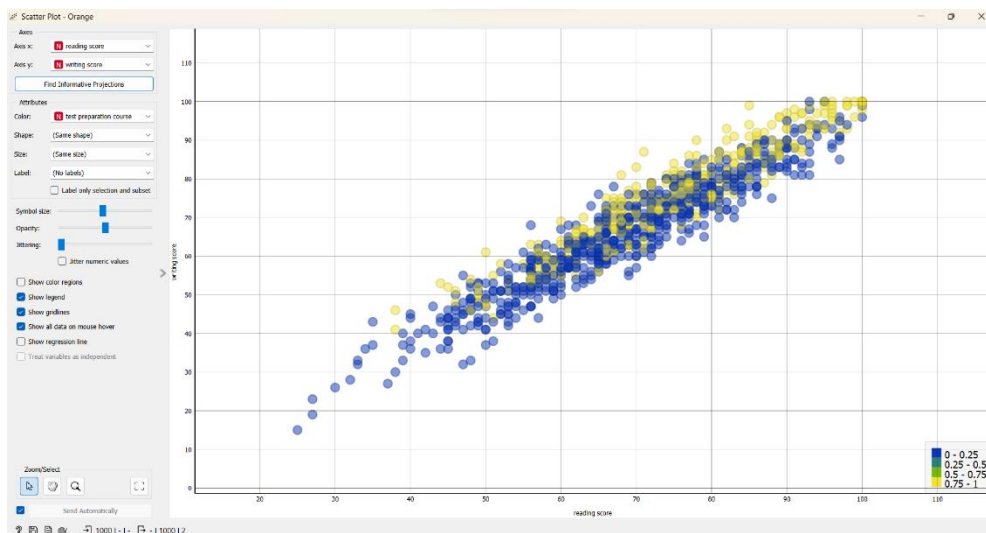
precision dan recall AdaBoost yaitu 0.638 dan 0.645. Sehingga dapat kita lihat bahwa metode Random forest berarti bahwa kinerja Random Forest dalam pengolahan data sangat baik dibandingkan Tree dan AdaBoost.

Instance	Tree	AdaBoost	Random Forest	Tree (Completed)	Tree (Time)	AdaBoost (Completed)	AdaBoost (Time)	Random Forest (Completed)	Random Forest (Time)	math score		
1	none	none	none	0	1	0.83754-07	0.99999	0.125	0.875	68		
2	none	none	none	0	1	1.20171e-08	1	0.032333	0.91667	70		
3	none	none	none	0	1	0.181818	0.81818	1.3243e-07	1	0	63	
4	completed	completed	none	none	0.66667	0.33333	1.7242e-05	0.99993	0.46111	0.28889	73	
5	completed	none	none	completed	0	1	2.6820e-07	1	0.62781	0.57019	64	
6	none	none	none	0	1	1.05814e-06	0.99999	0.25922	0.74078	1	88	
7	completed	none	none	completed	0	1	1.51382e-05	0.99985	0.525	0.475	73	
8	none	none	none	0	1	2.02092e-07	1	0.3775	0.8225	1	90	
9	none	none	none	0	1	7.91532e-07	0.99999	0.15	0.87	1	82	
10	completed	none	completed	completed	0	1	0.620731	0.371289	0.714643	0.285357	59	
11	none	completed	completed	none	0.5	0.5	0.99999	1.86146e-06	0.44328	0.55672	76	
12	none	none	none	0	1	0.945822	0.997273	0.902673	0	1	62	
13	none	completed	none	none	0.5	0.5	1.62227e-07	1	0.15667	0.84333	59	
14	completed	completed	none	completed	1	0	0.022459	0.97574	0.69085	0.30914	83	
15	completed	completed	none	none	0.66667	0.33333	2.8386e-10	1	0.17833	0.82167	75	
16	none	completed	completed	completed	1	0	0.99979	2.1545e-05	0.54833	0.45167	74	
17	none	none	none	0	1	2.41370e-08	1	0.081333	0.91867	1	66	
18	none	completed	completed	completed	1	0	0.99998	2.31589e-06	0.52	0.48	100	
19	completed	completed	completed	completed	1	0	3.37920e-07	0.616429	0.333571	1	82	
20	none	completed	none	completed	1	0	0.005111	0.99248	0.85262	0.14737	57	
21	none	completed	none	0	1	0.00034248	0.99987	0.40333	0.59667	1	70	
22	none	completed	completed	completed	0.5	0.5	0.975199	0.024801	0.75883	0.24117	75	
23	none	completed	none	0	1	0.000114761	0.99985	0.34	0.66	1	78	
24	completed	completed	completed	none	0.5	0.5	0.994207	0.0057337	0.20333	0.79667	62	
25	completed	completed	none	completed	0.66667	0.33333	0.0010785	0.907812	0.21667	0.78333	66	
26	none	none	completed	completed	0	1	0.790259	0.209741	0.56274	0.43726	74	
27	none	none	none	0	1	1.80229e-05	0.99998	0	1	1	67	
28	none	none	completed	completed	0.11111	0.88889	0.875215	0.12478	0.875	0.125	58	
29	none	none	none	0	1	2.17481e-07	1	0.033333	0.96667	1	64	
30	none	none	none	0	1	4.87868e-08	1	0.62	0.38	1	70	
31	none	completed	none	none	0.8	0.2	0.0011931	0.949807	0.35	0.64	67	
32	completed	none	none	0	1	0.808889	0.00052377	0.99907	0.00033	0.99967	63	
33	completed	completed	none	completed	0.66667	0.33333	0.99984	1.56803e-05	0.45167	0.54833	88	
34	completed	none	completed	none	0	1	0.782305	0.217695	0	1	62	
35	none	none	none	0	1	0.950877	0.94912	3.09386e-08	1	0.042571	0.95743	44
36	none	none	none	0	1	0.0004214	0.99958	0.00000	0.99999	0.76667	71	

Gambar 9 Data Tabel

Adapun kami juga menyediakan visualisasi data tabel yang berguna untuk melihat perbandingan hasil prediksi antara Tree, Random, AdaBoost, dan ground truth secara lebih mendetail. Di data tabel ini ditampilkan perincian dari 1000 data hasil dari metode ensemble per datanya. Jadi kita bisa mengetahui data mana saja yang diprediksi benar oleh Tree, Random Forest, dan AdaBoost.

5. LINEAR REGRESSION



Gambar 10 Scatter Plot Linear Regression

Pada metode *Linear Regression* pola titik-titik pada scatterplot tidak menyebar dan memiliki pola yang jelas. Titik-titik scatterplot pada metode ini berada hanya di atas angka 0 (di atas angka 10) pada sumbu Y, hal ini menandakan hasil dari scatterplot ini terdapat gejala heteroskedastisitas (adanya ketidaksamaan varian dari residual untuk semua pengamatan pada model regresi).

PERBANDINGAN HASIL ANALISIS

1. Perbandingan Analisis Hasil Test and Score

Model	AUC	CA	F1	Precision	Recall
kNN	0.889	0.823	0.819	0.821	0.823
Naïve Bayes	0.633	0.639	0.633	0.630	0.639
Tree	0.655	0.664	0.667	0.671	0.664
Random Forest	0.706	0.690	0.680	0.677	0.690
AdaBoost	0.635	0.645	0.641	0.638	0.645

Gambar 11 Perbandingan Test and Score

Dari tabel perbandingan hasil test and score di atas, kita dapat melihat bahwa nilai AUC, CA, F1, Precision, dan Recall yang paling besar adalah yang dihasilkan oleh algoritma klasifikasi kNN yaitu sebesar 0.889; 0.823 ;0.819; 0.821; dan 0.823. Sementara nilai AUC, CA, F1, Precision, dan Recall paling rendah adalah dihasilkan oleh algoritma Naive Bayes yaitu sebesar 0.633; 0.639; 0.633; 0.630; dan 0.639. Sehingga dapat kita lihat bahwa pengklasifikasian data dengan algoritma k-NN memiliki kinerja yang sangat baik dalam pengolahan data *Student Performance Prediction*. Sementara algoritma klasifikasi Naive Bayes memiliki kinerja yang paling buruk dalam pengolahan data *Student Performance Prediction*.

2. Perbandingan Analisis Hasil Confusion Matriks

Model	True Positive	True Negative	Total True	False Positive	False Negative	Total False
kNN	205	536	741	105	54	159
Naïve Bayes	144	495	639	200	161	361
Tree	191	473	664	153	183	336
Random Forest	153	537	690	191	119	310
AdaBoost	151	494	645	193	162	355

Gambar 12 Perbandingan Confusion Matriks

Dari tabel perbandingan analisis hasil confusion matriks dapat diketahui bahwa algoritma klasifikasi yang mampu memprediksi hasil yang benar dengan ground truth (yang sebenarnya) yaitu k-NN. Algoritma k-NN mampu memprediksi data dengan benar sebanyak 741 data, baik memprediksi hasil positive (completed) maupun hasil negative (none) yang sesuai dengan ground truth atau keadaan sebenarnya. Sementara itu algoritma klasifikasi yang paling sedikit memprediksi data sesuai dengan yang sebenarnya yaitu Naive Bayes yang prediksinya hanya benar 639 data saja, baik prediksi hasil positive (completed) maupun prediksi hasil negative

(none). Dari hal tersebut dapat kita lihat bahwa algoritma kNN memiliki kinerja yang sangat baik dalam pengolahan data *Student Performance Prediction*.

SIMPULAN

Setelah dilakukannya data mining menggunakan aplikasi orange dalam project ini, menghasilkan data berdasarkan lima model algoritma. Lima model algoritma tersebut terdiri dari k-Nearest Neighbor, Naive Bayes, Decision Tree, Ensemble, dan Linear Regression. Kelima model tersebut memiliki kriteria tujuan dan penggunaan masing-masing, sehingga dapat hasil data mining yang dihasilkan pun berbeda.

Berdasarkan hasil data mining dari kelime model tersebut dapat disimpulkan bahwa model algoritma yang paling akurat diantara kelimanya adalah model k-Nearest Neighbor (k-NN). Hal ini ditunjukkan pada hasil perbandingan *confusion matrix*, di dalamnya menunjukkan bahwa model ini berhasil menjawab benar dengan jumlah yang paling banyak. Adapun jika dilihat berdasarkan hasil perbandingan *test and score*, model k-NN juga menunjukkan nilai yang paling tinggi.

Saran yang dapat diberikan oleh tim penulis adalah diharapkan pada penelitian selanjutnya dengan menggunakan dataset yang serupa, diharapkan dapat dikembangkan lagi dengan menggunakan perbandingan dari beberapa algoritma klasifikasi yang lainnya selain k-NN, Naive Bayes, Decision Tree, Ensemble, dan Linear Regression.

DAFTAR PUSTAKA

<https://www.kaggle.com/datasets/rkiattisak/student-performance-in-mathematics>

Mardiani, Eri, Ferdan Akbar Ramadhan (2023). Design Information System Sales of Nuts and Bolts at PT. Catur Naga Steelindo. Jurnal SITEKIN: Jurnal Sains, Teknologi dan Industri , Vol. 20, No. 2, June 2023, pp.729 – 735

Mardiani, Eri, Nur Rahmansyah, Sari Ningsih, Endah Tri Esthi Handayani, Deny Hidayatullah (2023). PKM Meningkatkan Penjualan Umkm Bagi Pelaku Bisnis Usia Lanjut Melalui Inovasi Ecommerce Disaat Pandemi COVID 19. Jurnal Swadimas: Jurnal Pengabdian Masyarakat Volume 1, No 1, 2023.

Mardiani, Eri, Rahmansyah, N, Ningsih, Sari (2022). PKM Meningkatkan Penjualan UMKM Dengan E-Commerce Disaat Pandemi Covid 19. Jurnal Mindabaharu: Jurnal Pengabdian Masyarakat Volume 6, No 2 Desember, 2022. 234-243.

Mardiani, Eri., Rahmansyah, N., Wahyudi, N. M., Wijaya, Y. F., & Al Rizky, F. (2021). *Kumpulan Latihan PHP*. Elex Media Komputindo. Jakarta

- Mardiani, Eri, Nur Rahmansyah, Farid Al Rizky(2020).
<https://ejournal.upnvj.ac.id/index.php/informatik/article/download/2264/1135>
- Mardiani, Eri, Nur Rahmansyah, Hendra Kurniawan, Dana Indra Sensuse, Jayanta. 2016.
Kumpulan Latihan SQL, Jakarta : Elex Media Komputindo
- Mardiani, Eri, Nur Rahmansyah, Surniawan, Imam Budiawan, Muchammad Sholeh. 2016.
Aplikasi Penggajian Menggunakan Visual Basic, MySQL dan Data Report, Jakarta :
Elex Media Komputindo
- Mardiani, E,, Nur Rahmansyah, Satriawan Desmana,Ahmad Rifqi (2023). Analysis of Buyer's
Trust in E-Commerce Shop And Travel Web. Jurnal SITEKIN: Jurnal Sains, Teknologi
dan Industri , Vol. 20, No. 2, June 2023, pp.850 – 857
- Mardiani, E,, Nur Rahmansyah, Ira Kurniati (2023). Website Design At SDN Cipete Utara 07.
Jurnal SITEKIN: Jurnal Sains, Teknologi dan Industri , Vol. 20, No. 2, June 2023,
pp.891 – 898
- Mardiani, Eri, Nur Rahmansyah, Yanuardi, Dwi Sidik Permana, Hendra Kurniawan. 2019.
Membuat Aplikasi Penjualan Menggunakan Java Netbeans, Mysql, dan iReport,
Jakarta : Elex Media Komputindo