JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956

Penerapan Algoritma Naïve Bayes Untuk Pengelompokkan Predikat Peserta Uji Kemahiran Berbahasa Indonesia

Michal Dennis*, Rahmaddeni, Fransiskus Zoromi, M. Khairul Anam

Teknik Informatika, STMIK Amik Riau, Pekanbaru, Indonesia
Email: [1]michaldennis82@gmail.com, [2]rahmaddeni@sar.ac.id, [3]fransiskus_zoromi@sar.ac.id, [4]khairulanam@sar.ac.id
Email Penulis Korespondensi: michaldennis82@gmail.com

Abstrak−Uji Kemahiran Berbahasa Indonesia adalah uji kemahiran untuk mengukur kemahiran berbahasa seseorang dalam berkomunikasi dengan menggunakan bahasa Indonesia, baik penutur Indonesia maupun penutur asing. Pemeringkatan UKBI memiliki 7 kategori pemeringkatan yang terdiri dari peringkat istimewa, sangat unggul, unggul, madya, semenjana, marginal, dan terbatas. Jumlah Peserta yang mengikuti UKBI di Balai Bahasa Provinsi Riau sudah lebih dari 1000 namun belum ada yang mengelola data tersebut menjadi sebuah pengetahuan baru. Salah satu upaya yang bisa dilakukan dengan data tersebut adalah klasifikasi. Algoritma Naïve Bayes Classification merupakan salah satu algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat). Agar memperoleh hasil akurasi yang baik maka Algoritma Naive Bayes dikombinasikan dengan feature selection Adaboost dengan skema pengujian 70:30 dan 80:20. Hasil penelitian yang dilakukan menghasilkan nilai akurasi tertinggi yaitu 89% dimana melakukan kombinasi algoritma naive bayes dengan feature selection Adaboost dengan splitting data 70:30.

Kata Kunci: UKBI; Naive Bayes; Klasifikasi; Adaboost

Abstract−Indonesian Language Proficiency Test is a proficiency test to measure a person's language proficiency in communicating using Indonesian, both Indonesian speakers and foreign speakers. The UKBI rating has 7 rating categories consisting of special, very excellent, excellent, intermediate, poor, marginal, and limited. The number of participants who take UKBI at the Riau Province Language Center has more than 1000 but no one has managed the data into new knowledge. One of the efforts that can be done with the data is classification. The Naïve Bayes Classification Algorithm is a classification algorithm that is very effective (getting the right results) and efficient (the reasoning process is carried out by utilizing existing inputs in a relatively fast way). In order to obtain good accuracy results, the Naive Bayes Algorithm is combined with the Adaboost feature selection with a 70:30 and 80:20 test scheme. The results of the research carried out resulted in the highest accuracy value, namely 89% which combined the Naive Bayes algorithm with the Adaboost feature selection with 70:30 data splitting.

Keywords: UKBI; Naive Bayes; Classification; Adaboost

1. PENDAHULUAN

Bahasa Indonesia memiliki peranan penting dalam segala sendi kehidupan sejak ditetapkan sebagai bahasa nasional dan bahasa negara. Seiring berjalannya waktu, bahasa Indonesia mengalami berbagai perkembangan dan penambahan jumlah penutur, baik penutur jati maupun penutur asing. Selain itu, bahasa Indonesia juga menghadapi berbagai tantangan seperti maraknya penggunaan bahasa asing di ruang publik serta dominasi bahasa daerah sebagai media komunikasi sehari-hari. Namun tantangan tersebut tidak mengurangi peran strategis bahasa Indonesia dalam bidang pendidikan. Peranan bahasa Indonesia dalam bidang pendidikan dijelaskan dalam (UU RI No 24 Th 2009 Tentang Bendera, Bahasa, Dan Lambang Negara, Serta Lagu Kebangsaan, 2009). Di dalam undangundang tersebut, bahasa Indonesia wajib digunakan sebagai pengantar dalam dunia pendidikan nasional [1][2].

Uji Kemahiran Berbahasa Indonesia (UKBI) adalah sarana untuk mengukur kemahiran seseorang dalam berbahasa Indonesia baik lisan dan tulis. Uji yang hingga kini menjadi satu–satunya uji yang dapat digunakan untuk mengukur kemampuan seseorang dalam berbahasa Indonesia ini dapat diikuti oleh peserta dari berbagai profesi, latar belakang, maupun negara asal. UKBI dikembangkan sebagai tes standar yang berfungsi mengukur kemahiran berbahasa Indonesia baik bagi penutur jati maupun penutur asing. UKBI merupakan instrumen tes kemampuan bahasa Indonesia yang dikembangkan oleh Badan Pengembangan dan Pembinaan Bahasa yang telah teruji validitas dan reliabilitasnya [1][3]. Bagi para pegiat pendidikan, tes UKBI dapat menjadi tolak ukur kemampuan berbahasa seseorang, sama halnya dengan tes kebahasaan lainnya yang sudah ada lebih dahulu [5]. Dalam (Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 70 Tahun 2016 Tentang Standar Kemahiran Berbahasa Indonesia, 2016) bagi penutur jati, profesi jabatan pendidik dalam hal ini guru bahasa Indonesia, guru nonbahasa Indonesia, dosen, dan guru besar memiliki standar minimal kemahiran berbahasa Indonesia yang berbeda.

Hasil UKBI dapat dijadikan standar tingkat kemampuan seseorang dalam berbahasa Indonesia. Pemeringkatan UKBI memiliki 7 kategori pemeringkatan yang terdiri dari peringkat istimewa dengan skor (725-800), sangat unggul (641-724), unggul (578-640), madya (482-577), semenjana (405-481), marginal (326-404) dan terbatas (251-325) [4]. Dari data peserta yang mengikuti UKBI di Balai Bahasa Provinsi Riau belum ada yang mengelola data tersebut menjadi sebuah pengetahuan baru. Salah satu upaya yang bisa dilakukan dengan data tersebut adalah klasifikasi. Balai Bahasa Provinsi Riau perlu mengelompokkan predikat peserta UKBI untuk

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956

mendapatkan pengetahuan baru dalam menentukan suatu kebijakan, untuk dapat meningkatkan kualitas penutur bahasa Indonesia di masa yang akan datang [1].

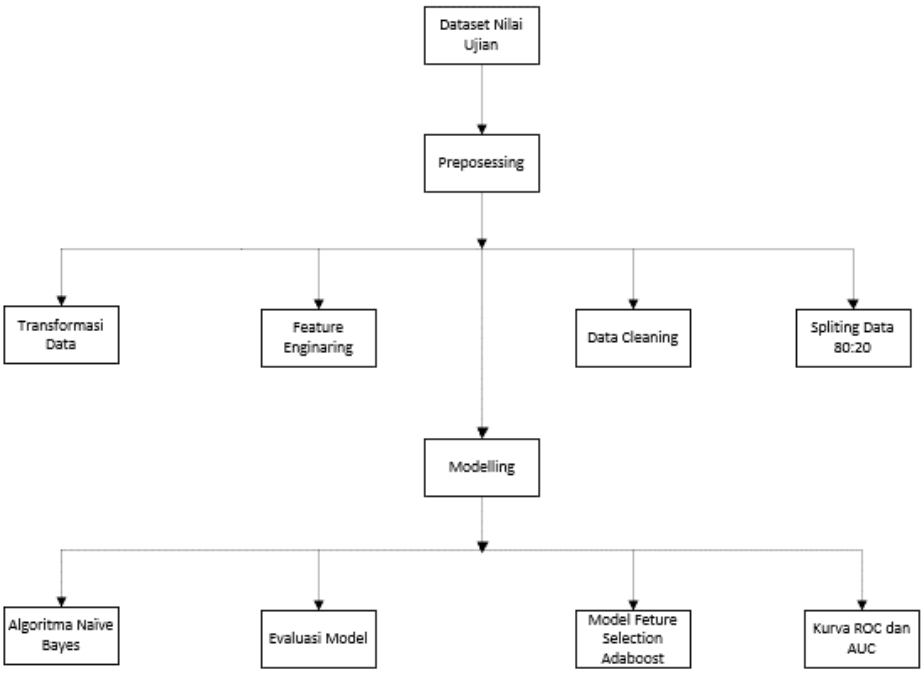
Klasifikasi merupakan pengelompokan objek kedalam kelas tertentu berdasarkan kelompoknya yang biasanya disebut dengan kelas (*class*). Proses klasifikasi terdapat banyak algoritma klasifikasi yang dapat digunakan. Suatu algoritma dikatakan paling baik menyelesaikan suatu permasalahan belum tentu baik juga untuk memecahkan permasalahan yang lain, tergantung pada jenis dan sifat datanya. *Naïve Bayes Classification* merupakan salah satu algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat) [4]. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara konditional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [5].

Penelitian sebelumnya yang menggunakan metode *Naïve Bayes* juga digunakan dalam memprediksi kelulusan mahasiswa dalam mengikuti *english proficiency test* di mana metode Naive Bayes berhasil mengklasifikasikan 49 data dari 50 data yang diuji [6]. Sehingga dengan demikian metode Naive Bayes ini berhasil memprediksi kelulusan mahasiswa dengan persentase keakuratan sebesar 98%. Metode Naïve Bayes juga digunakan dalam memprediksi tingkat kelulusan peserta sertifikasi *microsoft office specialist* dengan hasil memberikan solusi keputusan untuk memilih program sertifikasi yang tepat dengan membandingkan hasil prediksi uji sertifikasi antara word dan excel [7]. Selain itu, penelitian yang berkaitan dengan Naïve Bayes dengan topik klasifikasi berita twitter yang dilakukan oleh [8] yang menggunakan delapan kategori berita berbahasa Indonesia yaitu: ekonomi, entertainment, olahraga, teknologi, kesehatan, makanan, otomotif, dan travel. Hasil penelitian yang telah dilakukan didapatkan hasil nilai *precision* 0.962961, recall 0.789164 dan *f-measure* sebesar 0.862973. Penelitian yang berkaitan dengan peningkatan UKBI melalui program klinik bahasa juga dilakukan oleh [5]. Dimana metode yang digunakan dalam penelitian tersebut yaitu metode deskriptif kualitatif dan kuantitatif sederhana dengan data primer yang diperoleh melalui hasil survei. Hasil penelitian dari 51 peserta yang mengikuti kegiatan, 30 di antaranya mengikuti kegiatan simulasi UKBI dengan skor simulasi UKBI terendah adalah 60 dan tertinggi 100. Setelah mengikuti klinik bahasa, peserta yang sangat berminat mengikuti tes UKBI berjumlah 15 orang, berminat berjumlah 16 orang, dan kurang berminat berjumlah 2 orang atau berjumlah 92%.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam melakukan sebuah penelitian data dan informasi yang bersifat objektif yang akan digunakan sebagai titik acuan dalam penelitian, dengan adanya data-data tersebut di harapkan penelitian yang di hasilkan adalah penelitian yang berkualitas. Proses dalam melakukan penelitian ini digambarkan sebagai berikut [9]:



Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956

Gambar 1. Tahapan Penelitian

2.1.1 Pengumpulan Dataset

Dataset yang digunakan pada penelitian ini sebanyak 1320 data yang terdiri dari tahun 2017 – 2020 yang diperoleh dari UKBI Balai Bahasa Provinsi Riau. Berikut adalah bentuk tampilan data UKBI yang diperoleh.

No.	Nama Peserta	No.Peserta	Tempat dan Tgl.Lahir	Jenis Kelamin	Profesi	Telepon	Tgl.Ujian	TUKBI	Kode Soal	Nilai					Skor UKBI		Peringkat/Predikat	No.Sertifikat
										Seksi I	Seksi II	Seksi III	Seksi IV	Seksi V				
1	Yaiza Ike Aprilla	'4500060212200001	Langsat Hulu, 10 April 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	605	608	575	650	0	609		III (Unggul)	'0497/UKBI/06/2020
2	Rintani Hidayat	'4500060212200002	Pekanbaru, 4 September 2000	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	530	392	425	650	0	499		IV (Madya)	'0498/UKBI/06/2020
3	Wirdatul Hasanah	'4500060212200003	Dumai, 26 Agustus 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	530	464	545	590	0	532		IV (Madya)	'0499/UKBI/06/2020
4	Rani Komala Dewi	'4500060212200004	Sengkemang, 1 Agustus 1997	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	560	440	530	590	0	530		IV (Madya)	'0500/UKBI/06/2020
5	Ririn Anggraini	'4500060212200005	Rumbai, 1 Mei 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	440	416	440	500	0	449		V (Semenjana)	'0501/UKBI/06/2020
6	Dira Elfitri	'4500060212200006	Sialang Bawah, 21 Desember 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	500	536	695	710	0	610		III (Unggul)	'0502/UKBI/06/2020
7	Wulandari Eka Putri Nasution	'4500060212200007	Tanjungbalai, 17 September 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	575	560	575	620	0	582		III (Unggul)	'0503/UKBI/06/2020
8	Mutiara Chania	'4500060212200008	Pekanbaru, 28 November 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	515	536	500	530	0	520		IV (Madya)	'0504/UKBI/06/2020
9	Sri Ratika	'4500060212200009	Bandung, 20 Agustus 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	560	488	545	620	0	553		IV (Madya)	'0505/UKBI/06/2020
10	Roma Ito	'4500060212200010	Pekanbaru, 3 April 1997	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	425	368	560	530	0	471		V (Semenjana)	'0506/UKBI/06/2020
11	Siti Rahayu Darmica	'4500060212200011	Sedinginan, 4 Mei 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	575	512	515	500	0	525		IV (Madya)	'0507/UKBI/06/2020
12	Rezki Ayu	'4500060212200012	Sungaiapit, 25 Desember 1999	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	560	488	575	590	0	553		IV (Madya)	'0508/UKBI/06/2020
13	Melinda Antoni Putri	'4500060212200013	Pekanbaru, 16 September 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	470	440	620	740	0	568		IV (Madya)	'0509/UKBI/06/2020
14	Ami Septya Roza	'4500060212200014	Binjai, 5 September 1997	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	605	560	575	770	0	628		III (Unggul)	'0510/UKBI/06/2020
15	Humairoh	'4500060212200015	Muarakuantan, 5 September 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	515	512	515	530	0	518		IV (Madya)	'0511/UKBI/06/2020
16	Ria Permata Sari	'4500060212200016	Bengkalis, 15 Juni 1998	Perempuan	Mahasiswa		Rabu, 02-12-2020	Balai Bahasa Riau	45	575	608	530	620	0	583		III (Unggul)	'0512/UKBI/06/2020

Gambar 2. Dataset Hasil UKBI

2.1.2 Tahap Preprocessing

Tahapan ini terdiri dari beberapa proses, Tahap preprocessing dilakukan dengan menggunakan bantuan library pada Bahasa pemrograman Python3. Penerapan tahap preprocessing data pada penelitian ini dilakukan dengan melakukan 4 proses, di antaranya [10]:

1. Transformasi Data
   Transformasi Data dilakukan dengan tujuan utama untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis data.
2. Feature Engineering
   Feature Engineering mengacu pada proses penggunaan pengetahuan domain untuk memilih dan mengubah variabel yang paling relevan dari data mentah saat membuat model prediktif menggunakan pembelajaran mesin atau pemodelan statistik.
3. Data Cleaning
   Proses menyiapkan data untuk dilakukan analisis dengan cara menghapus atau memodifikasi data salah, tidak relevan, duplikat, dan tidak terformat
4. Splitting Data
   Splitting data digunakan membagi data menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Pada proses kali ini dicoba 3 kali percobaan. Percobaan pertama data dibagi sebesar 70% untuk data latih dan 30% untuk data uji dan percobaan kedua data dibagi sebesar 80% untuk data latih dan 20% untuk data uji.

2.1.3 Modelling

Pemodelan data dilakukan untuk mengetahui alur bagaimana proses-proses dan metode berjalan sebelum diimplementasikan ke dalam sebuah aplikasi dengan data yang ada. Berikut tahap pemodelan data untuk penentuan predikat UKBI menggunakan Naïve Bayes Classifier [11]:

1. Algoritma Naive Bayes
   Tahap pertama dari modelling adalah menerapkan algoritma Naive Bayes, untuk menerapkan algoritma Naive Bayes penulis menggunakan bahasa pemrograman Python dan menggunakan tools jupyter notebook.
2. Evaluasi Model
   Evaluasi model dilakukan untuk mengetahui performa dari metode Naive Bayes, maka dilakukan pengujian terhadap model yang telah dibuat. Teknik yang digunakan adalah splitting data dengan membandingkan data testing dan data training dengan dua skema yaitu 70:30 dan 80:20.
3. Model Feature Selection Adaboost
   Feature selection merupakan salah satu teknik terpenting dan sering digunakan dalam pre-processing. Teknik ini mengurangi jumlah fitur yang terlibat untuk menentukan suatu nilai kelas target dengan mengurangi fitur yang tidak relevan dan data berlebih. Feature selection yang digunakan adalah Adaboost [12].
4. Kurva ROC dan AUC

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

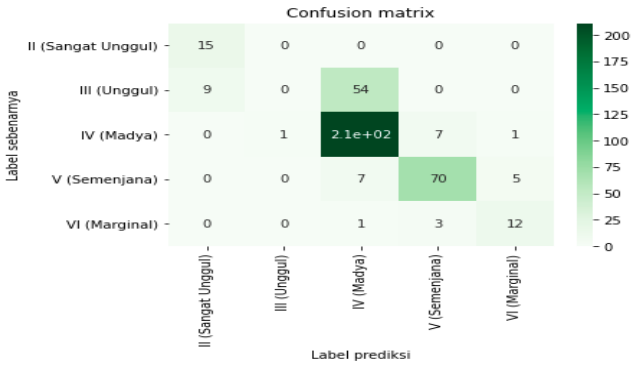
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956

Kurva ROC adalah grafik dua dimensi hubungan antara *True Positive Rate* (TPR) atau *Sensitivity* (sumbu Y) dengan *False Positive Rate* (FPR) atau 1- *Specificity* (sumbu X). AUC (*Area Under Curve*) merupakan daerah berbentuk persegi yang nilainya selalu berada diantara 0 dan 1. [13]

3. HASIL DAN PEMBAHASAN

Tahapan ini terdiri dari beberapa proses, Tahap preprocessing dilakukan dengan menggunakan bantuan library pada Bahasa pemrograman Python3. Penerapan tahap preprocessing data pada penelitian ini dilakukan dengan melakukan proses seperti yang dijelaskan pada gambar 1. Untuk menghasilkan nilai akurasi yang tinggi maka dilakukan proses data training dan testing. Adapun proses ini menggunakan skema 70:30 dan 80:20. [14]

Implementasi Algoritma Naive Bayes dengan Python dengan menggunakan tools jupyter notebook dengan mengimport *library sklearn naive bayes*. Selanjutnya untuk mengetahui performa dari metode Naive Bayes, maka dilakukan pengujian terhadap model yang telah dibuat. Hasil klasifikasi akan divisualisasi dalam bentuk confusion matrix [15]. Berikut adalah pengujian model klasifikasi dengan menggunakan library python sklearn.metric yang didalamnya memiliki confusion_matrix dan divisualisasi dengan menggunakan seaborn yang merupakan pustaka visualisasi dengan sumber terbuka dibangun diatas pustaka matplotlib. Gambar tabel 3 dibawah adalah visualisasi model Naïve Bayes 70 : 30.



Gambar 3. Visualisasi Confusion Matrix 70:30

Setelah diketahui Confusion Matrix dari model yang dibuat selanjutnya dilakukan perhitungan nilai akurasi dari model klasifikasi Naive Bayes yang telah dibuat menggunakan sklearn.metrics dengan mengimport accuracy score yang disediakan oleh library scikit.learn. Didapatkan hasil dari perhitungan matrix dengan kode program python yaitu sebesar 0.78.

Proses untuk menghasilkan *classification report* maka dilakukan beberapa tahapan yaitu memiliki memiliki data, kemudian melakukan proses import data, melakukan proses klasifikasi yaitu menggunakan naive bayes terdapat dalam package sklearn. Dalam pengklasifikasian ini dibutuhkan data testing dan data training [16]. Berikutnya adalah proses confusion matrix yang berisi ketepatan prediksi. Proses *classification report* yaitu menampilkan tingkat akurai dari klasifikasi dengan metode naive bayes yang dilakukan. Pada Tabel. 2 dibawah berikut adalah hasil *classification report* model Naive Bayes.

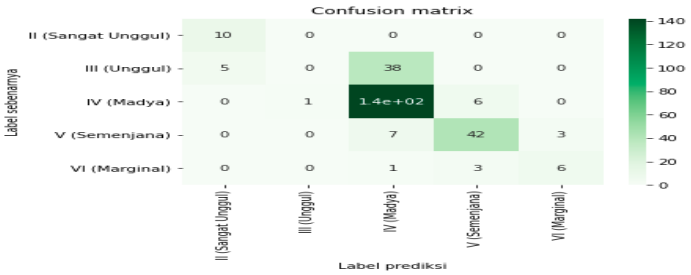
Tabel 2. Hasil Evaluasi Model Naïve Bayes 70 : 30

	precision	**recall**	**f1-score**	**support**
2	0,62	1,00	0,77	15
3	0,00	0,00	0,00	63
4	0,77	0,96	0,86	220
5	0,88	0,85	0,86	82
6	0,67	0,75	0,71	16
accuracy			0,78	396
macro avg	0,59	0,71	0,64	396
weighted avg	0,66	0,78	0,71	396

Berikutnya dilakukan juga splitting data dengan skema 80:20, setelah dilakukan pengolahan menggunakan Python maka diperoleh hasilnya seperti ditampilkan pada gambar 4 dibawah ini.

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956



Gambar 4. Visualisasi Confusion Matrix 80:20

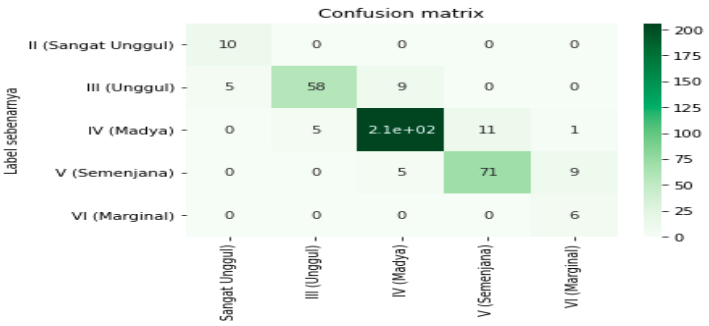
Gambar 4 diatas menunjukkan hasil visualisasi *Confusion Matrix*. Setelah diketahui *Confusion Matrix* dari model yang dibuat selanjutnya dilakukan perhitungan nilai akurasi dari model klasifikasi Naive Bayes yang telah dibuat menggunakan sklearn.metrics dengan mengimport *accuracy score* yang disediakan oleh library scikit.learn. Didapatkan hasil dari perhitungan matrix dengan kode program python yaitu sebesar 0.76.

Tabel 3. Hasil Evaluasi Model Naïve Bayes 80 : 20

	precision	recall	f1-score	support
2	0,67	1,00	0,80	10
3	0,00	0,00	0,00	43
4	0,76	0,95	0,84	149
5	0,82	0,81	0,82	52
6	0,67	0,60	0,63	10
accuracy			0,76	264
macro avg	0,58	0,67	0,62	264
weighted avg	0,64	0,76	0,69	264

Adaboost salah satu metode boosting yang mampu menyeimbangkan kelas dengan memberikan bobot pada tingkat error klasifikasi yang dapat merubah distribusi data [12]. Pada tabel 4 dibawah adalah kode yang digunakan untuk melakukan feature selection dengan AdaBoost.

Setelah model Naive Bayes dan Adaboost diperoleh nilainya maka selanjutnya dilakukan kembali splitting data dengan perbandingan data latih 70% dan data uji 30% yaitu sebagai berikut.



Gambar 5. Visualisasi model Naïve Bayes + AdaBoost 70 : 30

Gambar dibawah menunjukkan hasil sebesar 0.89 dengan melakukan kombinasi model *naive bayes dan feature selection adaboost*. Pada gambar 5 dibawah merupakan hasil perhitungan yang dilakukan.

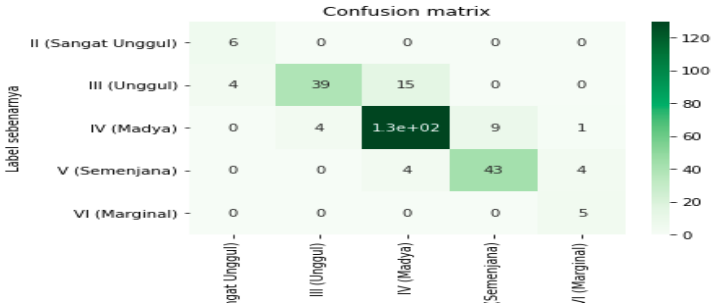
Tabel 5. Hasil Evaluasi Naive Bayes + Adaboost 70:30

	precision	**recall**	**f1-score**	**support**
2	0,67	1,00	0,80	10
3	0,92	0,81	0,86	72
4	0,94	0,92	0,93	223
5	0,87	0,84	0,85	85
6	0,38	1,00	0,55	6
accuracy			0,89	396
macro avg	0,75	0,91	0,80	396
weighted avg	0,90	0,89	0,80	396

Agar hasil model Naive Bayes dan Adaboost mendapatkan hasil yang maksimak maka dilakukan splitting data kembali dengan perbandingan data latih 80% dan data uji 20% yaitu sebagai berikut.

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956



Gambar 6. Visualisasi model Naïve Bayes + AdaBoost 80 : 20

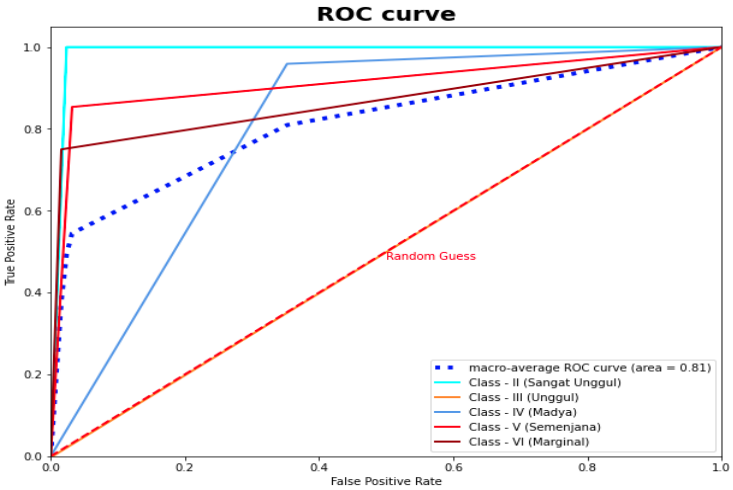
Hasil dari splitting data dengan melakukan kombinasi model naive bayes dan feature selection adaboost adalah sebesar 0.84. Pada gambar 4.11 dibawah merupakan hasil perhitungan yang dilakukan.

Tabel 6. Hasil Evaluasi Naive Bayes + Adaboost 80:20

	precision	recall	f1-score	support
2	0,60	1,00	0,75	6
3	0,91	0,67	0,77	58
4	0,87	0,90	0,89	144
5	0,83	0,84	0,83	51
6	0,50	1,00	0,67	5
accuracy			0,84	264
macro avg	0,74	0,88	0,78	264
weighted avg	0,86	0,84	0,84	264

Kurva ROC dibuat berdasarkan pada perhitungan dengan *Confusion Matrix* yaitu diantaranya *False Positive Rate* (FPR) dengan *True Positive Rate* (TPR).
Dimana :

1. *False Positive Rate* (FPR) = *False Positive* / (*False Positive* + *True Negative*)
2. *True Positive Rate* (TPR) = *True Positive* / (*True Positive* + *False Negative*)

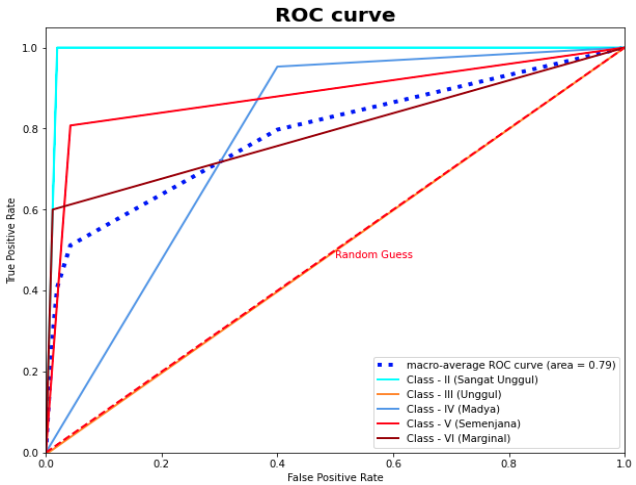


Gambar 7. Visualisasi Kurva ROC dan AUC 70 : 30

Pada gambar 7 diatas dapat dilihat 6 kurva, yaitu kurva putus-putus dengan warna merah terang, kurva dengan warna merah terang, kurva dengan warna merah gelap, kurva putus-putus dengan warna biru, kurva dengan warna biru dan kurva dengan warna biru muda. Dapat disimpulkan kinerja kurva berwarna biru muda lebih bagus dibandingkan kurva lainnya karena kurva ini yang paling mendekati titik 0,1. Kurva ROC dan AUC juga dilakukan splitting data dengan perbandingan data latih 80% dan data uji 20%. Pada gambar 8 berikut adalah kode python yang digunakan.

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956



Gambar 8. Visualisasi Kurva ROC dan AUC 80 : 20

Pada gambar 8 diatas dapat dilihat 6 kurva, dimana pada masing-masing warna ada keterangannya yaitu kurva putus-putus dengan warna merah terang, kurva dengan warna merah terang, kurva dengan warna merah gelap, kurva putus-putus dengan warna biru, kurva dengan warna biru dan kurva dengan warna biru muda. Dapat disimpulkan kinerja kurva berwarna biru muda lebih bagus dibandingkan kurva lainnya karena kurva ini yang paling mendekati titik 0,1. [17]. Selanjutnya dari hasil pengujian yang dilakukan maka diperoleh hasil penelitian seperti yang ditampilkan pada tabel 7 dibawah ini.

Tabel 7. Perbandingan Algoritma Naïve Bayes Berdasarkan Tingkat Akurasi

Splitting Data	Akurasi			
	Naïve Bayes	Naïve Bayes + AdaBoost	ROC Score Naïve Bayes	ROC Score Naïve Bayes + AdaBoost
70 : 30	0.78	0.89	0.81	0.86
80 : 20	0.76	0.84	0.79	0.85

4. KESIMPULAN

Hasil penelitian yang dilakukan dengan splitting data 70:30 hanya menggunakan model naive bayes mendapatkan nilai akurasi sebesar 78%, model naive bayes menggunakan feature selection adaboost memperoleh nilai akurasi sebesar 89%, kemudian model naive bayes dilakukan kombinasi dengan ROC Score memperoleh hasil akurasi sebesar 81%, dan dilakukan kombinasi model naive bayes dengan ROC Score dan adaboost memperoleh nilai akurasi 86%. Selanjutnya splitting data dengan 80:20 memperoleh nilai akurasi dengan model naive bayes sebesar 76%, model naive bayes dengan feature selection adaboost menghasilkan 84%, ROC Score dengan naive bayes menghasilkan nilai akurasi 79%, dan kombinasi ketiganya yaitu Naive Bayes, ROC Score dan Adaboost memperoleh nilai akurasi sebesar 85%.

REFERENCES

[1] S. Hudaa, "Peningkatan Keterampilan Berbahasa Indonesia Masyarakat Dengan Simulasi Tes UKBI," *SALINGKA, Maj. Ilm. Bhs. dan Sastra*, vol. 16, pp. 47–55, 2020.

[2] A. Suryadin, "Comparative Study Of Indonesian Language Skill Between PGSD And PJKR Students At STKIP Muhammadiyah Bangka Belitung," vol. 355, no. Pfeic, pp. 106–109, 2019.

[3] P. M. Pratama, "Peningkatan Kemahiran Berbahasa Indonesia melalui Program Klinik Bahasa UKBI Adaptif," vol. 7, no. 2, pp. 160–167, 2021.

[4] S. H. W. Admaja Dwi Herlambang, "Algoritma Naïve Bayes Untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif Di Smk Rumpun Teknologi Informasi Dan Komunikasi," vol. 6, no. 4, pp. 431–436, 2019, doi: 10.25126/jtiik.201961323.

[5] A. L. Shelly Janu Setyaning Tyas, Mita Febianah, Farkhatus Solikhah and W. A. A. Kamil, "Nalisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," vol. 8, no. 1, pp. 96–103, 2021.

[6] A. Saleh, "Penerapan Data Mining Dengan Metode Klasifikasi Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test (Studi Kasus : Universitas Potensi Utama)," *Konf. Nas. Sist. Informasi, Univ. Klabat,Manado,Indonesia, Vol. 2015*, no. June, pp. 1–6, 2015, [Online]. Available: https://www.researchgate.net/publication/304271255_PENERAPAN_DATA_MINING_DENGAN_METODE_KLASIFIKASI_NAIVE_BAYES_UNTUK_MEMPREDIKSI_KELULUSAN_MAHASISWA_DALAM_MENGIKUTI_ENGLISH_PROFICIENCY_TEST_Studi_Kasus_Universitas_Potensi_Utama.

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 6, Nomor 2, April 2022, Page 1183-1190
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v6i2.3956

[7] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
[8] B. Kurniawan, M. A. Fauzi, and A. W. Widodo, "Klasifikasi Berita Twitter Menggunakan Metode Improved Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 10, pp. 1193–1200, 2017.
[9] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
[10] I. Ardiyanto, "Introduction to Data Science and Machine Learning Tools : Python , Jupyter , and Google Colab."
[11] A. F. Cahyanti, "Penentuan Model Terbaik pada Metode Naive Bayes Classifier dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter," vol. 4, no. 1, pp. 28–35, 2015.
[12] A. Bisri, "Penerapan Adaboost untuk Penyelesaian Ketidakseimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode Decision Tree," vol. 1, no. 1, 2015.
[13] B. Wijonarko and T. Komputer, "PERBANDINGAN ALGORITMA DATA MINING NAIVE BAYES DAN BAYES," vol. 14, no. 1, pp. 21–26, 2020.
[14] G. F. Kelvin Hennry Loudry Malelak, I Made Dwi Ardiada, "IMPLEMENTASI KLASIFIKASI NAIVE BAYES DALAM MEMPREDIKSI LAMA DALAM MEMPREDIKSI LAMA STUDI MAHASISWA (STUDI KASUS : UNIVERSITAS DHYANA PURA)," *SINTECH*, vol. 4 No 2, no. October, pp. 202–209, 2021, doi: 10.31598/sintechjournal.v4i2.964.
[15] P. Terlaris and P. Penjualan, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 10, pp. 195–207, 2020.
[16] F. F. Zain, "Effectiveness of Naïve Bayes Weighted SVM Method in Movie Review Classification," vol. 5, no. 2, pp. 108–114, 2019.
[17] C. Dan and U. Klasifikasi, "ANALISIS DAN KOMPARASI ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI LOYALITAS PELANGGAN MNC PLAY KOTA SEMARANG," vol. 18, no. 2, pp. 161–172, 2021.

Submitted: **25/03/2022**; Accepted: **19/04/2022**; Published: **25/04/2022**