

ANALISIS SENTIMEN PADA *E-LEARNING* UNIVERSITAS XYZ MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE*

Jose Fernando¹⁾ dan Fathoni²⁾

^{1,2)} Fakultas Ilmu Komputer, Universitas Sriwijaya
Jalan Palembang-Prabumulih KM.32,Indralaya,Ogan Ilir, Sumatera Selatan
e-mail: joseyap123@gmail.com¹⁾, fathoni@unsri.ac.id²⁾

ABSTRAK

Pandemi covid-19 membuat para mahasiswa dan dosen harus melaksanakan belajar mengajar dari rumah. Maka dari itu, Universitas XYZ memfokuskan para mahasiswa untuk menggunakan *e-learning*. *E-learning* yang telah berjalan dan digunakan oleh para mahasiswa harus dievaluasi agar kegiatan belajar mengajar dapat berjalan dengan baik evaluasi dapat dilakukan dengan mengumpulkan opini berdasarkan fitur *e-learning* Universitas XYZ dari mahasiswa melalui kuesioner. Opini penilaian mahasiswa dapat dianalisis menggunakan metode klasifikasi yaitu metode *Naïve Bayes* dan *Support Vector Machine* sebagai perbandingan. Tahapan penelitian diawali pengumpulan data, *preprocessing* data, pelabelan dengan menggunakan *polarity*, menghitung frekuensi yang sering muncul dari per fitur *e-learning* nya, dan menghitung akurasi dari model *Complement Naïve Bayes* dan *Support Vector Machine*. Hasil penelitian yang dilakukan terhadap 1995 *dataset testing* menghasilkan opini mahasiswa yang bernilai positif sebanyak 1289 opini, bernilai negatif 372 opini, dan bernilai netral 364 opini. Diperkuat dengan perbandingan hasil akurasi model *Complement Naïve Bayes* dan *Support Vector Machine*. Dimana *Complement Naïve Bayes* sebesar 89%, *recall* sebesar 85,3%, *precision* sebesar 85%, dan *f1-score* sebesar 85%. Sedangkan hasil dari algoritma *Support Vector Machine* yang hasil akurasi model nya dibawah 11,1 % dari *Complement Naïve Bayes* yaitu hanya 74,46%. Hasil ini mengindikasikan bahwa dari 12 fitur *e-learning* Universitas XYZ, 8 fitur bernilai opini baik, 2 fitur memiliki opini kurang baik, dan 2 fitur memiliki opini netral.

Kata Kunci: *E-Learning*, Kuesioner, *Naïve Bayes*, Sentimen Analisis, *Support Vector Machine*

SENTIMENT ANALYSIS ON E-LEARNING UNIVERSITY XYZ WITH NAÏVE BAYES CLASSIFIER METHOD

Jose Fernando¹⁾ and Fathoni²⁾

^{1,2)}Computer Science Faculty, Sriwijaya University
Jalan Palembang-Prabumulih KM.32,Indralaya,Ogan Ilir, Sumatera Selatan
e-mail: joseyap123@gmail.com¹⁾, fathoni@unsri.ac.id²⁾

ABSTRACT

The covid-19 pandemic forced students and lecturers to carry out teaching and learning from home. Therefore, XYZ University focuses its students on using *e-learning*. *E-learning* that has been running and used by students must be evaluated, so that teaching and learning activities can run well. Evaluation can be done by collecting opinions based on the features of XYZ University *E-learning* on students through questionnaires. All opinions can be analyzed using classification method called *Naïve Bayes* and *Support Vector Machine* for comparison. The research started by collecting data, *preprocessing* data, labeling using *polarity*, calculating the frequency that often from each *e-learning* feature, and calculating the accuracy of the *Complement Naïve Bayes* model and *Support Vector Machine* model. The research results conducted on 1995 *dataset testing*, in student opinions with 1289 positive values, 372 negative values, and 364 neutral values. Reinforced by the comparison result of *Complement Naïve Bayes* and *Support Vector Machine*. When *Complement Naïve Bayes* model accuracy of 89%, *recall* 85,3%, and the *f1-score* 85%. While *Support Vector Machine* accuracy is lower 11,1% than *Complement Naïve Bayes* Model with only 74,4%. These results indicate that of the 12 features on XYZ University *E-learning*, 8 features have a good opinion, 2 feature have a bad opinion, and 2 feature have a neutral opinion.

Keywords: *E-Learning*, Sentiment Analysis, *Naïve Bayes*, Questionnaire, *Support Vector Machine*

I. PENDAHULUAN

SENTIMEN sentimen sering diungkapkan dari zaman dahulu. Ketidakadaannya internet membuat orang orang menyampaikan sentimen secara lisan ataupun tertulis. Misalnya dari sebuah tulisan dikertas,dan daun. Sentimen menjadi sebuah peranan penting dari aktivitas manusia, sentiment sering sekali terdapat pada organisasi untuk mengetahui konsumen. Saat adanya kampanye pemilihan pemerintahan juga dibutuhkan sentimen dari masyarakat. Bahkan, orang yang sedang dalam masalah akan meminta saran dari teman dan

keluarga secara langsung [1].

Berdasarkan Surat Edaran Nomor 15 Tahun 2020 Tentang Pedoman Penyelenggaraan Belajar Dari Rumah Dalam Masa Darurat Penyebaran *Corona Virus D/Sease* (COVID-19) menyatakan bahwa kegiatan belajar tetap dilaksanakan dengan metode belajar jarak jauh daring/luring dilaksanakan sesuai dengan pedoman surat edaran ini.[2]

Maka dari itu, pandemi *Covid-19* membuat kegiatan belajar mengajar menjadi terbatas. Para mahasiswa dan dosen menjadi tidak bisa bertemu secara tatap muka sehingga kegiatan belajar mengajar dilakukan menggunakan *E-learning*. Para mahasiswa di Universitas XYZ pun semua menggunakan E-Learning. Dengan kewajiban penggunaan *e-learning*, kualitas dan kesiapan dari *e-learning* itu sendiri sangat mempengaruhi hasil proses pembelajaran. Dalam penelitian [3], manfaat *e-learning* sangat banyak dimana para mahasiswa dapat mengakses materi walau tempatnya jauh dan mahasiswa dapat mengulang kembali materi yang menurut mereka kurang pahami. *E-learning* pada masa pandemi juga mengurangi beban biaya orang tua seperti kendaraan, uang saku, dan sebagainya. Sehingga kesiapan *e-learning* sangat diperlukan untuk para mahasiswa dan dosen. Karena jika tidak berjalan dengan baik maka kegiatan belajar akan terhambat. Maka dari itu, hal ini menjadi bagian motivasi untuk menganalisis opini dari para mahasiswa selama menggunakan *e-learning* Universitas XYZ. Pengalaman pribadi yang dialami oleh peneliti saat menggunakan *e-learning* XYZ seperti gangguan koneksi, lemot saat akses ke *e-learning*, tidak ada notifikasi absen, dan *bigbluebutton*(sejenis video conference) yang ada di Universitas XYZ sering terjadi gangguan.

Dalam upaya untuk mengetahui opini/sentimen para mahasiswa Universitas XYZ dilakukanlah survei dengan menggunakan kuesioner. Berdasarkan penelitian[4]. Kuesioner adalah cara pengumpulan data yang dilakukan dengan memberikan beberapa pertanyaan yang berhubungan dengan topik tertentu. Menurut Sugiyono (2017:142) kuesioner merupakan metode pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawabnya. Hasil kuesioner akan dianalisis apakah sentimen tersebut itu positif atau negatif dengan menggunakan metode *naive bayes*.

Naive bayes dikembangkan oleh ilmuwan yang bernama Thomas Bayes dari Inggris pada tahun 1702-1761. Thomas bayes sendiri belajar bagaimana cara mendistribusi parameter probabilitas dari distribusi binomial. metode *naive bayes* ini sederhana namun lebih efektif digunakan dalam hal melakukan proses klasifikasi, serta tingkat akurasi yang tinggi dalam klasifikasi teks[5][6]. *Naive bayes* juga terbukti cocok dengan database yang besar karena memiliki kecepatan dan akurasi tinggi[6].

Support vector machine adalah metode yang memiliki kemampuan generalisasi dalam mengklasifikasikan suatu *pattern*. *Support vector machine* juga merupakan metode dalam machine learning yang bekerja dengan prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *input space*. [7]

Penelitian mengenai analisis sentimen telah banyak dilakukan sebelumnya menggunakan algoritma klasifikasi. Adapun teknik klasifikasi yang digunakan meliputi *Naive Bayes* dan *Support vector machine* [8][9]. mengevaluasi menggunakan kuesioner sebagai media pengumpulan data [6] dan evaluasi terhadap fasilitas kampus menggunakan algoritma *support vector machine* [7]. Dengan membandingkan 2 metode klasifikasi yaitu *Naive Bayes* dan *Support Vector Machine*. Hasil yang didapat dapat dibandingkan dan dipilih mana yang terbaik untuk data opini mahasiswa universitas XYZ. Namun masih sedikit yang membahas mengenai penggunaan *e-learning* di kampus sehingga penelitian ini sangat cocok dilakukan dikarenakan juga masa pandemi sehingga penggunaan *e-learning* menjadi sangat tinggi.

II. PENELITIAN TERDAHULU

Metode *Naive Bayes Classifier* merupakan suatu metode klasifikasi statistik yang bisa digunakan untuk memprediksi sebuah probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel yang ada merupakan bagian dari kelas tertentu. *Naive Bayes Classification* adalah metode klasifikasi berdasarkan Teorema Bayes dengan asumsi independensi di antara para prediktor. *Naive Bayes Classifier* digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman yang ada sebelumnya oleh karena itu disebut Teorema Bayes [10]. Metode ini sangat cocok untuk data yang berjumlah kecil. Sedangkan *Support Vector Machine* adalah metode klasifikasi dan regresi yang umumnya dipakai untuk masalah *linear* dan *non linear*[11]. *SVM* juga memiliki kelebihan dengan data yang sedikit dapat membuat model klasifikasi yang baik[12].

Penelitian tentang aplikasi *fintech* Indonesia yang menggunakan *naive bayes* [8]. Disini data didapat dari review aplikasi *fintech* di Indonesia. Terdapat 2 data set yaitu A dan B. Data set A berisi review aplikasi berbahasa Indonesia dan inggris sedangkan data set B berisi review aplikasi Bahasa Indonesia. Didapat metode *naive bayes* cocok pada review aplikasi berbahasa indonesia dan berbahasa inggris.

Kemudian penelitian selanjutnya sama sama mengenai kepuasan namun sedikit berbeda objeknya. Pada penelitian yang diteliti adalah kepuasan mahasiswa pada layanan dan fasilitas kampus Universitas Telkom.

Metode pengambilan data juga diambil dengan kuesioner. Dalam penelitian ini menggunakan *Support Vector Machine* digunakan sebagai evaluasi terhadap fasilitas yang ada di kampus agar menjadi lebih baik lagi sehingga dapat meningkatkan kepuasan mahasiswa [7].

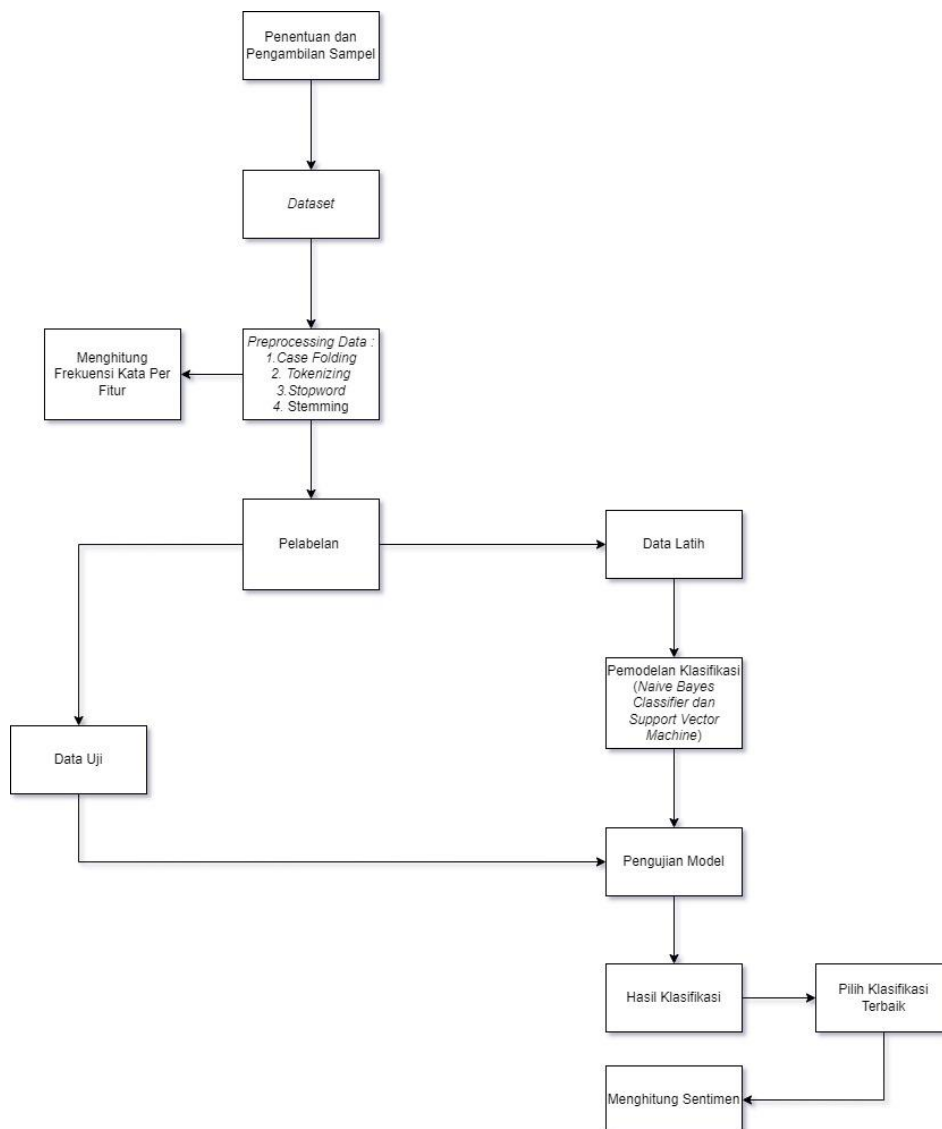
Penelitian selanjutnya adalah penelitian yang cukup mirip dengan menggunakan objek sistem daring pada masa pandemi, data diambil dengan kuesioner dan dianalisis dengan Metode *Naïve Bayes Classifier* dan *Lexicon Based*[6]. Dalam penelitian ini data diambil melalui kuesioner dengan memberikan 6 pertanyaan. Didapat sebanyak 265 jawaban sehingga diperoleh 1590 jawaban. Hasil pengujian didapat akurasi sebesar 53,8%. Hasil pengujian masih kurang baik. Penelitian selanjutnya adalah terdapat beberapa penelitian menggunakan kuesioner terhadap performa dosen namun, dengan 3 metode penelitian berbeda yaitu *Naïve Bayes Classifier*(NBC), *Support Vector Machine (SVM)*, dan *K-Means Clustering*. Dimana didapat masing masing akurasi adalah 73,9%, 82%, dan 88%. [5][13][14]

III. METODE PENELITIAN

A. Metode Penelitian

Metode penelitian yang dipakai dalam penelitian ini adalah metode kuantitatif yang mana penulis akan menyebarkan kuesioner yang digunakan untuk mengetahui sentimen positif dan negatif yang bisa dijadikan bahan evaluasi terhadap pengembangan *E-Learning* Universitas XYZ.

B. Desain Penelitian



Gambar 1. Tahapan Penelitian

1. Menentukan populasi dan sampel dalam penelitian dimana populasi adalah mahasiswa Universitas XYZ yang menggunakan *E-Learning* Universitas XYZ. Dan sampel diambil dengan metode *sample random sampling*.

2. Setelah data sampel didapatkan data sampel diubah menjadi file excel ataupun .csv yang akan menjadi dataset dari penelitian ini.
3. Kemudian dilakukan processing data untuk membuat teks agar menjadi lebih terstruktur yang meliputi 4 tahapan yaitu *case folding*, *tokenizing*, *stopword*, dan *stemming*.
4. Setelah dilakukan preprocessing akan dilakukan pengecekan frekuensi kata per fiturnya.
5. Kemudian dilakukan pelabelan dengan library python yaitu library Polarity.
6. Setelah selesai maka didapat data yang siap diujikan. Kemudian data tersebut dibagi menjadi 2 yaitu data latih (70%) dan data uji (30%).
7. Kemudian dilakukan pembuatan model klasifikasi *naïve bayes* dan *Support Vector Machine* dengan data latih agar hasil dari pelabelan semakin meyakinkan dari tingkat kebenaran perhitungannya.
8. Dengan model klasifikasi yang telah dibuat kemudian dilakukan penginputan data uji untuk menguji model yang telah dibuat sehingga didapat *accuracy*, *recall*, dan *precision*.
9. Setelah selesai dapat dilihat berapa jumlah data memiliki sentimen positif, netral, dan negatif saat model klasifikasi terbaik diuji. Positif adalah nilai kalimat yang memiliki kebenaran dan condong bernilai positif. Negatif memiliki nilai yang tidak benar dan kalimat condong negative. Apabila beban keduanya sama maka dari itu bersifat netral. Netral sendiri pada kalimat bersifat abu abu dan kurang jelas arahnya.

C. Model yang Digunakan

Model yang digunakan dalam penelitian ini adalah *naïve bayes classifier* dan *support vector machine*. Menurut penelitian [10] *naïve bayes classifier* dan *support vector machine* ini bertujuan untuk melakukan klasifikasi dengan mengidentifikasikan kata dari teks yang diberikan berdasarkan label yaitu positif, netral, dan juga negatif. Penulis sendiri menggunakan dua metode sekaligus ini dalam mengklasifikasi sentimen E-Learning yang diberikan para mahasiswa Universitas XYZ yang diambil dari kuesioner. Penulis akan melakukan pengujian untuk mendapatkan hasil *accuracy*, *recall*, dan *precision*. Serta memilih mana model yang lebih cocok terhadap data opini mahasiswa universitas XYZ.

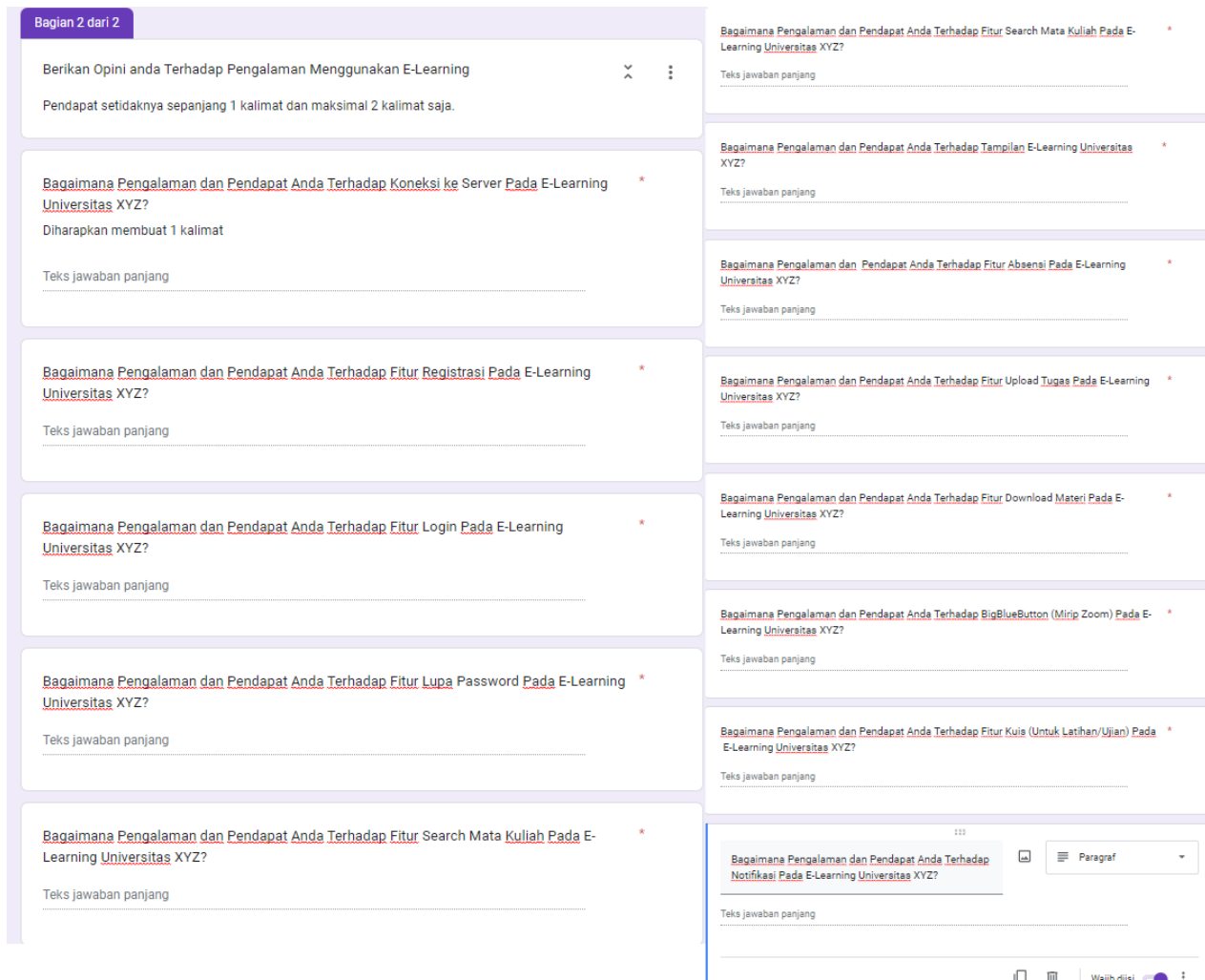
The image shows a digital questionnaire form with three main sections, each with a text input field for a short answer. The first section is titled 'Nama Lengkap' and has a text input field with the placeholder 'Teks jawaban singkat'. The second section is titled 'Jurusan *' and includes an example 'Contoh: Sistem Informasi' and a text input field with the placeholder 'Teks jawaban singkat'. The third section is titled 'Angkatan *' and features a list of radio button options for the years 2016, 2017, 2018, 2019, 2020, 2021, and 'Lainnya...'. At the bottom of the form, there is a navigation bar with the text 'Setelah bagian 1 Lanjutkan ke bagian berikut' and a right-pointing arrow.

Gambar 2 Kuesioner *Form* Pribadi.

IV. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Dalam pengumpulan data opini terhadap *e-learning* Universitas XYZ dimana peneliti tidak menggunakan *crawling* yang pada umumnya dilakukan saat ingin mencari sentiment. Dikarenakan data twitter yang sedikit, peneliti memutuskan untuk menggunakan kuesioner yang dibagikan kepada para mahasiswa Universitas XYZ. Kuesioner ini hanya memuat 3 data pribadi yaitu nama lengkap, jurusan, dan angkatan. Nama lengkap juga tidak harus diisi. Namun, untuk jurusan dan angkatan harus diisi.



Gambar 3 Kuesioner Pertanyaan Per Fitur

	A	B	C	D	E	F	G	H	I	J	K
1	Timestamp	Email Address	Nama Lengkap	NIM	Jurusan	Angkatan	Bagaimana Pengalaman	Bagaimana Pengalaman	Bagaimana Pengalaman	Bagaimana Pengalaman	Bagaimana Pengalaman
2	1/12/2022 16:55:00		ADDIEN PUTRA ARKANANTA		Sistem Informasi	2018	Baik	Belum pernah memaka	Baik	Baik namun pada bagi	Baik namun untuk bet
3	1/12/2022 16:55:33		Naberi Oktaria		Sistem informasi	2018	Sudah baik, dan tidak	Membantu	Karena tidak ada klesif	Mudah dimengerti dan	Sudah cukup baik
4	1/12/2022 17:07:51		Umi Perthwi		Sistem Informasi	2018	Pada fitur login sudah	Fitur lupa password su	Fitur search sudah mu	Tampilan elearning sud	Fitur absensi seabkn
5	1/12/2022 17:37:20		Dytha Ananda Widhiarso		Sistem Informasi	2018	Login cukup lancar, pa	Penanganan lupa pass	Fitur search cukup aki	Tampilan standar suatu	Absensi dilakukan de
6	1/12/2022 17:48:45		Radhaisyah pratwi		Sistem informasi	2019	Bagus	Bagus	Kurang bagus	Kurang bagus	Kurang efektif
7	1/12/2022 17:49:28		Shelly Putri		Sistem Informasi	2019	Baik	Baik	Form searchnya terliak	Baik	Baik
8	1/12/2022 18:09:05		Muhammad Fari Mukhlis		Sistem informasi	2019	sangat baik dan bisa d	iperantinya cukup mud	sangat kurang membar	cukup baik dan simpel	sedikit menyulitkan
9	1/12/2022 18:09:46		Chan		Sistem informasi	2018	Cukup mudah dipaham	Belum pernah menggu	Sangat bagus, mudah	(Ulinya terlihat kurang	b Absensi nya mudah d
10	1/12/2022 18:18:56		Ferzco Agni Melenio		Sistem Informasi	2018	fiturnya kurang menari	fiturnya bagus	fiturnya bagus dan cuk	tampilannya kurang me	fiturnya bagus dan an
11	1/12/2022 18:35:14		Khairunisa Adabbia		Sistem Informasi	2018	Sedikit kebingungan se	Belum pernah menggu	Tidak memunculkan ha	Tampilan mobilnya pe	Sejauh ini lancar, me
12	1/12/2022 18:47:28		Rahma Destriani		Sistem Informasi	2018	sudah baik	sudah baik	Fitur ini sudah bagus	Tampilannya sedang si	Absensi sangat ketat
13	1/12/2022 18:56:54		Fenia		Farmasi	2018	Cukup baik	Cukup baik	Mudah diakses	Cukup baik	Cukup baik
14	1/12/2022 18:58:18		Muhamad Riza Padlefi		Sistem Informasi	2019	Fitur Login menurut sa	Sudah bagus	Fiturnya sudah bagus	Tampilannya sedang si	Absensi sangat ketat
15	1/12/2022 19:00:06		Miftahurrohman Haque		Sistem Informasi	2018	Akan lebih baik jika isi	Belum pernah mencob	Ada beberapa mata ku	Sudah lumayan baik, ti	Baik
16	1/12/2022 19:03:19		Siti Raisah Adilah		Sistem Informasi	2018	Sangat bagus	tidak ada kesulitan	Tidak ada kesulitan	Cukup baik	Tidak ada kesulitan
17	1/12/2022 19:10:05		M. Agutri Dihan Prawira		Sistem Informasi	2018	Sangat baik	Sejauh ini belum perna	Sudah cukup baik	cukup baik dan mudah	Sudah cukup baik, nar
18	1/12/2022 19:13:44		A Wendi Saputra		Sistem Informasi Biling	2018	Mudah untuk di paham	Cukup membantu	Sangat membantu	Sudah cukup baik	Cukup baik, alangkah
19	1/12/2022 19:16:23		Tiara Dewangga		Sistem Informasi	2018	Sangat mudah digunak	Mungkin cukup mudah	Mudah digunakan, pad	Cukup baik.	Bagus dan mudah dig
20	1/12/2022 19:19:04		Sisilia Pratama		Teknik Kimia	2018	Baik	Baik	Baik	Cukup menarik	Cukup baik
21	1/12/2022 19:25:57		Muhammad Naufal Al Hafif		Teknik Informatika	2017	Cukup	Cukup	Baik	Baik	Baik
22	1/12/2022 19:54:46		panji arya kusuma		sistem informasi prodi	2018	cukup baik	menunggu verifikasi ya	sudah cukup baik	tampilan sangat tidak u	cukup baik walaupun
23	1/12/2022 20:43:04		Cindy Ardella		Sistem Informasi	2019	Sudah sangat baik	Belum pernah menggu	Fitur search tidak mem	Baik dan mudah dimen	Baik
24	1/12/2022 21:00:20		Nicole Jovanka Kristalisia		Teknik Kimia	2018	Menurut saya, akan let	Tidak ada kendala.	Menurut saya sudah c	Menurut saya sudah c	Menurut saya sudah c
25	1/12/2022 21:41:31		Muhammad amin abimayu		Teknikpertambangan	2018	Lumayan	Lumayan	Lumayan	Lumayan	Lumayan
26	1/12/2022 21:48:06		Christopher Marlo		Sistem Komputer	2018	Biasanya jika standby	Kurang membantu kar	Cukup baik namun tid	Cukup menarik dan sin	Pada dasarnya baik,
27	1/12/2022 22:05:21		M. Afat Hayatur Rizon		Sistem komputer	2018	Kurang menarik	Ribet di karena kan ha	Bagus	Sering down pada saat	Harus bisa di atur wal
28	1/12/2022 22:05:39		Julian aputra		Sistem Informasi	2018	Baik, tidak ada kendala	Tidak memiliki pengalai	Kurang baik, karena m	Cukup baik, tidak ada	Fitur absensi cukup b
29	1/12/2022 22:14:57		Nadva Anoraini		Sistem Informasi	2018	Cukup mudah dan lanc	Saat menalatkan fitur	Fitur ini ada di halam	Sudah cukup user frier	Sudah baik, bisa men

Gambar 4 Hasil Excel dari Opini Mahasiswa

Kuesioner terbagi menjadi 12 fitur yang akan diberikan sentimen oleh para mahasiswa, yang dimana 12 fitur tersebut adalah mengenai *login*, lupa password, *search* mata kuliah, tampilan, absensi, *upload*, *download*, *bigbluebutton*, notifikasi, kuis, registrasi, dan koneksi ke server.

2. Preprocessing Data

Preprocessing data dilakukan setelah mendapatkan data sentimen dimana dibagi menjadi 4 tahap yaitu *case folding*, *tokenizing*, *stopword*, dan *stemming*.

1. Case Folding

Data yang telah didapat dilakukan *case folding* yang bertujuan untuk mengubah huruf menjadi huruf kecil. *Case folding* juga menganggap karakter lain seperti tanda baca dan spasi dianggap sebagai delimiter. Delimiter adalah karakter karakter yang bertujuan untuk memisahkan teks.

Tabel I
Contoh potongan kode proses *case folding*

```

1 def case_folding(opini):
2     opini = str(opini)
3     opini = opini.lower()
4     return opini
5
6 text_df['Case Folding'] = text_df['Text'].apply(case_folding)
7 print('Hasil Case Folding : \n')
8 print(text_df['Case Folding'].head(10))
9 print('\n\n\n')
    
```

A	B
	Case Folding
0	baik
1	sudah baik, dan tidak ada kendala selama ini
2	pada fitur login sudah baik dan mudah juga untuk digunakan. namun dengan adanya pembagian fitur per tahun ajaran sering membuat mahasiswa salah dalam menekan fitur login sesuai dengan semesternya
3	login cukup lancar, password dapat di saved dan ada fitur keep me logged in
4	bagus
5	baik
6	sangat baik dan bisa di save
7	cukup mudah dipahami, tetapi uinya kurang rapi
8	fiturnya kurang menarik
9	sedikit kebingungan saat login dibagi pertahun ajaran
10	sudah baik
11	cukup baik
12	fitur login menurut saya sudah bagus, saran saya fitur session yang memungkinkan pengguna untuk tetap login pada waktu yang ditentukan agar ditambah.
13	akan lebih baik jika isian data username dan password diletakkan di tengah
14	sangat bagus
15	sangat baik
16	mudah untuk di pahami
17	sangat mudah digunakan.
18	baik
19	cukup
20	cukup baik
21	sudah sangat baik
22	menurut saya, akan lebih baik jika kita hanya login sekali saja dalam 1 device sehingga ketika buru" atau terlalu lama standby tidak perlu login kembali untuk menghemat waktu.
23	lumayan
24	biasanya jika standby lebih dari 2 jam maka logout dengan sendirinya, cukup mengganggu.
25	kurang menarik
26	baik, tidak ada kendala dan fitur login masih wajar tidak ada keanehan
27	cukup mudah dan lancar dalam melakukan login, sudah baik karena jika kita meninggalkan halaman dalam waktu yang lama, sistem akan otomatis log out dan meminta untuk melakukan login ulang, hal ini menu
28	kurang sederhana
29	fitur login ada kalanya terdapat gangguan sehingga akan lama loading

Gambar 5 Hasil *Case Folding* di Excel
Tabel II

Contoh potongan proses *tokenizing*

```

1 def tokenizing(opini):
2     opini = re.sub(r"s+", " ", str(opini))
3     opini = re.sub(r"d+", " ", str(opini))
4     opini = re.sub('&quot;', ' ', opini)
5     opini = re.sub(r"b[a-zA-Z]b", "", str(opini))
6     opini = re.sub(r"^[w\s]", " ", str(opini))
7     return opini
8
9 text_df['Hasil Token Sebelum Pisah Kalimat'] = text_df['Case Folding'].apply(tokenizing)
10 text_df['Hasil Tokenizing'] = text_df['Hasil Token Sebelum Pisah Kalimat'].apply(lambda x: x.split())
11 print(text_df['Hasil Tokenizing'].head(10))
    
```

Gambar 6 Hasil *Tokenizing* di Excel.

2. *Tokenizing*

Huruf yang telah diubah menjadi huruf kecil dalam tahap *Case Folding* dilanjutkan dengan melakukan *Tokenizing*. *Tokenizing* dilakukan dengan memisahkan kata per kata dari sebuah kalimat. Dalam tahap ini juga dilakukan penghapusan angka, spasi ganda menjadi satu spasi, menghapus simbol simbol, menghapus huruf yang hanya memiliki 1 karakter, menghapus tanda baca. Hasil akhir *tokenizing* adalah kalimat yang telah terpisah pisah menjadi satu kata.

3. *Normalisasi*

Normalisasi hanyalah tahapan tambahan dimana sentimen sentimen yang diberikan mahasiswa tidak semua kata katanya sempurna dan pasti memiliki kata yang disingkat ataupun *typo*. *Normalisasi* kata dapat meminimalisir kata kata yang *typo* ataupun disingkat dimana akan sangat membantu dikarenakan apabila kata yang dikeluarkan *typo* atau disingkat akan diubah menjadi kata yang lebih jelas. Sehingga saat *stopword* dan *stemming* akan menndapatkan hasil yang lebih baik.

4. *Stopwords*

Kalimat yang telah dipisah pisah dan dinormalisasi, akan dilakukan *stopwords* dimana didalam tahapan ini dilakukan penghapusan kata kata yang tidak penting. Kata kata yang tidak penting seperti kata penghubung yaitu di,atau,akan, dan lain lain. *Stopwords* ini digunakan agar saat melakukan *running* contohnya terhadap pelabelan hasil akan menjadi lebih cepat. *Stopword* yang saya gunakan diambil dari *library* *nltk.corpus* yang berbahasa Indonesia. Dimana dalam kamus *corpus stopwords* peneliti menambahkan dan mengurangi isi dalam kamus tersebut dikarenakan didalam kamus tersebut terdapat kata kata yang mengubah makna dari opini seseorang misalnya seorang mahasiswa berpendapat “*e-learning* Universitas XYZ baik”. Namun dalam kamus tersebut terdapat kata “baik”, apabila kata baik itu tidak dihilangkan . Maka kata “baik” akan hilang dari opini dari mahasiswa tersebut sehingga makna dari “*e-learning* Universitas XYZ baik” menjadi tidak memiliki makna karena kalimat yang terbaca hanya “*e-learning* Universitas XYZ”.

Tabel III
Contoh potongan proses *Stopword*

```

1 normalizad_word = pd.read_csv("normalisasi.csv")
2 normalizad_word_dict = { }
3
4 for index, row in normalizad_word.iterrows():
5     if row[0] not in normalizad_word_dict:
6         normalizad_word_dict[row[0]] = row[1]
7
8 def normalized_term(document):
9     return [normalizad_word_dict[term] if term in normalizad_word_dict else term for term in document]
10
11 text_df['Hasil Normalisasi'] = text_df['Hasil Tokenizing'].apply(normalized_term)
12 text_df['Hasil Normalisasi'].head(10)

```

E	
Hasil Normalisasi	
['baik']	
['sudah', 'baik', 'dan', 'tidak', 'ada', 'kendala', 'selama', 'ini']	
['pada', 'fitur', 'login', 'sudah', 'baik', 'dan', 'mudah', 'juga', 'untuk', 'digunakan', 'namun', 'dengan', 'adanya', 'pembagian', 'fitur', 'per', 'tahun', 'ajaran', 'sering']	
['login', 'cukup', 'lancar', 'password', 'dapat', 'di', 'saved', 'dan', 'ada', 'fitur', 'keep', 'me', 'logged', 'in']	
['bagus']	
['baik']	
['sangat', 'baik', 'dan', 'bisa', 'di', 'save']	
['cukup', 'mudah', 'dipahami', 'tetapi', 'user interface', 'kurang', 'rapi']	
['fitur', 'kurang', 'menarik']	
['sedikit', 'kebingungan', 'saat', 'login', 'dibagi', 'pertahun', 'ajaran']	
['sudah', 'baik']	
['cukup', 'baik']	
['fitur', 'login', 'menurut', 'saya', 'sudah', 'bagus', 'saran', 'saya', 'fitur', 'session', 'yang', 'memungkinkan', 'pengguna', 'untuk', 'tetap', 'login', 'pada', 'waktu', 'y']	
['akan', 'lebih', 'baik', 'jika', 'isian', 'data', 'username', 'dan', 'password', 'diletakkan', 'di', 'tengah']	
['sangat', 'bagus']	
['sangat', 'baik']	
['mudah', 'untuk', 'di', 'pahami']	
['sangat', 'mudah', 'digunakan']	
['baik']	
['cukup']	
['cukup', 'baik']	

Gambar 7 Hasil stopword di Excel.

Tabel IV
Contoh potongan proses stopwords

```

1 list_stopwords = stopwords.words('indonesian')
2 list_stopwords.extend(['wkwkwk', 'nya', 'was', 'da', 'an', 'sih', 'universitasindonesia',
3     'amerikaserikat', 'aja', 'by', 'ncaaaa', 'peng', 'ter', 'si', 'save',
4     'nya', 'dansebagainya', 'meng', 'danlainlain', 'so', 'far'])
5 list_stopwords.remove('tidak')
6 list_stopwords.remove('baik')
7 list_stopwords.remove('kurang')
8 list_stopwords.remove('belum')
9 list_stopwords.remove('ada')
10 list_stopwords.remove('lama')
11 list_stopwords.remove('cukup')
12 list_stopwords.remove('biasa')
13 list_stopwords.remove('masalah')
14
15 list_stopwords = set(list_stopwords)
16 def stopwords_removal(words):
17     return [word for word in words if word not in list_stopwords]
18
19 text_df['Hasil Stopword'] = text_df['Hasil Normalisasi'].apply(stopwords_removal)
20 print(text_df['Hasil Stopword'].head(10))
    
```

F	
Hasil Stopword	
['baik']	
['baik', 'tidak', 'ada', 'kendala']	
['fitur', 'login', 'baik', 'mudah', 'pembagian', 'fitur', 'ajaran', 'mahasiswa', 'salah', 'menekan', 'fitur', 'login', 'sesuai', 'semesternya']	
['login', 'cukup', 'lancar', 'password', 'saved', 'ada', 'fitur', 'keep', 'me', 'logged', 'in']	
['bagus']	
['baik']	
['baik']	
['cukup', 'mudah', 'dipahami', 'user interface', 'kurang', 'rapi']	
['fitur', 'kurang', 'menarik']	
['kebingungan', 'login', 'dibagi', 'pertahun', 'ajaran']	
['baik']	
['cukup', 'baik']	
['fitur', 'login', 'bagus', 'saran', 'fitur', 'session', 'pengguna', 'login', 'ditentukan', 'ditambah']	
['baik', 'isian', 'data', 'username', 'password', 'diletakkan']	
['bagus']	
['baik']	
['mudah', 'pahami']	
['mudah']	
['baik']	
['cukup']	
['cukup', 'baik']	

Gambar 8 Hasil stopwords di Excel.

5. *Stemming*

Stemming dapat dilakukan ketika kata/opini telah dinormalisasi sehingga kata dasar akan mudah didapatkan. *Stemming* sendiri bertujuan untuk menjadikan kata yang memiliki *suffix* dan *prefix* Kembali ke kata dasar. *Stemming* adalah tahap terakhir dari *preprocessing* sehingga kata/opini sudah bersih dan dilakukan pelabelan.

Tabel V
Contoh potongan proses *Stemming*

```

1 import swifter
2 factory = StemmerFactory()
3 stemmer = factory.create_stemmer()
4
5 def stemmed_wrapper(term):
6     return stemmer.stem(term)
7
8 term_dict = {}
9
10 for document in text_df['Hasil Stopword']:
11     for term in document:
12         if term not in term_dict:
13             term_dict[term] = ''
14
15 print(len(term_dict))
16 print("-----")
17
18 for term in term_dict:
19     term_dict[term] = stemmed_wrapper(term)
20     print(term,":",term_dict[term])
21
22 print(term_dict)
23 print("-----")
24
25 def get_stemmed_term(document):
26     return [term_dict[term] for term in document]
27
28 text_df['Hasil Stemming'] = text_df['Hasil Stopword'].swifter.apply(get_stemmed_term)
29 print(text_df['Hasil Stemming'])
    
```

Hasil Stemming	
'baik'	
'baik', 'tidak', 'ada', 'kendala'	
'fitur', 'login', 'baik', 'mudah', 'bagi', 'fitur', 'ajar', 'mahasiswa', 'salah', 'tekan', 'fitur', 'login', 'sesuai', 'semester'	
'login', 'cukup', 'lancar', 'password', 'saved', 'ada', 'fitur', 'keep', 'me', 'logged', 'in'	
'bagus'	
'baik'	
'baik'	
'cukup', 'mudah', 'paham', 'user interface', 'kurang', 'rapi'	
'fitur', 'kurang', 'tarik'	
'bingung', 'login', 'bagi', 'tahun', 'ajar'	
'baik'	
'cukup', 'baik'	
'fitur', 'login', 'bagus', 'saran', 'fitur', 'session', 'guna', 'login', 'tentu', 'tambah'	
'baik', 'isi', 'data', 'username', 'password', 'letak'	
'bagus'	
'baik'	
'mudah', 'paham'	
'mudah'	
'baik'	
'cukup'	
'cukup', 'baik'	
'baik'	
'baik', 'login', 'device', 'buru', 'lama', 'standby', 'tidak', 'login', 'hemat'	
'lumayan'	
'standby', 'jam', 'logout', 'cukup', 'ganggu'	
'kurang', 'tarik'	
'baik', 'tidak', 'ada', 'kendala', 'fitur', 'login', 'wajar', 'tidak', 'ada', 'aneh'	
'cukup', 'mudah', 'lancar', 'login', 'baik', 'tinggal', 'halaman', 'lama', 'sistem', 'otomatis', 'log', 'out', 'login', 'ulang', 'tingkat', 'aman', 'akun', 'guna'	
'kurang', 'sederhana'	
'fitur', 'login', 'ada', 'kala', 'ganggu', 'lama', 'loading'	

Gambar 9 Hasil *stemming* di Excel.

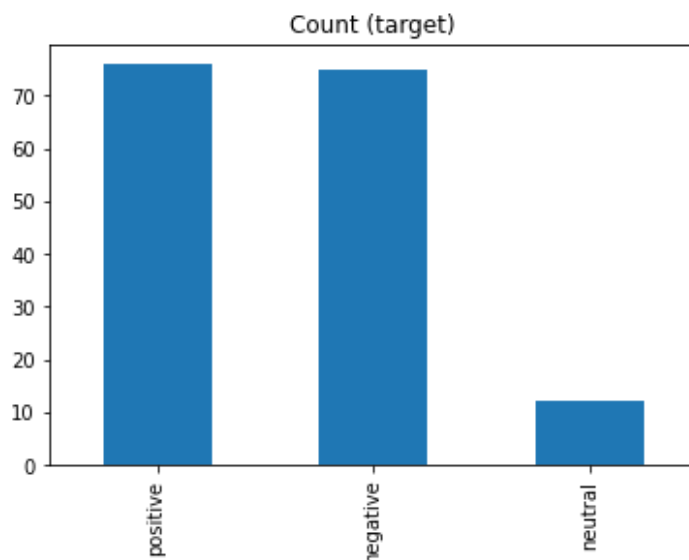
3. Frekuensi Kata dan *Wordcloud*

Frekuensi kata dan *Wordcloud* bertujuan untuk mengetahui kata apa yang sering muncul dan dapat dinilai apabila kata positif yang sering muncul artinya sentimen/opini terhadap sesuatu itu positif dan sebaliknya. *Wordcloud* sendiri merupakan penggambaran kata dimana kata yang terbanyak akan digambarkan paling besar sedang kata kata yang sedikit akan semakin mengecil.

Tabel VI
Contoh potongan total positif, negatif, dan netral dari fitur koneksi

```

1 target_count = df['Label'].value_counts()
2 target_count.plot(kind='bar', title='Count (target)')
3 target_count
    
```



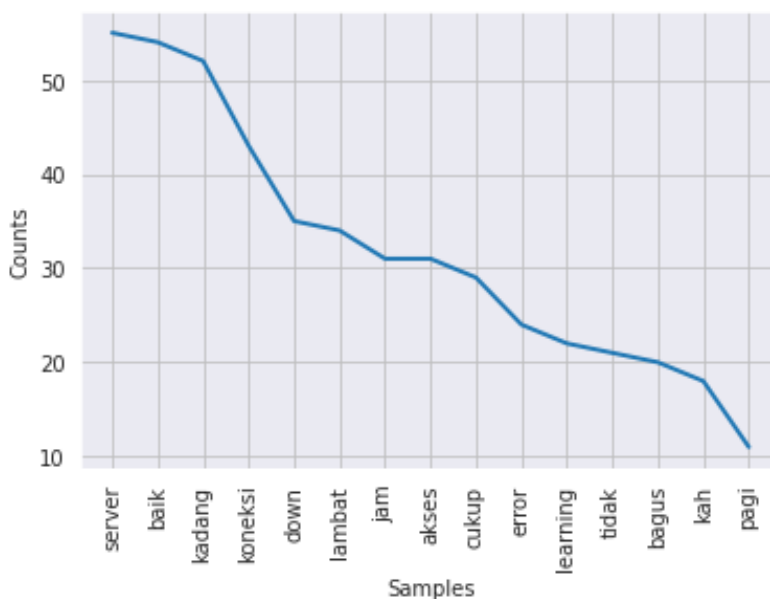
Gambar 10 Jumlah Positif, Negatif, dan Netral dari Fitur Koneksi

Tabel VII
Contoh potongan proses perhitungan koneksi

```

1 n_most_frequent_words = 100
2 opini_word_frequency = opini_stack.value_counts()
3 opini_word_frequency.head(n_most_frequent_words)

4 sns.set_style('darkgrid')
5 nlp_words=nlk.FreqDist(opini_stack)
6 nlp_words.plot(15);
    
```



Gambar 11 Grafik Frekuensi Kata yang Sering Muncul Pada Fitur Koneksi

Tabel VIII
Contoh potongan kode proses pembuat *wordcloud*

```

1 stopwords = set(STOPWORDS)
2 wordcloud = WordCloud(stopwords = stopwords , width=1600 ,
3 height=800,background_color="White",colormap="Set2").generate('.'.join(df["Text"]))
4 plt.figure(figsize=(20,10),facecolor='k')
5 plt.imshow(wordcloud,interpolation='bilinear')
6 plt.axis('off')
7 plt.tight_layout(pad=0)
    
```



Gambar 12 *Wordcloud* Fitur Koneksi

4. Pelabelan

Pelabelan digunakan untuk menilai suatu kalimat itu positif, negatif, ataupun netral. Pelabelan dilakukan dengan menggunakan *splitting* terhadap kalimat dikarenakan per kata akan dinilai satu per satu. *Library* yang peneliti gunakan adalah *library polarity*. *Library Polarity* lebih mendukung bahasa inggris sehingga sebelum melakukan penilaian positif, negatif, dan netral, peneliti melakukan *translate* file ke bahasa inggris untuk mendapatkan hasil yang lebih baik. Untuk *translate* sendiri penulis menggunakan *library TextBlob*.

5. Pemodelan Klasifikasi

A. *Naïve Bayes Classifier*

Tahapan selanjutnya adalah tahapan *training* dan *testing*. Dimana dari hasil *training* dan *testing* akan didapat hasil *recall*, *accuracy*, dan *precision*. *Recall* adalah sensitifitas dimana merupakan sebuah rasio prediksi *true positive* dengan keseluruhan data yang sebenarnya positif. *Accuracy* adalah prediksi benar (positif, netral, dan negatif) dibanding keseluruhan data. *Precision* merupakan rasio dari prediksi benar positif yang dibandingkan dengan hasil keseluruhan yang diprediksi positif. Dalam menjalankan *naïve bayes classifier* ini peneliti menggunakan *library sklearn*. *Library* ini bertujuan untuk membuat model *machine learning*.

Tabel IX
Contoh potongan proses pelabelan

```

1 data2=[]
2 for text in df['Text1']:
3     try:
4         text = ' '.join(text)
5         analysis= TextBlob(text)
6         an = analysis.translate(from_lang='id', to='en')
7         print(text)
8         print(an.sentiment)
9         polarity = 'neutral'
10        if(an.sentiment.polarity > 0.0):
11            polarity = 'positive'
12        if(an.sentiment.polarity < 0.0):
13            polarity = 'negative'
14        res = [text,polarity]
15        data2.append(res)
16        df=pd.DataFrame(data2)
17        df.head()
18    except:
19        pass
    
```

	A	B	C
1		0	1
2	0	baik	positive
3	1	baik tidak ada kendala	positive
4	2	fitur login baik mudah bagi fitur ajar mahasiswa salah	positive
5	3	login cukup lancar password saved ada fitur keep me	positive
6	4	bagus	positive
7	5	baik	positive
8	6	baik	positive
9	7	cukup mudah paham user interface kurang rapi	positive
10	8	fitur kurang tarik	negative
11	9	bingung login bagi tahun ajar	negative
12	10	baik	positive
13	11	cukup baik	positive
14	12	fitur login bagus saran fitur session guna login tentu t	positive
15	13	baik isi data username password letak	positive
16	14	bagus	positive
17	15	baik	positive
18	16	mudah paham	positive
19	17	mudah	positive
20	18	baik	positive
21	19	cukup	neutral
22	20	cukup baik	positive
23	21	baik	positive
24	22	baik login device buru lama standby tidak login hemat	positive
25	23	lumayan	positive
26	24	standby jam logout cukup ganggu	negative
27	25	kurang tarik	neutral
28	26	baik tidak ada kendala fitur login wajar tidak ada aneh	positive
29	27	cukup mudah lancar login baik tinggal halaman lama s	positive
30	28	kurang sederhana	negative
31	29	fitur login ada kala ganggu lama loading	negative

Gambar 13 Hasil Pelabelan di Excel.

Kemudian dilakukan *splitting* data dimana data dibagi menjadi data *training* dan data uji/*testing*. Peneliti membuat pembagian 70 % data *training* dan 30% data uji/*testing*. Kemudian melakukan import algoritma *Naïve Bayes* yaitu *ComplementNB*. Alasan menggunakan *Complement Naïve Bayes* adalah algoritma ini cocok untuk data yang tidak seimbang, dimana dapat dilihat dari jumlah total opini positif yang sangat tinggi dibanding netral dan negatif. Data training yang sudah displit dijadikan vector dengan bantuan *countvectorize*. Untuk mendapatkan sebuah model dilakukan training dengan *Complement Naïve Bayes*.

Tabel X
Contoh potongan kode *training* data *Complement Naïve Bayes*

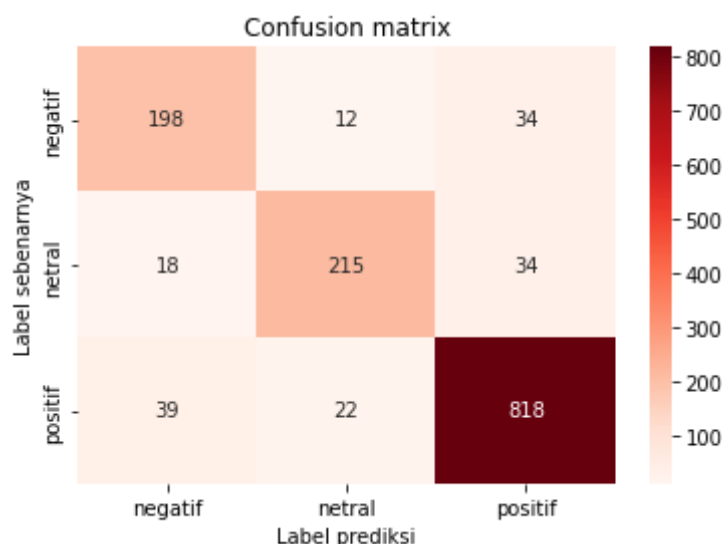
```

1 X = list(df['0'].values.astype(str))
2 y = np.array(list(df['1']))
3 X_train, X_test, y_train, y_test = train_test_split(X,y, random_state=1973, test_size=0.3)
4 count_vect = CountVectorizer(ngram_range=(1,1))
5 X_train_c = count_vect.fit_transform(X_train)
6 tfidf = TfidfTransformer()
7 X_train_tfidf = tfidf.fit_transform(X_train_c)
8 from sklearn.naive_bayes import ComplementNB
9 cnb_model = ComplementNB().fit(X_train_c,y_train)
10 y_train_pred = cnb_model.predict(X_train_c)
11 print('classification report')
12 print(classification_report(y_train,y_train_pred))
13 print('\n')
14 print('confusion matrix')
15 print(confusion_matrix(y_train,y_train_pred))
16 print('\n')
17 print('classification report')
18 print(accuracy_score(y_train,y_train_pred))
    
```

Tabel XI
Classification Report Naïve Bayes Classifier

	<i>Precision</i>	<i>Recall</i>	<i>f-1 score</i>	<i>Support</i>
Negatif	0.78	0.81	0.79	244
Netral	0.86	0.81	0.83	267
Positif	0.92	0.93	0.93	879
Akurasi			0.89	1390
<i>Macro avg</i>	0.85	0.85	0.85	1389
<i>Weighted avg</i>	0.89	0.89	0.89	1390

Didapat hasil dari *accuracy* dari model ini adalah sebesar 0,89 atau 89%. Dengan *precision* dari “negatif” sebesar 78%, “netral” 86%, dan “positif” 92%. Kemudian pada *recall* “negatif” sebesar 81%, “netral” sebesar 81%, dan “positif” 93%. Dari hasil ini dapat terlihat gambaran bahwa yang paling baik dalam kinerja model adalah positif dengan 93%, kedua netral dengan 83%, dan kinerja paling rendah ada di negative dengan 79%.



Gambar 14 Confusion Matrix Naïve Bayes Classifier.

Dalam gambar *Confusion Matrix* ditunjukkan kinerja dari model tersebut. Dari model ini mengklasifikasikan dengan benar positif 818 opini, 215 netral, dan 198 opini negatif. Model juga mengklasifikasikan salah positif hanya 39 opini yang dikira negatif dan 22 opini yang dikira netral. Sedangkan pada pengklasifikasian netral yang salah 18 opini dikira negatif dan 35 opini dikira positif. Pada pengklasifikasian yang negatif 12 opini dianggap sebagai netral dan sebanyak 34 opini dianggap positif.

B. Support Vector Machine

Tahapan dalam *support vector machine* hampir sama yaitu *mengimport library sklearn*. Kemudian dilakukan *splitting* data dimana data dibagi menjadi data *training* dan data uji/*testing*. Peneliti membuat pembagian 70 % data *training* dan 30% data uji/*testing*. Kemudian melakukan import algoritma *Support Vector Machine* yaitu *svm*. Data *training* yang sudah displit dijadikan vector dengan bantuan *countvectorize*. Untuk mendapatkan sebuah model dilakukan *training* dengan *Support Vector Machine* dalam pengujian ini penulis menggunakan kernel ‘linear’.

Didapat hasil dari *accuracy* dari model ini adalah sebesar 0,74 atau 74%. Dengan *precision* dari “negatif” sebesar 82%, “netral” 46%, dan “positif” 98%. Kemudian pada *recall* “negatif” sebesar 60%, “netral” sebesar 98%, dan “positif” 71%. Dari hasil ini dapat terlihat gambaran bahwa yang paling baik dalam kinerja model adalah positif dengan 83%, kedua negatif dengan 69%, dan kinerja paling rendah ada di netral dengan 62%.

Tabel XII
Contoh potongan kode training data *Support Vector Machine*

```

1 x = list(df['0'].values.astype(str))
2 y = np.array(list(df['1']))
3 x_train1, x_test1, y_train1, y_test1 = train_test_split(x,y, random_state=1973, test_size=0.3)

4 count_vect = CountVectorizer(ngram_range=(1,1))
5 X_train_tfidf = count_vect.fit_transform(x_train1)

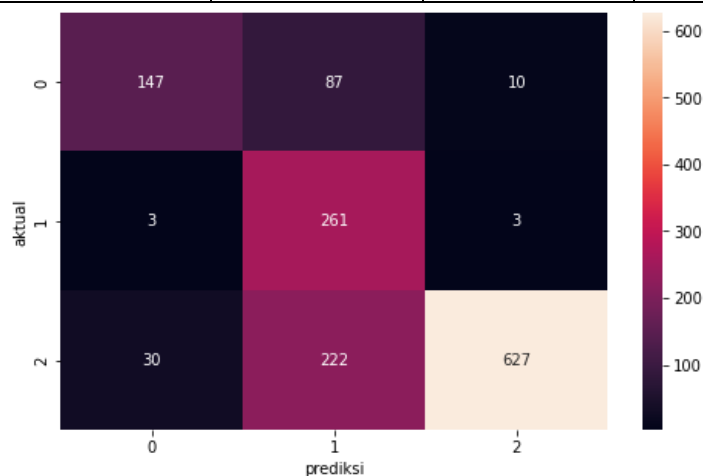
6 tfidf = TfidfTransformer()
7 X_train_tfidfc = tfidf.fit_transform(X_train_tfidf)

8 from sklearn import svm
9 SVM = svm.SVC(kernel='linear')
10 SVM.fit(X_train_tfidf, y_train1)
11 predictions_SVM = SVM.predict(X_train_tfidfc)

12 print('Confusion Matrix : ')
13 print(confusion_matrix)
14 print("\n Report Hasil : ")
15 print(classification_report(y_train1, predictions_SVM))
    
```

Tabel XIII
Classification Report Support Vector Machine

	<i>Precision</i>	<i>Recall</i>	<i>f-1 score</i>	<i>Support</i>
Negatif	0.82	0.60	0.69	244
Netral	0.46	0.98	0.62	267
Positif	0.98	0.71	0.83	879
Akurasi			0.74	1390
<i>Macro avg</i>	0.75	0.76	0.71	1389
<i>Weighted avg</i>	0.85	0.74	0.76	1390



Gambar 15. *Confusion Matrix Support Vector Machine.*

Dalam gambar *Confusion Matrix* ditunjukkan kinerja dari model tersebut. Dari model ini mengklasifikasikan dengan benar positif 627 opini, 261 netral, dan 147 opini negatif. Model juga mengklasifikasikan salah positif hanya 30 opini yang dikira negatif dan 222 opini yang dikira netral. Sedangkan pada

pengklasifikasian netral yang salah 3 opini dikira negatif dan 3 opini dikira positif. Pada pengklasifikasian yang negatif 87 opini dianggap sebagai netral dan sebanyak 10 opini dianggap positif.

6. Memilih Model Terbaik

Model yang dipilih merupakan model yang paling cocok untuk dataset opini mahasiswa Universitas XYZ yang memiliki tingkat *accuracy*, *precision*, *recall* yang paling baik. Berikut perbandingan *accuracy*, *precision*, dan *recall* dari *naïve bayes classifier* dan *support vector machine*.

Tabel XIV
Confusion Matrix Setiap Metode Algoritma

Metode	TP	FPNET	FPN	TNET	FNETP	FNETN	TN	FNNET	FNP
<i>Naïve Bayes</i>	818	22	39	215	34	18	198	12	34
<i>Support Vector Machine</i>	627	222	30	261	3	3	147	87	10

Keterangan:

- TP : *True Positive*
- FPNET : *False Positive (Neutral)*
- FPN : *False Positive (Negative)*
- TNET : *True Neutral*
- FNETP : *False Neutral (Positive)*
- FNETN : *False Neutral (Negative)*
- TN : *True Negative*
- FNNET : *False Negative (Neutral)*
- FNP : *False Negative (Positive)*

Tabel XV
Perbandingan *Performance* Metode Algoritma

Metode	<i>Accuracy</i>
<i>Naïve Bayes</i>	88,56%
<i>Support Vector Machine</i>	74,46%

Dari perbandingan yang ada pada Tabel XIV dan XV menunjukkan bahwa performa algoritma yang cocok untuk dataset opini mahasiswa Universitas XYZ adalah *Naïve Bayes Classifier*. Dimana dengan dataset yang sama *Naïve Bayes Classifier* memiliki akurasi lebih tinggi sebanyak 11,1 % dibanding *Support Vector Machine*.

7. Testing

Dikarenakan hasil akurasi yang didapat lebih tinggi pada model *complement naïve bayes* maka dari itu model ini yang digunakan untuk *testing dataset*. Dimana model akan mencoba memprediksi apakah sentimen/opini merupakan positif, negatif, ataupun netral. Berikut hasil dari tes kalimat yang dilakukan.

Dari hasil yang ada didapat ditunjukkan bahwa angka positif tetap tinggi di 1259 opini positif. Opini negatif diposisi kedua dengan 372 opini. Terakhir, netral sebanyak 364 opini. Ini menunjukkan model *Complement Naïve Bayes* sangat baik pada *dataset* yang tidak seimbang.

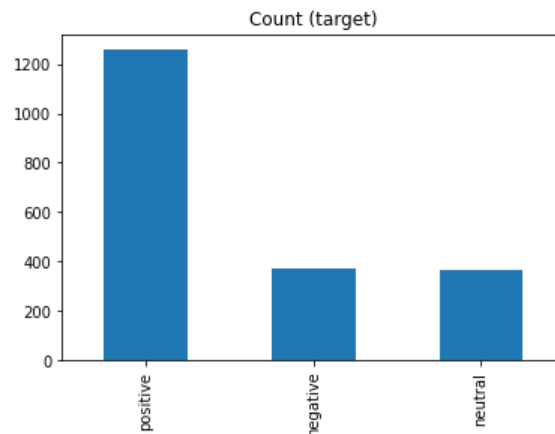
Tabel XVI
Contoh potongan proses testing

```

1 df1 = pd.read_csv('Only Text Preprocessing.csv')
2 test1 = df1['Data_After_Preprocessing']
3 predicted = []
4 for i in test1 :
5     a = count_vect.transform([i])
6     X_coba = tfidf.fit_transform(a).toarray()
7     y_pred = cnb_model.predict(X_coba)
8     acc = cnb_model.predict_proba(X_coba)
9     probmbnb = acc.max(axis=1)
10    predicted.append({"Opini": i , "label" : y_pred[0]})
11 df3 = pd.DataFrame(predicted)
12 df3.head()
    
```

A	B	C
	Opini	label
0	baik	positive
1	baik tidak ada kendala	positive
2	fitur login baik mudah bagi fitur ajar mahasiswa salah tekan fitur login sesuai semester	positive
3	login cukup lancar password saved ada fitur keep me logged in	neutral
4	bagus	positive
5	baik	positive
6	baik	positive
7	cukup mudah paham user interface kurang rapi	positive
8	fitur kurang tarik	negative
9	bingung login bagi tahun ajar	negative
10	baik	positive
11	cukup baik	positive
12	fitur login bagus saran fitur session guna login tentu tambah	positive
13	baik isi data username password letak	positive
14	bagus	positive
15	baik	positive
16	mudah paham	positive
17	mudah	positive
18	baik	positive
19	cukup	positive
20	cukup baik	positive
21	baik	positive
22	baik login device buru lama standby tidak login hemat	negative
23	lumayan	positive
24	standby jam logout cukup ganggu	negative
25	kurang tarik	negative
26	baik tidak ada kendala fitur login wajar tidak ada aneh	positive
27	cukup mudah lancar login baik tinggal halaman lama sistem otomatis log out login ulang tingkat	positive
28	kurang sederhana	negative
29	fitur login ada kala ganggu lama loading	negative

Gambar 15 Hasil *Testing* di Excel



Gambar 16 Hasil Positif, Negatif, dan Netral Setelah Model dibuat

V. KESIMPULAN DAN SARAN

Berdasarkan hasil yang didapatkan peneliti pada Analisis Sentimen Pada *E-Learning* Universitas XYZ menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine* maka peneliti berkesimpulan sebagai berikut :

- 12 Fitur dalam *E-Learning* Universitas XYZ memiliki hasil positif , negatif, dan netral. Terdapat 8 fitur yang bekerja dengan baik yaitu .Fitur *login* memiliki hasil yang baik dimana opini positif 122(74,39%) dari 166 opini. Untuk fitur *search* mata kuliah memiliki hasil yang cukup baik dimana opini positif 110(66,26%) dari 166 opini. Tampilan *E-Learning* sudah berjalan cukup baik dimana opini positif 117(70%) dari 167 opini. Fitur absensi juga berjalan dengan baik dimana opini positif 125(75%) dari 166 opini. Fitur *upload* telah berjalan dengan baik dimana opini positif 124(74,69%) dari 166 opini. Fitur *download* sangat baik dimana opini positif 146 (87%) dari 166 opini. Fitur kuis di *E-Learning* juga berjalan dengan sangat baik dimana opini positif 139 (84,24%) dari 165 opini. Dan fitur terakhir yang berjalan dengan baik adalah registrasi dengan 116 opini positif(72%). Terdapat 2 fitur yang tidak berjalan dengan baik yaitu fitur *BigBlueButton* dengan opini negatif 82(49%) dari 167 opini. Fitur koneksi tidak berjalan dengan baik dengan opini negatif 75(46%) dari 163 opini. Yang terakhir adalah fitur yang masih banyak orang belum pernah menggunakannya yaitu fitur lupa *password* dan notifikasi. Dimana fitur lupa *password* dengan opini netral 85(50,8%) dari 167 opini. Yang terakhir adalah notifikasi dengan opini netral 83(49,7%).
- Dari jumlah frekuensi kata menunjukkan bahwa fitur fitur seperti *login*, *search* mata kuliah, tampilan, absensi, *upload*, *download*, kuis, dan registrasi pada *e-learning* Universitas XYZ sudah cukup baik karena 8 fitur memiliki hasil yang positif. Hanya 2 fitur yang memiliki banyak hal

negatif yaitu fitur *BigBlueButton* dan koneksi sehingga perlu ditingkatkan kembali. Sedangkan terdapat 2 fitur yang banyak para mahasiswa belum gunakan adalah lupa password dan notifikasi.

3. Dari hasil penelitian menunjukkan bahwa metode algoritma *Naïve Bayes Classifier* lebih baik dan lebih cocok untuk dataset opini mahasiswa Universitas XYZ dibandingkan *Support Vector Machine* dengan akurasi masing masing 88,56% dan 74,46%. Perbedaan akurasinya kurang lebih 11,1 %.

Dari hasil dan kesimpulan yang ada peneliti memiliki saran dan masukan terhadap *e-learning* Universitas XYZ. Saran penulis adalah sebagai berikut :

1. Data yang didapat masih kurang banyak sehingga data yang paling banyak mendapatkan hasil yang lebih baik yaitu data positif. Apabila data yang ditraining lebih banyak model *naïve bayes* akan menjadi lebih baik.
2. Frekuensi kata hanya menunjukkan 1 kata saja sehingga hanya cukup membantu saja. Dikarenakan misalnya terdapat 2 kata seperti “tidak baik” maka yang muncul hanya kata “tidak” dan “baik” sehingga harus diperhatikan kembali

DAFTAR PUSTAKA

- [1] K. Khan *et al.*, “Urdu sentiment analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 646–651, 2018, doi: 10.14569/ijacsa.2018.090981.
- [2] Kementerian Pendidikan Dan Kebudayaan, “Pedoman Penyelenggaraan Belajar Dari Rumah Dalam Masa Darurat Penyebaran Corona Virus Disease (Covid-19),” *Surat Edaran Nomor 15 Tahun 2020*, no. 021, pp. 1–20, 2020.
- [3] A. S. AJIATMOJO, “Penggunaan E-Learning Pada Proses Pembelajaran Daring,” *Teach. J. Inov. Kegur. dan Ilmu Pendidik.*, vol. 1, no. 3, pp. 229–235, 2021, doi: 10.51878/teaching.v1i3.525.
- [4] A. G. Prawiyogi, T. L. Sadiyah, A. Purwanugraha, and P. N. Elisa, “Penggunaan Media Big Book untuk Menumbuhkan Minat Membaca di Sekolah Dasar,” *J. Basicedu*, vol. 5, no. 1, pp. 446–452, 2021, doi: 10.31004/basicedu.v5i1.787.
- [5] B. A. Sevsa and M. D. R. Wahyudi, “Analisis Sentimen pada Indeks Kinerja Dosen Fakultas SAINTEK UIN Sunan Kalijaga Menggunakan Naive Bayes Classifier,” *J. Buana Inform.*, vol. 10, no. 2, p. 112, 2019, doi: 10.24002/jbi.v10i2.2250.
- [6] C. Prianto *et al.*, “Sentimen Analisis Terhadap Pembelajaran Jarak Jauh Menggunakan Metode Naïve Bayes Classifier dan Lexicon Based,” vol. 14, no. 2, pp. 79–87, 2021.
- [7] I. Fakhri and R. F. Umbara, “Analisis Sentimen pada Kuisisioner Kepuasan Terhadap Layanan dan Fasilitas Kampus Universitas Dengan Menggunakan Klasifikasi Support Vector Machine (SVM),” *e-Proceeding Eng.*, vol. 6, no. 2, 2019.
- [8] R. R. Putra, M. E. Johan, and E. R. Kaburuan, “A naïve bayes sentiment analysis for fintech mobile application user review in Indonesia,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 1856–1860, 2019, doi: 10.30534/ijatcse/2019/07852019.
- [9] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [10] B. M. Pintoko and K. M. L., “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier,” *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [11] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [12] M. Ichwan, I. A. Dewi, and Z. M. S., “Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Mangga Berdasarkan Fitur Warna,” *MIND J.*, vol. 3, no. 2, pp. 16–23, 2019, doi: 10.26760/mindjournal.v3i2.16-23.
- [13] A. Firmansyah Sulaeman, A. Afif Supianto, and F. Abdurrachman Bachtiar, “Analisis Sentimen Opini Mahasiswa Terhadap Saran Evaluasi Kinerja Dosen Menggunakan TF-IDF dan Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 5647–5655, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [14] A. Zein and E. Suharyanto, “Sentimen Analisis Pada Komentar Pendek Evaluasi Dosen Oleh Mahasiswa (Edom) Program Studi Sistem Informasi Universitas ...,” *J. Ilmu Komput.*, vol. V, no. 01, 2022, [Online]. Available: <https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/113%0Ahttps://jurnal.pranataindonesia.ac.id/index.php/jik/article/download/113/66>