

**ANALISIS SENTIMEN MAHASISWA TERHADAP PERKULIAHAN  
DALAM JARINGAN MENGGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**SKRIPSI**

**OLEH  
BAGUS AZIZ RAHMATULLAH  
NIM. 17610014**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2021**

**ANALISIS SENTIMEN MAHASISWA TERHADAP PERKULIAHAN  
DALAM JARINGAN MENGGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**SKRIPSI**

**Diajukan Kepada  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh  
Bagus Aziz Rahmatullah  
NIM. 17610014**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2021**

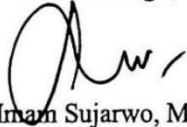
**ANALISIS SENTIMEN MAHASISWA TERHADAP PERKULIAHAN  
DALAM JARINGAN MENGGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**SKRIPSI**

**Oleh  
Bagus Aziz Rahmatullah  
NIM. 17610014**

Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal 16 November 2021

Pembimbing I,



Dr. Imam Sujarwo, M.Pd  
NIP. 19630502 198703 1 005

Pembimbing II



Erna Herawati, M.Pd  
NIDT. 19760723201802012222

Mengetahui,  
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005

**ANALISIS SENTIMEN MAHASISWA TERHADAP PERKULIAHAN  
DALAM JARINGAN MENGGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER***

**SKRIPSI**

**Oleh  
Bagus Aziz Rahmatullah  
NIM. 17610014**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 25 November 2021

Susunan Dewan Penguji:

Penguji Utama : Muhammad Khudzaifah, M.Si

Ketua Penguji : Angga Dwi Mulyanto, M.Si

Sekretaris Penguji : Dr. Imam Sujarwo, M.Pd

Anggota Penguji : Erna Herawati, M.Pd

Tanda Tangan

(.....)

(.....)

(.....)

(.....)

Mengetahui,  
Ketua Program Studi Matematika

  
Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005



## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Bagus Aziz Rahmatullah

NIM : 17610014

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Analisis Sentimen Mahasiswa Terhadap Perkuliahan  
Dalam Jaringan Menggunakan Metode *Naïve Bayes Classifier*

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Desember 2021  
Yang membuat pernyataan



Bagus Aziz Rahmatullah  
NIM 17610014

## **MOTTO**

*“Belajarlal mengucap syukur dari hal-hal baik di hidupmu dan belajarlal menjadi pribadi yang kuat dengan hal-hal buruk di hidupmu”*

*(B. J. Habibie)*

## **PERSEMBAHAN**

Skripsi ini penulis persembahkan untuk:

Ayah tercinta Totok Dwi Riyanto, Ibu tercinta Sumarni dan adik-adikku tersayang  
Dina Rahma Sityawati dan M. Muflih Surya Rizkillah yang selalu menjadi  
semangat dan motivasi bagi penulis dalam menuntut ilmu, mengabdikan, dan  
berjuang di tanah rantau.

## KATA PENGANTAR

*Bismillahirrahmanirrahim*

Puji dan syukur kepada Allah SWT yang telah memberi rahmat dan hidayah serta inayah-Nya kepada kita, khususnya kepada penulis sehingga mampu menyelesaikan skripsi dengan judul **ANALISIS SENTIMEN MAHASISWA TERHADAP PERKULIAHAN DALAM JARINGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER**.

Shalawat serta salam kepada baginda Nabi Besar Muhammad SAW, yang semoga kita semua senantiasa diberikan keberkahan dan syafaat-Nya hingga hari akhir nanti.

Penyusunan skripsi ini ditujukan sebagai salah satu persyaratan dalam menyelesaikan program Sarjana Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang serta sebagai partisipasi penulis untuk menerapkan ilmu yang telah diperoleh ketika penulis masih menimba ilmu di Program Studi Matematika Fakultas Sains dan Teknologi. Baik secara langsung maupun tidak langsung, oleh karena itu perkenankan penulis berterimakasih kepada:

1. Bapak Prof. Dr. M. Zainuddin, MA. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Ibu Dr. Sri Harini, M.Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Ibu Dr. Elly Susanti, M.Sc. selaku Ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.



4. Ibu Evawati Alisah, M.Pd. selaku Dosen Wali yang selalu memberi semangat penulis dalam menimba ilmu di kampus tercinta.
5. Bapak Dr. Imam Sujarwo, M.Pd. dan Ibu Erna Herawati, M.Pd. selaku dosen pembimbing yang telah membimbing serta mengarahkan penulis dalam menyusun skripsi.
6. Segenap dosen Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
7. Kedua orang tua penulis, Bapak Totok Dwi Ryanto, S.Pd. dan Ibu Sumarni yang selalu memberi semangat, doa, serta keteladanan baik secara moral maupun nyata sejak penulis menjalani perkuliahan sampai menapaki tugas akhir.
8. Adik tercinta Dina Rahma Sityawati dan M. Surya Rizkillah yang selalu menyemangati dan mengingatkan kakaknya.
9. Teman-teman Program Studi Matematika 2017 yang telah memberikan semangat dan dukungan dalam menyelesaikan skripsi ini.
10. Mahad Sunan Ampel Al-Aly, UPKM El-Ma'rifah, Diamond dan teman-teman yang ada di Ma'had yang telah memberikan pelajaran berharga tentang pengabdian, keikhlasan, dan arti bermasyarakat sesungguhnya.
11. Kepada semua pihak yang tidak dapat disebutkan satu persatu yang telah tulus membantu dalam menyelesaikan skripsi ini.

Dan akhirnya skripsi ini telah selesai, namun masih jauh dari kesempurnaan. Karena itu penulis mengharapkan kritik dan saran yang telah membangun dari semua belah pihak, demi kesempurnaan dan perbaikan karya ini.

Semoga Allah SWT melimpahkan rahmat dan karunia-Nya kepada kita semua.  
Selain itu, penulis juga berharap semoga skripsi ini bisa bermanfaat bagi penulis  
dan pembaca pada umumnya. *Aamiin*

*Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

Malang, 20 Desember 2021

Penulis

## DAFTAR ISI

HALAMAN JUDUL	
HALAMAN PENGAJUAN	
HALAMAN PERSETUJUAN	
HALAMAN PENGESAHAN	
HALAMAN PERNYATAAN KEASLIAN TULISAN	
HALAMAN MOTTO	
HALAMAN PERSEMBAHAN	
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
ABSTRAK .....	xv
ABSTRACT .....	xvi
المخلص .....	xvii

## BAB I PENDAHULUAN

1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Sistematika Penulisan .....	5

## BAB II KAJIAN PUSTAKA

2.1 Pembelajaran Daring .....	7
2.2 Analisis Sentimen .....	8
2.3 Teorema Bayes .....	9
2.4 <i>Naïve Bayes Classifier</i> .....	10
2.5 <i>Text Mining</i> .....	14
2.6 <i>Data Pre-processing</i> .....	15
2.7 <i>Bag of Words</i> .....	16
2.8 <i>Confusion Matrix</i> .....	17
2.9 <i>Cross Validation</i> .....	18
2.10 Kajian Keislaman .....	20

### **BAB III METODE PENELITIAN**

3.1 Pendekatan Penelitian .....	23
3.2 Variabel Penelitian .....	23
3.3 Jenis dan Sumber Data .....	23
3.4 Pra-pemrosesan Data .....	25
3.5 Analisis Data .....	29

### **BAB IV HASIL DAN PEMBAHASAN**

4.1 Pra-pemrosesan Data .....	32
4.1.1 <i>Case Folding</i> .....	32
4.1.2 <i>Cleansing</i> .....	34
4.1.3 <i>Stopword Removal</i> .....	36
4.1.4 <i>Stemming</i> .....	37
4.1.5 Tokenisasi .....	38
4.1.6 Model <i>Bag of Words</i> .....	39
4.1.7 Pemisahan Data <i>Set</i> Menjadi Data Latih dan Data Uji .....	40
4.2 Hasil Klasifikasi Sentimen Mahasiswa Terhadap Perkuliahan Daring .....	41
4.3 Akurasi Metode <i>Naïve Bayes Classifier</i> .....	48
4.4 Konsep Analisis Sentimen dan Menuntut Ilmu dalam Islam .....	50

### **BAB V PENUTUP**

5.1 Kesimpulan .....	52
5.2 Saran .....	52

<b>DAFTAR RUJUKAN</b> .....	54
-----------------------------	----

### **LAMPIRAN**

### **RIWAYAT HIDUP**

## DAFTAR TABEL

Tabel 2.1	Frekuensi Kata Pada Dokumen .....	17
Tabel 2.2	<i>Confusion Matrix</i> .....	17
Tabel 3.1	Data Responden .....	24
Tabel 4.1	Proses <i>Case Folding</i> .....	33
Tabel 4.2	Proses <i>Case Folding</i> .....	33
Tabel 4.3	Proses <i>Case Folding</i> .....	34
Tabel 4.4	Proses <i>Cleansing</i> .....	35
Tabel 4.5	Proses <i>Cleansing</i> .....	35
Tabel 4.6	Proses <i>Cleansing</i> .....	36
Tabel 4.7	<i>Stopword Removal</i> .....	37
Tabel 4.8	Proses <i>Stemming</i> .....	38
Tabel 4.9	Pemisahan Data Latih dan Data Uji .....	41
Tabel 4.10	Prediksi Pada Data Uji .....	42
Tabel 4.11	Contoh Data <i>Training</i> .....	43
Tabel 4.12	Perhitungan Probabilitas Setiap Kata Pada Komentar 1 .....	44
Tabel 4.13	Perhitungan Probabilitas Setiap Kata Pada Komentar 2 .....	44
Tabel 4.14	Perhitungan Probabilitas Setiap Kata Pada Komentar 3 .....	44
Tabel 4.15	Perhitungan Probabilitas Setiap Kata Pada Komentar 4 .....	45
Tabel 4.16	Perhitungan Probabilitas Setiap Kata Pada Komentar 5 .....	45
Tabel 4.17	Perhitungan Probabilitas Setiap Kata Pada Komentar 6 .....	45
Tabel 4.18	Contoh Data <i>Testing</i> .....	46
Tabel 4.19	Nilai Probabilitas Data <i>Testing</i> .....	48
Tabel 4.20	Model <i>Confusion Matrix</i> .....	49
Tabel 4.21	Nilai <i>Precision</i> , <i>Recall</i> , dan <i>f1-score</i> .....	49
Tabel 4.22	Nilai Dari Setiap Kelas .....	50

## DAFTAR GAMBAR

Gambar 2.1 Iterasi Pada K-Fold .....	19
Gambar 3.1 Proses <i>Case Folding</i> .....	26
Gambar 3.2 Proses <i>Cleansing</i> .....	26
Gambar 3.3 Proses <i>Stopword Removal</i> .....	27
Gambar 3.4 Proses <i>Stemming</i> .....	28
Gambar 3.5 Proses Tokenisasi .....	28
Gambar 3.6 <i>Flowchart</i> Klasifikasi Analisis Sentimen .....	31
Gambar 4.1 Visualisasi Kata Terpopuler dengan <i>Wordcloud</i> .....	39
Gambar 4.2 Data <i>Bag of Words</i> .....	39

## ABSTRAK

Rahmatullah, Bagus Aziz. 2021. **Analisis Sentimen Mahasiswa Terhadap Perkuliahan Dalam Jaringan Menggunakan Metode Naïve Bayes Classifier**. Skripsi. Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Islam Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Imam Sujarwo, M.Pd, (II) Erna Herawati, M.Pd

**Kata Kunci:** kuliah daring, Naïve Bayes, analisis sentimen, covid-19, akurasi

Sejak terjadinya pandemi virus Corona di Indonesia, pemerintah mengeluarkan surat edaran Kementerian Pendidikan dan Kebudayaan (Kemdikbud) Direktorat Pendidikan Tinggi No.1 tahun 2020 tentang pencegahan penyebaran Corona Virus Disease (Covid-19) di perguruan tinggi. Melalui surat edaran tersebut, pihak Kemdikbud memberikan instruksi kepada perguruan tinggi untuk menyelenggarakan pembelajaran jarak jauh dan menyarankan mahasiswa untuk belajar di rumah masing-masing. Perkuliahan daring yang dianggap sebagai strategi kemudian menjadi suatu kontroversi karena diperlukan adaptasi. Perubahan mendadak dari perkuliahan tatap muka menjadi perkuliahan daring menyebabkan berbagai macam respon dari para mahasiswa.

Tujuan dari penelitian ini adalah untuk menganalisis sentimen atau respon mahasiswa terhadap perkuliahan daring di masa pandemi Covid-19 di Indonesia dengan menggunakan data yang dikumpulkan lewat kuisioner dan diolah dengan menggunakan metode naïve bayes classifier. Penelitian ini adalah penelitian studi kasus deskriptif kuantitatif. Penelitian ini dilakukan dengan mengumpulkan data terlebih dahulu. Data dikumpulkan lewat kuisioner dengan pertanyaan mengenai tanggapan mereka terhadap perkuliahan dalam jaringan pada masa pandemi Covid-19. Data yang terkumpul ada 157 data opini mahasiswa tentang perkuliahan daring. Setelah data sudah terkumpul, data dibersihkan terlebih dahulu dari tanda baca dan kata-kata yang tidak berpengaruh dalam analisis sentimen. Setelah data menjadi bersih, selanjutnya akan diperlihatkan hasil klasifikasi serta akurasi yang diperoleh dari model naïve bayes classifier.

Hasil menunjukkan bahwa pembelajaran daring memiliki sentiment negatif lebih banyak ketimbang sentiment positif. Tingginya sentimen negatif disebabkan karena ketidaknyamanan mahasiswa dalam pembelajaran daring. Kata yang sering muncul adalah ‘tidak efektif’, ‘susah’, dan ‘tugas’. Hasil akurasi dari model ini adalah 75% di mana hasil akurasi ini merupakan hasil yang cukup baik dalam pengklasifikasian.

## ABSTRACT

Rahmatullah, Bagus Aziz. 2021. **Sentiment Analysis Student on Online Learning Using Naïve Bayes Classifier Method**. Thesis. Mathematics Department, Faculty of Science and Technology, Maulana Malik Ibrahim Malang State Islamic University of Malang. Advisors: (I) Dr. Imam Sujarwo, M.Pd, (II) Erna Herawati, M.Pd

**Keywords:** online learning, *naïve bayes*, sentiment analysis, covid-19, accuracy

Since the pandemic of Covid-19 was happened in Indonesia, the government shared the letter of The Ministry of Education and Higher Culture Education Directorate No.1, year 2020 about prevention of the spread of Corona Virus Disease (Covid-19) in higher education. Through the letter, The Ministry of Education and Culture gave an instruction for college to organize online learning and suggested students to study at their home. Online learning which was considered as a strategy then became a controversy because it needed adaption. This sudden change from normal learning to online learning caused many responses from students.

The aim of this research was to analyze student sentiment or responses on online learning in this pandemic era of Covid-19 in Indonesia by using data which had been collected using questionnaire and processed using naïve bayes classifier method. This research was case study descriptive quantitative research. The research was done by collecting the data first. The data was collected through questionnaire with the question about their opinion on online learning in this pandemic Covid-19 era. The data was 157 student's data opinion on online learning. After the data was collected, the data was cleaned first from question mark and the words which didn't give an effect in sentiment analysis. After the data was cleaned, then the result of the classification will be showed as well as the accuracy which the model earned.

The result showed that online learning had negative sentiment more than positive sentiment. The height of negative sentiment was caused by discomfort of student in online learning. The word which frequently showed was 'tidak efektif', 'susah', and 'tugas'. The accuracy of this model was 75% when the result of this accuracy was good result in classification.



## ملخص

رحمة الله، بكوس عزيز. ٢٠٢١. تحليل عواطف الطلبة على الدراسة عبر الانترنت باستخدام طريقة نيفي بيس جليسييفر (Bayes Classifier Naïve). البحث الجامعي. قسم الرياضيات. كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم مالانج. المشرف الأول : إمام سوجروو الماجستير. المشرف الثاني: إيرنا هيراوري

كلمة مفتاحية: الدراسة عبر الانترنت، نيفي بيس جليسييفر ، تحليل العواطف

منذ إنتشار جائحة كورونا في إندونيسيا، تنشر الحكومة رسالة التعميمية لوزارة التربية والثقافة، مديرية الجامعة رقم ١ سنة ٢٠٢١ عن نهي الانتشار فيروس كورونا (COVID-19) في الجامعة. بوصيلة هذه رسالة التعميمية يأمر وزيرة التربية والثقافة إلى الجامعة لإقامة التعليم عن بعد وتقتراح الطلبة للتعلم في البيت. ثم أصبحت المحاضرة عبر الانترنت كاستراتيجية عكسية لأنها مهتاج على التكيف. يسبب التغيير المفاجيء من المحاضرة وجها لوجه إلى المحاضرة عبر الانترنت إستجابة مختلفة من الطلبة. فأما أهداف هذا البحث لتحليل عواطف أو إستجابة الطلبة على الدراسة عبر الانترنت في عصر كورون في إندونيسيا باستخدام البيانات التي تم جمعها بالاستبانة وتحليلها باستخدام طريقة نيفي بيس جليسييفر (Bayes Classifier Naïve). هذا البحث هي دراسة الحالة الوصفية الكمية. قامت هذا البحث بجمع البيانات أولا. وجمعت البيانات عبر الاستبانة مع أسئلة عن إستجاباتهم إلى الدراسة عبر الانترنت في عصر فيروس كورونا. كان البيانات التي تم جمعها هي ١٥٧ لبيانات رأي الطلبة عن التعليم عبر الانترنت. وبعد تمام جمع البيانات، تفصل البيانات من علامة التقييم والكلمات التي لا تأثر في تحليل العواطف. وبعد ذلك ستعرض نتائج التصنيف والدقة المحسولة من نموذج نيفي بيس جليسييفر (Bayes Classifier Naïve).

تظهر النتائج أنّ لدى التعليم عبر الانترنت عواطف سلبية أكثر من عواطف إيجابية. وتسبب على أكثر هذه العواطف السلبية إزعاج الطلبة في التعليم عبر الانترنت. وتظهر الكلمات غالبا هي "غير فعالية" و "صعبة" و "واجبة". وكانت نتائج الدقة من هذه النموذج ٧٥% حيث تكون هذه نتائج الدقة نتيجة جيدة في التصنيف.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Peraturan Pemerintahan Republik Indonesia No. 30 tahun 1990 tentang perguruan tinggi mengatakan bahwa pendidikan tinggi adalah pendidikan jenjang yang lebih tinggi daripada pendidikan menengah di jalur pendidikan sekolah. Perguruan tinggi merupakan suatu pendidikan yang menjadi terminal akhir bagi seorang yang berpeluang belajar setinggi-tingginya melalui jalur pendidikan sekolah (Dardjowidjojo, 1991).

Ketika menuntut ilmu di perguruan tinggi, biasanya dosen dan mahasiswa melakukan kegiatan belajar mengajar di dalam kelas. Khusus mahasiswa yang melakukan praktikum, dosen dan mahasiswa melakukan kegiatan praktikum di *laboratorium*. Sejak terjadinya pandemi virus Corona di Indonesia, pemerintah mengeluarkan surat edaran Kementerian Pendidikan dan Kebudayaan (Kemdikbud) Direktorat Pendidikan Tinggi No.1 tahun 2020 tentang pencegahan penyebaran *Corona Virus Disease* (Covid-19) di perguruan tinggi. Melalui surat edaran tersebut pihak Kemdikbud memberikan instruksi kepada perguruan tinggi untuk menyelenggarakan pembelajaran jarak jauh dan menyarankan mahasiswa untuk belajar di rumah masing-masing atau yang dikenal sekarang dengan sebutan perkuliahan dalam jaringan (daring) ([kemdikbud.go.id](http://kemdikbud.go.id)).

Perkuliahan daring adalah pembelajaran interaktif. Dalam perkuliahan daring di masa pandemi membuat mahasiswa tidak diperkenankan untuk pergi ke

kampus. Dosen dan mahasiswa dapat berkomunikasi dengan menggunakan media internet (sevima.com).

Dosen juga bisa memberikan materi kuliah, baik berupa file atau video. Pada sistem perkuliahan daring, dosen dan mahasiswa juga memiliki jadwal tatap muka dengan menggunakan layanan komunikasi video seperti *Google Meet* atau *Zoom Video Communications*. Pada perkuliahan daring, tentu mahasiswa bisa menemukan berbagai masalah seperti sangat bergantung pada internet, harga gadget yang begitu mahal, dan ada beberapa mata kuliah yang tidak bisa diajarkan secara daring.

Perkuliahan daring yang dianggap sebagai strategi kemudian menjadi suatu kontroversi karena diperlukannya adaptasi. Perbedaan infrastruktur yang dimiliki, kualitas koneksi, perangkat yang digunakan, dan masih mahalnya kuota data internet menjadi hambatan utama perkuliahan daring ini. Perubahan mendadak dari kuliah tatap muka menjadi perkuliahan daring pada skala besar menyebabkan berbagai respon dari para mahasiswa (Samsir dkk, 2021). Penelitian ini bertujuan untuk menganalisis sentimen atau respon mahasiswa terhadap perkuliahan daring pada masa pandemi Covid-19 di Indonesia dengan menggunakan data yang dikumpulkan lewat kuisisioner dan diolah menggunakan *naïve bayes classifier* untuk menganalisis sentimen terhadap perkuliahan daring.

Analisis sentimen atau *opinion mining* adalah studi komputasional dari opini orang, sentimen, dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks (Liu, 2012). Analisis sentimen ini dapat mengelompokkan polaritas dari teks dalam kalimat atau dokumen tersebut apakah termasuk dalam kalimat positif atau negatif. Informasi yang didapatkan tentu sangat

berharga sebagai alat penentu kebijakan dan ini bisa dilakukan dengan *text mining*. *Text mining* juga dikenal sebagai *data mining* teks atau penemuan pengetahuan dari *database* tekstual dengan tujuan untuk mendapatkan informasi yang berguna dari sebuah dokumen.

Hal ini berkaitan dengan dengan salah satu hadist yang artinya:

فَلَمْ يَزَلْ عُمَرُ يُرَاجِعُنِي فِيهِ حَتَّى شَرَحَ اللَّهُ لِي ذَلِكَ صَدْرِي، وَرَأَيْتُ الَّذِي رَأَى عُمَرُ

“Umar senantiasa membujukku untuk mengevaluasi pendapatku dalam permasalahan itu hingga Allah melapangkan hatiku dan akupun berpendapat sebagaimana pendapat Umar” [HR. Bukhari].

Dalam riwayat tersebut, Imam Bukhari menceritakan usul Umar kepada Abu Bakar dalam mengumpulkan Al-Qur'an. Saat itu, Abu Bakar menolak usul tersebut namun Umar terus mendesak dan mengatakan bahwa hal itu adalah kebaikan. Akhirnya, Abu Bakar menerima hal tersebut. Hadist tersebut mengingatkan bahwa sebagai manusia tidak luput dari kesalahan, maka dari itu kita wajib untuk saling mengingatkan. Pada penelitian ini, penulis menggunakan metode *Naïve bayes*. *Naïve bayes* merupakan metode pengklasifikasian statistik yang bisa digunakan untuk memprediksi probabilitas keanggotaan dari suatu class. *Naïve bayes* memiliki nilai akurasi yang tinggi sehingga bisa digunakan untuk berbagai macam keperluan seperti klasifikasi dokumen dan masalah klasifikasi lainnya.

Pada penelitian yang ditulis oleh Tika dkk (2020) tentang Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Twitter Menggunakan *Naïve Bayes* mencoba untuk menganalisa opini masyarakat mengenai aturan larangan mudik dengan mengklasifikasikan opini menjadi positif dan negatif. Pada penelitian ini juga menunjukkan hasil akurasi yang tinggi dari metode *Naïve Bayes Classifier*, yaitu sebesar 83,38%.

Penelitian yang ditulis oleh Fitri dkk (2020) tentang Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naïve Bayes, Random Forest dan Support Vector Machine. Penelitian ini menunjukkan jumlah komentar positif lebih banyak diberikan dibandingkan dengan komentar negatif dari para pengguna aplikasi Ruangguru. Penelitian ini juga memperlihatkan bahwa tingkat keakurasian yang dihasilkan oleh *Naïve Bayes* sebesar 94,16%.

Pada penelitian sebelumnya (Fanissa dkk, 2018) yang meneliti tentang Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naïve Bayes Classifier* dan Seleksi Fitur Query Expansion Ranking mengungkapkan bahwa klasifikasi menggunakan *Naïve Bayes Classifier* mendapatkan akurasi terbaik yaitu 86,6%. Sehingga pada penelitian ini penulis akan melakukan analisis sentimen menggunakan metode *Naïve bayes*. Berdasarkan uraian di atas, penulis mengangkat tema permasalahan dan penyelesaian dalam bentuk penelitian yang berjudul “Analisis Sentimen Mahasiswa Terhadap Perkuliahan Daring Menggunakan Metode *Naïve Bayes Classifier*”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana hasil klasifikasi sentimen mahasiswa terhadap perkuliahan daring menggunakan metode *Naïve Bayes Classifier*?
2. Bagaimana akurasi metode *Naïve Bayes Classifier* dalam pengklasifikasian analisis sentimen mahasiswa terhadap perkuliahan daring?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang akan dibahas, tujuan penelitian ini adalah:

1. Untuk mengetahui hasil klasifikasi sentimen mahasiswa terhadap perkuliahan daring menggunakan metode *naïve bayes classifier*.
2. Untuk mengetahui akurasi yang diperoleh *naïve bayes classifier* dalam mengklasifikasikan sentimen mahasiswa terhadap perkuliahan daring.

### 1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah

1. Sebagai bahan referensi untuk penelitian lebih lanjut pada matematika terapan bidang komputasi.
2. Sebagai informasi untuk pengambil kebijakan penggunaan media daring dalam pembelajaran di universi.

### 1.5 Sistematika Penulisan

Adapun sistematika penulisan yang digunakan penulis terdiri dari lima bab yang masing-masing memiliki subbab diantaranya sebagai berikut:

#### Bab I Pendahuluan

Bab ini meliputi latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah.

#### Bab II Kajian Pustaka

Bab ini meliputi penelitian terdahulu, kajian dalam Islam, definisi, dan teori-teori yang mendukung yaitu, analisis sentimen, *text mining*, *data*

*preprocessing, Naïve Bayes Classifier, confusion matrix, cross validation.*

### Bab III Metode Penelitian

Bab ini meliputi pendekatan penelitian, variabel penelitian, jenis dan sumber data, dan analisis data.

### Bab IV Pembahasan

Bab ini menjelaskan tentang bagaimana klasifikasi sentimen menggunakan metode *Naïve Bayes Classifier* dan tingkat akurasi yang diperoleh.

### Bab V Penutup

Bab ini menyajikan poin-poin dari hasil dan pembahasan secara garis besar berupa kesimpulan dan saran.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Pembelajaran Daring**

Pembelajaran daring adalah pembelajaran yang memerlukan jaringan internet dengan aksesibilitas, konektivitas, fleksibilitas, dan kemampuan untuk memunculkan berbagai jenis interaksi pembelajaran (Moore dkk, 2011). Pembelajaran daring pada pelaksanaannya memerlukan dukungan perangkat-perangkat seperti *smartphone*, *tablet*, dan laptop yang digunakan untuk mengakses informasi di mana dan kapan saja (Gikas & Grant, 2013).

Penggunaan teknologi *mobile* memiliki kontribusi besar di dunia pendidikan, termasuk di dalamnya adalah untuk mencapai tujuan pembelajaran jarak jauh (Korucu & Alkan, 2011). Berbagai media lain juga bisa digunakan untuk mendukung pelaksanaan pembelajaran secara daring. Contohnya menggunakan kelas-kelas virtual menggunakan layanan *Google Classroom* (Iftakhar, 2016), Edmodo (Sicat, 2015), dan Schoology (Intergovernmental Panel on Climate Change, 2014), dan aplikasi pesan seperti WhatsApp (So, 2016).

Pembelajaran daring membuat mahasiswa jadi memiliki banyak waktu luang sehingga membuat mereka bisa belajar kapanpun dan di manapun. Selain itu, mahasiswa juga bisa berkomunikasi dengan dosen melalui *e-classroom*, *video conference*, zoom maupun *whatsapp group* (Dhull & Arora, 2019). Kegiatan pembelajaran tersebut merupakan suatu inovasi pendidikan untuk menjawab tantangan akan ketersediaan sumber belajar yang variatif (Fitriyani dkk, 2020).



Pembelajaran daring ternyata masih memiliki permasalahan yang kompleks dikalangan mahasiswa. Koneksi internet, kuota internet yang terbatas, kurang fokus, penggunaan media daring oleh dosen, tugas kuliah yang menumpuk, dan jadwal kuliah yang tidak teratur merupakan masalah yang dihadapi mahasiswa ketika melaksanakan pembelajaran daring. Sebagian mahasiswa tidak mampu mengikuti media daring yang diterapkan oleh dosen (Widodo & Nursaptini, 2020).

## 2.2 Analisis Sentimen

Analisis Sentimen atau bisa disebut *Opinion Mining* adalah pembelajaran tentang menganalisa pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, pelayanan, organisasi, individual, isu, kejadian, topik, dan atribut (Hermanto dkk, 2019). Analisis sentimen adalah suatu ilmu dari *text mining* yang menggunakan alat analitik untuk memahami ekstrak dan memproses data tekstual secara otomatis untuk memperoleh informasi yang terkandung dalam kalimat opini (Inyang dkk, 2017).

Analisis sentimen juga bisa disebut sebagai *opinion mining*, karena berfokus pada pendapat yang menyatakan positif atau negatif (Saputra dkk, 2020). Pada analisis sentimen, pengembangan data perlu dilakukan untuk menganalisis, memproses, dan mengekstrak data tekstual dalam suatu entitas, seperti layanan produk, individu, fenomena, atau topik tertentu (Liu, 2012). Proses analisis dapat mencakup teks ulasan, forum, tweet, atau blog, dengan data *preprocessing* mencakup proses *tokenization*, *stopword*, penghapusan, *stemming*, indentifikasi sentimen, dan klasifikasi sentimen.

### 2.3 Teorema Bayes

Teori probabilitas Bayesian digunakan untuk menghitung probabilitas terjadinya suatu peristiwa yang didasarkan oleh pengaruh yang didapat dari suatu pengujian. Probabilitas Bayes menerangkan hubungan antara probabilitas terjadinya hipotesis  $H_i$  dengan terdapat fakta (*evidence*)  $E$  yang telah terjadi dan probabilitas terjadinya *evidence*  $E$  dengan syarat hipotesis  $H_i$  telah terjadi.

Teori ini didasarkan pada prinsip bahwa jika terdapat tambahan informasi atau *evidence*, maka nilai probabilitas dapat diperbaiki. Oleh karena itu, teori ini bermanfaat untuk mengubah atau memperbaiki nilai kemungkinan yang ada menjadi lebih baik dengan didukung informasi atau *evidence* tambahan. Untuk rumus dari persamaan teori bayes adalah:

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^n p(E|H_k) \times p(H_k)} \quad (2.1)$$

Di mana:

$p(H_i|E)$  = probabilitas hipotesis  $H_i$  benar jika diberikan *evidence* atau fakta  $E$ .

$p(E|H_i)$  = probabilitas munculnya *evidence* (fakta)  $E$  jika diketahui hipotesis  $H_i$  benar.

$p(H_i)$  = probabilitas hipotesis  $H_i$  tanpa memandang *evidence* (fakta) apapun.

$n$  = jumlah hipotesis yang mungkin. (Budiharto & Suhartono, 2014)

Contoh:

Pada penyakit yang diderita oleh 0,1% populasi penduduk terdapat sebuah alat uji dengan akurasi 99% mendeteksi penderita penyakit tersebut. Namun alat tersebut memiliki kemungkinan 0,5% *false positive*, yaitu salah mendeteksi orang sakit

namun dideteksi sakit. Jika seseorang dideteksi sakit oleh alat ini, sebenarnya seberapa persen kemungkinan ia benar-benar sakit?

Jawaban:

$$P(\text{Sakit}|\text{Positif}) =$$

$$\frac{P(\text{Positif}|\text{Sakit}).P(\text{Sakit})}{P(\text{Positif}|\text{Sakit}).P(\text{sakit}) + P(\text{Positif}|\sim\text{Sakit}).P(\sim\text{Sakit})}$$

$$P(\text{Sakit}|\text{Positif}) = \frac{0,99 \cdot 0,001}{0,99 \cdot 0,001 + 0,005 \cdot 0,999}$$

$$P(\text{Sakit}|\text{Positif}) = 0,1654135338 = 16,54\%$$

## 2.4 Naïve Bayes Classifier

*Naïve Bayes* adalah salah satu metode klasifikasi yang begitu populer dalam *machine learning*. *Naïve Bayes* merupakan metode yang bergantung pada dua asumsi. Pertama, fiturnya tidak bergantung pada yang lain. Kedua, setiap fitur memiliki keunggulan yang sama. Berdasarkan asumsi-asumsi tersebut, *Naïve Bayes* berkerja berdasarkan pada probabilitas yang ada untuk menentukan probabilitas di masa depan (Hidayatillah dkk, 2019).

Supaya paham akan hal ini, diperlukan pemahaman mengenai Teorema Bayes. Teorema Bayes adalah algoritma yang bekerja berdasarkan kondisi probabilitas, yang mengukur suatu probabilitas yang akan terjadi berdasarkan kejadian yang sudah terjadi sebelumnya. Persamaan probabilitas yang mendasari *Naïve Bayes* adalah:

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)} \quad (2.2)$$

Di mana  $P(c|x)$  adalah *posterior probability* atau nilai probabilitas yang kita cari,  $P(c)$  adalah kelas probabilitas yang berdasarkan hipotesis (*prior probability*),  $P(x|c)$  adalah prediktor probabilitas yang berdasarkan pada kelas yang diberikan (*likelihood*).  $P(x)$  adalah prediktor probabilitas. Persamaan probabilitas Bayes dapat ditulis:

$$posterior = \frac{likelihood \times prior}{evidence} \quad (2.3)$$

Persamaan probabilitas dari Teorema Bayes dapat disubstitusikan dengan persamaan berikut:

$$P(class|data) = \frac{P(data|class) \times P(class)}{P(data)} \quad (2.4)$$

Di mana kelas adalah tingkat reaksi terdiri dari tiga kategori (*high, medium, low*). Selama data dimasukkan untuk menentukan kelas.

*Naïve Bayes* adalah salah satu metode yang cocok dalam *data mining* yang mengkombinasikan teorema Bayes dengan *Maximum A Posteriori* (MAP). Rumus MAP dapat ditulis sebagai berikut:

$$CMAP = \operatorname{argmax} P(X|c) \times P(c) \quad (2.5)$$

Di mana  $c$  adalah target class, dan  $P(X|c)$  probabilitas dari class yang berdasarkan hipotesis, sementara  $P(C)$  adalah probabilitas prior. Tugasnya adalah untuk mencari class yang benar dari setiap sampel (Hidayatillah dkk, 2019).

Berdasarkan fitur distribusi, *Naïve Bayes Classifier* dapat menjadi Gaussian, Bernoulli atau Multinomial. Gaussian *Naïve Bayes* digunakan ketika berhubungan dengan data kontinyu dengan asumsi bahwa fitur tersebut merupakan distribusi Gaussian.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp - \left( \frac{(x_i - \mu_y)^2}{2\sigma_y^2} \right) \quad (2.6)$$

Parameter  $\sigma_y$  dan  $\mu_y$  dinilai menggunakan *maximum likelihood*. *Bernoulli Naïve Bayes* digunakan untuk data yang didistribusikan menurut distribusi Bernoulli. Pada kasus klasifikasi teks menggunakan *multivariate event model*, vektor kemunculan kata, dari pada vektor perhitungan kata, sering digunakan untuk melatih dan menggunakan klasifikasinya. *Multinomial Naïve Bayes Classifier* digunakan untuk mendistribusi data secara multinomial. Multinomial Naïve Bayes menggunakan vektor perhitungan kata dari pada menggunakan vector pelatihan kata pada saat pelatihan dan menggunakan klasifikasi. Distribusi tersebut diparameterkan oleh vektor  $\theta_y = (\theta_{y1}, \dots, \theta_{yn})$  untuk setiap kelas  $y$ , di mana  $n$  adalah angka fitur (pada klasifikasi teks merupakan ukuran dari kosa kata) dan  $\theta_{yi}$  adalah probabilitas  $P(x_i|y)$  dari fitur  $i$  muncul pada sampel milik class  $y$ . Parameter  $\theta_y$  dinilai oleh versi halus dari *maximum likelihood*, yaitu perhitungan frekuensi relatif:

$$\hat{\theta}_y = \frac{N_{yi} + \alpha}{N_y + \alpha n} \quad (2.7)$$

Di mana,  $N_{yi} = \sum_{x \in T} x_i$  adalah berapa kali kemunculan fitur  $i$  pada sampel class  $y$  pada kumpulan data latih  $T$ , dan  $N_y = \sum_{i=1}^{|T|} N_{yi}$  adalah jumlah perhitungan dari semua fitur untuk class  $y$ . Peluang awal yang mulus  $\alpha \geq 0$  perhitungan untuk fitur yang tidak ada di pembelajaran sampel dan mencegah probabilitas nol pada komputasi lebih jauh. Pengaturan  $\alpha = 1$  disebut penghalusan *Laplace*, sementara  $\alpha < 1$  disebut penghalusan *Lindstone* (Inyang dkk, 2017).

Pada *naïve bayes* setiap *tweet* diwakilkan dalam pasangan atribut  $(a_1, a_2, a_3, \dots, a_n)$  di mana  $a_1$  merupakan kata pertama,  $a_2$  merupakan kata kedua, dan begitu seterusnya, sementara itu  $V$  merupakan himpunan kelas. Ketika klasifikasi, metode ini akan menghasilkan kelas atau kategori yang paling tinggi probabilitasnya ( $V_{MAP}$ ) dengan memasukkan atribut  $(a_1, a_2, a_3, \dots, a_n)$ . Untuk rumus dari  $V_{MAP}$  bisa dilihat pada persamaan (2.8).

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (2.8)$$

Dengan menggunakan teorema bayes, maka persamaan (2.8) bisa ditulis sebagai,

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} \frac{P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (2.9)$$

Nilai dari  $P(a_1, a_2, a_3, \dots, a_n)$  adalah konstan untuk semua  $v_j$  sehingga persamaan (2.9) bisa juga dinyatakan sebagai persamaan

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j) \quad (2.10)$$

*Naïve bayes* menyederhanakan hal tersebut dengan berasumsi bahwa di dalam setiap kategori, setiap atribut bebas bersyarat satu sama lain. Bisa dikatakan bahwa,

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (2.11)$$

Lalu jika persamaan (2.10) disubstitusikan ke persamaan (2.11), maka akan menjadi

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j) \times \prod_i P(a_i | v_j) \quad (2.12)$$

$P(v_j)$  dan probabilitas kata  $a_i$  untuk setiap kategori  $P(a_i | v_j)$  dihitung pada saat *training*, di mana:

$$P(v_j) = \frac{docs_j}{training} \quad (2.13)$$

$$P(a_i|v_j) = \frac{n_i + 1}{n + kosakata} \quad (2.14)$$

Di mana  $docs_j$  merupakan jumlah dokumen pada kategori  $j$  dan  $training$  merupakan jumlah dokumen yang digunakan pada saat proses *training*. Sedangkan  $n_i$  merupakan jumlah kemunculan kata  $a_i$  pada kategori  $v_j$ ,  $n$  merupakan jumlah kosakata yang muncul pada kategori  $v_j$  dan kosakata merupakan jumlah kata unik pada semua data *training* (Rodiyansyah & Winarko, 2013).

## 2.5 Text Mining

*Text Mining* adalah ilmu baru dan menarik dalam penelitian yang mencoba untuk memecahkan masalah dari informasi yang berlebihan dengan menggunakan teknik *data mining*, *machine learning*, *Natural Language Processing* (NLP), *Information Retrieval* (IR) dan pengetahuan manajemen. *Text mining* melibatkan tahapan *preprocessing* pengumpulan data seperti kategorisasi teks, ekstraksi informasi, ekstraksi istilah (Hermanto dkk, 2019).

Proses pengambilan informasi dalam *text mining* dapat menghasilkan analisis perasaan yang secara emosional mengidentifikasi pernyataan jika itu positif ataupun negatif. Objek *text mining* adalah dokumen yang tidak terstruktur atau semi terstruktur. *Text mining* secara efektif mengekstrak informasi yang diperlukan dari sejumlah dokumen (Samsir dkk, 2021).

## 2.6 Data *Pre-processing*

Sebelum melakukan klasifikasi, data akan melalui langkah preparasi terlebih dahulu supaya bisa menjadi data yang diinginkan. Proses ini disebut data *pre-processing*. Tujuannya adalah untuk mendapatkan data dengan representasi yang baik sehingga data menjadi layak. Data *pre-processing* adalah langkah yang harus dilalui dalam data *mining*. Ini dikarenakan, sering terjadinya masalah dalam mengestrak data dalam jumlah data yang besar. Sebagai contoh, informasi yang terkandung dalam data adalah informasi yang berbeda-beda, membuatnya susah dalam memproses (Hidayatillah dkk, 2019).

Dalam hal ini, kita melakukan proses yang namanya pembersihan data. Proses ini adalah proses di mana data tersebut masuk dalam proses *filtering*, *modifying*, dan *removing* data yang tidak diperlukan (Hidayatillah dkk, 2019).

Teknik *data preprocessing* yang digunakan pada analisis sentimen adalah:

1. *Lower text*. Tujuan utama dari *lower text* adalah untuk membuat kata seperti “Hello” dan “hello” tidak diperlakukan sebagai kata yang berbeda padahal kedua kata tersebut adalah kata yang sama. Hal ini juga membantu dalam mengurangi jumlah kata yang dibutuhkan kamus untuk dicari pada satu waktu.
2. *Removal of Punctuation*. Menghapus tanda baca atau kata alfanumerik apapun dari teks aslinya.
3. *Removing Whitespace*. Spasi tidak memiliki arti apapun pada teks, jadi itu dihapus untuk tujuan komputasi.
4. *Tokenization*. *Tokenization* dilakukan untuk memisahkan setiap kalimat pada teks.



5. *Stop word removal*. Kata seperti “a”, “an”, “the” tidak menyediakan arti atau makna apapun pada teks. Maka dari itu, kata-kata seperti itu dihapuskan. Sekarang teks yang relevan digunakan untuk tujuan sentimentil.
6. *Stemming*. *Stemming* adalah mengkonversi kata-kata menjadi bentuk asalnya. Hal ini mengurangi total kata dan membuat proses komputasi menjadi lebih cepat (Pradha dkk, 2019).

## 2.7 *Bag of Words*

Pada model ini, sebuah teks yang berupa kalimat atau dokumen diwakili sebagai kantung (*bag*) multiset kata-kata yang terkandung di dalamnya, tanpa memandang urutan kata dan tata bahasa tetapi tetap mempertahankan keberagamannya. Definisi lain dari model ini adalah model ini mempelajari suatu kosakata dari setiap dokumen, lalu memodelkan setiap dokumen dengan menghitung jumlah kemunculan tiap kata (Inyang dkk, 2017).

Berikut merupakan contoh teks yang diubah ke dalam bentuk *bag of words*.

(D1) “Aku merasa sehat hari ini”

(D2) “Aku merasa demam har ini”

(D3) “Aku berharap aku dapat bermain di luar”

Berdasarkan ketiga kalimat di atas, maka sebuah daftar dibuat untuk setiap dokumen sebagai berikut:

(D1) “aku”, “merasa”, “sehat”, “hari”, “ini”

(D2) “aku”, “merasa”, “demam”, “hari”, “ini”

(D3) “aku”, “berharap”, “aku”, “dapat”, “bermain”, “di”, “luar”.

Lalu setiap kata ini akan dihitung frekuensinya dan dipetakan kembali ke dalam dokumen.

Tabel 2.1 Frekuensi Kata Pada Dokumen

aku	merasa	sehat	hari	ini	demam	berharap	dapat	bermain	di	luar
1	1	1	1	1	0	0	0	0	0	0
1	1	0	1	1	1	0	0	0	0	0
2	0	0	0	0	0	1	1	1	1	1

Tabel di atas menggambarkan fitur *training* yang mengandung frekuensi kata dari setiap kalimat pada sebuah dokumen. Dengan menggunakan pendekatan ini, tata bahasa dan urutan kata tidak dipentingkan, hanya jumlah kemunculan kata yang diperhatikan.

## 2.8 Confusion Matrix

*Confusion Matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. *Confusion Matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan (Rahman dkk, 2017).

Tabel 2.2 *Confusion Matrix*

	<i>Classified as</i>	
	<i>Predicted “+”</i>	<i>Predicted “-”</i>
<i>Actual “+”</i>	<i>True Positives</i>	<i>False Negatives</i>
<i>Actual “-”</i>	<i>False Positives</i>	<i>True Negatives</i>

Berdasarkan tabel *Confusion Matrix* di atas:

- a. ***True Postives (TP)*** adalah jumlah *record* data positif yang diklasifikasikan sebagai nilai positif.
- b. ***False Positives (FP)*** adalah jumlah *record* data negatif yang diklasifikasikan sebagai nilai positif.
- c. ***False Negatives (FN)*** adalah jumlah *record* data positif yang diklasifikasikan sebagai nilai negatif.
- d. ***True Negatives (TN)*** adalah jumlah *record* data negatif yang diklasifikasikan sebagai nilai negatif.

Nilai yang dihasilkan melalui metode *Confusion Matrix* adalah berupa evaluasi sebagai berikut:

- a. *Accuracy*, presentase jumlah *record* data yang diklasifikasikan (prediksi) secara benar oleh algoritma. Rumusnya adalah  $(TP + TN) / \text{Total Data} = \text{Accuracy}$ .
- b. *Misclassification (Error) Rate*, presentase jumlah *record* data yang diklasifikasikan (prediksi secara salah oleh algoritma). Untuk rumusnya,  $(FP + FN) / \text{Total Data} = \text{Misclassification Rate}$ .

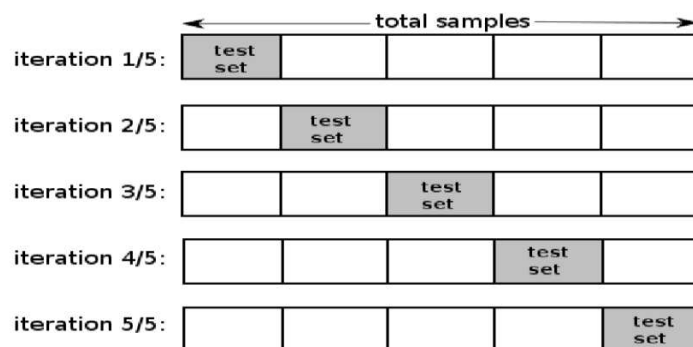
## 2.9 Cross Validation

*Cross Validation* adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma di mana data dipisahkan menjadi dua *subset* yaitu data proses pembelajaran dan data validasi/evaluasi. Model atau algoritma dilatih oleh *subset* pembelajaran dan divalidasi oleh subset validasi (Permana dkk, 2018). Validasi merupakan proses untuk mengevaluasi akurasi

prediksi dari *data mining*. *K-Fold Cross Validation* adalah teknik validasi dengan membagi data secara acak menjadi  $k$  bagian dan setiap bagian akan dilakukan proses klasifikasi.

Contoh pada pengujian data 500 *tweet* data akan dibagi menjadi 5 bagian atau  $k=5$  sehingga data yang diperoleh adalah 500 dibagi 5 lipatan dengan isinya masing-masing adalah 100 data. Selain itu akan ditentukan mana yang termasuk data *testing* dan mana yang termasuk data *training* dengan perbandingan 80-20, sehingga didapatkan 400 data sebagai data *training* dan 100 sisanya sebagai data *testing*. Berdasarkan 5 lipatan yang sudah ditentukan maka akan ada 4 lipatan dikalikan  $100=400$  data *training* dan sisanya 1 lipatan dikalikan  $100 = 100$  data *testing*.

Pengujian menggunakan data yang sudah dipartisi akan diulang sebanyak 5 kali ( $k=5$ ) dengan posisi data tes berbeda disetiap iterasinya. Misalnya iterasi pertama data tes pada posisi awal, iterasi kedua di posisi kedua begitu seterusnya sampai iterasi yang kelima. Gambaran  $k$ -fold dapat dilihat pada gambar di bawah ini.



Gambar 2.1 Iterasi Pada K-Fold

## 2.10 Kajian Keislaman

Agama Islam menyuruh kita untuk menuntut ilmu karena begitu banyak keutamaan yang kita dapat ketika menuntut ilmu. Ketika menuntut ilmu hendaknya kita bersungguh-sungguh dan selalu mempunyai niat untuk mendapatkan ilmu yang bermanfaat. Tuntutlah ilmu seolah-olah tidak pernah kenyang dengan ilmu yang diperoleh, hendaklah kita selalu berkeinginan untuk mencari ilmu. Sebagaimana disebutkan dalam hadits Anas:

مَنْهُوَ مَنْ لَا يَشْبَعَانِ : طَالِبٌ عِلْمٍ وَطَالِبٌ دُنْيَا

*“Ada dua orang yang begitu rakus dan tidak pernah merasa kenyang: yaitu orang yang rakus terhadap ilmu dan tidak pernah kenyang dengannya dan orang yang rakus terhadap dunia dan tidak pernah kenyang dengannya.” (HR. Al-Hakim dalam Al-Mustadrok 1:92. Dishahihkan oleh Al-Hakim dan disepakati oleh Imam Adz-Dzahabi).*

Orang yang tergila-gila tentu tidak akan merasa puas sehingga terus mencari dunia dan dunia. Seperti yang dikatakan dalam hadits Ibnu Abbas, beliau berkata bahwa Rasulullah shallallahu alaihi wa sallam bersabda,

لَوْ أَنَّ لِابْنِ آدَمَ وَادِيًّا مِنْ ذَهَبٍ أَحَبَّ أَنْ يَكُونَ لَهُ وَادِيَانِ ، وَلَنْ يَمْلَأَ فَاهُ إِلَّا التُّرَابُ ، وَيَتُوبُ

اللَّهُ عَلَى مَنْ تَابَ

*“Seandainya seorang anak Adam memiliki satu lembah emas, tentu ia menginginkan dua lembah lainnya, dan sama sekali tidak akan memenuhi mulutnya (merasa puas) selain tanah (yaitu setelah mati) dan Allah menerima taubat orang-orang yang bertaubat.” (Muttafaqun ‘alaih. HR. Bukhari no. 6439 dan muslim no.1048).*

Sedangkan untuk seorang penuntut ilmu akan terus mencari ilmu dan ilmu setiap hari, setiap waktu, dan setiap tempat saking penting dan agungnya ilmu itu.

Agama Islam mengenal istilah sentimen dengan arti pendapat atau pandangan berdasarkan perasaan terhadap sesuatu. Berpendapat adalah potensi dasar yang sebaiknya dikembangkan oleh manusia. Dengan kata lain, Islam mengajarkan bahwa setiap manusia memiliki hak untuk berpendapat dan itu merupakan hal yang

tidak dapat dipisahkan dari potensi sekaligus perintah Allah SWT agar manusia senantiasa berpendapat. Dalam hal berpendapat, Allah SWT memberikan penjelasan dalam surat Asy Syura: 38 yang artinya:

وَالَّذِينَ اسْتَجَابُوا لِرَبِّهِمْ وَأَقَامُوا الصَّلَاةَ وَأَمْرُهُمْ شُورَى بَيْنَهُمْ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ

*“Dan (bagi ) orang-orang yang menerima (mematuhi) seruan Tuhannya dan mendirikan salat, sedang urusan mereka (diputuskan) dengan musyawarah antara mereka; dan mereka menafkahkan sebagian dari rezeki yang Kami berikan kepada mereka.” (Asy-Sura:38)*

Menurut tafsir Zubdatut Tafsir Min Fathil Qadir menyatakan bahwa mereka bermusyawarah dalam urusan-urusan yang umum dan yang khusus tanpa mementingkan dan memaksakan pendapat individu, seperti urusan kepemimpinan, wilayah, masalah hukum, dan perkara-perkara yang khusus. Artinya bahwa musyawarah adalah suatu hal yang lazim dalam menyelesaikan masalah. Oleh Ismail bin Umar Al-Quraisy bin Katsir Al-Bashri Ad-Dimasyqi: Firman Allah Subhanahu wa ta’ala:

*“Dan (bagi) orang-orang yang menerima (mematuhi) seruan Tuhannya.”*

Yakni mereka mengikuti rasul-rasul Allah dan taat kepada perintah-perintah-Nya serta menjauhi larangan-larangan-Nya. “dan mendirikan shalat. Salat adalah ibadah yang paling besar. “sedangkan urusan mereka (diputuskan) dengan musyawarah antara mereka.” Artinya, mereka tidak pernah memutuskan sesuatu urusan melainkan terlebih dahulu mereka musyawarahkan di antara sesamanya agar masing-masing dari mereka mengemukakan pendapatnya. Seperti dalam menghadapi urusan perang dan lain sebagainya yang penting, sebagaimana yang disebutkan dalam ayat lain melalui firman-Nya surah Al-Imran ayat 159:

Karena itulah Rasulullah selalu bermusyawarah dengan para sahabat saat menghadapi peperangan dan urusan penting lainnya, sehingga dengan demikian hati mereka merasa senang dan lega. Hal yang sama telah dilakukan oleh Khalifah Umar ibnul Kahttab r.a. saat menjelang ajalnya karena tertusuk, ia menjadikan urusan kekhalifaan sesudahnya agar dimusyawarahkan di antara sesama mereka untuk memilih salah seorang dari enam orang berikut, yaitu Usman, Ali, Talhah, Az-Zubair, Sa'ad, dan Abdurrahman Ibnu Auf. Maka Akhirnya pendapat semua sahabat sepakat menunjuk sahabat Usman ibnu Affan r.a. sebagai khalifah sesudah Umar.r.a. “dan mereka menafkahkan sebagian dari rezeki yang kami berikan kepada mereka.” Yang demikian itu terealisasi dengan berbuat kebaikan makhluk Allah yang paling dekat dengan mereka, lalu berikutnya adalah orang-orang yang dekat dengan mereka.

Orang yang bermusyawarah haruslah orang yang berilmu pengetahuan. Bagaimana sesuai hadits.

أَخْرَجَهُ الْخَطِيبُ عَنْ أَبِي هُرَيْرَةَ مَرْفُوعًا اسْتَزْشِدُوا الْعَاقِلَ تَرْشِدُوا وَلَا تَعْصُوهُ فَتَنْدَمُوا

*“Hadits di-ikharj-kan oleh Al-Khatib dari Abi Hurairah marfu’ : “mintalah petunjuk kepada orang yang berakal maka kamu akan mendapatkan petunjuk, jangan berbuat maksiat nanti kamu akan menyesal.”*

Hadits ini memerintahkan agar memintak petunjuk dan arahan kepada orang yang berilmu pengetahuan. Menetapkan peserta musyawarah haruslah orang yang profesional dan berilmu pengetahuan agar mendapatkan suatu kepastian.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Pendekatan Penelitian**

Pada penelitian ini, pendekatan yang digunakan adalah dengan pendekatan studi kasus deskriptif kuantitatif. Pada studi kasus yaitu dengan pengumpulan data menggunakan kuisioner yang disebarakan kepada responden. Pada penelitian ini penulis menggunakan *Google Form*. Sedangkan pendekatan deskriptif kuantitatif menganalisis data dan menyusun data yang telah ada agar sesuai dengan kebutuhan peneliti.

#### **3.2 Variabel Penelitian**

Pada penelitian ini, variabel penelitian dibagi menjadi dua, yaitu variabel bebas dan variabel terikat. Pada variabel bebas yaitu beberapa opini mahasiswa tentang pembelajaran daring, sedangkan untuk variabel terikatnya adalah label positif dan negatif yang diberikan sendiri oleh para responden.

#### **3.3 Jenis dan Sumber Data**

Pada penelitian ini data yang digunakan adalah data yang berjenis teks opini mahasiswa terhadap perkuliahan daring dan responden juga yang menentukan label dari opini tersebut, apakah opini tersebut termasuk opini positif atau opini negatif. Data bersumber dari responden menggunakan instrumen berupa kuisioner. Pada instrumen tersebut responden akan menjawab pertanyaan tentang bagaimana pendapat mereka tentang perkuliahan daring. Data yang digunakan hanya sebagai



simulasi pada program saja. Responden pada penelitian ini adalah mahasiswa dari beberapa universitas.

Tabel 3.1 Data Responden

No.	Universitas	Jumlah
1	UIN Maulana Malik Ibrahim Malang	84
2	Universitas Sam Ratulangi	30
3	Politeknik Negeri Manado	8
4	Universitas Gunadarma	7
5	Universitas Brawijaya	2
6	Politeknik Penerbangan Surabaya	1
7	Universitas Terbuka	1
8	Universitas Muhammadiyah Yogyakarta	2
9	Universitas Hasanudin	1
10	Univeristas Darussalam Gontor	1
11	Universitas Negeri Gorontalo	1
12	UIN Sunan Gunung Jati Bandung	1
13	Universitas Negeri Malag	2
14	PTIQ Jakarta	1
15	Asia Pasific University Malaysia	1
16	Universitas Pesantren Tinggi Darul Ulum Jombang	1
17	Universitas Padjadjaran	1
18	Universitas Internasional Semen Indonesia	1
19	UNIKOM	1

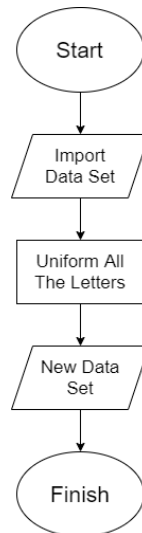
20	Universitas Syiah Kuala	2
21	Universitas Diponegoro	1
22	Universitas Maritim Raja Ali Haji	1
23	Respati Indonesia	1
24	Universitas Ahmad Dahlan	1
25	IAIN Kediri	1
26	IPB	1
27	Universitas Indonesia	1
28	UNIDA	1
Total		157

### 3.4 Pra-pemrosesan Data

Sebelum melakukan klasifikasi, data dipreparasi dulu sehingga mendapatkan data yang diinginkan. Tahap ini dinamakan pra-pemrosesan data. Tahap untuk melakukan pra-pemrosesan data adalah sebagai berikut:

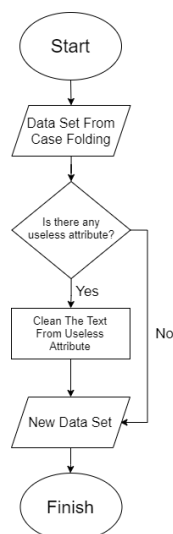
#### 1. *Case Folding*

Pada tahap ini, semua dokumen akan diubah menjadi huruf kecil. Hal ini dilakukan agar tidak terjadi *redudansi* atau pengulangan data yang hanya berbeda pada hurufnya saja. Berikut merupakan *flowchart* dari *case folding* yang bisa dilihat pada Gambar 3.1.

Gambar 3.1 Proses *Case Folding*

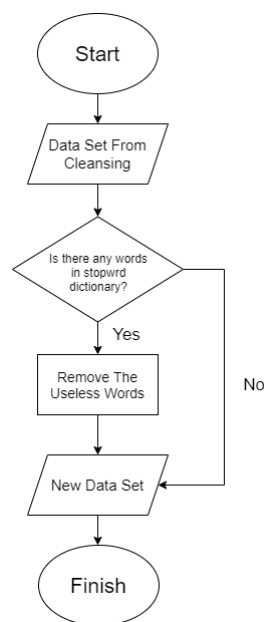
## 2. *Cleansing*

Pada tahap ini, semua kata dibersihkan dari kata-kata yang tidak berpengaruh sama sekali dengan analisis sentimen. Komponen pada dokumen memiliki berbagai macam atribut seperti karakter simbol atau tanda baca. Atribut yang tidak berhubungan ini akan dihilangkan dari dokumen dan diganti dengan karakter spasi. Berikut merupakan *flowchart* dari proses *cleansing* yang bisa dilihat pada Gambar 3.2.

Gambar 3.2 Proses *Cleansing*

### 3. *Stopword Removal*

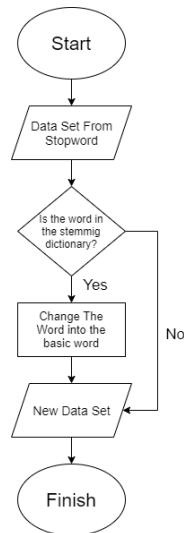
Tahap ini merupakan tahap di mana semua kata yang tidak sesuai dengan topik dihilangkan, jika ada kata yang tidak memengaruhi akurasi klasifikasi maka kata akan dihilangkan. Kata-kata yang termasuk pada *stopword* diantaranya “dengan”, “ada”, “harus”, “kepada”, dan lain-lain. Pada penelitian ini, peneliti menghapus kata ‘tidak’ pada kamus *stopwords* karena kata ‘tidak’ merupakan kata yang mengandung makna negatif. Berikut merupakan *flowchart* dari proses *stopword removal* yang bisa dilihat pada Gambar 3.3.



Gambar 3.3 Proses *Stopword Removal*

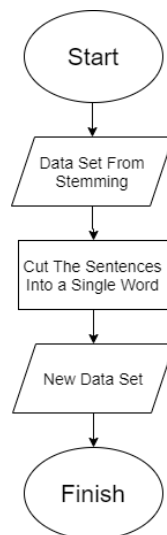
### 4. *Stemming*

Pada tahap ini, semua kata diubah ke dalam bentuk dasarnya dengan menghapus imbuhan pada depan kata maupun imbuhan pada belakang kata. Berikut merupakan *flowchart* dari proses *stemming* yang bisa dilihat pada Gambar 3.4.

Gambar 3.4 Proses *Stemming*

## 5. Tokenisasi

Pada tahap ini, kalimat dipecah menjadi beberapa bagian yang disebut token. Sebuah token dianggap sebagai bentuk kata atau frasa yang memiliki arti. Berikut merupakan *flowchart* dari proses tokenisasi yang bisa dilihat pada Gambar 3.5.



Gambar 3.5 Proses Tokenisasi

### 3.5 Analisis Data

Data yang digunakan yaitu data teks yang dikumpulkan dari tanggal 17 Oktober 2020 sampai 6 Desember 2020. Jumlah data yang dikumpulkan adalah sebanyak 100 data teks opini tentang perkuliahan daring. Data terdiri dari nama mahasiswa, jenis kelamin, universitas, fakultas/Program Studi/angkatan, komentar, dan jenis komentar apakah positif atau negatif. Pada data tersebut responden lah yang menentukan apakah opini yang diberikan positif atau negatif.

Setelah data diproses, data tersebut akan diuji dengan menggunakan metode *Naïve Bayes Classifier* pada Bab 4 hasil dan pembahasan. Adapun langkah-langkah yang digunakan untuk menganalisis sentimen mahasiswa terhadap perkuliahan daring, yaitu sebagai berikut:

1. Menguji model *Naïve Bayes Classifier* pada data uji

Pada tahap ini dilakukan proses klasifikasi menggunakan *Naïve Bayes Classifier*. Pada tahap ini dilakukan proses *training* terlebih dahulu, selanjutnya akan dilakukan proses *testing* dengan mengacu pada probabilitas dari data *training*. Untuk persamaan pada proses *training* menggunakan persamaan (6) dan (7) pada kajian teori.

2. Prediksi hasil *test set*

Pada tahap ini dihitung dan dicari probabilitas tertinggi untuk setiap kata sentimen. Pada tahap ini menggunakan perhitungan *Vmap*. *Vmap* adalah perhitungan yang digunakan *naïve bayes* untuk menentukan probabilitas data uji masing-masing kelas berdasarkan proses *learning*. Nilai probabilitas yang akan dipilih. (Seminar Nasional Teknologi dan Informatika, 2017). Pada tahap ini menggunakan persamaan (5) pada kajian teori.

### 3. Membuat *confusion matrix*

*Confusion matrix* diperlukan untuk melihat akurasi dari analisis sentimen.

*Confusion matrix* menggambarkan kinerja model klasifikasi pada data uji yang nilainya sudah diketahui. Beberapa contoh dari *confusion matrix* adalah, [1,1], [1,0], [0,0], [0,1] di mana [1,1] adalah hasil klasifikasi positif, [0,0] adalah klasifikasi negative, [0,1] adalah klasifikasi salah negative dan [1,0] adalah prediksi salah positif.

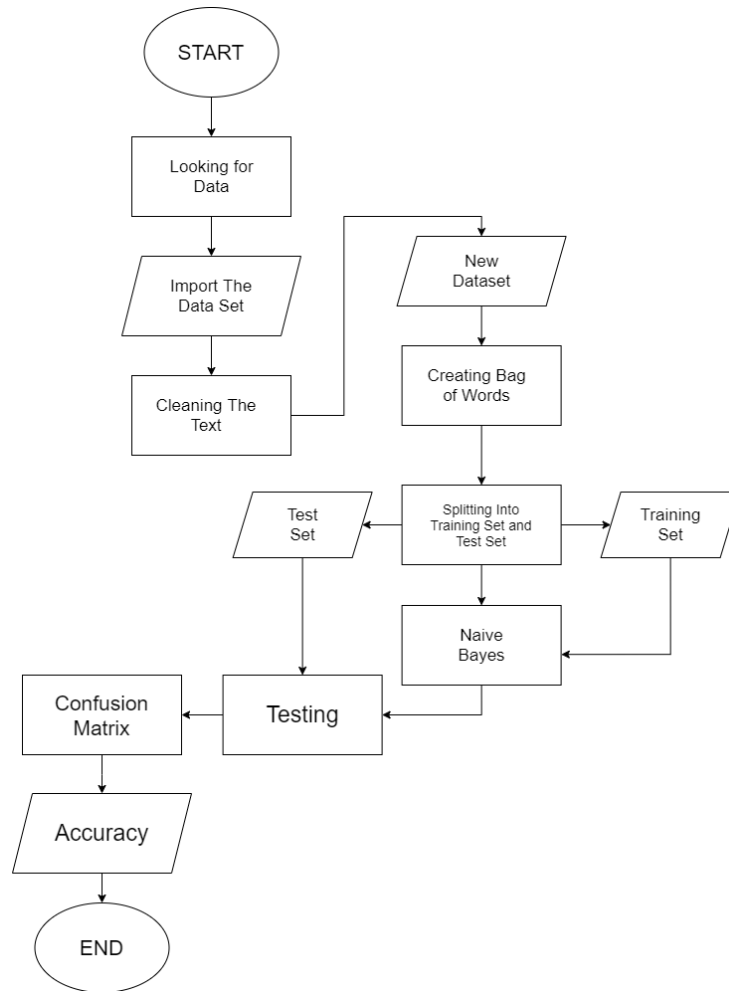
### 4. Menghitung nilai akurasi

Nilai akurasi dihitung dari berdasarkan *confusion matrix*. Untuk rumus yang dipakai adalah:

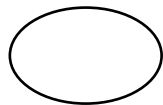
$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

Untuk menentukan hasil uji yang maksimal dilakukan pengujian kedua dengan menggunakan *k-fold cross validation*.

Gambar 3.1 merupakan *flowchart* klasifikasi analisis sentimen mahasiswa terhadap perkuliahan daring menggunakan metode *Naïve Bayes Classifier*.

Gambar 3.6 *Flowchart* Klasifikasi Analisis Sentimen

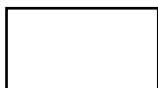
Keterangan:



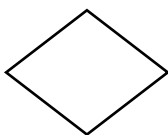
= Simbol untuk permulaan atau akhir dari suatu kegiatan



= Simbol yang menyatakan proses input atau output tanpa tergantung dengan jenis peralatannya



= Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer



= Perbandingan pernyataan, penyeleksian data yang memberikan pilihan untuk langkah selaniutnva



## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Bab ini menjelaskan tentang implementasi terhadap tahapan-tahapan yang sudah dipaparkan sebelumnya pada bab 3 mengenai metodologi. Kemudian, setiap tahapan akan dijelaskan dan ditunjukkan hasilnya.

#### **4.1 Pra-pemrosesan Data**

Pada tahap ini akan dilakukan pra-pemrosesan terhadap teks komentar mahasiswa yang digunakan sebagai data. Untuk menghilangkan *noise* dan memperjelas fitur, proses tokenisasi dan konversi data perlu dilakukan dalam pra-pemrosesan data. Tahap ini bertujuan untuk membersihkan data dan menghapus atribut-atribut yang tidak diperlukan seperti tanda baca. Proses ini akan mengelolah teks menjadi data yang mudah diterima oleh sistem. Tahap pra-pemrosesan data diperlukan sebelum melakukan tahap klasifikasi.

Kode yang dibuat untuk mengimplementasikan tahap ini ditulis dengan menggunakan bahasa pemrograman *python* 3.7. Untuk kode bisa dilihat pada bagian lampiran penelitian ini.

##### **4.1.1 Case Folding**

*Case folding* dilakukan untuk menyeragamkan huruf yang digunakan menjadi huruf kecil. Berikut merupakan 3 dokumen yang sudah diterapkan *case folding*.

Tabel 4.1 Proses *Case Folding*

Sebelum	Sesudah
menurut saya, selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah. Tapi, dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan.	menurut saya, selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah. <b>tapi</b> , dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan.

Tabel di atas menunjukkan penyeragaman semua kata menjadi huruf kecil semua. Kata yang ditunjukkan pada data di atas adalah kata yang ditandai dengan warna merah yaitu kata “tapi”.

Tabel 4.2 Proses *Case Folding*

Sebelum	Sesudah
Tidak nyaman saat kuliah daring karena materi yang di berikan tidak dimengerti sehingga sulit masuk ke kepala juga sering stress di rumah karena terlalu bosan hingga pada saat kuliah daring materi tidak diserap dengan baik.	<b>tidak</b> nyaman saat kuliah daring karena materi yang di berikan tidak dimengerti sehingga sulit masuk ke kepala juga sering stress di rumah karena terlalu bosan hingga pada saat kuliah daring materi tidak diserap dengan baik.

Sementara pada tabel kedua menunjukkan penyeragaman semua kata menjadi huruf kecil. Kata yang diubah pada tabel di atas adalah kata yang ditandai dengan warna merah yaitu kata “tidak”.

Tabel 4.3 Proses *Case Folding*

Sebelum	Sesudah
Kuliah online memiliki kekurangan dan kelebihan tersendiri. Untuk kelebihanannya salah satunya untuk memutus rantai penyebaran covid-19 dan lebih mengajarkan kepada mahasiswa untuk lebih mandiri lagi seperti belajar mandiri pokoknya semua pekerjaan dilakukan sendiri dan untuk diri sendiri. Untuk kekurangannya kuliah online ini sedikit menyusahkan mahasiswa, dimana kita tahu kondisi semua mahasiswa itu tidak sama. Kadang kuliah online ini bagi mahasiswa bisa menjadi beban namun juga bisa menjadi sarana pembelajaran jarak jauh .	<b>kuliah</b> online memiliki kekurangan dan kelebihan tersendiri. <b>untuk</b> kelebihanannya salah satunya untuk memutus rantai penyebaran covid-19 dan lebih mengajarkan kepada mahasiswa untuk lebih mandiri lagi seperti belajar mandiri pokoknya semua pekerjaan dilakukan sendiri dan untuk diri sendiri. <b>untuk</b> kekurangannya kuliah online ini sedikit menyusahkan mahasiswa, dimana kita tahu kondisi semua mahasiswa itu tidak sama. Kadang kuliah online ini bagi mahasiswa bisa menjadi beban namun juga bisa menjadi sarana pembelajaran jarak jauh .

Pada tabel ketiga menunjukkan penyeragaman semua kata menjadi huruf kecil. Kata yang diubah adalah kata yang ditandai dengan warna merah kata “kuliah” dan “untuk”.

#### 4.1.2 *Cleansing*

Setelah tahap *case folding* dilakukan, tahap berikutnya adalah *cleansing* untuk menghilangkan tanda baca, simbol, angka, atau karakter yang bukan huruf. Berikut merupakan 3 dokumen yang sudah diterapkan tahap *cleansing*.

Tabel 4.4 Proses *Cleansing*

Sebelum	Sesudah
menurut saya, selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah. tapi, dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan.	menurut saya selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah tapi dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan

Pada tabel di atas menunjukkan suatu data yang sudah bersih dari tanda baca titik dan koma.

Tabel 4.5 Proses *Cleansing*

Sebelum	Sesudah
tidak nyaman saat kuliah daring karena materi yang di berikan tidak dimengerti sehingga sulit masuk ke kepala juga sering stress di rumah karena terlalu bosan hingga pada saat kuliah daring materi tidak diserap dengan baik.	tidak nyaman saat kuliah daring karena materi yang di berikan tidak dimengerti sehingga sulit masuk ke kepala juga sering stress di rumah karena terlalu bosan hingga pada saat kuliah daring materi tidak diserap dengan baik

Pada tabel di atas menunjukkan suatu data yang sudah bersih dari tanda baca. Tanda baca yang dihilangkan adalah tanda titik.

Tabel 4.6 Proses *Cleansing*

Sebelum	Sesudah
kuliah online memiliki kekurangan dan kelebihan tersendiri. untuk kelebihannya salah satunya untuk memutus rantai penyebaran covid-19 dan lebih mengajarkan kepada mahasiswa untuk lebih mandiri lagi seperti belajar mandiri pokoknya semua pekerjaan dilakukan sendiri dan untuk diri sendiri. untuk kekurangannya kuliah online ini sedikit menyusahkan mahasiswa, dimana kita tahu kondisi semua mahasiswa itu tidak sama. kadang kuliah online ini bagi mahasiswa bisa menjadi beban namun juga bisa menjadi sarana pembelajaran jarak jauh .	kuliah online memiliki kekurangan dan kelebihan tersendiri untuk kelebihannya salah satunya untuk memutus rantai penyebaran covid dan lebih mengajarkan kepada mahasiswa untuk lebih mandiri lagi seperti belajar mandiri pokoknya semua pekerjaan dilakukan sendiri dan untuk diri sendiri untuk kekurangannya kuliah online ini sedikit menyusahkan mahasiswa dimana kita tahu kondisi semua mahasiswa itu tidak sama kadang kuliah online ini bagi mahasiswa bisa menjadi beban namun juga bisa menjadi sarana pembelajaran jarak jauh

Pada tabel di atas menunjukkan bahwa data sudah bersih dari tanda baca, huruf dan karakter yang bukan huruf. Tanda baca yang dihilangkan adalah tanda baca titik dan koma, angka yang dihilangkan adalah angka 19 pada kata covid-19, dan karakter bukan huruf yang dihilangkan adalah karakter “-” pada kata covid-19.

#### 4.1.3 *Stopword Removal*

Pada tahap ini, kata-kata yang kurang bermakna atau tidak memiliki arti akan dilakukan penghapusan. Hal ini dilakukan karena kata pada *stopword* tidak memengaruhi akurasi dalam klasifikasi sentimen. Jika ada kata pada dokumen yang sama dengan kata dalam *stopword*, maka kata tersebut akan dihilangkan dan diganti dengan spasi. Berikut adalah 3 dokumen yang sudah diterapkan tahap *stopword removal*.

Tabel 4.7 *Stopword Removal*

Sebelum	Sesudah
menurut saya selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah tapi dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan	kuliah online sistem belajar teratur kuliah kegiatan belajar dirumah kuliah online kadang mengalami kendala jaringan tidak jaringan daerah susah jaringan
tidak nyaman saat kuliah daring karena materi yang di berikan tidak dimengerti sehingga sulit masuk ke kepala juga sering stress di rumah karena terlalu bosan hingga pada saat kuliah daring materi tidak diserap dengan baik	tidak nyaman kuliah daring materi tidak dimengerti sulit masuk kepala stress rumah bosan kuliah daring materi tidak diserap
kuliah online memiliki kekurangan dan kelebihan tersendiri untuk kelebihannya salah satunya untuk memutus rantai penyebaran covid dan lebih mengajarkan kepada mahasiswa untuk lebih mandiri lagi seperti belajar mandiri pokoknya semua pekerjaan dilakukan sendiri dan untuk diri sendiri untuk kekurangannya kuliah online ini sedikit menyusahkan mahasiswa dimana kita tahu kondisi semua mahasiswa itu tidak sama kadang kuliah online ini bagi mahasiswa bisa menjadi beban namun juga bisa menjadi sarana pembelajaran jarak jauh	kuliah online memiliki kekurangan kelebihan tersendiri kelebihannya salah satunya memutus rantai penyebaran covid mengajarkan mahasiswa mandiri belajar mandiri pokoknya pekerjaan kekurangannya kuliah online menyusahkan mahasiswa dimana kondisi mahasiswa tidak kadang kuliah online mahasiswa beban sarana pembelajaran jarak
Dst...	Dst...

#### 4.1.4 *Stemming*

Setelah tahap *stopword removal* dilakukan, tahap selanjutnya adalah melakukan *stemming*. Pada tahap *stemming*, semua kata diubah ke dalam bentuk

dasarnya dengan menghapus kata imbuhan di depan maupun di belakang kata.

Berikut adalah 3 dokumen yang sudah diterapkan tahap *stemming*.

Tabel 4.8 Proses *Stemming*

Sebelum	Sesudah
kuliah online sistem belajar teratur kuliah kegiatan belajar dirumah kuliah online kadang mengalami kendala jaringan tidak jaringan daerah susah jaringan	kuliah online sistem ajar atur kuliah giat ajar rumah kuliah online kadang alami kendala jaring tidak jaring daerah susah jaring
tidak nyaman kuliah daring materi tidak dimengerti sulit masuk kepala stress rumah bosan kuliah daring materi tidak diserap	tidak nyaman kuliah daring materi tidak erti sulit masuk kepala stress rumah bosan kuliah daring materi tidak serap
kuliah online memiliki kekurangan kelebihan tersendiri kelebihannya salah satunya memutus rantai penyebaran covid mengajarkan mahasiswa mandiri belajar mandiri pokoknya pekerjaan kekurangannya kuliah online menyusahkan mahasiswa dimana kondisi mahasiswa tidak kadang kuliah online mahasiswa beban sarana pembelajaran jarak	kuliah online milik kurang lebih sendiri lebih salah satu putus rantai sebar covid ajar mahasiswa mandiri ajar mandiri pokok kerja kurang kuliah online susah mahasiswa mana kondisi mahasiswa tidak kadang kuliah online mahasiswa beban sarana ajar jarak
Dst...	Dst...

#### 4.1.5 Tokenisasi

Tokenisasi merupakan langkah di mana memisahkan kata, simbol, frase, atau entitas penting yang lain dari sebuah teks dokumen. Berdasarkan hasil tokenisasi yang dilakukan, maka penulis juga ingin mengetahui frekuensi kata yang paling banyak dibicarakan oleh para mahasiswa. Untuk itu, penulis melakukan visualisasi menggunakan *wordcloud*. Hasil visualisasi tokenisasi bisa dilihat pada Gambar 4.1.





Gambar di atas mengandung frekuensi kata dari tiap kalimat dalam sebuah dokumen. Dengan pendekatan ini, tata bahasa dan urutan kata tidak begitu penting, hanya jumlah kata yang diperhatikan. Pada model *bag of words*, setiap dokumen dijadikan seperti vektor dengan elemen, jika terdapat kata yang bersangkutan dengan dokumen maka diberikan nilai 1, jika tidak ada maka diberikan nilai 0.

#### **4.1.7 Pemisahan Data Set Menjadi Data Latih dan Data Uji**

Memisahkan data menjadi merupakan hal yang diperlukan dalam mengevaluasi suatu model. Sebagian besar data akan menjadi *training set* dan sisanya akan menjadi *test set*. Pemisahan ini menggunakan cara komputasi dengan membuat dua variabel yaitu variabel X dan variabel Y. Variabel X adalah data dari sentimen yang sudah melewati tahap pra-pemrosesan sebelumnya, sedangkan variabel Y adalah label dari sentimen. Data dipisahkan menjadi 80% untuk data latih dan 20% untuk data uji. Setelah dibuat programnya dengan menggunakan bahasa pemrograman *python* maka hasilnya akan menjadi seperti ini:

Tabel 4.9 Pemisahan Data Latih dan Data Uji

Data Latih	Data Uji
Data latih X = [[0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0], ..., [0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0]]	Data uji X = [[0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 1, 0, 0], [0, 0, 0, ..., 3, 0, 0], ..., [0, 0, 0, ..., 0, 1, 0], [0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0]]
Data latih Y = [0, 1, 0, 0, 0, 1, 0, 0, ..., 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0]	Data uji Y = [0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0]

## 4.2 Hasil Klasifikasi Sentimen Mahasiswa Terhadap Perkuliahan Daring

Klasifikasi sentimen ini dilakukan secara otomatis dengan mengimplementasikan metode *naïve bayes*. Berdasarkan tabel data 4.10, diprediksi hasil pada data uji sebagai berikut:

Tabel 4.10 Prediksi Pada Data Uji

Positif	Negatif
[[1,1], [1,1], [1,1]]	[[0,0], [0,0]]

Hasil di atas adalah hasil prediksi sentimen positif dan negatif menggunakan *naïve bayes*. [1,1] adalah sentimen positif dan [0,0] adalah sentimen negatif. Sejauh ini, tidak semua kalimat terklasifikasi sesuai dengan sentimen yang diinginkan. Namun, sebagian besar sudah terklasifikasi sesuai dengan hasil yang diinginkan. Dari 20% data yang diuji, 21 sentimen diprediksi dalam kategori negatif, 3 sentimen diprediksi ke dalam kategori yang positif, 5 sentimen merupakan prediksi yang salah dari kategori positif, dan 3 sentimen sisanya merupakan prediksi yang salah dari kategori yang negatif.

Sebagai contoh pada perhitungan manual terdapat 6 data yang berhasil didapatkan dan sudah melalui tahap *pre-processing* seperti pada tabel 4.11.

Tabel 4.11 Contoh Data *Training*

No.	Komentar	Kategori
1.	kuliah online tidak efektif	Negatif
2.	tidak kuliah online	Negatif
3.	kuliah online gila	Negatif
4.	semoga kuliah online	Positif
5.	terlanjur nyaman kuliah online	Positif
6.	nilai bagus kuliah online	Positif

Dari tabel di atas dibuat suatu model probabilitas dengan menggunakan persamaan (6) pada kajian teori.

$$P(v_{positif}) = \frac{3}{6} = \frac{1}{2}$$

$$P(v_{negatif}) = \frac{3}{6} = \frac{1}{2}$$

Selanjutnya dihitung probabilitas setiap kata dengan menggunakan persamaan (7) pada kajian teori. Jika dibuat menjadi tabel, maka probabilitas setiap kata pada data *training* akan seperti pada tabel 4.12, 4.13, 4.14, 4.15, 4.16, dan 4.17.

Tabel 4.12 Perhitungan Probabilitas Setiap Kata Pada Komentar 1

Kategori	$P(v_j)$	$P(a_i v_j)$			
		kuliah	online	tidak	efektif
Positif	$\frac{1}{2}$	$\frac{4}{21}$	$\frac{4}{21}$	$\frac{1}{21}$	$\frac{1}{21}$
Negatif	$\frac{1}{2}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{3}{20}$	$\frac{2}{20}$

Tabel 4.13 Perhitungan Probabilitas Setiap Kata Pada Komentar 2

Kategori	$P(v_j)$	$P(a_i v_j)$			
		tidak	kuliah	online	lagi
Positif	$\frac{1}{2}$	$\frac{1}{21}$	$\frac{4}{21}$	$\frac{4}{21}$	$\frac{1}{21}$
Negatif	$\frac{1}{2}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{2}{20}$

Tabel 4.14 Perhitungan Probabilitas Setiap Kata Pada Komentar 3

Kategori	$P(v_j)$	$P(a_i v_j)$		
		kuliah	online	gila
Positif	$\frac{1}{2}$	$\frac{4}{21}$	$\frac{4}{21}$	$\frac{1}{21}$
Negatif	$\frac{1}{2}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{2}{20}$

Tabel 4.15 Perhitungan Probabilitas Setiap Kata Pada Komentar 4

Kategori	$P(v_j)$	$P(a_i v_j)$		
		semoga	kuliah	online
Positif	$\frac{1}{2}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{4}{21}$
Negatif	$\frac{1}{2}$	$\frac{1}{20}$	$\frac{4}{20}$	$\frac{4}{20}$

Tabel 4.16 Perhitungan Probabilitas Setiap Kata Pada Komentar 5

Kategori	$P(v_j)$	$P(a_i v_j)$			
		terlanjur	nyaman	kuliah	online
Positif	$\frac{1}{2}$	$\frac{2}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{4}{21}$
Negatif	$\frac{1}{2}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{4}{20}$	$\frac{4}{20}$

Tabel 4.17 Perhitungan Probabilitas Setiap Kata Pada Komentar 6

Kategori	$P(v_j)$	$P(a_i v_j)$			
		nilai	bagus	kuliah	online
Positif	$\frac{1}{2}$	$\frac{2}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{4}{21}$
Negatif	$\frac{1}{2}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{4}{20}$	$\frac{4}{20}$

Hasil perhitungan probabilitas di atas digunakan sebagai model probabilistik yang selanjutnya digunakan sebagai data acuan dalam menentukan data testing. Berikut merupakan contoh dari data testing yang sudah melalui tahap *pre-processing*.

Tabel 4.18 Contoh Data *Testing*

No.	Komentar	Kategori
1.	kuliah online lulus	?
2.	tidak suka kuliah online	?
3.	nilai baik kuliah online	?

Pada proses ini, dihitung probabilitas dan dicari probabilitas tertingginya menggunakan persamaan (5) pada kajian teori.

$$P(Komentar_1|Positif)$$

$$= P(a_{kuliah}|V_{positif}) \times P(a_{online}|V_{positif}) \times P(a_{lulus}|V_{positif})$$

$$= \frac{4}{21} \times \frac{4}{21} \times 1 \times \frac{1}{2}$$

$$= 0,01814$$

$$P(Komentar_1|Negatif)$$

$$= P(a_{kuliah}|V_{negatif}) \times P(a_{online}|V_{negatif}) \times P(a_{lulus}|V_{negatif})$$

$$= \frac{4}{20} \times \frac{4}{20} \times 1 \times \frac{1}{2}$$

$$= 0,02$$

$$P(Komentar_2|V_{positif})$$

$$= P(a_{tidak}|V_{positif}) \times P(a_{suka}|V_{positif}) \times P(a_{kuliah}|V_{positif})$$

$$\times P(a_{online}|V_{positif})$$

$$= \frac{1}{21} \times 1 \times \frac{4}{21} \times \frac{4}{21} \times \frac{1}{2}$$

$$= 0,00086$$

$$\begin{aligned}
& P(Komentar_2|V_{negatif}) \\
&= P(a_{tidak}|V_{negatif}) \times P(a_{suka}|V_{negatif}) \times P(a_{kuliah}|V_{negatif}) \\
&\times P(a_{online}|V_{negatif}) \\
&= \frac{3}{20} \times 1 \times \frac{4}{20} \times \frac{4}{20} \times \frac{1}{2} \\
&= 0,003P(Komentar_3|V_{positif}) \\
&= P(a_{nilai}|V_{positif}) \times P(a_{baik}|V_{positif}) \times P(a_{kuliah}|V_{positif}) \\
&\times P(a_{online}|V_{positif}) \\
&= \frac{2}{21} \times 1 \times \frac{4}{21} \times \frac{4}{21} \times \frac{1}{2} \\
&= 0,00173 \\
&P(Komentar_3|V_{positif}) \\
&= P(a_{nilai}|V_{negatif}) \times P(a_{baik}|V_{negatif}) \times P(a_{kuliah}|V_{negatif}) \\
&\times P(a_{online}|V_{negatif}) \\
&= \frac{1}{20} \times 1 \times \frac{4}{20} \times \frac{4}{20} \times \frac{1}{2} \\
&= 0,001
\end{aligned}$$

Setelah menghitung probabilitas dari setiap data *testing* maka didapatkan hasil pada tabel 4.19.



Tabel 4.19 Nilai Probabilitas Data *Testing*

Komentar	Probabilitas	
	Positif	Negatif
Komentar 1	0,01814	<b>0,02</b>
Komentar 2	0,00086	<b>0,003</b>
Komentar 3	<b>0,00173</b>	0,001

Pada tabel di atas dapat dianalisis dua dari ketiga data *testing* memiliki nilai probabilitas negatif lebih besar sehingga dapat disimpulkan bahwa ketiga komentar di atas termasuk komentar yang negatif. Sementara untuk komentar 3 memiliki nilai probabilitas positif yang lebih besar, sehingga komentar 3 merupakan komentar yang positif atau mendukung.

Setelah proses klasifikasi dilakukan dengan menghitung probabilitas antar kalimat terhadap setiap kelas, barulah bisa didapatkan hasil prediksi dari data yang dimasukkan. Setelah semua proses dilakukan, selanjutnya penenliti akan menghitung performa atau akurasi yang didapatkan dari model yang digunakan.

### 4.3 Akurasi Metode *Naïve Bayes Classifier*

Untuk mengetahui performa dari *Naïve Bayes Classifier*, maka dilakukan pengujian terhadap model. Hasil dari klasifikasi akan dimunculkan dalam bentuk *confusion matrix*. Kelas dari *confusion matrix* terdiri dari kelas *predicted* dan *actual*. Model dari *confusion matrix*  $2 \times 2$  ditunjukkan pada tabel 4.20.

Tabel 4.20 Model *Confusion Matrix*

Correct Classification		Predict Class	
		Predicted “+”	Predicted “-”
Actual Class	Actual “+”	21	3
	Actual “-”	5	3

Untuk nilai akurasi dari suatu model diperoleh dari jumlah data yang tepat hasil klasifikasi dibagi dengan total data, untuk rumus yang digunakan adalah

$$\begin{aligned}
 & \frac{TP + TF}{TP + FP + TN + FN} \\
 &= \frac{21 + 3}{21 + 3 + 3 + 5} \\
 &= 0,75
 \end{aligned}$$

Nilai akurasi yang didapatkan sebesar **0,75** yang dihitung berdasarkan jumlah nilai diagonal dari *confusion matrix* dibagi dengan jumlah data. Untuk performa metode klasifikasi dari setiap kelas dapat dilihat dari nilai *precision*, *recall*, dan *f1-score* pada setiap kelas. Hasil dari nilai *precision*, *recall*, dan *f1-score* mempunyai nilai sebesar 0-1. Semakin tinggi nilainya maka semakin baik. Hasil dari keseluruhan proses evaluasi model dapat dilihat pada tabel 4.21.

Tabel 4.21 Nilai *Precision*, *Recall*, dan *f1-Score*

<i>precision</i>	<i>recall</i>	<i>f1-Score</i>
0,8750	0,8077	0,84

Untuk hasil *precision*, *recall*, dan *f1-score* di setiap kelasnya bisa dilihat pada tabel 4.22.

Tabel 4.22 Nilai Dari Setiap Kelas

Klasifikasi	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
Positif	0,50	0,38	0,43
Negatif	0,81	0,88	0,84

Hasil dari evaluasi model dapat dilihat dari nilai precision dan recall di setiap kelasnya. Dapat dikatakan tingkat kemampuan model dalam mencari ketepatan suatu informasi yang diminta oleh pengguna adalah sebesar 50% untuk kelas positif dan 81% untuk kelas negatif. Sedangkan kemampuan suatu model dalam menemukan kembali suatu informasi adalah sebesar 38% untuk kelas positif dan 88% untuk kelas negatif. Sebagai kesimpulan, kemampuan suatu model dalam menemukan kembali suatu informasi sangat rendah.

Untuk mengevaluasi suatu model dilakukan pengujian kedua dengan menggunakan *k-fold cross validation*. Di mana jumlah *k* yang digunakan pada penelitian ini adalah 10. Dalam *10-fold cross validation*, data dibagi menjadi 10 bagian yang kira-kira berukuran sama, sehingga memiliki 10 *subset* data untuk mengevaluasi suatu model. Dengan menggunakan *cross validation* nilai akurasi menjadi **0,61**. Cross validation dilakukan agar hasil akurasi yang didapat menjadi lebih valid dibanding dengan memisahkan data 80% untuk data latih dan 20% untuk data uji.

#### 4.4 Konsep Analisis Sentimen dan Menuntut Ilmu dalam Islam

Pada dasarnya, analisis sentimen adalah studi komputasi untuk mengenali atau mengekspresikan opini, sentimen, sikap, emosi, evaluasi, penilaian atau

pandangan yang terdapat pada suatu teks. Pada zaman sekaran ini begitu banyak pendapat yang dikemukakan oleh masyarakat di sosial media. Untuk itu dibutuhkanlah analisis sentimen untuk mengetahui mana yang lebih banyak antara pendapat yang positif dan pendapat yang negatif.

Menurut Kamus Besar Bahasa Indonesia (KBBI) sentimen merupakan pendapat atau pandangan yang didasarkan pada perasaan yang berlebihan terhadap sesuatu. Jadi, sentimen bisa dikatakan sebagai pendapat atau pikiran.

Berpendapat merupakan suatu kebebasan yang dimiliki setiap manusia. Kebebasan ini merupakan hak individu yang bisa dipandang dari beberapa urusan baik umum maupun khusus. Kebebasan berpendapat ini sudah dipraktekkan oleh umat muslim sejak dulu. Contohnya ketika seorang sahabat Habab bin Mundzir yang memberikan pendapat secara pribadi dalam mengatur strategi perang Badar. Kendati tak sesuai pendapat Rasul, namun Rasul tetap mengikuti pendapat sahabat itu.

Selain hak berpendapat tadi, agama Islam juga memberikan hak untuk menuntut ilmu bahkan memberikan kewajiban bagi seluruh umatnya untuk menuntut ilmu. Tidak ada pengecualian karena sesungguhnya seluruh umat Islam wajib untuk menuntut ilmu. Hal ini telah dijelaskan dalam hadits riwayat Ibnu Majah no. 224 yang artinya menuntut ilmu itu wajib atas setiap muslim.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil pengujian menggunakan *naïve bayes classifier* yang telah dilakukan, ada beberapa hal yang dihasilkan:

1. Hasil klasifikasi pada penelitian ini menunjukkan bahwa sentimen negatif lebih banyak dibandingkan sentimen positif. Ketika data dipisah 20% untuk menguji suatu model, 21 sentimen diklasifikasikan sebagai sentimen yang negatif sedangkan 3 sentimen diklasifikasikan sebagai sentimen yang positif.
2. Pada penelitian ini, metode *naïve bayes classifier* mendapatkan hasil akurasi sebesar 75%. Hasil ini bisa dikatakan cukup baik. Untuk memastikan hasil penelitian ini dilakukan juga pengujian menggunakan *k-fold cross validation* dengan k yaitu 10 mendapatkan nilai akurasi sebesar 61% sehingga lebih akurasi lebih bagus menggunakan pemisahan 80% untuk data latih dan 20% untuk data uji.

#### **5.2 Saran**

Dari hasil yang dilakukan masih begitu banyak kekurangan, dengan demikian penelitian ini untuk dikembangkan dengan beberapa saran dari penulis, diantaranya:

1. Menggunakan metode klasifikasi yang lain sehingga dapat membandingkan hasil uji model yang dilakukan agar bisa menemukan metode klasifikasi yang terbaik.
2. Menambah data uji dan menyeimbangkan jumlah data pada setiap kelasnya.
3. Melakukan *pre-processing* dengan lebih baik lagi agar teks yang dihasilkan lebih bersih sehingga nilai akurasinya bisa lebih baik lagi.

## DAFTAR RUJUKAN

- Admin Sevima. 2020. "Kuliah Daring, Kelebihan dan Persiapan yang Harus Dilakukan Kampus", <https://sevima.com/kuliah-daring-kelebihan-dan-persiapan-yang-harus-dilakukan-kampus/>, diakses pada 17 April 2020.
- Al-Qur'an dan Terjemahannya. 1998. Al Basyir. Semarang: ASY-SYIFA
- Budiharto, W., & Suhartono, D. (2014). *Artificial Intelligence: Konsep dan Penerapannya*. Andi Publisher. <https://research.binus.ac.id/publication/95BD64D6-904F-4D97-BE58-E2BA778F5422/artificial-intelligence-konsep-dan-penerapannya/>
- Dardjowidjojo, S. (1991). *Pedoman Pendidikan Tinggi*. PT. Gramedia Widiasarana Indonesia.
- Dhull, I., & Arora, S. (2019). *Online Learning*. 3(8), 32–34.
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770.
- Fitriyani, Y., Fauzi, I., & Sari, M. Z. (2020). Motivasi Belajar Mahasiswa Pada Pembelajaran Daring Selama Pandemi Covid-19. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, 6(2), 165. <https://doi.org/10.33394/jk.v6i2.2654>
- Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education*, 19, 18–26. <https://doi.org/10.1016/j.iheduc.2013.06.002>
- Hermanto, H., Kuryanti, S. J., & Khasanah, S. N. (2019). Comparison of Naïve Bayes Algorithm, C4.5 and Random Forest for Classification in Determining Sentiment for Ojek Online Service. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 3(2), 266–274. <https://doi.org/10.33395/sinkron.v3i2.235>
- Hidayatillah, R., Mirwan, M., Hakam, M., & Nugroho, A. (2019). Levels of Political Participation Based on Naive Bayes Classifier. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), 73. <https://doi.org/10.22146/ijccs.42531>
- Iftakhar, S. (2016). GOOGLE CLASSROOM: WHAT WORKS AND HOW? *Journal of Education and Social Sciences*, 3(1), 12–18.

- Intergovernmental Panel on Climate Change (Ed.). (2014). Summary for Policymakers. In *Climate Change 2013—The Physical Science Basis* (pp. 1–30). Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.004>
- Inyang, I., Ozuomba, S., & Ezenkwu, C. (2017b). Comparative analysis of Mechanisms for Categorization and Moderation of User Generated Text Contents on a Social E-Governance Forum. *Mathematical and Software Engineering*, 3(1), 78–86.
- Korucu, A. T., & Alkan, A. (2011). Differences between m-learning (mobile learning) and e-learning, basic terminology and usage of m-learning in education. *Procedia - Social and Behavioral Sciences*, 15, 1925–1930. <https://doi.org/10.1016/j.sbspro.2011.04.029>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2), 129–135. <https://doi.org/10.1016/j.iheduc.2010.10.001>
- Pengelola Web Kemdikbud. 2020. "Mendikbud Terbitkan SE tentang Pelaksanaan Pendidikan dalam Masa Darurat Covid-19, <https://www.kemdikbud.go.id/main/blog/2020/03/mendikbud-terbitkan-se-tentang-pelaksanaan-pendidikan-dalam-masa-darurat-covid19>, diakses pada 24 Maret 2020.
- Permana, A. Y., Ismasari, & Effendi, M. M. (2018). Optimasi Stemming Porter KBBI dan Cross Validation Naïve Bayes untuk Klasifikasi Topik Soal UN Bahasa Indonesia. *Jurnal Ilmiah Komputasi*, 17(4). <https://doi.org/10.32409/jikstik.17.4.2492>
- Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data. *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–8. <https://doi.org/10.1109/KSE.2019.8919368>
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- Rodiyansyah, S. F., & Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 6(1). <https://doi.org/10.22146/ijccs.2144>



- Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021b). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 157. <https://doi.org/10.30865/mib.v5i1.2580>
- Saputra, C. B., Muzakir, A., & Udariansyah, D. (2020). ANALISIS SENTIMEN MASYARAKAT TERHADAP #2019GANTIPRESIDEN BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER. *Bina Darma Conference on Computer Science, Vol 1 No 2 (2019): Bina Darma Conference on Computer Science (BDCCS)*, 403–4013.
- Seminar Nasional Teknologi dan Informatika (Ed.). (2017). *Prosiding Seminar Nasional Teknologi dan Informatika, 2017: Kudus, 25 Juli 2017*.
- Sicat, A. S. (2015). *Enhancing College Students Proficiency in Business Writing Via Schoology*. 3(1), 159–178.
- Silmi, Adawiya. 2019. "Dua Orang Rakus yang Tidak Kenal Rasa Kenyang", <https://bincangsyariah.com/kalam/orang-tidak-kenal-rasa-kenyang/>, diakses pada 7 Oktober 2019
- So, S. (2016). Mobile instant messaging support for teaching and learning in higher education. *The Internet and Higher Education*, 31, 32–42. <https://doi.org/10.1016/j.iheduc.2016.06.001>
- Tafsir Zubdatut Tafsir Min Fathil Qadir / Syaikh Dr. Muhammad Sulaiman Al Asyqar (On-Line), tersedia pada <https://tafsirweb.com/9128-surat-asy-sy-ura-ayat-38.html>
- Yansyah, Yudi. 2020. "Mimbar Dakwah Sesi 39: Cara Muhasabah Diri", <https://jabar.kemenag.go.id/portal/read/mimbar-dakwah-sesi-39-cara-muhasabah-diri->, diakses pada 18 Agustus 2020
- Widodo, A., & Nursaptini. (2020). *PROBLEMATIKA PEMBELAJARAN DARING DALAM PERSPEKTIF MAHASISWA*. 4. <http://dx.doi.org/10.30651/else.v4i2.5340>

## LAMPIRAN

### Lampiran 1. Script Memasukkan Data

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#import the data set
dataset = pd.read_csv('ResponSentimen.csv')

label=[]
for index, row in dataset.iterrows():
    if row['Komentar anda termasuk komentar yang...'] == 'Positif':
        label.append(1)
    else:
        label.append(0)

dataset['label']=label
dataset=dataset.drop(columns=['Komentar anda termasuk komentar yang...'])
data_proses = dataset.copy()
data_proses = data_proses.drop(columns=['Timestamp', 'Nama', 'Gender', 'Universitas', 'Fakultas/Jurusan/Angkatan'])
#s_1 = data_proses[data_proses['label']==0].sample(50, replace=True)
#s_2 = data_proses[data_proses['label']==1].sample(50, replace=True)
#data_proses=pd.concat([s_1,s_2])
```

### Lampiran 2. Script Pemebersihan Teks

```
corpus = []
for i in range(0,157):
    review = re.sub('[^a-zA-Z]', ' ', data_proses['Komentar anda tentang kuliah online'][i])
    review = review.lower()
    review = review.split()
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    all_stopwords = stopwords.words('indonesian')
    all_stopwords.remove('tidak')
    review = [stemmer.stem(word) for word in review if not word in set(all_stopwords)]
    review = ' '.join(review)
    corpus.append(review)
```

### Lampiran 3. Script Visualisasi Kata dengan Wordcloud

```
corpusss = pd.Series(corpus).str.cat(sep=' ')
from wordcloud import WordCloud

wordcloud = WordCloud(width=1600, height=800, max_font_size=200).generate(corpusss)
plt.figure(figsize=(12,10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

### Lampiran 4. Script Membuat Model Bag of Words

```
#Creating The Bag of Words Model
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
cv = CountVectorizer(max_features = 500)
bow_article = cv.fit(corpus)
count_tokens = bow_article.get_feature_names()
article_vect = bow_article.transform(corpus)
df_count_vect = pd.DataFrame(data=article_vect.toarray(), columns=count_tokens)
X = cv.fit_transform(corpus).toarray()
y = dataset.iloc[:, -1].values
```

## Lampiran 5. Script Uji Model

```
# Splitting the dataset into the Training set and Test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)

# Training the Naive Bayes model on the Training set
from sklearn.naive_bayes import MultinomialNB
classifier = MultinomialNB()
classifier.fit(X_train, y_train)
```

## Lampiran 6. Script Prediksi Klasifikasi

```
# Predicting the Test set results
y_pred = classifier.predict(X_test)
hasil=np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1)
print(hasil)
```

## Lampiran 7. Script Membuat *Confusion Matrix*

```
# Making the Confusion Matrix
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

## Lampiran 8. Script Melihat Nilai Akurasi dari Suatu Model

```
cm_matrix = pd.DataFrame(data=cm, columns=['Actual Negative:0', 'Actual Positive:1'],
                        index=['Predict Negative:0', 'Predict Positive:1'])

sns.heatmap(cm_matrix, annot=True, fmt='d', cmap='YlGnBu')

TP = cm[0,0]
TN = cm[1,1]
FP = cm[0,1]
FN = cm[1,0]

precision = TP / float(TP + FP)

print('Precision : {0:0.4f}'.format(precision))

recall = TP / float(TP + FN)

print('Recall or Sensitivity : {0:0.4f}'.format(recall))

#See The Accuracy
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))

from sklearn.model_selection import cross_val_score
scores = cross_val_score(classifier, X, y, cv=10)
print(scores)
print(scores.mean())

TP = cm[0,0]
TN = cm[1,1]
FP = cm[0,1]
FN = cm[1,0]

precision = TP / float(TP + FP)

print('Precision : {0:0.4f}'.format(precision))

recall = TP / float(TP + FN)

print('Recall or Sensitivity : {0:0.4f}'.format(recall))

fscore = (2*recall*precision/(recall+precision))
print('f1-score:', fscore)
```



KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI  
MAUALANA MALIK IBRAHIM MALANG  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Gajayana No. 50 Dinoyo Malang Telp./Fax.(0341)558933

### BUKTI KONSULTASI SKRIPSI

Nama : Bagus Aziz Rahmatullah  
NIM : 17610014  
Fakultas/  
Program Studi : Sains dan Teknologi / Matematika  
Judul Skripsi : Analisis Sentimen Mahasiswa Terhadap Perkuliahan  
Dalam Jaringan Menggunakan Metode *Naïve Bayes Classifier*  
Pembimbing I : Dr. Imam Sujarwo, M.Pd  
Pembimbing II : Erna Herawati, M.Pd

No	Tanggal	Hal	Tanda Tangan
1	15 Maret 2021	Konsultasi Kajian Keagamaan	1.
2	18 Maret 2021	Konsultasi Bab I, Bab II, dan Bab III	2.
3	25 Maret 2021	Konsultasi Bab I, Bab II, dan Bab III	3.
4	26 Maret 2021	Konsultasi Kajian Keagamaan	4.
5	16 April 2021	Konsultasi Bab I, Bab II, Bab III, dan Bab IV	5.
6	21 April 2021	Konsultasi Bab IV	6.
7	6 Mei 2021	Konsultasi Kajian Keagamaan	7.
8	6 September 2021	Konsultasi Bab I, Bab II, Bab III, Bab IV, dan Bab V	8.
9	20 Desember 2021	Konsultasi dan ACC Keseluruhan	9.

Malang, 20 Desember 2021

Mengetahui,  
Ketua Program Studi  
Matematika



Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005

## RIWAYAT HIDUP



Bagus Aziz Rahmatullah lahir di Kota Manado pada 14 November 1999. Memiliki nama panggilan Bagus. Alamatnya berada di Jalan Politeknik Perum Buhas Asri Banua 3, Blok. B, No.12, Kelurahan Buha, Kecamatan Mapanget, Kota Manado, Sulawesi Utara. Merupakan anak pertama dari Bapak Totok Dwi Riyanto dan Ibu Sumarni.

Pendidikan yang pernah ditempuh adalah TK At-Taqwa. Kemudian melanjutkan sekolahnya di SD Negeri 30 Manado dan lulus pada tahun 2012. Menempuh Pendidikan SMP di Sekolah MTs Negeri Unggulan Manado dan lulus pada tahun 2014. Melanjutkan Pendidikan SMA di Madrasah Aliyah Negeri 1 Plus Keterampilan Manado dan lulus pada tahun 2017.

Tahun 2017 melanjutkan Pendidikan ke jenjang starta 1 di Universitas Islam Negeri Maulana Malik Ibrahim Malang mengambil program studi Matematika Fakultas Sains dan Teknologi. Aktif mengikuti kegiatan pengabdian dan organisasi serta komunitas yang ada di dalam dan di luar kampus, seperti menjadi Musyrif di Ma'had Sunan Ampel Al-Aly (2018-2022), Pengurus Mathematics English Club (2018-2020), Pengurus UPKM El-Ma'rifah (2018-2021).

Prestasi-prestasi yang pernah diikuti adalah juara 4 pidato Bahasa Inggris pada tingkat SMA dan juara 3 duta Bahasa Inggris yang diselenggarakan oleh Pusat Bahasa Universitas Islam Negeri Maulana Malik Ibrahim Malang pada tahun 2019.