

Sentimen Analisis Terhadap Pembelajaran Jarak Jauh Menggunakan Metode *Naïve Bayes Classifier* dan *Lexicon Based*

Cahyo Prianto^{a1}, Woro Isti Rahayu^{a2}, Nurul Izza Hamka^{b3}

^aTeknik Informatika, Politeknik Pos Indonesia
Bandung, Indonesia

¹cahyoprianto@poltekpos.ac.id (Corresponding author)

²woroisti@poltekpos.ac.id

^bTeknik Informatika, Politeknik Pos Indonesia
Bandung, Indonesia

³nurulizzahamka@gmail.com

Abstract

Since the beginning of the COVID-19 pandemic, all fields have been affected, especially in the field of education, where the learning process is currently carried out remotely. This research was conducted to find out the responses from the community within the scope of education such as students, students and teachers. The number of respondents who gave their responses as many as 265 through filling out a questionnaire in the form of a google form. Based on this research, it is known that of the 265 people who responded with a total of 6 statements, then obtained 1,590 different answers. From 1,590 data were reprocessed so that the final data was 1,468. As for the results of labeling with this Lexicon based dictionary, 162 were positive, 516 were negative, and 790 were neutral. The test results with nave Bayes obtained an accuracy rate of 53.8% by using the measurement of the effectiveness of the confusion matrix.

Keywords: Covid-19, Sentiment Analysis, Distance Learning, Naïve Bayes Classifier, Lexicon Based

Abstrak

Sejak awal pandemic covid-19 ini segala bidang terkena dampak khususnya dibidang pendidikan, dimana saat ini proses pembelajaran dilakukan secara jarak jauh. Penelitian ini dilakukan untuk mengetahui tanggapan dari masyarakat yang berada dalam ruang lingkup pendidikan seperti siswa, mahasiswa serta pengajar. Adapun jumlah responden yang memberikannya yaitu sebanyak 265 melalui pengisian kuesioner dalam bentuk *google form*. Berdasarkan penelitian ini diketahui bahwa dari 265 orang yang memberikan respon dengan jumlah pernyataan sebanyak 6, maka diperoleh 1.590 jawaban yang berbeda. Dari 1.590 data diolah kembali sehingga data akhir sebanyak 1.468. Adapun hasil dari pelabelan dengan kamus lexicon based ini berjumlah 162 positif, 516 bernilai negatif, dan 790 netral. Hasil pengujian dengan naïve bayes diperoleh tingkat akurasi 53.8% dengan menggunakan pengukuran efektifitas *confusion matrix*.

1. Pendahuluan

Kondisi saat ini seluruh dunia salah satunya Indonesia dikejutkan dengan munculnya virus covid-19 sejak maret 2021. Covid-19 ini adalah virus yang penularannya sangat cepat. Dampak yang diberikan bagi manusia adalah segala aktivitas menjadi terhambat dan harus terbatas [1]. Menurut Ismawati pandemic ini juga memberikan dampak dan mempengaruhi aktivitas kehidupan khususnya dalam dunia bidang pendidikan. Pendidikan sendiri adalah salah satu sarana terdepan dalam memberikan kesetaraan sosial dan juga salah satu pengalaman optimis serta memenuhi kehidupan individu [2].

Pandemic Covid-19 sampai saat ini juga belum berhenti dan masih berlangsung, hal ini membawa dampak perubahan pada metode dalam bidang Pendidikan yaitu proses pembelajaran jarak jauh (PJJ) [3]. Sekolah dan perguruan tinggi sampai dengan perusahaan ataupun kantor dituntut untuk melakukan belajar dan bekerja dari rumah atau WFH (*Work from home*). Menurut Latip dalam penelitiannya saat ini kebijakan yang berlaku adalah dari rumah atau yang dikenal dengan PJJ, dan pada pelaksanaannya PJJ ini memberikan tantangan baru bagi pelajar dan juga pengajar bahkan masyarakat secara umum seperti orang tua [4]. Orang tua sangat berperan luas dan penting untuk mendampingi kesuksesan anaknya selama proses belajar dilakukan dari rumah [5].

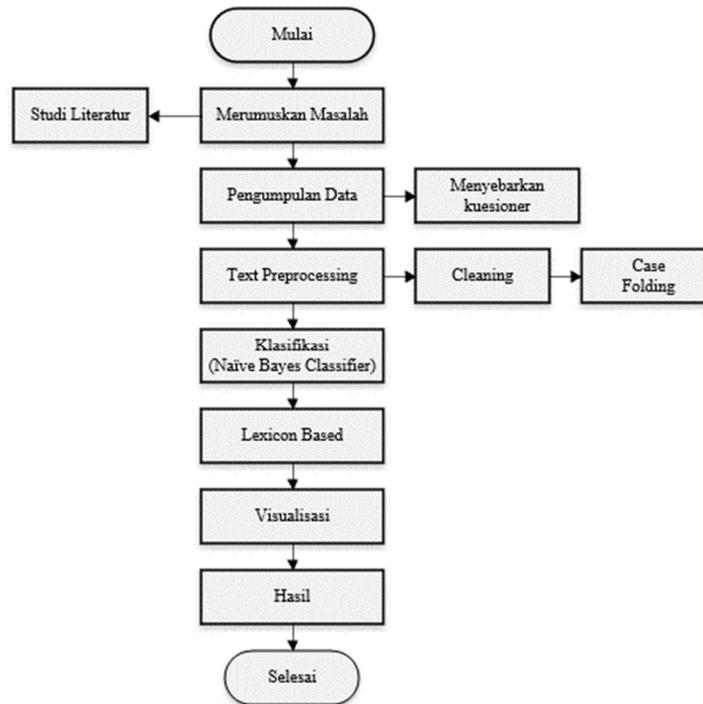
Metode pembelajaran ini ada beberapa pihak yang menganggap bahwa pembelajaran saat ini lebih praktis, cepat dan juga aman dan tepat. Namun disisi lain yang juga banyak yang kontra dan merasa terjadinya kesenjangan dalam proses pembelajaran secara *online* ini, karena multimedia ini sepenuhnya belum mampu beradaptasi dalam bidang Pendidikan [6].

Analisis sentimen ini dilakukan untuk melihat opini atau pendapat terhadap objek apakah cenderung positif, negatif, ataupun netral. Pada penelitian ini akan menggunakan data dari opini melalui pengisian kuesioner. Kemudian dilakukan analisis sentimen menggunakan metode *Naïve Bayes Classifier* (NBC). Menurut Suryani, 2019 metode NBC ini sederhana namun lebih efektif digunakan dalam hal melakukan proses klasifikasi, serta tingkat akurasi yang tinggi dalam klasifikasi teks. Sedangkan metode *Lexicon Based* metode ini digunakan untuk mengelompokkan pendapat ataupun opini berdasarkan sentimennya, yaitu positif dan negatif atau netral [7].

2. Metode Penelitian

2.1 Diagram Alur Metode Penelitian

Pada diagram alur metodologi penelitian ini menjelaskan setiap alur atau proses dalam penelitian yang dilakukan. Dalam penelitian ini terdapat ruang lingkup agar pembahasan ini terstruktur dengan baik. Adapun tahapan kerangka penelitian yang akan dilakukan dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Diagram Alur Metode Penelitian

2.2 Tahapan-Tahapan Diagram Alur Metode Penelitian

Proses dalam penelitian ini menggunakan salah satu metode yaitu *naïve bayes classifier*. Analisis sentiment ini untuk menentukan kategori tanggapan positif, negative, dan netral. Adapapun alur penelitian dalam sentiment analisis ini merupakan tahapan atau gambaran secara umum yang menunjukkan tahapan kerja yang dilakukan. Pertama dilakukan studi literatur, ini dilakukan untuk mendapatkan referensi mengenai landasan teori yang berkaitan dengan masalah yang di angkat yang mana bersumber dari buku literatur. Berdasarkan tinjauan Pustaka dari penelitian sebelumnya, penulis menemukan solusi pemecahan tentang analisis sentimen menggunakan metode *Naïve Bayes* dan *Lexicon Based*.

Dalam proses penelitian ini pengumpulan data yang akan digunakan diambil dari tanggapan kuesioner yang disebar dalam rentang waktu bulan April - Mei 2021. Penyebaran kuesioner ini penulis lakukan untuk mendapatkan data berupa opini atau pendapat dari pelajar atau pengajar untuk dilakukan pengelolaan data.

Selanjutnya data yang diperoleh dilakukan proses cleaning yang bertujuan untuk menghilangkan komponen-komponen yang tidak berhubungan dengan informasi dan juga memperbaiki kata-kata yang tidak sesuai dengan kaidah, selain itu dilakukan juga proses cleaning yaitu case folding yang bertujuan untuk mengubah karakter huruf kapital menjadi huruf kecil. Ketika proses cleaning dan case folding selesai, selanjutnya menentukan skor sentiment dengan kamus *lexicon* dan berikutnya adalah klasifikasi *naïve bayes* menggunakan pengujian *confusion matrix*.

3. Hasil Dan Pembahasan

3.1 Pengumpulan Data

Pengambilan data ini dilakukan dengan cara pengumpulan data tanggapan atau opini tentang proses pembelajaran jarak jauh yang terjadi selama masa pandemic ini. Data yang terkumpul dari enam pernyataan dalam kuesioner sebanyak 1.468 dari 265 orang yang memberikan tanggapan. Data tersebut diperoleh dalam rentang waktu bulan April - Mei 2021.

Tabel 1. Data Awal dari Kuesioner

| Nama | Status | Gender | Respon |
|-----------------------|-----------|-----------|--|
| Andi Husnul Khatimah | Mahasiswa | Perempuan | Pembelajaran secara online saya kurang bisa memahami materi yang diberikan dosen |
| Nur Ummul Annisa | Siswa | Perempuan | Terlalu banyak tugas selama online |
| Chandra Kirana Poetra | Mahasiswa | Laki-Laki | Lebih efektif tatap muka karena lebih fokus dan jelas |
| Ilham Muhammad Ariq | Mahasiswa | Laki-Laki | Sulit menyerap ilmu yang disampaikan dosen |
| Wirda Nengsi,S.Pd | Pengajar | Perempuan | kurangnya minat dan partisipasi peserta didik dalam pembelajaran online, karena keterbatasan media dan jaringan internet |
| Nia zuliana | Mahasiswa | Perempuan | Saya rindu pembelajaran offline, lebih menyenangkan dan efektif. |

Setelah waktu memberikan tanggapan berakhir, selanjutnya tanggapan yang diperoleh disimpan kedalam excel dengan format .xlsx. Setelah data disimpan kedalam excel, maka selanjutnya diinputkan kedalam R Studio.

3.2 Pre-Processing

Cleaning data dilakukan untuk menghapus karakter seperti tanda baca, angka, dan juga URL. Sebelum melakukan proses *cleaning* data maka terlebih dahulu dilakukan *corpus data*. Setelah

proses corpus selanjutnya proses *case folding* dan *cleaning data*. Selanjutnya data disimpan kedalam dataset baru dengan ekstensi csv. Berikut *syntax* pada proses *cleaning data*:

```

2 docs <- Corpus(VectorSource(docs))
3 inspect(docs)
4
5 tospace <- content_transformer(function(x , pattern ) gsub(pattern, " ", x))
6 docs <- tm_map(docs, tospace, "/"")
7 docs <- tm_map(docs, tospace, "@")
8 docs <- tm_map(docs, tospace, "\\|")
9
10 docs <- tm_map(docs, content_transformer(tolower))
11
12 docs <- tm_map(docs, tospace, "[[:digit:]]")
13 docs <- tm_map(docs, tospace, "[[:punct:]]")

```

Gambar 2. Syntax Tahap Preprocessing

Pada baris ke 2-3 merupakan proses corpus data yang berguna untuk merubah data frame menjadi data corpus. Kemudian pada baris ke 5-8 berguna untuk mengunah tanda / @ dan \\ menjadi space. Pada baris ke 10 adalah proses *case folding* yang berguna untuk mengubah semua kata menjadi huruf kecil. Terakhir pada baris ke 12 berguna untuk menghapus nomor dan baris ke 13 untuk menghapus tanda baca.

Tabel 2. Hasil Data Processing

| No | Respon |
|----|--|
| 1. | pembelajaran online saya kurang memahami materi diberikan dosen |
| 2. | banyak tugas selama online |
| 3. | lebif efektif tatap muka fokus jelas |
| 4. | sulit menyerap ilmu disampaikan dosen |
| 5. | kurangnya minat partisipasi peserta didik pembelajaran karena keterbatasan media jaringan internet |
| 6. | saya rindu pembelajaran offline menyenangkan efektif |

3.3 Lexicon Based

Setelah tahap preprocessing, maka tahap selanjutnya adalah proses pelabelan data untuk mengetahui yang mana saja tanggapan positif, netral, dan negatif. *Lexicon based* ini dapat digunakan pada proses pelabelan data atau tanggapan dengan cara menghitung setiap skor sentiment untuk mempermudah dalam proses klasifikasi.

Proses menghitung setiap kata berdasarkan kamus positif serta negatif dalam suatu kalimat tersebut, yaitu dengan menjumlahkan nilai opini. Jumlah nilai opini buat sentiment positif yaitu nilai 1 ataupun lebih, netral yaitu bernilai = 0, demikian juga untuk sentiment negatif bernilai -1 ataupun lebih [8]. Proses pengelolaan terdiri dari dua tahap, yaitu: pertama, mengimport kamus kata *lexicon* positif dan juga negatif. Dalam penelitian ini kamus *lexicon* ini berupa *file* berekstensi .text. yang diambil melalui github.com. proses selanjutnya yaitu penentuan skor sentiment.

| | | |
|----|----|---|
| 11 | 1 | jujur saya setuju pjj |
| 12 | 0 | saya setuju pembelajaran online efektif serta kebanyakan gu... |
| 13 | 0 | saya setuju pembelajaran secara online materi ajarkan dose... |
| 14 | 4 | baiknya tetap secara offline fokus menerima pembelajaran ... |
| 15 | -3 | saya setuju pembelajaran online tatap muka saya efektif jara... |

Gambar 3. Hasil Proses Pelabelan

3.4 Klasifikasi Naïve Bayes

Berdasarkan penelitian, dari data yang diperoleh sentiment yang masuk kedalam *class* positif sebanyak 162 tanggapan, untuk *class negative* sebanyak 516 tanggapan, dan *class netral* sebanyak 790 tanggapan. Selanjutnya diidentifikasi sebagai kedalam *confusion matrix*. *Confusion matrix* sendiri merupakan tabel yang diprediksi jumlah data uji yang diperoleh dengan klasifikasi benar dan data uji dengan klasifikasi yang salah [9].

3.4.1 Pengujian Menggunakan Excel

a. Perhitungan Probabilitas Kelas Sentimen

Tabel 3. Probabilitas Kelas Sentiment

| PROBABILITAS KELAS | |
|--------------------|--------|
| Sentiment | Jumlah |
| Positif | 162 |
| Negatif | 516 |
| Netral | 790 |
| Jumlah | |

Berdasarkan pada gambar diatas adalah hasil perhitungan dari probabilitas pada kelas sentiment. Data yang diambil sebanyak 1.468. Kemudian dilakukan perhitungan seberapa banyak komentar positif, negatif, dan netral.

b. Perhitungan Probabilitas Gender

Tabel 4. Perhitungan Probabilitas Gender

| Gender | Positif | Prob | Negatif | Prob | Netral | Prob |
|-----------|---------|-------------|---------|-------------|--------|-------------|
| Laki-Laki | 43 | 0,265432099 | 144 | 0,279069767 | 223 | 0,282278481 |
| Perempuan | 119 | 0,734567901 | 372 | 0,720930233 | 567 | 0,717721519 |
| Jumlah | 162 | | 516 | | 790 | |

Berdasarkan pada gambar diatas adalah hasil perhitungan dari probabilitas pada kelas gender. Data yang diambil sebanyak 1.468. Kemudian dilakukan perhitungan untuk positif yaitu jumlah sentiment positif pada gender laki-laki di bagi dengan jumlah sentiment positif secara keseluruhan. Selanjutnya untuk probabilitas negatif dan netral juga akan dijumlahkan banyaknya masing-masing komentar pada laki-laki dan perempuan lalu dibagi dengan jumlah keseluruhan komentar.

c. Perhitungan Probabilitas Status

Tabel 5. Perhitungan Probabilitas Status

| Status | Positif | Prob | Negatif | Prob | Netral | Prob |
|-----------|---------|-------------|---------|-------------|--------|-------------|
| Siswa | 12 | 0,074074074 | 48 | 0,093023256 | 72 | 0,091139241 |
| Mahasiswa | 114 | 0,703703704 | 249 | 0,48255814 | 382 | 0,483544304 |
| Pengajar | 36 | 0,222222222 | 219 | 0,424418605 | 336 | 0,425316456 |
| Jumlah | 162 | | 516 | | 790 | |

Berdasarkan pada gambar diatas adalah hasil perhitungan dari probabilitas status. pertama dilakukan perhitungan untuk jumlah dari masing-masing banyak komentar pada siswa,

mahasiswa dan juga pengajar. Selanjutnya jumlah komentar pada siswa, mahasiswa dan engajar dibagi dengan banyaknya komentar pada tiap sentiment untuk mendapatkan nilai probabilitas.

3.4.2 Pengujian Menggunakan Bahasa Pemrograman R

Pada proses ini adalah untuk mencocokkan hasil probabilitas yang dilakukan secara manual di excel dan juga pada R. Berikut adalah tahapan pada pengujian pada R:

a. Read File Excel

```
3 library(readxl)
4 dataku <- read_excel("D:/bismillah/UntukAkurasi.xlsx",
5                       sheet = "sheet1")
```

Gambar 4. Syntax Read File Excel

Syntax diatas adalah tahap awal yaitu untuk membaca *file* excel dengan nama file UntukAkurasi.xlsx

b. View Data

```
6 view(dataku)
7 str(dataku)
```

Gambar 5. Syntax View Data

Pada baris ke 6 adalah untuk melihat data kita yang nama variabelnya “dataku”, sedangkan pada baris ke 7 adalah untuk melihat struktur data.

c. Mengubah Data Menjadi Kategorik

```
10 dataku$`gender`=as.factor(dataku$`gender`)
11 dataku$`status`=as.factor(dataku$`status`)
12 dataku$`klasifikasi`=as.factor(dataku$`klasifikasi`)
13 colnames(dataku)=c("text", "gender", "status", "klasifikasi")
14
```

Gambar 6. Syntax Mengubah Data Menjadi Kategorik

Pada baris ke 10 sampai 12 adalah untuk mengubah setiap kelas atau atribut menjadi kategorik.

d. Klasifikasi *Naïve Bayes*

```
15 library(e1071)
16 library(caret)
17
```

Gambar 7. Library Klasifikasi Naive bayes

Pada baris ke 15 dan 16 terdapat *library* “e1072” untuk metode *naïve bayes* itu sendiri, sedangkan *library* “caret” adalah untuk mengetahui seberapa baik pemodelan metode yang digunakan dan untuk proses confusion Matrix.

e. Confusion Matrix

```
19 model1<- naiveBayes(klasifikasi~gender+Status, data=dataku)
20 model1
21 pred1<-predict(model1,dataku)
22 pred1
23 confusionMatrix(pred1, dataku$klasifikasi)
```

Gambar 8. Syntax Confusion Matrix

Pada baris ke 18 dan 19 adalah pemodelan *naïve bayes*, dan diambil variable independent gender + kulit dengan data fram “dataku”. Selanjutnya pada baris 20 sampai 22 adalah proses melakukan prediksi menggunakan *confusion matrix*.

f. Hasil Pemodelan *Naïve Bayes*

```

Conditional probabilities:
gender
Y      Laki-Laki Perempuan
Negatif 0.2790698 0.7209302
Netral  0.2822785 0.7177215
Positif 0.2654321 0.7345679

Status
Y      Mahasiswa Pengajar Siswa
Negatif 0.48255814 0.42441860 0.09302326
Netral  0.48354430 0.42531646 0.09113924
Positif 0.70370370 0.22222222 0.07407407
    
```

Gambar 9. Hasil Pemodelan *Naïve Bayes*

Dapat dilihat dari gambar 4.9 yang merupakan hasil probabilitas menggunakan R, jika dilihat pada perhitungan excel sebelumnya nilai probabilitas yang diperoleh sama persis

g. Prediksi *Confusion Matrix*

```

Confusion Matrix and Statistics

Reference
Prediction Negatif Netral Positif
Negatif      0      0      0
Netral      516    790    162
Positif      0      0      0

Overall statistics

Accuracy : 0.5381
95% CI : (0.5122, 0.5639)
No Information Rate : 0.5381
P-Value [Acc > NIR] : 0.5107
    
```

Gambar 10. Hasil Prediksi *Confusion Matrix*

Berdasarkan gambar diatas dari pengujian confusion matrix diperoleh akurasi sebesar 0.5381, jika di presentasekan maka sebesar 53.8%.

h. Prediksi *Multiclass Confusion Matrix*

Penelitian Ini menggunakan *multiclass confusion matrix* 3x3, karena output ada 3 yaitu positif, negatif, dan netral. Berikut tabel dari jumlah data klasifikasi:

Tabel 6. Prediksi *Multiclass Confusion Matrix*

| | | Prediksi | | |
|--------|---------|----------|---------|--------|
| | | Positif | Negatif | Netral |
| Aktual | Positif | 0 | 0 | 0 |
| | Negatif | 0 | 0 | 0 |
| | Netral | 162 | 516 | 790 |

$$\text{Akurasi} = \frac{0 + 790 + 0}{0 + 0 + 0 + 0 + 0 + 516 + 162 + 0 + 790}$$

References

- [1] I. P. Y. Purandina and I. M. A. Winaya, "Pendidikan Karakter di Lingkungan Keluarga Selama Pembelajaran Jarak Jauh pada Masa Pandemi COVID-1," *Jurnal Ilmu Pendidikan*, vol. 3 (2), p. 271, 2020.
- [2] D. Ismawati and I. Prasetyo, "Efektivitas Pembelajaran Menggunakan Video Zoom Cloud Meeting pada Anak Usia Dini Era Pandemi Covid-19," *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, vol. 5, no. 1, p. 666, 2021.
- [3] R. M. Napitupulu, "Dampak pandemi Covid-19 terhadap kepuasan pembelajaran jarak jauh," *Jurnal Inovasi Teknologi Pendidikan*, vol. 7(1), p. 24, 2020.
- [4] A. Latip, "Peran Literasi Teknologi Informasi Dan Komunikasi Pada Pembelajaran Jarak Jauh Di Masa Pandemi Covid-19," *EduTeach : Jurnal Edukasi dan Teknologi Pembelajaran*, vol. 1, no. 2, p. 107, 2020.
- [5] E. Kurniati, D. K. N. Alfaeni and F. Andriani, "Analisis Peran Orang Tua dalam Mendampingi Anak di Masa Pandemi Covid-19," *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, vol. 5, no. 1, p. 242, 2021.
- [6] F. Angelica, K. Tan, A. Lauw, W. Rosalya, S. and W. Fitri, "Dampak Penyebaran Covid-19 Terhadap Dunia Pendidikan Dan Sanitasi Di Indonesia," *Syntax Idea*, vol. 3, no. 1, p. 101, 2021.
- [7] N. P. S. M. Suryani, L. and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 145, 2019.
- [8] Prasetyo, H. Buntoro and M. , "Analisis Sentimen Pada Channel Autonetmagz Terhadap Review Mobil Almaz 2019 dengan Metode Naive Bayes Classifier dan Lexicon Based," *Komputek*, p. 63, 2020.
- [9] M. Iqbal, "Penerapan Fast Correlation Based Filter dan Naive Bayes Untuk Klasifikasi Tingkat Kemiskinan," *Pekanbaru*, 2021.
- [10] G. F. Ramdani and I. Aknuranda, "Pengembangan WebGIS Untuk Analisis Dan Pemodelan Data Menggunakan Teknik Regresi Spasial Dan R-Shiny Web Framework (Studi Kasus: Data Kemiskinan dan Zakat Jawa Timur)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 3, p. 1292, 2018.