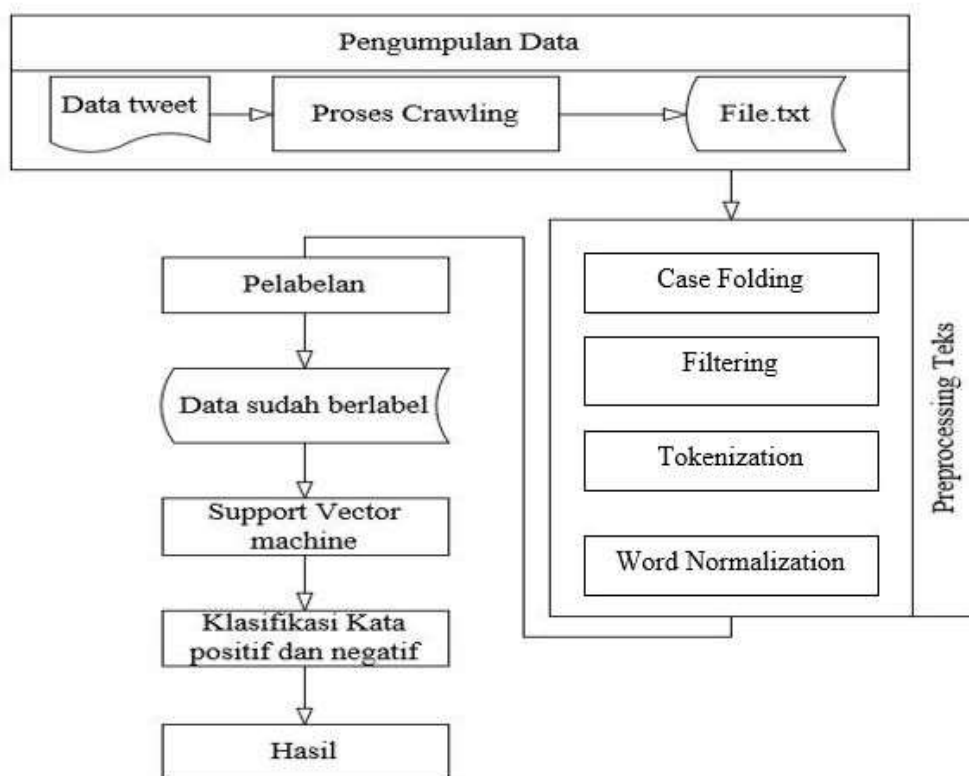


## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Analisis Sistem

Dalam implementasi algoritma *support vector machine* pada kasus sentiment analisis dengan topik *Covid-19*, dataset yang digunakan diambil dari hasil *crawling* dengan topik *Covid-19*. Pada data latih aspek sudah ditentukan bersama sentimennya pada setiap kalimat. Sebelum masuk pada pelatihan *support vector machine* untuk klasifikasi sentimen, perlu dilakukan proses *preprocessing*. Tahap *preprocessing* ini meliputi tahap *case folding*, *filtering*, *tokenization*, dan *word normalization*. Selain kalimat pada data latih dilakukan *preprocessing* sebelum masuk tahap pengujian. Adapun gambaran umum dari sistem yang akan dibangun dapat dilihat pada Gambar 4.1:



Gambar 4. 1 Gambaran Umum Sistem yang Akan Dibangun

Berikut merupakan penjelasan dari Gambar 4.1:

1. Pada data tweet harus dilakukan proses crawling terlebih dahulu dengan topik Virus covid-19 untuk mendapatkan data tweet
2. Data hasil crawling dilakukan preprocessing teks dengan tahapan *case folding, filtering, tokenization dan word normalization*.
3. Setelah dilakukan preprocessing teks data diberi label untuk dapat memudahkan proses klasifikasi teks
4. Tahapan selanjutnya ialah klasifikasi teks menggunakan algoritma *Support Vector Machine*

#### 4.1.1 Proses Crawling Data

Proses Data crawling pada penelitian ini merupakan Data Primer bersifat eksplorasi yang menggunakan bantuan dari *twitter API* yaitu #TAGS yang berfungsi sebagai penghubung ke database *twitter*, program API #TAGS yang digunakan dalam penelitian ini merupakan *twitter archived google spreadsheet* yang menggunakan *twitter API* resmi tetapi merupakan aplikasi pihak ketiga dengan kemudahan yang juga bersifat public dengan versi yang paling terbaru *twitter API #TAGS v6.1.9.1*. Proses tahapan pengumpulan data yang menggunakan *twitter API* ini merupakan proses penambahan data atau yang umum disebut dengan (crawling data). Proses crawling data yang menggunakan *twitter API* ini dilakukan dengan mencari kata kunci yang berhubungan dengan virus covid-19. Dikarenakan program yang dipakai merupakan bersifat *free* tentu didapat limitasi dalam menarik data *tweet* yang ada, proses pengambilan data hanya dibatasi sampai 7 hari saja paling maksimal, di karenakan *twitter API* ini sangat dibatasi kecuali ingin membayar untuk meningkatkan fitur yang didapat, pengambilan *tweet*-nya dibatasi tidak lebih dari 10.000 atau umumnya para pengguna API ini hanya mengambil sebagian dari tweet yang disediakan. Berikut ini tampilan dari *script python* yang digunakan untuk melakukan crawling data pada *twitter*. Berikut ini script dari proses crawling pada gambar 4.2

```

180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Gambar 4. 2 Script Python Crawling

#### 4.1.2 Analisis Data Masukan

Data masukan yang akan digunakan pada penelitian ini diambil dari hasil *crawling* terhadap data *tweet* dengan topik covid-19. Data hasil *crawling* terdapat 74544 kemudian dilakukan *filter* yang hanya mendapatkan data yang berkualitas 36924 kemudian dilakukan pembagian data latih dan uji, untuk data latih 80% dari jumlah data sehingga didapatkan 29539 dan data uji 20% dari jumlah data sehingga didapatkan 7385 kalimat untuk data uji.

#### 4.1.3 Preprocessing Teks

*Preprocessing teks* dilakukan untuk mengurangi *noise* dan menyeragamkan data supaya memudahkan proses pelatihan dan proses pengujian. Adapun tahapan *preprocessing* yang akan dilakukan yaitu, *case folding*, *filtering*, *tokenization*, *stemming*. Gambaran umum tahap *preprocessing* dapat dilihat pada Gambar 4.2:



Gambar 4. 3 Tahapan *Preprocessing*

#### 4.1.4 Case Folding

*Case folding* adalah proses konversi keseluruhan teks ke dalam bentuk standar (biasanya huruf kecil atau *lowercase*). Pada penelitian ini semua huruf pada data masukan diubah menjadi huruf kecil agar mempermudah pengenalan kata pada proses algoritma utama. Contoh kata setelah dilakukan *case folding* “BELAJAR”, “belajAr” atau “BeLaJaR” diubah ke huruf kecil menjadi “belajar”. Proses *case folding* dapat dilihat pada Gambar 4.3 berikut ini:

Gambar 4. 4 Proses *Case Folding*

Berikut ini hasil dari *case folding* pada Tabel 4.1

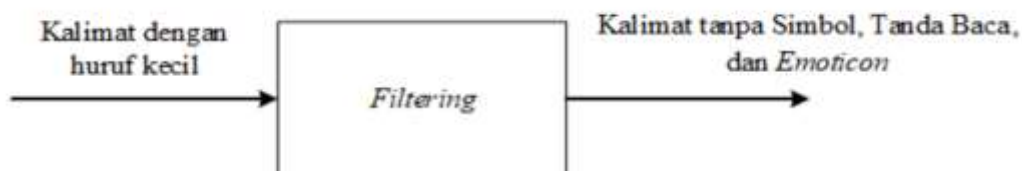
Tabel 4. 1 Case Folding

No	Data Tweet	Case Folding
1	<a href="https://bit.ly/2L6CcbB">https://bit.ly/2L6CcbB</a> Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="https://pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>	<a href="https://bit.ly/2l6ccbb">https://bit.ly/2l6ccbb</a> seharusnya saat ini pemerintah lampung bisa memberikan bantuan kepada mahasiswa tersebut. mengingat begitu derasnya donasi yang masuk ke pemprov lampung melalui tim gugus tugas covid-19 <a href="https://pic.twitter.com/andu9vxtjg">pic.twitter.com/andu9vxtjg</a>

#### 4.1.5 Filtering

*Filtering* adalah proses penghilangan simbol-simbol yang tidak diperlukan dalam kalimat seperti tanda baca, angka dan *emoticon*. Proses ini

dilakukan supaya menghilangkan noise yang terdapat pada kalimat agar tidak mengganggu proses algoritma utama. Proses *filtering* dapat dilihat pada Gambar 4.4. Setelah melalui proses *filtering* maka kalimat sudah tidak memiliki simbol-simbol seperti tanda baca, angka, dan *emoticon* yang tidak diperlukan.



Gambar 4. 5 Proses *Filtering*

Berikut ini hasil dari *Filtering* pada Tabel 4.2

Tabel 4. 2 *Filtering*

No	Data Tweet	Filtering
1	<a href="https://bit.ly/2L6CcbB">https://bit.ly/2L6CcbB</a> Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="https://pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>	seharusnya saat ini pemerintah lampung bisa memberikan bantuan kepada mahasiswa tersebut. mengingat begitu derasnya donasi yang masuk ke pemprov lampung melalui tim gugus tugas covid-19

#### 4.1.6 Tokenization

*Tokenization* adalah proses pemisahan kata-kata dalam kalimat supaya menjadi token kalimat. Pemisahan tersebut dilakukan dengan mengandalkan karakter spasi pada kalimat. Setelah melalui proses *tokenization* kata-kata dalam kalimat dipecah menjadi sebuah token. Berikut kata yang akan di proses kedalam *tokenization* “yang terdapat pada tabel 4.3 berikut ini:”.

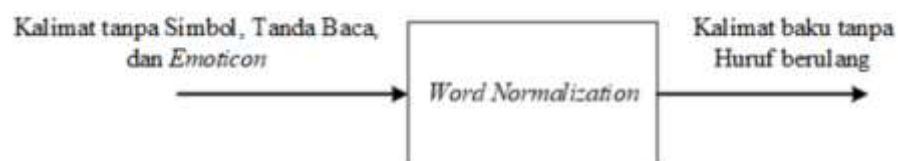
Berikut ini hasil dari *Tokenization* pada Tabel 4.3

Tabel 4. 3 Tokenization

No	Data Tweet	Tokenization
1	<a href="https://bit.ly/2L6CcbB">https://bit.ly/2L6CcbB</a> Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>	Seharusnya Saat Ini Pemerintah Lampung Bisa Memberikan Bantuan kepada mahasiswa tersebut. Mengingat Begitu Derasnya Donasi Yang Masuk Ke Pemprov Lampung Melalui Tim Gugus Tugas covid-19

#### 4.1.6 Word Normalization

*Word Normalization* adalah proses mengubah kata-kata tidak baku menjadi kata baku. Pada proses ini kata-kata yang tidak baku seperti kata “enakkkk” akan diubah menjadi “enak” dan kata “ga”, “engga”, “gak”, “enggak” dan “gk” menjadi “tidak” agar pada proses pembentukan kamus kata, kata tersebut tidak dianggap berbeda. Gambaran proses *word normalization* ditunjukkan pada Gambar 4.5:

Gambar 4. 6 Proses *Word Normalization*

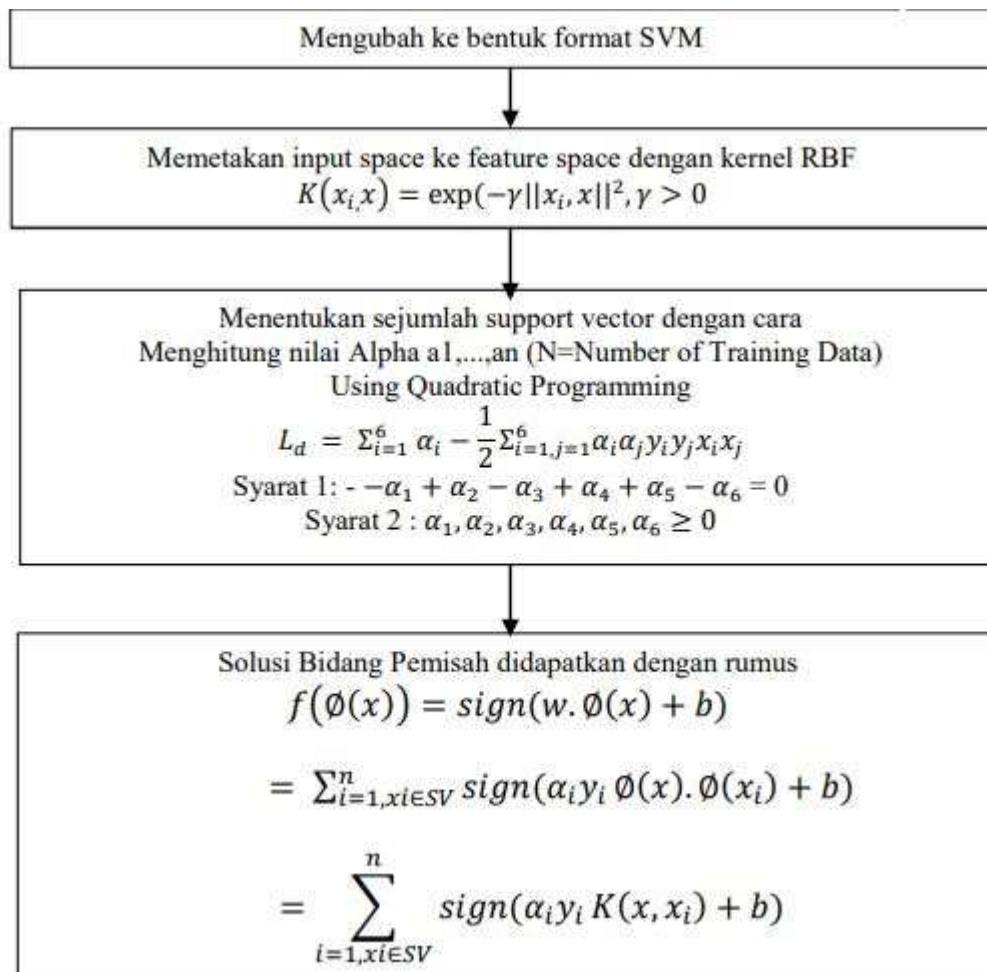
Berikut ini hasil dari *Word Normalization* pada Tabel 4.4.

Tabel 4. 4 Word Normalization

No	Data Tweet	Word Normalization
1	Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="https://pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>	Harus Saat Ini Pemerintah Lampung Bisa berikan Bantuan kepada mahasiswat tersebut ingat Begitu Derasnya Donasi Yang Masuk lalui tim gugus tugas

#### 4.1.7 Analisa Algoritma *Support Vector Machine*

Pelatihan SVM bertujuan untuk menemukan vektor  $\alpha$ , nilai  $W$  dan konstanta  $b$  untuk mendapatkan hyperplane terbaik. Pada penelitian kali ini, data yang digunakan sebagai pelatihan adalah data komentar. Dalam pelatihan SVM, setiap model klasifikasi dilatih pada data dari dua kelas ke- $i$  dan kelas- $j$ . Data masukan yang akan digunakan untuk proses pelatihan yaitu data komentar P1 sampai P6 telah diberi kelas dan telah melalui tahapan preprocessing. Sesuai dengan data masukan, data komentar diberi kelas positif dan kelas negatif, lalu diberikan label kelas 1 atau -1 yang dimana kelas -1 merupakan kelas negatif sedangkan kelas 1 merupakan kelas positif. Berikut alur proses pelatihan SVM dapat dilihat pada gambar 4.7.



Gambar 4. 7 Algoritma *Support Vector Machine*

Simulasi Contoh Soal dengan Perhitungan Metode *Support Vector Machine*. Misalkan terdapat 1 buah dokumen yang telah melalui tahapan *preprocessing* seperti terdapat 2 tweet

Tweet 1 :

jokowi sudah mau selesai 2 periodenya covid yang belum selesai

Tweet 2 :

kasus covid-19 beneran ada, dihimbau untuk jangan beranggapan sudah selesai



Kemudian akan dilakukan pemisahan terhadap kata positif dan negatif seperti pada Table 3.10

Tabel 3. 1 Pemisahan Kata

Tweet 1 (positif)	tweet 2 (negatif)
kasus	beranggapan
ada	selesai
Yang	dihimbau
belum	
Cerita	
Periode	

Kemudian akan melakukan pembobotan dengan term frekuensi seperti pada Table 3.11

Tabel 3. 2 Pembobotan Kata

No	Kata	Positif	Negatif
1	kasus	1	0
2	ada	1	0
3	Yang	1	0
4	belum	1	0
5	Cerita	1	0
6	Priode	1	0
7	beranggapan	0	1
8	selesai	0	1
9	dihimbau	0	1
No	Jumlah	6	3

Kemudian didapatkan count positif berjumlah 6 dan count negatif berjumlah 3 dengan total kata 9 kata. selanjutnya proses klasifikasi dan data latih

1. Hitung probabilitas prior setiap kategori yang menjadi kategori ada 2 kategori yaitu positif dan negatif

$$p(\text{positif}) = \frac{fx(\text{positif})}{|c|} = \frac{1}{6} = 0,14$$

$$p(\text{negatif}) = \frac{fx(\text{negatif})}{|c|} = \frac{1}{3} = 0,14$$

2. Menghitung probabilitas likelihood setiap term dari semua dokumen. jumlah seluruh kata 13 dan 7 untuk positif 6 untuk negatif. banyaknya term tergantung proses pada data

$$p(w|\text{positif/negatif}) = \frac{(n.(\text{positif/netral}) + 1}{(n.\text{positif/negatif})} \dots\dots\dots$$

1. Probabilitas kata "mau"

$$p(\text{mau/positif}) = \frac{1 + 1}{3 + 9} = 0.1$$

$$p(\text{mau/negatif}) = \frac{0 + 1}{3 + 9} = 0.0523$$

2. Probabilitas kata "selesai"

$$p(\text{selesai/positif}) = \frac{1 + 1}{6 + 9} = 0.1$$

$$p(\text{selesai/negatif}) = \frac{0 + 1}{3 + 9} = 0.0523$$

3. Probabilitas kata "sudah"

$$p(\text{sudah/positif}) = \frac{1 + 1}{6 + 9} = 0.1$$

$$p(\text{suda/negatif}) = \frac{0 + 1}{3 + 9} = 0.0523$$

4. Probabilitas kata "bisa"

$$p(\text{bisa/positif}) = \frac{1 + 1}{6 + 9} = 0.1$$

$$p(\text{bisa/negatif}) = \frac{0 + 1}{3 + 9} = 0.0523$$

5. Probabilitas kata "cerita"

$$p(\text{cerita/positif}) = \frac{1 + 1}{6 + 9} = 0.1$$

$$p(\text{cerita/negatif}) = \frac{0 + 1}{3 + 9} = 0.0523$$

3. Probabilitas kata "beneran"

$$p(\text{beneran/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{beneran/negatif}) = \frac{1 + 1}{3 + 9} = 0.10$$

6. Probabilitas kata "apalagi"

$$p(\text{apalagi/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{apalagi/negatif}) = \frac{1 + 1}{3 + 9} = 0.10$$

8. Probabilitas kata "bawa"

$$p(\text{bawa/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{bawa/negatif}) = \frac{1 + 1}{3 + 9} = 0.10$$

9. Probabilitas kata "isu"

$$p(\text{isu/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{isu/negatif}) = \frac{1 + 1}{3 + 9} = 0,10$$

10. Probabilitas kata " berapa"

$$p(\text{berapa/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{berapa/negatif}) = \frac{1 + 1}{3 + 9} = 0,05$$

11. Probabilitas kata " periode"

$$p(\text{periode/positif}) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{periode/negatif}) = \frac{1 + 1}{3 + 9} = 0,10$$

12. Probabilitas kata " tidak"

$$p\left(\frac{\text{tidak}}{\text{positif}}\right) = \frac{0 + 1}{6 + 9} = 0.05$$

$$p(\text{tidak/negatif}) = \frac{1 + 1}{3 + 9} = 0,10$$

Berikut ini table data uji yang terdapat pada Table 3.12

Tabel 3. 3 Data Uji

No	Data Uji
1	Virus covid sudah mau selesai 2 tahun ini

Pada data uji kata "sudah, selesai"

$$p(\text{uji}|\text{positif})$$

$$= p(\text{positif}) \times P(\text{sudah}|\text{positif}) \times P(\text{selesai}|\text{positif}) \times P(\text{cerita}|\text{positif})$$

$$= 0,14 \times 0,1 \times 0,1 \times 0,1$$

$$= 0,00014$$

$$p(\text{uji}|\text{negatif})$$

$$= p(\text{negatif}) \times P(\text{periode}|\text{negatif})$$

$$= 0,14 \times 0,10$$

$$= 0,014$$

Kesimpulan nilai probabilitas tertinggi yaitu sebesar 0,014 pada  $p(\text{uji}|\text{negatif})$  sehingga tweet tersebut diklasifikasikan kedalam kelas negative

UNIVERSITAS ISLAM NEGERI  
SUMATERA UTARA MEDAN

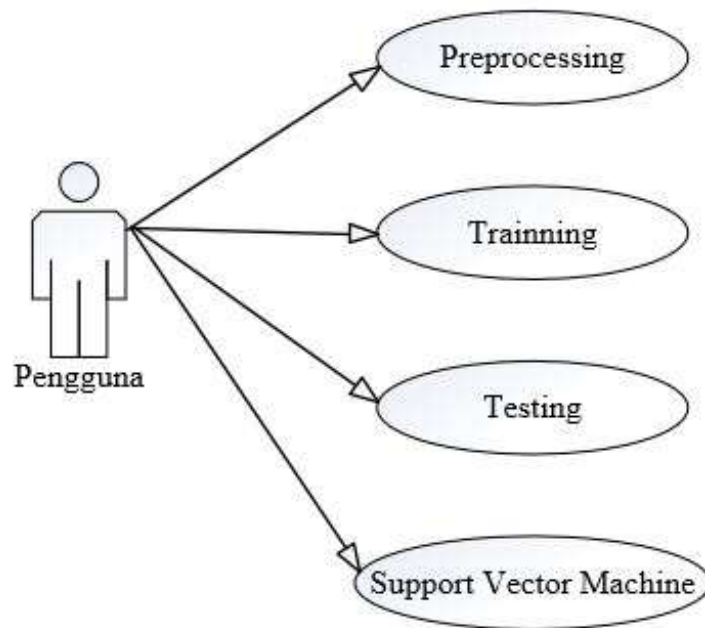
## 4.2 Perancangan Sistem

Pada tahapan perancangan sistem ini akan di jelaskan proses–proses pembuatan sebuah sistem sentimen analisis terhadap topik *Covid-19* dengan algoritma *SVM* agar proses alurnya lebih jelas dengan menggunakan model perancangan sistem *Unified Modern Language* (UML).

### 1. Use Case Diagram

*Use case diagram* menggambarkan fungsionalitas yang di harapkan dari sebuah sistem sentimen analisis terhadap topik covid-19 dengan algoritma

*SVM* sehingga terdapat *use case* yang mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan proses yang akan di buat, berikut ini perancangan *use case* sistem yaitu:



Gambar 4. 8 *Usecase* Diagram Penelitian

Dalam *Use Case Diagram* di atas dijelaskan bahwa dalam *user* terdapat aktivitas seperti *preporcessing*, pelatihan, pengujian dan algoritma SVM dan hasil yang berfungsi sebagai proses dalam melakukan sentimen analisis terhadap topik *covid-19* positif, netral dan negatif. Skenario Use Case memiliki komponen-komponen: nama use case, aktor (actor), deskripsi, kondisi awal (pre-condition), skenario utama (main flow) dan skenario normal. Berikut ini skenario usecase yang terdapat pada tabel 4.1

Tabel 4. 5 Skenario preprocessing

Identifikasi use case	
Nama use case	Preprocessing
Deskripsi Use case	Use Case ini memungkinkan proses preprocvessing dat

Aktor	Pengguna
Kondisi awal	Pengguna melakukan preprocessing
Skenario utama	
Aksi-aktor	Reaksi sistem
1. Aktor melakukan preprocessing	2. Menampilkan tahapan dari preprocessing
3. pengguna melihat hasil preprocessing	4. selesai

Tabel 4. 6 Skenario training

Identifikasi use case	
Nama use case	training
Deskripsi Use case	Use Case ini memungkinkan proses training data
Aktor	Pengguna
Kondisi awal	Pengguna melakukan training data
Skenario utama	
Aksi-aktor	Reaksi sistem
1. Aktor melakukan training data	2. Menampilkan tahapan dari training
3. pengguna melihat akurasi training	4. selesai

UNIVERSITAS ISLAM NEGERI  
SUMATERA UTARA MEDAN

Tabel 4. 7 Skenario testing

Identifikasi use case	
Nama use case	Testing
Deskripsi Use case	Use Case ini memungkinkan proses testing data
Aktor	Pengguna
Kondisi awal	Pengguna melakukan testing data
Skenario utama	

Aksi-aktor	Reaksi sistem
1. Aktor melakukan testing data	2. Menampilkan tahapan dari testing
3. pengguna melihat akurasi testing	4. selesai

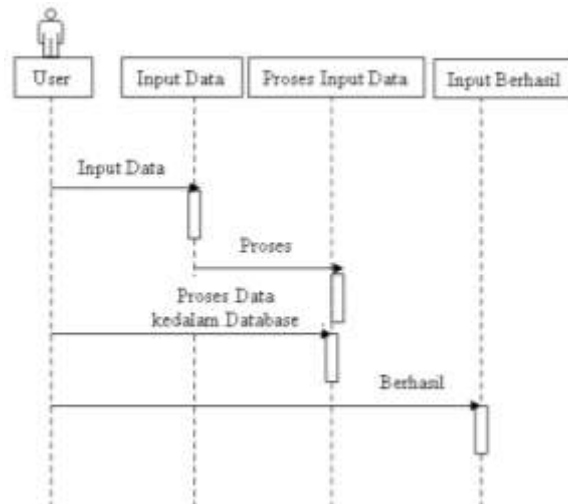
Tabel 4. 8 Skenario SVM

Identifikasi use case	
Nama use case	Support vector machine
Deskripsi Use case	Use Case ini memungkinkan proses klasifikasi dengan Support vector machine
Aktor	Pengguna
Kondisi awal	Pengguna melakukan klasifikasi data
Skenario utama	
Aksi-aktor	Reaksi sistem
1. Aktor melakukan klasifikasi dengan Support vector machine	2. Menampilkan tahapan dari klasifikasi
3. pengguna melihat hasil dari Support vector machine	4. selesai

## 2. *Sequence Diagram Input Data*

*Sequence Diagram* input data menggambarkan prosedur informasi input data yang dilakukan oleh pengguna. Berikut adalah *Sequence Diagram* input data yang dirancang penulis pada Gambar 4.9:

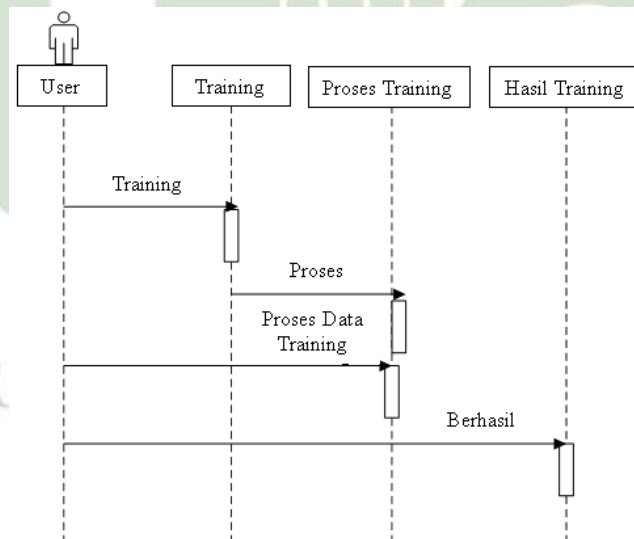




Gambar 4. 9 *Sequence Diagram Input Data*

### 3. *Sequence Diagram Training*

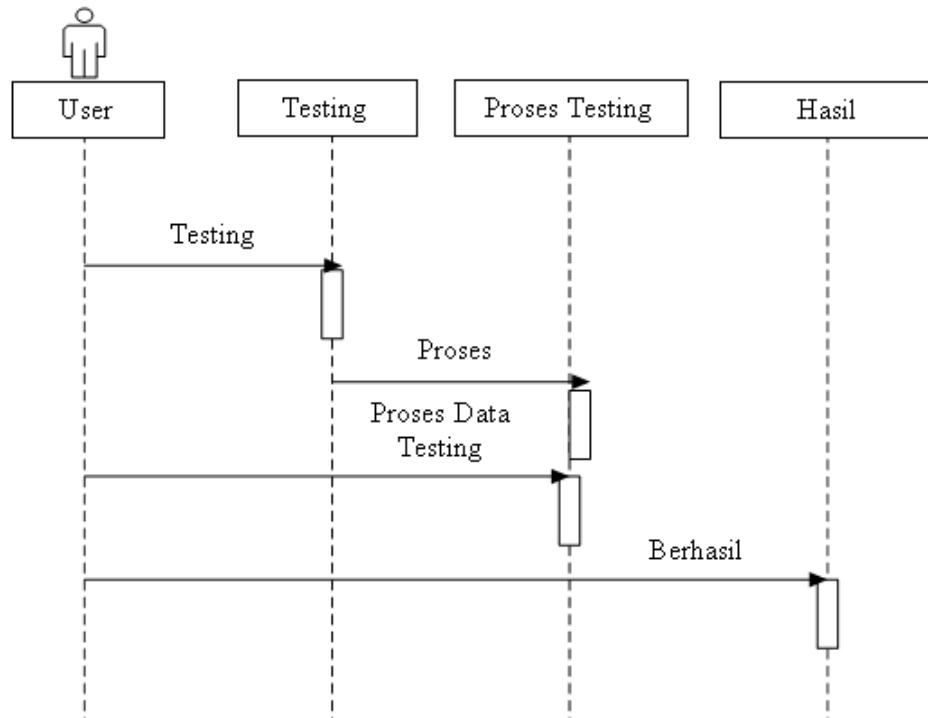
*Sequence Diagram Training* menggambarkan prosedur informasi *training* data input yang dilakukan. Berikut adalah hasil data yang dirancang penulis pada Gambar 4.10:



Gambar 4. 10 *Sequence Diagram Training*

### 4. *Sequence Diagram Testing*

*Sequence Diagram Testing* menggambarkan prosedur informasi *testing* data input yang dilakukan. Berikut adalah input data yang dirancang penulis pada Gambar 4.11:



Gambar 4. 11 *Sequence Diagram Testing*

### 4.3 Hasil dan Implementasi

Pada bagian ini akan dipaparkan hasil dari implementasi proses pembuatan dan implementasi sistem klasifikasi kata positif, netral dan negatif pada *tweet* covid 19 dengan teknik sentimen analisis menggunakan algoritma support vector machine

#### 4.3.1 Proses Akuisisi Data

Berikut ini tabel dari hasil *crawling* terhadap data *tweet* yang mengandung kata yang berkaitan dengan topik yang dianalisis:

```
df = pd.read_csv('daring.csv', usecols=['text'])
df.head()
```

Gambar 4. 12 Menampilkan Hasil *Crawling*

Tabel 4. 9 Data *Tweet* Covid 19

No	Data <i>Tweet</i>
----	-------------------

1	<a href="https://bit.ly/2L6CcbB">https://bit.ly/2L6CcbB</a>   Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="https://pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>
2	Klo kata gw Pemerintah tuh lagi menerapkan Herd immunity karena kemampuan pemerintah nihil soal pemberantasan covid. Gambling udah siapa kuat dia bertahan hidup yang lemah maka mati. Padahal dinegara lain begitu serius menangani virus tsb.
3	Saat ini yang bisa saya lakukan hanya menyiapkan makanan untuk berbuka dan sahur untuk 10 kepala keluarga terdampak covid di Tasik dan tidak dpt bantuan dr pemerintah krn tidak masuk di data. Untuk akhir Ramadhan kita sedang rencanakan paket sembako lengkap. Semoga ada rejekinya.
4.	Satu Warga Positif Corona, Bupati: Kondisi ini Menuntut Kebersamaan Kita: RAKYAT NEWS, BANTAENG “ Pemerintah Kabupaten Bantaeng melalui tim gugus tugas penanganan Covid-19 di Bantaeng merilis [â€¦] <a href="http://dlvr.it/RWBnbD">http://dlvr.it/RWBnbD</a> #rakyatdotnews #referensiterkini #rakyatnews <a href="https://pic.twitter.com/jOwYbh3Txw">pic.twitter.com/jOwYbh3Txw</a>
5	emosi bgt, lg kondisi begini gue disuruh liputan tentang covid. ini dosen gue bercanda ga si? pemerintah aja nyuruh #dirumahaja masa gue disuruh liputan keluar rumah ðŸ™¸
6	ANTISIPA penyebaran pandemi COVID-19 membutuhkan kerjasama antara pemerintah dan masyarakat. Pandemi ini adalah musibah berkatagori non alam.

Pada Tabel 4.8 dijelaskan dokumen yang diakuisisi yaitu dokumen berlabel teks media sosial *Twitter* yang berasal dari *tweet* pengguna tentang Covid 19. Kemudian dokumen tersebut dimasukkan sebagai proses pembelajaran, pada tahap

ini data masih utuh atau belum dibersihkan sehingga data teks yang akan diolah masih bercampur dengan karakter-karakter lain masih melekat pada data tersebut.

### 4.3.2 Preprocessing

Tahapan pengolahan data asli yang telah diakuisisi berupa data teks sebelum data tersebut diolah lebih lanjut, tujuan dari *preprocessing* adalah menghilangkan *noise*, memperjelas fitur, mengkonversi data asli agar sesuai dengan kebutuhan serta memperbesar atau memperkecil data. Dalam *preprocessing* sendiri terdapat beberapa tahapan antara lain:

### 4.3.3 Proses Pembersihan Data

Pada proses pembersihan dokumen akan melakukan pembersihan terhadap data-data hasil *crawling* tentang Covid 19. Berikut ini hasil dari pembersihan dokumen pada Tabel 4.2:

```
stopword = set(stopwords.words('indonesian'))

with open('combined_slang_words.txt') as f:
    slang_words = f.read()
    slang_words = ast.literal_eval(slang_words)

def text_cleaning(teks, from_tokenize=False, delete_username=False):
    if from_tokenize:
        teks = ' '.join(teks)

    if not from_tokenize:
        # Remove tab new line
        pattern_0 = r'[\n\t\r]'
        teks = re.sub(pattern_0, ' ', teks)
        teks = re.sub('""', '', teks)

    if not from_tokenize:
        # Remove non ASCII (emoticon, chinese word (if exists))
        teks = teks.encode('ascii', 'replace').decode('ascii')
```

Tabel 4. 10 Pembersihan data

No	Data <i>Tweet</i>
1	Pemerintah lampung bantuan mahasiswa derasnya donasi masuk pemerintah provinsi Lampung tim gugus tugas covid 19
2	pemerintah menerapkan herd immunity kemampuan pemerintah nihil pemberantasan covid gambling kuat bertahan hidup lemah mati negara serius menangani virus

3	Lakukan makanan berbuka sahur 10 kepala keluarga terdampak covid tasik bantuan pemerintah masuk data ramadan rencanakan paket sembako lengkap semoga rejekinya
4.	Warga Positif Corona Bupati Kondisi Menuntut Kebersamaan Rakyat News Bantaeng Pemerintah Kabupaten Bantaeng Tim Gugus Tugas Penanganan Covid 19 Bantaeng Merilis Rakyatdotnews Referensiterkini Rakyatnews
5.	Emosi Banget Kondisi Disuruh Liputan Covid Dosen Bercanda Sih Pemerintah Menyuruh Dirumahaja Disuruh Liputan Rumah
6.	Antisipa Penyebaran Pandemi Covid 19 Membutuhkan Kerja Pemerintah Masyarakat Pandemi Musibah Berkatagori Non Alam Prosesnewsid Psbb Gorontalo

#### 4.3.4 Case Folding

Pada Tabel 4.9 akan dijelaskan tahapan penyeragaman bentuk huruf dalam dokumen, huruf kapital diubah menjadi huruf kecil dan huruf-huruf diseragamkan dari A sampai Z dan selain huruf akan dihilangkan karena dianggap delimiter.

```
# Case folding
teks = teks.lower()

if not from_tokenize:
    # Remove url
    pattern_1 = r'((\w|\W)\.|\.\s+)|(https?:\/\/[\s]+)'
    teks = re.sub(pattern_1, '', teks)

if not from_tokenize:
    # Replace Username
    pattern_2 = r'@[\s]+'
    teks = re.sub(pattern_2, '#USERNAME', teks)

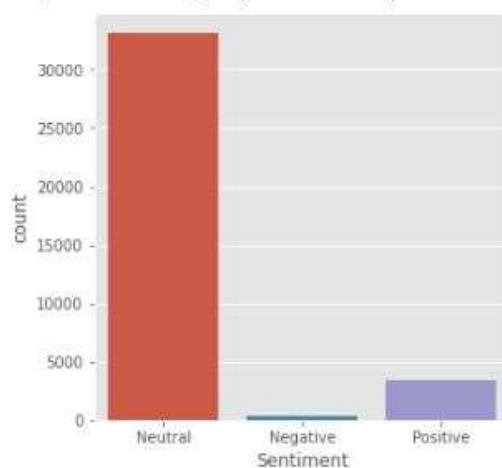
if not from_tokenize:
    # Replace Username
    pattern_3 = 'rt'
    teks = re.sub(pattern_3, '#USERNAME', teks)
```

Tabel 4. 11 Proses *Case Folding*

No	Data <i>Tweet</i>
1	pemerintah lampung bantuan mahasiswa derasnya donasi masuk pemerintah provinsi lampung tim gugus tugas covid 19
2	pemerintah menerapkan herd immunity kemampuan pemerintah nihil pemberantasan covid gambling kuat bertahan hidup lemah mati negara serius menangani virus
3	lakukan makanan berbuka sahur 10 kepala keluarga terdampak covid tasik bantuan pemerintah masuk data ramadan rencanakan paket sembako lengkap semoga rejekinya
4	warga positif corona bupati kondisi menuntut kebersamaan rakyat news bantaeng pemerintah kabupaten bantaeng tim gugus tugas penanganan covid 19 bantaeng merilis rakyatdotnews referensiterkini rakyatnews
5	emosi banget kondisi disuruh liputan covid dosen bercanda sih pemerintah menyuruh dirumahaja disuruh liputan rumah
6	antisipasi penyebaran pandemi covid 19 membutuhkan kerja pemerintah masyarakat pandemi musibah berkatagori non alam prosesnewsid psbb gorontalo

#### 4.4 Distribusi Kata

Distribusi kata merupakan kata yang paling banyak frekuensi yang terdapat pada data-data *tweet* hasil *crawling* tentang covid 19, berikut ini distribusi kata yang akan ditampilkan dengan diagram pada Gambar 4.13:



Gambar 4. 13 Distribusi Kata





Gambar 4. 15 Word Cloud Kata Tweet dengan label negatif



Gambar 4. 16 Word Cloud Kata Tweet dengan label Netral



#### 4.5 Support Vector Machine

Berikut ini pengujian dalam melakukan klasifikasi dengan algoritma *support vector machine* dalam melakukan klasifikasi terdapat data pelatihan dan pengujian seperti pada Tabel 4.6 berikut ini:

Tabel 4. 12 Data Pengujian dan Data Pelatihan

Perbandingan Data		Jumlah data	
Data Pelatihan	Data Pengujian	Data Pelatihan	Data Pengujian
90%	10%	33229	3693

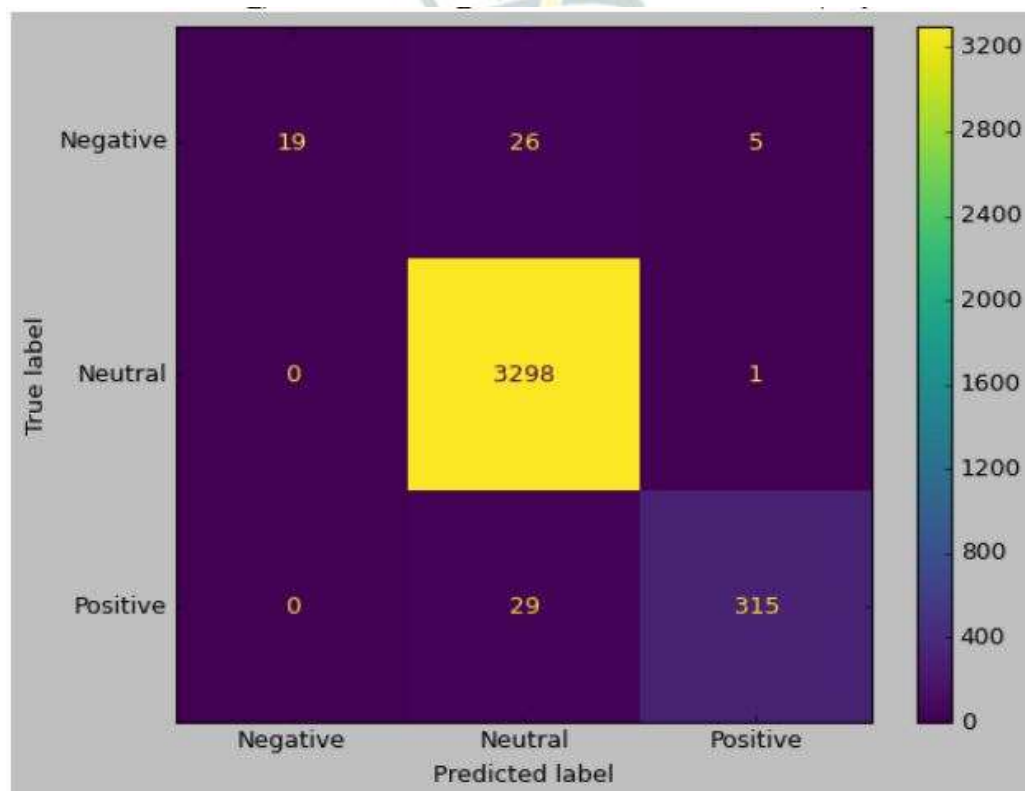
Pada Tabel 4.7 dijelaskan dari hasil akuisisi data kemudian lewat tahap *preprocessing* data yang ada sebanyak 33229 yang dibagi kedalam 3 kelas sentimen yaitu positif 0, netral 1 dan negatif 2. Data yang sudah dinormalisasi sebelum dimasukan kemesin klasifikasi, data tersebut dibagi menjadi 2 yaitu data pelatihan dan data pengujian. Data yang akan dilatihkan maupun diujikan telah dibagi dan ditip percobaannya dilakukan menggunakan algoritma SVM data yang sudah dibagi pada setiap percobaannya dilakukan klasifikasi dengan fungsi *polynomial* kernel yang memetakan data *non-linear* sehingga mendapat data set baru model pembelajaran pada tiap-tiap percobaan. Hasil dari model pembelajaran diklasifikasikan dengan percobaan pengujian diperoleh *matrix* dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi.

#### 4.6 Confusion Matrix

*Confusion Matrix* merupakan pengukuran performa untuk masalah klasifikasi terhadap penelitian yang penulis bahas tentang dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 6 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada enam istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu *True Positif*, *True Netral*, *True Negatif*, *False Positif*, *False Netral* dan *False Negatif*

	precision	recall	f1-score	support
Negative	1.00	0.38	0.55	50
Neutral	0.98	1.00	0.99	3299
Positive	0.98	0.92	0.95	344
accuracy			0.98	3693
macro avg	0.99	0.77	0.83	3693
weighted avg	0.98	0.98	0.98	3693

Gambar 4. 17 Hasil Precision, Recall, dan F1-score



Gambar 4. 18 Confussion Matrix

Keterangan Gambar 4.18 akan menjelaskan bahwa gambar diatas menunjukkan sebaran prediksi data TP (*True Positive*), TNe (*True Netral*), TN (*True Negatif*), FP (*False Positif*), FNe (*False Netral*), FN (*False Negatif*).. Berdasarkan Gambar 4.18 akan dihitung nilai akurasi sebagai berikut:

Keterangan:

$$TP + TNe + TN / TP + TNe + TN + FP + FNe + FN * 100 =$$

$$19 + 3298 + 315 = 3632 / 19 + 26 + 5 + 0 + 3298 + 1+0+29+315 = 3693$$

$$3632 / 3693 = 0,98 * 100 = 98\%$$

Dan akurasi yang didapat pada klasifikasi terhadap topik covid-19 yang di analisis, yaitu:

$$\text{Akurasi Skor} = 0,98 * 100 = 98\%$$

