



# **ANALISIS POPULARITAS IPB BERDASARKAN DATA TWITTER MENGGUNAKAN AWAN KATA, GEOVISUALISASI, DAN PEMODELAN KLASIFIKASI**

© Hak cipta milik IPB (Institut Pertanian Bogor)

**ABDULLAH GHIFARI**



**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2018**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



**PERNYATAAN MENGENAI SKRIPSI DAN  
SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa skripsi berjudul Analisis Popularitas IPB berdasarkan Data Twitter Menggunakan Awan Kata, Geovisualisasi, dan Pemodelan Klasifikasi adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2018

*Abdullah Ghifari*  
NIM G14140017

10

Bab I  
Latar Belakang

Bogor Agricultural University

- Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

  - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - Pengutipan tidak mengiklan kepentingan yang wajar IPB.

2. Dilarang menggumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## ABSTRAK

ABDULLAH GHIFARI. Analisis Popularitas IPB berdasarkan Data Twitter Menggunakan Awan Kata, Geovisualisasi, dan Pemodelan Klasifikasi. Dibimbing oleh KHAIRIL ANWAR NOTODIPUTRO dan BAGUS SARTONO.

Twitter merupakan media sosial yang sangat aktif digunakan di Indonesia. Twitter menjadi suatu media untuk para data analis untuk menggali informasi mengenai masalah-masalah yang ada di media sosial. Karakteristik yang dilihat adalah mengenai topik yang sering dibahas mengenai IPB, pemetaan frekuensi *tweets* pada berbagai wilayah di Indonesia, dan perbandingan *tweets* dengan 9 universitas lain terbaik di Indonesia. Hal yang akan dibandingkan adalah program pendidikan, sumber daya manusia, dan kemahasiswaan. Pelabelan respon tersebut akan menggunakan otomatisasi dengan memanfaatkan metode pemodelan klasifikasi seperti regresi logistik, *support vector machine*, *multilayer perceptron*, bayes naïf, dan hutan acak. Penelitian ini akan menggunakan awan kata (*word cloud*) untuk mengetahui topik yang akan di bahas. Kemudian, *tweets* akan dipetakan secara geovisualisasi berdasarkan lokasi pada bio pengguna pada suatu *tweets*. Hasil dari penelitian ini menjelaskan bahwa IPB memiliki respon yang positif terhadap topik kemahasiswaan namun masih cenderung negatif pada beberapa topik yang lain. Manfaat dari penelitian ini diharapkan dapat menjadi bahan pertimbangan untuk Humas IPB dalam melakukan strategi sosialisasi IPB kedepannya.

Kata kunci: awan kata, geovisualisasi, IPB, klasifikasi, twitter

## ABSTRACT

ABDULLAH GHIFARI. IPB Popularity Analysis using Word Cloud, Geovisualization, and Classification in Twitter Data. Supervised by KHAIRIL ANWAR NOTODIPUTRO and BAGUS SARTONO.

Twitter is a very active social media used in Indonesia. Twitter became a platform for data analysts to get information about the problems that exist in social media. Characteristics seen are on topics frequently discussed about IPB, mapping the frequency of tweets in various regions of Indonesia, and comparison of tweets with 10 other best universities in Indonesia. Things that will be compared are education programs, human resources, and student affairs. Labeling of the category will use automation by utilizing classification modeling methods such as logistic regression, support vector machine, multilayer perceptron, naive bayes, and random forest. This research will use awan kata to know the topic to be discussed. Then, the tweets will be geovisualized based on the location on the user's bio on a tweets. The results of this study explain that IPB has a positive response to the topic of student affairs and can still be used on several other topics. The benefits of this research are expected to be a consideration for IPB Public Relations in conducting IPB socialization strategy in the future.

Keywords: classification, geovisualization, IPB, twitter, word cloud



# **ANALISIS POPULARITAS IPB BERDASARKAN DATA TWITTER MENGGUNAKAN AWAN KATA, GEOVISUALISASI, DAN PEMODELAN KLASIFIKASI**

© Hak cipta milik IPB (Institut Pertanian Bogor)

**ABDULLAH GHIFARI**

Skripsi  
sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Statistika  
pada  
Departemen Statistika

**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2018**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Judul Skripsi: Analisis Popularitas IPB berdasarkan Data Twitter Menggunakan  
Awan Kata, Geovisualisasi, dan Pemodelan Klasifikasi  
: Abdullah Ghifari  
: G14140017

Hak Cipta Dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan hakikat atau tinjauan suatu masalah  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Disetujui oleh



Hak cipta milik IPB

Dr Ir Khairil Anwar Notodiputro, MS

Pembimbing I

Dr Bagus Sartono, MSI

Pembimbing II

Diketahui oleh



Dr Anang Kurnia, MSI

Ketua Departemen

Tanggal Lulus: 16 AUG 2018

Bogor Agri



## PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2017 ini ialah analisis data pada media sosial, dengan judul Analisis Popularitas IPB berdasarkan Data Twitter Menggunakan Awan Kata, Geovisualisasi, dan Pemodelan Klasifikasi.

Pada kesempatan ini, penulis mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan arahan selama proses penulisan skripsi ini berlangsung, di antaranya:

1. Keluarga yang selalu memberikan dukungan dan doa untuk penulis.
2. Bapak Prof Dr Ir Khairil Anwar Notodiputro, MS dan Bapak Dr Bagus Sartono, MSi selaku Dosen Pembimbing Skripsi Departemen Statistika FMIPA IPB.
3. Ibu Dr Utami Dyah Syafitri selaku Dosen Pengaji Skripsi Departemen Statistika FMIPA IPB.
4. Seluruh dosen Departemen Statistika FMIPA IPB yang telah membantu melancarkan penulisan karya ilmiah ini.
5. Hesti yang sudah mendukung dan mendoakan penulis dalam menyelesaikan karya ilmiah ini.
6. Junjun, Ayunda, Muhyi, Albasia, Rere dan Ado yang sudah mendukung penulis dalam menyelesaikan karya ilmiah ini.
7. Mahasiswa/i Departemen Statistika FMIPA IPB angkatan 51.
8. Seluruh pihak terkait yang telah membantu penulis menyusun yang tidak dapat dituliskan satu persatu.

Semoga karya ilmiah ini bermanfaat bagi seluruh pihak yang membutuhkan. Penulis mohon maaf atas kekurangan dan kesalahan yang terdapat pada karya ilmiah ini.

Bogor, Agustus 2018

*Abdullah Ghifari*

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## DAFTAR ISI

|                               |      |
|-------------------------------|------|
| DAFTAR TABEL                  | vii  |
| DAFTAR GAMBAR                 | viii |
| DAFTAR LAMPIRAN               | ix   |
| DAFTAR ISTILAH                | ix   |
| PENDAHULUAN                   | 1    |
| Latar Belakang                | 1    |
| Tujuan Penelitian             | 2    |
| TINJAUAN PUSTAKA              | 2    |
| Media Sosial                  | 2    |
| Twitter                       | 2    |
| Penambangan Teks              | 3    |
| Awan Kata                     | 3    |
| <i>Support Vector Machine</i> | 4    |
| Regresi Logistik              | 5    |
| Bayes Naif                    | 6    |
| Hutan Acak                    | 7    |
| <i>Multilayer Perceptron</i>  | 8    |
| Ukuran Kebaikan Model         | 8    |
| K-Lipat Validasi Silang       | 10   |
| Geovisualisasi                | 10   |
| METODE                        | 11   |
| Data                          | 11   |
| Prosedur Analisis             | 11   |
| Lingkungan Pengembangan       | 16   |
| HASIL DAN PEMBAHASAN          | 17   |
| Eksplorasi Data               | 17   |
| Awan Kata                     | 19   |
| Geovisualisasi                | 20   |
| Persiapan Data                | 21   |
| Regresi Logistik              | 22   |
| Hutan Acak                    | 24   |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



|  |                               |    |
|--|-------------------------------|----|
|  | <i>Support Vector Machine</i> | 26 |
|  | Bayes Naif                    | 28 |
|  | <i>Multilayer Perceptron</i>  | 30 |
|  | Perbandingan Model            | 32 |
|  | Prediksi                      | 34 |
|  | <b>SIMPULAN DAN SARAN</b>     | 48 |
|  | Simpulan                      | 48 |
|  | Saran                         | 48 |
|  | <b>DAFTAR PUSTAKA</b>         | 48 |
|  | <b>LAMPIRAN</b>               | 51 |
|  | <b>RIWAYAT HIDUP</b>          | 56 |

© Hak Cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## DAFTAR TABEL

|   |    |
|---|----|
| 1 Ilustrasi tabel klasifikasi   | 9  |
| 2 Contoh <i>tweets</i> yang telah diambil   | 13 |
| 3 Ilustrasi <i>tweets</i> yang telah diberikan label  | 14 |
| 4 Ilustrasi <i>tweets</i> yang telah dilakukan prapemrosesan  | 15 |
| 5 Ilustrasi bentuk <i>document term matrix</i>  | 15 |
| 6 Banyaknya <i>tweets</i> setiap kelas pada respon topik  | 22 |
| 7 Banyaknya tweets setiap kelas pada respon sentimen  | 22 |
| 8 Akurasi setiap lipatan dengan menggunakan metode regresi logistik (dalam persentase)  | 23 |
| 9 Akurasi setiap lipatan dengan menggunakan metode hutan acak (dalam persentase)  | 25 |
| 10 Akurasi setiap lipatan dengan menggunakan metode <i>support vector machine</i> (dalam persentase)  | 27 |
| 11 Akurasi setiap lipatan dengan menggunakan metode bayes naif (dalam persentase)   | 28 |
| 12 Akurasi setiap lipatan dengan menggunakan metode <i>multilayer perceptron</i> (dalam persentase)   | 31 |
| 13 Hasil akurasi pada respon topik dengan menggunakan 10-lipat validasi silang (dalam persentase)   | 32 |
| 14 Hasil AUC pada respon respon topik   | 33 |
| 15 Hasil akurasi pada respon sentimen dengan menggunakan 10-lipat validasi silang (dalam persentase)  | 33 |
| 16 Hasil AUC pada respon respon sentimen  | 33 |
| 17 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk IPB  | 35 |
| 18 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk IPB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal | 35 |
| 19 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UGM  | 36 |
| 20 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UGM dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal | 37 |
| 21 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UI   | 38 |
| 22 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UI dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal  | 38 |
| 23 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk ITB  | 39 |
| 24 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk ITB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal | 40 |
| 25 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk ITS  | 41 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

|    |   |    |
|----|---|----|
| 26 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk ITS dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal  | 41 |
| 27 | Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UB  | 42 |
| 28 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal   | 42 |
| 29 | Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNAIR   | 43 |
| 30 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNAIR dengan nilai-p di atas diagonal dan statistik uji dibawah diagonal | 43 |
| 31 | Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNDIP   | 44 |
| 32 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNDIP dengan nilai-p di atas diagonal dan statistik uji dibawah diagonal | 45 |
| 33 | Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNHAS   | 46 |
| 34 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNHAS dengan nilai-p di atas diagonal dan statistik uji dibawah diagonal | 46 |
| 35 | Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNY   | 47 |
| 36 | Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNY dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal  | 48 |

## DAFTAR GAMBAR

|    |  |    |
|----|--|----|
| 1  | Ilustrasi pembentukan <i>hyperplane</i> dengan memaksimumkan margin            | 4  |
| 2  | Ilustrasi grafik ROC dengan AUC pada masing masing kelas                       | 9  |
| 3  | Ilustrasi k-lipat validasi silang dengan k=5                                   | 10 |
| 4  | Diagram alir tahapan analisis  | 12 |
| 5  | Banyaknya <i>tweets</i> setiap universitas                                     | 17 |
| 6  | Frekuensi <i>tweets</i> dengan topik IPB setiap hari                           | 18 |
| 7  | Frekuensi <i>tweets</i> dengan topik UI setiap hari                            | 18 |
| 8  | Frekuensi <i>tweets</i> dengan topik UGM setiap hari                           | 19 |
| 9  | Awan kata dengan topik IPB   | 20 |
| 10 | Peta banyaknya <i>tweets</i> tentang IPB                                       | 21 |
| 11 | Contoh 5 <i>tweets</i> hasil prapemrosesan menjadi <i>document-term matrix</i> | 21 |
| 12 | Kurva ROC untuk respon topik dengan metode regresi logistik                    | 23 |
| 13 | Kurva ROC untuk respon sentimen dengan metode regresi logistik                 | 24 |
| 14 | Kurva ROC untuk respon topik dengan metode hutan acak                          | 25 |
| 15 | Kurva ROC untuk respon sentimen dengan metode hutan acak                       | 26 |
| 16 | Kurva ROC untuk respon topik dengan metode <i>support vector machine</i>       | 27 |



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

|    |  |    |
|----|--|----|
| 17 | Kurva ROC untuk respon sentimen dengan metode <i>support vector machine</i>      | 28 |
| 18 | Kurva ROC untuk respon topik dengan metode bayes naif                            | 29 |
| 19 | Kurva ROC untuk respon sentimen dengan metode bayes naif                         | 30 |
| 20 | Kurva ROC untuk respon topik dengan metode <i>multilayer perceptron</i>          | 31 |
| 21 | Kurva ROC untuk respon sentimen dengan metode <i>multilayer perceptron</i>       | 32 |
| 22 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas IPB   | 35 |
| 23 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UGM   | 36 |
| 24 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UI    | 37 |
| 25 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas ITB   | 39 |
| 26 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas ITS   | 40 |
| 27 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UB    | 41 |
| 28 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNAIR | 43 |
| 29 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNDIP | 44 |
| 30 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNHAS | 45 |
| 31 | Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNY   | 47 |

## DAFTAR LAMPIRAN

|   |   |    |
|---|---|----|
| 1 | Kata kunci yang digunakan untuk pengumpulan data  | 51 |
| 2 | Frekuensi <i>tweets</i> setiap universitas setiap hari 14 Januari 2018 sampai dengan 27 Februari 2018 | 52 |
| 3 | Awan kata setiap universitas  | 54 |



## DAFTAR ISTILAH

|   |                         |                         |
|---|-------------------------|-------------------------|
|   | Bahasa Indonesia        | Bahasa Inggris          |
| 1 | Awan Kata               | Word Cloud              |
| 2 | Bayes Naif              | Naive Bayes             |
| 3 | Geovisualisasi          | Geovisualization        |
| 4 | Hutan Acak              | Random Forest           |
| 5 | Kata Kunci              | Keywords                |
| 6 | K-Lipat Validasi Silang | K-Fold Cross Validation |
| 7 | Penarikan Data          | Crawling                |
| 8 | Pengambilan Contoh Acak | Sampling                |
| 9 | Penambangan Data        | Text Mining             |

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



# PENDAHULUAN

## Latar Belakang

Popularitas merupakan bagian dari elektabilitas. Dengan adanya modal popularitas maka akan lebih mudah bagi seseorang atau figur tersebut untuk mencuri perhatian masyarakat, melalui pemberitaan media yang diharapkan nantinya akan mempunyai nilai tambah untuk meningkatkan atau mendongkrak elektabilitas (Nimmo dan Rakhmat 2018). Oleh karena itu, untuk meningkatkan elektabilitas IPB diperlukan peningkatan dalam hal popularitas.

Peningkatan popularitas dapat dilakukan dalam berbagai hal. Terutama aktif pada media sosial. Karena menurut Puntoadi (2011), media sosial merupakan salah satu tempat membangun citra diri seseorang di depan orang lain. Sehingga, peran media sosial sangat berpengaruh terhadap popularitas suatu organisasi atau individu. Berita UGM (2017) menyatakan bahwa berdasarkan *Virtual Social Media Index* yang dirilis *Center for Digital Society FISIPOL UGM*, penggunaan media sosial di Indonesia berada pada peringkat 6 pada tahun 2017. Angka ini cukup tinggi untuk menyatakan bahwa Indonesia berperan aktif di media sosial.

Salah satu media yang sangat aktif di Indonesia adalah twitter. Menurut Dwi Ardiansyah sebagai *Head of Business Development Twitter South East and Australia*, dalam Berita Satu (2017) Indonesia merupakan lima besar pengguna twitter aktif di dunia. Berdasarkan data yang dirilis 77% pengguna twitter di Indonesia merupakan pengguna aktif. Dengan jumlah *tweets* sepanjang tahun 2016 mencapai 4,1 miliar *tweets*. Angka tersebut merupakan angka yang cukup besar. Dengan angka tersebut penyebaran informasi yang ada di Indonesia sangatlah aktif. Hal ini menjadi salah satu daya tarik bagi para peneliti untuk meneliti bagaimana pandangan masyarakat Indonesia pada suatu isu tertentu.

Penggunaan twitter yang tinggi menghasilkan ukuran data yang cukup tinggi. Untuk mengetahui karakteristik popularitas suatu topik tidak akan mudah dengan menggunakan teknik secara tradisional. Oleh karena itu, dibutuhkan automatisasi dalam pengolahan data twitter. Pada penelitian yang dilakukan oleh Choudhury (2012) mampu mengresponkan pengguna twitter secara otomatis dengan menggunakan berbagai metode klasifikasi untuk mendapatkan ketepatan yang tinggi.

Penelitian ini akan menghasilkan gambaran bagaimana sosialisasi IPB terkait program pendidikan, sumber daya manusia, dan kegiatan kemahasiswaan. Penelitian ini akan mampu membandingkan bagaimana popularitas IPB berdasarkan 3 faktor tersebut terhadap universitas lain. Hasil dari penelitian ini diharapkan menjadi bahan pertimbangan untuk Humas IPB dalam melakukan strategi sosialisasi IPB kedepannya.

Sesuai dengan uraian di atas, untuk menjamin efektifitas dan efisiensi penelitian maka berbagai pertanyaan berikut akan dijawab:

1. Seberapa jauh tingkat kepopuleran IPB secara umum?
2. Bagaimana tingkat kepopuleran IPB dibandingkan 9 perguruan tinggi terbaik lainnya yaitu UGM, ITS, ITB, UI, UNDIP, UNAIR, UB, UNHAS, dan UNY (Kemenristek Dikti 2017)?
3. Seberapa jauh tingkat kepopuleran IPB di berbagai wilayah di Indonesia?

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



- 2
4. Seberapa besar tingkat kepopuleran IPB dari sisi program pendidikan?
  5. Seberapa jauh tingkat kepopuleran IPB dari sisi sumber daya manusia?
  6. Seberapa jauh tingkat kepopuleran IPB dari sisi kemahasiswaan?

## Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut :

Mengetahui apa saja yang menjadi topik pembicaraan masyarakat Indonesia mengenai IPB.

Mengetahui frekuensi *tweet* tentang IPB di berbagai wilayah di Indonesia.

Membandingkan popularitas IPB dengan 9 universitas lain di twitter berdasarkan 3 faktor respon yaitu program pendidikan akademik, sumber daya masyarakat, dan kemahasiswaan.

## TINJAUAN PUSTAKA

### Media Sosial

Media sosial merupakan alat komunikasi yang memudahkan para pengguna untuk berinteraksi dengan efektif dan efisien. (Paramitha 2011). Menurut Evans (2008), media sosial selalu mengalami perubahan yang signifikan sehingga penerima dan penyebar informasi mampu berperan aktif ketika menggunakan media sosial. Sehingga, informasi dapat menyebar secara cepat.

Media sosial dapat membangun citra diri seseorang di depan orang lain. Popularitas dan eksistensi diri menjadi kepuasan tersendiri bagi mayoritas orang. Selain itu media sosial juga dapat menjadi sarana yang tepat untuk pemasaran. Kemudahan para pengguna dalam mengakses informasi menggunakan media sosial memudahkan pemasar untuk berkomunikasi dengan konsumen. Perkembangan ini menjadi pemicu para pengguna yang semakin banyak. (Puntoadi 2011)

### Twitter

Twitter merupakan situs jejaring sosial yang masih populer hingga saat ini. Twitter dapat membuat sebuah *blog* yang berukuran kecil. *Blog* tersebut dapat digunakan oleh berbagai pengguna yang telah mendaftar pada aplikasi Twitter. Twitter didirikan oleh Jack Dorsey pada bulan Maret 2006. Twitter membatasi kata yang akan dipost sebanyak 140 karakter. Namun, tidak hanya tulisan media lain dapat diunggah ke jejaring sosial tersebut, seperti foto, video, *url* dan lain lain. Twitter banyak digunakan untuk berbagi informasi, menjalin relasi bisnis, menuangkan isi hati dan pikiran dalam bentuk tulisan.

- Hak Cipta Dilindungi Undang-Undang  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Pada aplikasi twitter disediakan sebuah *search engine* yang dapat digunakan oleh pengguna twitter. Dengan *search engine* tersebut kita bisa mendapatkan informasi terkait *tweets* yang ada pada aplikasi twitter. Informasi banyak digunakan untuk mencari wawasan terkait permasalahan yang sedang dihadapi. Informasi yang diberikan dapat berupa *tweets*, pengguna yang menuliskan *tweets* tersebut, lokasi *tweets* tersebut, dan lain-lain. (Adiyana dan Hakim 2015).

### Penambangan Teks

Penambangan teks merupakan salah satu variasi dari *data mining* yaitu menggali pola atau informasi penting dari data berupa teks. Penambangan teks melibatkan proses *data mining*, *information retrieval*, *machine learning*, statistika dan *computational linguistics*. *Information retrieval* atau temu kembali informasi adalah ilmu tentang pencarian dokumen atau informasi di dalam dokumen. Dokumen bisa berupa teks atau multimedia (Han *et al.* 2011).

Penambangan teks melibatkan proses penataan teks *input* (*parsing*, bersama dengan penambahan beberapa fitur linguistik turunan dan penghilangan beberapa diantaranya, dan penyisipan *subsequent* ke dalam *database*), menentukan pola dalam data terstruktur, dan mengevaluasi serta menginterpretasikan output. Proses Penambangan teks yang khas meliputi responsasi teks, *text clustering*, ekstraksi konsep/entitas, produksi taksonomi granular, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas (yaitu, pembelajaran hubungan antara entitas bernama).

Untuk melakukan pengresponan pada data teks yang besar membutuhkan otomatisasi. Penelitian yang dilakukan oleh Chourdhury (2012) mampu mengresponkan pengguna twitter secara otomatis.

### Awan Kata

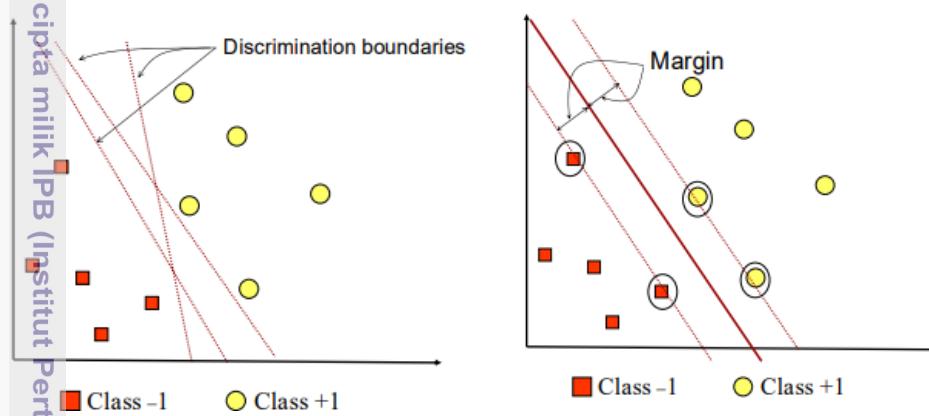
Awan kata adalah kumpulan kata-kata yang paling banyak muncul dalam data teks yang dianalisis. Kata-kata tersebut terkumpul seperti sebuah gumpalan awan yang berisi kata-kata sehingga disebut awan kata. Intensitas keseringan kata yang digunakan, ditunjukkan dengan ukuran huruf pada kata. Semakin besar huruf dari kata yang terdapat di awan kata menunjukkan semakin sering kata tersebut muncul. Tampilan awan kata ini lebih menarik serta cepat untuk menemukan kata-kata yang sering muncul, akan tetapi awan kata memiliki kekurangan yaitu tidak dapat menunjukkan frekuensi kata-kata muncul dalam suatu teks yang dianalisis (Heimerl 2014).

Awan kata memiliki salah satu dari dua respon yang akan difokuskan, yaitu kemudahan serta wawasan dan peningkatan serta perluasan. Kemudahan serta wawasan dapat dilakukan hanya dengan menghitung frekuensi kata kemudian memvisualisasi dalam bentuk awan kata. Kemudahan serta wawasan yang diambil merupakan analisis yang umum untuk digunakan.

Pada penelitian Sinclair dan Cardew-Hall (2008) membandingkan awan kata dengan sebuah *user interface* yang sederhana pada sebuah kotak pencarian. Awan kata mampu memberikan suatu gambaran data dengan kesan yang baik. Serta mampu untuk diinterpretasi lebih mudah dan memberikan informasi yang bermanfaat.

### **Support Vector Machine**

*Support Vector Machine* adalah sebuah metode klasifikasi yang dikembangkan oleh ilmu komputer pada tahun 1990. Metode ini cukup populer digunakan hingga saat ini. *Support Vector Machine* memberikan performa yang cukup baik, namun tidak selalu baik pada semua dataset.



Gambar 1 Ilustrasi pembentukan *hyperplane* dengan memaksimumkan margin

Konsep yang digunakan oleh *Support Vector Machine* adalah memaksimumkan margin pada metode klasifikasi. Metode ini memanfaatkan *hyperplane* untuk memisahkan kelas data respon. Apabila data memiliki  $p$ -ukuran dimensi akan membuat *hyperplane* yang berukuran  $p - 1$  dimensi. Misalkan diperoleh data yang memiliki peubah penjelas  $x_1$  dan  $x_2$  maka kita akan membentuk *hyperplane* berukuran 1 dimensi yang akan memisahkan  $x_1$  dan  $x_2$  menjadi 2 kelas yang berbeda. Ilustrasi pembentukan *hyperplane* dengan memaksimumkan margin terdapat pada Gambar 1. (Friedman J et al. 2001)

Misalkan terdapat  $y_1, y_2, \dots, y_n \in \{-1, 1\}$  maka untuk membentuk *hyperplane* sebagai berikut.

$$\{x : f(x) = x^T \beta + \beta_0 = 0\}, \quad (1)$$

Dimana  $\beta$  adalah vektor satuan:  $\|\beta\| = 1$ .

Untuk memaksimumkan margin maka akan digunakan

$$\max_{\beta, \beta_0, \|\beta\|=1} M \\ y_i(x_i^T \beta + \beta_0) \geq M, i = 1, 2, 3, \dots, N \quad (2)$$

dimana,

$$M = \frac{2}{\|\beta\|} \quad (3)$$

Sehingga akan dicari

$$\begin{aligned} & \min_{\beta, \beta_0} \frac{1}{2} \|\beta\|^2 \\ & \text{st. } y_i(\beta^T x_i + \beta_0) \geq 1, \text{ untuk } i = 1, 2, 3, \dots, N \end{aligned} \quad (4)$$

Dengan menggunakan fungsi *lagrange* yang meminimumkan  $\frac{1}{2} \|\beta\|^2$  diperoleh fungsi *lagrange* akhir

$$L_P = \frac{1}{2} \|\beta\|^2 - \sum_{i=1}^N \alpha_i [y_i(\beta^T x_i + b) - 1] \quad (5)$$

Dengan menggunakan turunan diperoleh

$$\beta = \sum_{i=1}^N \alpha_i y_i x_i \quad (6)$$

$$0 = \sum_{i=1}^N \alpha_i y_i \quad (7)$$

Akan diterapkan optimisasi menggunakan

$$\min_{\beta, \beta_0, \xi_i} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i \quad (8)$$

$$\text{st. } y_i(\beta^T x_i + \beta_0) \geq 1 - \xi_i, \xi_i \geq 0, \text{ untuk } i = 1, 2, 3, \dots, N$$

Dengan C merupakan cost parameter dengan ketentuan:

Semakin kecil C maka margin akan semakin lebar

Semakin besar C maka margin akan semakin sempit

C merupakan *hyperparameter*

## Regresi Logistik

Regresi logistik adalah salah satu metode klasifikasi yang dibangun dengan menggunakan *odds ratio* terhadap masing-masing kelas respon ke-i terhadap kelas respon ke-K. Model menggunakan fungsi linear dari peubah penjelas (x) terhadap *odds ratio* kelas respon. Bentuk fungsi regresi logistik adalah sebagai berikut: (Friedman *et al.* 2001)

$$\begin{aligned} \log \frac{\Pr(G = 1|X = x)}{\Pr(G = K|X = x)} &= \beta_{10} + \beta_1^T x \\ \log \frac{\Pr(G = 2|X = x)}{\Pr(G = K|X = x)} &= \beta_{20} + \beta_2^T x \\ &\vdots \\ \log \frac{\Pr(G = K-1|X = x)}{\Pr(G = K|X = x)} &= \beta_{(K-1)0} + \beta_{K-1}^T x \end{aligned} \quad (9)$$



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Model tersebut memanfaatkan fungsi logit sehingga nilai regresi dari peubah  $X$  berada pada selang 0 sampai 1. Berikut adalah fungsi peluang yang digunakan:

$$\Pr(G = i|X = x) = \frac{\exp(\beta_i + \beta_i^T x)}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}, i = 1, \dots, K-1, \quad (10)$$

$$\Pr(G = K|X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}$$

Misalkan  $\theta = \{\beta_{10}, \beta_1^T, \dots, \beta_{(k-1)0}, \beta_{(k-1)}^T\}$  dan diasumsikan kelas pada peubah respon menyebar *multinomial*. Maka untuk mencari penduga parameter  $\theta$  terbaik, akan digunakan fungsi *likelihood* sebagai berikut dengan jumlah data sebanyak  $N$ .

$$l(x) = \sum_{i=1}^N \log p_{Gi}(x_i|\theta) \quad (11)$$

Jika  $y$  adalah vektor kelas respon ( $y_i$ ), matriks  $X$  dengan ukuran  $N \times (p+1)$  matriks yang berisi vektor  $x_i$ ,  $p$  adalah vektor yang berisi peluang kemunculan kelas target pada data ke-i ( $p(x_i; \underline{\beta^{old}})$ ) dan matriks  $W$  adalah matriks diagonal berukuran  $N \times N$  yang berisi pembobot  $p(x_i; \underline{\beta^{old}})(1 - p(x_i; \underline{\beta^{old}}))$ . Maka diperoleh turunan pertama.

$$\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}} = X^T (\underline{y} - \underline{p}) \quad (12)$$

$$\frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T} = -X^T W X \quad (13)$$

Dengan menggunakan newton raphson diperoleh pendugaan parameter sebagai berikut:

$$\begin{aligned} \underline{\beta^{new}} &= \underline{\beta^{old}} + (X^T W X)^{-1} X^T (\underline{y} - \underline{p}) \\ &= (X^T W X)^{-1} X^T W (X \underline{\beta^{old}} + W^{-1} (\underline{y} - \underline{p})) \\ &= (X^T W X)^{-1} X^T W z \end{aligned} \quad (14)$$

Dengan  $z = X \underline{\beta^{old}} + W^{-1} (\underline{y} - \underline{p})$ .

## Bayes Naif

Bayes naif merupakan suatu metode klasifikasi yang populer akhir-akhir ini dengan performa yang cukup baik. (McCallum 1998). Metode ini menggunakan aturan *Bayesian*. *Bayesian Classification* merupakan klasifikasi secara statistik dengan memprediksi suatu data terprediksi ke dalam kelas tertentu (Han et al. 2012). Berikut adalah aturan bayesian yang akan digunakan. Misalkan label adalah suatu

kelas target kemudian peubah adalah peubah penjelas yang menjelaskan peubah target. N adalah banyaknya peubah penjelas yang ada pada data.

$$P(label|peubah) = \frac{P(label) * P(peubah|label)}{P(peubah)} \quad (15)$$

Dengan menggunakan asumsi bahwa setiap peubah merupakan peubah yang saling bebas. Menghasilkan persamaan:

$$P(label|peubah) \propto P(label) \prod_{i=1}^N P(peubah_i|label) \quad (16)$$

Sehingga persamaan pada asumsi di atas menghasilkan.

$$P(label|peubah) = \frac{1}{Z} P(label) \prod_{i=1}^N P(peubah_i|label) \quad (17)$$

Dengan  $Z = P(peubah) = \sum_{i=1}^k P(peubah|label_i)P(label_i)$  sehingga Z merupakan konstanta yang nilainya tidak akan berubah. Dimana k adalah banyaknya kelas yang ada pada peubah target. Untuk membangun model klasifikasi maka persamaan akhir yang akan digunakan adalah

$$\hat{y} = argmax_k P(C_k) \prod_{i=1}^N P(peubah_i|label) \quad (18)$$

## Hutan Acak

Hutan acak merupakan metode *ensemble* dengan menggunakan *decision tree*. Hutan acak memanfaatkan metode *ensemble* untuk mendapatkan akurasi yang tinggi dengan ragam yang rendah. Hutan acak merupakan pengembangan dari metode *ensemble bagging* karena pada metode *bagging* antar pohon yang dibangun memiliki kemiripan, namun pada metode hutan acak akan dibangun pohon yang berbeda pada setiap iterasi. Hutan acak memiliki algoritma untuk melakukan prediksi sebagai berikut.

Pertama akan dilakukan iterasi sebanyak K kali. Pada setiap iterasi akan dilakukan percontohan berukuran  $n$  pada data latih sebanyak N dan akan dipilih  $m$  peubah dari  $p$  peubah yang ada pada data latih.

Kemudian akan dibangun pohon pada setiap iterasi dengan menggunakan  $p$  peubah dan  $n$  data. Kriteria pembentukan pohon yang dibangun menggunakan *gini index*. Misalkan ada G adalah kelas  $k$  pada *node-j* maka *gini index* memiliki formula sebagai berikut. (Friedman J et al. 2001)

$$Gini = \sum_{i=1}^G p_{ij}(1 - p_{ij}) \quad (19)$$

Rumus *gini* serupa dengan menghitung ragam pada sebaran *bernoulli*. Semakin besar *gini* maka semakin heterogen kelas yang ada pada *node* tertentu semakin kecil nilai *gini*. Sehingga nilai *gini* yang baik adalah nilai *gini* yang lebih kecil.



Untuk melakukan prediksi dengan menggunakan hutan acak maka setiap pohon ke- $k$  akan melakukan prediksi. Kemudian hasil akan menggunakan *majority vote* pada hasil prediksi seluruh pohon.

### Multilayer Perceptron

*Multilayer perceptron* adalah nama lain dari *feedforward neural networks*. *Multilayer perceptron* merupakan salah satu pemodelan *deep learning* yang sering digunakan untuk menangani data yang tidak terstruktur. Tujuan dari *multilayer perceptron* adalah untuk menduga fungsi  $\hat{f}$ . Untuk memodelkan klasifikasi maka *multilayer perceptron* akan menghasilkan fungsi dugaan terhadap respon yang mendekati fungsi sebenarnya dengan parameter tertentu. (Goodfellow I *et al.* 2016)

$$G \approx \hat{f}(x; w) \quad (20)$$

*Multilayer perceptron* memodelkan dengan konsep menyusun beberapa fungsi berbeda secara bersamaan. Hal ini dapat dilakukan dengan menggunakan *directed acyclic graph*. Misalkan fungsi  $f_1, f_2$ , dan  $f_3$  akan terhubung seperti rantai dengan bentuk  $f(x) = f_3(f_2(f_1(x)))$ . Pada kasus ini  $f_1$  dapat disebut *first layer*,  $f_2$  *second layer* dan selanjutnya. Lapisan terakhir disebut dengan *output layer*. Dikarenakan setiap lapisan tidak dapat memperlihatkan keluarannya. Maka lapisan ini disebut *hidden layer*.

Pada *multilayer perceptron* membutuhkan sebuah fungsi aktivasi. Salah satu fungsi aktivasi yang akan digunakan adalah *RELU* dengan fungsi sebagai berikut  $g(z) = \max\{0, z\}$ . Kemudian untuk mendapatkan nilai-prediksi adalah dengan menggunakan  $\hat{y} = \underset{i \in 1, \dots, N}{\operatorname{argmax}} g(z)$ .

Setelah mendapatkan hasil prediksi dibutuhkan *cost function*. *Cost function* yang akan digunakan adalah *cross entropy loss*.

$$l(w) = - \sum_{i \in 1, \dots, N} y_i \log(g(z)) \quad (21)$$

Untuk menentukan bobot pada *multilayer perceptron* dapat menggunakan *Adam Optimizer* sebagai. Dengan menggunakan *cross entropy* sebagai *cost function* maka *Adam Optimizer* akan mendapatkan bobot akhir yaitu bobot yang konvergen.

### Ukuran Kebaikan Model

Model klasifikasi dibangun diharapkan mampu memprediksi *dataset* dengan benar. Namun, kenyataannya tidak semua model dapat melakukan klasifikasi dengan benar. Oleh karena itu, pemodelan klasifikasi memerlukan ukuran kebaikan model. Salah satu alat untuk mengukur kebaikan model adalah tabel klasifikasi (Choudhury *et al.* 2012). Tabel klasifikasi adalah matriks yang digunakan untuk melihat bagaimana model mampu menghasilkan target dugaan yang dibandingkan dengan target sebenarnya. Ilustrasi tabel klasifikasi dapat dilihat pada Tabel 1.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Tabel 1 Ilustrasi tabel klasifikasi

| Kelas<br>Aktual | Kelas Faktual |         |         |         |
|-----------------|---------------|---------|---------|---------|
|                 | Kelas=0       | Kelas=1 | Kelas=2 | Kelas=3 |
| Kelas=0         | a             | b       | c       | d       |
| Kelas=1         | e             | f       | g       | h       |
| Kelas=2         | i             | j       | k       | l       |
| Kelas=3         | m             | n       | o       | p       |

Hasil dari tabel klasifikasi dapat digunakan untuk menghitung akurasi. Akurasi adalah banyaknya prediksi benar pada semua data yang diprediksi. Akurasi merupakan metode ukuran kebaikan metode yang umum digunakan untuk pemodelan klasifikasi.

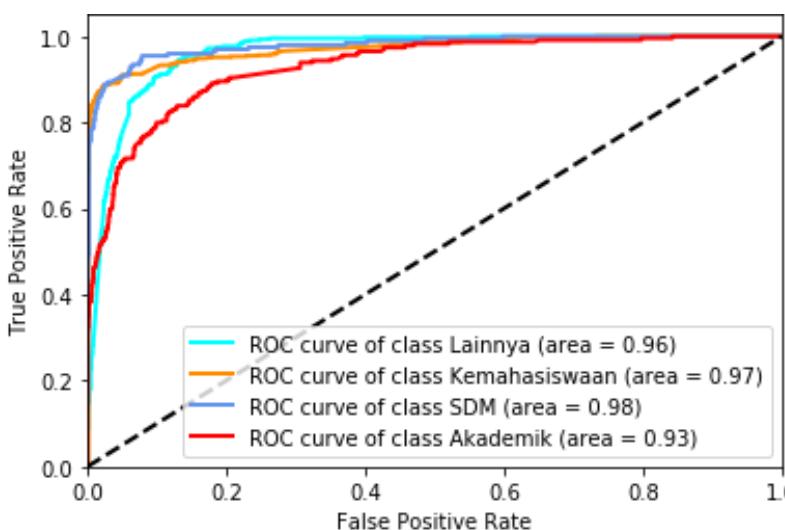
$$Akurasi = \frac{a + f + k + p}{a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p} \quad (22)$$

Selain menggunakan akurasi, ukuran kebaikan model dapat dilihat dengan menggunakan grafik *receiver operating characteristic* (ROC). Dimana grafik ROC menampilkan 2 tipe sisaan untuk setiap *thresholds*. Grafik ROC dapat menghasilkan informasi performa sebuah model dengan menggunakan nilai *area under ROC curve* (AUC). AUC yang lebih besar menandakan performa model klasifikasi yang baik.

ROC dibentuk dengan menempatkan peluang terprediksi positif benar sebagai sumbu-Y dengan peluang terprediksi negatif salah sebagai sumbu-X. Apabila telah diperoleh maka akan dihitung auc dengan rumus sebagai berikut:

$$A = \int_{-\infty}^{\infty} TPR(T) FPR'(T) dT \quad (23)$$

dengan,  $TPR(T) = \int_T^{\infty} f_1(x) dx$  dan  $FPR(T) = \int_T^{\infty} f_0(x) dx$ . Ilustrasi grafik ROC dengan AUC berada pada Gambar 2.



Gambar 2 Ilustrasi grafik ROC dengan AUC pada masing masing kelas

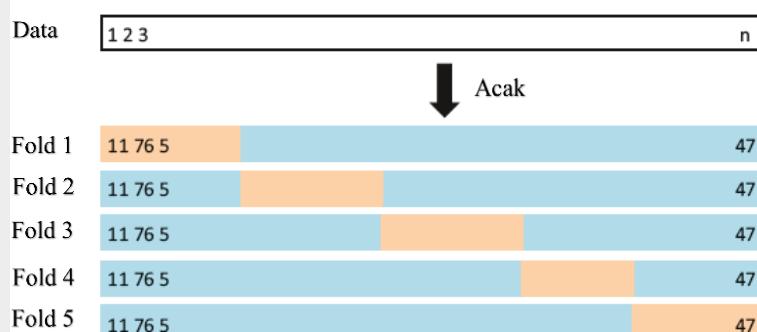
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## K-Lipat Validasi Silang

K-lipat validasi silang merupakan metode yang membagi himpunan contoh secara acak menjadi K himpunan bagian. Dimana pada bagian ke-i menjadi data uji kemudian bagian lainnya menjadi data latih. Kemudian berlangsung hingga  $i=k$ . Dalam penelitian ini, metode yang digunakan dalam proses pemisahan data latih dan data uji adalah metode k-lipat validasi silang dengan nilai  $k=10$ . Karena menurut Han *et al.* (2012) k-lipat validasi silang dengan  $k = 10$  merupakan pembagian himpunan yang baik dalam penentuan nilai K. Berikut adalah ilustrasi k-lipat validasi silang.

$$\text{Evaluasi} = \frac{1}{k} \sum_{i=1}^k \text{Akurasi}_i \quad (8)$$

Gambar 3 merupakan ilustrasi k-lipat validasi silang. Kolom yang berwarna jingga menjadi data uji yang digunakan untuk menguji kebaikan model. Begitupun kolom yang berwarna biru adalah data latih yang digunakan untuk melatih model. Kemudian perhitungan evaluasi akhir yang akan digunakan pada k-lipat validasi silang menjadi rata-rata akurasi pada setiap bagian data uji.



Gambar 3 Ilustrasi k-lipat validasi silang dengan  $k=5$

## Geovisualisasi

Geovisualisasi merupakan analisis data spasial yang digunakan dengan memanfaatkan informasi lokasi. Informasi yang digunakan dapat berupa *longitude* dan *latitude*. Dengan memanfaatkan informasi tersebut dapat diketahui karakteristik daerah berdasarkan lokasi yang ditentukan. Visualisasi ini menggunakan *point pattern*, *line pattern* dan *area pattern* (Yasobant *et al.* 2015).

Namun penelitian ini menggunakan file data berbentuk kelas SpatialPolygonsDataframe dengan ekstensi file berbentuk .rds. Kelas file tersebut dapat diakses menggunakan aplikasi R dengan library *sp*. Library *sp* merupakan salah satu library yang berfungsi untuk mengolah data dengan kelas dan metode yang berbentuk spasial. Library ini dikembangkan terakhir dengan versi 1.2-7 pada tanggal 19 Januari 2018 oleh Edzer Pebesma bersama timnya.

File data berbentuk kelas SpatialPolygonsDataframe sudah memiliki beberapa objek seperti data, *polygons*, *plotOrder*, *bbox*, dan *proj4string*. Objek

tersebut terdapat fungsi-fungsi yang saling terintegrasi dalam membentuk geovisualisasi. *File* tersebut memiliki kemudahan dalam melakukan visualisasi data.

## METODE

### Data

Data diperoleh dengan proses penarikan data di twitter dengan menggunakan R. Penarikan data dilakukan dari tanggal 14 Januari 2018 hingga 27 Februari 2018 di Bogor. Penarikan data dilakukan dengan menggunakan kata kunci yang berkaitan dengan 10 universitas terbaik menurut Kemenristek Dikti (2017).

Dalam melakukan penarikan data akan digunakan *longitude* sebesar 119.503407 dan *latitude* sebesar -3.648612 dengan radius 3000 KM. Data yang diperoleh sebanyak 226 192 *tweets*.

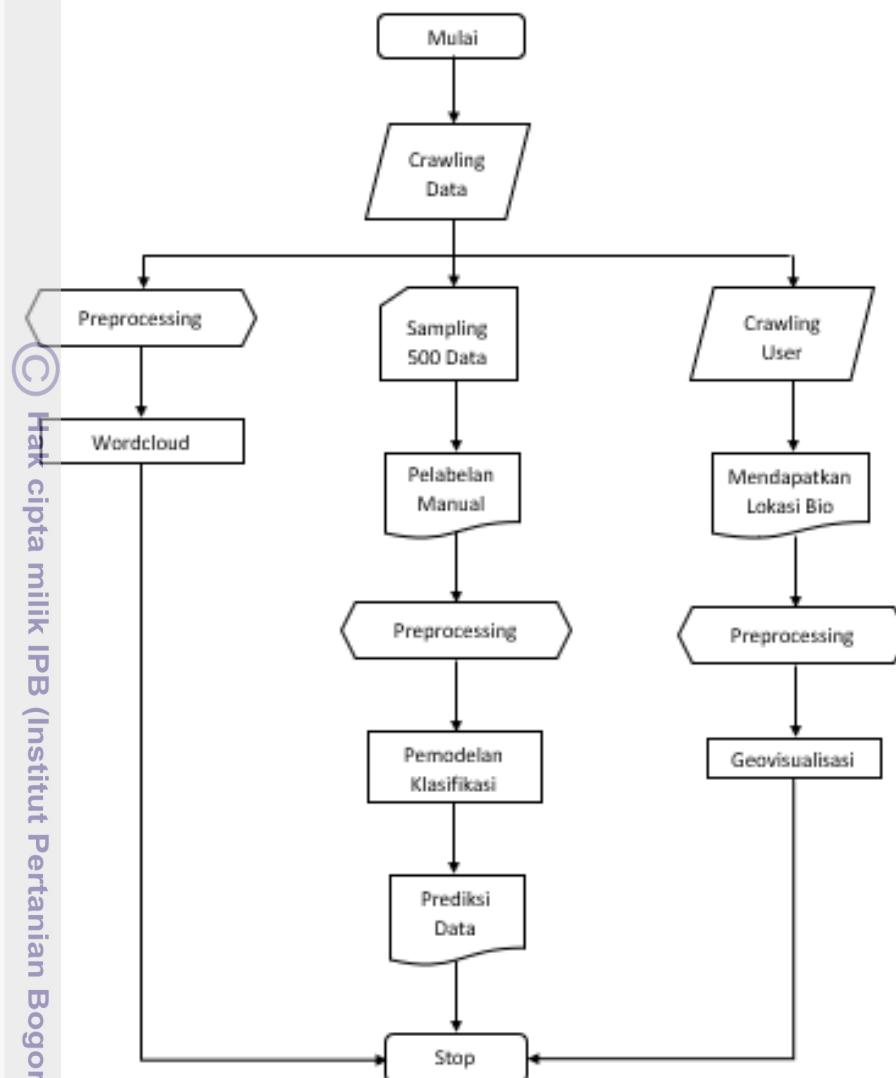
### Prosedur Analisis

Luaran pokok yang akan dihasilkan oleh penelitian ini ada 3 luaran, yaitu: awan kata, geovisualisasi IPB, dan prediksi tweets yang belum terlabelkan dengan respon berdasarkan topik ujaran dan respon berdasarkan sentimen ujaran. Oleh karena itu, terdapat tahapan analisis untuk masing-masing luaran.

Tahapan awal sebelum melakukan analisis adalah pengambilan data. Teknik pengambilan data yang digunakan adalah teknik penarikan data. Teknik penarikan data ini membutuhkan kata kunci yang harus dipersiapkan terlebih dahulu. Kata kunci tersebut digunakan sebagai kunci dalam pengambilan data twitter agar *tweets* yang diperoleh relevan dengan topik yang dibahas.

Tahapan selanjutnya adalah membentuk luaran pertama yaitu awan kata. Tahapan ini membutuhkan prapemrosesan kemudian membentuk *document-term matrix*. Setelah itu, akan dibentuk tabel frekuensi untuk menghitung kemunculan kata pada masing-masing universitas. Kemunculan kata yang dibentuk menggunakan awan kata diharapkan dapat merepresentasikan topik pembicaraan untuk masing-masing universitas.

Tahapan selanjutnya adalah membentuk luaran kedua yaitu geovisualisasi. Tahapan awal untuk membentuk geovisualisasi adalah menentukan daftar *id user* unik pada universitas IPB yang pernah membuat *tweets*. Setelah mengetahui daftar *id user*, maka daftar *id user* digunakan untuk memperoleh profil biodata *user* dengan menggunakan penarikan data terhadap *id user*. Profil tersebut memiliki lokasi yang dicantumkan oleh *user*. Kemudian lokasi tersebut diambil dan dilakukan pembersihan agar sesuai dengan provinsi yang ada di Indonesia. Setelah itu, banyaknya *tweets* tersebut dipetakan sesuai dengan kemunculan di provinsi tersebut.



Gambar 4 Diagram alir tahapan analisis

Tahapan selanjutnya adalah membentuk luaran ketiga yaitu tahapan membentuk model prediksi dan melakukan prediksi terhadap *tweets* yang belum memiliki label. Proses ini dimulai dengan cara melakukan tahap pemrosesan sehingga membentuk *document-term matrix*. Kemudian matriks tersebut digunakan untuk membangun pemodelan klasifikasi dengan respon berdasarkan topik ujaran dan respon berdasarkan sentimen ujaran. Setelah mendapatkan model terbaik maka model tersebut digunakan untuk melakukan prediksi terhadap *tweets* yang belum memiliki label. Kemudian hasil dari label tersebut dianalisis menggunakan pengujian proporsi. Diagram alir untuk tahapan prosedur analisis dapat dilihat pada Gambar 4. Tahapan prosedur analisis dibahas lebih lengkap pada subbab di bawah.

### Penentuan Kata Kunci

Kata kunci adalah kata kunci yang digunakan untuk mengambil data dari twitter. Misalkan, kata kunci yang digunakan adalah "IPB" maka apabila sebuah

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a.

b.

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

#### Hak Cipta Dilindungi Undang-Undang

*tweets* memiliki kata yang berisi “IPB” maka *tweets* tersebut akan terambil oleh sistem. Penentuan kata kunci merupakan suatu langkah yang penting untuk mengambil data di twitter. Penentuan kata kunci yang tidak tepat dapat memberikan *tweets* yang tidak relevan dengan permasalahan didapatkan. Sehingga dibutuhkan penentuan kata kunci yang tepat agar *tweets* yang diperoleh relevan dengan topik permasalahan.

Pertimbangan yang digunakan dalam menentukan kata kunci adalah sebagai berikut :

1. Singkatan universitas
2. Singkatan universitas yang memiliki kemiripan dengan konteks lain ditambahkan kata yang dan kampus
3. Kepanjangan universitas
4. Akun resmi universitas

Menggunakan kata kunci singkatan universitas diharapkan untuk mendapatkan *tweets* yang membicarakan suatu universitas dengan menyebutkan singkatannya saja. Namun, terdapat singkatan universitas yang memiliki kesamaan secara tulisan namun berbeda secara makna. Misalkan, kata “ui” akan memiliki banyak 2 makna, seperti *user interface* dan universitas indonesia. Oleh karena itu, untuk kata kata yang memiliki kemungkinan tercampur maknanya maka ditambahkan kata “yang” dengan harapan *tweets* yang terpilih berbahasa indonesia. Adapun untuk kata “kampus” diharapkan *tweets* yang terpilih memiliki hubungan antara singkatan dengan kampus. Kepanjangan universitas digunakan untuk mendapatkan *tweets* yang menyebutkan kepanjangan universitas dalam *tweets*-nya. Kemudian, akun resmi universitas diharapkan *tweets* yang didapatkan membicarakan universitas yang bersangkutan melalui *mention* ke akun resmi universitas tersebut. Daftar kata kunci dapat dilihat pada Lampiran 1.

#### Penarikan Data

Pengambilan data *tweet* dilakukan dengan cara mengambil data pada data twitter. Pengambilan data tersebut membutuhkan perizinan dari pihak twitter. Perizinan tersebut dapat dilakukan dengan cara mendaftarkan aplikasi menggunakan akun twitter. Pendaftaran dapat dilakukan pada link berikut <https://apps.twitter.com/>.

Pendaftaran aplikasi digunakan untuk mendapatkan kode berupa *consumer key*, *consumer secret*, *access token*, dan *access secret*. Keempat kode tersebut akan digunakan untuk melakukan *setup twitter oauth*. Sehingga dapat diakses untuk melakukan penarikan data terhadap data twitter.

Tabel 2 Contoh *tweets* yang telah diambil

| No | Tweets   |
|----|--|
| 1  | RT @humasbpblbatam: Kegiatan Mahasiswa/i magang IPB, UBRI, UNIV.BORNEO TARAKAN dan SMKN 4 BATAM sedang melakukan pemberian pakan ikan di KJ...<br>2 Bikin Roti Cane Dari Ampas Tahu, Mahasiswi IPB Ini Tersohor di Malaysia – Berita Terbaru <a href="https://t.co/zaTpQkHD7w">https://t.co/zaTpQkHD7w</a> |
| 3  | Hebat, Doktor IPB Ciptakan Drone Kargo untuk Angkut Barang <a href="https://t.co/GW8ncskqva">https://t.co/GW8ncskqva</a>   |



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Setelah mendapatkan akses untuk melakukan penarikan data, setelan untuk melakukan penarikan data diterapkan *longitude* sebesar 119.503407 dan *latitude* sebesar -3.648612 dengan *radius* 3 000 KM. Kemudian, penarikan data dilakukan berdasarkan setiap kata kunci yang telah ditentukan. Proses penarikan data dapat berlangsung sekitar 4 – 5 jam. Proses tersebut tergantung dengan kecepatan internet yang digunakan dan aturan yang telah ditentukan oleh pihak twitter. Ilustrasi data yang telah diambil dapat dilihat pada Tabel 2.

### Percontohan

Setelah diperoleh data sebanyak 226 192 *tweets*. Data *tweets* tersebut dilakukan pengambilan contoh sebanyak 5 000 data untuk dilakukan pelabelan terhadap data. Teknik percontohan yang digunakan adalah penarikan contoh acak sederhana. Kemudian dilakukan pemodelan klasifikasi terhadap 5 000 data tersebut.

### Pelabelan Data

Berdasarkan data yang telah dilakukan penarikan contoh diberikan label secara manual terhadap 5000 data. (De Choudhury *et al.* 2012) Label yang diberikan berupa 4 kelas yaitu Mawa untuk *tweets* yang mengandung topik kemahasiswaan yang mengacu pada ketentuan dari kemenristek dikt mengenai fungsi direktorat kemahasiswaan, Akademik untuk *tweets* yang mengandung topik program akademik mengacu pada Akreditasi Institusi Perguruan Tinggi Baru oleh BAN-PT (2017), SDM untuk *tweets* yang mengandung topik sumber daya manusia pada perguruan tinggi tersebut pelabelan mengacu kepada Akreditasi Institusi Perguruan Tinggi Baru oleh BAN-PT (2017) , dan Lainnya untuk label yang tidak relevan dengan 3 respon di atas.

Tabel 3 Ilustrasi *tweets* yang telah diberikan label

| No | Tweets   | Label |
|----|--|-------|
| 1  | RT @humasbpblbatam: Kegiatan Mahasiswa/i magang IPB, Mawa UBRI, UNIV.BORNEO TARAKAN dan SMKN 4 BATAM sedang melakukan pemberian pakan ikan di KJ...    | Mawa  |
| 2  | Bikin Roti Cane Dari Ampas Tahu, Mahasiswi IPB Ini Tersohor di Malaysia – Berita Terbaru <a href="https://t.co/zaTpQkHD7w">https://t.co/zaTpQkHD7w</a> | Mawa  |
| 3  | Hebat, Dosen IPB Ciptakan Drone Kargo untuk Angkut Barang <a href="https://t.co/GW8ncskqva">https://t.co/GW8ncskqva</a>                                | SDM   |

Sedangkan untuk sentimen positif, negatif dan netral dilakukan pelabelan secara manual dengan ketentuan sebagai berikut. Sentimen positif adalah perkataan positif yang diungkapkan, seperti kebahagiaan, kegembiraan, kebaikan dan lain-lain. Sentimen negatif adalah perkataan yang memiliki unsur negatif seperti kebencian, kesedihan, kekecewaan, diskriminasi dan lain-lain. Sedangkan untuk netral adalah tweets yang tidak tergolong pada sentimen positif dan negatif, seperti informasi netral, url saja, tweet yang hanya bentuk unicode dan lain-lain. Ilustrasi pelabelan *tweets* setelah dilakukan prapemrosesan dapat dilihat pada Tabel 3.

### Prapemrosesan

Seluruh data dilakukan prapemrosesan terhadap data teks. Teks yang mengandung unsur *url*, *hashtag*, *mention*, bukan alfabet (seperti bilangan, tanda baca, huruf china, huruf jepang dan lain-lain) diganti dengan spasi. Setelah itu, semua huruf kapital diubah menjadi huruf kecil. Kemudian, peneliti menghapus kata-kata pada tweets yang mengandung *stopword*. Pada penelitian ini didapatkan 768 kata *stopword* berdasarkan penelitian Tala (2003). Ilustrasi prapemrosesan dapat dilihat pada Tabel 4.

Tabel 4 Ilustrasi *tweets* yang telah dilakukan prapemrosesan

| N | Tweets   | Label |
|---|--|-------|
| 1 | rt kegiatan mahasiswa i magang ipb ubri univ borneo tarakan smkn 4 batam melakukan pakan ikan kj | Mawa  |
| 2 | bikin roti cane ampas tahu mahasiswa ipb tersohor malaysia berita terbaru                        | Mawa  |
| 3 | hebat dosen ipb ciptakan drone kargo angkut barang   | SDM   |

Setelah dilakukan proses tersebut, teks dilakukan partisi sehingga menjadi bentuk *document term matrix*. Ilustrasi *document term matrix* dapat dilihat pada Tabel 5. *Document term matrix* adalah matriks frekuensi yang memiliki baris berupa dokumen dan kolom berupa kata. Label merupakan peubah respon dan kata yang tedapat pada kolom merupakan peubah penjelasan.

Tabel 5 Ilustrasi bentuk *document term matrix*

| Dokumen | Kata |          |           |     |       |        |        | Label  |
|---------|------|----------|-----------|-----|-------|--------|--------|--------|
|         | rt   | kegiatan | mahasiswa | ... | kargo | angkut | barang |        |
| 1       | 1    | 1        |           | 1   | ...   | 0      | 0      | 0 Mawa |
| 2       | 0    | 0        |           | 1   | ...   | 0      | 0      | 0 Mawa |
| 3       | 0    | 0        |           | 0   | ...   | 1      | 1      | 1 SDM  |

### Awan Kata

Tiap kolom yang berada pada *document term matrix* dijumlahkan terhadap seluruh baris. Sehingga tabel menghasilkan frekuensi pada setiap kata. Kemudian, kata yang memiliki frekuensi tertinggi divisualisasikan membentuk awan. Semakin tinggi frekuensi sebuah kata maka kata tersebut divisualisasikan dengan sangat besar. Semakin rendah frekuensi sebuah kata maka kata tersebut divisualisasikan dengan sangat kecil. Visualisasi akan dilakukan pada setiap universitas.

### Geovisualisasi

Proses untuk melakukan geovisualisasi adalah dengan cara melakukan penarikan data terhadap pengguna yang memasang *tweets* yang telah diperoleh. Kemudian, pengguna diidentifikasi lokasi berdasarkan profil yang dicantumkan pada kolom deskripsi. Setelah itu, lokasi tersebut dilakukan pembersihan data dengan cara mengidentifikasi nama lokasi pada masing-masing profil tersebut menjadi nama provinsi yang ada di Indonesia. Nama lokasi pada masing-masing



16

pengguna sangat beragam ada yang mencantumkan "Bali", "Garut", "Hamba Allah", "dibawah telapak kaki ibu", dan lain-lain. Sehingga perlu dilakukan pembersihan data pada masing-masing pengguna.

Setelah itu, lokasi yang telah dibersihkan dihitung dengan banyaknya *tweets* pada setiap provinsi. Sehingga proses tersebut menghasilkan vektor banyaknya *tweets* dengan ukuran 34 unsur. Vektor tersebut bersesuaian dengan masing-masing nama provinsi yang ada di Indonesia.

Kemudian, vektor tersebut dipanggil *file* peta Indonesia dengan ekstensi .rds yang diunduh pada alamat berikut <http://gadm.org/data.html> dengan *library* *sp*. Setelah itu *file* tersebut dipetakan vektor banyaknya *tweets* dengan *file* peta Indonesia yang telah dipanggil.

### Pemodelan Klasifikasi

Berdasarkan 5 000 data yang telah dilakukan prapemrosesan akan dilakukan pemodelan klasifikasi dengan peubah penjelas adalah *document term matrix* dan peubah respon adalah respon yang berisi topik pembicaraan (Kemahasiswaan, Akademik, SDM dan Lainnya) dan respon yang berisi sentimen (negatif, netral dan positif). Pemodelan menggunakan 10-lipat validasi silang dengan menggunakan beberapa metode klasifikasi. Kemudian akurasi dilihat berdasarkan model yang sedang dibentuk. Model akhir yang digunakan adalah dengan menggunakan 5 000 data. Kemudian akan dilakukan prediksi terhadap tweets yang belum memiliki label pada respon topik dan respon sentimen.

### Prediksi

Model terbaik yang telah dibangun pada proses pemodelan klasifikasi dilakukan prediksi terhadap *tweets* yang belum memiliki label. Kemudian label tersebut akan dibandingkan setiap universitas secara eksploratif dan inferensia. Analisis eksploratif dilakukan dengan cara menampilkan diagram batang yang menunjukkan frekuensi *tweets* yang muncul untuk setiap respon topik dan respon sentimen. Inferensia dilakukan dengan menggunakan pengujian proporsi untuk sentimen positif dan negatif.

## Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

1. Perangkat keras menggunakan laptop dengan spesifikasi:
  - Processor Intel i5-7200U.
  - RAM 12GB.
  - 500GB HDD.
  - 250GB SSD
2. Perangkat lunak yang digunakan yaitu:
  - R 3.4.3 sebagai bahasa pemrograman pada penelitian.
  - Python 3.6 sebagai bahasa pemrograman pada penelitian pada *platform* anaconda.
  - *Library* twitteR untuk mengambil data *tweet* menggunakan API Twitter.
  - *Library* awan kata untuk membentuk awan kata.

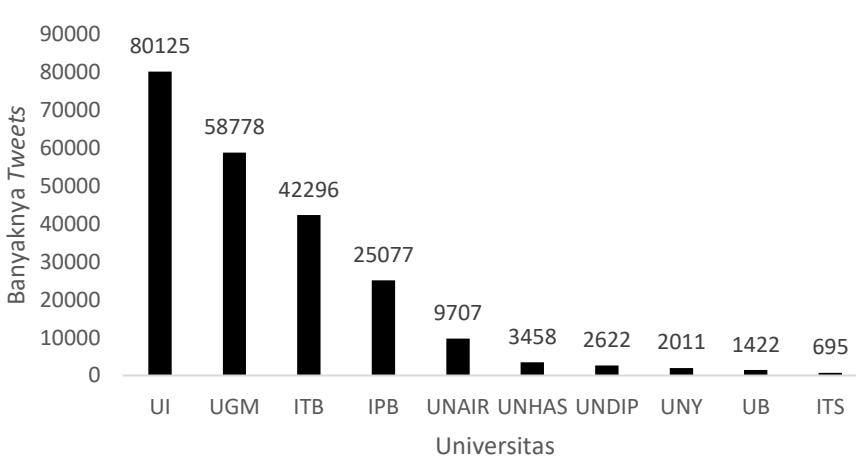
- *Library sp* untuk membentuk geovisualisasi.
- *Library scikit-learn* untuk membangun model klasifikasi.

## HASIL DAN PEMBAHASAN

### Eksplorasi Data

Data *tweet* yang telah diambil dari tanggal 14 Januari 2018 Hingga 27 Februari 2018 menghasilkan *tweet* total sebanyak 226 192 *tweet*. Kemudian, untuk mengetahui frekuensi *tweet* tiap universitas akan dilakukan eksplorasi menggunakan diagram batang. Berikut adalah *diagram batang* yang dihasilkan untuk menggambarkan frekuensi *tweet* setiap universitas di Indonesia.

Diagram batang pada Gambar 5 dapat dilihat Universitas Indonesia memiliki frekuensi yang tertinggi karena UI pernah mengalami suatu kejadian khusus yang membuat UI menjadi bahan pembicaraan di Indonesia. Pembicaraan tersebut terkait kartu kuning Jokowi yang diberikan oleh salah satu Mahasiswa UI bernama Zaadit. Hal tersebut menjadi isu yang cukup meningkatkan frekuensi UI di twitter.



Gambar 5 Banyaknya *tweets* setiap universitas

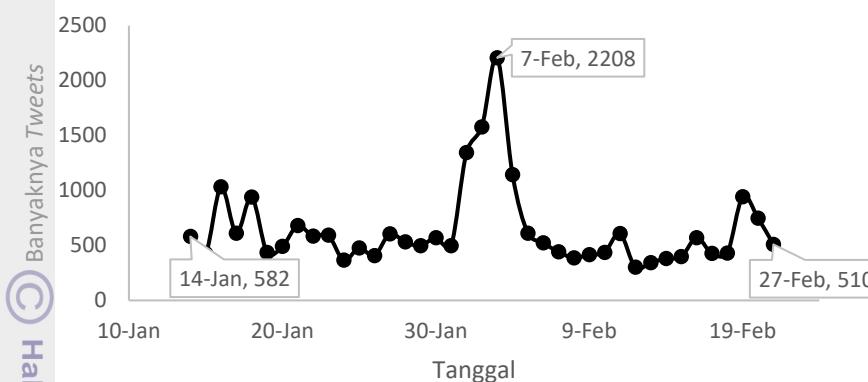
Frekuensi tertinggi kedua diperoleh oleh UGM. Hal tersebut terjadi karena UGM merupakan salah satu universitas terkenal di Indonesia. Namun, terdapat suatu kejadian khusus dimana UGM menjadi isu pembicaraan yang menarik di twitter. Isu tersebut terjadi ketika seluruh mahasiswa setiap universitas hadir pada acara Mata Najwa. Kemudian, presiden mahasiswa UGM mendapatkan banyak sentimen positif karena cara penyampaian dia yang cukup memukau di acara Mata Najwa.

Namun, untuk IPB berada pada peringkat 4. Hal ini terjadi karena pada topik IPB tidak terjadi suatu kejadian khusus yang menarik perhatian pengguna twitter di Indonesia. Namun, karena IPB merupakan universitas yang cukup terkenal di Indonesia sehingga menempatkan IPB di peringkat 4. Setelah mengetahui frekuensi

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

- Hak Cipta Dilindungi Undang-Undang © Hak cipta milik IPB (Institut Pertanian Bogor)**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

*tweet* setiap universitas di Indonesia, diperiksa bagaimana pola dari banyaknya *tweet* setiap hari pada setiap universitas.

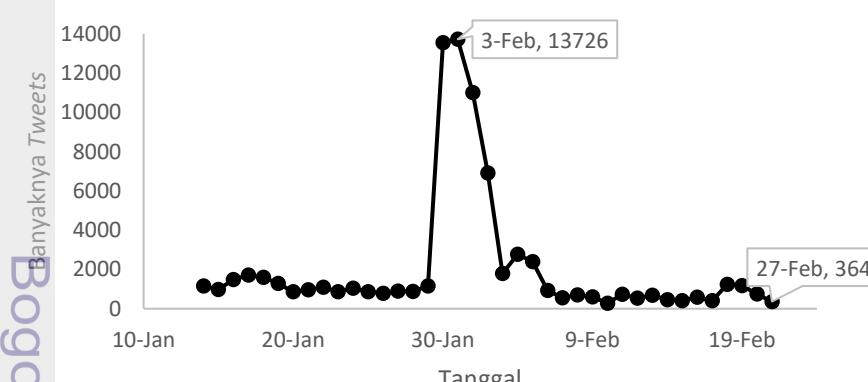


Gambar 6 Frekuensi *tweets* dengan topik IPB setiap hari

Pada Gambar 6 dapat dilihat bahwa frekuensi *tweets* di IPB berada pada sekitar 500 *tweets* perhari. Namun, peningkatan jumlah *tweet* terjadi pada tanggal 4 Februari 2018 sebanyak 1 346 *tweets*. Hal ini disebabkan karena pada tanggal itu sedang hangat isu kartu kuning Jokowi. Hal tersebut cukup berdampak terhadap beberapa universitas di Indonesia khususnya IPB. Karena IPB menjadi salah satu perwakilan universitas di acara Mata Najwa. Sehingga frekuensi *tweet* mengenai IPB menjadi meningkat.

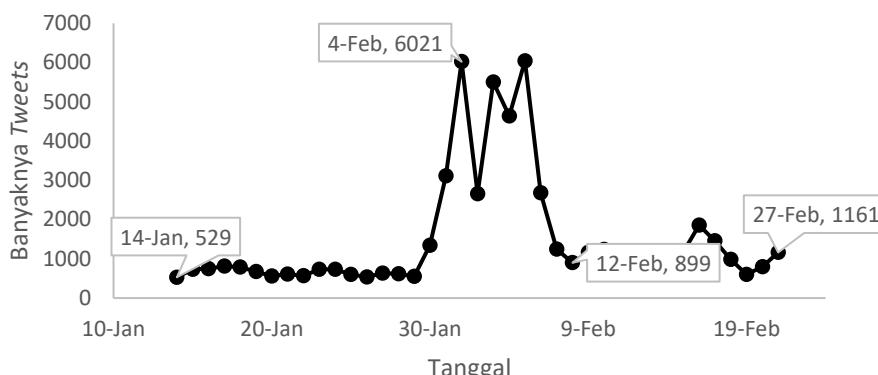
Peningkatan isu kartu kuning jokowi mampu bertahan selama 4 hari dalam peningkatan frekuensi *tweet* mengenai IPB. Namun ketika tanggal 9 Februari frekuensi *tweet* kembali normal.

Gambar 7 mendeskripsikan frekuensi *tweet* dengan topik UI setiap harinya. Dapat dilihat bahwa frekuensi *tweet* UI dalam keadaan normal jarang mencapai lebih dari 2 000. Namun, ketika ada kejadian kartu kuning jokowi UI mengalami peningkatan jumlah *tweet* yang cukup signifikan. Pencapaian *tweet* maksimum ketika kejadian kartu kuning jokowi terjadi pada tanggal 2 Februari 2018 sebanyak 13 553. Isu ini berlangsung selama 8 hari dan kembali normal pada tanggal 10 Februari 2018.



Gambar 7 Frekuensi *tweets* dengan topik UI setiap hari

Dampak isu kartu kuning jokowi ditemukan juga pada topik UGM. Karena, presiden mahasiswa UGM mendapatkan beberapa pujian dikarenakan penyampaian dia yang karismatik di acara Mata Najwa. Grafik 8 adalah grafik frekuensi *tweet* dengan topik UGM setiap hari.



Gambar 8 Frekuensi *tweets* dengan topik UGM setiap hari

Frekuensi *tweet* dengan topik ugm sebelum adanya isu kartu kuning jokowi selalu berada pada di bawah 1 000 *tweets* perhari. Namun setelah adanya isu kartu kuning jokowi, tweets mengenai UGM mengalami peningkatan maksimum hingga 6 041 *tweets* perhari.

Frekuensi *tweets* untuk masing-masing universitas setiap harinya ditampilkan pada Lampiran 1. Pola yang dihasilkan untuk setiap universitas hampir sama. Namun berbeda untuk universitas yang tidak terkait dengan kasus kartu kuning jokowi. Universitas yang tidak terkait isu tersebut adalah UNAIR, UNHAS, UB, ITS, UNDIP, UNY.

### Awan Kata

Awan kata dibentuk dengan menghitung frekuensi kemunculan kata yang ada pada kumpulan *tweets*. Kata yang memiliki frekuensi yang cukup banyak ditampilkan dengan ukuran yang lebih besar. Berikut adalah keluaran awan kata yang dihasilkan dengan menggunakan program R.

Awan kata yang dihasilkan dengan topik IPB sesuai dengan Gambar 9. Kata yang sering muncul dari awan kata di atas adalah kata beasiswa. Hal ini menunjukkan bahwa informasi terkait beasiswa sangat banyak disebutkan di twitter. Begitu pula dengan ketua bem. Hal ini disebabkan karena ketua BEM IPB pernah mengikuti acara mata najwa yang disebabkan kasus kartu kuning jokowi. Sehingga, topik terkait ketua bem menjadi bahan pembicaraan yang cukup menarik oleh masyarakat Indonesia. Kemudian, topik mengenai bank menjadi banyak pembicaraan hal ini dimungkinkan terjadi karena pandangan masyarakat Indonesia mengenai lulusan IPB banyak yang menjadi pegawai bank.

Topik-topik yang dibahas setiap universitas dapat dilihat pada Lampiran 2. Lampiran tersebut berisi tentang awan kata untuk masing-masing universitas. Berdasarkan awan kata tersebut dapat diketahui kata-kata yang sering muncul pada *tweets* mengenai universitas tersebut.



Gambar 9 Awan kata dengan topik IPB

## Geovisualisasi

Untuk mengetahui penyebaran *tweet* mengenai suatu topik pada data twitter dapat menggunakan *library sp* yang ada pada R. Pemetaan dapat menggunakan *longitude* dan *latitude* yang telah disediakan oleh *file* dengan ekstensi .rds. Kemudian banyaknya *tweets* dari setiap lokasi dipetakan sesuai nama lokasi berupa nama provinsi.

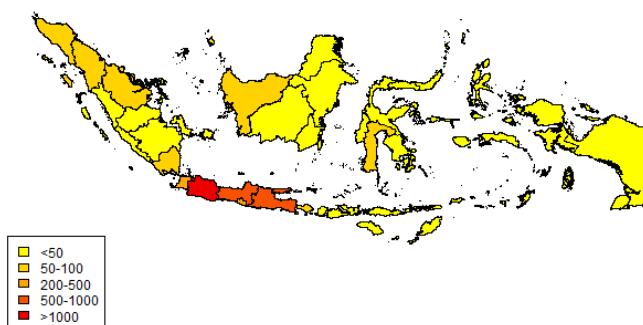
Pemetaan pada Gambar 10 merupakan pemetaan dengan topik IPB. Dari pemetaan tersebut dapat dideskripsikan bahwa provinsi jawa barat dan jakarta merupakan provinsi yang paling banyak membicarakan tentang IPB di twitter. Sedangkan Jawa Tengah dan Jawa Timur merupakan daerah yang cukup banyak membicarakan IPB di twitter. Namun, Sehingga dapat disimpulkan bahwa *tweet* yang membicarakan tentang IPB dominan berasal dari pulau Jawa.

Pulau yang cukup banyak membicarakan tentang IPB setelah pulau jawa adalah pulau Sumatera. Provinsi di pulau sumatera yang membicarakan IPB banyaknya di antara 50-100 adalah provinsi Aceh, Sumatera Utara, Lampung dan Riau. Namun, untuk provinsi Kepulauan Riau, Sumatera Barat, Bangka Belitung dan Jambi banyaknya *tweets* berada di bawah 50 *tweets*.

Pada pulau kalimantan dapat kita lihat bahwa kalimantan barat merupakan provinsi yang paling banyak membicarakan tentang IPB di twitter. Namun, pada pulau sulawesi dapat kita lihat provinsi sulawesi selatan merupakan provinsi yang sering membicarakan IPB di twitter. Sedangkan untuk wilayah timur untuk provinsi NTT, NTB, Maluku Utara, Maluku, Papua, Papua Barat memiliki jumlah *tweets* yang kurang dari 50. Namun untuk provinsi Bali banyaknya *tweets* diantara 50-100.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

PETA BANYAKNYA TWEETS TENTANG IPB

Gambar 10 Peta banyaknya *tweets* tentang IPB

### Persiapan Data

Sebelum melakukan pemodelan klasifikasi, data perlu dipersiapkan agar data mampu untuk diterapkan menggunakan pemodelan klasifikasi. Berdasarkan 5 000 *tweets* yang telah dilakukan penarikan contoh acak terdapat 5 *tweets* yang tidak memiliki *tweets* sama sekali. *Tweets* tersebut perlu untuk dihapus. Sehingga, *Tweets* yang digunakan berjumlah 4 995 *tweets* yang akan digunakan untuk melakukan pemodelan klasifikasi.

*Tweets* tersebut dilakukan prapemrosesan sehingga membentuk *document-term matrix*. *Document-term matrix* adalah matriks yang berisi dokumen sebagai baris dalam kasus ini dokumen adalah *tweets* dan kata sebagai kolom. *Document-term matrix* berisi banyaknya kata yang muncul yang bersesuaian dengan masing-masing *tweets*.

Prapemrosesan dilakukan dengan cara menjadikan huruf menjadi kecil. Selanjutnya menghilangkan URL, bentuk HTML, *mention*, *hash-tags*, dan angka. Serta menghilangkan kata-kata stopwords. Stopword adalah kata-kata yang umum digunakan yang tidak memiliki makna.

*Document-term matrix* yang telah dilakukan prapemrosesan dibentuk sehingga berukuran  $4\ 995 \times 7\ 660$ . Matriks tersebut akan digunakan sebagai prediktor untuk menerapkan pemodelan klasifikasi. Ilustrasi hasil *document-term matrix* dapat dilihat pada Gambar 11.

|       | aa  | aaaaaa | aaero | aamiiin | aapl | ab  | abad | abah | abang | abbiamo | ... | zokudams | zon | zona | zonk | zoo | zu  | zum | zur | zvali | zwelithini |
|-------|-----|--------|-------|---------|------|-----|------|------|-------|---------|-----|----------|-----|------|------|-----|-----|-----|-----|-------|------------|
| 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 | 0.0 | 0.0 | 0.0   |            |
| 1 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 | 0.0 | 0.0   |            |
| 2 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 | 0.0 | 0.0   |            |
| 3 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 | 0.0 | 0.0   |            |
| 4 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 | 0.0 | 0.0   |            |

5 rows x 7660 columns

Gambar 11 Contoh 5 *tweets* hasil prapemrosesan menjadi *document-term matrix*

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pemodelan klasifikasi dibangun untuk memprediksi *tweets* dengan respon yang berisi topik dan respon yang berisi sentimen. Banyaknya kelas pada masing-masing respon dapat dilihat pada Tabel 6 untuk respon topik dan Tabel 7 untuk respon sentimen.

Tabel 6 Banyaknya *tweets* setiap kelas pada respon topik

| No | Kelas         | Banyak Tweets |
|----|---------------|---------------|
| 1  | Kemahasiswaan | 2164          |
| 2  | Akademik      | 696           |
| 3  | SDM           | 670           |
| 4  | Lainnya       | 1465          |

Tabel 7 Banyaknya tweets setiap kelas pada respon sentimen

| No | Kelas   | Banyak Tweets |
|----|---------|---------------|
| 1  | Positif | 1375          |
| 2  | Netral  | 2145          |
| 3  | Negatif | 1475          |

## Regresi Logistik

Pemodelan dengan menggunakan regresi logistik dilakukan dengan menggunakan *document-term matrix*. *Document-term matrix* dibentuk berdasarkan tahapan pada subbab persiapan data. Regresi logistik diterapkan terhadap respon yang berisi topik ujaraan dan respon yang berisi sentimen ujaran.

Pemodelan regresi logistik menggunakan 10-lipat validasi silang untuk mengetahui tingkat akurasi dan kestabilan model. Setelah menggunakan 10-lipat validasi silang, evaluasi pemodelan regresi logistik dilakukan dengan melihat grafik ROC. Dengan menggunakan grafik ROC model dapat dilihat ketepatan memprediksi terhadap masing-masing kelas.

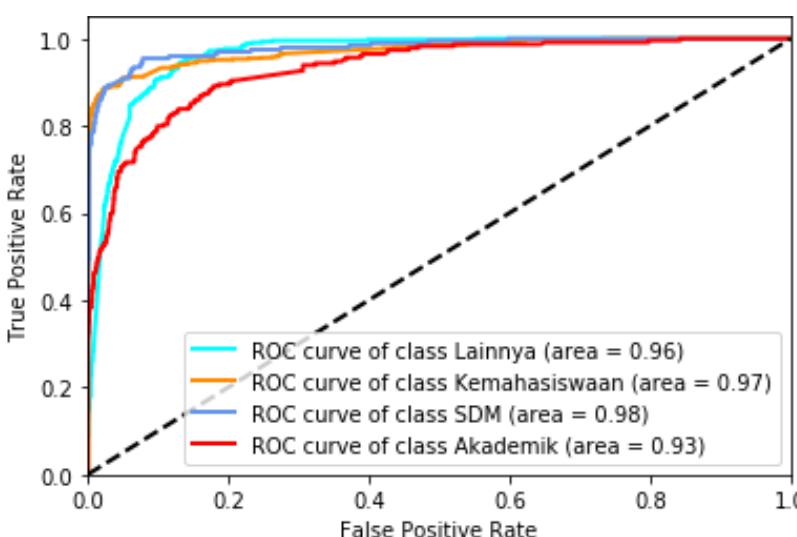
Rata-rata akurasi metode regresi logistik dengan menggunakan 10-lipat validasi silang memiliki nilai yang cukup baik pada masing-masing respon. Rata-rata akurasi tersebut bernilai 86.29% dengan simpangan baku 1.23% untuk respon topik. Sementara itu, untuk respon sentimen rata-rata akurasi bernilai sebesar 84.56% dengan simpangan baku 1.36% untuk respon sentimen. Rata-rata akurasi untuk respon sentimen lebih kecil dibandingkan dengan rata-rata akurasi untuk respon topik. Hal ini menunjukkan bahwa respon sentimen lebih sulit untuk diprediksi dengan baik dibandingkan dengan respon topik dengan menggunakan regresi logistik.

Pada respon topik, regresi logistik mendapatkan performa terbaik dengan lipatan ke-8 yaitu dengan akurasi sebesar 89%. Akurasi ini cukup baik karena akurasi mendekati 90%. Begitupun dengan respon sentimen, performa terbaik diperoleh pada lipatan ke-8 dengan akurasi sebesar 87%. Akurasi setiap lipatan secara lengkap dapat dilihat pada Tabel 8.

Tabel 8 Akurasi setiap lipatan dengan menggunakan metode regresi logistik (dalam persentase)

| Respon   | Lipatan |    |    |    |    |    |    |    |    |    | Rata-rata |
|----------|---------|----|----|----|----|----|----|----|----|----|-----------|
|          | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |           |
| Topik    | 86      | 86 | 86 | 86 | 87 | 85 | 85 | 89 | 88 | 85 | 86.29     |
| Sentimen | 85      | 84 | 85 | 83 | 85 | 85 | 83 | 87 | 86 | 82 | 84.56     |

Setelah mengetahui tingkat ketepatan dan kestabilan model yang dibentuk oleh regresi logistik, maka model akan diperiksa kemampuannya untuk memprediksi masing-masing kelas dengan melihat grafik ROC. Grafik ROC dibentuk dengan menampilkan kurva untuk masing-masing kelas pada masing-masing respon.



Gambar 12 Kurva ROC untuk respon topik dengan metode regresi logistik

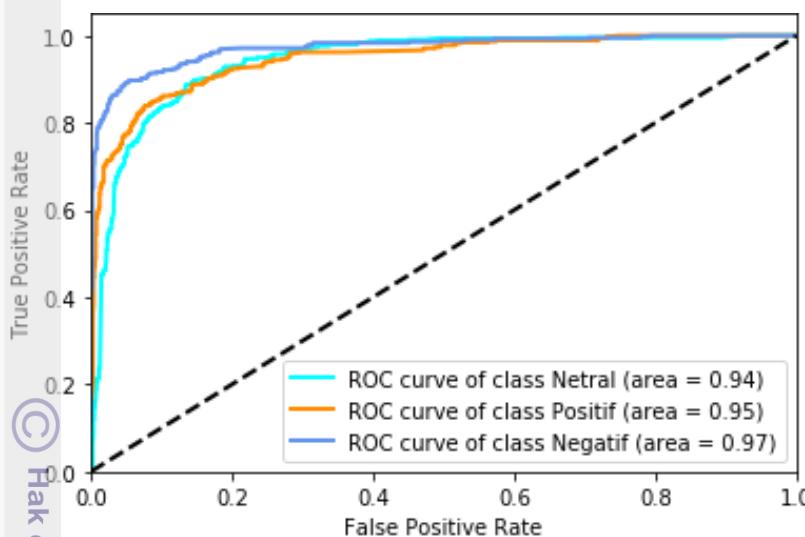
ROC yang dihasilkan oleh regresi logistik pada respon topik menghasilkan nilai AUC yang cukup baik untuk setiap kelas. Namun terdapat kesulitan dalam melakukan pengklasifikasian terhadap kelas akademik. Kurva yang dihasilkan oleh kelas akademik terletak lebih rendah dibandingkan dengan kurva kelas yang lain pada respon topik. AUC yang dihasilkan untuk kelas akademik sebesar 0.93. Sedangkan AUC untuk kelas selain akademik memiliki nilai yang cukup besar di atas 0.95. Respon sentimen memiliki grafik ROC yang hampir sama untuk setiap kelas. Grafik ROC untuk respon topik dapat dilihat pada Gambar 12 dan untuk respon sentimen dapat dilihat pada Gambar 13.

- Hak Cipta Dilindungi Undang-Undang  
© Hak cipta milik IPB (Institut Pertanian Bogor)
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 13 Kurva ROC untuk respon sentimen dengan metode regresi logistik

### Hutan Acak

Pemodelan klasifikasi yang dibangun dengan hutan acak memiliki setelan yang harus diatur terlebih dahulu. Setelan untuk hutan acak berupa jumlah pohon yang dibangun, maksimum peubah yang digunakan pada setiap pohon, banyaknya contoh minimum untuk pemisahan, banyaknya contoh minimum yang ada pada daun, dan kriteria pemisahan. Setelan tersebut disebut dengan *hyperparameter*.

*Hyperparameter* yang digunakan untuk hutan acak merupakan *hyperparameter* yang disarankan oleh *library* scikit-learn. Jumlah pohon yang digunakan adalah 10 pohon. Kemudian, maksimum peubah yang digunakan untuk setiap pohon adalah sebanyak akar dari banyaknya peubah yang tersedia. Maka, banyaknya peubah yang digunakan adalah sebanyak 87 peubah. Setelah itu, banyaknya contoh minimum yang digunakan untuk pemisahan sebanyak 2. Kemudian, banyaknya contoh minimum yang ada pada daun adalah sebanyak 1.

Setelah setelan dipersiapkan, maka hutan acak diterapkan dengan menggunakan data yang telah dibentuk pada subbab Persiapan Data. Respon yang digunakan adalah respon berdasarkan topik ujaran dan respon berdasarkan sentimen ujaran. Pemodelan hutan acak menggunakan 10-lipat validasi silang untuk memeriksa tingkat ketepatan dan kestabilan model.

Rata-rata akurasi metode hutan acak dengan menggunakan 10-lipat validasi silang memiliki nilai yang cukup baik pada masing-masing respon. Rata-rata akurasi tersebut bernilai 83.70% dengan simpangan baku 1.39% untuk respon topik. Sementara itu, untuk respon sentimen rata-rata akurasi bernilai sebesar 84.20% dengan simpangan baku 1.82% untuk respon sentimen. Rata-rata akurasi untuk respon sentimen lebih besar dibandingkan dengan rata-rata akurasi untuk respon topik. Hal ini menunjukkan bahwa respon sentimen lebih mudah untuk diprediksi dengan baik dibandingkan dengan respon topik dengan menggunakan hutan acak.

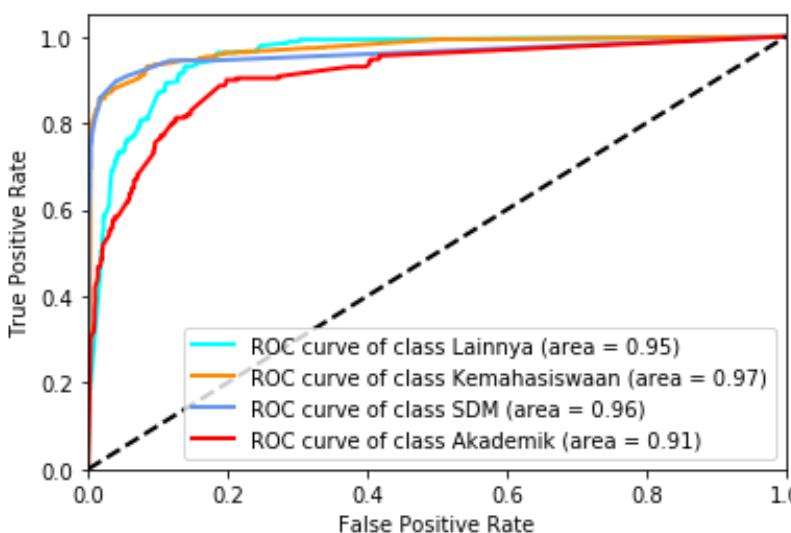
Pada respon topik, regresi logistik mendapatkan performa terbaik dengan lipatan ke-8 yaitu dengan akurasi sebesar 86%. Begitupun dengan respon sentimen,

performa terbaik diperoleh pada lipatan ke-8 dengan akurasi sebesar 87%. Akurasi setiap lipatan secara lengkap dapat dilihat pada Tabel 9.

Tabel 9 Akurasi setiap lipatan dengan menggunakan metode hutan acak (dalam persentase)

| Respon   | Lipatan |    |    |    |    |    |    |    |    |    | Rata-rata |
|----------|---------|----|----|----|----|----|----|----|----|----|-----------|
|          | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |           |
| Topik    | 82      | 85 | 84 | 84 | 84 | 81 | 83 | 86 | 84 | 85 | 83.70     |
| Sentimen | 85      | 85 | 85 | 84 | 83 | 86 | 82 | 87 | 85 | 81 | 84.20     |

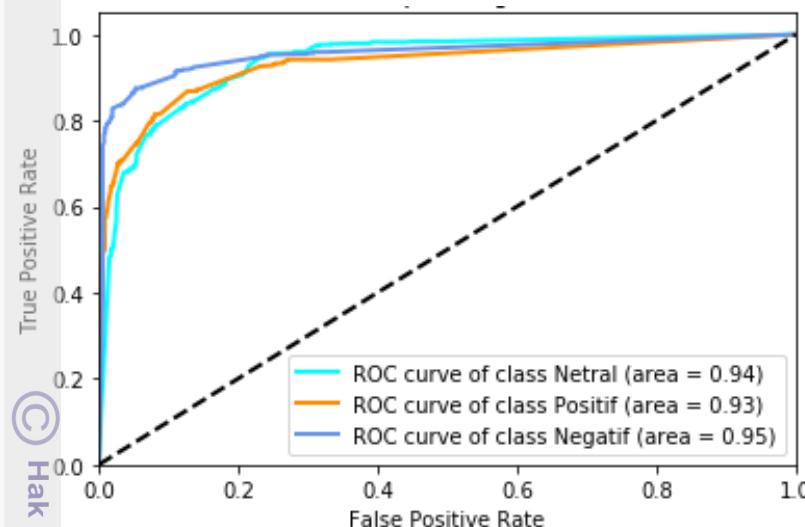
Setelah mengetahui tingkat ketepatan dan kestabilan model yang dibentuk oleh hutan acak, maka model akan diperiksa kemampuannya untuk memprediksi masing-masing kelas dengan melihat grafik ROC. Grafik ROC dibentuk dengan menampilkan kurva untuk masing-masing kelas pada masing-masing respon.



Gambar 14 Kurva ROC untuk respon topik dengan metode hutan acak

ROC yang dihasilkan oleh hutan acak pada respon topik menghasilkan nilai AUC yang cukup baik untuk setiap kelas. Namun terdapat kesulitan dalam melakukan pengklasifikasian terhadap kelas akademik. Kurva yang dihasilkan oleh kelas akademik terletak lebih rendah dibandingkan dengan kurva kelas yang lain pada respon topik. AUC yang dihasilkan untuk kelas akademik sebesar 0.91. Sedangkan AUC untuk kelas selain akademik memiliki nilai yang cukup besar di atas 0.95. Respon sentimen memiliki grafik ROC yang hampir sama untuk setiap kelas. Grafik ROC untuk respon topik dapat dilihat pada Gambar 14 dan untuk respon sentimen dapat dilihat pada Gambar 15.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 15 Kurva ROC untuk respon sentimen dengan metode hutan acak

### ***Support Vector Machine***

Penggunaan *support vector machine* pada pemodelan klasifikasi menggunakan fungsi kernel linear. Pemodelan support vector machine dibangun dengan fungsi kernel linear disebabkan oleh komputasi yang sangat tinggi apabila menggunakan fungsi kernel lainnya, seperti: *radial basis function*, polinomial, *sigmoid*, dan selainnya. Oleh karena itu, dipilih fungsi kernel linear untuk membentuk model dengan metode *support vector machine*.

Setelah menentukan fungsi kernel, metode *support vector machine* memiliki *hyperparameter* yang perlu disetel. *Hyperparameter* yang disetel untuk metode *support vector machine* adalah C. C merupakan *hyperparameter* yang mengendalikan tingkat toleransi yang diberikan oleh model yang akan dibangun. Sehingga toleransi tersebut akan berpengaruh dalam penentuan margin. Semakin besar C maka semakin sempit margin yang akan dibentuk. Begitupula sebaliknya semakin kecil C maka margin akan semakin luas. Berdasarkan setelan standar maka C yang akan digunakan sebesar 1.

Metode *support vector machine* dibangun dengan peubah penjelas yang dibentuk pada subbab Persiapan Data. Kemudian peubah yang dijadikan respon pertama adalah respon berdasarkan topik ujaran dan untuk respon kedua adalah respon berdasarkan sentimen ujaran.

Tingkat ketepatan dan kestabilan model yang dibangun menggunakan *support vector machine* diperiksa dengan metode 10-lipat validasi silang. Dengan menggunakan 10-lipat validasi silang, metode *support vector machine* akan menghasilkan 10 akurasi untuk setiap lipatan yang dibangun. Akurasi yang dihasilkan oleh setiap lipatan dapat dilihat pada Tabel 10 Akurasi tertinggi untuk respon topik dan respon sentimen terjadi pada lipatan yang sama yaitu lipatan ke-8 dengan nilai yang sama yaitu sebesar 87%. Rata-rata akurasi yang dihasilkan oleh metode *support vector machine* dengan respon topik sebesar 85.29% dengan simpangan baku 0.86%. Sedangkan rata-rata akurasi untuk respon sentimen sebesar

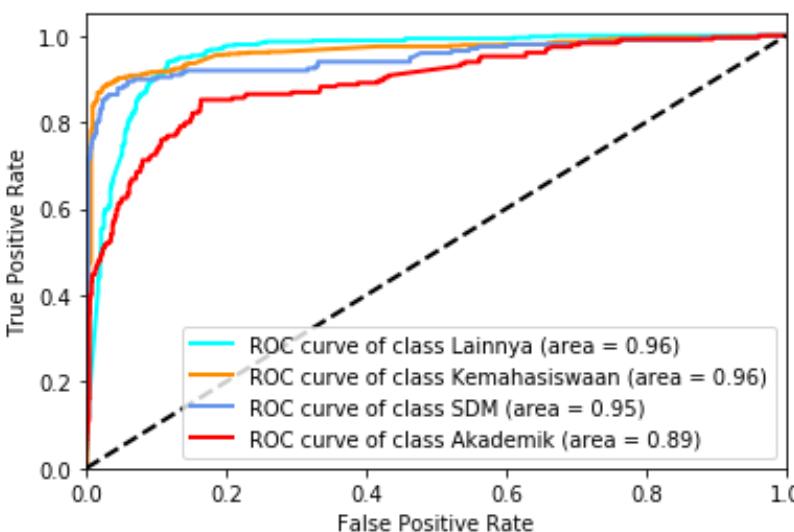
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

84.36% dengan simpangan baku 1.90%. Hasil akurasi dari metode *support vector machine* cukup tepat dan stabil.

Tabel 10 Akurasi setiap lipatan dengan menggunakan metode *support vector machine* (dalam persentase)

| Respon   | Lipatan |    |    |    |    |    |    |    |    |    | Rata-rata |
|----------|---------|----|----|----|----|----|----|----|----|----|-----------|
|          | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |           |
| Topik    | 85      | 85 | 85 | 86 | 85 | 84 | 85 | 87 | 86 | 84 | 85.29     |
| Sentimen | 84      | 86 | 85 | 84 | 85 | 86 | 81 | 87 | 85 | 81 | 84.36     |

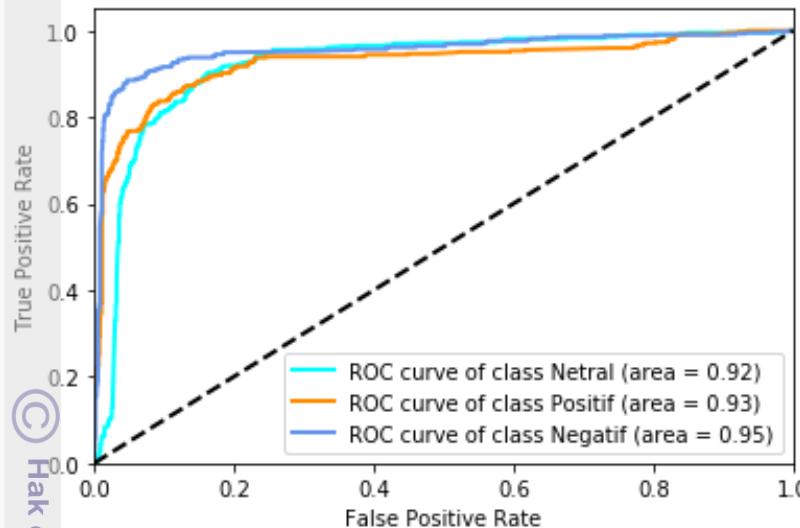
Kemampuan untuk mengklasifikasi dengan menggunakan metode *support vector machine* pada setiap kelas dapat diidentifikasi dengan menggunakan Grafik ROC. Berdasarkan grafik ROC pada Gambar 16, untuk metode *support vector machine* pada respon topik memiliki kemampuan klasifikasi yang baik untuk setiap kelas. Namun, untuk kelas akademik kurang mampu mengklasifikasikan dengan baik dibandingkan kelas selainnya. AUC yang dihasilkan untuk kelas akademik sebesar 0.89. Sedangkan untuk kelas selain akademik nilai AUC yang dihasilkan berada di atas 0.95



Gambar 16 Kurva ROC untuk respon topik dengan metode *support vector machine*

Setelah mengetahui ROC yang dihasilkan oleh respon topik, maka untuk respon sentimen juga diperiksa untuk setiap kelasnya. Berdasarkan Grafik ROC pada Gambar 17 dapat dilihat bahwa hasil klasifikasi untuk setiap kelas hampir sama yaitu mampu diklasifikasikan dengan baik. Rataan AUC yang dihasilkan pada respon sentimen adalah sebesar 0.93. Nilai ini cukup baik untuk menerapkan metode *support vector machine* pada kasus ini.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Kurva ROC untuk respon sentimen dengan metode *support vector machine*

### Bayes Naif

Bayes naif adalah salah satu pemodelan yang dapat digunakan untuk melakukan klasifikasi terhadap data teks. Namun, terdapat asumsi yang dimiliki oleh bayes naif. Asumsi tersebut adalah setiap peubah penjelas yang digunakan untuk membangun model klasifikasi harus saling bebas.

Pemodelan ini dibangun dengan menggunakan peubah penjelas sesuai tahapan pada subbab Persiapan Data. Respon yang digunakan adalah respon topik dan respon sentimen. Metode evaluasi ketepatan dan kestabilan yang digunakan adalah dengan menggunakan 10-lipat validasi silang.

Rataan akurasi yang dihasilkan oleh metode bayes naif cukup rendah yaitu sebesar 76.54% dengan simpangan baku sebesar 1.43% untuk respon topik. Sedangkan untuk respon sentimen menghasilkan rataan akurasi yang cukup tinggi sebesar 82.60% dengan simpangan baku sebesar 1.46%. Akurasi tertinggi diperoleh pada lipatan ke-4 dan ke-8 untuk respon topik dengan akurasi tertinggi sebesar 79%. Sedangkan untuk respon sentimen akurasi tertinggi berada pada lipatan ke-2, ke-3 dan ke-5 yaitu sebesar 84%. Keseluruhan akurasi setiap lipatan dapat dilihat pada Tabel 11.

Tabel 11 Akurasi setiap lipatan dengan menggunakan metode bayes naif (dalam persentase)

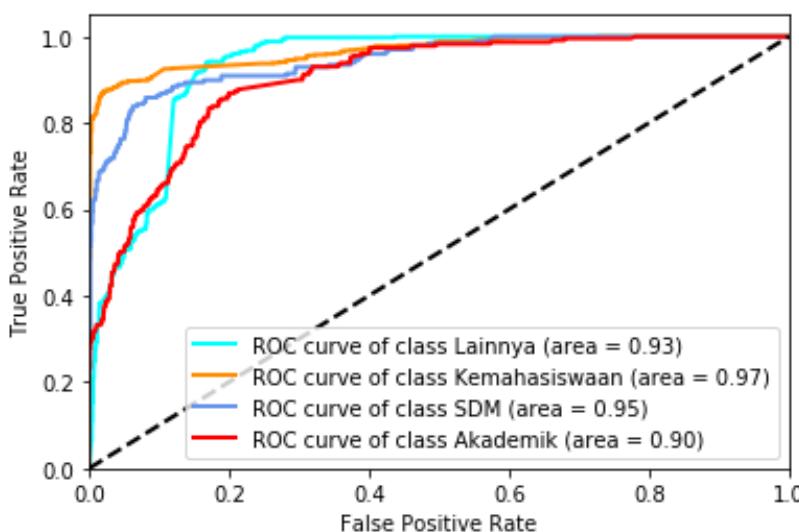
| Respon   | Lipatan |    |    |    |    |    |    |    |    |    | Rata-rata |
|----------|---------|----|----|----|----|----|----|----|----|----|-----------|
|          | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |           |
| Topik    | 76      | 75 | 75 | 79 | 75 | 76 | 78 | 79 | 77 | 76 | 76.54     |
| Sentimen | 83      | 84 | 84 | 82 | 84 | 83 | 83 | 83 | 80 | 80 | 82.60     |

Akurasi yang dihasilkan oleh metode bayes naif cukup rendah. Hasil ini bisa disebabkan karena adanya korelasi antar kemunculan kata. Apabila antar kata

memiliki korelasi yang tinggi maka peubah penjelas yang dihasilkan tidak saling bebas. Tidak saling bebasnya peubah penjelas mengakibatkan asumsi yang dibutuhkan metode bayes naif dilanggar. Sehingga performa yang dihasilkan metode bayes naif rendah.

Kemampuan prediksi untuk setiap kelas dapat dilihat dengan menggunakan grafik ROC seperti pada Gambar 18. Kelas yang mampu terprediksi dengan baik untuk respon topik hanya kelas kemahasiswaan dan kelas sumber daya manusia. Namun, untuk kelas selainnya belum cukup baik untuk melakukan prediksi dengan metode bayes naif.

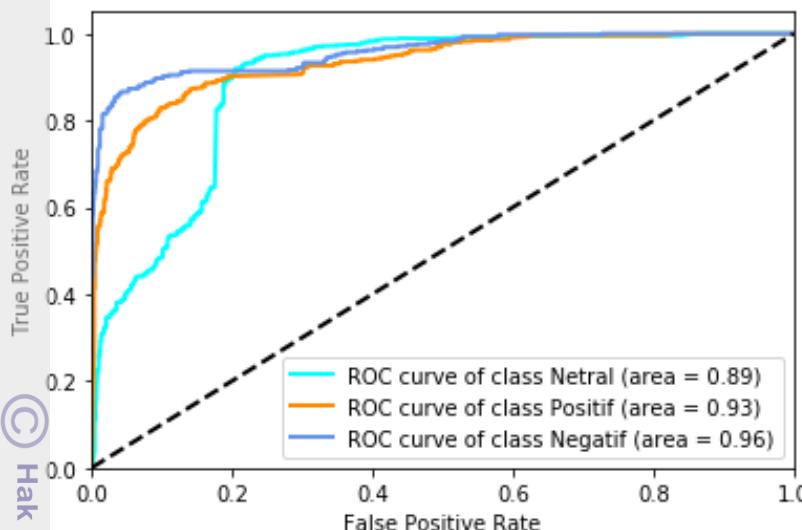
Kelas kemahasiswaan mampu terprediksi dengan baik memiliki AUC yang cukup tinggi sebesar 0.97. Namun untuk kelas akademik belum mampu terprediksi dengan baik memiliki AUC yang rendah sebesar 0.90. Rataan AUC untuk keseluruhan kelas pada respon topik dengan menggunakan metode bayes naif adalah sebesar 0.9375.



Gambar 18 Kurva ROC untuk respon topik dengan metode bayes naif

Grafik ROC yang dihasilkan oleh metode bayes naif dengan respon sentimen memiliki kesulitan dalam melakukan klasifikasi salah satu kelas. Grafik tersebut dapat dilihat pada Gambar 19. Kelas yang cukup sulit untuk dilakukan prediksi adalah kelas dengan label netral. AUC yang dihasilkan oleh kelas netral adalah sebesar 0.89. Nilai AUC ini lebih rendah dibandingkan dengan kelas yang lain. Kelas yang mampu diprediksi dengan baik adalah kelas dengan label negatif. Nilai AUC yang dihasilkan adalah 0.96. Pemodelan dengan menggunakan bayes naif akan lebih baik untuk kelas negatif namun akan sulit memprediksi dengan tepat untuk kelas netral. Secara keseluruhan rataan AUC yang dihasilkan adalah sebesar 0.9267.

- Hak Cipta Dilindungi Undang-Undang  
Hak Cipta milik IPB (Institut Pertanian Bogor)  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 19 Kurva ROC untuk respon sentimen dengan metode bayes naif

### Multilayer Perceptron

Pemodelan dengan menggunakan *multilayer perceptron* memiliki tingkat komputasi yang cukup tinggi dibandingkan dengan metode yang lainnya. Hal ini dikarenakan kompleksitas yang dibentuk oleh model sangatlah tinggi dibandingkan metode yang lainnya.

*Hyperparameter* yang perlu diatur untuk pemodelan menggunakan *multilayer perceptron* lebih banyak dibandingkan dengan pemodelan lainnya. Metode ini menggunakan *hidden layer* sebanyak 100. Sehingga komputasi yang dibutuhkan sangat tinggi. Fungsi aktivasi yang digunakan adalah *RELU* berguna untuk menyederhanakan komputasi. Kemudian agar pemodelan klasifikasi dapat diterapkan fungsi aktivasi pada lapisan terakhir adalah *sigmoid*. Optimasi yang digunakan untuk memperbarui bobot adalah *adam optimizer*. *Adam optimizer* memiliki kelebihan pada dataset yang besar dan komputasi yang cepat. Karena *adam optimizer* mampu mendapatkan bobot yang konvergen dengan cepat dan lebih baik dibandingkan yang lainnya. Dengan menggunakan *adam optimizer* maka diperlukan inisial untuk *learning rate*. *Learning rate* adalah langkah yang digunakan untuk memperbarui bobot. *Learning rate* yang digunakan adalah 0.001. Setelah itu, diperlukan regularisasi dengan menggunakan metode L2 yaitu alpha sebesar 0.0001. Maksimum iterasi yang digunakan adalah 200 iterasi.

Pembentukan metode *multilayer perceptron* menggunakan data yang telah disiapkan pada subbab Persiapan Data. Respon yang digunakan adalah respon topik dan respon sentimen. Evaluasi model untuk metode ini menggunakan 10-lipat validasi silang dan grafik ROC.

Tingkat ketepatan dan kestabilan model yang dihasilkan oleh metode *multilayer perceptron* memiliki akurasi yang cukup baik. Rata-rata akurasi yang dihasilkan metode *multilayer perceptron* untuk respon topik adalah sebesar 85.07% dengan simpangan baku 1.17%. Begitupun dengan respon sentimen, akurasi yang dihasilkan cukup tinggi yaitu sebesar 84.46% dengan simpangan baku 1.91%. Kemampuan prediksi dengan menggunakan *multilayer perceptron* memiliki

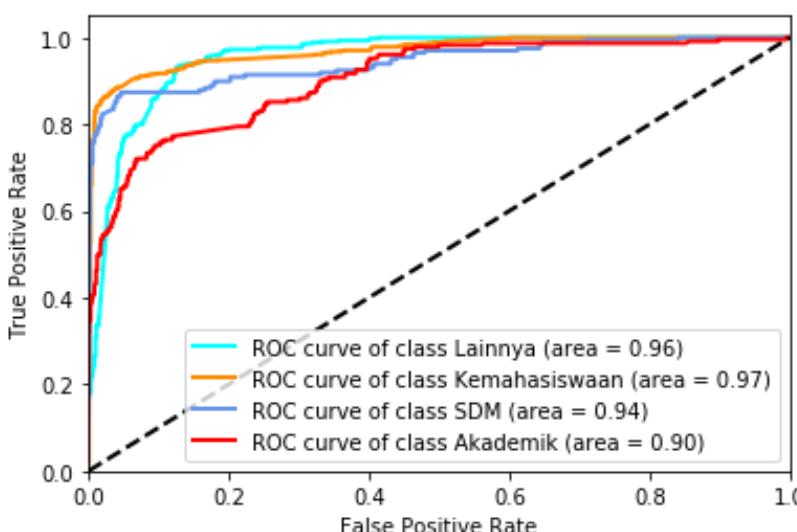
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

ketepatan yang cukup tinggi dan model yang dihasilkan cukup stabil. Hasil tersebut dapat dilihat secara lebih lengkap pada Tabel 12.

Tabel 12 Akurasi setiap lipatan dengan menggunakan metode *multilayer perceptron* (dalam persentase)

| Respon   | Lipatan |    |    |    |    |    |    |    |    |    | Rata-rata |
|----------|---------|----|----|----|----|----|----|----|----|----|-----------|
|          | 1       | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |           |
| Topik    | 86      | 85 | 85 | 83 | 86 | 84 | 85 | 87 | 86 | 84 | 85.07     |
| Sentimen | 85      | 86 | 86 | 84 | 84 | 86 | 81 | 87 | 85 | 81 | 84.46     |

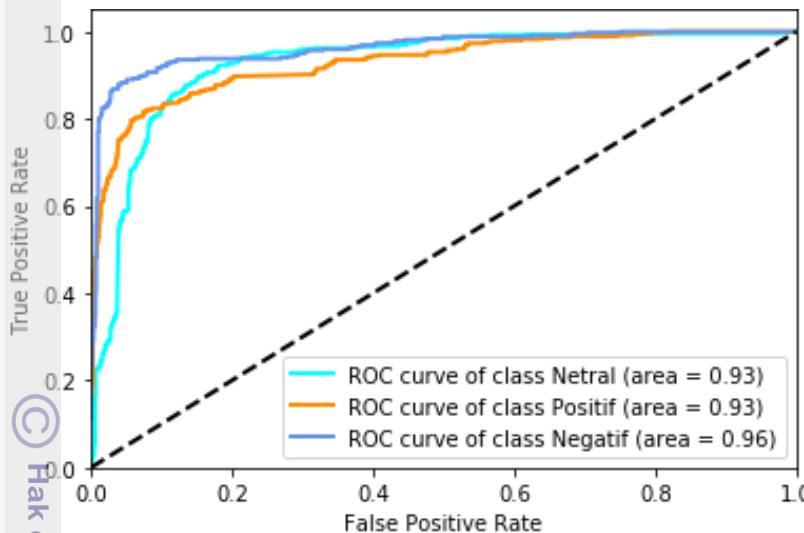
Model yang dibentuk oleh *multilayer perceptron* mengalami kesulitan dalam melakukan prediksi terhadap kelas akademik pada respon topik. Berdasarkan grafik ROC pada Gambar 20 *multilayer perceptron* tidak mampu memprediksi kelas akademik dengan baik dibandingkan kelas selain akademik. AUC yang dihasilkan oleh kelas akademik sebesar 0.90. Apabila dibandingkan dengan kelas selain akademik, AUC yang dihasilkan oleh kelas selainnya cukup besar. Kelas dengan label kemahasiswaan memiliki nilai AUC yang paling besar dibandingkan dengan kelas lainnya. Nilai AUC untuk kelas kemahasiswaan adalah sebesar 0.97.



Gambar 20 Kurva ROC untuk respon topik dengan metode *multilayer perceptron*

Respon yang berdasarkan sentimen ujaran memiliki performa yang baik untuk diklasifikasikan menggunakan metode *multilayer perceptron*. Seperti terlihat pada Gambar 21, masing-masing kelas memiliki kemampuan yang hampir seimbang dalam melakukan prediksi menggunakan metode *multilayer perceptron*. Nilai AUC yang tertinggi ketika melakukan prediksi untuk kelas negatif dengan nilai sebesar 0.96. Namun untuk kelas netral dan kelas positif memiliki nilai AUC yang cukup besar juga yaitu sebesar 0.93. Sehingga, rataan AUC secara keseluruhan yang dihasilkan oleh metode *multilayer perceptron* untuk respon sentimen adalah sebesar 0.94. Nilai ini cukup baik untuk melakukan prediksi terhadap respon sentimen.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar

1 Kurva ROC untuk respon sentimen dengan metode *multilayer perceptron*

### Perbandingan Model

Hasil perbandingan akurasi keseluruhan model dengan respon topik dapat dilihat pada Tabel 13. Berdasarkan tabel tersebut dapat dilihat bahwa regresi logistik memiliki akurasi yang paling tinggi dibandingkan dengan metode klasifikasi yang lain. Simpangan baku dari akurasi pun relatif kecil yang berarti bahwa akurasi yang dihasilkan oleh regresi logistik cukup stabil.

Tabel 13 Hasil akurasi pada respon topik dengan menggunakan 10-lipat validasi silang (dalam persentase)

| No | Metode                       | Akurasi | Simpangan Baku |
|----|------------------------------|---------|----------------|
| 1  | Regresi Logistik             | 86.29   | 1.23           |
| 2  | Hutan Acak                   | 83.70   | 1.39           |
| 3  | SVM                          | 85.29   | 0.86           |
| 4  | Bayes Naif                   | 76.54   | 1.43           |
| 5  | <i>Multilayer Perceptron</i> | 85.07   | 1.17           |

Adapun untuk kriteria selanjutnya adalah membandingkan AUC pada masing-masing model. Hal ini dilakukan dengan cara melakukan pembagian data latih dan data uji. Setelah itu, data latih akan diterapkan beberapa metode klasifikasi. Kemudian pada data latih akan dibentuk grafik ROC untuk masing-masing metode.

Grafik ROC pada masing-masing metode dapat dilihat pada gambar pada subbab masing-masing metode. Tabel 14 adalah nilai AUC yang dihasilkan oleh masing-masing metode pada respon topik. Berdasarkan Tabel 14 kita dapat melihat bahwa rataan AUC yang paling tinggi dengan menggunakan metode regresi logistik. Hasil ini memiliki perbedaan yang cukup signifikan antara metode regresi logistik dengan metode yang lain dalam membangun model klasifikasi. Sehingga untuk

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

melakukan prediksi terhadap respon respon topik akan digunakan metode regresi logistik.

Tabel 14 Hasil AUC pada respon respon topik

| No | Metode Klasifikasi           | Kemahasiswaan | Akademik | SDM  | Lainnya | Rataan |
|----|------------------------------|---------------|----------|------|---------|--------|
| 1  | Regresi Logistik             | 0.97          | 0.93     | 0.98 | 0.96    | 0.96   |
| 2  | SVM                          | 0.96          | 0.89     | 0.95 | 0.96    | 0.94   |
| 3  | Hutan Acak                   | 0.97          | 0.91     | 0.96 | 0.95    | 0.95   |
| 4  | <i>Multilayer Perceptron</i> | 0.97          | 0.90     | 0.94 | 0.96    | 0.94   |
| 5  | Bayes Naif                   | 0.97          | 0.90     | 0.95 | 0.93    | 0.94   |

Untuk respon sentimen hasil akurasi pemodelan klasifikasi dengan menggunakan 10-lipat validasi silang dapat dilihat pada Tabel 15. Berdasarkan Tabel 15 untuk respon dengan respon sentimen dapat kita lihat bahwa regresi logistik memiliki akurasi yang paling tinggi dibandingkan dengan metode klasifikasi yang lain. Simpangan baku dari akurasi pun relatif kecil yang berarti bahwa akurasi yang dihasilkan oleh regresi logistik cukup stabil.

Tabel 15 Hasil akurasi pada respon sentimen dengan menggunakan 10-lipat validasi silang (dalam persentase)

| No | Metode                       | Akurasi | Simpangan Baku |
|----|------------------------------|---------|----------------|
| 1  | Logistik Regresi             | 84.56   | 1.36           |
| 2  | Hutan Acak                   | 84.20   | 1.82           |
| 3  | SVM                          | 84.36   | 1.90           |
| 4  | Bayes Naif                   | 82.60   | 1.46           |
| 5  | <i>Multilayer Perceptron</i> | 84.46   | 1.91           |

Adapun untuk kriteria selanjutnya adalah membandingkan AUC pada masing-masing model. Hal ini dilakukan serupa dengan respon topik Tabel 16 adalah nilai AUC yang dihasilkan oleh masing-masing metode pada respon sentimen.

Tabel 16 Hasil AUC pada respon respon sentimen

| No | Metode Klasifikasi           | Positif | Netral | Negatif | Rataan |
|----|------------------------------|---------|--------|---------|--------|
| 1  | Regresi Logistik             | 0.95    | 0.94   | 0.97    | 0.95   |
| 2  | SVM                          | 0.93    | 0.92   | 0.95    | 0.93   |
| 3  | Hutan Acak                   | 0.93    | 0.94   | 0.95    | 0.94   |
| 4  | <i>Multilayer Perceptron</i> | 0.93    | 0.93   | 0.96    | 0.94   |
| 5  | Bayes Naif                   | 0.93    | 0.89   | 0.96    | 0.93   |

Berdasarkan Tabel 16 kita dapat melihat bahwa rataan AUC yang paling tinggi dengan menggunakan metode regresi logistik. Hasil ini memiliki perbedaan yang cukup signifikan antara metode regresi logistik dengan metode yang lain dalam

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

membangun model klasifikasi. Sehingga untuk melakukan prediksi terhadap respon sentimen akan digunakan metode regresi logistik.

## Prediksi

Metode pemodelan klasifikasi terbaik telah ditentukan pada subbab Perbandingan Model. Setelah itu, data yang belum memiliki label akan dilakukan prediksi berdasarkan metode klasifikasi terbaik. Model klasifikasi terbaik untuk data ini adalah regresi logistik. Model akan dibentuk pada masing-masing respon (respon topik dan respon sentimen) dengan menggunakan 4 995 data yang telah memiliki label.

Setelah mendapatkan label hasil prediksi dengan menggunakan metode terbaik, maka label tersebut diuji dengan menggunakan uji proporsi. Pengujian ini berfungsi untuk melakukan inferensi terhadap label yang telah diperoleh. Pengujian tersebut dilakukan dengan cara membandingkan banyaknya *tweets* yang mengandung label positif dengan keseluruhan *tweets* yang mengandung label positif dan negatif. Proporsi tersebut dihitung untuk setiap label pada respon topik. Kemudian diuji dengan hipotesis nol proporsi tersebut tidak lebih dari 50%. Taraf nyata yang digunakan untuk kasus ini sebesar 5%.

Kemudian dilanjutkan dengan pengujian dua proporsi antar label pada respon yang berdasarkan topik ujaran. Pengujian ini dibentuk dengan menggunakan tabulasi silang antar label. Yang mana isi dari tabel tersebut adalah selisih dari proporsi pada label yang ada di kolom dengan proporsi pada label yang ada di baris. Apabila isi dari nilai itu positif maka proporsi kolom lebih besar dari proporsi baris berlaku sebaliknya apabila isi dari nilai tersebut negatif maka proporsi kolom lebih kecil dari proporsi baris. Kemudian dilakukan pengujian proporsi 2 populasi yang akan menghasilkan tabel yang berisi nilai-p pada masing-masing selisih proporsi.

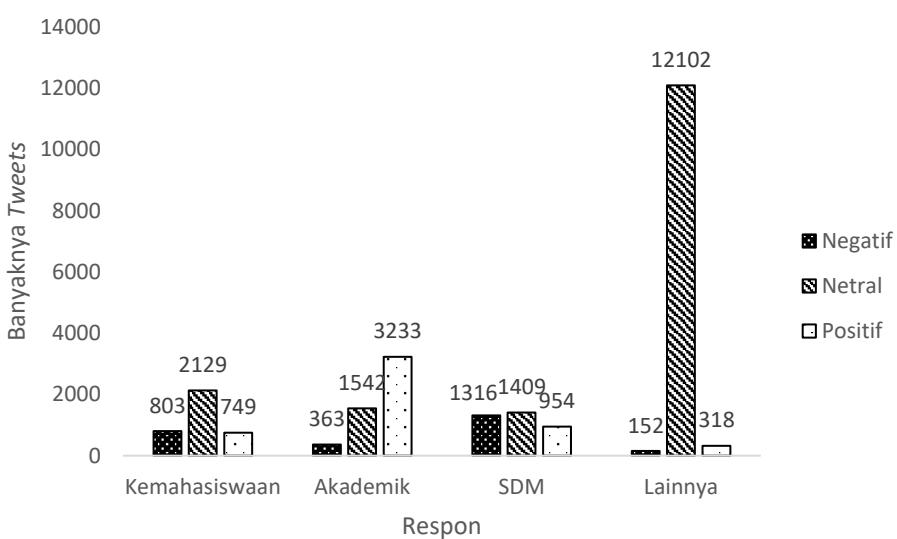
## IPB

Pada Gambar 22 dapat dilihat pada grafik batang yang dibentuk oleh *tweets* dengan topik IPB menghasilkan label lainnya yang cukup tinggi dibandingkan dengan label yang lain. Pada topik akademik IPB mendapatkan sentimen yang cukup positif. Dapat dilihat pada grafik batang dengan label akademik positif lebih tinggi dibandingkan sentimen yang lainnya. Namun, untuk sumber daya manusia IPB sempat mengalami sentimen negatif dari pengguna *tweets*. Sedangkan untuk kemahasiswaan, sentimen yang dihasilkan pada topik IPB netral saja.

Perbandingan tweets yang mengandung unsur positif dan negatif dapat dilihat pada Tabel 17. Dapat dilihat bahwa label kemahasiswaan dan label SDM untuk topik IPB tidak cukup bukti untuk menyatakan bahwa proporsi sentimen positif yang dihasilkan lebih dari 50%. Sedangkan untuk label lainnya dan akademik cukup bukti untuk menyatakan bahwa proporsi sentimen positif lebih dari 50%.

Hal ini menunjukkan bahwa IPB memiliki sentimen positif untuk keseluruhan *tweets*. Sentimen positif tersebut terjadi untuk topik akademik dan lainnya. Sedangkan sentimen negatif terjadi pada topik kemahasiswaan dan sumber daya manusia. Hal ini bisa disebabkan oleh kasus yang terjadi saat kartu kuning jokowi untuk label kemahasiswaan. Untuk topik sumber daya manusia bisa disebabkan karena terjadinya isu alumni yang bermasalah dengan civitas akademik IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 22 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas IPB

Tabel 17 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk IPB

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.48     | 0.92    |
| Akademik      | 0.90     | 0.00    |
| SDM           | 0.42     | 1.00    |
| Lainnya       | 0.68     | 0.00    |
| Total         | 0.67     | 0.00    |

Setelah mengetahui sentimen pada masing-masing topik, maka perbandingan sentimen positif untuk masing-masing topik dilakukan dengan menggunakan pengujian dua proporsi. Seperti yang terlihat pada Tabel 18. bahwa proporsi sentimen positif untuk masing-masing topik berbeda nyata.

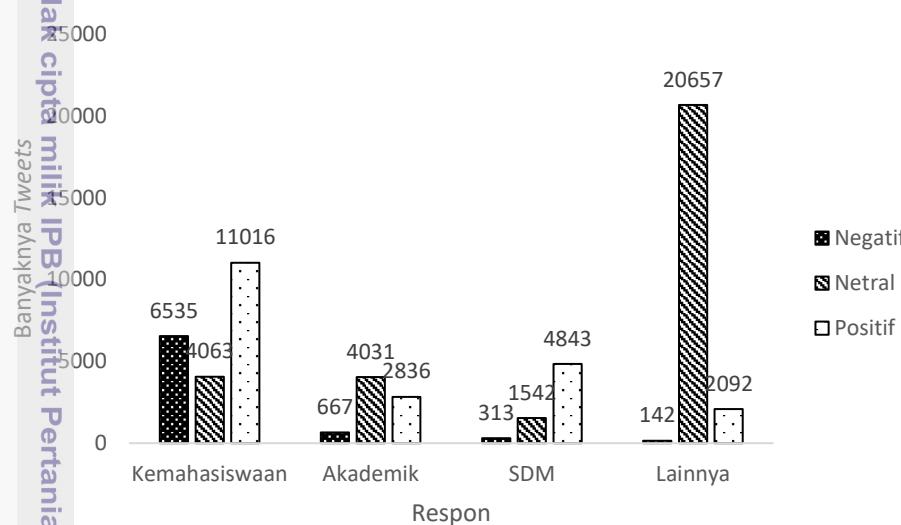
Tabel 18 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk IPB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM   | Lainnya |
|---------------|---------------|----------|-------|---------|
| Kemahasiswaan |               | 0.00     | 0.00  | 0.00    |
| Akademik      | -0.42         |          | 0.00  | 0.00    |
| SDM           | -0.18         | 0.48     |       | 0.00    |
| Lainnya       | -0.19         | 0.22     | -0.26 |         |

Label kemahasiswaan memiliki proporsi sentimen positif yang lebih kecil dibandingkan dengan topik lainnya. Sedangkan untuk label akademik memiliki proporsi sentimen positif yang lebih besar dibandingkan dengan label lainnya.

- UGM**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Topik sumber daya manusia memiliki proporsi sentimen positif yang lebih besar dibandingkan dengan kemahasiswaan namun lebih kecil dari topik akademik dan topik lainnya.



Gambar 3 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UGM

Untuk melakukan inferensi, maka pengujian proporsi dilakukan. Pengujian ini dilakukan untuk melihat sentimen yang diperoleh UGM pada masing-masing kategori untuk respon topik. Dapat dilihat pada Tabel 19, nilai-p seluruhnya yang dihasilkan kurang dari 0.05. Hal ini menunjukkan bahwa UGM mendapatkan sentimen positif pada keseluruhan *tweets* dan juga mendapatkan sentimen positif pada setiap label pada respon topik.

Tabel 19 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UGM

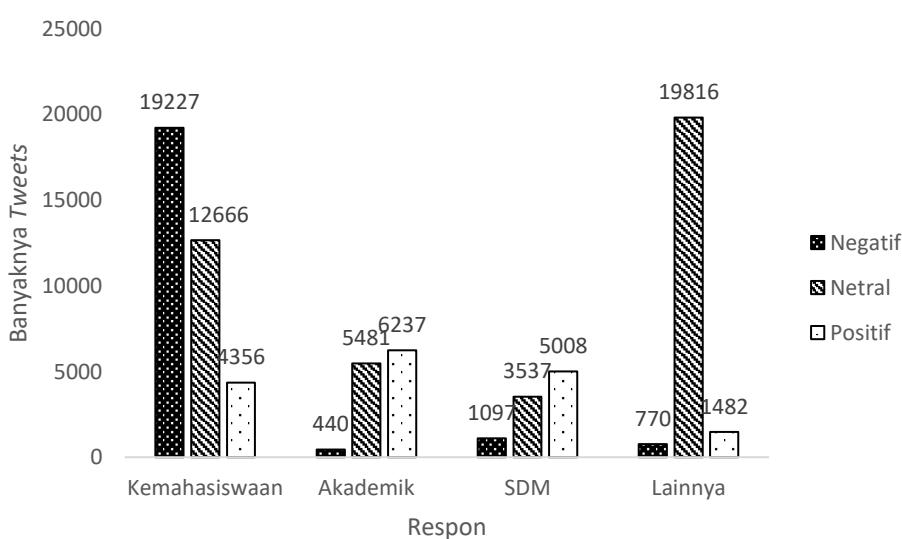
| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.63     | 0.00    |
| Akademik      | 0.81     | 0.00    |
| SDM           | 0.94     | 0.00    |
| Lainnya       | 0.94     | 0.00    |
| Total         | 0.73     | 0.00    |

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Tabel 20 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UGM dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.00    |
| Akademik      | -0.18         |          | 0.00 | 0.00    |
| SDM           | -0.10         | -0.13    |      | 0.32    |
| Lainnya       | -0.31         | -0.13    | 0.00 |         |

Hasil prediksi untuk UI dapat dilihat pada Gambar 24. Hasil tersebut menunjukkan bahwa topik yang terprediksi paling banyak adalah respon dengan sentimen netral dan topik lainnya. Kemudian untuk respon kemahasiswaan menghasilkan sentimen yang cenderung negatif. Hal ini bisa disebabkan karena banyaknya pemberitaan mengenai UI terkait isu kartu kuning Jokowi. Kemudian, pada respon program akademik topik UI cukup mendapatkan sentimen positif. Sedangkan untuk respon sumber daya manusia UI adalah topik yang paling jarang dibahas. Namun sumber daya alam memiliki sentimen yang cukup positif.



Gambar 24 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UI

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

Tabel 21 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UI

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.18     | 1.00    |
| Akademik      | 0.93     | 0.00    |
| SDM           | 0.82     | 0.00    |
| Lainnya       | 0.66     | 0.00    |
| Total         | 0.44     | 1.00    |

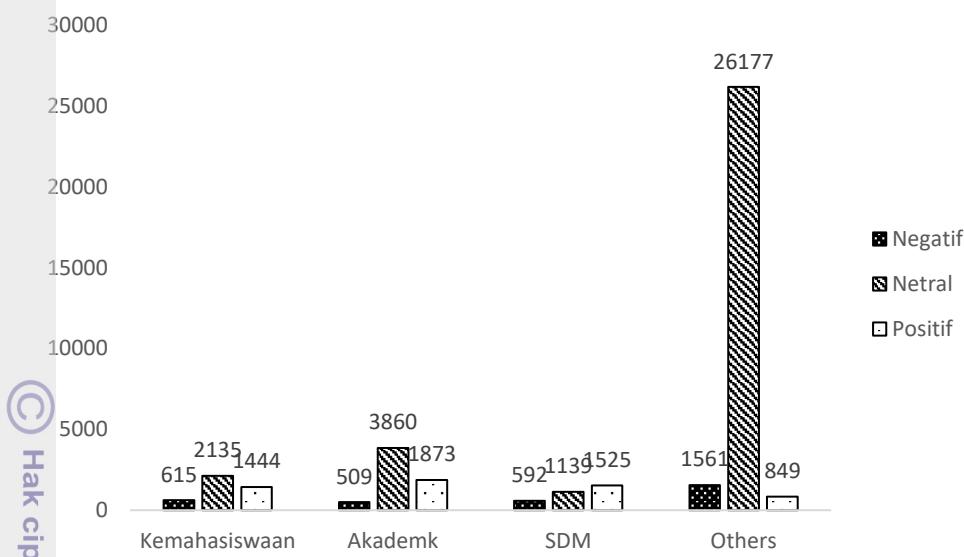
Pengujian untuk perbandingan antar label dapat dilihat pada Tabel 22. Hasil dari pengujian tersebut menunjukkan bahwa setiap selisih proporsi positif antar label saling nyata. Hal ini berarti bahwa proporsi positif antar label saling berbeda. Proporsi positif untuk label kemahasiswaan memiliki proporsi positif yang lebih kecil dibandingkan dengan akademik, sumber daya manusia, dan lainnya. Sedangkan untuk label akademik memiliki proporsi sentimen positif yang lebih besar dibandingkan dengan sumber daya manusia dan lainnya. Kemudian untuk label sumber daya manusia dan lainnya proporsi sentimen positif lebih besar untuk label SDM.

Tabel 22 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UI dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.00    |
| Akademik      | -0.75         |          | 0.00 | 0.00    |
| SDM           | -0.26         | 0.11     |      | 0.00    |
| Lainnya       | -0.47         | 0.28     | 0.16 |         |

## ITB

Prediksi yang dihasilkan untuk ITB dapat dilihat pada Gambar 25. *Tweets* yang terprediksi untuk universitas ITB memiliki label paling banyak adalah label netral untuk respon sentimen dan label lainnya untuk. Sedangkan untuk label kemahasiswaan, akademik, cenderung netral begitupun dengan label lainnya. Sedangkan untuk label sumber daya manusia menghasilkan sentimen yang positif.



Gambar 25 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas ITB

Tabel 23 menjelaskan bahwa sentimen yang dihasilkan oleh setiap label memiliki sentimen yang positif kecuali pada label lainnya. Hal ini ditunjukkan dengan nilai-p yang berinilai 0 untuk seluruh label namun bernilai 1 untuk label lainnya. Hal ini menunjukkan bahwa ITB mendapatkan sentimen positif untuk keseluruhan *tweets* terutama untuk label kemahasiswaan, akademik, dan sumber daya manusia. Namun untuk label lainnya mendapatkan sentimen negatif.

Tabel 23 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk ITB

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.70     | 0.00    |
| Akademik      | 0.79     | 0.00    |
| SDM           | 0.72     | 0.00    |
| Lainnya       | 0.35     | 1.00    |
| Total         | 0.63     | 0.00    |

Pengujian yang dilakukan untuk membandingkan proporsi sentimen positif antar label dapat dilihat pada Tabel 24. Berdasarkan tabel tersebut dapat dilihat bahwa nilai selisih proporsi dengan sentimen positif untuk setiap label memiliki nilai-p sebesar 0. Hal ini berarti bahwa proporsi sentimen positif antar label saling berbeda nyata. Berdasarkan tabel tersebut kita dapat menyimpulkan bahwa label kemahasiswaan memiliki proporsi sentimen positif yang lebih kecil dibandingkan akademik namun lebih besar dibandingkan label sumber daya manusia dan label lainnya. Namun, untuk label akademik memiliki nilai-proporsi sentimen positif yang lebih besar dibandingkan label sumber daya manusia dan label lainnya. Begitupun dengan sumber daya manusia memiliki proporsi sentimen positif yang lebih besar dibandingkan label lainnya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

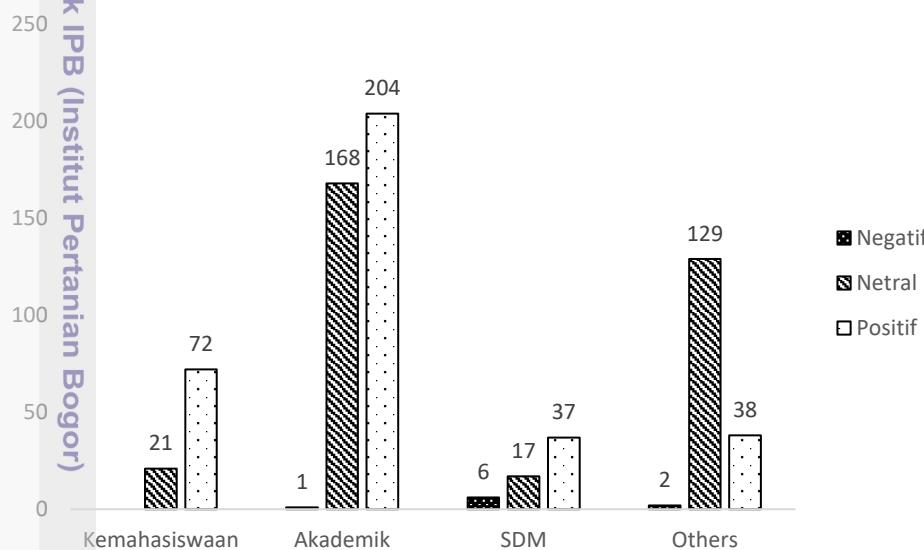
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Tabel 24 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk ITB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.00    |
| Akademik      | -0.08         |          | 0.00 | 0.00    |
| SDM           | 0.07          | 0.07     |      | 0.00    |
| Lainnya       | 0.35          | 0.43     | 0.37 |         |

ITS

Hasil prediksi yang dihasilkan untuk universitas ITS memiliki label yang positif untuk setiap kelasnya. Namun kelas yang paling banyak terprediksi adalah label akademik. Sangat jarang *tweets* yang membicarakan universitas ITS dengan sentimen negatif. Hasil ini ditunjukkan berdasarkan Gambar 26.



Gambar 26 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas ITS

Pernyataan di atas dapat dibuktikan dengan pengujian yang dilakukan pada Tabel 25. Nilai-p untuk setiap label bernilai 0. Hal ini menunjukkan bahwa untuk setiap label respon topik dan secara keseluruhan ITS mendapatkan sentimen yang cenderung positif.

Hasil pengujian proporsi sentimen positif antar label dapat dilihat pada Tabel 26. Label yang tidak berbeda nyata adalah label kemahasiswaan dengan sumber daya manusia, label kemahasiswaan dengan label lainnya, label akademik dengan label sumber daya manusia, dan label sumber daya manusia dengan label lainnya. Namun, label kemahasiswaan dengan akademik dan label akademik dengan label lainnya berbeda nyata. Hal ini ditunjukkan dengan nilai-p yang lebih kecil dari 0.025. Hal ini bisa membuktikan bahwa proporsi sentimen positif untuk

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

label kemahasiswaan lebih besar dibandingkan dengan label akademik. Begitupun dengan label lainnya memiliki nilai-proporsi sentimen positif yang lebih besar dibandingkan dengan label akademik.

Tabel 25 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk ITS

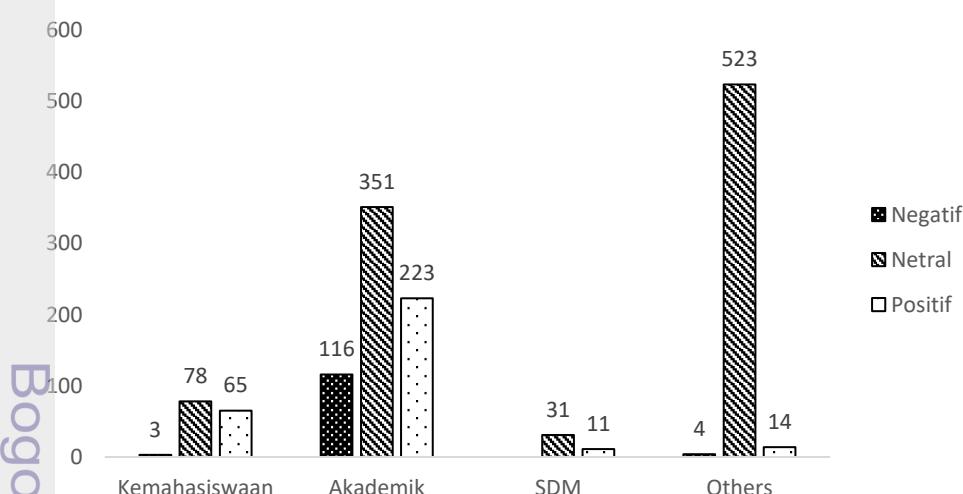
| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 1.00     | 0.00    |
| Akademik      | 0.99     | 0.00    |
| SDM           | 0.86     | 0.00    |
| Lainnya       | 0.95     | 0.00    |
| Total         | 0.97     | 0.00    |

Tabel 26 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk ITS dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM   | Lainnya |
|---------------|---------------|----------|-------|---------|
| Kemahasiswaan |               | 0.00     | 0.23  | 0.15    |
| Akademik      | 0.14          |          | 0.08  | 0.01    |
| SDM           | 0.02          | -0.09    |       | 0.10    |
| Lainnya       | 0.01          | -0.13    | -0.04 |         |

UB

Hasil prediksi untuk universitas UB dapat ditunjukkan dengan menggunakan Gambar 27. Gambar tersebut dapat menjelaskan bahwa kelas sumber daya manusia sangat jarang diprediksi dibandingkan dengan kelas selainnya. Label akademik mendapatkan sentimen yang cenderung netral. Begitupun dengan label lainnya dan kemahasiswaan.



Gambar 27 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UB

- Hak Cipta Dilindungi Undang-Undang  
© Hak cipta milik IPB (Institut Pertanian Bogor)
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a.

b. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pengujian untuk setiap label dilakukan untuk melihat label mana yang memiliki sentimen positif. Berdasarkan Tabel 27. dapat disimpulkan bahwa setiap label memiliki sentimen positif. Begitupun dengan *tweets* secara keseluruhan, sentimen yang dihasilkan cenderung positif.

Tabel 27 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UB

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.96     | 0.00    |
| Akademik      | 0.66     | 0.00    |
| SDM           | 1.00     | 0.00    |
| Lainnya       | 0.78     | 0.00    |
| Total         | 0.72     | 0.00    |

Setelah melihat sentimen untuk setiap label maka proporsi sentimen positif antar label akan dibandingkan dengan menggunakan pengujian proporsi 2 populasi. Pengujian tersebut menghasilkan tabel nilai-p seperti pada Tabel 28. Berdasarkan tabel tersebut dapat disimpulkan bahwa label kemahasiswaan berbeda nyata dengan seluruh label kecuali label akademik karena memiliki nilai-p di bawah 0.038233. Kemudian untuk label akademik berbeda nyata dengan label sumber daya manusia dan label lainnya. Namun untuk label sumber daya manusia tidak berbeda nyata dengan label lainnya. Tabel tersebut menunjukkan bahwa kemahasiswaan memiliki proporsi sentimen positif yang lebih besar dibandingkan dengan sumber daya manusia dan label lainnya. Kemudian akademik memiliki proporsi sentimen positif yang lebih besar dibandingkan dengan sumber daya manusia dan label lainnya. Untuk label selainnya tidak dapat disimpulkan saling berbeda nyata.

Tabel 28 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UB dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

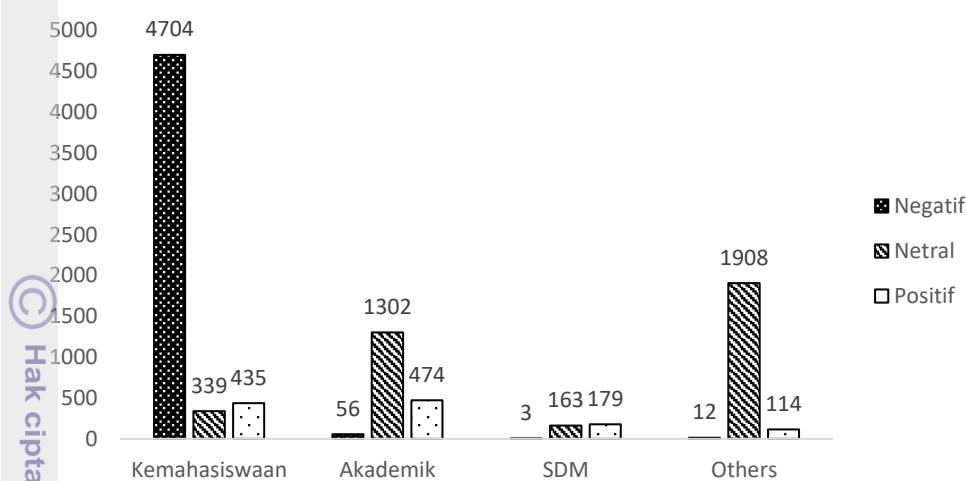
|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.04     | 0.01 | 0.00    |
| Akademik      | 0.01          |          | 0.02 | 0.00    |
| SDM           | 0.23          | 0.08     |      | 0.12    |
| Lainnya       | 0.16          | 0.01     | 0.10 |         |

## UNAIR

Prediksi yang dihasilkan pada universitas UNAIR dapat dilihat pada Gambar 28. Hal yang dapat dilihat pada gambar tersebut adalah pada label kemahasiswaan memiliki nilai yang cenderung negatif dibandingkan dengan label lainnya. Label akademik dan label lainnya memiliki sentimen yang cenderung netral. Sedangkan untuk label sumber daya manusia memiliki sentimen yang cenderung positif.

Pengujian yang dilakukan untuk melihat label mana yang memiliki sentimen positif dapat dilihat pada Tabel 29. Berdasarkan tabel tersebut dapat dilihat bahwa hanya label kemahasiswaan yang memiliki sentimen yang cenderung negatif. Begitupun dengan keseluruhan *tweets*, keseluruhan *tweets* memiliki sentimen yang

cenderung negatif. Namun untuk label akademik, sumber daya manusia dan lainnya memiliki sentimen yang cenderung positif.



Gambar 28 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNAIR

Tabel 29 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNAIR

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.08     | 1.00    |
| Akademik      | 0.89     | 0.00    |
| SDM           | 0.98     | 0.00    |
| Lainnya       | 0.90     | 0.00    |
| Total         | 0.20     | 1.00    |

Tabel 30 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNAIR dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.00    |
| Akademik      | -0.90         |          | 0.01 | 0.00    |
| SDM           | -0.12         | 0.08     |      | 0.36    |
| Lainnya       | -0.81         | 0.09     | 0.01 |         |

Perbandingan proporsi sentimen positif antar label dilakukan untuk melihat label mana yang memiliki proporsi sentimen lebih positif dibandingkan label selainnya. Perbandingan tersebut ditunjukkan pada Tabel 30. Berdasarkan tabel tersebut dapat disimpulkan bahwa hanya label sumber daya manusia dan label lainnya yang tidak berbeda nyata. Besarnya perbedaan label kemahasiswaan dengan setiap label selainnya menunjukkan bahwa proporsi sentimen positif untuk label kemahasiswaan lebih kecil dibandingkan label selainnya. Sedangkan untuk

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## UNDIP

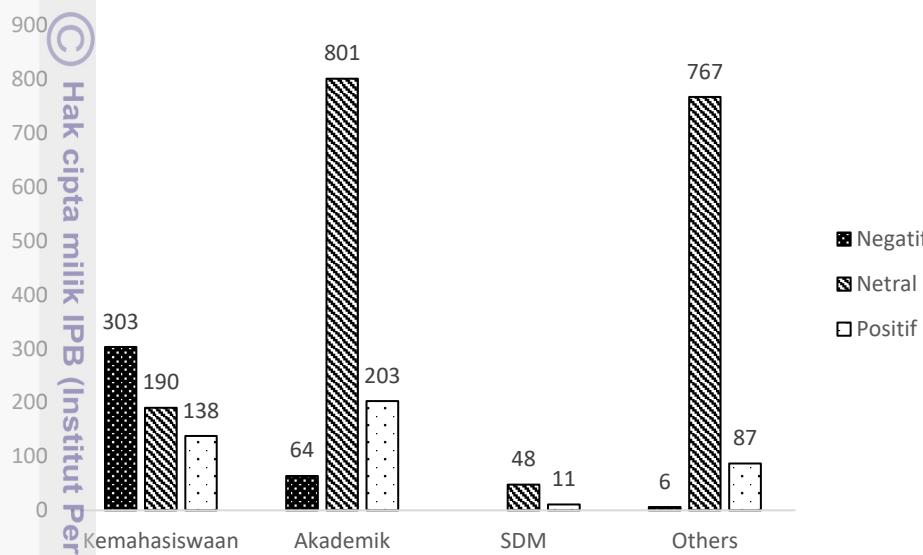
label akademik memiliki proporsi sentimen positif yang lebih besar dibandingkan label selainnya. Kemudian, untuk label sumber daya manusia hanya unggul terhadap label lainnya.

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 9 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNDIP

Hasil pengujian label mana yang memiliki sentimen yang cenderung positif ditunjukkan pada Tabel 31. Tabel tersebut menunjukkan bahwa seluruh tweets memiliki sentimen yang cenderung positif. Begitupun dengan setiap labelnya kecuali label kemahasiswaan.

Tabel 31 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNDIP

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.31     | 1.00    |
| Akademik      | 0.76     | 0.00    |
| SDM           | 1.00     | 0.00    |
| Lainnya       | 0.94     | 0.00    |
| Total         | 0.54     | 0.01    |

Selanjutnya adalah membandingkan proporsi sentimen positif antar label. Hasil dari pengujian proporsi dua populasi ditunjukkan pada Tabel 32. Tabel tersebut menunjukkan bahwa antar label memiliki perbedaan yang berbeda nyata.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

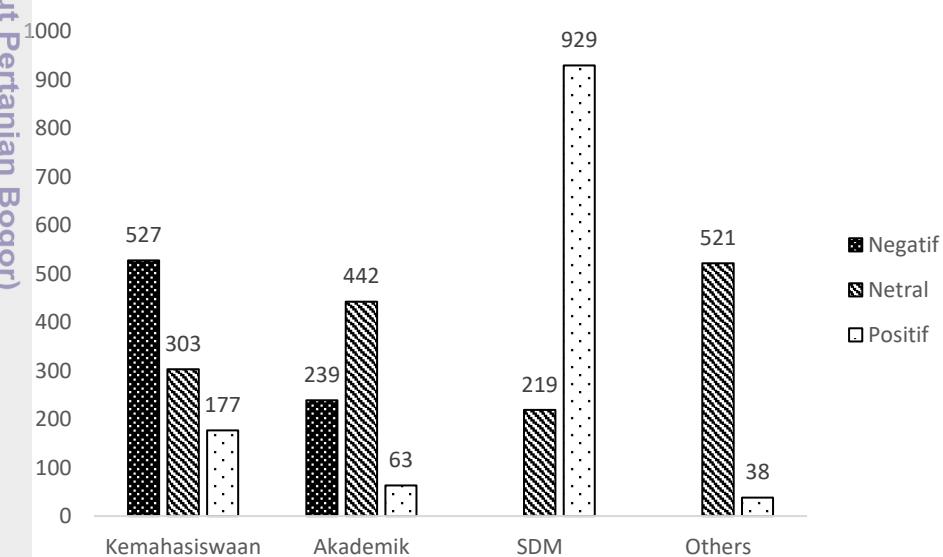
Hasil tersebut menunjukkan bahwa label kemahasiswaan memiliki proporsi sentimen positif yang lebih kecil dibandingkan dengan label lainnya. Label akademik memiliki proporsi sentimen positif yang lebih besar apabila dibandingkan dengan label sumber daya manusia dan label lainnya. Kemudian label sumber daya manusia memiliki nilai-proporsi sentimen positif yang lebih besar dibandingkan dengan label lainnya.

Tabel 32 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNDIP dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.00    |
| Akademik      | -0.69         |          | 0.01 | 0.00    |
| SDM           | -0.23         | 0.06     |      | 0.00    |
| Lainnya       | -0.45         | 0.24     | 0.18 |         |

## UNHAS

UNHAS memiliki hasil prediksi yang ditampilkan pada Gambar 30. Label kemahasiswaan memiliki sentimen yang cenderung negatif. Label sumber daya manusia memiliki sentimen yang cenderung positif. Sedangkan untuk label akademik dan lainnya cenderung memiliki sentimen netral.



Gambar 30 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNHAS

Sentimen positif diidentifikasi untuk setiap label dengan cara menguji proporsi sentimen positif. Hasil dari pengujian untuk setiap label tersebut ditunjukkan pada Tabel 33. Berdasarkan tabel tersebut kita dapat menyimpulkan bahwa keseluruhan tweets yang membahas universitas UNHAS memiliki sentimen positif. Namun tidak seluruh label memiliki sentimen yang positif. Label kemahasiswaan dan label

akademik tidak cenderung memiliki sentimen yang positif. Berbeda dengan halnya label lainnya dan label sumber daya manusia memiliki sentimen yang positif.

Tabel 33 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNHAS

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.25     | 1.00    |
| Akademik      | 0.21     | 1.00    |
| SDM           | 1.00     | 0.00    |
| Lainnya       | 1.00     | 0.00    |
| Total         | 0.61     | 0.00    |

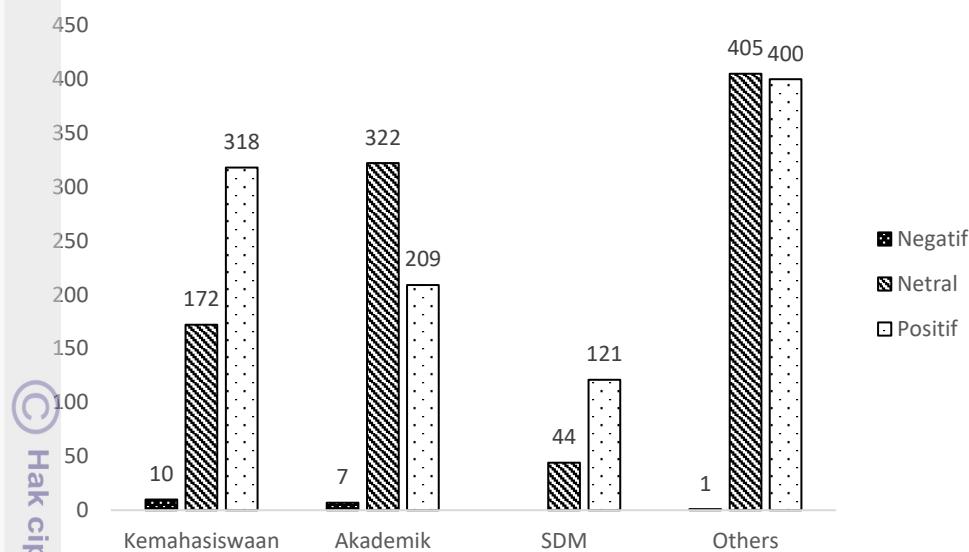
Setelah mengetahui sentimen untuk setiap label maka pengujian untuk setiap label dilakukan dengan cara menguji dengan menggunakan pengujian proporsi dua populasi. Proporsi yang digunakan adalah proporsi sentimen positif. Proporsi tersebut dibandingkan untuk setiap label. Hasil dari pengujian tersebut disajikan pada Tabel 34. Berdasarkan tabel tersebut antar tabel memiliki perbedaan proporsi sentimen positif kecuali pada label kemahasiswaan dengan label lainnya dan label akademik dengan label sumber daya manusia. Label kemahasiswaan memiliki proporsi sentimen positif yang lebih kecil apabila dibandingkan dengan label akademik dan label sumber daya manusia. Untuk label akademik dan label sumber daya manusia memiliki perbedaan yang sama dengan perbedaan yang sama dengan label lainnya, yaitu proporsi sentimen positif untuk label akademik dan label sumber daya manusia lebih besar apabila dibandingkan dengan label lainnya.

Tabel 34 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNHAS dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.00     | 0.00 | 0.07    |
| Akademik      | -0.75         |          | 1.00 | 0.00    |
| SDM           | -0.36         | 0.00     |      | 0.00    |
| Lainnya       | 0.04          | 0.79     | 0.79 |         |

## UNY

Prediksi dengan topik universitas UNY ditampilkan pada Gambar 31. Berdasarkan gambar tersebut UNY mendapatkan sentimen positif untuk kelas kemahasiswaan dan sumber daya manusia. Sedangkan untuk kelas lainnya dan kelas akademik memiliki sentimen netral. UNY sangat jarang mendapatkan sentimen negatif untuk masing-masing kelasnya



Gambar 31 Sebaran hasil prediksi untuk respon sentimen setiap topik pada universitas UNY

Setelah mengetahui secara eksploratif maka diperlukan pengujian untuk menarik kesimpulan. Pengujian tersebut dilakukan untuk melihat label yang memiliki sentimen positif. Hasil dari pengujian ini disajikan pada Tabel 35. Tabel 35 menjelaskan bahwa seluruh *tweets* yang membicarakan UNY memiliki sentimen yang positif. Begitupun dengan setiap label pada respon topik, setiap label memiliki sentimen yang positif.

Tabel 35 Pengujian proporsi sentimen positif untuk masing-masing label pada respon topik untuk UNY

| Label         | Proporsi | Nilai-p |
|---------------|----------|---------|
| Kemahasiswaan | 0.97     | 0.00    |
| Akademik      | 0.97     | 0.00    |
| SDM           | 1.00     | 0.00    |
| Lainnya       | 1.00     | 0.00    |
| Total         | 0.98     | 0.00    |

Setiap label dibandingkan dengan menggunakan pengujian proporsi dua populasi. Hasil dari perbandingan tersebut ditampilkan pada Tabel 36. Tabel tersebut menjelaskan bahwa label yang memiliki perbedaan proporsi adalah label akademik dengan sumber daya manusia, label kemahasiswaan dengan label lainnya dan label akademik dengan lainnya. Tabel 36 menunjukkan bahwa label kemahasiswaan memiliki proporsi yang lebih kecil apabila dibandingkan dengan label lainnya. Label akademik memiliki proporsi sentimen positif yang lebih kecil apabila dibandingkan dengan label sumber daya manusia dan label lainnya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Tabel 36 Pengujian perbandingan proporsi sentimen positif antar label pada respon topik untuk UNY dengan nilai-p di atas diagonal dan statistik uji di bawah diagonal

|               | Kemahasiswaan | Akademik | SDM  | Lainnya |
|---------------|---------------|----------|------|---------|
| Kemahasiswaan |               | 0.45     | 0.08 | 0.01    |
| Akademik      | 0.01          |          | 0.01 | 0.01    |
| SDM           | -0.01         | -0.03    |      | 0.16    |
| Lainnya       | -0.03         | -0.03    | 0.01 |         |

## SIMPULAN DAN SARAN

### Simpulan

Topik tentang beasiswa yang sering dibahas pada media sosial twitter. Sedangkan, topik lain yang dibahas adalah ketua BEM KM IPB, himpunan alumni, lulusan IPB yang menjadi pegawai bank, dan lain-lain.

Frekuensi *tweets* yang membicarakan IPB sangat banyak di daerah Pulau Jawa. Kemudian, pulau Sumatera merupakan pulau terbanyak kedua yang membicarakan topik IPB. Sedangkan, untuk wilayah lainnya tidak banyak yang membahas mengenai IPB.

Topik tentang IPB cenderung membahas selain topik kemahasiswaan, program akademik, dan sumber daya manusia. Selanjutnya, program pendidikan IPB mendapatkan sentimen yang positif. Sedangkan untuk sumber daya manusia IPB mendapatkan sentimen negatif. Kemudian untuk respon kemahasiswaan IPB mendapatkan sentimen yang netral.

### Saran

Pelabelan akan lebih baik hasilnya jika dilakukan oleh para ahli sehingga respon yang diberikan bukan subjektifitas dari penulis. Menambahkan normalisasi kata singkatan sangat diperlukan untuk mengubah kata menjadi baku. Sangat disarankan menggunakan aplikasi berbayar untuk mengakses API yang disediakan oleh pihak twitter. API yang berbayar akan mendapatkan akses yang lebih banyak dibandingkan dengan menggunakan aplikasi yang disediakan secara gratis.

## DAFTAR PUSTAKA

- Adiyana I, Hakim F. 2015. Implementasai Text Mining Pada Mesin Pencarian Twitter Untuk Menganalisis Topik – Topik Terkait “KPK dan JOKOWI”. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015*. Jakarta, Indonesia. Jakarta(ID) : Universitas Islam Indonesia. hlm: 570-581





Health. *Journal of Geographic Information System*. Wuhan (CN): Scientific Research Publishing. hlm 20-28.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## LAMPIRAN

Lampiran 1 Kata kunci yang digunakan untuk pengumpulan data

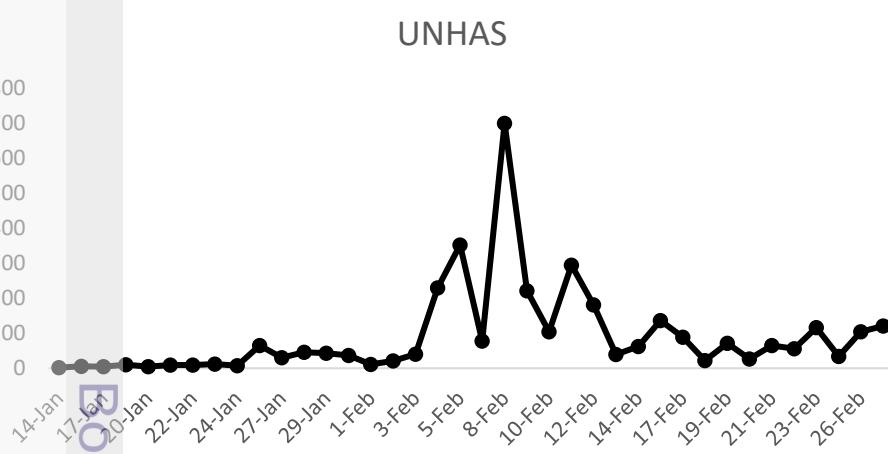
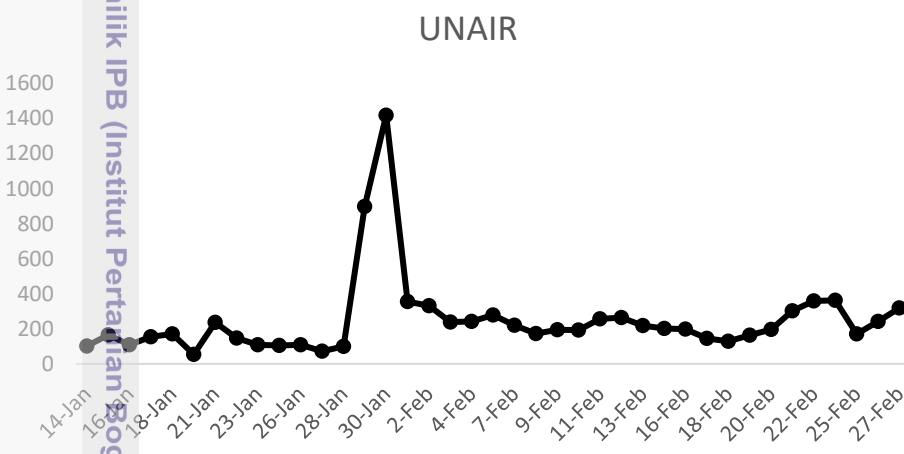
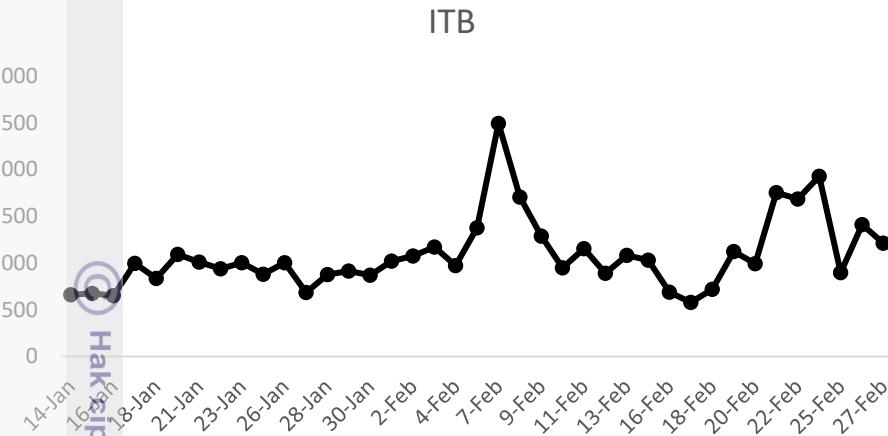
| No | Kata kunci                          |
|----|-------------------------------------|
| 1  | IPB                                 |
| 2  | UNDIP                               |
| 3  | UGM                                 |
| 4  | ITB                                 |
| 5  | UI+kampus                           |
| 6  | UI+yang                             |
| 7  | UNAIR+kampus                        |
| 8  | UNAIR+yang                          |
| 9  | UNHAS+kampus                        |
| 10 | UNHAS+yang                          |
| 11 | UNY+kampus                          |
| 12 | UNY+yang                            |
| 13 | UB+kampus                           |
| 14 | UB+yang                             |
| 15 | ITS+kampus                          |
| 16 | ITS+yang                            |
| 17 | Institut Pertanian Bogor            |
| 18 | Universitas Indonesia               |
| 19 | Universitas Gajah Mada              |
| 20 | Institut Teknologi Bandung          |
| 21 | Universitas Diponegoro              |
| 22 | Universitas Airlangga               |
| 23 | Universitas Hasanuddin              |
| 24 | Universitas Negeri Yogyakarta       |
| 25 | Universitas Brawijaya               |
| 26 | Institut Teknologi Sepuluh Nopember |
| 27 | @ipbofficial                        |
| 28 | @its_campus                         |
| 29 | @itb_official                       |
| 30 | @univ_indonesia                     |
| 31 | @UGMYogyakarta                      |
| 32 | @univ_diponegoro                    |
| 33 | @unair_official                     |
| 34 | @UB_official                        |
| 35 | @unyofficial                        |
| 36 | @hasanuddin_univ                    |

Lampiran 2

Frekuensi *tweets* setiap universitas setiap hari 14 Januari 2018 sampai dengan 27 Februari 2018

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



### Lampiran 2 Lanjutan

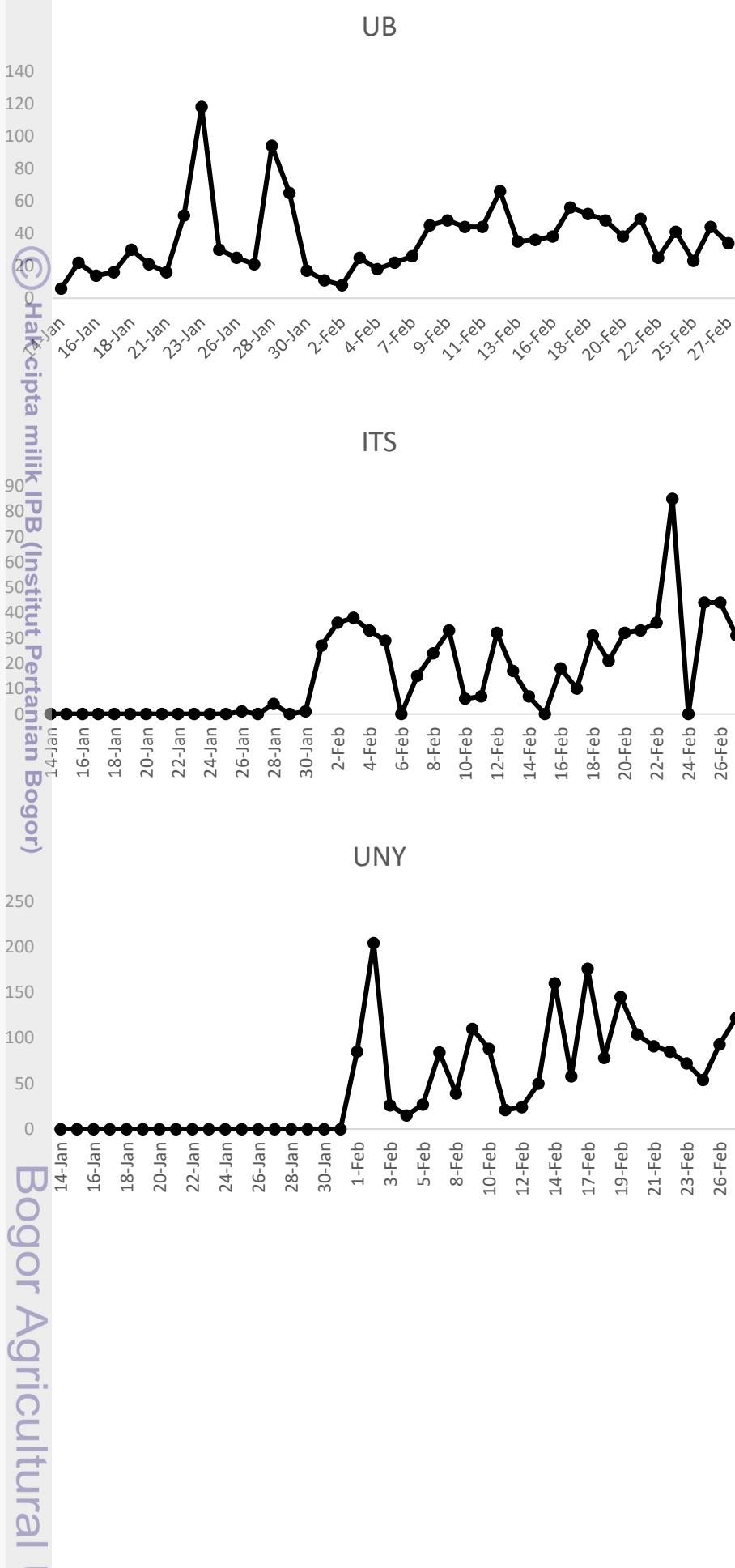
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.





### Lampiran 3 Awan kata setiap universitas

**1. Dilarang mengutip sebagian atau seluruh karya tulis**

- a. Pengutipan hanya untuk Repentingan pendidikan, penelitian, penulisan Riset ilmiah, penyusunan laporan, penulisan R

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbaikanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



IPB



UGM



UNAIR



ITB



U



UNDIP



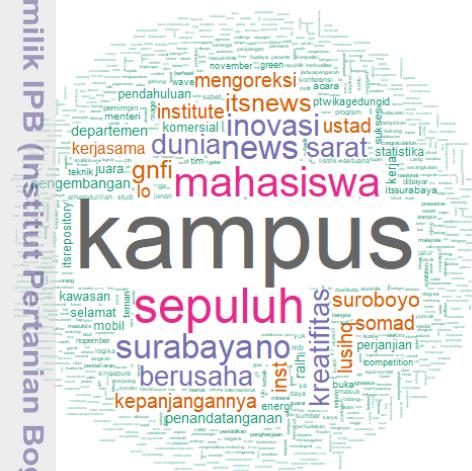
## Lampiran 3 Lanjutan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak mengikuti kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



UNHAS



ITS



UNY



UE



56

## RIWAYAT HIDUP

Penulis dilahirkan di Bandung pada tanggal 30 Agustus 1997 dan merupakan anak pertama dari empat bersaudara dari pasangan Bapak Wanwan Agung Nurwanto dan Ibu Laelasari Nuzuliah. Pada tahun 2014 penulis menyelesaikan pendidikan menengah atas di SMAN 1 Garut. Pada tahun yang sama penulis diterima di Institut Pertanian Bogor dengan mayor Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Institut Pertanian Bogor (IPB) melalui Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN).

Penulis bergabung dengan Dewan Perwakilan Mahasiswa (DPM) FMIPA pada periode 2015-2016 sebagai anggota komisi 5. Pada periode yang sama saya juga bergabung dengan Majelis Permusyawaratan Mahasiswa Keluarga Mahasiswa (MPM KM) IPB sebagai Badan Pekerja 4 (BP 4). Kemudian pada periode 2016-2017 saya bergabung dengan DPM FMIPA IPB sebagai ketua komisi 5. Penulis berkesempatan praktik lapang (PL) di PT. Independence yang merupakan anak perusahaan dari PT. Deka Marketing Research Jakarta Selatan pada Juli sampai September 2017.

Penulis meraih juara 1 pada perlombaan *National Data Days* (NDD) *Telkom University* tingkat nasional pada tahun 2017. Penulis juga meraih juara 3 pada perlombaan Statistika Ria 12 (SR 12) IPB tingkat nasional pada tahun 2017 dan *Data Analysis Competition* (DAC) Institut Teknologi Sepuluh Nopember (ITS) tingkat Asia Tenggara pada tahun 2016. Penulis mendapatkan penghargaan sebagai finalis pada perlombaan DIVACTORY yang diadakan oleh *Data Science Indonesia* (DSI) yang bekerja sama dengan PT. Telkom Indonesia.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.