# **BAB 4**

**HASIL PENELITIAN**



## ***Testing Environment***

### **Spesifikasi Sistem**

#### **Device**

Penelitian yang diimplementasikan ke dalam sebuah web aplikasi dibangun dan dijalankan pada perangkat pc dengan spesifikasi sebagai berikut.

|  |  |  |
| --- | --- | --- |
| Operating System | : | Windows 10 Pro Edition |
| CPU | : | Intel Core i7-4710HQ CPU @2.50 GHz |
| RAM | : | 8.00 GB |
| Internal Memory | : | 1 TB |



#### 

#### **Server**

Sistem web aplikasi ini dijalankan dengan server Apache

### **Perangkat Lunak**

Berikut adalah perangkat lunak yang digunakan dalam penelitian ini

1. Eclipse IDE
2. Notepad++
3. Microsoft Excel 2010
4. Xampp Server

### ***Library***

Berikut adalah library yang digunakan dalam pembangunan sistem prediksi.

1. Keras
2. Numpy
3. Sci-kit Learn

## **Hasil**

* + 1. **Distribusi Data**

Distribusi data dari dataset myPersonality dapat dilihat pada Tabel 4.1 dengan total sebanyak 250 data user Facebook.

**Tabel 4.1** Distribusi data myPersonality

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Value** | **Openness** | **Conscientiousness** | **Extraversion** | **Agreeableness** | **Neuroticism** |
| Yes | 176 | 130 | 96 | 134 | 99 |
| No | 74 | 120 | 154 | 116 | 151 |

Distribusi data dari dataset manual gathering dapat dilihat pada Tabel 4.2 dengan total sebanyak 150 data user Facebook yang didapatkan secara manual.

**Tabel 4.2** Distribusi data manual gathering

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Value** | **Openness** | **Conscientiousness** | **Extraversion** | **Agreeableness** | **Neuroticism** |
| Yes | 97 | 63 | 38 | 81 | 50 |
| No | 53 | 87 | 112 | 69 | 100 |

Distribusi data dari gabungan kedua dataset di atas dapat dilihat pada Tabel 4.3 dengan total sebanyak 400 data user.

**Tabel 4.3** Distribusi data gabungan myPersonality dan manual data gathering

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Value** | **Openness** | **Conscientiousness** | **Extraversion** | **Agreeableness** | **Neuroticism** |
| Yes | 273 | 193 | 134 | 215 | 149 |
| No | 127 | 207 | 266 | 185 | 251 |

* + 1. **Faktor Skenario Percobaan**

Beberapa faktor skenario percobaan yang akan digunakan selama proses testing untuk mendapatkan hasil terbaik dari implementasi machine learning ini terbagi dua yaitu:

* + - 1. **Features Selection**

Feature utama yang digunakan dari penelitian ini yaitu LIWC sebanyak 85 features, SPLICE sebanyak 94 features, dan SNA features sebanyak 7 features akan melewati proses feature selections atau pemilihan fitur selama proses testing berjalan untuk mendapatkan feature yang paling optimal dalam menghasilkan tingkat akurasi paling tinggi.

Modul feature selection ini berjalan secara otomatis dan dinamis selama proses berjalan dan feature paling optimal akan muncul di akhir bersamaan dengan tingkat akurasinya. Features selection hanya akan digunakan untuk skenario implementasi machine learning.

* + - 1. **Resampling**

Resampling adalah proses dimana dataset yang ada dilakukan manipulasi data dan duplikasi data untuk membuat sebuah data memiliki elemen yang seimbang. Proses resampling terbagi menjadi 2 yaitu:

* Oversampling

Menambahkan data ke data dengan persentase minoritas. Tujuan oversampling pada penelitian ini karena ada dataset dimana elemennya tidak seimbang. Sebagai contoh terdapat 70% pria dan 30% wanita. Hasil dari testing biasanya akan buruk dikarenakan faktor tersebut. Untuk itu, oversampling akan menambahkan elemen yang menjadi minoritas agar data menjadi seimbang. Seperti contoh diatas dimana 70% pria (70 orang) dan 30% wanita (30 orang) dari total 100 data menjadi 50% pria (75 pria) dan 50% wanita (75 wanita) dengan total data menjadi 150 data.

* Undersampling

Sistem undersampling hanya berbanding terbalik dengan oversampling. Jika oversampling menambahkan data, undersampling mengurangi data yang memiliki persentase mayoritas untuk menyeimbangkan data. Undersampling sebenarnya cocok digunakan untuk data dalam jumlah yang besar karena tidak terlalu berpengaruh jika dataset dikurangi. Tetapi jika digunakan pada dataset yang sedikit, akan membuat jumlah dataset semakin sedikit lagi. Namun, dalam penelitian ini, peneliti tetap mencoba untuk melakukan kedua proses resampling ini dan melihat hasil serta perbandingan keduanya.

Faktor skenario resampling ini akan digunakan pada skenario implementsi machine learning dan juga skenario implementasi deep learning.

* + 1. **Implementasi Machine Learning**
       1. **Skenario**

Implementasi machine learning dalam testing ini akan menggunakan 5 jenis algoritma (Support Vector Machine, Naïve Bayes, Logistic Regression, Gradient Boosting, dan LDA) dan 3 features utama (LIWC, SPLICE, dan SNA). Kemudian, dataset yang digunakan terbagi menjadi 3 seperti dijelaskan pada Bagian 4.2.1 mengenai distribusi data.

Dari faktor skenario diatas pada Bagian 4.2.2, machine learning akan menggunakan kedua faktor tersebut yaitu features selection (Bagian 4.2.2.1) dan resampling (Bagian 4.2.2.2) sehingga dapat dibentuk skenario percobaan untuk machine learning sebagai berikut:

* + 1. Penggunaan dataset myPersonality, percobaan tanpa menggunakan Features Selection dan tanpa menggunakan Resampling.
    2. Penggunaan dataset myPersonality, percobaan dengan menggunakan Features Selection dan tanpa menggunakan Resampling.
    3. Penggunaan dataset myPersonality, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Undersampling.
    4. Penggunaan dataset myPersonality, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Oversampling.
    5. Penggunaan dataset myPersonality, percobaan dengan menggunakan Features Selection dan dengan menggunakan Undersampling.
    6. Penggunaan dataset myPersonality, percobaan dengan menggunakan Features Selection dan dengan menggunakan Oversampling.
    7. Penggunaan dataset manual gathering, percobaan tanpa menggunakan Features Selection dan tanpa menggunakan Resampling.
    8. Penggunaan dataset manual gathering, percobaan dengan menggunakan Features Selection dan tanpa menggunakan Resampling.
    9. Penggunaan dataset manual gathering, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Undersampling.
    10. Penggunaan dataset manual gathering, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Oversampling.
    11. Penggunaan dataset manual gathering, percobaan dengan menggunakan Features Selection dan dengan menggunakan Undersampling.
    12. Penggunaan dataset manual gathering, percobaan dengan menggunakan Features Selection dan dengan menggunakan Oversampling.
    13. Penggunaan dataset gabungan, percobaan tanpa menggunakan Features Selection dan tanpa menggunakan Resampling.
    14. Penggunaan dataset gabungan, percobaan dengan menggunakan Features Selection dan tanpa menggunakan Resampling.
    15. Penggunaan dataset gabungan, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Undersampling.
    16. Penggunaan dataset gabungan, percobaan tanpa menggunakan Features Selection dan dengan menggunakan Oversampling.
    17. Penggunaan dataset gabungan, percobaan dengan menggunakan Features Selection dan dengan menggunakan Undersampling.
    18. Penggunaan dataset gabungan, percobaan dengan menggunakan Features Selection dan dengan menggunakan Oversampling.

Keterangan pada table percobaan:

|  |  |
| --- | --- |
|  | = Tabel percobaan yang menggunakan dataset myPersonality |
|  | = Tabel percobaan yang menggunakan dataset manual gathering |
|  | = Tabel percobaan yang menggunakan dataset gabungan |
|  |  |
| O | = Traits Openness |
| C | = Traits Conscientiousness |
| E | = Traits Extraversion |
| A | = Traits Agreeableness |
| N | = Traits Neuroticism |

* + - 1. **myPersonality dataset**
         1. **Skenario 1 (Percobaan tanpa menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.4** Hasil Percobaan Skenario 1 tanpa Features Selection dan tanpa Resampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 64.80% | 50.00% | 51.60% | **54.80%** | 54.40% |
| SVM | 70.00% | 50.00% | 60.00% | 46.80% | 58.00% |
| Logistic Regression | **70.40%** | 50.00% | 61.60% | 51.20% | 60.40% |
| GRADIENT BOOSTING | 63.20% | 53.20% | 57.60% | 53.60% | 56.80% |
| LDA | 62.00% | **54.00%** | **64.40%** | 53.60% | **60.80%** |
| SPLICE | NAÏVE BAYES | 59.20% | **52.00%** | 53.60% | 53.60% | 47.20% |
| SVM | 62.00% | 51.60% | 53.20% | 54.00% | 50.40% |
| Logistic Regression | 63.60% | 51.60% | **58.00%** | 51.60% | 54.40% |
| GRADIENT BOOSTING | 62.00% | 50.00% | 54.40% | 50.80% | 54.00% |
| LDA | **64.00%** | 51.60% | 54.00% | **58.00%** | **55.20%** |
| SNA | NAÏVE BAYES | 58.00% | 54.80% | **68.80%** | 45.60% | 52.40% |
| SVM | 58.80% | 53.60% | 51.60% | 42.00% | 38.80% |
| Logistic Regression | **70.00%** | 54.00% | 68.40% | 52.00% | **58.40%** |
| GRADIENT BOOSTING | 63.20% | 48.80% | 68.00% | **55.20%** | 54.00% |
| LDA | 69.20% | **56.40%** | 66.00% | 54.80% | 57.60% |
| LIWC + SPLICE | NAÏVE BAYES | 64.80% | 50.80% | 53.60% | **54.00%** | 46.80% |
| SVM | 59.60% | 50.80% | 54.00% | 51.20% | 50.80% |
| Logistic Regression | 63.60% | 52.40% | 56.80% | 51.20% | 54.00% |
| GRADIENT BOOSTING | **65.60%** | **54.40%** | 54.80% | 47.60% | **60.00%** |
| LDA | 58.00% | 54.00% | **63.20%** | 49.60% | 50.00% |
| LIWC + SNA | NAÏVE BAYES | 58.00% | **55.20%** | **68.80%** | 45.60% | 52.40% |
| SVM | 58.80% | 53.60% | 51.60% | 41.20% | 38.80% |
| Logistic Regression | **70.00%** | 54.00% | 68.40% | 51.20% | 59.20% |
| GRADIENT BOOSTING | 64.80% | 50.40% | 60.00% | **55.20%** | **60.40%** |
| LDA | 62.40% | 53.60% | 66.40% | 52.80% | 59.60% |
| SPLICE + SNA | NAÏVE BAYES | 40.00% | 49.60% | **64.00%** | 51.20% | 44.00% |
| SVM | 62.80% | 50.80% | 63.20% | **54.00%** | 47.20% |
| Logistic Regression | **68.40%** | **54.80%** | 63.60% | 45.20% | **59.20%** |
| GRADIENT BOOSTING | 64.80% | 50.00% | 62.40% | 48.40% | 55.60% |
| LDA | 60.80% | 53.20% | 56.40% | 53.20% | 56.40% |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 40.00% | 49.60% | 64.00% | 51.20% | 44.00% |
| SVM | 62.80% | 50.80% | 63.20% | 54.00% | 47.20% |
| Logistic Regression | **68.40%** | **56.40%** | **65.60%** | 46.80% | 57.60% |
| GRADIENT BOOSTING | 63.60% | 53.60% | 60.80% | **52.80%** | **60.00%** |
| LDA | 60.40% | 54.40% | 63.20% | 47.60% | 52.40% |

* + - * 1. **Skenario 2 (Percobaan dengan menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.5** Hasil Percobaan Skenario 2 dengan Features Selection

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 70.00% | 48.80% | 59.60% | 47.60% | 51.20% |
| SVM | **70.40%** | **56.00%** | 61.60% | 52.40% | **60.40%** |
| Logistic Regression | **70.40%** | 53.60% | 61.60% | 53.60% | **60.40%** |
| GRADIENT BOOSTING | 62.00% | 54.40% | 60.80% | **56.40%** | 58.80% |
| LDA | 69.60% | 52.80% | **62.80%** | 53.60% | **60.40%** |
| SPLICE | NAÏVE BAYES | 59.20% | 51.60% | 53.60% | **54.00%** | 49.60% |
| SVM | 44.00% | 48.40% | 51.60% | 51.20% | 54.80% |
| Logistic Regression | **65.60%** | 52.00% | **56.40%** | 52.00% | **57.60%** |
| GRADIENT BOOSTING | 63.60% | 50.80% | 54.40% | 51.20% | 53.60% |
| LDA | 62.40% | **54.00%** | 53.60% | **54.00%** | **57.60%** |
| SNA | NAÏVE BAYES | 58.00% | **54.80%** | **68.80%** | 45.60% | 52.40% |
| SVM | 58.80% | 53.60% | 51.60% | 41.60% | 38.80% |
| Logistic Regression | **70.00%** | 53.60% | 68.40% | 50.80% | 57.60% |
| GRADIENT BOOSTING | 61.20% | 48.40% | 68.00% | 50.40% | **59.20%** |
| LDA | **70.00%** | 52.40% | 68.00% | **56.80%** | 58.40% |
| LIWC + SPLICE | NAÏVE BAYES | 60.80% | 52.00% | 52.80% | **53.60%** | 48.80% |
| SVM | 45.20% | 51.60% | 54.00% | 50.40% | 54.00% |
| Logistic Regression | **65.20%** | 52.00% | 56.40% | 51.60% | **56.80%** |
| GRADIENT BOOSTING | 63.20% | 52.00% | **58.00%** | 50.80% | 53.60% |
| LDA | 62.00% | **55.20%** | 55.60% | 52.80% | 56.40% |
| LIWC + SNA | NAÏVE BAYES | 58.00% | **54.80%** | **68.80%** | 45.60% | 52.40% |
| SVM | 58.80% | 53.60% | 51.60% | 41.60% | 38.80% |
| Logistic Regression | **70.00%** | 54.00% | 68.40% | 52.00% | 58.00% |
| GRADIENT BOOSTING | 62.40% | 56.00% | 66.80% | **60.40%** | **57.20%** |
| LDA | 68.80% | 51.60% | 68.40% | 54.80% | 56.40% |
| SPLICE + SNA | NAÏVE BAYES | 40.40% | 50.00% | 64.80% | 52.00% | 44.40% |
| SVM | 62.80% | 50.80% | 63.20% | **56.80%** | 46.40% |
| Logistic Regression | **66.40%** | **54.80%** | **65.20%** | 44.80% | 58.00% |
| GRADIENT BOOSTING | 65.20% | 50.80% | 61.60% | 52.80% | 54.00% |
| LDA | 64.00% | 52.80% | 60.40% | 54.00% | **59.20%** |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 40.40% | 50.00% | **64.80%** | 52.00% | 44.00% |
| SVM | 62.80% | 50.80% | 63.20% | **56.80%** | 47.20% |
| Logistic Regression | **68.80%** | 54.80% | 63.60% | 46.00% | 57.60% |
| GRADIENT BOOSTING | 61.60% | 50.00% | 60.40% | 50.00% | **60.00%** |
| LDA | 61.60% | **56.40%** | 60.00% | 53.60% | 52.40% |

* + - * 1. **Skenario 3 (Percobaan tanpa menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.6** Hasil Percobaan Skenario 3 dengan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 35.60% | 50.00% | 46.80% | **55.60%** | 49.20% |
| SVM | 44.00% | 51.60% | 53.60% | 46.80% | 48.40% |
| Logistic Regression | 43.20% | **54.40%** | 49.60% | 48.40% | 52.40% |
| GRADIENT BOOSTING | 44.00% | 51.60% | 55.60% | 52.80% | 53.60% |
| LDA | **49.60%** | 54.00% | **56.00%** | 52.40% | **58.00%** |
| SPLICE | NAÏVE BAYES | 37.60% | 50.00% | 42.40% | **54.00%** | 43.60% |
| SVM | 39.60% | 51.60% | 48.00% | 50.40% | 50.80% |
| Logistic Regression | 45.20% | 50.80% | **51.60%** | 50.80% | 50.40% |
| GRADIENT BOOSTING | 43.20% | 49.20% | 44.80% | 49.60% | **54.00%** |
| LDA | **53.60%** | **52.40%** | **51.60%** | 52.80% | 52.80% |
| SNA | NAÏVE BAYES | 38.00% | 57.60% | **68.00%** | 48.00% | 50.80% |
| SVM | **46.40%** | 51.20% | 54.80% | 40.00% | 42.00% |
| Logistic Regression | 41.20% | 52.00% | 66.40% | 52.80% | 58.00% |
| GRADIENT BOOSTING | 41.60% | 53.20% | 55.20% | **56.40%** | 53.60% |
| LDA | **46.40%** | **58.40%** | 56.40% | 53.60% | **60.00%** |
| LIWC + SPLICE | NAÏVE BAYES | 41.20% | 50.40% | 45.60% | **55.60%** | 45.60% |
| SVM | 44.00% | 53.60% | 46.00% | 41.60% | **54.40%** |
| Logistic Regression | 45.20% | 51.60% | 49.60% | 49.60% | 50.00% |
| GRADIENT BOOSTING | 46.80% | **55.20%** | 48.80% | 50.00% | 51.60% |
| LDA | **52.00%** | 52.40% | **57.60%** | 50.40% | 54.00% |
| LIWC + SNA | NAÏVE BAYES | 38.00% | **57.60%** | **68.00%** | 48.00% | 50.80% |
| SVM | 47.20% | 57.20% | 61.20% | 42.40% | 45.60% |
| Logistic Regression | 42.00% | 52.00% | 66.40% | 52.80% | **58.40%** |
| GRADIENT BOOSTING | 48.40% | 50.80% | 51.20% | 50.40% | 57.60% |
| LDA | **48.80%** | 54.80% | 56.80% | **54.00%** | 50.80% |
| SPLICE + SNA | NAÏVE BAYES | 40.80% | 51.60% | 52.80% | 52.40% | 42.00% |
| SVM | 44.80% | 54.00% | 52.00% | 50.40% | 52.00% |
| Logistic Regression | 48.40% | **55.20%** | **60.00%** | 50.80% | 56.40% |
| GRADIENT BOOSTING | 47.60% | 51.20% | 55.60% | 53.20% | **56.80%** |
| LDA | **53.20%** | 54.80% | 50.80% | **54.80%** | 53.60% |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 42.40% | 51.60% | 52.80% | 52.40% | 42.00% |
| SVM | 44.40% | 49.60% | 54.80% | 51.60% | 50.00% |
| Logistic Regression | 45.20% | 51.60% | **58.80%** | **55.60%** | 54.00% |
| GRADIENT BOOSTING | 50.00% | 49.60% | 55.60% | 49.60% | **56.80%** |
| LDA | **54.40%** | **54.80%** | 54.00% | 48.80% | 46.80% |

* + - * 1. **Skenario 4 (Percobaan tanpa menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.7** Hasil Percobaan Skenario 4 dengan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 48.80% | 49.20% | 50.00% | 52.00% | 50.80% |
| SVM | 54.80% | 50.40% | 52.80% | 49.20% | 53.20% |
| Logistic Regression | 56.00% | 53.20% | 50.80% | 48.40% | 53.60% |
| GRADIENT BOOSTING | **57.20%** | **56.00%** | 57.20% | 51.60% | 57.60% |
| LDA | 55.20% | 54.80% | **60.00%** | **53.20%** | **58.80%** |
| SPLICE | NAÏVE BAYES | 41.20% | **52.00%** | 52.00% | **56.40%** | 43.20% |
| SVM | 49.60% | 49.60% | 49.60% | 48.40% | 54.00% |
| Logistic Regression | 54.00% | 51.20% | **54.80%** | 51.60% | 48.80% |
| GRADIENT BOOSTING | **60.80%** | 50.40% | 51.60% | 54.40% | **56.80%** |
| LDA | 57.60% | **52.00%** | 52.00% | **56.40%** | 55.60% |
| SNA | NAÏVE BAYES | 37.20% | **59.20%** | **65.60%** | 44.00% | 49.60% |
| SVM | 50.40% | 55.20% | 47.20% | 45.20% | 38.40% |
| Logistic Regression | 45.20% | 54.40% | 61.20% | 49.20% | **58.00%** |
| GRADIENT BOOSTING | **57.20%** | 48.80% | 64.80% | **52.80%** | 55.20% |
| LDA | 52.40% | 58.40% | 59.60% | 52.00% | 57.20% |
| LIWC + SPLICE | NAÏVE BAYES | 46.40% | 50.80% | 51.20% | **56.40%** | 41.60% |
| SVM | 50.00% | 52.00% | 52.40% | 46.00% | 51.60% |
| Logistic Regression | 54.80% | 50.00% | 57.20% | 51.20% | 49.60% |
| GRADIENT BOOSTING | **62.80%** | 51.20% | 54.00% | 51.60% | **59.20%** |
| LDA | 54.40% | **54.40%** | **61.20%** | 49.60% | 48.80% |
| LIWC + SNA | NAÏVE BAYES | 37.20% | **59.20%** | **65.60%** | 44.40% | 49.60% |
| SVM | 48.40% | 55.20% | 46.80% | 45.20% | 38.40% |
| Logistic Regression | 46.40% | 54.40% | 61.20% | 48.80% | **58.40%** |
| GRADIENT BOOSTING | 56.40% | 54.80% | 57.20% | 53.20% | 57.20% |
| LDA | **56.80%** | 54.40% | 62.80% | **53.60%** | 55.20% |
| SPLICE + SNA | NAÏVE BAYES | 41.20% | 50.40% | 60.00% | 49.20% | 43.20% |
| SVM | 50.80% | 51.20% | 52.40% | 48.40% | 52.00% |
| Logistic Regression | 54.00% | 53.60% | 59.60% | 49.20% | **57.20%** |
| GRADIENT BOOSTING | **56.80%** | 46.80% | **60.80%** | 51.60% | 55.60% |
| LDA | **56.80%** | **55.60%** | 53.20% | **54.00%** | 52.00% |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 41.20% | 50.40% | 60.00% | 49.20% | 43.20% |
| SVM | 50.80% | 51.20% | 52.40% | 48.40% | 52.00% |
| Logistic Regression | 55.60% | 53.60% | 61.60% | 48.80% | 59.20% |
| GRADIENT BOOSTING | **61.20%** | 53.20% | 58.00% | **53.60%** | **60.40%** |
| LDA | 56.40% | **55.60%** | **64.40%** | 50.40% | 50.40% |

* + - * 1. **Skenario 5 (Percobaan dengan menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.8** Hasil Percobaan Skenario 5 dengan Features Selection dan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 48.00% | 48.80% | 57.20% | 46.80% | 36.80% |
| SVM | 55.20% | 52.40% | 54.00% | 53.60% | **54.40%** |
| Logistic Regression | **56.40%** | 50.80% | 54.80% | 53.20% | 52.00% |
| GRADIENT BOOSTING | 42.40% | **53.20%** | 46.00% | **63.20%** | 46.80% |
| LDA | 50.40% | **53.20%** | **59.20%** | 52.80% | 50.00% |
| SPLICE | NAÏVE BAYES | 40.00% | 50.80% | 42.00% | 54.00% | 45.60% |
| SVM | 50.00% | **55.20%** | **52.40%** | **56.80%** | 41.60% |
| Logistic Regression | 45.20% | 53.60% | 50.80% | 51.20% | 48.80% |
| GRADIENT BOOSTING | 46.80% | 51.20% | 47.60% | 52.80% | 54.00% |
| LDA | **51.20%** | 51.60% | 51.60% | 52.00% | **54.40%** |
| SNA | NAÏVE BAYES | 38.00% | **57.60%** | **67.60%** | 48.00% | 50.80% |
| SVM | **44.40%** | 53.60% | 57.60% | 43.60% | 46.00% |
| Logistic Regression | 41.60% | 52.00% | 66.00% | 52.80% | **58.80%** |
| GRADIENT BOOSTING | 43.60% | 50.80% | 62.80% | **55.20%** | 52.80% |
| LDA | 43.60% | 54.40% | 58.40% | 54.80% | 56.80% |
| LIWC + SPLICE | NAÏVE BAYES | 40.00% | 50.40% | 42.00% | 54.00% | 46.40% |
| SVM | 46.80% | 52.40% | 52.80% | **54.80%** | 45.20% |
| Logistic Regression | 44.80% | **52.80%** | 50.00% | 52.00% | 48.80% |
| GRADIENT BOOSTING | 45.20% | 48.80% | 46.00% | 54.00% | **54.80%** |
| LDA | **53.20%** | **52.80%** | **55.20%** | 52.80% | 51.20% |
| LIWC + SNA | NAÏVE BAYES | 38.00% | **57.60%** | **68.00%** | 48.00% | 50.80% |
| SVM | **47.20%** | 49.60% | 51.60% | 43.20% | 45.60% |
| Logistic Regression | 42.80% | 52.00% | 66.00% | 52.80% | **58.00%** |
| GRADIENT BOOSTING | **47.20%** | 55.60% | 59.20% | **56.80%** | 55.20% |
| LDA | 46.00% | 54.40% | 63.60% | 54.80% | 55.60% |
| SPLICE + SNA | NAÏVE BAYES | 41.60% | 51.60% | 53.20% | 52.80% | 42.40% |
| SVM | 44.40% | 51.60% | 58.80% | 52.00% | 50.80% |
| Logistic Regression | 44.00% | **54.80%** | **61.60%** | 52.80% | **57.20%** |
| GRADIENT BOOSTING | **48.40%** | 50.40% | 56.00% | **54.40%** | 56.80% |
| LDA | 48.00% | 52.00% | 53.20% | 53.60% | 55.60% |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 41.60% | 51.60% | 53.20% | 52.80% | 42.00% |
| SVM | 44.40% | 52.00% | 56.80% | 49.20% | 50.00% |
| Logistic Regression | 49.20% | 54.80% | **61.20%** | 51.60% | 54.00% |
| GRADIENT BOOSTING | **52.00%** | 50.80% | 53.60% | 53.20% | **56.80%** |
| LDA | 49.20% | **55.60%** | 55.20% | **54.80%** | 46.80% |

* + - * 1. **Skenario 6 (Percobaan dengan menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.9** Hasil Percobaan Skenario 6 dengan Features Selection

dan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 54.40% | 48.40% | 56.00% | 47.20% | 37.60% |
| SVM | 55.20% | 52.80% | 54.80% | 52.80% | 51.60% |
| Logistic Regression | 54.80% | 50.00% | 54.00% | 53.60% | 50.80% |
| GRADIENT BOOSTING | **59.20%** | **56.40%** | 51.20% | **59.60%** | **53.20%** |
| LDA | 51.20% | 54.80% | **59.20%** | 55.20% | 50.40% |
| SPLICE | NAÏVE BAYES | 41.20% | 51.20% | 50.80% | **56.00%** | 44.00% |
| SVM | 44.40% | 50.40% | 54.40% | 52.40% | 48.40% |
| Logistic Regression | 51.60% | 53.20% | **55.20%** | 52.00% | 50.80% |
| GRADIENT BOOSTING | **60.40%** | **54.00%** | 54.00% | 52.40% | 55.20% |
| LDA | 58.80% | 53.60% | 54.00% | 51.20% | **57.20%** |
| SNA | NAÏVE BAYES | 37.20% | **59.20%** | 65.60% | 44.00% | 49.60% |
| SVM | 50.40% | 55.20% | 49.60% | 45.20% | 38.40% |
| Logistic Regression | 45.20% | 54.40% | 61.20% | 48.80% | 58.00% |
| GRADIENT BOOSTING | **55.20%** | 50.00% | **66.80%** | **52.00%** | 56.40% |
| LDA | 45.60% | 54.00% | 61.20% | 51.20% | **58.80%** |
| LIWC + SPLICE | NAÏVE BAYES | 41.60% | 52.00% | 53.20% | **56.00%** | 43.60% |
| SVM | 53.60% | 54.00% | 54.40% | 48.80% | 44.80% |
| Logistic Regression | 52.40% | 52.80% | 55.20% | 51.20% | 50.80% |
| GRADIENT BOOSTING | **58.80%** | 50.80% | 54.00% | 54.00% | **53.20%** |
| LDA | 53.60% | **56.00%** | **56.00%** | 52.80% | 52.80% |
| LIWC + SNA | NAÏVE BAYES | 37.20% | **59.20%** | **65.60%** | 44.00% | 49.60% |
| SVM | 50.40% | 55.20% | 49.60% | 45.20% | 38.40% |
| Logistic Regression | 45.20% | 54.00% | 61.20% | 49.20% | **58.00%** |
| GRADIENT BOOSTING | **56.40%** | 57.60% | 59.60% | **59.60%** | 57.60% |
| LDA | 55.60% | 52.40% | 62.80% | 54.40% | 52.80% |
| SPLICE + SNA | NAÏVE BAYES | 41.20% | 50.40% | 60.80% | 50.00% | 43.60% |
| SVM | 50.40% | 51.20% | 54.40% | 49.20% | 49.60% |
| Logistic Regression | 56.80% | **54.80%** | **61.20%** | 48.80% | **58.40%** |
| GRADIENT BOOSTING | **61.20%** | 48.00% | 59.20% | **54.80%** | 56.00% |
| LDA | 56.00% | 54.40% | 58.00% | 52.40% | 57.60% |
| LIWC + SPLICE + SNA | NAÏVE BAYES | 41.20% | 50.40% | **60.80%** | 50.00% | 43.20% |
| SVM | 50.40% | 51.20% | 54.40% | 49.20% | 52.00% |
| Logistic Regression | 56.40% | 55.20% | **60.80%** | 46.40% | 59.20% |
| GRADIENT BOOSTING | **58.00%** | 51.60% | 58.80% | 52.00% | **60.40%** |
| LDA | 56.00% | **56.40%** | 58.00% | **52.80%** | 50.40% |

* + - 1. **Manual gathering dataset**
         1. **Skenario 7 (Percobaan tanpa menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.10** Hasil Percobaan Skenario 7 tanpa Features Selection

dan tanpa Resampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 60.67% | 62.67% | 73.33% | 52.00% | 59.33% |
| SVM | 62.67% | 65.33% | 76.00% | 56.67% | **69.33%** |
| Logistic Regression | 64.00% | 57.33% | 74.00% | **58.00%** | 66.67% |
| GRADIENT BOOSTING | **67.33%** | 62.67% | 70.67% | **58.00%** | 66.67% |
| LDA | 58.00% | **67.33%** | **79.33%** | 57.33% | 57.33% |
| SPLICE | NAÏVE BAYES | 42.00% | 44.00% | **70.67%** | 47.33% | 43.33% |
| SVM | 60.67% | 58.00% | 46.67% | 56.67% | 43.33% |
| Logistic Regression | **65.33%** | **66.00%** | 66.00% | 55.33% | 56.00% |
| GRADIENT BOOSTING | 58.67% | 57.33% | 66.67% | **58.67%** | **58.67%** |
| LDA | 58.00% | 61.33% | 61.33% | 54.00% | 58.00% |
| LIWC + SPLICE | NAÏVE BAYES | 42.00% | 44.00% | 69.33% | 46.00% | 44.00% |
| SVM | 58.00% | 55.33% | 47.33% | 58.00% | 54.00% |
| Logistic Regression | 66.00% | **66.00%** | 65.33% | 56.00% | 57.33% |
| GRADIENT BOOSTING | **67.33%** | 61.33% | **72.00%** | 56.00% | **65.33%** |
| LDA | 54.00% | 47.33% | 48.67% | **62.67%** | 56.00% |

* + - * 1. **Skenario 8 (Percobaan dengan menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.11** Hasil Percobaan Skenario 8 dengan Features Selection

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 52.67% | 54.67% | 72.67% | 49.33% | **70.00%** |
| SVM | **64.67%** | 57.33% | 74.67% | 53.33% | 66.67% |
| Logistic Regression | **64.67%** | **58.00%** | 74.67% | **54.00%** | 66.67% |
| GRADIENT BOOSTING | 51.33% | 48.67% | 76.00% | 52.00% | 61.33% |
| LDA | 60.00% | 57.33% | **78.00%** | **54.00%** | 66.67% |
| SPLICE | NAÏVE BAYES | 41.33% | 44.00% | **70.00%** | 46.67% | 40.67% |
| SVM | 60.67% | 54.00% | 52.00% | 55.33% | 54.67% |
| Logistic Regression | **65.33%** | **66.67%** | 66.00% | **59.33%** | 57.33% |
| GRADIENT BOOSTING | 61.33% | 54.67% | 68.00% | 55.33% | 57.33% |
| LDA | 56.67% | 62.00% | 64.00% | 54.00% | **62.00%** |
| LIWC + SPLICE | NAÏVE BAYES | 41.33% | 44.00% | 70.00% | 46.67% | 40.67% |
| SVM | 60.67% | 54.00% | 58.00% | 55.33% | 50.67% |
| Logistic Regression | **66.67%** | **66.67%** | 66.00% | **59.33%** | 56.67% |
| GRADIENT BOOSTING | 62.00% | 56.67% | 72.67% | 58.00% | **64.67%** |
| LDA | 54.67% | 63.33% | **74.00%** | 53.33% | 64.00% |

* + - * 1. **Skenario 9 (Percobaan tanpa menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.12** Hasil Percobaan Skenario 9 dengan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 52.67% | 58.00% | **67.33%** | 50.67% | 46.00% |
| SVM | 55.33% | 58.67% | 56.67% | 56.00% | **59.33%** |
| Logistic Regression | **62.00%** | 54.00% | 50.00% | 57.33% | 52.67% |
| GRADIENT BOOSTING | 53.33% | **62.00%** | 52.67% | 55.33% | 50.67% |
| LDA | 54.67% | 58.67% | 58.00% | **60.67%** | 52.00% |
| SPLICE | NAÏVE BAYES | 41.33% | 44.67% | **60.00%** | 47.33% | 40.67% |
| SVM | 46.00% | 51.33% | 56.00% | 56.00% | 53.33% |
| Logistic Regression | **62.67%** | **62.00%** | 59.33% | **58.00%** | 52.00% |
| GRADIENT BOOSTING | 56.00% | 57.33% | 53.33% | 56.00% | **58.00%** |
| LDA | 52.00% | 57.33% | 51.33% | 50.67% | 52.67% |
| LIWC + SPLICE | NAÏVE BAYES | 40.67% | 44.67% | **62.67%** | 46.00% | 42.00% |
| SVM | 51.33% | 53.33% | 56.00% | 51.33% | 51.33% |
| Logistic Regression | **61.33%** | 62.00% | 58.00% | 57.33% | 53.33% |
| GRADIENT BOOSTING | 58.67% | **66.00%** | 54.67% | **59.33%** | **59.33%** |
| LDA | 50.00% | 64.67% | 62.00% | 46.00% | 57.33% |

* + - * 1. **Skenario 10 (Percobaan tanpa menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.13** Hasil Percobaan Skenario 10 dengan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 52.67% | 60.67% | 73.33% | 53.33% | 46.00% |
| SVM | 62.00% | 62.00% | 68.00% | 56.00% | **64.00%** |
| Logistic Regression | 61.33% | 60.00% | 67.33% | 55.33% | 60.67% |
| GRADIENT BOOSTING | **67.33%** | 62.67% | 72.00% | 57.33% | 62.67% |
| LDA | 59.33% | **64.00%** | **78.00%** | **59.33%** | 57.33% |
| SPLICE | NAÏVE BAYES | 43.33% | 44.00% | 58.00% | 47.33% | 42.67% |
| SVM | 55.33% | 60.67% | 52.67% | 57.33% | 44.67% |
| Logistic Regression | **62.67%** | **64.67%** | 62.67% | 56.67% | 57.33% |
| GRADIENT BOOSTING | 62.00% | 60.67% | **65.33%** | **58.00%** | **58.67%** |
| LDA | 56.00% | 63.33% | 54.67% | 52.00% | 54.67% |
| LIWC + SPLICE | NAÏVE BAYES | 42.67% | 43.33% | 63.33% | 46.67% | 40.67% |
| SVM | 60.67% | 60.67% | 60.00% | 56.00% | 44.67% |
| Logistic Regression | 63.33% | 66.00% | 61.33% | 56.00% | 58.00% |
| GRADIENT BOOSTING | **64.67%** | **66.67%** | **73.33%** | 55.33% | **67.33%** |
| LDA | 52.67% | 46.67% | 48.67% | **62.67%** | 56.67% |

* + - * 1. **Skenario 11 (Percobaan dengan menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.14** Hasil Percobaan Skenario 11 dengan Features Selection dan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 43.33% | 52.67% | 51.33% | 52.67% | 60.00% |
| SVM | 47.33% | 50.67% | **72.67%** | 55.33% | **64.00%** |
| Logistic Regression | **48.67%** | 52.67% | 69.33% | 53.33% | **64.00%** |
| GRADIENT BOOSTING | 46.67% | **53.33%** | 64.00% | 52.67% | 55.33% |
| LDA | 43.33% | 49.33% | 68.00% | **54.00%** | **64.00%** |
| SPLICE | NAÏVE BAYES | 41.33% | 43.33% | **60.00%** | 46.00% | 40.67% |
| SVM | 54.67% | 54.00% | 50.67% | **60.67%** | 46.00% |
| Logistic Regression | **62.00%** | **62.67%** | 52.00% | 58.67% | 52.67% |
| GRADIENT BOOSTING | 58.00% | 61.33% | 51.33% | 52.00% | **53.33%** |
| LDA | 56.67% | 60.00% | 48.00% | 50.00% | **55.33%** |
| LIWC + SPLICE | NAÏVE BAYES | 41.33% | 43.33% | 60.67% | 46.00% | 40.67% |
| SVM | 54.00% | 54.00% | 56.00% | 57.33% | 48.67% |
| Logistic Regression | **61.33%** | 62.67% | 50.67% | 58.67% | 53.33% |
| GRADIENT BOOSTING | 52.67% | 56.00% | 59.33% | **60.67%** | 62.00% |
| LDA | 54.67% | **64.67%** | **64.00%** | 55.33% | **64.00%** |

* + - * 1. **Skenario 12 (Percobaan dengan menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.15** Hasil Percobaan Skenario 12 dengan Features Selection

dan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 45.33% | **54.00%** | 66.00% | 50.00% | 62.67% |
| SVM | 47.33% | 48.00% | 68.67% | 56.67% | 62.00% |
| Logistic Regression | 47.33% | 49.33% | 66.00% | **59.33%** | 62.67% |
| GRADIENT BOOSTING | **53.33%** | 48.67% | 70.00% | 52.67% | 59.33% |
| LDA | 42.67% | 51.33% | **70.67%** | 52.00% | **63.33%** |
| SPLICE | NAÏVE BAYES | 42.67% | 42.67% | 57.33% | 46.67% | 40.00% |
| SVM | 54.00% | 58.00% | 53.33% | 54.00% | 41.33% |
| Logistic Regression | **63.33%** | **66.67%** | 59.33% | **59.33%** | 56.67% |
| GRADIENT BOOSTING | 60.67% | 62.00% | **66.67%** | 54.67% | **60.67%** |
| LDA | 54.67% | 62.00% | 53.33% | 54.67% | 56.00% |
| LIWC + SPLICE | NAÏVE BAYES | 43.33% | 42.67% | 58.00% | 47.33% | 40.00% |
| SVM | 50.00% | 60.00% | 56.67% | 56.00% | 41.33% |
| Logistic Regression | **66.00%** | **66.00%** | 59.33% | 58.00% | 57.33% |
| GRADIENT BOOSTING | **66.00%** | 58.67% | **72.00%** | **60.67%** | 60.00% |
| LDA | 52.00% | 64.00% | 70.67% | 56.67% | **63.33%** |

* + - 1. **Combined Dataset**
         1. **Skenario 13 (Percobaan tanpa menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.16** Hasil Percobaan Skenario 13 tanpa Features Selection

dan tanpa Resampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 61.50% | 55.50% | 62.00% | 52.00% | 49.25% |
| SVM | 67.25% | **58.75%** | 66.50% | 54.50% | **62.50%** |
| Logistic Regression | **68.50%** | 54.25% | 66.25% | 51.00% | **62.50%** |
| GRADIENT BOOSTING | 65.25% | 57.25% | 64.25% | 55.75% | 59.75% |
| LDA | 62.50% | 58.00% | **68.50%** | **56.00%** | 55.00% |
| SPLICE | NAÏVE BAYES | 34.50% | 50.25% | 61.25% | 46.50% | 40.75% |
| SVM | 44.75% | 51.50% | 52.00% | **52.50%** | 48.75% |
| Logistic Regression | 63.00% | 52.25% | **63.25%** | 48.50% | 61.25% |
| GRADIENT BOOSTING | 61.50% | 54.50% | 60.75% | 50.00% | 57.25% |
| LDA | **64.25%** | **56.25%** | 60.00% | 51.25% | **63.75%** |
| LIWC + SPLICE | NAÏVE BAYES | 35.00% | 50.50% | 61.50% | 46.00% | 40.00% |
| SVM | 58.25% | 49.25% | 58.00% | 48.75% | 52.25% |
| Logistic Regression | **63.50%** | 55.00% | 62.50% | 47.75% | **63.75%** |
| GRADIENT BOOSTING | **63.50%** | 56.25% | 62.50% | **56.50%** | 59.50% |
| LDA | 58.00% | **59.25%** | **64.25%** | 54.75% | 59.25% |

* + - * 1. **Skenario 14 (Percobaan dengan menggunakan Features Selection dan tanpa Resampling)**

**Tabel 4.17** Hasil Percobaan Skenario 14 dengan Features Selection

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 65.50% | 51.25% | 63.50% | 45.50% | 57.00% |
| SVM | 68.00% | 52.25% | 66.25% | 52.50% | 62.25% |
| Logistic Regression | **68.25%** | 50.25% | **66.50%** | 53.75% | **62.75%** |
| GRADIENT BOOSTING | 62.00% | 52.50% | 59.50% | 54.75% | 56.25% |
| LDA | 68.00% | **54.25%** | 66.25% | **55.25%** | 61.50% |
| SPLICE | NAÏVE BAYES | 34.25% | 51.00% | 60.75% | 46.25% | 39.50% |
| SVM | 58.25% | 54.75% | 54.75% | 50.25% | 49.00% |
| Logistic Regression | 63.00% | 54.25% | 64.50% | 49.25% | 62.00% |
| GRADIENT BOOSTING | 62.25% | 54.25% | 58.50% | **51.75%** | 59.50% |
| LDA | **63.50%** | **58.50%** | **61.50%** | 51.50% | **65.25%** |
| LIWC + SPLICE | NAÏVE BAYES | 34.00% | 51.00% | 61.25% | 46.00% | 39.25% |
| SVM | 53.50% | 51.50% | 50.50% | **53.50%** | 48.25% |
| Logistic Regression | **63.25%** | 55.00% | **64.25%** | 49.50% | 62.00% |
| GRADIENT BOOSTING | 63.00% | 55.75% | 57.00% | 53.25% | 57.25% |
| LDA | 62.50% | **58.75%** | 63.00% | 52.00% | **65.00%** |

* + - * 1. **Skenario 15 (Percobaan tanpa menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.18** Hasil Percobaan Skenario 15 dengan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 39.75% | 55.50% | 49.25% | 54.25% | 39.50% |
| SVM | 42.50% | 58.75% | 52.25% | 52.75% | **52.50%** |
| Logistic Regression | 46.25% | 58.50% | 50.75% | 51.50% | 50.00% |
| GRADIENT BOOSTING | **49.00%** | 55.75% | 52.00% | **56.75%** | 50.50% |
| LDA | 46.00% | **59.50%** | **59.75%** | 56.00% | 49.50% |
| SPLICE | NAÏVE BAYES | 36.75% | 50.50% | **59.75%** | 46.75% | 41.00% |
| SVM | 47.25% | 51.25% | 45.00% | 49.50% | 53.50% |
| Logistic Regression | 48.25% | 52.00% | 52.50% | 50.75% | 51.50% |
| GRADIENT BOOSTING | 47.50% | **56.00%** | 47.25% | **52.00%** | 50.75% |
| LDA | **49.50%** | 55.25% | 49.25% | 51.75% | **54.75%** |
| LIWC + SPLICE | NAÏVE BAYES | 37.75% | 50.50% | **61.00%** | 46.75% | 41.00% |
| SVM | 49.00% | 50.50% | 48.25% | 51.25% | 51.00% |
| Logistic Regression | 47.50% | 54.25% | 52.25% | 50.00% | 53.75% |
| GRADIENT BOOSTING | 48.00% | **58.25%** | 56.25% | **56.00%** | 53.75% |
| LDA | **54.50%** | 57.75% | 60.00% | 53.75% | **56.00%** |

* + - * 1. **Skenario 16 (Percobaan tanpa menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.19** Hasil Percobaan Skenario 16 dengan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | 50.25% | 55.00% | 55.00% | 53.25% | 41.00% |
| SVM | **57.25%** | **60.75%** | 58.50% | 52.75% | 54.00% |
| Logistic Regression | 56.50% | 57.50% | 53.50% | 47.75% | **56.50%** |
| GRADIENT BOOSTING | 57.00% | 59.25% | **65.25%** | **58.25%** | 54.25% |
| LDA | 57.00% | 59.25% | 64.75% | 56.50% | 53.00% |
| SPLICE | NAÏVE BAYES | 37.00% | 50.50% | **60.50%** | 46.25% | 39.50% |
| SVM | **61.25%** | 50.50% | 49.50% | 51.00% | 55.50% |
| Logistic Regression | 54.25% | 52.00% | 53.25% | 49.00% | 51.50% |
| GRADIENT BOOSTING | 60.00% | 54.75% | 56.50% | 48.50% | 56.00% |
| LDA | 54.25% | **55.25%** | 52.00% | **52.25%** | **56.75%** |
| LIWC + SPLICE | NAÏVE BAYES | 36.75% | 50.75% | 60.75% | 45.00% | 38.75% |
| SVM | 57.25% | 51.25% | 44.00% | **55.00%** | 56.25% |
| Logistic Regression | 55.00% | 54.75% | 56.50% | 49.75% | 52.25% |
| GRADIENT BOOSTING | **61.50%** | **59.75%** | **62.25%** | 54.75% | **57.75%** |
| LDA | 54.75% | 58.75% | 59.50% | 53.50% | 56.00% |

* + - * 1. **Skenario 17 (Percobaan dengan menggunakan Features Selection dan dengan Undersampling)**

**Tabel 4.20** Hasil Percobaan Skenario 17 dengan Features Selection

dan Undersampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | **55.75%** | 52.00% | **63.50%** | 45.75% | 37.50% |
| SVM | 51.75% | 50.75% | 58.25% | 52.75% | 51.25% |
| Logistic Regression | 51.75% | 49.75% | 57.00% | 54.50% | 52.00% |
| GRADIENT BOOSTING | 49.00% | 51.50% | 51.00% | 53.00% | **52.50%** |
| LDA | 49.50% | **53.50%** | 57.50% | **57.00%** | 51.75% |
| SPLICE | NAÏVE BAYES | 36.75% | 51.25% | **59.25%** | 47.25% | 41.00% |
| SVM | 49.75% | 50.25% | 53.50% | **54.25%** | 54.50% |
| Logistic Regression | 48.50% | 52.75% | 51.50% | 50.25% | 51.25% |
| GRADIENT BOOSTING | 47.25% | **56.25%** | 47.25% | 51.50% | 51.00% |
| LDA | **50.25%** | 56.00% | 48.00% | 50.50% | **55.50%** |
| LIWC + SPLICE | NAÏVE BAYES | 37.00% | 51.00% | **59.25%** | 47.00% | 40.75% |
| SVM | 46.75% | 53.25% | 43.50% | 51.25% | 54.50% |
| Logistic Regression | 47.75% | 53.50% | 51.50% | 50.75% | 51.50% |
| GRADIENT BOOSTING | 48.25% | 57.00% | 50.50% | **53.50%** | 50.75% |
| LDA | **48.50%** | **58.50%** | 50.00% | 52.25% | **58.25%** |

* + - * 1. **Skenario 18 (Percobaan dengan menggunakan Features Selection dan dengan Oversampling)**

**Tabel 4.21** Hasil Percobaan Skenario 18 dengan Features Selection

dan Oversampling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Algorithm** | **O** | **C** | **E** | **A** | **N** |
|
| LIWC | NAÏVE BAYES | **61.75%** | **52.25%** | **64.25%** | 45.50% | 43.25% |
| SVM | 52.75% | 50.50% | 60.00% | 55.50% | 53.75% |
| Logistic Regression | 54.25% | 49.25% | 57.25% | 54.50% | **56.25%** |
| GRADIENT BOOSTING | 56.50% | 51.50% | 53.00% | 55.75% | 51.50% |
| LDA | 53.00% | 52.00% | 59.25% | **58.00%** | 53.25% |
| SPLICE | NAÏVE BAYES | 37.25% | 51.00% | **60.50%** | 46.25% | 39.00% |
| SVM | 51.25% | 54.75% | 52.00% | 49.25% | 53.50% |
| Logistic Regression | 55.00% | 54.25% | 53.00% | **50.75%** | 53.75% |
| GRADIENT BOOSTING | **57.50%** | 55.00% | 54.25% | **50.75%** | 55.00% |
| LDA | 53.00% | **57.50%** | 54.00% | 50.25% | **56.50%** |
| LIWC + SPLICE | NAÏVE BAYES | 37.25% | 51.00% | **60.75%** | 46.00% | 38.50% |
| SVM | 50.25% | 50.25% | 57.25% | 53.25% | 51.75% |
| Logistic Regression | 54.50% | 55.50% | 54.50% | 51.00% | 54.75% |
| GRADIENT BOOSTING | **59.75%** | 54.50% | 57.00% | **53.75%** | 58.50% |
| LDA | 53.50% | **58.50%** | 54.00% | 51.00% | **59.25%** |

* + 1. **Implementasi Deep Learning**
       1. **Skenario**

Implementasi *deep learning* pada testing kali ini akan menggunakan 4 model arsitektur yaitu MultiLayer Perceptron (MLP), LSTM, CNN dan GRU. Fitur yang digunakan dalam implementasi *deep learning* hanya dengan metode *open vocabulary*. Proses *testing* pada *deep learning* juga akan diaplikasikan ke masing-masing dataset berbeda seperti halnya implementasi *deep learning* pada Bagian 4.2.2 diatas. Distribusi data yang akan digunakan dapat kembali di lihat di Bagian 4.2.1.

Dari faktor skenario diatas pada Bagian 4.2.2, *deep learning* hanya akan menggunakan satu faktor tersebut yaitu Resampling (Bagian 4.2.2.2) sehingga dapat dibentuk skenario percobaan untuk *deep learning* sebagai berikut:

* + 1. Penggunaan dataset myPersonality, percobaan tanpa menggunakan Resampling.
    2. Penggunaan dataset myPersonality, percobaan dengan menggunakan Undersampling.
    3. Penggunaan dataset myPersonality, percobaan dengan menggunakan Oversampling.
    4. Penggunaan dataset manual gathering, percobaan tanpa menggunakan Resampling.
    5. Penggunaan dataset manual gathering, percobaan dengan menggunakan Undersampling.
    6. Penggunaan dataset manual gathering, percobaan dengan menggunakan Oversampling.
    7. Penggunaan dataset gabungan, percobaan tanpa menggunakan Resampling.
    8. Penggunaan dataset gabungan, percobaan dengan menggunakan Undersampling.
    9. Penggunaan dataset gabungan, percobaan dengan menggunakan Oversampling.
       1. **myPersonality dataset**
       2. **Manual gathering dataset**
       3. **Combined dataset**
    10. **Tampilan Layar Aplikasi**
  1. **Evaluasi** 
     1. **Evaluasi Subjektif**
     2. **Evaluasi Objektif**