## 4. Evaluasi

### Hasil Pengujian

Hasil eksekusi dari masing-masing skenario pengujian dapat dilihat pada tabel 5. Perhitungan F1-Measure dan akurasi dibulatkan dengan 3 desimal. Algoritma dengan performa terbaik ditandai dengan *shading* berwarna hijau.

| Skenario | Algoritma | Feature Extraction | Stopwords | Random Over-Sampling | Akurasi & F1-Measure | Akurasi & F1-Measure optimized |
|---|---|---|---|---|---|---|
| 1 | SVM | TF-IDF | Default | N | 0.453<br>0.429 | 0.445<br>0.438 |
| | NB | TF-IDF | Default | N | 0.423<br>0.421 | 0.445<br>0.434 |
| 2 | SVM | TF-IDF | Modifikasi | N | 0.474<br>0.452 | 0.438<br>0.434 |
| | NB | TF-IDF | Modifikasi | N | 0.401<br>0.421 | 0.467<br>0.479 |
| 3 | SVM | TF-IDF | Default | Y | 0.799<br>0.798 | 0.839<br>0.837 |
| | NB | TF-IDF | Default | Y | 0.772<br>0.772 | 0.815<br>0.815 |
| 4 | SVM | TF-IDF | Modifikasi | Y | 0.819<br>0.818 | 0.842<br>0.842 |
| | NB | TF-IDF | Modifikasi | Y | 0.782<br>0.781 | 0.829<br>0.829 |
| 5 | SVM | TFPOS-IDF | Default | N | 0.438<br>0.43 | 0.438<br>0.43 |
| | NB | TFPOS-IDF | Default | N | 0.431<br>0.463 | 0.453<br>0.45 |
| 6 | SVM | TFPOS-IDF | Modifikasi | N | 0.445<br>0.431 | 0.504<br>0.491 |
| | NB | TFPOS-IDF | Modifikasi | N | 0.401<br>0.43 | 0.467<br>0.472 |
| 7 | SVM | TFPOS-IDF | Default | Y | 0.815<br>0.814 | 0.836<br>0.836 |
| | NB | TFPOS-IDF | Default | Y | 0.735<br>0.732 | 0.795<br>0.793 |
| 8 | SVM | TFPOS-IDF | Modifikasi | Y | 0.826<br>0.825 | 0.846<br>0.846 |
| | NB | TFPOS-IDF | Modifikasi | Y | 0.752<br>0.748 | 0.812<br>0.81 |

**Tabel 1** *Hasil Pengujian*

### Analisis Hasil Pengujian

Berdasarkan hasil pengujian, penggunaan *stopwords* versi modifikasi PySastrawi secara rata-rata berpengaruh baik pada hasil F1-measure dan akurasi untuk algoritma SVM dan NB. Hal ini didukung oleh pernyataan Mohammed

et al. [12] bahwa beberapa *stopwords* dapat berdampak signifikan dalam menentukan tingkat kesulitan sebuah soal. Ekstraksi fitur menggunakan metode TFPOS-IDF juga secara rata-rata berdampak baik pada hasil F1-measure dibandingkan dengan metode TF-IDF. Melakukan pembobotan pada kata berdasarkan POSTag dapat membantu algoritma klasifikasi untuk meningkatkan performansinya.

Dataset yang melalui proses *random oversampling* mampu menghasilkan skor F1-measure dan akurasi yang lebih baik pada semua skenario pengujian dibandingkan data yang tidak melalui proses *random oversampling*. Hal ini dikarenakan algoritma klasifikasi dapat dilatih dengan data yang lebih banyak, sehingga dapat melakukan klasifikasi pada data pengujian secara lebih baik. Sementara itu, menggunakan parameter hasil optimasi GridSearchCV secara rata-rata mampu meningkatkan skor F1-measure dan akurasi untuk kedua algoritma. Untuk hasil optimasi yang menunjukkan penurunan skor seperti pada skenario 2 dengan algoritma SVM dan skenario 5 dengan algoritma NB, dapat disebabkan karena cara kerja GridSearchCV yang menentukan parameter terbaik berdasarkan rata-rata tertinggi dari hasil *cross validation.*

SVM menghasilkan performa paling baik pada skenario 8 dengan parameter C = 10 dan kernel = 'linear'. Sementara itu, NB menghasilkan performa paling baik pada skenario 4 dengan parameter alpha = 0. Hasil kesalahan klasifikasi kelas pada kedua algoritma tersebut dapat dilihat pada Tabel 6, sementara untuk kesalahan klasikasi berdasarkan mata pelajaran dapat dilihat pada Tabel 7.

Berdasarkan Tabel 6, kelas C3 merupakan kelas dengan kesalahan klasifikasi terbanyak untuk kedua algoritma. Hal ini dapat disebabkan oleh strategi random oversampling yang digunakan adalah 'not majority', sehingga kelas C3 yang merupakan mayoritas kelas pada dataset ini tidak dilakukan random oversampling. Urutan kelas berikutnya yang dengan kesalahan klasifikasi terbanyak secara urut adalah C2, C4 dan C1 untuk kedua algoritma. Kemudian, pada Tabel 7 dapat dilihat bahwa urutan mata pelajaran yang paling banyak terdapat kesalahan klasifikasi secara urut adalah bahasa indonesia, matematika dan ipa untuk kedua algoritma.

| | C1 | C2 | C3 | C4 | C5 | C6 |
|---|---|---|---|---|---|---|
| SVM | 4 | 16 | 21 | 8 | 0 | 0 |
| NB | 5 | 14 | 21 | 9 | 0 | 3 |

**Tabel 2** *Jumlah Salah Prediksi per Kelas*

| | Bahasa Indonesia | IPA | Matematika |
|---|---|---|---|
| SVM | 22 | 9 | 18 |
| NB | 26 | 9 | 19 |

**Tabel 3** *Jumlah Salah Prediksi per Mata Pelajaran*