1. **Deskripsi Masalah**

Implementasi metode Support vector machines (SVM) untuk klasifikasi analisis sentimen. Menggunakan dataset review femaledaily yang berjumlah 3080. Untuk evaluasi hasil klasifikasi menggunakan Accuracy, Precision, Recall, dan F-Measure. Kemudian akan di analisis pengaruh persentase data latih dan data uji, waktu pelatihan yang diperlukan, dan data yang salah diklasifikasikan.

1. **Perancangan Sistem**

* PreProcessing

Sebelum masuk tahap klasifikasi, data akan di cleaning terlebih dahulu dari simbol, angka, url, dan double space. Kemudian semua huruf di lowercase.

* TF IDF

Dilakukan pembobotan TF-IDF untuk menghitung bobot setiap kata yang umum digunakan. Pada eksperimen kali ini menggunakan ‘max\_features=1000’, yang artinya hanya mengambil top 1000 terms.

* Klasifikasi

Untuk klasifikasi analisis sentiment menggunakan metode algoritma SVM, dengan kernel linear.

1. **Hasil dan Analisis**

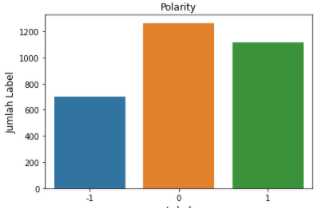
Dari 3080 data text review, dengan 3 kelas yang masing-masing berjumlah sebagai berikut :

0 = Netral

1 = Positive

-1 = Negative





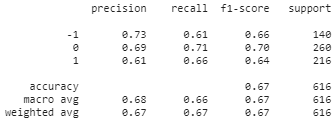
Dilakukan 2 eksperimen, yaitu perbedaan perbandingan split data train dan data test. Pada eksperimen pertama perbandingan data train : data test adalah 80 : 20, pada eksperimen kedua adalah 70 : 30.

Pada eksperimen pertama, untuk proses training dan predict dibutuhkan waktu sebagai berikut :



Kemudian hasil predict dari 616 data testing dihasilkan :



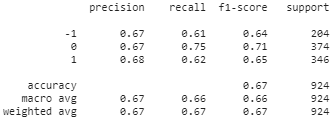


Pada eksperimen kedua, untuk proses training dan predict dibutuhkan waktu sebagai berikut :



Kemudian hasil predict dari 924 data testing dihasilkan :





Dari kedua hasil eksperimen tidak terlalu menunjukan perbedaan yang signifikan. Hanya proses training dan predict yang membutuhkan waktu yang berbeda dikarenakan jumlah data pada eksperimen pertama lebih banyak sehingga membutuhkan waktu lebih banyak. Kemudian nilai akurasinya pun sama, hanya ada perbedaan tipis di nilai precision, recall, dan f1-measure.

Dari table confusion matrix cukup lumayan banyak data yang tidak sesuai dengan actual labelnya. Sehingga nilai precision, recall, f1-measure tidak terlalu bagus. Nilai precision rendah maka model tidak cukup baik untuk memprediksi data yang positif, sedangkan nilai recall rendah maka model tidak cukup mengenali data yang positif. Kemudian nilai f1-measure adalah harmonic mean dari precision dan recall dikarenakan dapat terjadi tradeoff antara precision dan recall.

Dikarenakan jumlah data antara masing-masing kelas tidak terlalu jauh jumlahnya, maka nilai precision recall dan f1-measure masing-masing kelas tidak terlalu jauh

Cukup lumayan banyak data yang salah prediksi. Hal ini dapat terjadi karena terdapat beberapa review yang ambigu. Dalam satu review membahas kelebihan dan kekurangan suatu produk sehingga dapat terjadi salah pelabelan. Ada beberapa terms yang termasuk positif atau negatif, namun dalam beberapa kalimat terdapat penambahan negasi di depan terms tersebut, misalkan ‘tidak berjerawat’, ‘tidak cocok’, ‘tidak bagus’, dll.

Berikut beberapa contoh data yang salah klasifikasi :

