**Nama** : M. Fikri Hasani / Andreas Novian Dwi Triastanto

**NIM** : 23518014 / 23518002

**Kuliah** : IF5282 – Pemrosesan Bahasa Alami

**Klasifikasi Teks Keluhan**

Arsitektur dari sistem klasifikasi teks keluhan yang kami buat adalah sebagai berikut:

1. ***Preprocessing*** : menggunakan algoritma buatan sendiri untuk menghilangkan link, tweet mention, dan tanda baca. Kelompok kami tidak menggunakan fitur normalisasi kata karena mengalami kesulitan dalam menggunakan API dari Prosa.ai. Selain itu, kami menemukan sebuah algoritma normalisasi buatan Muhammad Nur Yasir Utomo (<https://github.com/yasirutomo/text-normalization>), namun setelah kami coba gunakan, hasil akurasi dari proses pelatihan dengan algoritma *machine learning* sangat buruk, yaitu sekitar 0,5.
2. ***Feature extraction*** : menggunakan modul CountVectorizer yang dimiliki oleh library scikit-learn (<https://scikit-learn.org/>) yang dikombinasikan dengan modul TF-IDF dari library yang sama sehingga menghasilkan Document-Term matrix yang akan digunakan dalam algoritma *machine learning*.
3. **Algoritma *machine learning*** : menggunakan tiga jenis algoritma *machine learning*, yaitu Naïve-Bayes, Random Forest, dan Support Vector Machines (SVM) yang semuanya diambil dari library scikit-learn. Untuk algoritma Naïve-Bayes menggunakan Multinomial Naïve-Bayes dengan semua parameter bernilai *default* (*alpha=1.0, fit\_prior=True, class\_prior=None*). Untuk algoritma Random Forest menggunakan Random Forest Classifier dengan parameter n\_estimators=1000 dan random\_state=0, sedangkan parameter sisanya bernilai *default.* Untuk algoritma Support Vector Machines menggunakan SVC dengan parameter kernel="linear" sedangkan parameter lainnya bernilai *default*.

Hasil pengujian akurasi dari data latih yang diberikan untuk masing-masing algoritma *machine learning* yang digunakan adalah sebagai berikut:

* Akurasi Naïve-Bayes: 0.8625498007968128
* Akurasi Random Forest: 0.8911022576361222
* Akurasi SVM: 0.8877822045152722

**Referensi :**

1. https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a