**LAPORAN TUGAS 2**

Implementasi Oversampling dan Undersampling Methods pada Imbalance Datasets

**Kelompok 6**

Muhammad Adib Arinanda – 5115100111

Wahyu Ivan Satyagraha – 5115100155

Alfindio Muhammad Abdallah – 5115100164

**Library yang digunakan :**

1. Pandas
2. Numpy
3. Scikit-learn
4. Matplotlib

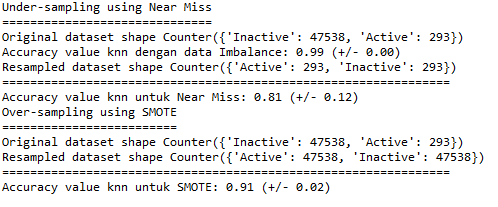
**Datasets yang digunakan :**

PubChem Bioassay Data Dataset

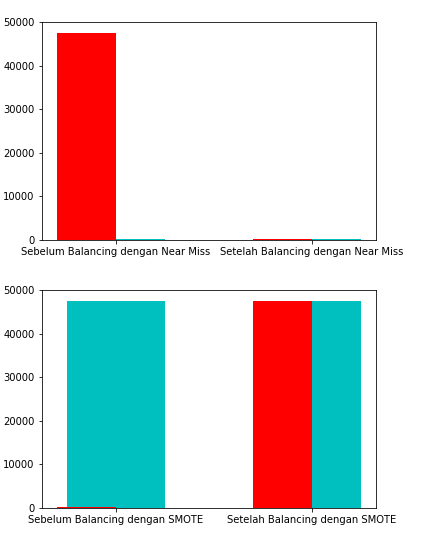
**STEP :**

1. Load dataset.
2. Memisahkan atribut kelas dengan non kelas.
3. Melakukan balancing data dengan metode oversampling (SMOTE) dan undersampling (Near Miss).
4. Menghitung akurasi.
5. Melakukan Feature Reduction dengan PCA.

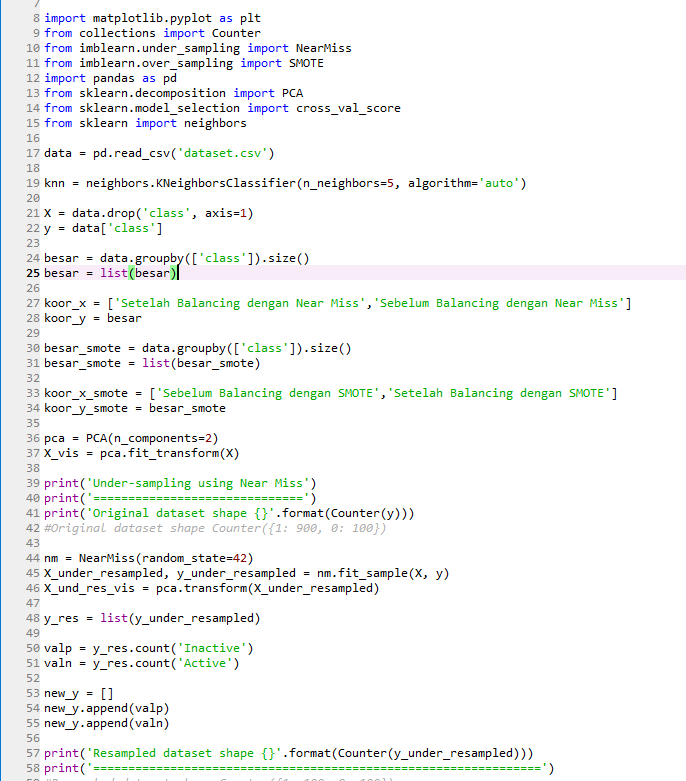
**HASIL BALANCING DATA DAN AKURASI :**

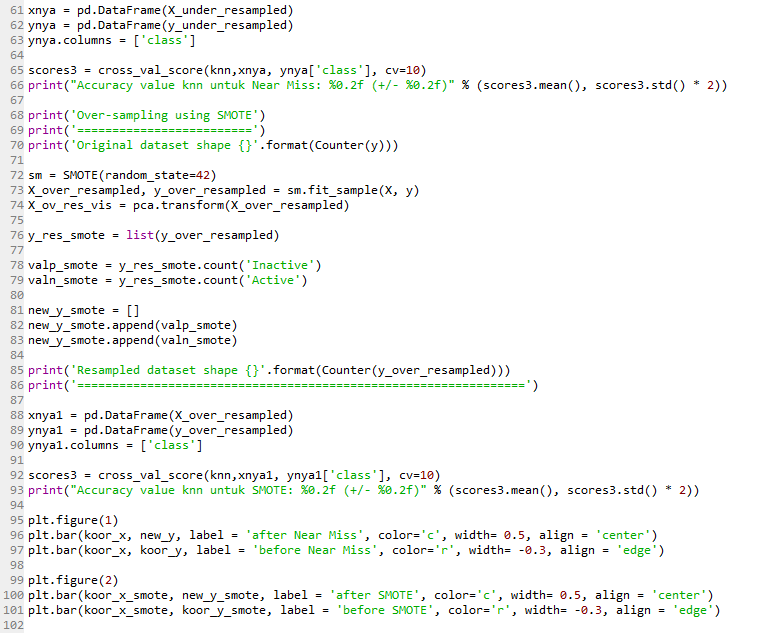


**PLOT :**



**CODE :**





**KESIMPULAN**

Dalam menentukan sebuah akurasi data mining, metode balancing data sangat diperlukan untuk ke akuratan hasil yang di peroleh, pada data set ini, oversampling memiliki akurasi yang lebih tinggi namun dengan compile time yang lebih lama disbanding undersampling karena data yang di olah menjadi lebih banyak. Pada data imbalance akurasi yang didapat lebih tinggi namun dalam keadaan imbalance, hasil dari akurasi tersebut masih harus dipertanyakan.