

*Dataset* dirancang secara artifisial menggunakan *package* sklearn pada python menggunakan fungsi `make_classification`. Data yang dihasilkan umumnya bersifat numerik dan memerlukan proses klasterisasi untuk mendapatkan data *nominal*. Pada data ini didefinisikan 10 kolom fitur dan 1 kolom label. Selain itu juga dibuat 1000 baris data dengan sebaran 50% untuk label “False” dan 50% untuk label “True”. Maksud dari “True” adalah siswa melanjutkan ke perguruan tinggi sedangkan “False” berarti tidak melanjutkan ke perguruan tinggi. Berikut adalah fitur-fitur yang ada pada dataset: