

Abstrak

Permasalahan data dengan distribusi kelas tidak seimbang atau yang dikenal dengan *class imbalance problem* banyak ditemukan dalam permasalahan dunia nyata. Dalam beberapa kasus, data minoritas dalam dataset bernilai lebih potensial dibandingkan data mayoritas. Pengklasifikasian dengan cara umum tidak dapat mengatasi permasalahan ini karena menganggap distribusi setiap kelas adalah sama pentingnya. Oleh karena itu diperlukan suatu metoda untuk melakukan proses klasifikasi yang dapat mengklasifikasikan data minoritas dengan baik.

Algoritma PNrul adalah sebuah algoritma berdasarkan rule yang didesain untuk dapat menangani permasalahan data imbalance multi-class. Permasalahan multi-class ditangani dengan membangun *binary classifier* untuk masing-masing kelas pada dataset dengan kelas target adalah kelas yang sedang dipelajari. Algoritma ini terdiri dari tahapan P, tahapan N, dan mekanisme scoring. Setiap tahapan memiliki peranan masing-masing terhadap pembentukan rule akhir dari tiap kelas.

Dataset yang digunakan pada tugas akhir ini berjumlah lima buah dan diambil dari *UCI Machine Learning Repository*. Setiap dataset merupakan *multi-class* dengan distribusi tiap kelas berbeda-beda. Pengukuran performansi dari algoritma PNrul menggunakan *recall*, *precision*, dan *F_measure*. Dari hasil pengujian dan analisis diperoleh kesimpulan bahwa algoritma PNrul dapat dengan baik mengklasifikasikan data *imbalance multi-class* apabila menggunakan nilai parameter yang tepat.

Kata kunci : *class imbalance*, *multi-class*, algoritma PNrul