

INTISARI

PENERAPAN METODE *DECISION TREE* C4.5 PADA KLASIFIKASI DATA STATUS KERJA YANG TIDAK SEIMBANG DENGAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE)

(Studi Kasus : Data Survei Angkatan Kerja Nasional di Kabupaten Karawang

Agustus 2021)

Afifi Indrasari

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Pengangguran merupakan salah satu permasalahan yang dihadapi oleh negara-negara berkembang salah satunya di Indonesia. Besarnya jumlah penduduk disertai dengan pertumbuhan yang sangat cepat berakibat pada meningkatnya jumlah angkatan kerja. Angka pengangguran akan meningkat seiring dengan meningkatnya jumlah angkatan kerja yang tidak terserap. Pengangguran di Indonesia merupakan masalah yang serius. Pada pelaporan BPS periode Agustus 2021, Kabupaten Karawang memiliki tingkat pengangguran terbuka sebesar 11.83 persen. Hal ini cukup kontradiktif, mengingat Kabupaten Karawang dikenal sebagai kawasan industri yang strategis dan dipengaruhi secara spasial oleh Provinsi DKI Jakarta. Penelitian ini menerapkan algoritma C4.5 sebagai bagian dari metode *decision tree*. Dari data SAKERNAS, kelas pengangguran memiliki jumlah jauh lebih sedikit dibandingkan kelas bekerja. Sehingga dilakukan penanganan ketidakseimbangan data terhadap kelas minoritas menggunakan teknik *resampling* yaitu SMOTE. Tujuan dari penelitian ini adalah untuk mengeksplorasi variabel-variabel yang paling berkontribusi terhadap proses pengklasifikasian dan membandingkan model sebelum dan sesudah SMOTE untuk mendapatkan model klasifikasi terbaik. Berdasarkan pengujian yang telah dilakukan, model setelah SMOTE dengan *split* data 80%:20% memberikan performa terbaik dibandingkan tanpa SMOTE dengan nilai akurasi, *sensitivity*, *specificity*, *g-means*, dan AUC berturut-turut sebesar 0.789, 0.817, 0.757, 0.786 dan 0.787. Selain itu, *root* teratas yang memiliki kontribusi besar dalam menentukan status kerja penduduk di Kabupaten Karawang adalah variabel status perkawinan.

Kata Kunci : *decision tree*, algoritma C4.5, klasifikasi, ketidakseimbangan data, SMOTE, pengangguran, angkatan kerja