

ABSTRAK

Rasio pajak Indonesia yang rendah jika dibandingkan dengan negara Asia Pasifik lainnya mengindikasikan adanya praktik penghindaran pajak. Penghindaran pajak yang terus berlanjut dapat merusak kepercayaan publik dan berdampak negatif terhadap perekonomian secara keseluruhan. Oleh karena itu, mendeteksi adanya indikasi kecurangan dalam pembayaran pajak merupakan salah satu prioritas utama otoritas pajak. Dengan perkembangan teknologi yang pesat, muncul istilah *data mining* yang merupakan proses menemukan pola dan tren yang berguna dalam kumpulan data.

Tujuan dari penelitian ini adalah untuk menyusun model prediksi aktivitas penghindaran pajak perusahaan di BEI dengan menggunakan pendekatan *data mining*. *Corporate Tax to Turnover Ratio* dijadikan sebagai atribut target yang menggambarkan perilaku pembayaran pajak perusahaan. Penelitian ini menggunakan data seluruh perusahaan BEI yang terdaftar mulai tahun 2018 hingga 2022 dengan jumlah 4.219 baris data. Metodologi CRISP-DM digunakan untuk melakukan penyusunan model. Hasilnya, model *Naive Bayes* mampu menghasilkan nilai *recall* sebesar 80,99%. Menurut model *Decision Tree*, GPM, OPM, dan ROA memiliki hubungan negatif terhadap rasio pembayaran pajak perusahaan, sedangkan DAR dan TATO memiliki hubungan positif. Terdapat perbedaan antara hasil evaluasi model menggunakan data secara keseluruhan dengan pengujian per industri dan per tahun. Oleh karena itu, penyusunan model sebaiknya dilakukan per klaster industri dan per tahun.

Kata kunci: *data mining*, *machine learning*, penghindaran pajak.