

ABSTRAK

Pertumbuhan jumlah investor pasar modal di Indonesia mendorong kebutuhan terhadap metode analisis berbasis data yang tidak hanya mengandalkan pergerakan harga saham, tetapi juga memperhatikan aktivitas transaksi *broker*. Aktivitas tersebut merupakan pola perdagangan transaksi saham yang membentuk akumulasi dan distribusi yang menggambarkan kecenderungan pembelian dan penjualan saham. Penelitian ini bertujuan menerapkan algoritma *eXtreme Gradient Boosting* (XGBoost) untuk memprediksi pola akumulasi dan distribusi *broker* saham berdasarkan data historis *broker summary*. Objek penelitian ini menggunakan saham PT. Aneka Tambang Tbk. dengan memanfaatkan data yang diperoleh secara harian sebanyak 238 *file* dan diolah menjadi satu dari platform IPOT dengan periode 2 Januari 2025 hingga 2 Januari 2026 atau sebanyak 14.761 data. Penelitian ini menggunakan metodologi *Sample, Explore, Modify, Model, and Assess* (SEMMA) dan pemodelan dilakukan menggunakan lima skenario pembagian data secara kronologis, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Proses optimasi hyperparameter dilakukan menggunakan *RandomizedSearchCV* dengan validasi *TimeSeriesSplit*. Hasil pengujian menunjukkan bahwa skenario pembagian data 80:20 menghasilkan kinerja model lebih baik daripada yang lain dengan nilai *F1-score* sebesar 62,11%. Meskipun evaluasi model yang ditunjukkan memiliki nilai yang sedang, kecenderungan model dalam mempelajari data historis *broker summary* dinilai memiliki kemampuan prediksi yang masih terbatas. Hasil evaluasi model kemudian diimplementasikan ke dalam antarmuka berbasis web untuk mempermudah analisis evaluasi dan hasil prediksi dan belum dapat digunakan secara langsung sebagai dasar keputusan investasi.

Kata Kunci: Prediksi *Broker*, *eXtreme Gradient Boosting*, SEMMA, Akumulasi dan Distribusi Saham, Investasi, Saham

ABSTRACT

The growing number of capital market investors in Indonesia has driven the need for data driven analytical methods that not only rely on stock price movements but also take into account broker trading activity. These activities constitute stock trading patterns that form accumulation and distribution, which reflect trends in stock buying and selling. This study aims to apply the eXtreme Gradient Boosting (XGBoost) algorithm to predict stock broker accumulation and distribution patterns based on historical broker summary data. The subject of this study is the stock of PT. Aneka Tambang Tbk., utilizing 238 daily data files processed into a single dataset from the IPOT platform, covering the period from January 2, 2022 to January 2, 2026 that has 14,761 data. This study employs the Sample, Explore, Modify, Model, and Assess (SEMMA) methodology, and modeling was conducted using five chronological data split scenarios: 50:50, 60:40, 70:30, 80:20, and 90:10. Hyperparameter optimization was performed using RandomizedSearchCV with TimeSeriesSplit validation. Test results show that the 80:20 data split scenario produced better model performance than the others with an F1-score of 62.11%. Although the evaluation of the model shown indicates moderate performance, the model's tendency to learn from historical broker summary data is assessed as having limited predictive capability. The results of the model evaluation were then implemented into a web-based interface to facilitate the analysis of evaluation results and predictions; however, they cannot yet be used directly as a basis for decisions to buy or sell stocks.

Keywords: *Broker Forecasts, eXtreme Gradient Boosting, SEMMA, Stock Accumulation and Distribution, Investing, Stock*