

## ABSTRAKSI

Salah satu tantangan yang dihadapi oleh PT Telkom Indonesia dalam proses transisi ke layanan FTTH adalah bagaimana mempertahankan pelanggan saat ini dan menarik pelanggan untuk bermigrasi ke produk Indihome. Telkom saat ini memiliki pelanggan telepon rumah sekitar 8,3 juta dan hanya sekitar 26% yang terprofil dengan baik. Kondisi ini menyebabkan perusahaan sulit untuk mengidentifikasi dan mengenali pelanggan yang potensial. Oleh karena itu, perusahaan memerlukan model segmentasi pelanggan untuk menentukan prioritas proses migrasi pelanggan yang sudah ada ke Indihome. Berdasarkan aturan Pareto 80/20, perusahaan yang dapat menemukan 20% pelanggan yang menguntungkan akan dapat mempertahankan pendapatan usahanya.

Dalam penelitian ini, proses pengelompokan pelanggan menggunakan analisis komponen utama (PCA) dikombinasikan dengan K-Means, kemudian proses klasifikasi menggunakan customer lifetime value (CLV) dikombinasikan dengan support vector machine dan logistic regression. Sumber data adalah call data record dan pembayaran pelanggan saat ini. Proses dimulai dengan data primer dianalisis menggunakan PCA. Beberapa atribut akan diseleksi berdasarkan nilai signifikan hasil PCA kemudian K-Means akan membentuk kluster. Kluster yang optimal ditentukan oleh index Calinski Harabasz dan Davies Bouldin. Pada setiap cluster, skor CLV akan dihitung untuk menentukan apakah kluster tersebut adalah kluster atas yang memiliki potensi keuntungan masa depan atau kluster yang lebih rendah dimana membutuhkan lebih banyak upaya dalam proses migrasi. Proses terakhir adalah pembangunan model klasifikasi menggunakan SVM dan regresi logistik untuk data baru.

Model terbaik adalah analisis komponen utama (PCA) sebagai teknik seleksi fitur. Metode pengukuran clustering untuk menentukan kluster yang optimal menggunakan Calinski Harabasz. Sementara, untuk membangun model klasifikasi dengan algoritma logistic regression. Model ini memberikan hasil akurasi yang lebih baik dan peningkatan kinerja algoritma dengan skor F-Test 99,16%, akurasi sekitar 99,93%, dan presisi 98,84%. Analisis komponen utama sebagai seleksi dapat meningkatkan kinerja logistic regression hingga 2% dibandingkan dengan model tanpa PCA dan model RFM. Sementara, daftar teratas dari pelanggan menguntungkan yang dapat disimpulkan dengan menggunakan model ini adalah 33,62% dari total pelanggan saat ini.

**Kata kunci:** analisis komponen utama, Customer lifetime value, K-Means, support vector machine, regresi logistik