

ABSTRAK

Pada perkembangan teknologi saat ini, data memiliki peran penting untuk mendukung pencapaian tujuan bagi perusahaan. Pentingnya data bagi perusahaan yaitu untuk pemenuhan kualitas menunjang kebutuhan bisnis perusahaan. Kualitas tinggi yang dimiliki oleh data bernilai penting bagi perusahaan. Selain itu laporan keuangan perusahaan sangat penting bagi perusahaan karna dengan laporan keuangan perusahaan dapat mengetahui perkembangan perusahaan. Laporan keuangan dapat dibuat apabila terdapat data keuangan perusahaan sehingga diperlukan pendekatan big data untuk mengolah data perusahaan yang cukup besar. Penelitian ini membahas tentang penerapan analisa *roll rate* model pada data transaksi pelanggan di perusahaan telekomunikasi XYZ. Sebelum dilakukan proses implementasi *roll rate* diperlukan tahapan untuk mengolah big data yaitu proses ETL dan *data preprocessing* sehingga data yang mentah tersebut dapat diterapkan kepada proses *roll rate* model dengan baik. Teknologi yang digunakan untuk proses ETL dan *data preprocessing* adalah teknologi Apache Spark dengan modul Spark Scala, SparkSQL, dan PySpark untuk memperoleh hasil yang efektif dan efisien. Terdapat tiga tahapan yang akan dilakukan di antaranya ETL, *data preprocessing*, dan *roll rate* model. Proses ETL yang dilakukan mendapatkan waktu lebih cepat dari pada proses eksisting yang ada pada perusahaan. Selain itu tingkat evaluasi jumlah nilai *roll rate* didapatkan pada setiap periode dibandingkan hasil dari proses eksisting memiliki tingkat persamaan 100%.

Kata kunci—***ETL, Data Preprocessing, Roll Rate Model, Spark***