ABSTRACT

In today's technological developments, data has an important role to support the goals for the company. Data is important for companies, namely to meet the quality that supports the company's business needs. The high quality that data has is of critical value to the company. In addition, the company's financial statements are important for companies because company reports can find out the company's development. Financial reports can be made if there is company financial data so that a big data approach is needed to process large company data. This study discusses the application of roll rate model analysis on customer transaction data at XYZ telecommunications company. Before the roll rate implementation process is carried out, steps are needed to process big data, namely the ETL process and preprocessing data so that the raw data can be applied to the roll rate model process properly. The technology used for ETL processing and data preprocessing is Apache Spark technology with Spark Scala, SparkSQL, and PySpark modules to obtain effective and efficient results. There are three stages that will be carried out including ETL, data preprocessing, and roll rate model. The ETL process that is carried out gets faster time than the existing processes in the company. In addition, the level of evaluation of the total roll rate value obtained in each period is compared with the results of the existing process having a 100% level equation.

*Keywords —**ETL, Data Preprocessing, Roll Rate Model, Spark***