

ABSTRACT

ANALYSIS AND DETECTION OF FRAUD ON CALL DATA USING K-NEAREST NEIGHBOR ALGORITHM (CASE STUDY: PT XYZ)

By

EPTA RIZKY ALFIYYAH

NIM : 1202164077

This study discusses one of the problems that occur in Indonesia, namely telecom fraud. One of the State-Owned Enterprises (SOEs), PT XYZ, is the party responsible for dealing with telecom fraud. Telecom fraud is an activity of using telecommunications facilities that is carried out illegally and intentionally in various forms of fraud, fraud, or embezzlement by certain people or organizations whose purpose is to obtain these services and avoid service fees or tracking records of invoices that are carried out illegally. The purpose of this study is to analyze and detect fraud in call data that is indicated as SIMBox fraud where this has harmed PT XYZ who has the task of handling the fraud problem. This research was conducted using the data mining process and the K-Nearest Neighbor algorithm. There are several types of telecom fraud, SIMBox fraud is one of the types of telecom fraud that will be discussed in this study. The data used in this study are data calls made from June to August 2017, where this data will be processed using the k-nearest neighbor algorithm. This data call in the form of CSV data that is processed or processed using the data mining process. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases. The results of this study will be in the form of the accuracy of the implementation of the K-Nearest Neighbor algorithm in classifying phone number data that is indicated as SIMBox fraud. The study was conducted using the Jupyter Notebook tool. After testing the K-Nearest Neighbor algorithm by using $k = 1$, $k = 3$, $k = 5$, $k = 7$ and $k = 9$, the highest accuracy value is 74.2% with $k = 9$, ratio 0.1, precision 74%, recall 74 %, and f1-score 74%.

Key words : Telecom Fraud, SIMBox fraud, Data Mining, K-Nearest Neighbor.