May - June 2020
ISSN: 0193-4120 Page No. 4399-4401

Detection of Fraudulent Behavior in Water Consumption using ARIMA Model

[1]Navya S, [2]Pooja K, [3]Shilpa N R, [4]Nethravathi M

[1,2,3,4]Department of Computer Science
REVA University, Bangalore, India
[1]navyas3616@gmail.com, [2]poojakannanv98@gmail.com, [3]shilpanr@reva.edu.in,
[4]nethravathi849@gmail.com

Article Info
Volume 83
Page Number: 4399-4401
Publication Issue:
May - June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract
Companies that supply water face numerous issues when customers deceit the organization by involving in fraudulent activities. It is a challenging task for the company to figure out the culprit. There has been constant research going on to resolve the issue, and here we stress upon a more profound level to support the companies. So, this particular model assists in finding out the behaviour of the customers. The research focuses principally on forecasting techniques where an algorithm called ARIMA is used to detect the behaviour. By applying this time series algorithm, it exhibits accurate results and improves efficiency, thereby improving the revenue and reducing the non-technical loss of the company.

Keywords: *Time Series, Water Consumption, forecasting, ARIMA Model*

1. Introduction

Water is the most fundamental element for all living beings. It plays a major role in the day to day life, agriculture, and industries. Insufficiency of water is a huge problem faced by many countries around the world, which leads to a threat that affects all the domains that depend on the water for their growth and success.

This problem is the major drawback to developing countries in an economical way. This situation is due to an increase in the population over the decade. Many governments have come up with new plans and techniques to improve the services provided to the people by providing new sources, reducing the water rates that doesn't emit revenue, constructing the network again, detect loss of supplied water and regulate water usage.

Many water supplying agencies and companies face significant losses due to the fraudulent behavior of the customer in water consumption. losses can be classified into two categories, problems in the production system, transferring water through water networks, network breakages are called Technical Loss (TL), and the second category Tampering of the water meter, thieving by bypassing the meter, making sure of false reading, and ensuring billing irregularities are called Non-Technical Loss which leads to the loss in revenue. A large number of Non-Technical Losses is due to the fraudulent activities by the customer. To detect all these fraudulent behavior the current process is expensive and not effective or efficient.

This paper focuses on detecting the fraud customer using the customer's historical data. The main objective of this entire project is to use a well-known machine learning algorithm called ARIMA model to identify suspicious fraudulent customers and forecast future traits.

2. Literature Survey

This section reviews some of the techniques used to detect domains related to the detection of fraudulent behavior in water consumption.

For example, the paper [4] proposed a model for detecting fraud in water consumption using the data mining technique. To build this model an industry-standard data mining methodology was used. The entire process is based on the customer's historical metered consumption data. The customers billing data is taken in account that is maintained by the billing system which is implemented with a tool based billing system where the bills were computed and stored in the system. The project

May - June 2020
ISSN: 0193-4120 Page No. 4399-4401

was built using datasets that include thousands of customer profiles. Support Vector Machine and KNN eminent data mining algorithms are used to classify the customer profile and help in detecting where the customer's label that was not balanced were known as fraud customers.

Coming to paper [5] they asserted a model that detects the behaviour of customers who tamper the water meters taking all the previous year's data in Gaza City. He used a data mining algorithm such as SVM and compared his model using the other two algorithms which are ANN and KNN. The chief method he used involves collecting the data, cleaning it and finally testing it to get accurate results.

In paper [6] they designed a model using a pattern-based technique where it detects the behaviour taking weeks and days. It is natural to have variations every day mainly because of differences in the number of people due to vacations or relatives coming to people's houses. So, keeping all these variations in mind, the model tried to detect precisely irrespective of large differences in consumption. The forecasting is done based on set patterns such as the differences in the weekdays and weekends due to having work during weekdays. The patterns may be identical or different, but all these patterns are taken into account and the nearest neighbor along with the Euclidean distance is calculated.

3. Problem Definition

Quantitative forecasting methods such as the time series forecasting technique greatly help in determining the true nature of the model. Using the previous years data of the user, the future value is predicted and classified as fraud or non-fraud. A machine learning approach, namely ARIMA Model, is applied to detect the customers in water consumption. Statistical principles and concepts are applied. Using the concepts, the future data of the variable is predicted by feeding the algorithm the historical data of the water consumption by that user.

4. Architecture

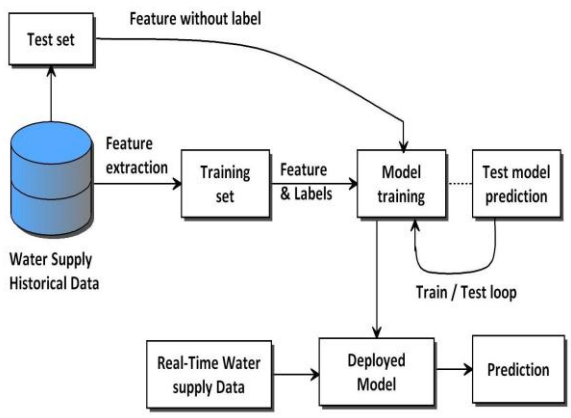

Figure 1: Architecture

As our project deals with identifying fraudulent behavior in the consumption of water by every individual in the society. To solve this problem we are using the our model which is a time-series Machine learning algorithm that takes non-labeled data and gives accurate results with feature enhancing results.

Forecasting is divided into two main parts:

1. One where we measure with quality such as size, color, etc..
2. Second where we measure in meters. Our project comes under the second category.

Some principles and concepts are applied to a data set(we are including water consumption rate of a hundred users of several years) to forecast the feature requirement of every individual. Some of the time series forecasting techniques used in our project are mentioned below:

1. Elimination of recessive characters comparing the previous readings.
2. Calculating the average from the previous subset from the data set we took.
3. Seasonal Regression Model
4. Distributed Lag Model.

Here we are converting Non-Stationary data into Stationary data based on the series difference data observed on previous readings which are denoted by y^ t.

Sometime-series forecasting methods used in our project are:

1. **Auto Regression**: In this, the extraction of lagged data from the data set is done based on the properties it shows. This value is shown as p.
2. **Make category:** Here, we convert non-stationary data into stationary data and we remove the trend. It is represented by d in the algorithm.
3. **Calculate Average:** In our project this value is shown as q. the number of the lagged value of error-term. These steps are calculated using the correlation with lag function and Partial correlation with lag Function.

The following algorithm explains the flow of our project

1. Plot the graph.
2. Estimate the values from the data set provided in our project.
3. Plot the graph using values in step 2.

If the graph is coming with no lag, predict the future output using these values else repeat all the above procedure until you get accurate results.

This model depicts the algorithm(p,d,q)of y^t. Once p,q,d is selected, we check the functionality of our model on a training data set and then we use these values to calculate the future requirement of the different individuals based on their consumption. Once it is done we cross-check with the previous value to whether our estimated values are working correctly or not.

May - June 2020
ISSN: 0193-4120 Page No. 4399-4401

5. Experimental Results

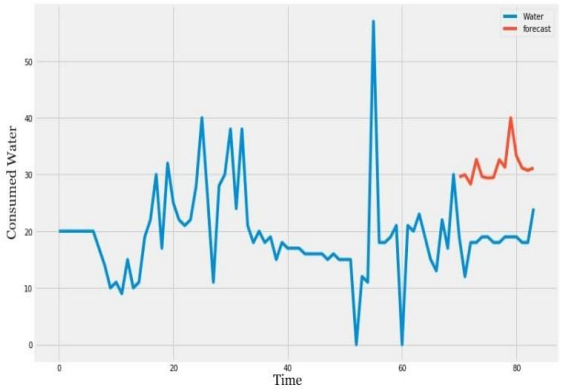

Figure 2: Output when test data is compared with the predicted value

The given graph is plotted over Time in the x-axis and Consumed Water in the y-axis. Here, it shows that in the test data which we're giving, there is a spike in the water consumption. If the graph(red color) would have displayed similar to the predicted value, then there would be no suspicious behavior. But since there is a sudden increase in the value, the graph indicates the same and displays that, that particular user is a fraud.

6. Conclusion And Future Enhancement

In this paper, we have applied a machine learning algorithm to detect fraudulent behavior in water consumption. We have used ARIMA model to classify the suspicious fraud behavior of the customer using the historical data of his consumption and also the moving average helps us in figuring out the future consumption rates of the customer in a way that, with changing times the algorithm will get adapted to it and hence classify the customers accurately and efficiently.

In the project, we emphasize more on detecting the behavior rather than detecting the water leak of the system. Therefore, detection of water leak can be thought of in the future to avoid loss of water. Once the leak is detected, the issue can be fixed instantly which otherwise go unnoticed.

References

[1] "Jordan Water Sector Facts & Figures, Ministry of Water and irrigation of Jordan". Technical Report. 2015.

[2] N/A, "Water Reallocation Policy, Ministry of Water and irrigation of Jordan". Technical Report. 2016.

[3] C. Ramos , A. Souza , J. Papa and A. Falcao, "Fast non-technical losses identification through optimum-path forest". In Proc. of the 15th Int. Conf. Intelligent System Applications to Power Systems, 2009, pp.1-5.

[4] Qasem A. Al-Radaideh, Mahmoud M. Al-Zoubi "A Data Mining Based Model for Detection of Fraudulent Behaviour in Water Consumption", Expert Systems with Applications.

[5] Eyad Hashem S. Humaid "A Data Mining Based Fraud Detection Model for Water Consumption Billing System in MOG

[6] Rafael Benítez ,Carmen Ortiz-Caraballo , Juan Carlos Preciado , José M. Conejero, Fernando Sánchez Figueroa and Álvaro Rubio-Largo , "A Short-Term Data Based Water Consumption Prediction Approach

[7] N. Carneiro, G. Figueira and Costa M., "A data mining based system for credit-card fraud detection in e-tail decision support systems", Decision Support Systems, 2017, 95(C): 91-101.

[8] Ortega P., Figueroa C., and Ruz G. "A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile", In proc of DMIN, 2006.

[9] B. Kusaksizoglu, "Fraud detection in mobile communication networks using data mining", Bahcesehir University, The Department of computer engineering, Master Thesis. 2006.

[10] C. Liang-Chun, H. Chien-Lung, L.Nai-Wei, Y. Kuo-Hui and L. PingHsien, "Fraud analysis and detection for real-time messaging.