Chapter 23
Churn Prediction and Fraud Detection in Dairy Sector Using Machine Learning

Hitarth Deepak Shah
Charotar University of Science and Technology, India

Chintan M. Bhatt
Charotar University of Science and Technology, India

Shubham Mitul Patel
Charotar University of Science and Technology, India

Jayshil Bhavin Khajanchi
Charotar University of Science and Technology, India

Jaimin Narendrakumar Makwana
Charotar University of Science and Technology, India

ABSTRACT

India has globally been the largest milk-producing country in the world for two decades. About 400 million litres of milk is produced every day. It is the responsibility of a dairy sector to look after the farmers by providing them with various services for their livelihood. The growing financial capital of the dairy industry has enticed various fraudulent behaviour. The majority of suspicious activities are seen during the collection at local collection centres, fake farmer entries, tempered quantity and fat entries manually, and adulteration are the profound malpractices exercised by farmers. So, in this research work, the authors present a profound study on the most popular machine learning methods applied to the problems of farmer churn prediction and fraud detection in the dairies. They applied a plethora of machine learning algorithms to get accurate results for churn and fraud detection. XGBoost Classifier was the best for churn prediction with 93% accuracy, while random forest classifier turns out to be effective for fraud detection with 94% accuracy.

DOI: 10.4018/978-1-7998-6618-3.ch023

INTRODUCTION

In today's era, the farmer attrition is considered to be the biggest loss for the dairy industry. The peasants play a very crucial role in producing, collecting and delivering the milk to the centres or organizations on which they are dependent. So it is must to look after them by providing all the necessary services and stop them from getting churn. As the farmers stop collecting or delivering the milk the whole chain of supplying of milk to the consumers break and it can cause a huge loss to the milk consumers and dairy sector. Our ultimate goal is to know which farmers have already churned and who are likely to churn in near future by applying various machine learning algorithms and generating the classification report of each and every algorithm to know which algorithm gives the most accurate result.

Today milk industries and farmers are facing many challenges due to various malpractices carried out by few farmers and milk collection authorities to fill their own pockets. Due to these practices, industries are facing financial loss, due to that they are not able to provide sufficient facilities and money to farmers. These industries must minimise the fraudulent activities practices by farmers and milk collection authorities to offer more comprehensive benefits to the farmer and to serve better quality milk and dairy products to its customers. The central aim of this system is to detect as many suspicious entries in the milk collection dataset, which would ultimately help dairy industries to maximize their profit and serve better quality dairy products to its customers.

Fraud entries can be categorized as:

1. Milkamresult:-
    a. By this feature, we can identify the kind of farmer whose entry is done manually in terms of percentage. So, we can watch on this kind of farmers, whose entry would have done manually on their total entry, which helps to identify their fat and quantity weights records were inserted manually or automatic.
    b. We can identify those kinds of the farmer from this, whose fat entry was done more than 80% manually.
    c. We can identify those kinds of the farmer from this, whose weight quantity was done more than 50% manually.
    d. Identify those kinds of the farmer whose both fat and qty weight entry is done manually combine percent is more than 45%
    e. Advantage of this feature is that, the company can prevent losses which would eventually lead to profit maximization.
2. Editresult:-
    a. By this feature, we can identify the data entry mistakes, which would be done automatically taken by the system. We considered those entry which updated after 60 minutes(This is default criteria you can update it by given external value).
    b. Advantage of this feature is that, we can detect the entry mistake how many times it directly leads to huge amount of loss. So, by using this we can save the cost of expenses.
3. Sqmilk:-
    a. By this feature, we can calculate first highest to fifth highest most repeat fat and qty count and its total percentage, if it's more than 50% then it display the result.
    b. We can also individually filter out personal farmer fat and qty count greater than 50%.
    c. We can detect two type of behavior either it is adulteration or some other suspicious activity.