

A advance account fraud detection and budget prediction system using machine learning

Tapaprabha Chowdhury

Abstract: Advance account fraud detection system is an application which can monitor the expenses of a company based on every year and month. The application is built using a machine learning algorithm. The algorithm anlysis companies past few years records and current market values of product. Based on that analysis the system can predict a particular years or months expenses records or it can check the records entered by accountant if there is any wrong in the record.

Keywords: IP, IR, IF, PIP

INTRODUCTION

Accounts department is one of the major and important departments in any company or organization. This account department maintains all the records of companies expenses. The company maintains its monthly or annual budget through this department. If there is any wrong or mistake then the company may face a huge loss. It will be also better for a company if the annual budget of a year is predicted at the beginning of a year. This paper is based on a smart account monitoring system which can predict the yearly or monthly expenses of a company and analyze the budget of a year or month put by the accountant and analyze if there is any wrong or not.

ALGORITHM

Machine Learning: In Machine learning, a machine learns from its past experience and performs tasks based on that learning without being properly programmed.

Supervised learning: In supervised learning, a machine can learn from a labeled data and gives an output of label data.

In this paper we will learn a supervised learning algorithm which can predict the overall expenses of a company in a particular year or month or can detect fraud or mistake in accounts.

Step 1:

At first user has to put past few years expense records. From this record the algorithm will first find out the IP (Increment Percentage).

y_1 is the expense of year1

y_2 is the expense of year2

IP_1 is the increment percentage of year1 and year2

$$IP_1 = \frac{y_2 - y_1}{y_1} * 100$$

$$IP_2 = \frac{y_3 - y_2}{y_1} * 100$$

If there is n data in the records then

$$IP_{n-1} = \frac{y_n - y_{n-1}}{y_1} * 100$$

Step 2:

Algorithm will find the increment rate IR. IR is the difference between two IP

$$IR_1 = IP_2 - IP_1$$

$$IR_2 = IP_3 - IP_2$$

If there is number of data then

$$IR_{n-2} = IP_{n-1} - IP_{n-2}$$

Step 3:

Algorithm will find out the Increment Factor IF, PIP

IF is the mean value of all IR

PIP is predicted increment percentage

If there is n number of data then

$$IF = \frac{IR_1 + IR_2 + IR_{n-2}}{n-2}$$

$$PIP = IP_{n-1} + IF$$

Step 4:

Algorithm will predict PTy and PTm based on the analysis.

PTy is the predicted total of current year

PTm is the predicted total of current month

$$PTy=yn+((yn*PIP)/100)$$

$$PTm= (yn+((yn*PIP)/100))/12$$

Algorithm will predict the expens of each and every field of the record based on this process.

Step 6:

The algorithm will compare the actual data put by accountant with the predicted value.

It will give the result in three category- red,yellow,green.

If the data is marked as green then the data is correct.

If the data is marked as yellow then the algorithm has doubt on this data.

If the data is marked as red then the algorithm is sure that there is something wrong in this data.

IMPLEMENTATION

I have collected some sample data on which I applied this algorithm.

The data are divided into two parts-i>test data ii>train data

Train data:This is the train data.By analysing these data the algorithm will get an idea what the expens of a perticular year and perticular month should be.

| YEAR | FOOD | ELECTRICITY | SERVANT | OTHERS | TOTAL | IP | IR |
|------|---------|-------------|---------|--------|-------|----|----|
| 2019 | 6000000 | 36000 | 240000 | 25000 | | | |
| 2020 | 6250000 | 37000 | 248000 | 25500 | | | |
| 2021 | 6310000 | 37800 | 256500 | 26100 | | | |
| 2022 | 6370000 | 38500 | 261000 | 26700 | | | |
| 2023 | 6410000 | 39100 | 267000 | 27100 | | | |

Test data:

Test data are the actual data put by the accountant which will be checked by the machine the data are write or not.

I have used advance excel to implement this algorithm.

In the above project the algorithm analysis the past 5 yaers data-2019,2020,2021,2022,2023 and find out the IP,IR,PIP,IF and predict the current years data .

It has also check data put by user and marks red yellow and green.

Green-No problem

Yellow-doubt

Red-problem

CONCLUSION

This model will help any company to maintain its expens in right way.If there is any mistake or corruption in account depertment then it will be detected. Beside this the model will predict the expens of a perticular month and year from which company will get an idea of the expens.