International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Prediction System for Customer Subscription to The Term Deposit

Kumar Gaurav*, Aditya Negi
Department of Computer Applications, Maharaja Surajmal Institute, New Delhi, Delhi, India

Article Info
Volume 7, Issue 4
Page Number: 230-234
Publication Issue :
July-August-2021
Article History
Accepted : 10 July 2021
Published : 18 July 2021

ABSTRACT

Telephonic campaign are one of the most effective ways to reach individuals to sell products but many a time or most of the time is wasted on customers that do not need the products. But this algorithm simply based on the data can tell whether the customer will buy the term deposit or not accurately, which reduces the time wastage and increases efficiency.

Keywords : Prediction, Algorithm, Campaign

I. INTRODUCTION

Telephonic selling campaigns still stay one in all the foremost effective thanks to reach dead set individuals. However, they need vast investment as giant decision centers area unit employed to truly execute these campaigns. Hence, it's crucial to spot the shoppers possibly to convert beforehand in order that they'll be specifically targeted via decision. you're supplied with the shopper information such as: age of the shopper, their job sort, their legal status, etc. together with the shopper information, you're conjointly supplied with the knowledge of the decision like the period of the decision, day and month of the decision, etc. Given this info, your task is to predict if the shopper/customer will purchase term deposit.

II. METHODS AND MATERIAL

Proposed Solution

- **Supervised Machine Learning**

Supervised learning means you've got input variables (x) and output variables (Y), and use algorithms to find out the mapping function from input to output.

$$Y = f(X)$$

The goal is to approximate the mapping function well in order that once you have new input file (x), you'll predict the output variable (Y) of that data [1]. It's called supervised learning because the method of learning algorithms from the training data set are often considered the teacher of the supervised learning process. We all know the right answer, and therefore the algorithm iteratively makes predictions on the training data, which are corrected by the teacher.

When the algorithm reaches a suitable level of performance, learning stops. Regression and

classification issues are two types of supervised learning tasks.

- **Classification**:

A classification problem is when the output variable may be a category, like "red" or "blue" or "disease" and "no disease" [2].

- **Regression**:

A regression problem is when the output variable may be a real value, like "dollars" or "weight". [3]

Some common problem types built on classification and regression include recommendation and time series forecasting, respectively. Since we are dealing with classification-related issues, this is why we will use supervised machine learning.

Data Preprocessing

The information collected from the field contains numerous superfluous things that lead to mistake examination. For illustration, the information may contain purge areas, may contain columns that are not significant to the current examination, and so on. Subsequently, the information must be pre-processed to meet the necessities of the sort of examination you're trying to find. This can be exhausted during the preprocessing module.

Go to **open file** ... option under the **Preprocess** tag select your file.[4][5]

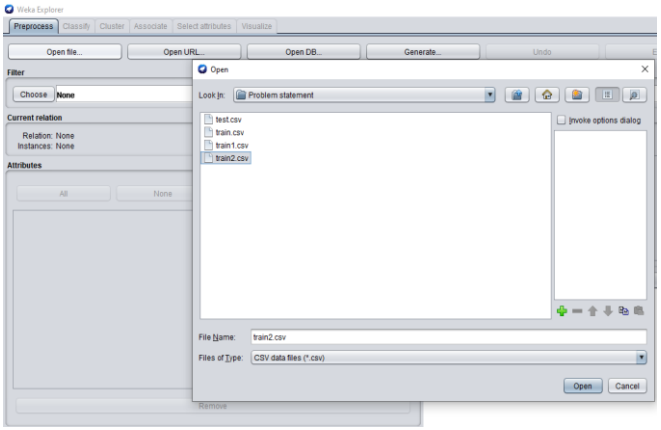

Figure 1: Preprocess windows

When the file is opened, your screen will look like this:

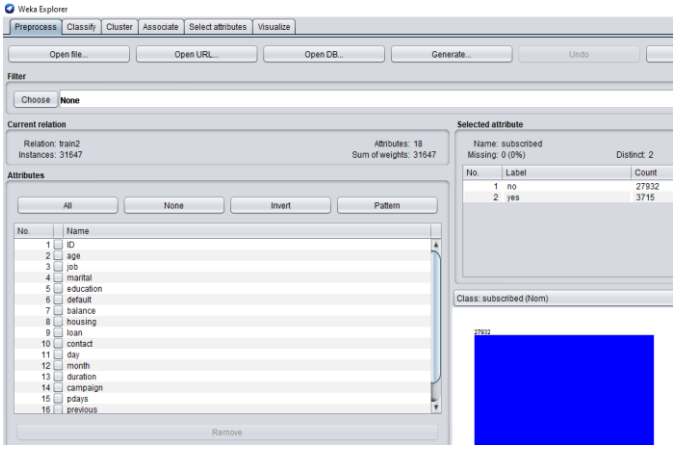

Figure 2: File view

This screen shows details about data we have.

Data

You are provided with following files:

1. **train2.csv:** Use this dataset to train the model. The client and call information, as well as the target variable "subscribed," are all contained in this file. You must use this file to train your model.
2. **test.csv:** Use the trained model to predict whether a new set of clients will subscribe the term deposit.

The dataset is downloaded from UCI Repository

Link: https://archive.ics.uci.edu/ml/index.php [6]

Data Dictionary

Here is the description of all the variables:

Variable	Definition
ID	Unique client ID
age	Age of the client
job	Type of job
marital	Marital status of the client
education	Education level
default	Credit in default.
housing	Housing loan
loan	Personal loan
contact	Type of communication
month	Contact month
day_of_week	Day of week of contact
duration	Contact duration
campaign	number of contacts performed during this campaign to the client
pdays	number of days that passed by after the client was last contacted
previous	number of contacts performed before this campaign
poutcome	outcome of the previous marketing campaign
Subscribed (target)	has the client subscribed a term deposit?

Understanding Data:

Let us to begin with see at the current relationship sub-window. It shows the title of the right now stacked database. You'll conclude two focuses from this sub-window.

There are 31647 occasions, the number of columns within the table. The table contains 18 attributes-fields.

On the left, take note the sub window that shows the properties of each field within the database.

After you tap to choose an trait from the list, more nitty gritty information about the quality itself will be shown on the correct.

The visual representation of the lesson values can be found at the foot of the window.

The algorithm: J48 Decision Tree Algorithm [7]

Reason for choosing J48 Decision Tree Algorithm:

1. Compared to other it requires less effort for data preparation when pre-processing.
2. A decision tree does not require normalization of data.
3. A decision tree does not require scaling of data as well.
4. Missing values in the data also does NOT affect the process of building decision tree to any considerable extent.
5. A Decision trees model is very intuitive and easy to explain to technical teams as well as stakeholders.

III. RESULTS AND DISCUSSION

- **J48 decision tree algorithm with cross-validations.**

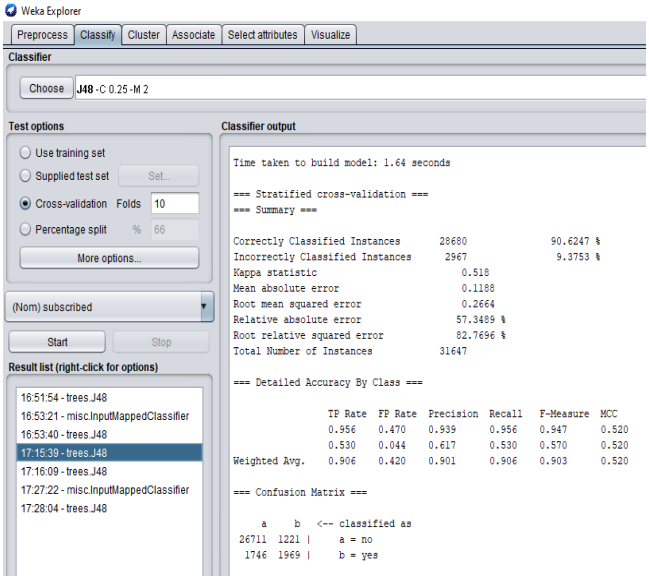

Figure 3: Cross validation

Results

Correctly classifies instances: 90.6247%

incorrectly classifies instances: 9.3753%

Predicted values:-
customer will not subscribe: 26711 . false positive:1221
customer will subscribe: 1969 . false positive: 1746

J48 decision tree algorithm with a 66% data split (for training and testing).

Results

Correctly classifies instances: 90.4554%
incorrectly classifies instances: 9.5446%

predicted values:-
customer will not subscribe: 9078 . false positive:421
customer will subscribe: 655 . false positive: 606

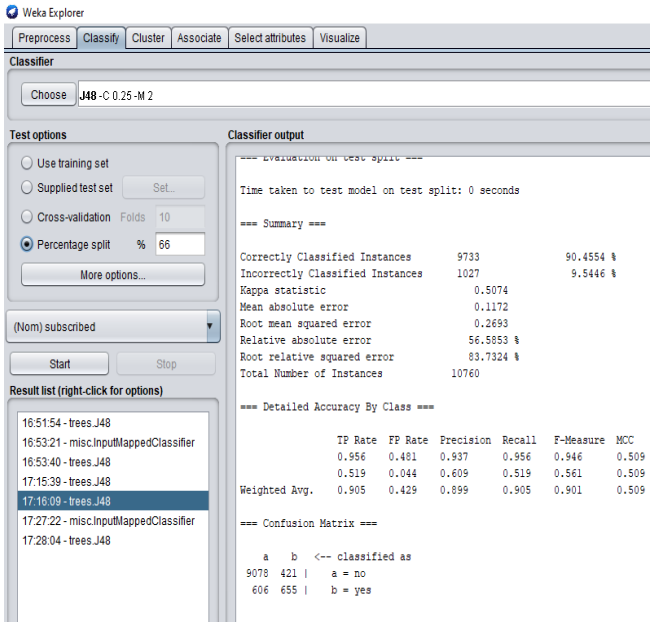

Figure 4: Percentage split

J48 decision tree algorithm with training set.

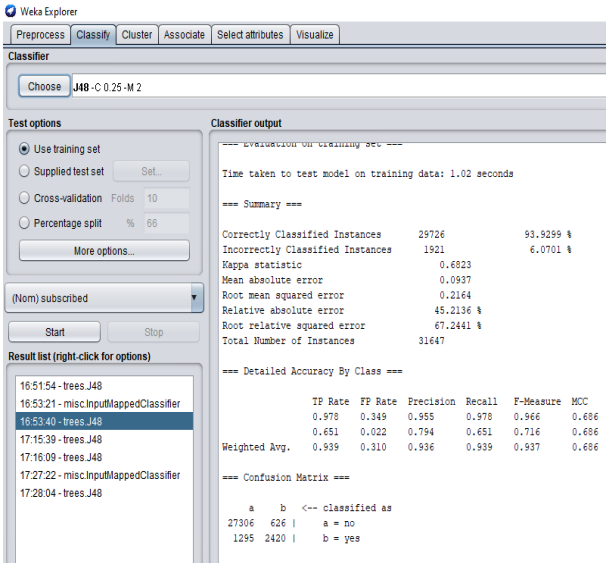

Figure 5: Using training set

Results

Correctly classifies instances: 93.9299%
Incorrectly classifies instances: 6.0701%

Predicted values:-
Customers will not subscribe: 27306 . false positive:626
Customers will subscribe: 2420 . false positive: 1295

Result of test

Figure 6: Result File

From the diagram we can see and conclude that the last two rows i.e. predicted result and actual result came out same, which means that our algorithm, which we trained, was able to correctly take the inputs and give the same results as that which came with the data, which proves how fast and accurate these algorithms is.

IV. CONCLUSION

One of the most successful ways to reach out to people is through telemarketing campaigns. They do however necessitate a significant expenditure because enormous call centres are used to carry out the campaigns. As a result, it's critical to identify the clients who are most likely to convert before making a call to them.

So, by this take a look at, we tend to foreseen however seemingly the client is to buy the term deposit that's provided by the retail banking establishment. The establishment can solely target those customers and therefore the total investment is going to be reduced. This may save time and cash of the establishment to a big quantity.

V. REFERENCES

[1]. Nasteski, Vladimir. (2017). An overview of the supervised machine learning methods. HORIZONS.B. 4. 51-62. 10.20544/HORIZONS.B.04.1.17.P05.

[2]. S. Chowdhury and M. P. Schoen, "Research Paper Classification using Supervised Machine Learning Techniques," 2020 Intermountain Engineering, Technology and Computing (IETC), 2020, pp. 1-6, doi: 10.1109/IETC47856.2020.9249211.

[3]. https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/last accessed on 19/April/2021

[4]. WEKA Explorer: Visualization, Clustering, Association Rule Mining. Software Testing Help. (2021, May 25). https://www.softwaretestinghelp.com/weka-explorer-tutorial/.

[5]. rushdishams. (2021,May 25). Weka Tutorial 03: Classification 101 using Explorer (Classification). YouTube. https://www.youtube.com/watch?v=bPrTeUAS6_I

[6]. UCI Machine Learning Repository. (2021, May 25). https://archive.ics.uci.edu/ml/index.php.

[7]. Bhargava, N., Girja Sharma, Ritu Bhargava and M. Mathuria. “Decision Tree Analysis on J48 Algorithm for Data Mining.” (2013).

Cite this article as :

Kumar Gaurav, Aditya Negi, "Prediction System for Customer Subscription to The Term Deposit", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 230-234, July-August 2021. Journal URL : https://ijsrcseit.com/CSEIT217449