START ML

DOCUMENTATION

TEAM MEMBERS:

PANJALA POOJITHA 2203A52047

LEKKALA VYSHNAVI 2203A52035

YENAGANDULA RAMYA 2203A52064

# PREDICTING ANALYSIS OF BANK MARKETING FOR FINANCIAL DECISION SUPPORT

ABSTRACT

In the dynamic landscape of the banking industry, effective marketing strategies play a pivotal role in attracting and retaining customers while driving business growth. This motive builds us a software tool for predictive analysis of bank marketing based on the data of customers. the success of telemarketing depends on various factors, the customer age, job, loan, status, education, marital, Housing loan, duration, pdays etc. hence, this factors contains various features analysed by data mining to predict customer tendency with respective marketing campaigns. machine learning help us to predict whether a person gets a loan or not based on person details and hence provide good decision marketing. the accuracy of logistic regression model is 0.767 and 0.769. the accuracy of KNN is 0.782. it aims to improve the bank marketing by targeting more customers, hitting right audiences.

1.INTRODUCTION:

Bank marketing refers to the strategic activities and process employed by financial institutions, such as bank to promote their produces and services to existing and potential customers. It helps us to get a loan in a bank by persons details, before only we can predict it. there are few feature like age, job, education, martial, duration etc… in this we include some model like logistic regression, perceptron, KNN, SVM(support vector machine), PCA(principal component analysis). It helps in improve in the banking sector. By improving the this customers can deposit they funds in it. Machine learning help us in predicting in the bank marketing.

2.Literature Review:

Mem, B., & wireko , has done his research on this bank marketing predicting in 2016. Now a days technology playing a import role , banking is positive relationship between the customers satisfaction. He tried to do the a new model of bank marketing to satifive the customers. They also wants banks to become the market leader in this competitive environment , it became more innovation in product and services.

Machogu, A.M., & Okiko, he also done research on this banking marketing in 2015. He showed the result that factories are leading customers satisfactions particularly in this banking .

3.Methodology

About data set:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sno | age | job | marital | education | balance | housing |
| 1 | 30 | unemployed | married | primary | 1789 | no |
| 2 | 55 | service | married | secondary | 1244 | Yes |
| 3 | 25 | management | single | tertiary | 3554 | no |
| 4 | 33 | management | married | tertiary | 4646 | no |
| 5 | 22 | blue-collar | married | secondary | 3232 | yes |

The dataset comes from the UCI Machine Learning repository, the data consists of certain no of columns they are:

Age, job, marital, education, default, balance, housing loan, contact, day, month, duration, campaign, pdays, previous, poutcome, deposit

From the above columns Age, job, marital, education, default, balance, housing loan, contact, day, month, duration, campaign, pdays, previous, poutcome are the features of the data and target variable is deposit.

#### Bank client data:

#### There are the few features of the client which check in the bank

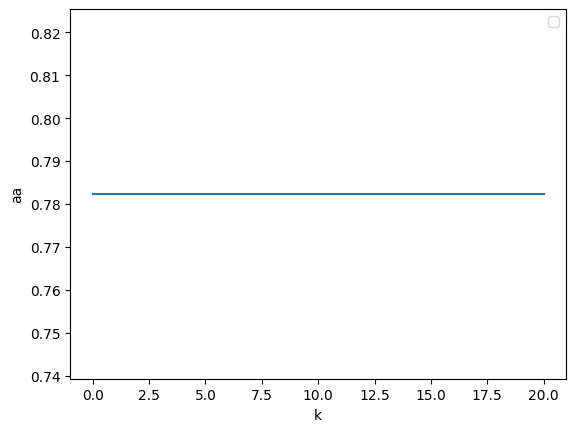
Sector there are: age is a column it contain age of person it ,job in this sector there are should be blue-collar , students, technical , unemployed etc.. there are marital the categorical are divorced, single, married . the education the categorical are secondary or primary. They are default are yes or no. the housing categorical also includes yes or no . contact either there telephone or cellular number should give it. Loan categorical either they have other loan in any other bank yes or no. days either mon, tue, wed, thus, fri, sat. duration like last contact duration or in seconds (numeric). Campaign the number of contact are done in during this campaign. p days the number of days data passed by the client was last contact done. Previous the number of call performed before this campaign for client also. Pout comes its also a column it contains failure or non-existent , success.

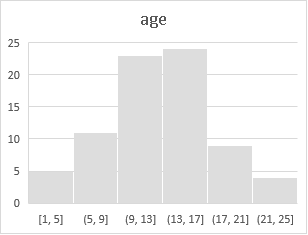
Target variable :

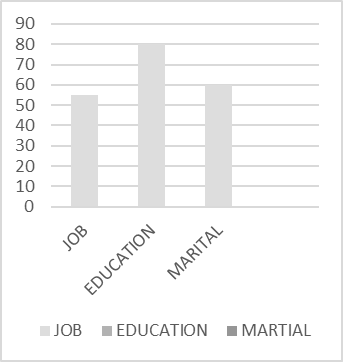
Deposit it’s a column it contain yes or no .

#### 

GRAPHS:



 Fig 1



THE ALGORITMS:

Logistic Regression:

it is a binary classification. if the target values are catogorial then we can apply supervised learning. we apply logistic regression.

Logistic regression follows the parametric method approach. Following equation with known and unknown variables.

Y=mx + c

Z=wo+w1x1+w2x2……………. Wnxn

Z values are continuous values so, to convert catogorial we required a activation function.

-y log (yp) – (1-y) log (1-yp)

Where,

(variance) Yp=1/1+e^-z ---🡪 sigmoid function

In logistic regression the activation function is sigmoid.

The output sigmoid equation contains only 0 to 1.

In logistic regression we will calculate cress entropy loss (log error).



**.: Logistic regression model graph**

**4**



3



**2**



**1**



**Perceptron:**

**Perceptron is inspired by the biologycial newron. it can do binary classification**

**constant**

**weights**

W0

weight sum

W1



W2

step function out

W3

The data set like x1, x2, x3……. X n. weights like w1, w2, w3 or random selected and one constant bias b.

Weighted sum => z=wi xi+b

Where z= w1 x1+ w2 x2…………wn xn+b

The activation function will convert linear dats to the non-linear data .in perceptron learning the activation fun is sigmoid. the output of the sigmoid function is 0 to 1.

SVM (support vector meachine):

SVM is the strongest model in the meachine learning . it will be applied for supervised learning data. it follow the parametric method approach. It can be used for both regression and classification .



support vector



fig1

the equation is x1^+ - x1^-=z/w

Hyper plane is a line which classific the given data correctly. + hyper and – hyper plan will be used to find out the supporting vectors. Positive side and negative side. With respect to the supporting vectors we are making hyper plane extractly middle.

Margine is the distance between positive hyper plane to negative hyper plane.

KNN(Nearest Neighbour):

k-Nearest Neighbour is a non-parametric method approach finds k -nearest neighbours from the all data points. In this method our target is to find optional k-value for the given data set. K value may be k=1,2,3—n here n is no of rows. In knn we will calculate distances between given data points to all data points from the given data set. after calculating distances we will select knn and the average y-value of knn is the y-value of given data point. For calculating the distances we can use Eucledian or manhatlern distance |x2-x1|+|y2-y1|



PCA(Principal component analysis):

Principle component analysis is used to reduce the dimension like x1 and x2 are the two co-ordinates we can convert into one coordinates (dimension).pca is an unsupervised learning.to reduce dimensionality it will calculate eigen values and eigen vectors.

Steps to implement pca(principle component analysis) are:

1.take the original data like x1,x2----

2.calculate the mean of individual features.

3.find co-variance matrix.

4.find out eigen values and eigen vector for the co-variance matrix.

5.normalize the given eigen values.

6.find principle components.

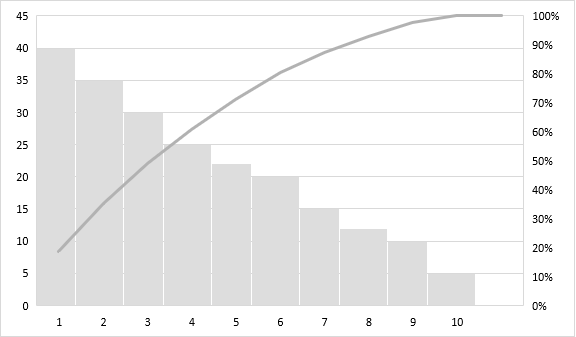


Fig1

4.RESULT:

RESULT TABLE BEFORE USING PCA:

|  |  |  |
| --- | --- | --- |
| SNO | MODEL | ACCURACY |
| 1 | LOGISTIC REGRESSION | 0.767129422301 |
| 2 | SUPPORT VECTOR MEACHINR | 0.79516183223 |
| 3 | KNN | 0.7823555754590238 |
| 4 | PERCEPTRON | 0.729064039408867 |

RESULT TABLE AFTER USING PCA:

|  |  |  |
| --- | --- | --- |
| SNO | MODEL | ACCURACY |
| 1 | LOGISTIC REGRESSION | 0.5893416927899686 |
| 2 | SUPPORT VECTOR MEACHINR | 0.6318853560232871 |
| 3 | PERCEPTRON | 0.5320197044334976 |

5.CONCLUSION:

Bank marketing is a virtual for attracting and retaining bank customer. A customer needs to evaluate bank continually adapt there marketing strategies and building trust and creating a new experience for customers and bank marketing offering personalized solutions for the customers.

ACKNOWLEDGMENT:

We thank SR University and Ramesh sir.

REFERENCE:

1.kaggle (UCI respiratory machine learning).

2.Ramesh sir.