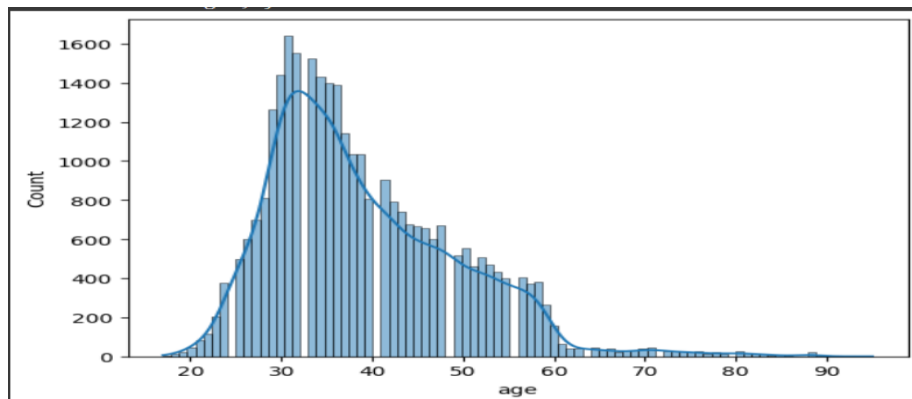


Customer Subscription Prediction with Machine Learning

- 1) What is the distribution of the customer ages ?

Customer age ranges from 17 – 98 years in the bank data .



- 2) What is the relationship between customer age and subscription ?

According to the correlation matrix , the relation between the age and subscription is Inverse . Chances decrease when age is more . The Correlation coefficient is -0.003803.

- 3) Are there any other factors that are correlated with subscription ?

Yes, all the factors are related with the subscription but contact , duration , pdays , previous , poutcome , emp_var_rate , cons_price_idx , euribor3m , nr_employed

These create more stronger dependency of the subscription. Correlation values are : -

age	-0.003803
job	0.014110
marital	0.048599
education	0.058077
housing	0.011219
loan	-0.003005
contact	-0.137026
month	-0.021121
day_of_week	0.020395
duration	0.405113
campaign	-0.063838
pdays	-0.321491
previous	0.220896
poutcome	0.140126
emp_var_rate	-0.288201
cons_price_idx	-0.112447
cons_conf_idx	0.024003
euribor3m	-0.299976
nr_employed	-0.348121

4) What is the accuracy of the logistic regression model ?

The Accuracy of the Logistic Regression model maximises to 93 % if we keep the training and testing data random . Mostly it is around 90%.

5) What are the most important features for the logistic regression model ?

The importance of a feature depends on the its Coefficient value magnitude .

For bank data the following are the values of most important features : -

campaign	----->	1.441
cons_price_idx	----->	0.9522
emp_var_rate	----->	0.8208
euribor3m	----->	0.7188
education	----->	0.3178
month	----->	0.2015
cons_conf_idx	----->	0.1967
previous	----->	0.1804

6) What is the precision of the logistic regression model ?

The precision of the Logistic regression model ranges from 60 % to 70% for randomly selected data .

7) What is the recall of the logistic regression model ?

The Recall score of Logistic regression model is from 35% to 55% for random data.

8) What is the f1-score of the logistic regression model ?

The F-1 Score of the Logistic Regression Model is around 50% .

9) How can you improve the performance of the logistic regression model ?

Performance can be improved by the following methods : -

- Removing noise and missing values.
- Normalizing the data.
- Hyperparameter Tuning and balancing classes.
- Using Optimizers.

10) What are the limitations of the logistic regression model ?

Limitations are : -

- Logistic regression assumes that the relationship between the independent and dependent variables is linear
- Logistic regression is prone to overfitting, especially when there are a large number of independent variables relative to the number of observations
- Multicollinearity can make it difficult to interpret the results of the model and can also lead to overfitting.