The Given Dataset shows us about the “ **Purchased Product** ”which customers had purchased the particular product “ **ABC** ” and based on the given data we will train the model and predict that whether the particular customer will buy the product or not.

We can see that the customers belong to different group and Category and each has individual rating.

Overall, The data given is Categorical (**Group,Category,Rating,Purchased\_Product\_ABC**) and Continuous data (**Var1,Var2**)so I have used Decision tree as it is one of the best model to use.

A decision tree is a map of the possible outcomes of a series of related choices. It allows to weigh possible actions against one another based on their costs, probabilities, and benefits. They can be used either to drive informal discussion or to map out an algorithm that predicts the best choice mathematically.

**Confusion Matrix**

It is the easiest way to measure the performance of a classification problem where the output can be of two or more type of classes. A confusion matrix is nothing but a table with two dimensions. “Actual” and “Predicted” and furthermore, both the dimensions have True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN)

Here in our model

|  |  |  |
| --- | --- | --- |
| **1** | **1**  **True Positive**  **2734** | **0**  **False Positive**  **1150** |
| **0** | **False Negative**  **1121** | **True Negative**  **2703** |

Out of 7708 we predicted 5437 are correctly predicted 2271 were Wrongly predicted.

**Classification Accuracy**

Accuracy=TP+TN/TP+FP+FN+TN = 2734+2703/2734+1150+1121+2703 = 0.705

Accuracy of our model : 70.5%

**Classification Report**

This report consists of the scores of Precisions, Recall, F1 and Support. They are explained as follows −

Precision

Precision, used in document retrievals, may be defined as the number of correct documents returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Precision=TP/TP+FP =

Recall or Sensitivity

Recall may be defined as the number of positives returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Recall=TP/TP+FN

Specificity

Specificity, in contrast to recall, may be defined as the number of negatives returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Specificity=TN/TN+FP

Support

Support may be defined as the number of samples of the true response that lies in each class of target values.

F1 Score

This score will give us the harmonic mean of precision and recall. Mathematically, F1 score is the weighted average of the precision and recall. The best value of F1 would be 1 and worst would be 0. We can calculate F1 score with the help of following formula −

𝑭𝟏 = 𝟐 ∗ (𝒑𝒓𝒆𝒄𝒊𝒔𝒊𝒐𝒏 ∗ 𝒓𝒆𝒄𝒂𝒍𝒍) / (𝒑𝒓𝒆𝒄𝒊𝒔𝒊𝒐𝒏 + 𝒓𝒆𝒄𝒂𝒍𝒍)

F1 score is having equal relative contribution of precision and recall.

.

### AUC (Area Under ROC curve)

AUC (Area Under Curve)-ROC (Receiver Operating Characteristic) is a performance metric, based on varying threshold values, for classification problems. As name suggests, ROC is a probability curve and AUC measure the separability. In simple words, AUC-ROC metric will tell us about the capability of model in distinguishing the classes. Higher the AUC, better the model.

Variable 2 (Var2) has high correlation so and it has more influence on sales.

Group1 members are showing less interest in products only 48% purchased our products so we can reach them to increase sales of product ABC.