DATA ANALYTICS WITH COGNOS – GROUP 3

CUSTOMER CHURN PREDICTION

PHASE4

DEVELOPMENT PART 2

MODEL BUILDING:

We have completed the data preprocessing in the previous phase, now we build a model that can predict the churning of customers.

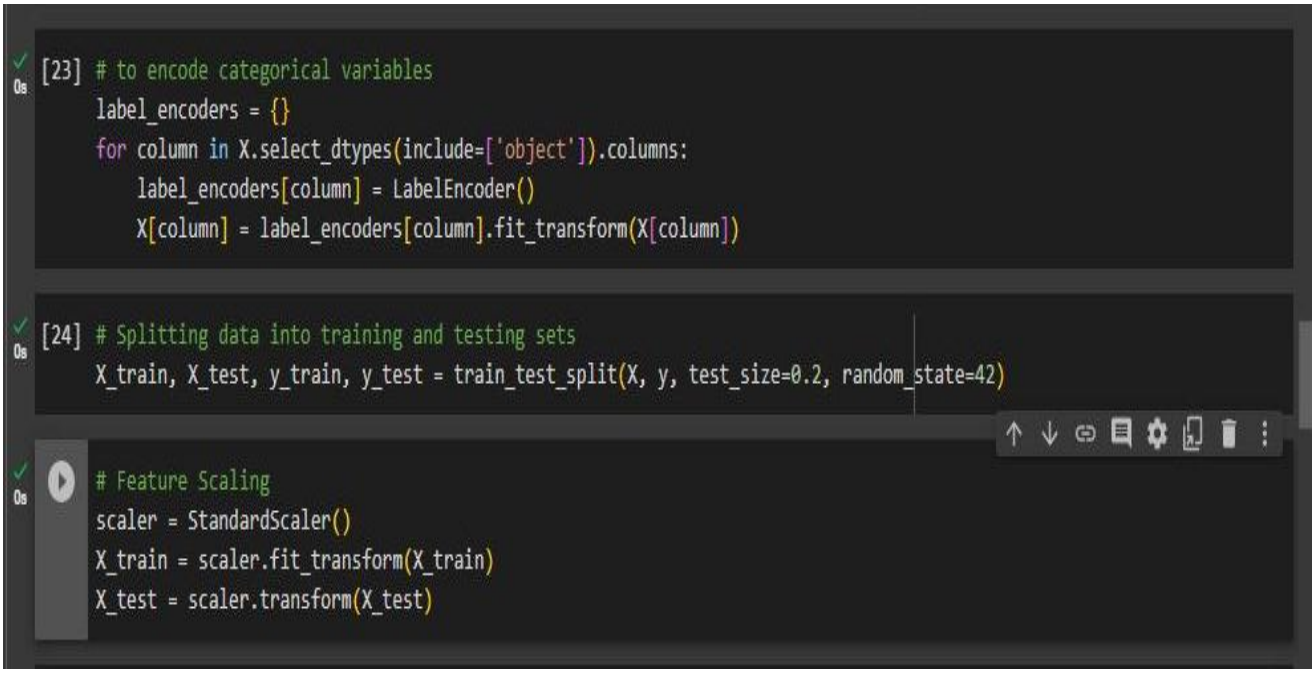
STEP1: IMPORTING LIBRARIES

We import the libraries necessary for train and test split of data, to encode categorical variables, for feature scaling, to import the model and the metrics to calculate the accuracy and confusion matrix of the model.

We have imported the Random Forest Classifier model from sklearn.ensemble as the algorithm we have chosen for our prediction is Random Forest Algorithm. We have also imported the train_test_split from sklearn.model_selection in order to split the data into training and validation set for the model to learn the training data.

STEP2: FITTING THE DATA

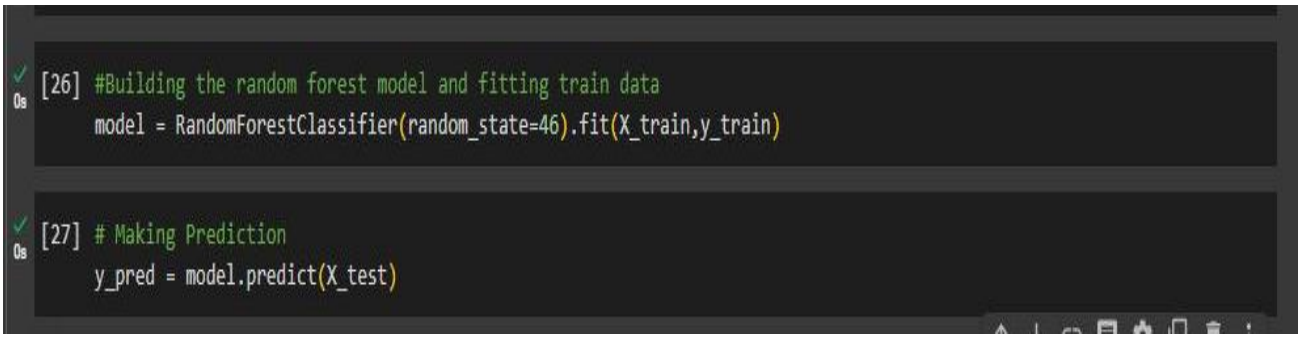
For the data to fit the model we encode the categorical variables using label encoder, Split the data into training and testing set for the model to learn the training data using train_test_split, we perform feature scaling using standard scaler.



STEP3: MODEL BUILDING AND PREDICTION

We build the RandomForestClassifier model and assign the random_state as 46 for the consistency of the model and fit the X_train and y_train data into the model.

Then we make prediction using the model.predict() function by giving the X_test as the input.



STEP4: EVALUATION OF MODEL

After the prediction is over we evaluate the data by attaining its accuracy using accuracy_score(), confusion matrix using confusion_matrix() and the overall classification report using classification_report().

```
[21] #attaining the overall classification report
     print(classification_report(y_test,y_pred))

                  precision    recall  f1-score   support

              No       0.83      0.91      0.87      1036
             Yes       0.66      0.47      0.55       373

        accuracy                           0.79      1409
       macro avg       0.74      0.69      0.71      1409
    weighted avg       0.78      0.79      0.78      1409
```

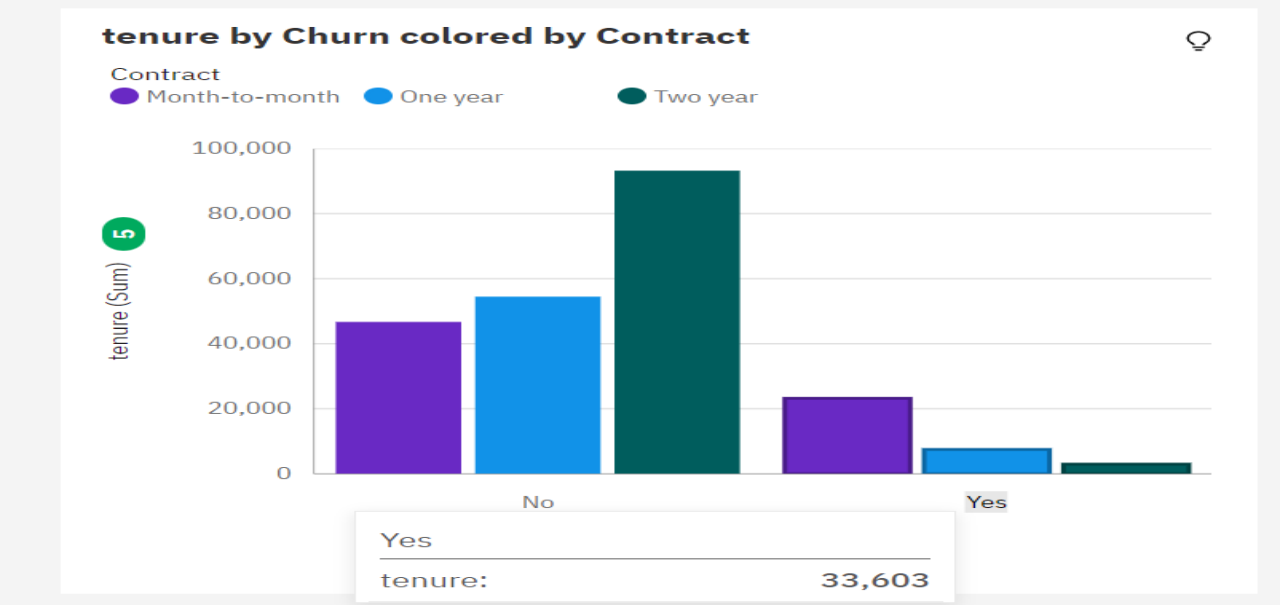
DATA ANALYSIS AND VISUALIZATION:

1) Payment method and churn:

To analyze the effect of payment method on the churners we create a bullet chart to visualize the relationship between them.

It is analyzed that the churners count is higher in the Electronic check payment method than any other payment method.

2) TENURE AND CONTRACT WITH RESPECT TO CHURN:



Using the bar plot,We analyze the relationship between churners and their tenure and contract period.

It is clearly observed that the churner decreases when the tenure is high and the contract is of long period like two year.

CONCLUSION:

In this phase we have built the model and had fit the data to the model and had made prediction. We have also made few analyses by visualizing using Cognos.