



# Machine Learning Techniques KAI-651

## **PRACTICAL-02**

```

import pandas as pd
df = pd.read_csv('/content/churnData.csv')
df.head()
df.info()
churn_df = df[['tenure', 'age', 'address', 'income', 'ed', 'employ', 'equip',
'callcard', 'wireless', 'churn']]
churn_df.head()
from sklearn.model_selection import train_test_split
train_set, test_set = train_test_split(churn_df, test_size=0.2, random-
state=42)
train_set.head()
X_train = train_set[['tenure', 'age', 'address', 'income', 'ed', 'employ',
'equip', 'callcard', 'wireless']]
Y_train = train_set['churn']
X_test = test_set[['tenure', 'age', 'address', 'income', 'ed', 'employ',
'equip', 'callcard', 'wireless']]
for col in X_train.columns.to_list():
    X_train[col] = (X_train[col] - X_train[col].mean()) / X_train[col].std()
for col in X_test.columns.to_list():
    X_test[col] = (X_test[col] - X_test[col].mean()) / X_test[col].std()
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix
LR = LogisticRegression(C=0.01, solver='liblinear').fit(X_train, Y_train)
LR
Yhat = LR.predict(X_test)
Yhat
Yhat_prob = LR.predict_proba(X_test)
Yhat_prob

```

## Experiment-2 [ Logistic Regression ]

**Aim:** The aim of this experiment is to create and test a Logistic regression model on the 'customer churn' dataset to predict how many customers will stay in the company.

**Dataset:** The dataset used for this experiment is 'ChurnData.csv'.

**Objective:**

1. Utilize scikit Logistic Regression to classify
2. Understand confusion matrix.

### Procedure.

#### 1. Import Necessary Libraries

Import the required libraries for data manipulation and model implementation.

#### 2. Get Data

Load the dataset 'ChurnData.csv' into a pandas Dataframe.

#### 3. Explore Data

Inspect the dataset to understand its structure, datatypes & any missing values.

#### 4. Select Specific Features and create New Dataframe 'churn\_df'

Choose relevant features for analysis and convert the target variable to integer type.

#### 5. Train-Test Split



Teacher's Signature : \_\_\_\_\_

```
from sklearn.metrics import jaccard_score
```

```
jaccard_score(y_test, yhat, pos_label=0)
```

```
cnf_matrix = confusion_matrix(y_test, yhat, labels=[1,0])
```

```
cnf_matrix
```



Divide the dataset into training and testing sets to train the model on a subset of data and evaluate its performance on unseen data.

6. Normalize Feature Values

Normalize the feature values to ensure that all features contribute equally to the model.

7. Training

Train the Logistic Regression model using the training data.

8. Predict Target Variable on Test Set

Use the trained model to predict the target variable on the test set.

9. Predicted Probabilities

Obtain the predicted probabilities of the target variable from the model.

10. Evaluation

Evaluate the performance of the model using metrics such as Jaccard index.

11. Confusion Matrix

Generate a confusion matrix to visualise the model's performance in terms of true positive, false positive, true negative and false negative predictions.

Conclusion:

Summarised the findings of the experiment and the effectiveness of the logistic regression model in predicting customer churn.

