Assignment

Project 1 : KNN Algorithm

Program: To Find the KNN Accuracy score & Plot the Graph for one dependant variable & others as Independent variables.

```python
import pandas as pd

from sklearn import preprocessing

from sklearn.model_selection import train_test_split

from sklearn.metrics import confusion_matrix

from sklearn import neighbors

import matplotlib.pyplot as plt


SVM_train = pd.read_csv("D:/AI_ML_Course/Day24/train.csv")

def cal_Knn(SVM_train):

    SVM_train = SVM_train.drop(["Name", "Ticket", "Cabin", "PassengerId"], axis = 1)


    le = preprocessing.LabelEncoder()

    le.fit(SVM_train["Sex"])

    SVM_train["Sex"] = le.transform(SVM_train["Sex"])

    SVM_train["Embarked"] = le.fit_transform(SVM_train["Embarked"])

    SVM_train["Age"] = SVM_train["Age"].astype(int)

    SVM_train["Fare"] = SVM_train["Fare"].astype(int)

    SVM_var = list(SVM_train.columns)

    print(SVM_var)

    # For Loop to iterate through all the variables
    for i in SVM_var:

        y = SVM_train[i]
```

```python
    X = SVM_train.drop([i], axis=1)

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)

    Knn_score = []

    iteration = []

    # While loop to iterate K values from 1 to 5

    n_neighbors = 1

    while n_neighbors < 6:

        KNN = neighbors.KNeighborsClassifier(n_neighbors)

        Knn_ascore = KNN.fit(X_train, y_train).score(X_test, y_test)

        print("The KNN Accuracy score for iteration:",n_neighbors, "is:", Knn_ascore, "for the variable
as", i)

        Knn_score.append(Knn_ascore)

        iteration.append(n_neighbors)

        n_neighbors += 1

        y_pred = KNN.predict(X_test)

        c_m = confusion_matrix(y_test, y_pred)

        print("The confusion matrix is\n", c_m, "for the variable as:",i)


    # Print the KNN accuracy score

    print("The Knn accuracy score for", i ,"all the iteration is:",Knn_score)

    # Plot the Bar Graph

    plt.bar(iteration, Knn_score, tick_label=iteration, width=0.2, color=["Red", "Green"])

    plt.ylabel(i)

    plt.xlabel('K-Value')

    plt.show()

cal_Knn(SVM_train)
```

['Survived', 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']

The KNN Accuracy score for iteration: 1 is: 0.700374531835206 for the variable as Survived

The confusion matrix is

 [[123 34]

 [46 64]] for the variable as: Survived

The KNN Accuracy score for iteration: 2 is: 0.6554307116104869 for the variable as Survived

The confusion matrix is

 [[142 15]

 [77 33]] for the variable as: Survived

The KNN Accuracy score for iteration: 3 is: 0.6891385767790262 for the variable as Survived

The confusion matrix is

 [[122 35]

 [48 62]] for the variable as: Survived

The KNN Accuracy score for iteration: 4 is: 0.6591760299625468 for the variable as Survived

The confusion matrix is

 [[133 24]

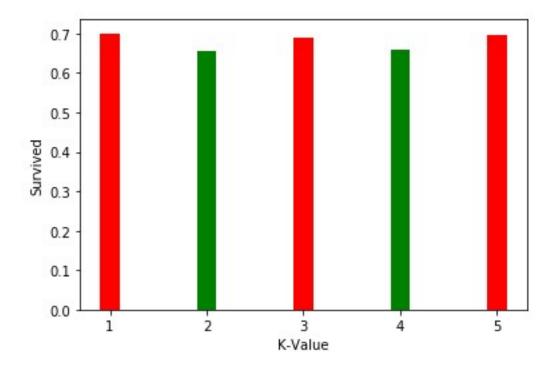 [67 43]] for the variable as: Survived

The KNN Accuracy score for iteration: 5 is: 0.6966292134831461 for the variable as Survived

The confusion matrix is

 [[125 32]

 [49 61]] for the variable as: Survived

The Knn accuracy score for Survived all the iteration is: [0.700374531835206, 0.6554307116104869, 0.6891385767790262, 0.6591760299625468, 0.6966292134831461]

Output: For K=1, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.8876404494382022 for the variable as Pclass

The confusion matrix is

 [[63 6 1]

 [3 36 10]

 [1 9 138]] for the variable as: Pclass

The KNN Accuracy score for iteration: 2 is: 0.8614232209737828 for the variable as Pclass

The confusion matrix is

 [[66 3 1]

 [8 37 4]

 [5 16 127]] for the variable as: Pclass

The KNN Accuracy score for iteration: 3 is: 0.8576779026217228 for the variable as Pclass

The confusion matrix is

 [[62 6 2]

[8 28 13]

[3   6 139]] for the variable as: Pclass

The KNN Accuracy score for iteration: 4 is: 0.8239700374531835 for the variable as Pclass

The confusion matrix is

[[60   7   3]

[8 31 10]

[5 14 129]] for the variable as: Pclass

The KNN Accuracy score for iteration: 5 is: 0.8164794007490637 for the variable as Pclass
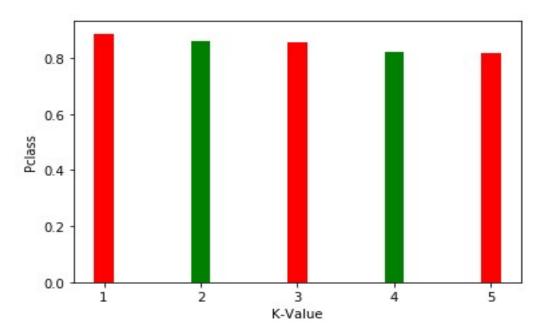
The confusion matrix is

[[57 10   3]

[8 26 15]

[4   9 135]] for the variable as: Pclass

The Knn accuracy score for Pclass all the iteration is: [0.8876404494382022, 0.8614232209737828, 0.8576779026217228, 0.8239700374531835, 0.8164794007490637]

Output: For K=1, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.6966292134831461 for the variable as Sex

The confusion matrix is

 [[57 41]

 [40 129]] for the variable as: Sex

The KNN Accuracy score for iteration: 2 is: 0.5917602996254682 for the variable as Sex

The confusion matrix is

 [[70 28]

 [81 88]] for the variable as: Sex

The KNN Accuracy score for iteration: 3 is: 0.6404494382022472 for the variable as Sex

The confusion matrix is

 [[49 49]

 [47 122]] for the variable as: Sex

The KNN Accuracy score for iteration: 4 is: 0.6367041198501873 for the variable as Sex

The confusion matrix is

 [[64 34]

 [63 106]] for the variable as: Sex

The KNN Accuracy score for iteration: 5 is: 0.704119850187266 for the variable as Sex
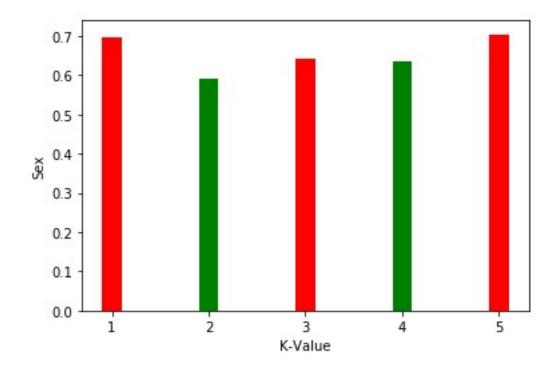
The confusion matrix is

 [[56 42]

 [37 132]] for the variable as: Sex

The Knn accuracy score for Sex all the iteration is: [0.6966292134831461, 0.5917602996254682, 0.6404494382022472, 0.6367041198501873, 0.704119850187266]

Output: For K=5, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.11610486891385768 for the variable as Age

The confusion matrix is

[[0 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Age

The KNN Accuracy score for iteration: 2 is: 0.06367041198501873 for the variable as Age

The confusion matrix is

[[1 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Age

The KNN Accuracy score for iteration: 3 is: 0.08239700374531835 for the variable as Age

The confusion matrix is

[[1 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Age

The KNN Accuracy score for iteration: 4 is: 0.09737827715355805 for the variable as Age

The confusion matrix is

[[1 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Age

The KNN Accuracy score for iteration: 5 is: 0.12359550561797752 for the variable as Age

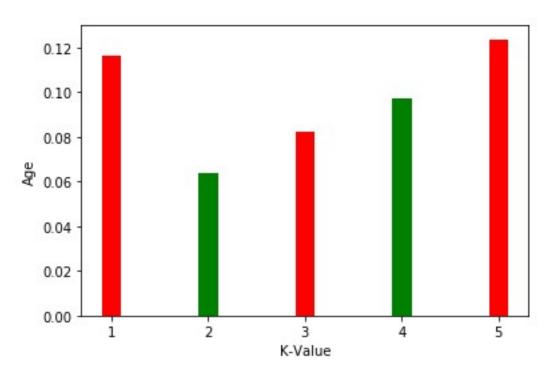The confusion matrix is

[[1 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Age

The Knn accuracy score for Age all the iteration is: [0.11610486891385768, 0.06367041198501873, 0.08239700374531835, 0.09737827715355805, 0.12359550561797752]

Output: For K=5, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.7265917602996255 for the variable as SibSp

The confusion matrix is

[[149 26 5 0 2 0 0]

[25 35 5 0 0 0 0]

[4 1 3 0 0 0 0]

[0 3 1 3 0 0 0]

[0 0 1 0 1 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the variable as: SibSp

The KNN Accuracy score for iteration: 2 is: 0.7265917602996255 for the variable as SibSp

The confusion matrix is

[[165 15 0 0 2 0 0]

[40 22 3 0 0 0 0]

[5 2 1 0 0 0 0]

[2 2 1 2 0 0 0]

[0 0 1 0 1 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the variable as: SibSp

The KNN Accuracy score for iteration: 3 is: 0.7340823970037453 for the variable as SibSp

The confusion matrix is

[[157 22 1 0 2 0 0]

[27 31 3 2 2 0 0]

[5 1 1 1 0 0 0]

[1 3 0 3 0 0 0]

[0 1 0 0 1 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the variable as: SibSp

The KNN Accuracy score for iteration: 4 is: 0.7078651685393258 for the variable as SibSp

The confusion matrix is

[[163 16 1 0 2 0 0]

[38 20 3 2 2 0 0]

[5 1 0 2 0 0 0]

[1 4 0 2 0 0 0]

[0 0 1 0 1 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the variable as: SibSp

The KNN Accuracy score for iteration: 5 is: 0.7265917602996255 for the variable as SibSp

The confusion matrix is

[[159 20 1 0 2 0 0]

[30 30 3 0 1 0 1]

[5 1 1 1 0 0 0]

[2 5 0 0 0 0 0]
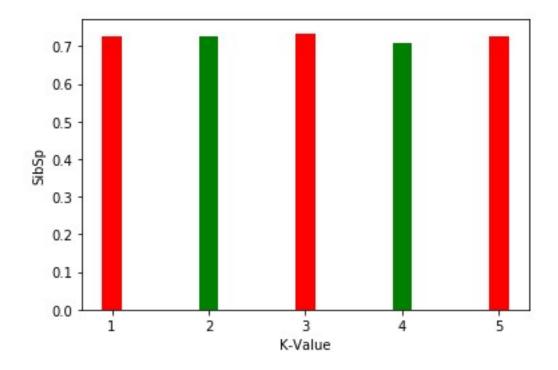
[0 0 1 0 1 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the variable as: SibSp

The Knn accuracy score for SibSp all the iteration is: [0.7265917602996255, 0.7265917602996255, 0.7340823970037453, 0.7078651685393258, 0.7265917602996255]

Output: For K=3, the accuracy score is highest

The KNN Accuracy score for iteration: 1 is: 0.7677902621722846 for the variable as Parch

The confusion matrix is

[[180 12 4 0 2 0]

 [21 13 4 0 0 2]

 [7 8 12 0 0 0]

 [1 0 0 0 0 0]

 [1 0 0 0 0 0]

 [0 0 0 0 0 0]] for the variable as: Parch

The KNN Accuracy score for iteration: 2 is: 0.7902621722846442 for the variable as Parch

The confusion matrix is

[[192 5 1 0 0]

 [27 12 1 0 0]

 [15 5 7 0 0]

 [1 0 0 0 0]

[1 0 0 0 0]] for the variable as: Parch

The KNN Accuracy score for iteration: 3 is: 0.7865168539325843 for the variable as Parch

The confusion matrix is

[[190 7 1 0 0]

[26 13 1 0 0]

[15 5 7 0 0]

[1 0 0 0 0]

[1 0 0 0 0]] for the variable as: Parch

The KNN Accuracy score for iteration: 4 is: 0.7865168539325843 for the variable as Parch

The confusion matrix is

[[193 4 1 0 0]

[29 11 0 0 0]

[15 6 6 0 0]

[0 1 0 0 0]

[1 0 0 0 0]] for the variable as: Parch

The KNN Accuracy score for iteration: 5 is: 0.7865168539325843 for the variable as Parch

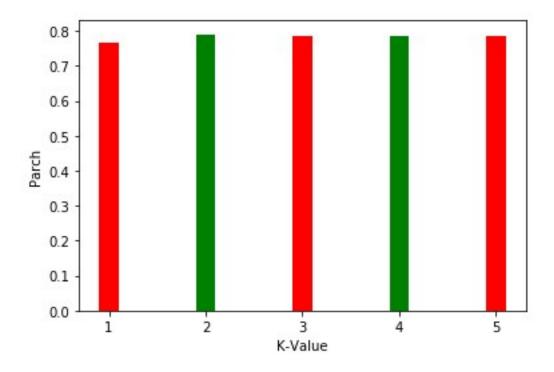The confusion matrix is

[[193 4 1 0 0]

[28 9 3 0 0]

[15 4 8 0 0]

[0 1 0 0 0]

[1 0 0 0 0]] for the variable as: Parch

The Knn accuracy score for Parch all the iteration is: [0.7677902621722846, 0.7902621722846442, 0.7865168539325843, 0.7865168539325843, 0.7865168539325843]

Output: For K=2, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.27340823970037453 for the variable as Fare

The confusion matrix is

[[3 0 2 ... 0 0 0]

[0 0 2 ... 0 0 0]

[0 0 37 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 1]] for the variable as: Fare

The KNN Accuracy score for iteration: 2 is: 0.2883895131086142 for the variable as Fare

The confusion matrix is

[[3 0 2 ... 0 0 0]

[0 0 3 ... 0 0 0]

[0 0 53 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Fare

The KNN Accuracy score for iteration: 3 is: 0.27715355805243447 for the variable as Fare

The confusion matrix is

[[3 0 1 ... 0 0 0]

[0 0 2 ... 0 0 0]

[0 0 51 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 1 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Fare

The KNN Accuracy score for iteration: 4 is: 0.27340823970037453 for the variable as Fare

The confusion matrix is

[[3 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]

[0 0 0 ... 0 0 0]] for the variable as: Fare

The KNN Accuracy score for iteration: 5 is: 0.2696629213483146 for the variable as Fare
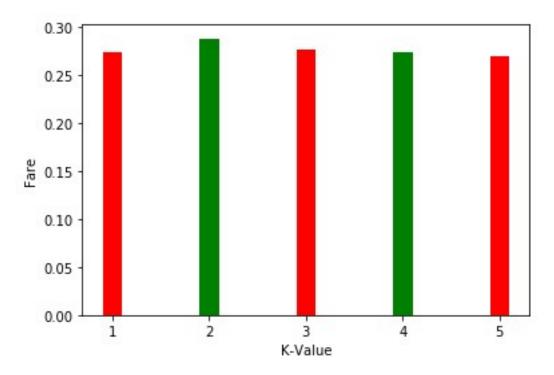
The confusion matrix is

[[2 0 1 ... 0 0 0]

[0 0 2 ... 0 0 0]

[1 1 51 ... 0 0 0]

...

[0 0 0 ... 0 0 0]

[0 0 1 ... 0 0 0]

[0 0 0 ... 0 0 1]] for the variable as: Fare

The Knn accuracy score for Fare all the iteration is: [0.27340823970037453, 0.2883895131086142, 0.27715355805243447, 0.27340823970037453, 0.2696629213483146]

Output: For K=2, the accuracy score is highest.

The KNN Accuracy score for iteration: 1 is: 0.7528089887640449 for the variable as Embarked

The confusion matrix is

[[31 2 16]

[2 10 10]

[28   8 160]] for the variable as: Embarked

The KNN Accuracy score for iteration: 2 is: 0.6554307116104869 for the variable as Embarked

The confusion matrix is

[[38   1 10]

[7   5 10]

[54 10 132]] for the variable as: Embarked

The KNN Accuracy score for iteration: 3 is: 0.700374531835206 for the variable as Embarked

The confusion matrix is

[[17   3 29]

[1   8 13]

[30   4 162]] for the variable as: Embarked

The KNN Accuracy score for iteration: 4 is: 0.7228464419475655 for the variable as Embarked

The confusion matrix is

[[24   5 20]

[2 10 10]

[27 10 159]] for the variable as: Embarked

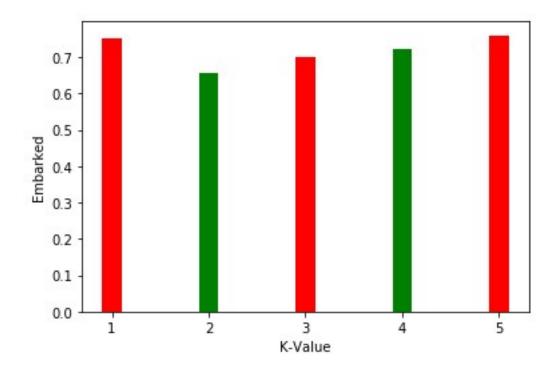The KNN Accuracy score for iteration: 5 is: 0.7602996254681648 for the variable as Embarked

The confusion matrix is

[[21   5 23]

[1   9 12]

[18   5 173]] for the variable as: Embarked

The Knn accuracy score for Embarked all the iteration is: [0.7528089887640449, 0.6554307116104869, 0.700374531835206, 0.7228464419475655, 0.7602996254681648]

Output: For K=5, the accuracy score is highest.

Project 2: SVM (Support Vector Machine)

Program: To Find the SVM Accuracy score for One Dependent variable & others as Independent variables

```python
import pandas as pd

from sklearn import preprocessing

from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score

from sklearn.metrics import confusion_matrix

from sklearn import svm as svm


SVM_train = pd.read_csv("D:/AI_ML_Course/Day24/train.csv")

def cal_Knn(SVM_train):

    SVM_train = SVM_train.drop(["Name", "Ticket", "Cabin"], axis = 1)

    le = preprocessing.LabelEncoder()

    le.fit(SVM_train["Sex"])


    SVM_train["Sex"] = le.transform(SVM_train["Sex"])

    SVM_train["Embarked"] = le.fit_transform(SVM_train["Embarked"])

    SVM_train["Age"] = SVM_train["Age"].astype(int)

    SVM_train["Fare"] = SVM_train["Fare"].astype(int)


    SVM_var = list(SVM_train.columns)

    print(SVM_var)

    # For Loop to iterate through all the variables

    for i in SVM_var:

        y = SVM_train[i]
```

```python
        X = SVM_train.drop([i, "PassengerId"], axis=1)

        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)


        clf = svm.SVC(gamma=0.01, C=100)

        clf.fit(X_train, y_train)


        y_pred = clf.predict(X_test)

        acc_score = accuracy_score(y_test, y_pred, normalize=True)

        print("The SVM accuracy score is", acc_score, "for the dependant variable as:", i)

        c_m = confusion_matrix(y_test, y_pred)

        print("The SVM confusion matrix is:\n", c_m, "for the dependant variable as:", i)


cal_Knn(SVM_train)
```

Solution

['PassengerId', 'Survived', 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']

The SVM accuracy score is 0.0 for the dependant variable as: PassengerId

The SVM confusion matrix is:

 [[0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 ...

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]] for the dependant variable as: PassengerId

The SVM accuracy score is 0.7415730337078652 for the dependant variable as: Survived

The SVM confusion matrix is:

 [[124 33]

 [36 74]] for the dependant variable as: Survived

The SVM accuracy score is 0.9063670411985019 for the dependant variable as: Pclass

The SVM confusion matrix is:

 [[63   5   2]

 [4 36   9]

 [0   5 143]] for the dependant variable as: Pclass

The SVM accuracy score is 0.7153558052434457 for the dependant variable as: Sex

The SVM confusion matrix is:

 [[62 36]

 [40 129]] for the dependant variable as: Sex

The SVM accuracy score is 0.17228464419475656 for the dependant variable as: Age

The SVM confusion matrix is:

 [[2 0 1 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 ...

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]

 [0 0 0 ... 0 0 0]] for the dependant variable as: Age

The SVM accuracy score is 0.7640449438202247 for the dependant variable as: SibSp

The SVM confusion matrix is:

[[152 22 6 0 2 0 0]

[20 41 4 0 0 0 0]

[4 1 3 0 0 0 0]

[0 3 1 3 0 0 0]

[0 0 0 0 2 0 0]

[0 0 0 0 0 1 0]

[0 0 0 0 0 0 2]] for the dependant variable as: SibSp

The SVM accuracy score is 0.7827715355805244 for the dependant variable as: Parch

The SVM confusion matrix is:

[[184 8 6 0 0 0]

[22 14 3 0 0 1]

[9 7 11 0 0 0]

[0 0 1 0 0 0]

[1 0 0 0 0 0]

[0 0 0 0 0 0]] for the dependant variable as: Parch

The SVM accuracy score is 0.38202247191011235 for the dependant variable as: Fare

The SVM confusion matrix is:

[[3 0 3 … 0 0 0]

[0 0 3 … 0 0 0]

[0 1 61 … 0 0 0]

…

[0 0 0 … 0 0 0]

[0 0 0 … 0 0 0]

[0 0 0 … 0 0 0]] for the dependant variable as: Fare

The SVM accuracy score is 0.7490636704119851 for the dependant variable as: Embarked

The SVM confusion matrix is:

 [[21 2 26]

 [0 8 14]

 [20 5 171]] for the dependant variable as: Embarked