**Case Study : Classify the type of avocado using SVM**[**¶**](#gjdgxs)

The avocado is a rather unique fruit. While most fruit consists primarily of carbohydrate, avocado is high in healthy fats. Numerous studies show that it has powerful health benefits. The dataset contains 14 feature Target column is Type

In [1]:

**import** **pandas** **as** **pd**

In [2]:

df = pd.read\_csv('avocado.csv')

In [3]:

df.tail(9)

Out[3]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Unnamed: 0** | **Date** | **AveragePrice** | **Total Volume** | **4046** | **4225** | **4770** | **Total Bags** | **Small Bags** | **Large Bags** | **XLarge Bags** | **type** | **year** | **region** |
| **18240** | 3 | 2018-03-04 | 1.54 | 17393.30 | 1832.24 | 1905.57 | 0.00 | 13655.49 | 13401.93 | 253.56 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18241** | 4 | 2018-02-25 | 1.57 | 18421.24 | 1974.26 | 2482.65 | 0.00 | 13964.33 | 13698.27 | 266.06 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18242** | 5 | 2018-02-18 | 1.56 | 17597.12 | 1892.05 | 1928.36 | 0.00 | 13776.71 | 13553.53 | 223.18 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18243** | 6 | 2018-02-11 | 1.57 | 15986.17 | 1924.28 | 1368.32 | 0.00 | 12693.57 | 12437.35 | 256.22 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18244** | 7 | 2018-02-04 | 1.63 | 17074.83 | 2046.96 | 1529.20 | 0.00 | 13498.67 | 13066.82 | 431.85 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18245** | 8 | 2018-01-28 | 1.71 | 13888.04 | 1191.70 | 3431.50 | 0.00 | 9264.84 | 8940.04 | 324.80 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18246** | 9 | 2018-01-21 | 1.87 | 13766.76 | 1191.92 | 2452.79 | 727.94 | 9394.11 | 9351.80 | 42.31 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18247** | 10 | 2018-01-14 | 1.93 | 16205.22 | 1527.63 | 2981.04 | 727.01 | 10969.54 | 10919.54 | 50.00 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18248** | 11 | 2018-01-07 | 1.62 | 17489.58 | 2894.77 | 2356.13 | 224.53 | 12014.15 | 11988.14 | 26.01 | 0.0 | organic | 2018 | WestTexNewMexico |

In [4]:

df.shape

Out[4]:

(18249, 14)

In [5]:

df=df.drop(["Unnamed: 0","Date"],axis=1)  
df.head(3)

Out[5]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AveragePrice** | **Total Volume** | **4046** | **4225** | **4770** | **Total Bags** | **Small Bags** | **Large Bags** | **XLarge Bags** | **type** | **year** | **region** |
| **0** | 1.33 | 64236.62 | 1036.74 | 54454.85 | 48.16 | 8696.87 | 8603.62 | 93.25 | 0.0 | conventional | 2015 | Albany |
| **1** | 1.35 | 54876.98 | 674.28 | 44638.81 | 58.33 | 9505.56 | 9408.07 | 97.49 | 0.0 | conventional | 2015 | Albany |
| **2** | 0.93 | 118220.22 | 794.70 | 109149.67 | 130.50 | 8145.35 | 8042.21 | 103.14 | 0.0 | conventional | 2015 | Albany |

In [6]:

df.describe()

Out[6]:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AveragePrice** | **Total Volume** | **4046** | **4225** | **4770** | **Total Bags** | **Small Bags** | **Large Bags** | **XLarge Bags** | **year** |
| **count** | 18249.000000 | 1.824900e+04 | 1.824900e+04 | 1.824900e+04 | 1.824900e+04 | 1.824900e+04 | 1.824900e+04 | 1.824900e+04 | 18249.000000 | 18249.000000 |
| **mean** | 1.405978 | 8.506440e+05 | 2.930084e+05 | 2.951546e+05 | 2.283974e+04 | 2.396392e+05 | 1.821947e+05 | 5.433809e+04 | 3106.426507 | 2016.147899 |
| **std** | 0.402677 | 3.453545e+06 | 1.264989e+06 | 1.204120e+06 | 1.074641e+05 | 9.862424e+05 | 7.461785e+05 | 2.439660e+05 | 17692.894652 | 0.939938 |
| **min** | 0.440000 | 8.456000e+01 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.000000 | 2015.000000 |
| **25%** | 1.100000 | 1.083858e+04 | 8.540700e+02 | 3.008780e+03 | 0.000000e+00 | 5.088640e+03 | 2.849420e+03 | 1.274700e+02 | 0.000000 | 2015.000000 |
| **50%** | 1.370000 | 1.073768e+05 | 8.645300e+03 | 2.906102e+04 | 1.849900e+02 | 3.974383e+04 | 2.636282e+04 | 2.647710e+03 | 0.000000 | 2016.000000 |
| **75%** | 1.660000 | 4.329623e+05 | 1.110202e+05 | 1.502069e+05 | 6.243420e+03 | 1.107834e+05 | 8.333767e+04 | 2.202925e+04 | 132.500000 | 2017.000000 |
| **max** | 3.250000 | 6.250565e+07 | 2.274362e+07 | 2.047057e+07 | 2.546439e+06 | 1.937313e+07 | 1.338459e+07 | 5.719097e+06 | 551693.650000 | 2018.000000 |

In [7]:

df['type'].value\_counts()

Out[7]:

conventional 9126  
organic 9123  
Name: type, dtype: int64

In [8]:

**from** **sklearn.preprocessing** **import** LabelEncoder  
var\_mod = ['type','year','region']  
le = LabelEncoder()  
**for** i **in** var\_mod:  
 df[i] = le.fit\_transform(df[i])

In [9]:

df.head(3)

Out[9]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AveragePrice** | **Total Volume** | **4046** | **4225** | **4770** | **Total Bags** | **Small Bags** | **Large Bags** | **XLarge Bags** | **type** | **year** | **region** |
| **0** | 1.33 | 64236.62 | 1036.74 | 54454.85 | 48.16 | 8696.87 | 8603.62 | 93.25 | 0.0 | 0 | 0 | 0 |
| **1** | 1.35 | 54876.98 | 674.28 | 44638.81 | 58.33 | 9505.56 | 9408.07 | 97.49 | 0.0 | 0 | 0 | 0 |
| **2** | 0.93 | 118220.22 | 794.70 | 109149.67 | 130.50 | 8145.35 | 8042.21 | 103.14 | 0.0 | 0 | 0 | 0 |

In [10]:

X = df.drop(['type'], axis = 1)  
y = df['type']

In [11]:

**from** **sklearn.model\_selection** **import** train\_test\_split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 0)

In [12]:

**from** **sklearn.preprocessing** **import** StandardScaler   
sc=StandardScaler()  
X\_train=sc.fit\_transform(X\_train)  
X\_test=sc.transform(X\_test)

D:\anaconda3\lib\site-packages\sklearn\preprocessing\data.py:645: DataConversionWarning: Data with input dtype int32, int64, float64 were all converted to float64 by StandardScaler.  
 return self.partial\_fit(X, y)  
D:\anaconda3\lib\site-packages\sklearn\base.py:464: DataConversionWarning: Data with input dtype int32, int64, float64 were all converted to float64 by StandardScaler.  
 return self.fit(X, \*\*fit\_params).transform(X)  
D:\anaconda3\lib\site-packages\ipykernel\_launcher.py:4: DataConversionWarning: Data with input dtype int32, int64, float64 were all converted to float64 by StandardScaler.  
 after removing the cwd from sys.path.

In [13]:

**from** **sklearn.svm** **import** SVC  
my\_model = SVC(kernel = 'rbf', random\_state = 0)  
result = my\_model.fit(X\_train, y\_train)

In [14]:

predictions = result.predict(X\_test)  
predictions

Out[14]:

array([1, 0, 1, ..., 1, 1, 1])

In [15]:

**from** **sklearn** **import** metrics  
print("Accuracy:",metrics.accuracy\_score(y\_test, predictions))

Accuracy: 0.9408219178082192

In [16]:

**import** **seaborn** **as** **sn**  
**from** **sklearn.metrics** **import** confusion\_matrix  
conf\_matrix =confusion\_matrix(predictions,y\_test)  
confusion\_df = pd.DataFrame(conf\_matrix, index=['Actual 0','Actual 1'], columns=['Predicted 0','Predicted 1'])  
confusion\_df  
*#matrix=sn.heatmap(confusion\_df, cmap='coolwarm', annot=True)*

Out[16]:

|  |  |  |
| --- | --- | --- |
|  | **Predicted 0** | **Predicted 1** |
| **Actual 0** | 1631 | 72 |
| **Actual 1** | 144 | 1803 |

In [17]:

new\_pred= list(result.predict([[1.33,64236.62,1036.74,54454.85,48.16,8696.87,8603.62,93.25,0.0,0,0]]))  
new\_pred

Out[17]:

[0]

In [ ]: