CLASSIFYMEISTER

Credit Score Classification using Logistic Regression KNN & SVM

Diptansu Poddar (210347)

Harshit Bhushan(210420)

Rutvik Nawade (210659)

Yuval Bansal (221232)

Objective

Using a machine learning model to classify credit scores has the goal of automating the process of determining a person's creditworthiness. The specific goal of applying machine learning to credit rating systems is to create a model that can accurately classify individuals into various credit score categories (e.g., high and low) based on their income and other relevant features. This classification helps lenders make informed decisions about whether to approve a loan application, set appropriate interest rates, or establish credit limits. The ultimate goal of using an ML model for credit scoring is to improve the efficiency and fairness of the credit evaluation process, reduce human bias, and provide consistent and reliable credit decisions to lenders and borrowers.

Methodology

We will be using multiple classification models over the dataset and comparing the results of each of them

For that, we have splitted the data into test data and training data, scaled the features included in it then fitted the model on the training data and applied it on the test data.

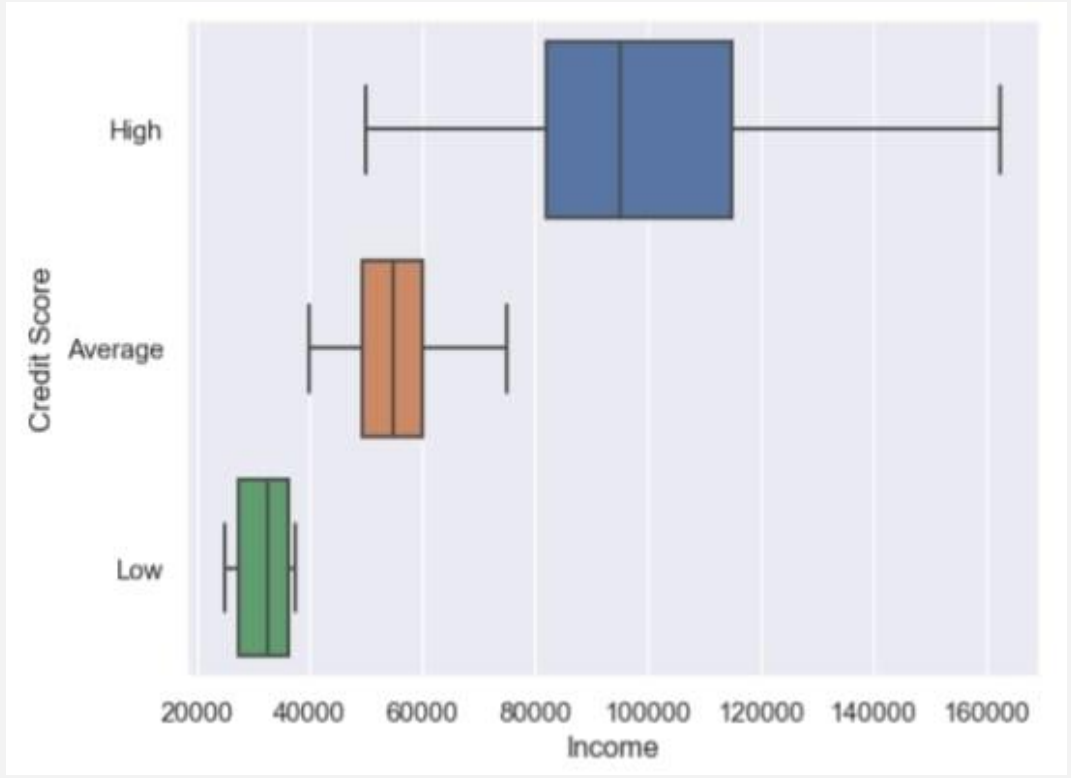
Dataset Used

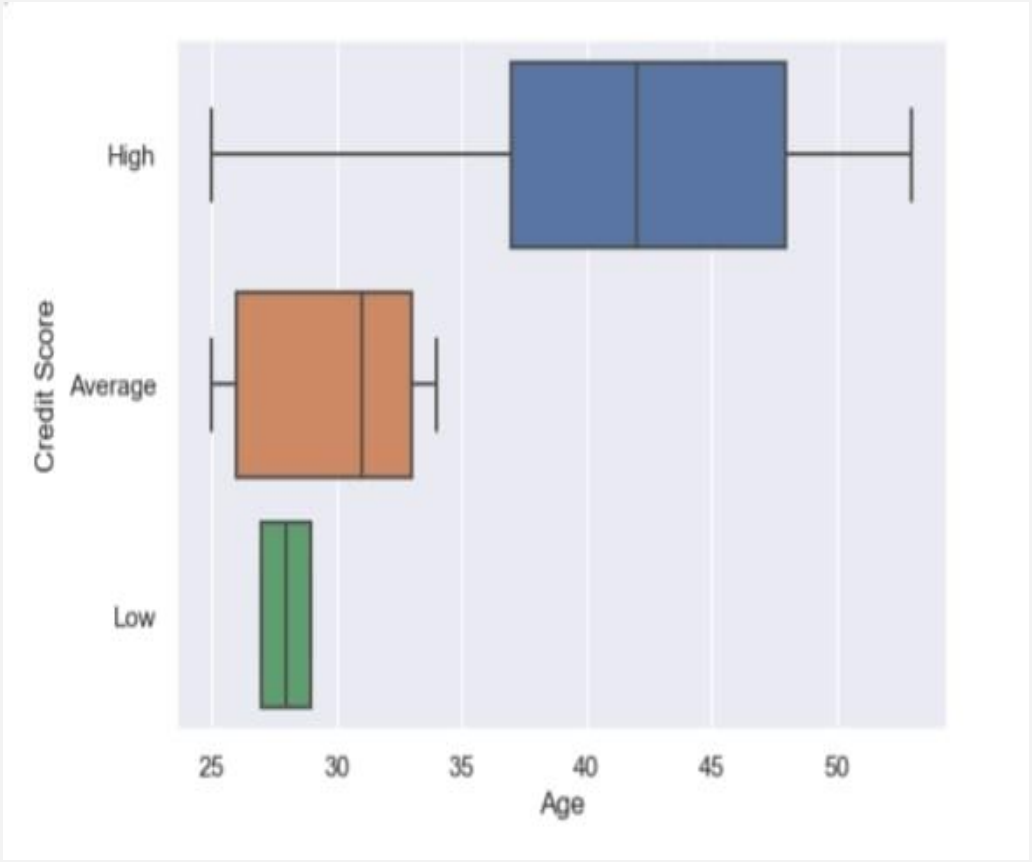
Credit Score Classification Dataset is used to make this Machine learning models.

It contains a lot of different information about people, other features and their credit scores. We will have to alter some of the data points in the dataset to make it feasible to fit on a classification model.

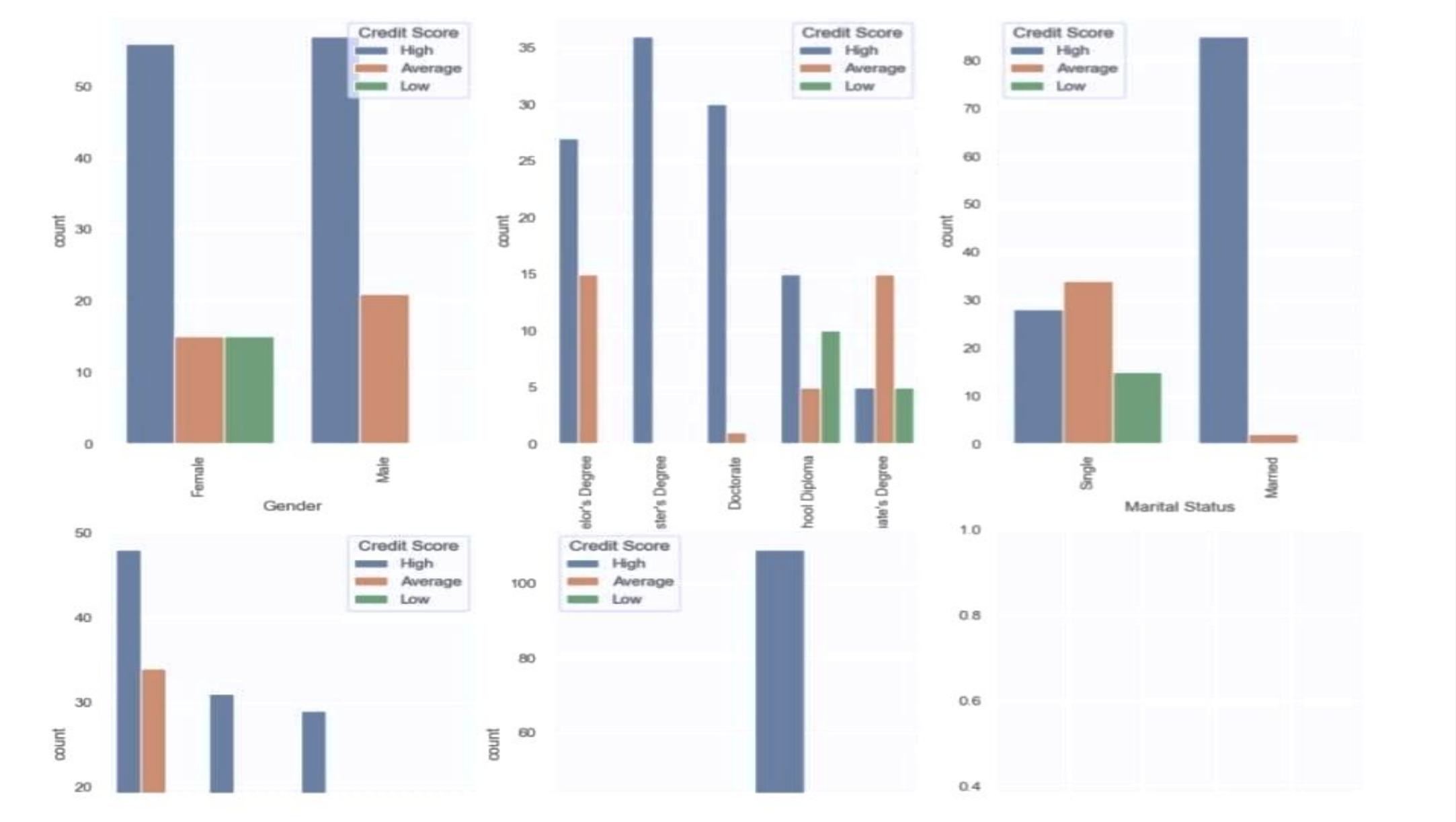
	Age	Gender	Income	Education	Marital Status	Number of Children	Home Ownership	Credit Score
0	25	Female	50000	Bachelor's Degree	Single	0	Rented	High
1	30	Male	100000	Master's Degree	Married	2	Owned	High
2	35	Female	75000	Doctorate	Married	1	Owned	High
3	40	Male	125000	High School Diploma	Single	0	Owned	High
4	45	Female	100000	Bachelor's Degree	Married	3	Owned	High

Data Visualization:





Data Visualization:



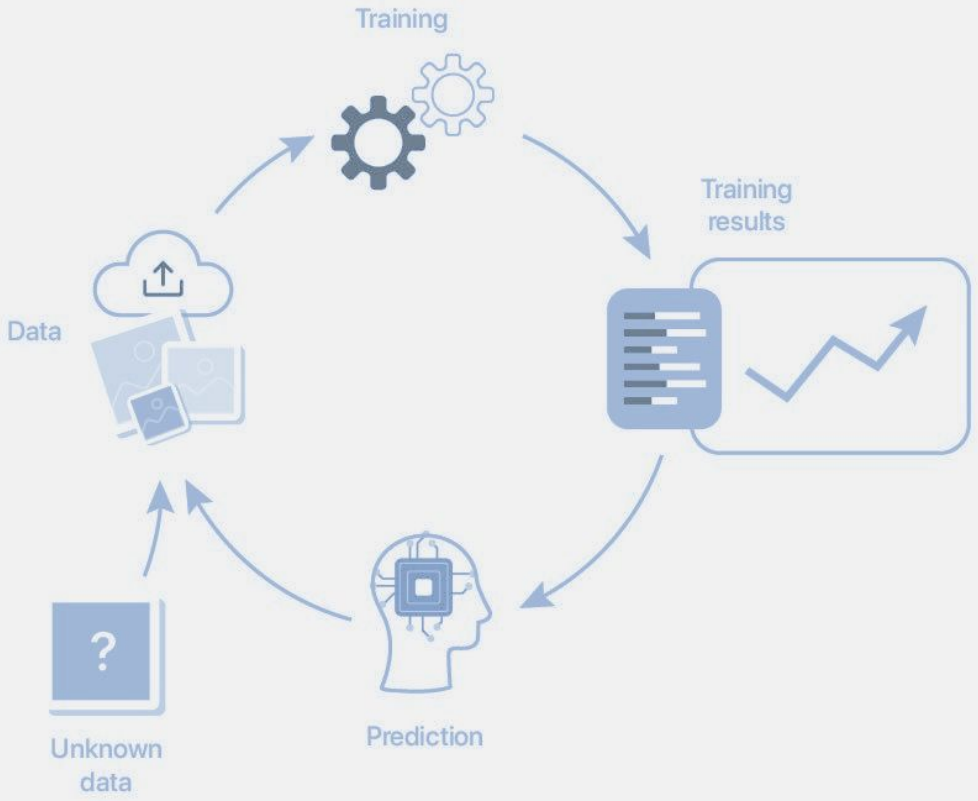
Model Training

The classification models used in the classification of the dataset are:

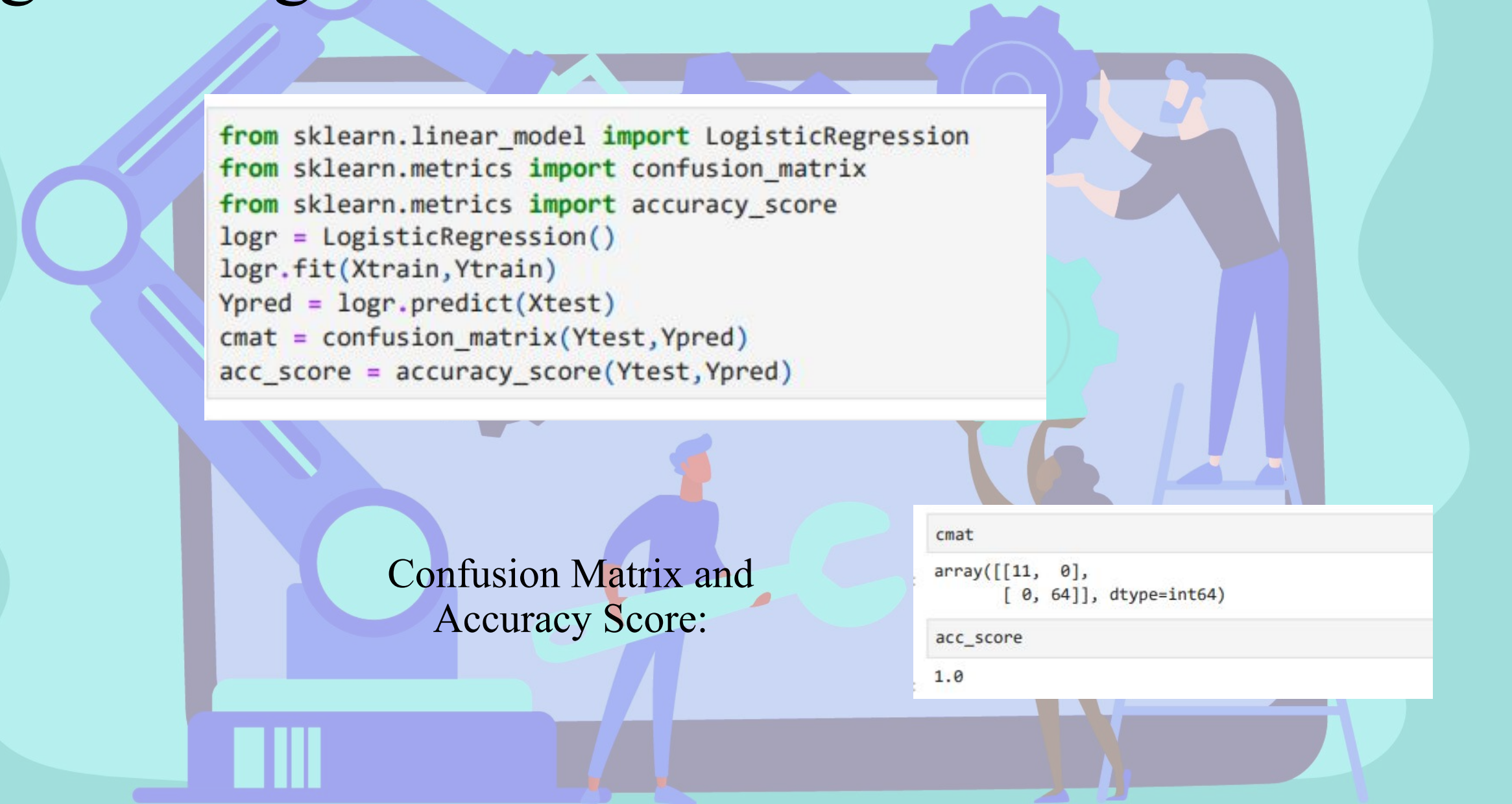
Logistic Regression: Logistic regression is a statistical model that utilizes a logistic sigmoid (S curve) function to transform input variables into a range between 0 and 1, representing the likelihood of belonging to a particular category.

K-nearest neighbors (KNN) is a machine learning algorithm that classifies new data points based on the majority vote of their K number of nearest neighbors. It measures similarity using distance metrics to assign a label.

Support Vector Machines or SVM is another algorithm that runs on the idea to find the best boundary (or hyperplane) that separates the different classes of data. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.



Logistic Regression



```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
logr = LogisticRegression()
logr.fit(Xtrain,Ytrain)
Ypred = logr.predict(Xtest)
cmat = confusion_matrix(Ytest,Ypred)
acc_score = accuracy_score(Ytest,Ypred)
```

Confusion Matrix and
Accuracy Score:

```
cmat
```

```
array([[11,  0],
       [ 0, 64]], dtype=int64)
```

```
acc_score
```

```
1.0
```

K- Nearest Neighbours

```
from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier()
knn.fit(Xtrain,Ytrain)
Ypred = knn.predict(Xtest)
cmat = confusion_matrix(Ytest,Ypred)
acc_score = accuracy_score(Ytest,Ypred)
```

Confusion Matrix and Accuracy Score:

```
cmat
```

```
array([[11,  0],
       [ 0, 64]], dtype=int64)
```

```
acc_score
```

```
1.0
```

Support Vector Machines

```
from sklearn import svm
```

```
clf = svm.SVC(kernel='linear',C=0.1)
clf.fit(X,Y)
```

```
SVC(C=0.1, kernel='linear')
```

```
Ypred = clf.predict(Xtest)
cmat = confusion_matrix(Ytest,Ypred)
acc_score = accuracy_score(Ytest,Ypred)
```

Confusion Matrix and Accuracy Score:

```
cmat
```

```
array([[11,  0],
       [ 0, 64]], dtype=int64)
```

```
acc_score
```

```
1.0
```

Results

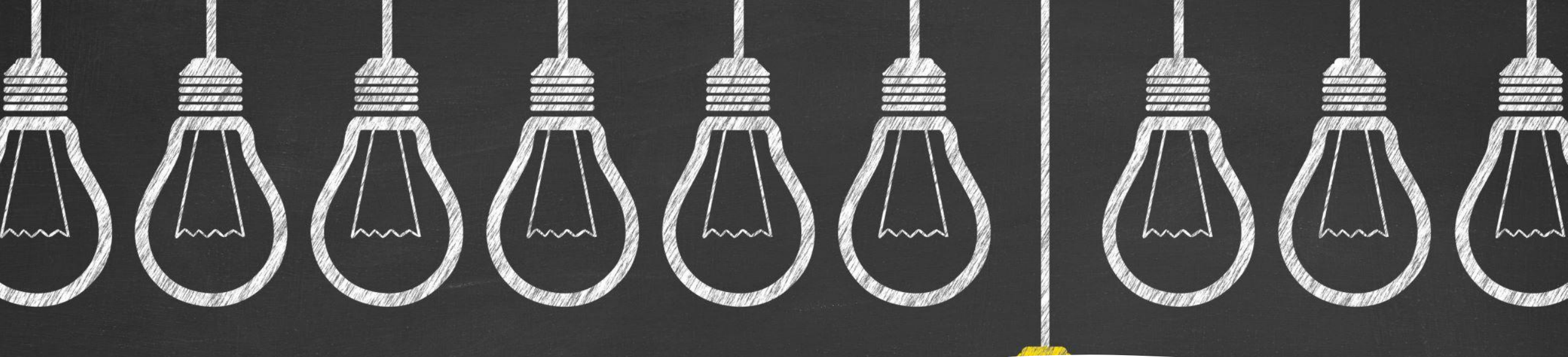
We have used the accuracy_score and confusion_matric functions to evaluate the classification models that we have used.

```
cmat
```

```
array([[11,  0],
       [ 0, 64]], dtype=int64)
```

```
acc_score
```

```
1.0
```



Conclusion

Thus, we trained the machine learning models utilizing Logistic Regression, K-nearest neighbors, and Support Vector Machines and gave the prediction for the credit scores, based on the dataset which contained features like Income, Age, etc.