In [2]:
```
pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in d:\users\neetu27\anaconda3\lib\site-packages (1.2.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in d:\users\neetu27\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: joblib>=1.1.1 in d:\users\neetu27\anaconda3\lib\site-packages (from scikit-learn) (1.1.1)
Requirement already satisfied: numpy>=1.17.3 in d:\users\neetu27\anaconda3\lib\site-packages (from scikit-learn) (1.23.5)
Requirement already satisfied: scipy>=1.3.2 in d:\users\neetu27\anaconda3\lib\site-packages (from scikit-learn) (1.10.0)
Note: you may need to restart the kernel to use updated packages.
```

In [4]:
```
import numpy as np
import pandas as pd
import random as rnd

# visualization
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

# machine learning
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC, LinearSVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import Perceptron
from sklearn.linear_model import SGDClassifier
from sklearn.tree import DecisionTreeClassifier
```

In [6]:
```
df_movie=pd.read_csv(r"C:\Users\neetu27\Downloads\archive (2)\Netflix_Dataset_Movie.csv")
df_movie.columns =['MovieIDs','MovieName','Category']
df_movie.dropna(inplace=True)
df_movie.head()
```

Out[6]:

	MovieIDs	MovieName	Category
0	1	2003	Dinosaur Planet
1	2	2004	Isle of Man TT 2004 Review
2	3	1997	Character
3	4	1994	Paula Abdul's Get Up & Dance
4	5	2004	The Rise and Fall of ECW

In [9]:
```
df_rating = pd.read_csv(r"C:\Users\neetu27\Downloads\archive (2)\Netflix_Dataset_Rating.csv")
df_rating.columns =['ID','MovieID','Ratings']
df_rating.dropna(inplace=True)
df_rating.head()
```

Out[9]:

	ID	MovieID	Ratings
0	712664	5	3
1	1331154	4	3
2	2632461	3	3
3	44937	5	3
4	656399	4	3

In [11]:
```
df_user = pd.read_csv(r"C:\Users\neetu27\Downloads\archive (2)\Netflix_Dataset_Rating.csv")
df_user.columns =['UserID','Gender','Age']
df_user.dropna(inplace=True)
df_user.head()
```

Out[11]:

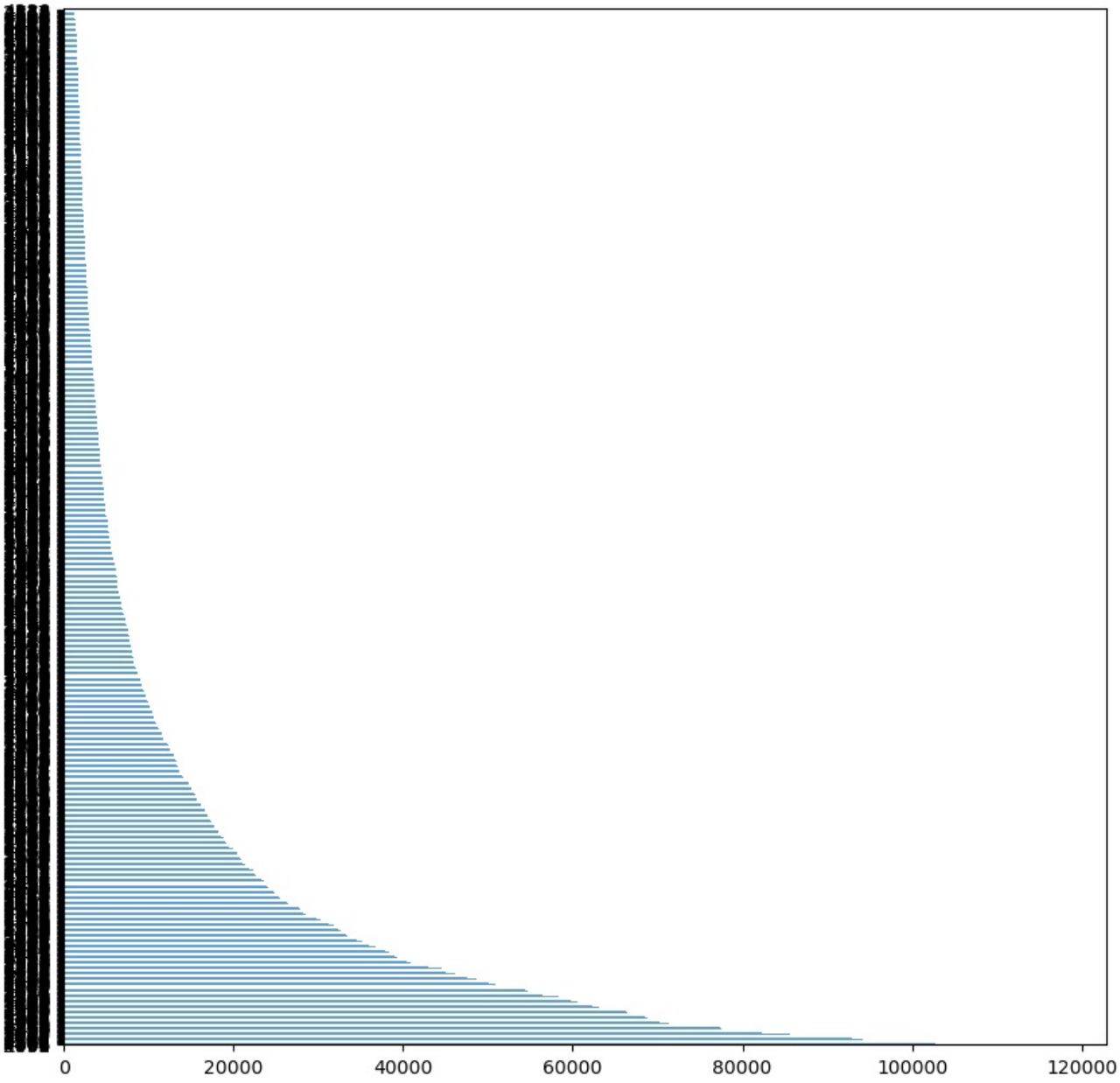
	UserID	Gender	Age
0	712664	5	3
1	1331154	4	3
2	2632461	3	3
3	44937	5	3
4	656399	4	3

In [12]:
```
df = pd.concat([df_movie, df_rating,df_user], axis=1)
df.head()
```

Out[12]:

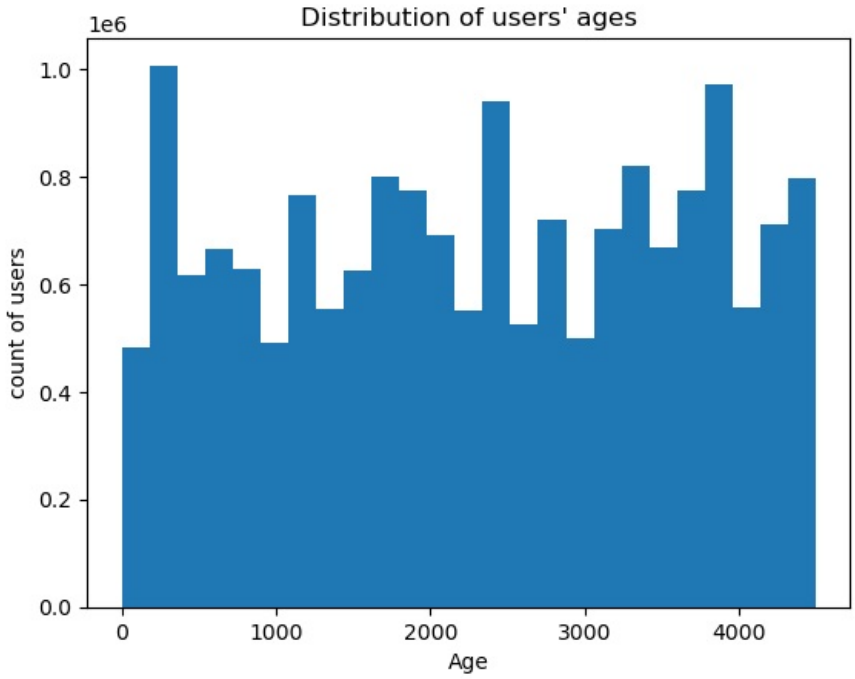
	MovieIDs	MovieName	Category	ID	MovieID	Ratings	UserID	Gender	Age
0	1.0	2003.0	Dinosaur Planet	712664	5	3	712664	5	3
1	2.0	2004.0	Isle of Man TT 2004 Review	1331154	4	3	1331154	4	3
2	3.0	1997.0	Character	2632461	3	3	2632461	3	3
3	4.0	1994.0	Paula Abdul's Get Up & Dance	44937	5	3	44937	5	3
4	5.0	2004.0	The Rise and Fall of ECW	656399	4	3	656399	4	3

In [20]:
```
df['Age'].value_counts().plot(kind='barh',alpha=0.7,figsize=(10,10))
plt.show()
```



In [21]:
```
df.Age.plot.hist(bins=25)
plt.title("Distribution of users' ages")
plt.ylabel('count of users')
plt.xlabel('Age')
```

Out[21]:
```
Text(0.5, 0, 'Age')
```

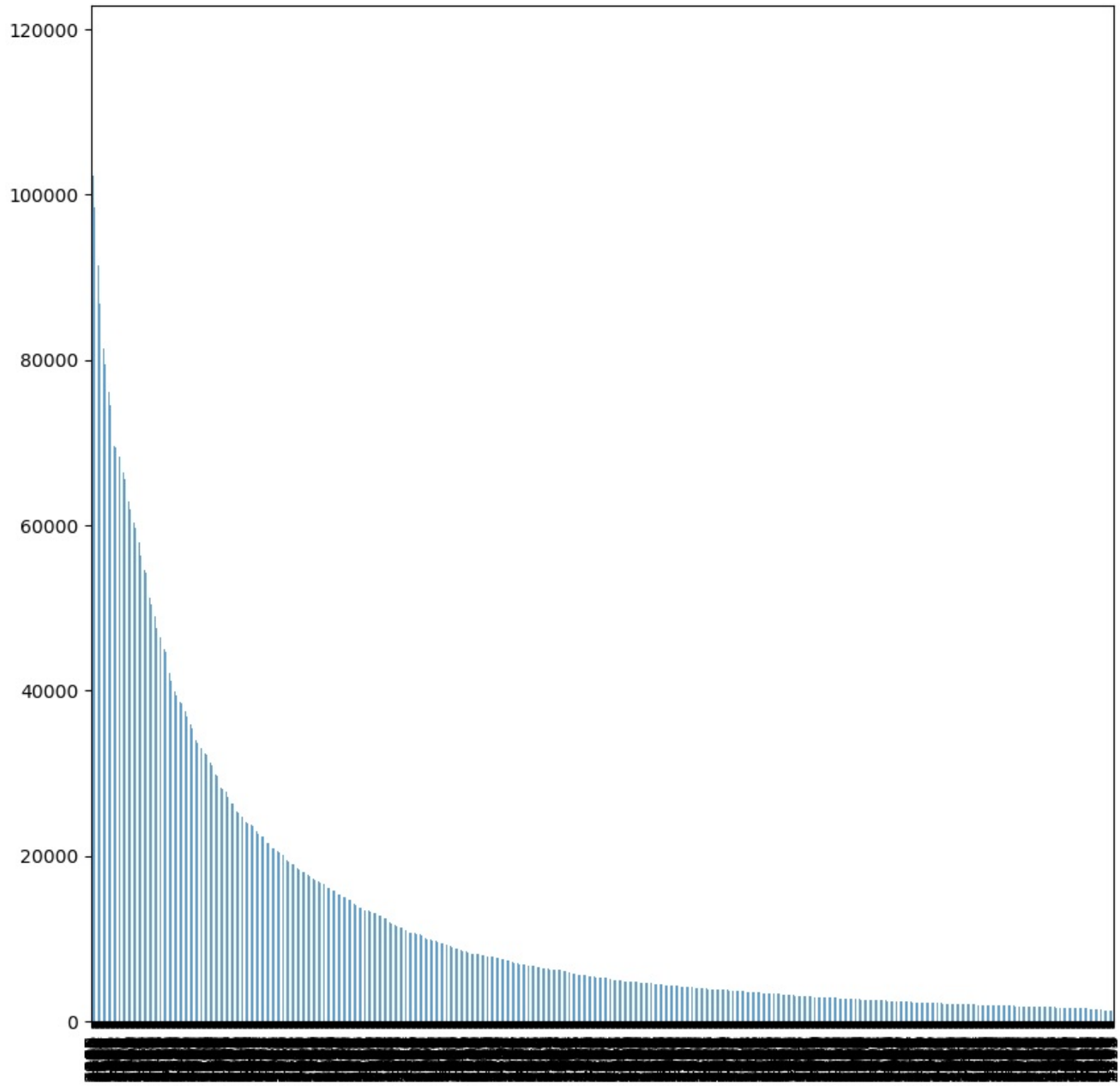


In [22]:
```
labels = ['0-9', '10-19', '20-29', '30-39', '40-49', '50-59', '60-69', '70-79']
df['age_group'] = pd.cut(df.Age, range(0, 81, 10), right=False, labels=labels)
df[['Age', 'age_group']].drop_duplicates()[:10]
```

Out[22]:

	Age	age_group
0	3	0-9
1524	8	0-9
10903	16	10-19
13420	17	10-19
17585	18	10-19
27004	26	20-29
31266	28	20-29
63663	30	30-39
141165	32	30-39
142794	33	30-39

In [23]:
```
df['Ratings'].value_counts().plot(kind='bar',alpha=0.7,figsize=(10,10))
plt.show()
```



In [34]:
```
groupedby_movieName = df.groupby('MovieName')
groupedby_rating = df.groupby('Ratings')
groupedby_uid = df.groupby('UserID')
#groupedby_age = df.loc[most_50.index].groupby(['MovieName', 'age_group'])
```

In [36]:
```
first_500 = df[500:]
first_500.dropna(inplace=True)
```

```
C:\Users\neetu27\AppData\Local\Temp\ipykernel_1244\3388945548.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ret
urning-a-view-versus-a-copy
  first_500.dropna(inplace=True)
```

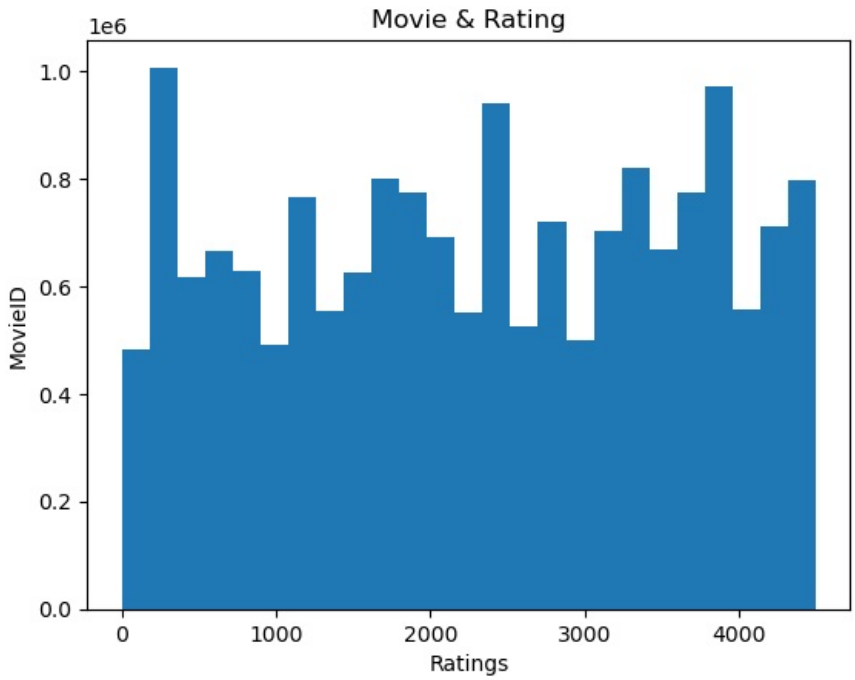
In [37]:
```
features = first_500[['MovieID','Age']].values
```

In [38]:
```
labels = first_500[['Ratings']].values
```

In [39]:
```
train, test, train_labels, test_labels = train_test_split(features,labels,test_size=0.33,random_state=42)
```

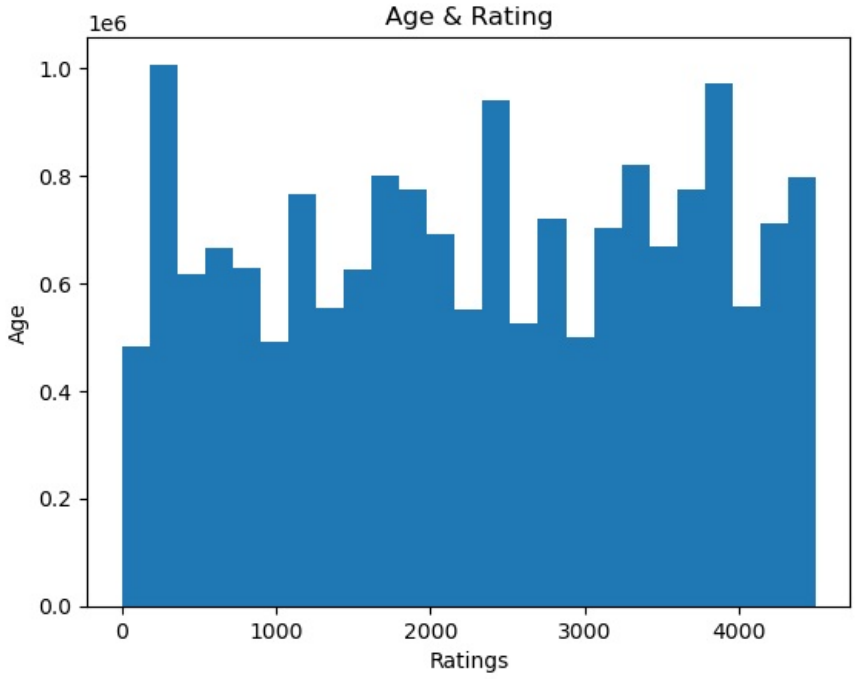
In [40]:
```
df.Age.plot.hist(bins=25)
plt.title("Movie & Rating")
plt.ylabel('MovieID')
plt.xlabel('Ratings')
```

Out[40]:
```
Text(0.5, 0, 'Ratings')
```



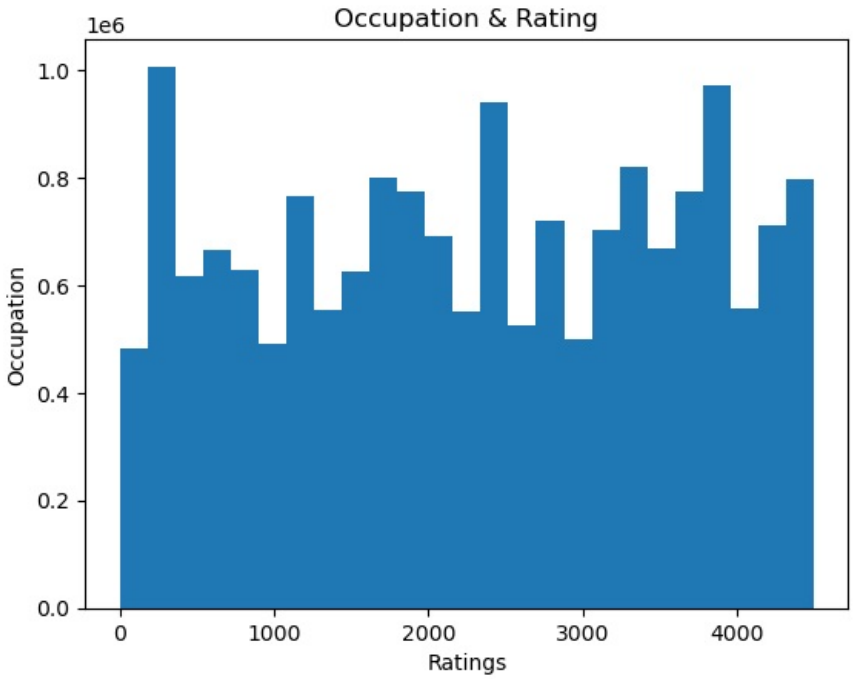
In [41]:
```
df.Age.plot.hist(bins=25)
plt.title("Age & Rating")
plt.ylabel('Age')
plt.xlabel('Ratings')
```

Out[41]:
```
Text(0.5, 0, 'Ratings')
```



In [42]:
```
df.Age.plot.hist(bins=25)
plt.title("Occupation & Rating")
plt.ylabel('Occupation')
plt.xlabel('Ratings')
```

Out[42]:
```
Text(0.5, 0, 'Ratings')
```



In [43]:
```
logreg = LogisticRegression()
logreg.fit(train, train_labels)
Y_pred = logreg.predict(test)
acc_log = round(logreg.score(train, train_labels) * 100, 2)
acc_log
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:458: ConvergenceWarning: lbfgs f
ailed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[43]:
```
98.94
```

In [44]:
```
svc = SVC()
svc.fit(train, train_labels)
Y_pred = svc.predict(test)
acc_svc = round(svc.score(train, train_labels) * 100, 2)
acc_svc
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
```

Out[44]:
```
100.0
```

In [45]:
```
knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(train, train_labels)
Y_pred = knn.predict(test)
acc_knn = round(knn.score(train, train_labels) * 100, 2)
acc_knn
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\neighbors\_classification.py:215: DataConversionWarning: A
column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for exam
ple using ravel().
  return self._fit(X, y)
```

Out[45]:
```
100.0
```

In [46]:
```
gaussian = GaussianNB()
```

```
gaussian.fit(train, train_labels)
Y_pred = gaussian.predict(test)
acc_gaussian = round(gaussian.score(train, train_labels) * 100, 2)
acc_gaussian
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
```

Out[46]:
```
100.0
```

In [47]:
```
perceptron = Perceptron()
perceptron.fit(train, train_labels)
Y_pred = perceptron.predict(test)
acc_perceptron = round(perceptron.score(train, train_labels) * 100, 2)
acc_perceptron
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
```

Out[47]:
```
55.13
```

In [48]:
```
linear_svc = LinearSVC()
linear_svc.fit(train, train_labels)
Y_pred = linear_svc.predict(test)
acc_linear_svc = round(linear_svc.score(train, train_labels) * 100, 2)
acc_linear_svc
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\svm\_base.py:1244: ConvergenceWarning: Liblinear failed to
converge, increase the number of iterations.
  warnings.warn(
```

Out[48]:
```
92.98
```

In [49]:
```
sgd = SGDClassifier()
sgd.fit(train, train_labels)
Y_pred = sgd.predict(test)
acc_sgd = round(sgd.score(train, train_labels) * 100, 2)
acc_sgd
```

```
D:\Users\neetu27\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example us
ing ravel().
  y = column_or_1d(y, warn=True)
```

Out[49]:
```
98.94
```

In [50]:
```
decision_tree = DecisionTreeClassifier()
decision_tree.fit(train, train_labels)
Y_pred = decision_tree.predict(test)
acc_decision_tree = round(decision_tree.score(train, train_labels) * 100, 2)
acc_decision_tree
```

Out[50]:
```
100.0
```

In [51]:
```
random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(train, train_labels)
Y_pred = random_forest.predict(test)
random_forest.score(train, train_labels)
acc_random_forest = round(random_forest.score(train, train_labels) * 100, 2)
acc_random_forest
```

```
C:\Users\neetu27\AppData\Local\Temp\ipykernel_1244\3903029764.py:2: DataConversionWarning: A column-vector y wa
s passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  random_forest.fit(train, train_labels)
```

Out[51]:
```
100.0
```

In [ ]: