



LOVELY
PROFESSIONAL
UNIVERSITY

INT-254 Fundamentals Of Machine Learning

C.A. I – Project

Topic – Customer Prediction Through ML

Submitted By: -

Member 1: -

Name - Aayush Deep

Registration Number - 12019059

Roll Number - RKM117B57

Member 2: -

Name – Aman Gupta

Registration Number – 12015361

Roll Number - RKM117B51

Under the guidance of: - Dr. Pawan Kumar Mall Sir

GitHub Link - <https://github.com/Aayush2609/CA-Customer-Prediction>

Acknowledgement

We would like to express our sincere appreciation to Mr. Pawan Kumar Mall Sir for his invaluable advice and recommendations, which we much appreciated as we finished our project on 'Customer Prediction through ML' to be more précised we have talked about 'Customer Behavior Prediction Through ML' in the project. His comments and guidance were extremely helpful to us at various stages of the project's completion. In this aspect, we shall be eternally grateful to you.

We also want to thank Lovely Professional University, our own university, for providing us with a project like this, which helped us learn new technologies and develop our skills.

We made sure that we didn't copy anyone else's work when we finished this job. In closing, we want to express our gratitude to all of our classmates for their invaluable assistance.

Introduction

Artificial Intelligence is the ability of a machine to learn like a human, thus achieving the level of human intelligence and much more.

With advancements in the field of AI, it has led to improvements across several industries like Automation, Supply chain, eCommerce, Manufacturing, and many more.

Understanding customer behavior: -

Customer Behavior, in its broadest sense, is concerned with how customers select, decide, use and dispose of goods and services. It covers individuals, groups, or organizations of any vertical.

It gives a good idea and insights about customer's emotions, attitudes, and preferences which affect buying behavior. Thus, helping marketers to understand the needs of customers, bringing value to the customers, and in return generating revenue for the company.

Predicting the customer behavior: -

Big companies understand that predicting customer behavior fills the gap in the markets and identifies products that are needed, and which could generate bigger revenue.

Objective: -

The objective of this project is to understand the behaviors of your customers based on their purchasing data. This helps the marketing team to understand and plan new strategies accordingly.

Understanding dataset: -

In this dataset, we have information related to customers like:

1. CustomerID - ID of the customer
2. Gender - Gender of the customer
3. Age - Age of the customer
4. AnnualIncome - annual income of the customer
5. SpendingScore - score assigned based on the customer's

behavior and their purchasing data You can download the dataset :-

<https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorialin-python>

Customer behavior prediction can be done by:

1. **Segmentation:** separating customers into smaller groups based on buying behaviors. This helps in the separation of concerns, which in turn helps us identify the region of the market.
2. **Predictive Analytics:** we use statistical techniques to analyze previous historical data to predict the future behavior of customers.

Customer Prediction Through Machine Learning

Importing Libraries:-

For data exploration, it is mandatory to have a few Python libraries installed

```
In [6]: import numpy as np
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing KMeans algorithm to be used for data analytics

```
In [37]: from sklearn.cluster import KMeans
```

Importing warnings library to clean my further project by ignoring the warnings

```
In [38]: import warnings
warnings.filterwarnings("ignore")
```

View Dataset:-

To view the dataset we have to import by reading the CSV file as shown below

```
In [39]: df = pd.read_csv('Mall_Customers.csv')
df.head()
```

```
Out[39]:
```

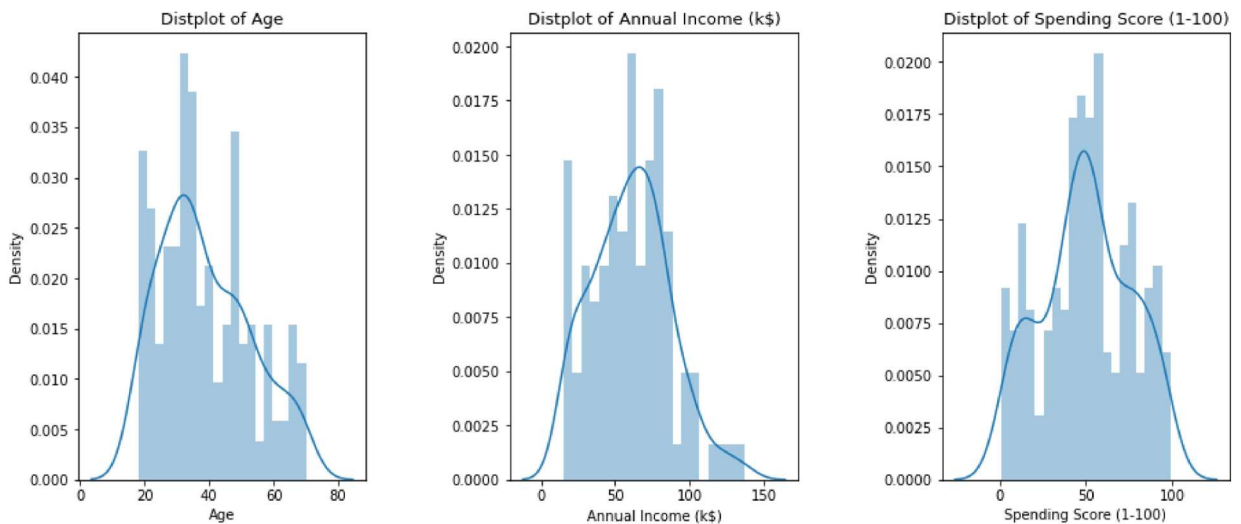
	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Data Visualizations:-

Correlation between Age, Income and Spending scores

Analyzing the spending patterns is the better strategy for marketing. Here, we will try to analyze and find how the age, annual incomes and spending scores of the customers are.

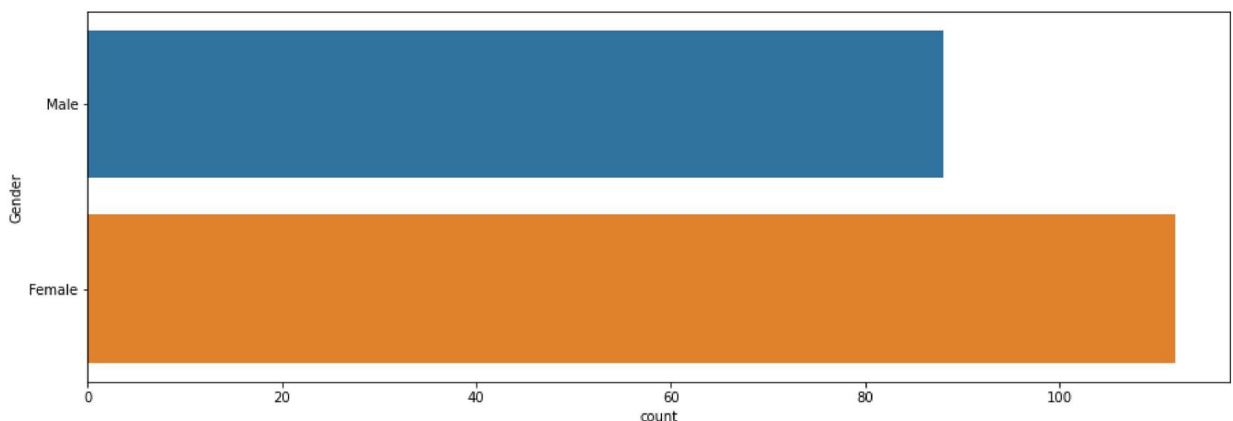
```
In [40]: plt.figure(1, figsize = (15, 6)) # sets the dimensions of image
n = 0
for x in ['Age', 'Annual Income (k$)', 'Spending Score (1-100)']:
    n += 1
    plt.subplot(1, 3, n) # creates 3 different sub-plots
    plt.subplots_adjust(hspace = 0.5, wspace = 0.5)
    sns.distplot(df[x], bins = 20) # creates a distribution plot
    plt.title('Distplot of {}'.format(x)) # sets title for each plot
plt.show() # displays all the plots
```



Gender Analysis:-

The second most important thing in deciding the strategy is to analyze the spending patterns based on Gender. Here, we find that Females tend to purchase more than Males do.

```
In [41]: plt.figure(1, figsize = (15, 5))
sns.countplot(y = 'Gender', data = df)
plt.show()
```



Customer segmentation:-

Segmentation helps in dividing a set of large data into groups of smaller observations that are similar in specific ways relevant to marketing. Each group contains individuals that are similar in-between themselves, and different from individuals from the other groups.

Segmentation is widely used as a marketing tool to create a clusters of clients and adapt a relevant strategy for each of them.

Segmentation using Age and Spending Score:-

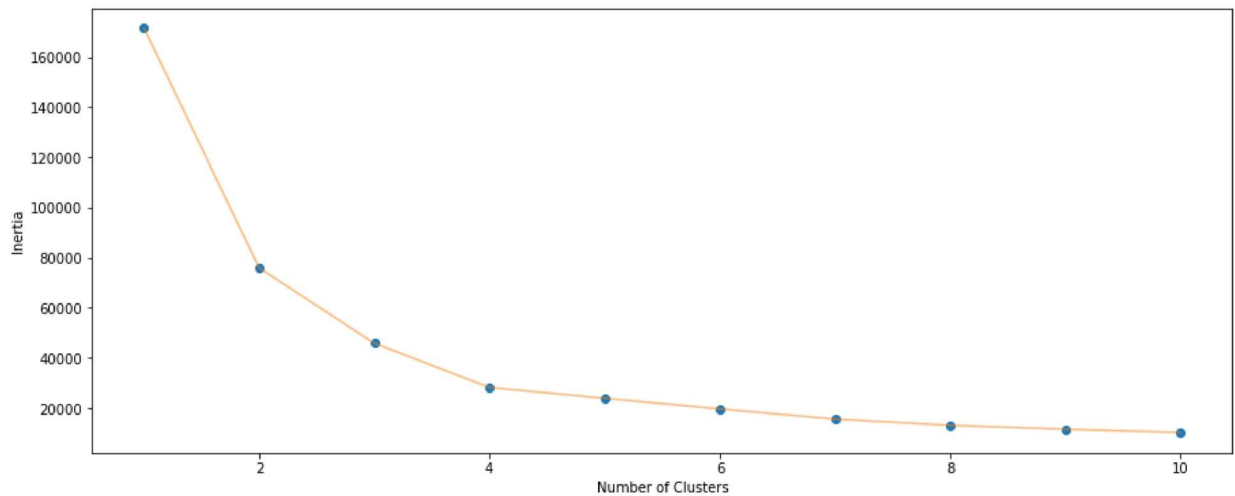
Here we will be segmenting the customers based on their age and their spending scores. This will help us to understand the age category of the customers, which could possibly improve spending score, thereby increasing the revenue for the company.

In this case, we have to decide the possible number of clusters(segments) that would return the best results. To do that, we loop through 1 to 11, and find which customer would be the right choice.

```
In [42]: X_age_spending = df[['Age' , 'Spending Score (1-100)']].iloc[:, :].values # extracts
inertia = []
for n in range(1 , 11):
    model_1 = (KMeans(n_clusters = n , init='k-means++', n_init = 10 , max_iter=300,
                      tol=0.0001, random_state= 111 , algorithm='elkan')) # use pr
    model_1.fit(X_age_spending) # fit the data into the model
    inertia.append(model_1.inertia_)
```

After using segmentation on the dataset, we need to visualize this by plotting the graph:-

```
In [43]: plt.figure(1 , figsize = (15 ,6)) # set dimension of image
plt.plot(np.arange(1 , 11) , inertia , 'o') # Mark the points with a solid circle
plt.plot(np.arange(1 , 11) , inertia , '-' , alpha = 0.5) # connect remaining points v
plt.xlabel('Number of Clusters') , plt.ylabel('Inertia') # Label the x and y axes
plt.show() # display
```

After cluster 4 , the line graph starts becoming stable. This is known as Elbow Method.

```
In [34]: model_2 = (KMeans(n_clusters = 4 ,init='k-means++', n_init = 10 ,max_iter=300,
                           tol=0.0001, random_state= 111 , algorithm='elkan') ) # set r
model_2.fit(X_age_spending) # fit the model
labels1 = model_2.labels_
centroids1 = model_2.cluster_centers_
```

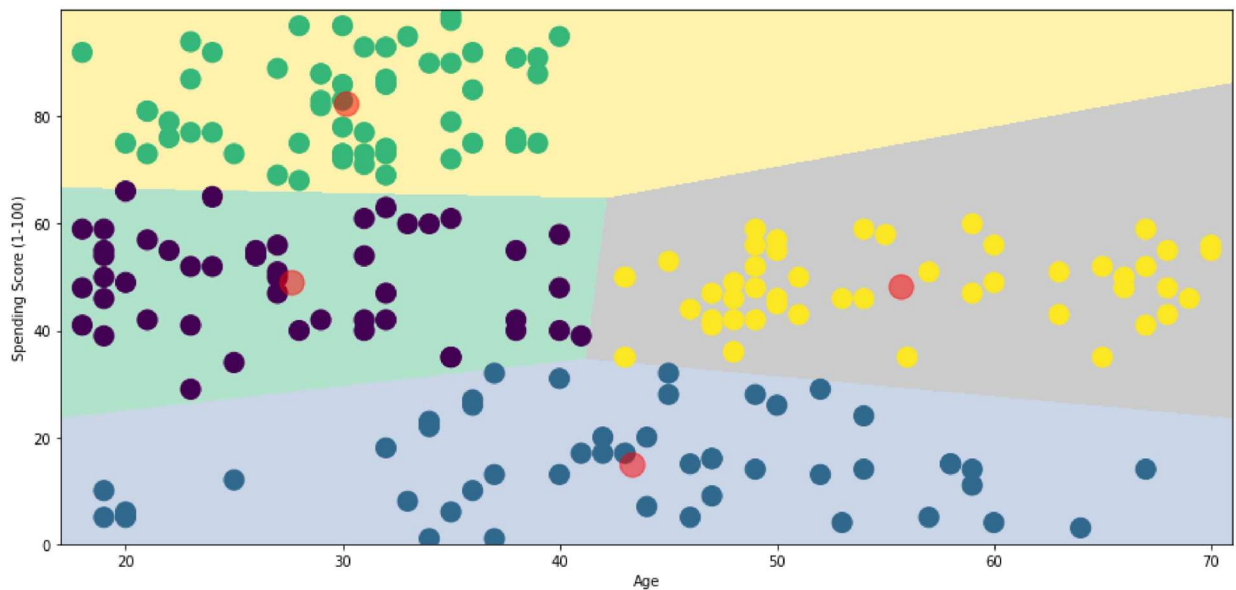
Some prerequisites for plotting a graph - max and min ranges of values, initialising a meshgrid(), and so on:-

```
In [35]: h = 0.02
x_min, x_max = X_age_spending[:, 0].min() - 1, X_age_spending[:, 0].max() + 1
y_min, y_max = X_age_spending[:, 1].min() - 1, X_age_spending[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, h), np.arange(y_min, y_max, h))
Z = model_2.predict(np.c_[xx.ravel(), yy.ravel()]) # returns flattened 1D array
```

Plotting the graph:-

```
In [46]: plt.figure(1 , figsize = (15 , 7) )
plt.clf()
Z = Z.reshape(xx.shape)
plt.imshow(Z , interpolation='nearest',
           extent=(xx.min(), xx.max(), yy.min(), yy.max()),
           cmap = plt.cm.Pastel2, aspect = 'auto', origin='lower')

plt.scatter( x = 'Age' , y = 'Spending Score (1-100)' , data = df , c = labels1 ,
            s = 200 )
plt.scatter(x = centroids1[:, 0] , y = centroids1[:, 1] , s = 300 , c = 'red' , alr
plt.ylabel('Spending Score (1-100)' ) , plt.xlabel('Age')
plt.show()
```

K-Means with 4 clusters

(Graph includes Four clusters with the Red Points referring to the centroids of the respective clusters)

So now we can deduce the following informatin about the spending patterns:

1. The average spending score irrespective of age would be around 20.
2. In the topmost cluster, Customers below age 40 has the highest spending scores. The cluster is less sparse.
3. Above age 40, the spending score remains consistently within the range of 30 - 60.

Conclusion

Data is an essential part of any enterprise or business. It is critical to collect, process, and analyze data flow in a timely and accurate manner to discover information useful for decision making in business.

The data science industry is rapidly expanding. Data volume can be large, making information handling difficult and time consuming.

As we learned from the above simple case-study, we find that AI has played a significant role in almost all the industries. With the rise in the trend of data analysis, the customers behavior is being continuously monitored for improving strategies and taking better decisions.