CUSTOMER SEGMENTION

INNOVATION

**INTRODUCTION**

Customer segmentation is the practice of dividing a company's customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.

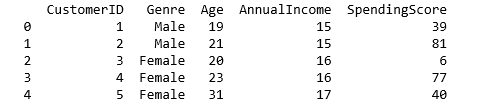
**DATASET**

**The dataset given for customer segmention is implemented to**

**The program by**



**And we got the Mall\_customer data**

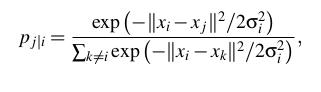


**T-SNE ALGORITHM**

(T-SNE)-Distributed Stochastic Neighbor Embedding is a non-linear dimensionality reduction algorithm used for exploring high-dimensional data. It maps multi-dimensional data to two or more dimensions suitable for human observation. With help of the t-SNE algorithms, you may have to plot fewer exploratory data analysis plots next time you work with high dimensional data.

Stochastic Neighbor Embedding (SNE) starts by converting the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities. The similarity of datapoint to datapoint is the conditional probability, , would pick as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian centered at.

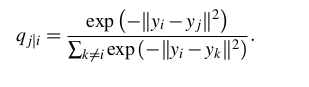
For nearby datapoints, is relatively high, whereas for widely separated datapoints, will be almost infinitesimal (for reasonable values of the variance of the Gaussian,). Mathematically, the conditional probability is given by



where is the variance of the Gaussian that is centered on datapoint

If you are not interested in the math, think about it in this way, the algorithm starts by converting the shortest distance (a straight line) between the points into probability of similarity of points. Where, the similarity between points is: the conditional probability that would pick  as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian (normal distribution) centered at.

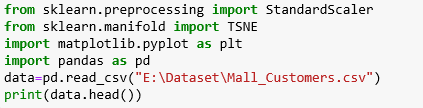
the low-dimensional counterparts and of the high-dimensional datapoints and  it is possible to compute a similar conditional probability, which we denote by .



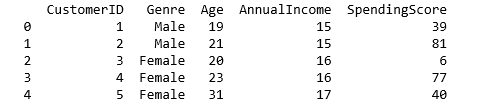
Note that, pi|i and pj|j are set to zero as we only want to model pair wise similarity.

**PROGRAM**

**Mall\_Customers data set is taken and let implement the libraries**

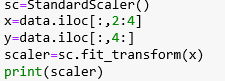


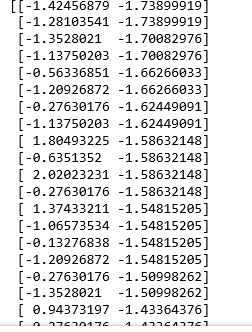
And the data are



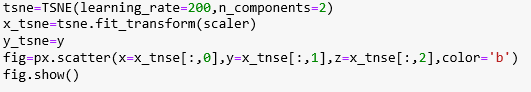
In this project we are going to change the dimension of the data set using T-SNE

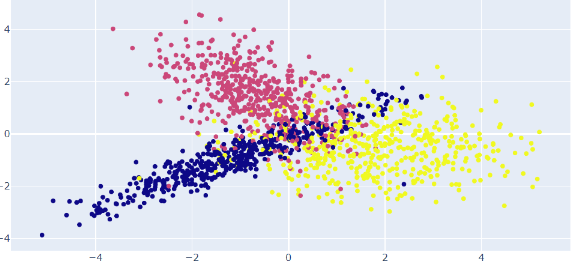
Algorithm . let combine the age and AnnualIncome data to predict the Spending score for customers’s product sales.





Then





CONCLUSION

From the data set we can predict the customer segmentation ,the goal is to better understand and target different groups of customers with tailored marketing strategies, product offerings and services.