**Unsupervised Machine Learning**

**(Clustering)**

By

Dhanraj A Tiwari

Data Science, Almabetter

**Abstract**

Data mining can efficiently deal with the

large amount of historical and current data,

from the database can find some potential,

useful and valuable information for the retail

stores. The paper takes a large retail data as its

study object, uses data mining methods

to retail enterprise customer segments,

and then use association rules to different

groups of customers and get rules about

customer characteristics to make customer

characteristic analysis. Finally, give some

references to the supermarket's marketing

and management work.

**Introduction**

In this project I am taking customer segments

on a transnational data set which contains

all the transactions occurring between

01/12/2010 and 09/12/2011 for

a UK-based and registered store online

retail. So by using the rfm technique we

are checking how much recent and

frequent the product has been purchased

by a customer. And then using KMeans clustering

technique to take our overall data in one

cluster.

**Keywords: Machine Learning, KMeans Clustering**

**The Approach Used to Solve the Problem**

**●** To get meaningful insights from this dataset, an approach of exploratory data analysis helped a lot.

.

● The dataset was imported and converted into a Pandas dataframe. After that doing an Exploratory Data Analysis.

● To get a correct trend we used KMeans Clustering technique

**Libraries used for analysis**

1. Pandas : To load the data into a dataframe object and analyse.

2. Matplotlib : To help visualise the data.

3. Seaborn : For added functionality to matplotlib.

4. Numpy : To use the numpy functions in analysis.

5. Sklearn: To do any machine learning model Scikit learn is very useful to make the correct trend.

**Data preparation before analysis**

We have online retail dataset which consist

Of 541909 rows and 8 columns. And with that

all features we have to predict our model by taking

Insights from that data.

**Conclusion**

We calculated the rfm of our given dataset which is nothing but recency frequency and monetary. And after calculating all this we went for KMeans clustering.