ANALYSING LEAD TIMES USING KMEANS TO IDENTIFY THE RE-ORDER PERIOD

Dr.M.RAJESWARI[1]
Assistant Professor, Department of B.Com (Business Analytics)
PSGR Krishnammal College for Women
Coimbatore, India
rajeshwarim@psgrkc.ac.in

C. SOUNDARYA[2]
UG Scholar, Department of B.Com (Business Analytics)
PSGR Krishnammal College for Women
Coimbatore, India
sound40670@gmail.com

ABSTRACT

Lead time refers to the time taken to complete a process. In this study the lead time of production that is the time taken to complete the production process of various products is analyzed. *KMeans clustering* algorithm is an unsupervised machine learning algorithm used to form groups of data points known as, clusters. The clusters are formed based on similarities in mean values and number of clusters defined by the end user. The dataset used in this study is about logistics. The main aim of this study is to identify the possible levels of *re-order periods* based on the lead times of production of various products. The KMeans clustering algorithm is used to analyze the lead times of production and is further processed using *if-else* statements to identify the levels of re-order period.

Key words: KMeans, clustering, lead times, re-order period, cluster

I. INTRODUCTION

Machine learning is a part of artificial intelligence. The algorithms that are developed automatically through experience and use of data are the key concept of machine learning. Earlier Machine learning was used only by large companies and institutions which had technical experts. But recently, Machine learning technologies have developed more such that anyone can use it on their data and achieve required results. This has increased the competition in business world. The two main areas of Machine learning are: Supervised Machine learning and Unsupervised Machine learning. Clustering is a type of Machine learning technique where data points are grouped together. These groups are called as clusters. Clustering is a most common data analysis technique which is used to identify the structure of the data. Each cluster differ from each other, but the data points inside a single cluster are similar to each other. The clusters are formed on the basis of similar mean values of data points and the number of clusters defined by the end user. KMeans is an unsupervised machine learning algorithm. It is an iterative algorithm that creates 'k' number of clusters using 'n' number of data points. Each data point belongs to only one cluster. K-means is a centroid-based algorithm, or a distance-based algorithm, where we calculate the distances to assign a point to a cluster. In K-Means, each cluster is associated with a centroid. The main objective of the K-Means algorithm is to minimize the sum of distances between the points and their respective cluster centroid.[7]

II. OBJECTIVE

Re-order period is the automated time or number of days when the purchase order for the product is to be placed regularly. This can be identified based on life of the product, usage of the product, sales and demand for the product, lead times of production of the product, etc. The lead times of production process is used in this study to identify the various levels of re-order periods for the products. Lead time is the time taken to complete a process. The lead times is calculated by taking the difference between production start date and production completion date. This gives the number of days taken to complete the production process of the product. This study helps to reduce cost and time taken to re-order and re-stock the products.

III. RELATED WORK

Manufacturing lead times are a critical measure of manufacturing performance that have not received a great deal of attention in the literature. Lead times are affected by many factors including capacity, loading, batching and scheduling, and themselves affect many aspects of costs, and control. [9]

One type of demand is manufacturing lead-time-sensitive, and higher prices prevail for short lead times; the other type of demand is lead-time-insensitive. The objective is to maximize the expected profit. The form of the optimal policies in terms of manufacturing lead times and customer order-acceptance rates is characterized for the general model developed. It is shown that the optimal policy has the effect of both providing short manufacturing lead times and increasing system utilization rates, while the expected profit is maximized.[10]

In many production environments where demand and lead times are variable, significant levels of safety stock inventory are required to assure timely production and delivery of the final product. Traditional models to determine the appropriate safety stock level may result in

more safety stocks at sub-assembly and finished goods levels than necessary and thus lead to higher inventory carrying costs than desired. Such models generally incorrectly assume that the demand during the lead time follows a normal distribution.[1]

It can be difficult to set the reorder point in an inventory system because often one does not have much knowledge of the lead-time demand (LTD)distribution. A frequent practice is to assume a "standard" distribution, such as the Normal. The reorder point is then taken as the p-th fractile of that standard distribution, where (1 −p) is the specified probability of stockout during a replenishment cycle. When the desired service level p is high and the "true" LTD distribution is skewed, previous research has shown that the reorder point and inventory costs are strongly affected by the shape of the assumed LTD distribution. Ideally, no assumptions about this distribution should be necessary. [4]

Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document. Computational notebooks have been around for decades, but Jupyter in particular has exploded in popularity over the past couple of years. This rapid uptake has been aided by an enthusiastic community of user-developers and a redesigned architecture that allows the notebook to speak dozens of programming languages -- a fact reflected in its name, which was inspired, according to co-founder Fernando Pérez, by the programming languages Julia (Ju), Python (Py) and R.[5]

Artificial intelligence is seen as one of the major enablers for Smart Logistics and Smart Production initiatives. [6] For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. [3]

IV. METHODOLOGY

KMEANS ALGORITGHM

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.[2]

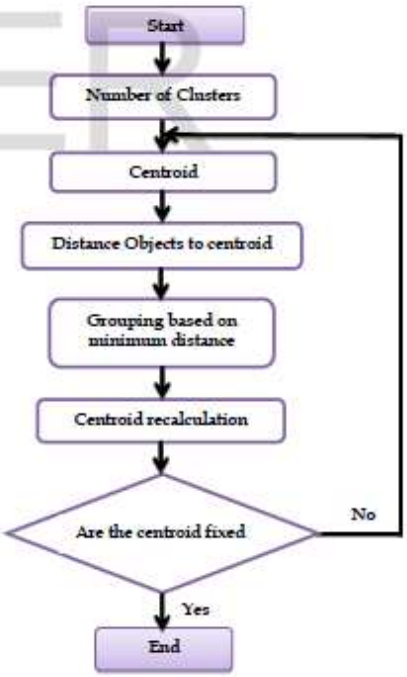

fig1: kmeans workflow[8]

The Fig1 shows the workflow of KMeans clustering algorithm. It shows the process of how the clusters are formed. First number of clusters are defined by the user. Then the centroids are calculated for each cluster. Then the distance between data points and centroids are calculated. Then the data points with minimum distance are grouped together. After grouping the centroids are recalculated. Then it is checked if the centroid values are fixed or not. If yes then clustering process is over, if no then again from the calculation of centroid value all the steps are repeated.

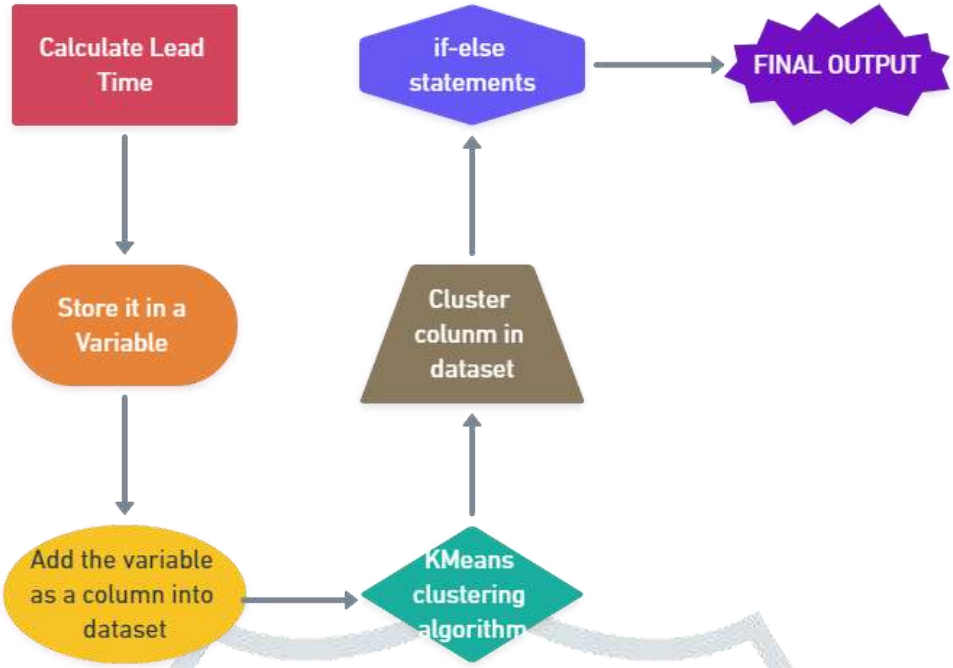

fig 2: workflow of study

The Fig 2 shows the workflow of the study, after loading data, from calculation of lead time up to the final result of the study.

V. RESULT

LEAD TIME:

```
for ind, row in data.iterrows() :
    data.loc[ind,"l"] = row['c'] - row['s']
import numpy as np
L=(data.l / np.timedelta64(1, 'D')).astype(int)
data['lead']=L
data
```

where,

data is the dataset loaded variable
l is the variable storing lead time in days format
c is the production completed date
s is the production start date
l is the variable storing lead time in integer format
lead is the column created in the dataset

Calculate lead times by taking the difference of production start date and production completed date. The resultant value will be in days format. So convert it into integer so that it can be used in KMeans clustering algorithm. And store it in a variable. Then add that as a column into the dataset.

LINE GRAPH

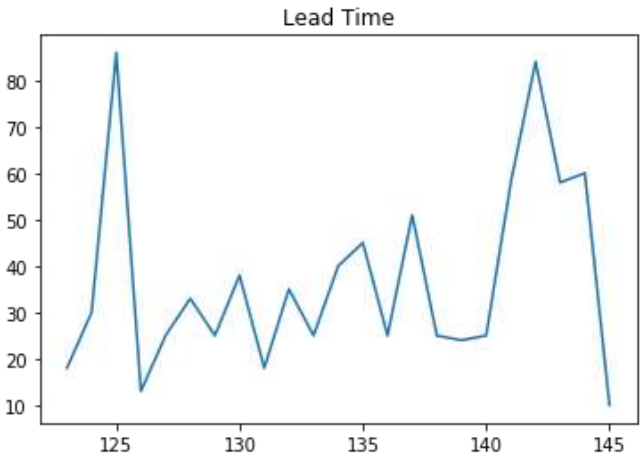

fig 3: line graph of lead times

The Fig3 shows the line graph of lead times calculated. The x axis is product id and the y axis is the lead times. From this we can identify the range of lead times for various products.

ELBOW METHOD GRAPH

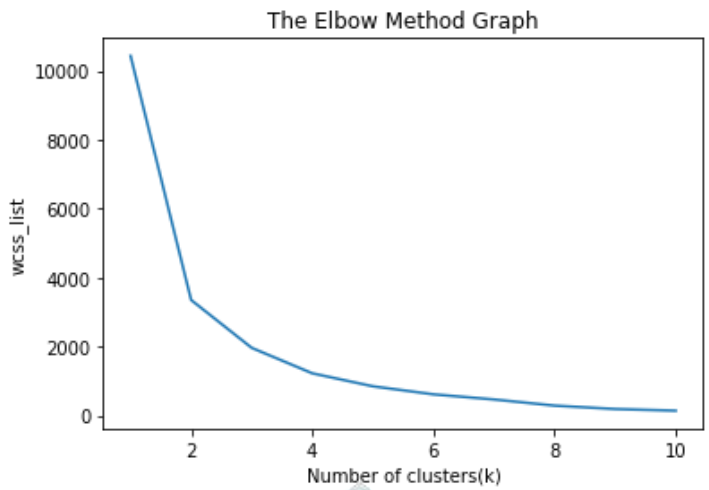

fig4: the elbow method graph

The Fig4 shows the elbow method graph on the basis of mean square values calculated for clustering process. This graph suggests the number of clusters which can be formed using the given set of attributes. The above figure shows bends in 2,3 and 4 which means a minimum of 2 clusters can be formed and a maximum of 4 clusters can be formed.

SCATTER PLOT

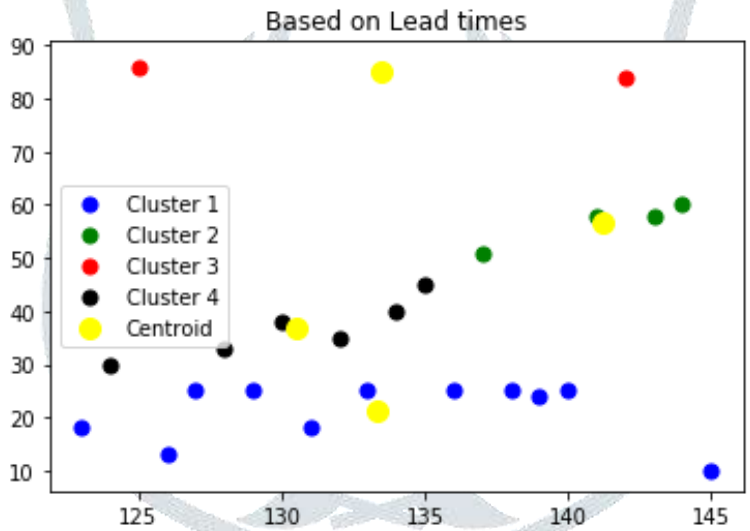

fig 5: scatter plot of clusters

The Fig5 shows the scatter plot of various clusters formed using KMeans clustering algorithm with their respective centroid points. Different clusters are shown in different colours like blue, green, red and black. And their centroids are shown in yellow colour.

CLUSTER COLUMN

Out[25]:

	L	cluster
0	18	0
1	30	3
2	86	2
3	13	0
4	25	0
5	33	3
6	25	0
7	38	3
8	18	0
9	35	3
10	25	0

fig5: cluster column

The Fig5 shows the lead times and cluster columns of the dataset. This helps to identify the range of lead times in each cluster.

RE-ORDER PERIOD

Lead_time	clusterit	lres
18	0	Re-order period can be longer
30	3	Re-order period can be medium
86	2	Re-order period should be shorter
13	0	Re-order period can be longer
25	0	Re-order period can be longer
33	3	Re-order period can be medium
25	0	Re-order period can be longer
38	3	Re-order period can be medium
18	0	Re-order period can be longer
35	3	Re-order period can be medium
25	0	Re-order period can be longer
40	3	Re-order period can be medium
45	3	Re-order period can be medium
25	0	Re-order period can be longer
51	1	Re-order period should be shorter
25	0	Re-order period can be longer
24	0	Re-order period can be longer
25	0	Re-order period can be longer
58	1	Re-order period should be shorter
84	2	Re-order period should be shorter
58	1	Re-order period should be shorter

fig 6 : re-order period

The Fig 6 shows the final result of lead times, cluster values and reorder period. The levels of re-order period are determined using if-else statements based on cluster values.

VI. CONCLUSION AND FURTHER WORK

From this study the levels of re-order period are identified by analyzing the lead times of production of products. When the lead time is higher, the re-order period should be shorter; when the lead time is lower, the re-order period can be longer and when the lead time is in medium value, the re-order period can also be a medium value. The determination of levels of re-order period on the basis of lead times of production helps in reduction of cost incurred on ordering, saves time for re-stocking the products, helps in matching the demand and supply of products and helps in effective inventory control and optimization.

Further this work can be elaborated by finding the exact number of days for re-order based on main factors affecting the re-order period like sales of product, demand for the product, lead times of production, etc. For proceeding with this some more extra data related to the factors are necessary so that the final result will be accurate and not approximate.

VII. REFERENCE

[1] Alex J. Ruiz-Torres &Farzad Mahmoodi, “Safety stock determination based on parametric lead time and demand information” International Journal of Production Research Volume48 2010 issue 10, Pages 2841-2857

[2] Imad Dabbura, “K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks”, www.towardsdatascience.com

[3] Jake VanderPlas, Python data science handbook: Essential tools for working with data, O'Reilly Media, Inc., 2016

[4] JAMES H. BOOKBINDER &ANNE E. LORDAHL, “Estimation of Inventory Re-Order Levels Using the Bootstrap Statistical Procedure” IIE Transactions Volume 21, 1989 – Issue 4

[5] Jeffrey M. Perkel, “Why Jupyter is data scientists' computational notebook of choice”, ,Nature Publishing Group(Vol. 563, Issue 7732), Nov2018.

[6]Manuel Woschank, Erwin Rauch, Helmut Zsifkovits, A Review of Further Directions for Artificial Intelligence, Machine Learning, and Deep Learning in Smart Logistics, 2020

[7] PULKIT SHARMA, “The Most Comprehensive Guide to K-Means Clustering You’ll Ever Need” ,AUGUST 19, 2019, www.analyticsvidhya.com

[8] Sreedevi Kadiyala, Chandra Srinivas Potluri, “Analyzing the Student’s Academic Performance by using Clustering Methods in Data Mining”, International Journal of Scientific & Engineering Research, ISSN 2229-5518, Volume 5, Issue 6, June-2014.

[9] Uday S.Karmarkar “Chapter 6 Manufacturing lead times, order release and capacity loading” Handbooks in Operations Research and Management Science Volume 4, 1993, Pages 287-329

[10] Z.KevinWeng, “Manufacturing lead times, system utilization rates and lead-time-related” European Journal of Operational Research Volume 89, Issue 2, 8 March 1996, Pages 259-268