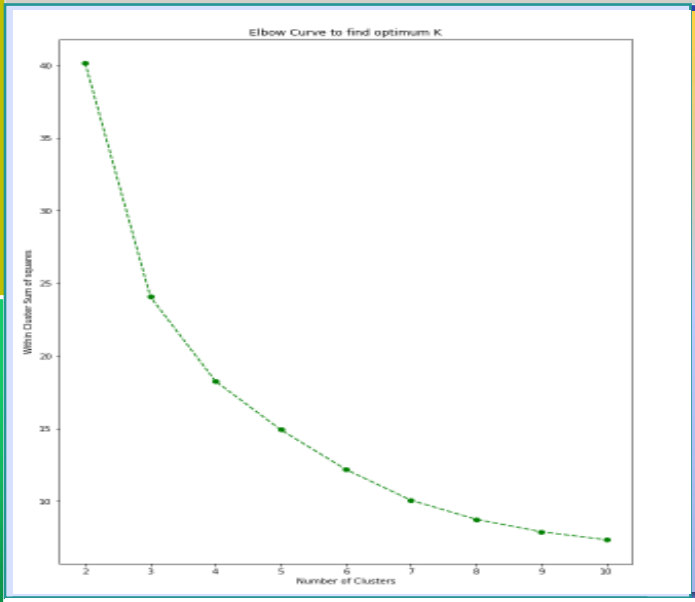
APPLIED DATA SCIENCE 1

Introduction

Clustering is one of the most important tasks in machine learning and clustering is the work that helps to divide the data points into the same or other data points. The report is based on the clustering and also curve fitting and in this report, the entire work has been performed in the Jupyter notebook platform and with the help of the python language. In this report, the entire technical work is based on k-means clustering. It is known that K-means clustering is a particular type of unsupervised learning and this process of machine learning has been used for the unlabeled data.

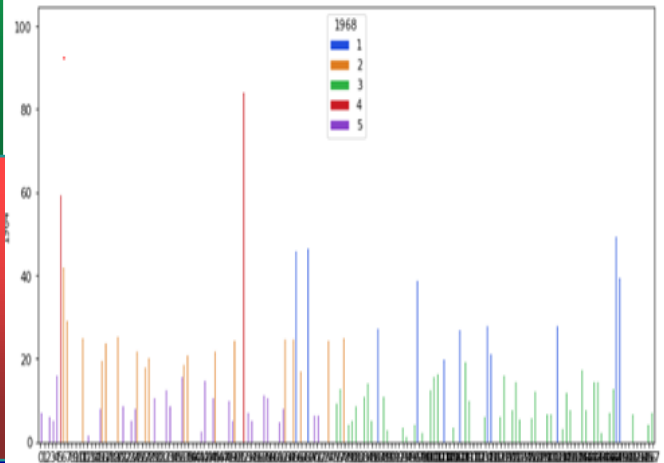
Technical analysis

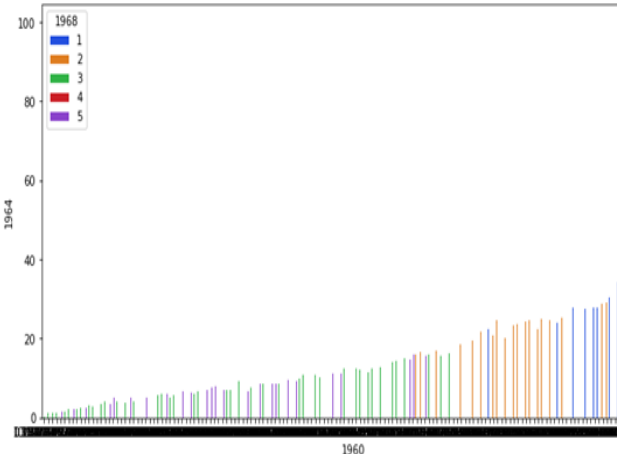
In this part of the report, the analysis of the technical part has been elaborated and the entire technical work has been performed in the jupyter notebook platform. The name of the language that has been used for performing the entire technical work in python. The dataset that has been used in the technical part is based on the carbon-di-oxide and GDP values of 30 or 40 years. The entire k means clustering has been performed on the basis of this dataset (Brunton and Kutz, 2022).



Contd...

It can be seen that the required relevant code has been generated for the calculation of centroid operation where the total values of clusters can be centrealled. In the variable centroid, the dataframe and the table of 12 columns with mentioned each column name individually separated by single quotation has been passed and the overall table and data has been arranged. This centroid method is a procedure or a process as well as for the representation of the cluster centre position for either real or imaginary location (Mehta and et al., 2018).





Conclusion

K- Means clustering is one of the most important parts that are included in machine learning. K- Means clustering is one of the most important techniques that are included in data clustering. The entire report is based on the k-means clustering and the curve plotting that has been performed in the jupyter notebook platform with the help of python language. It is so easy to implement the k-means algorithm in the python language. K means clustering helps to satisfy the unlabeled data into the predetermined number of the particulate that is based on k. The dataset that has been used in performing the K-means clustering consists of GDP and carbon-di-oxide values for 30 to 40 years. This is one of the most important algorithms that is used in the field of different machine learning parts.

References

Efron, B. and Hastie, T., 2021. Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science (Vol. 6). Cambridge University Press.

Mehta, S., Shen, X., Gou, J. and Niu, D., 2018. A new nearest centroid neighbor classifier based on k local means using harmonic mean distance. Information, 9(9), p.234.