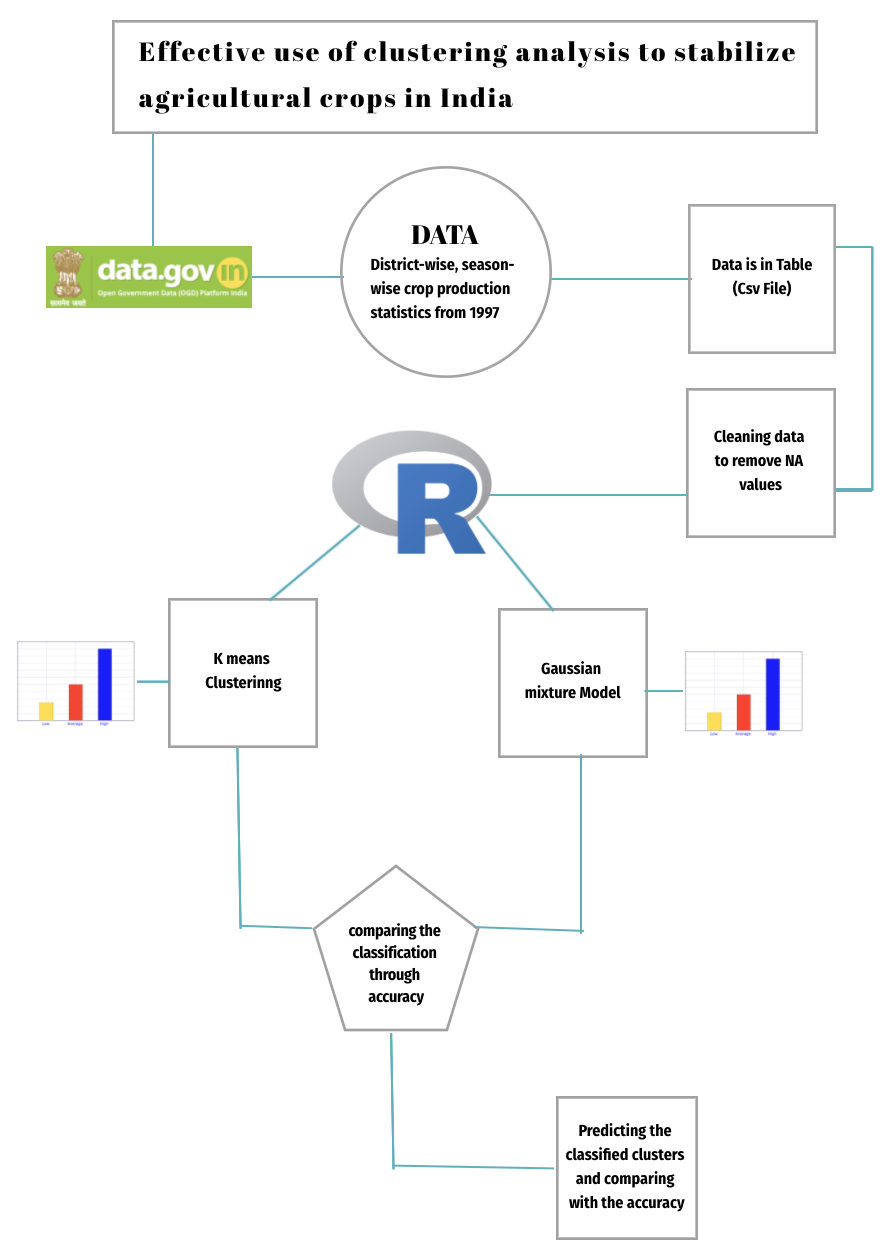
# Artefact Design & Development

Figure 1 : Artefact Design

# Explanation of Artefact

Artefact gives the study flow in a planned manner, where each phase is necessary to follow and needs to be followed in order to achieve the research results. In Figure 1 it shows that Crop Production data is taken from the website of Data.gov.in.

In this research data such as Production and Area of a specific year are derived from Crop Data and these variables are used as variables for the Classification of Crop Production and Area of crop region (all production and area for a particular year in different states). The data was then pre-processed to remove NA values in the Crop Dataset table as the data was having a lot of problems and then needs to be cleaned, which is done in r studio, after that the cleaned data has been pushed / stored in another table where the data can be used for further analysis.

R is used to execute various algorithms on the data set in a variety of scenarios, such as running an algorithm on the data set and obtaining the answer from it. Explain the findings and what the outcomes are. Seeking a method for classifying datasets into clusters and predicting clusters and also comparing the two cluster algorithms. Further, after classifying into 3 clusters high, normal and low values, we can know the low production and high production from different states. And also Checking even the accuracy of the 2 models after the classification of the actual dataset that generates the outcome according to the model. Predicting the classified clustered output to know whether there is a change in the low crop production states and high crop production states. Moreover, checking the output using the graph method and know the comparison between the clusters. Furthermore, comparing the 3 predicted clusters on both the algorithms using accuracy.

# Contribution of Artefact in this Research

The proposed artefact is unique and needs to be implemented as given, since each step is important and provides value to the research in which it is difficult to achieve the outcome, as shown in the end 2 clustering algorithms are used in R i.e. k means clustering and gaussian mixture modelling needs to be performed and then compared to see which algorithm is best suited for this research. Cluster analysis is usually performed to compare cases of groups where grouping composition was not identified before the analysis. Cluster analysis is a technique for grouping in unknown classes of items or objects. In steps the classification proceeds from the most general to the most unique. Cluster classification is a simple tool for defining homogeneous sets of objects or clusters. Objects share certain features in a particular cluster, but are rather dissimilar to objects without relating to that cluster. Today, K-means is a popular machine learning and data mining method. Despite its long history, k-means continues to enjoy widespread importance (Jain 2010); it has been listed among the top 10 algorithms in data mining (Wu et al. 2008). Today, there are several opportunities for ongoing research to promote, strengthen, and further improve k-means. Since clustering is always at the core of larger data mining activities, new developments in k-means have a strong influence. A Gaussian Mixing Model (GMM) is a parametric probability density function defined as a weighted sum of Gaussian component densities. A Gaussian mixture model (GMM) is useful for analysing data from one or many categories: categories can be different from each other, but data points within the same category can be well-modelled by a Gaussian distribution. If only one algorithm is used and no comparison is made with another existing algorithm, the Correct Findings may / may not be obtained and the performance of other specific algorithms may be more / less efficient based on conditions and datasets used. Also the change of dataset even the efficiency is affected. A single change can be a factor to create a large graph deflection which can also cause a major change in performance.