

DWDM – KMeans Clustering write up:

1. Describe output.

The KMeans Clustering algorithm will provide an out in the form of groups (clusters). The elements in these clusters are all related. For our dataset, it was run on countries and suicide numbers (input). It is anticipated there they may/may not be some unknown relationship/grouping between certain countries and suicide rates. The output (clusters) given at the end would give the answer.

2. Specify who would find this useful

The results of this would be of particular interest to policy makers (government).

Mental Health advocates and Professionals

3. Scenario of usefulness

Policy Makers (Government) - For example, in Japan, it was discovered that there was extremely high suicide rates among the working class, so the government implemented work limit policy and also incentivized taking vacations and leave. Similarly, this insight can be used to implement policies to lessen the rates. Advocates and Professionals - can use this data to raise

awareness about the issue and see what underlying factors are contributing to the issue and see how they can mitigate it respectively.

Autoregressive Integrated Moving Average Model(Arima) Time series Analysis

Write up:

1. Describe output.

The ARIMA model can be used to forecast future time steps. The predict function on the ARIMA results object to make predictions. It accepts the index of the time steps to make predictions as arguments. These indexes are relative to the start of the training dataset used to make predictions. The ARIMA model is created by calling its method and passing the p,d,q parameters. The p parameter is the number of lag observations included in the model, also called the lag order. The d parameter is The number of times that the raw observations are differenced, also called the degree of differencing. Finally the q parameter is the size of the moving average window, also called the order of moving average. A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model. The model is then trained by calling the fit function. After which predictions can be made specifying the index of time or times to be predicted.

2. Specify who would find this useful

Medical professionals, mental healthcare practitioners/advocates, and the general public would find this information useful. Using ARIMA to carry time series analysis thereby predicting possible future outcomes healthcare professionals can have an idea of the degree to

which suicides may increase in a specific country demographic etc. This information then allows them to set up mitigation strategies which could lead to a more controlled suicide rate.

3. Scenario of usefulness

An example of this information proves useful is if medical professionals were to learn that in the United States of America, suicides among males between the ages of 15 - 24 increased significantly around the 5th month of every year. What these professionals could now do is analyze the activities that happen around that time every year which may be an influential factor for suicides. Obtaining this information would then allow them to setup a strategy to lower the suicides in the demographic. It could also allow them to try targeting that demographic by providing counsellors since all these males would more or less be students in some institution.