Literature survey

INTRODUCTION

This literature review is based on a IEEE paper “ A Quality Based Automated Admission System for Educational Domain”. Follow are the key points extracted from the survey

The entire student admission system can be automated with the help of machine learning applying knowledge mining. This research focuses on analysis of k-means clustering algorithm as a simple and efficient tool to analyse the admission taken by the students in previous years

Methodology used in IEEE paper



Examples

Methods used

*V=V1, V2,… Vn be the set of data centres.*

*1. Let ‘C’ be the number of cluster centres selected randomly.*

*2. Calculate the Euclidian distance between each data set and*

*Cluster centre.*

*3. Assign each data point to its respective nearest cluster*

*Centre.*

*4. Recalculate the new cluster centre by using the following*

*Formula: Vi= (1/Ci) \*(sum of data entries)*

1

*Where Ci represents the number of data points in ith cluster.*

*5. Recalculate the distance between each data point and all the*

*Newly obtained cluster centres.*

*6. If no data point was re-assigned then stop the process, else*

*Go to step 3.*

*B. Benefits of K-Means Algorithm*

Latency is low and is more robust and easily

Understandable.

Comparatively simple and efficient than other

Algorithms.

It yields optimum result when dataset provided are

Distinct and quite separated from each other

Our dataset comprises 13 different features collected during student admission process. A collection of 129 records were aggregated by consulting different parents



Comparison

|  |  |
| --- | --- |
| IEEE paper | Our project |
| Uses a classifier model | Uses a regression model |
| Uses the kmeans classifier | Uses multiple regression  (multiple has the best results when compared to polynomial ,decision tree etc) |
| Provides binary output .If the student will get admitted or not | Provides a probabilistic output, which is more accurate |
| Sum of squared error for 129 data entries is 669.0  This is because the input data provide contained response variables to be of binary format.Hence not provide adequate information to the | Sum of squared errors of 200 data entries 0.0017.  Our project the dataset had response variables which were probabilistic which provided more information to the regressor model and  Hence our model appears to be more successful |

Scope for future work

This system model will also be helpful in finding the faulty areas of institute due to which admissions

have reduced over time and hence will be beneficial in formulating steps to improve admission performance in the subsequent sessions. Thus our proposed model is helpful in listing out students on which main focus is to be drawn to transform inquiry into admission