



UNIVERSITY WITH A PURPOSE





# MINOR-1

## Student Segmentation using Clustering Algorithms

MEMBER'S NAME	ROLL NUMBER	SAP ID	BRANCH
Rohan Nyati	R177219148	500075940	BTech CSE-AIML
Rajneesh	R177219143	500076347	BTech CSE-AIML
Shantanu Jaswal	R177219170	500075224	BTech CSE-AIML

Under the guidance of  
Sujoy Chatterjee

# Introduction

1. Achieving Student Segmentation using clustering algorithms
2. In this project we are using these clustering algorithms :
  - Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a base algorithm for density-based clustering. It can discover clusters of different shapes and sizes from a large amount of data, which is containing noise and outliers.
  - K-means clustering algorithm computes the centroids and iterates until we it finds optimal centroid. It assumes that the number of clusters are already known. It is also called flat clustering algorithm. The number of clusters identified from data by the algorithm is represented by 'K' in K-means.

# Motivation

The Motivation behind opting for this is that for the past one year, students have been studying in online mode and because of which they are facing difficulties in some areas. So our motivation here is to access the college in managing such students and assisting them to get back on track.

As a ML enthusiast, we wanted to explore the core of these algorithms using a structured programming approach to gain the pure insights and working of this algorithm.

# Problem Statement

If there are 1,2,3,.....,n number of students :

1. How to segment those students into different clusters ( $C_1, C_2, C_3 \dots$ )
2. How to map new students into particular cluster ( $P_1, P_2, \dots$ )
3. If m number of new students join then how we can cluster them without using algorithm again.

# Literature Review

1. DBSCAN is a density based algorithm [1] that requires only one input parameter and supports the user in determining an appropriate value for it.
2. INCREMENTAL DBSCAN CLUSTERING [2] insertion of some new data items into the already existing clusters. outliers which fulfil the Minpts & eps criteria , combinly can form clusters using DBSCAN.
3. Incremental DBSCAN clustering algorithm [3] is used to handle dynamic databases.
4. The K-Means algorithm based on dividing [4] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition.
5. The incremental K-means algorithm[5] presented in this paper is similar to the block sequential algorithm with the exception that each block is accessed only one time. Each block is going through a set of  $l$  epochs of K-means where the final centers of block  $i$  are used as the initial centers for block  $i+1$ .

# Objective

1. Clear understanding of Clustering algorithms
2. To create a way for colleges to obtain a better idea about how much help does a student requires after returning to college to improve their academic status .
3. To create clusters of students on the basis of their current academic knowledge .

# Methodology

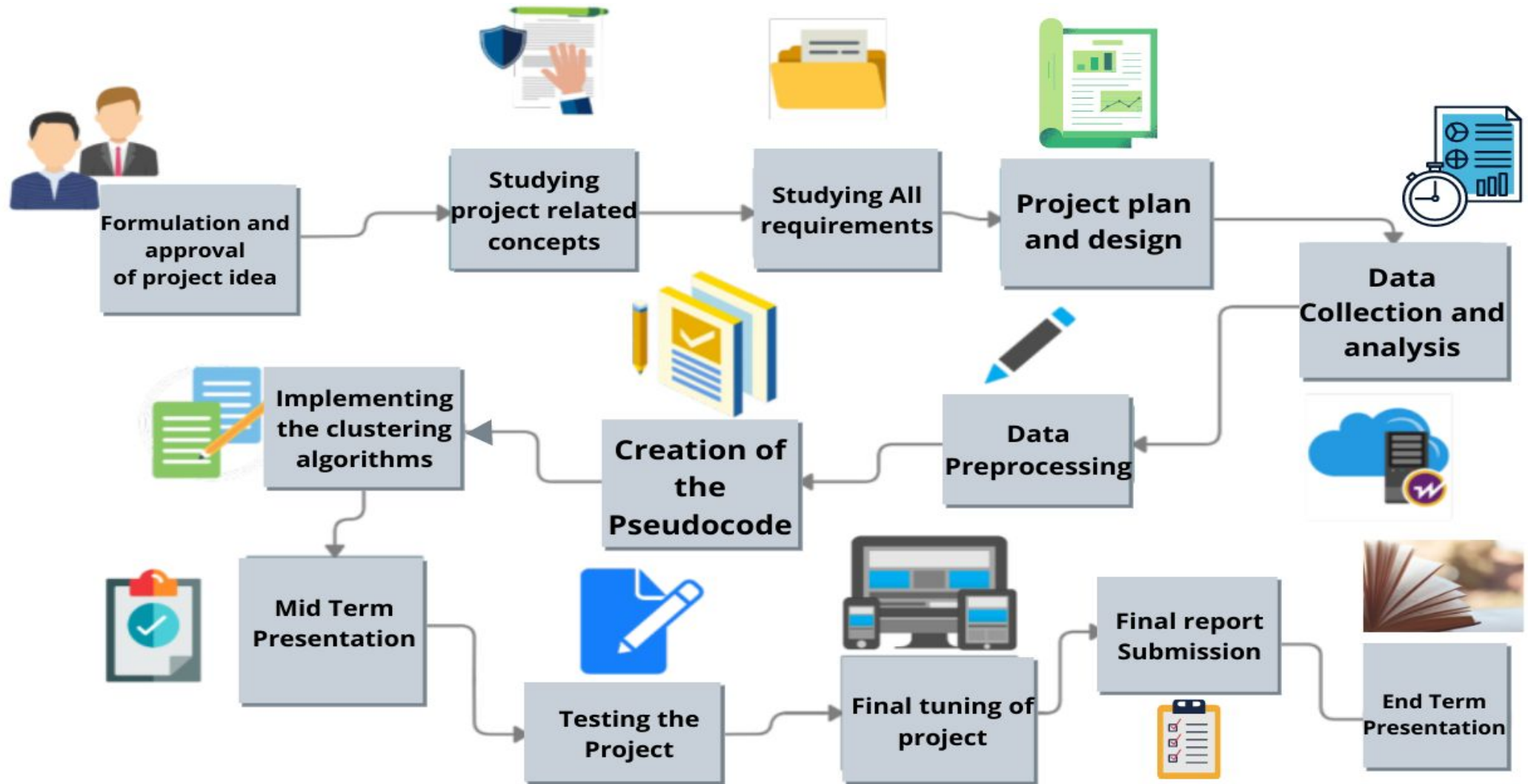
Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data.

Steps -

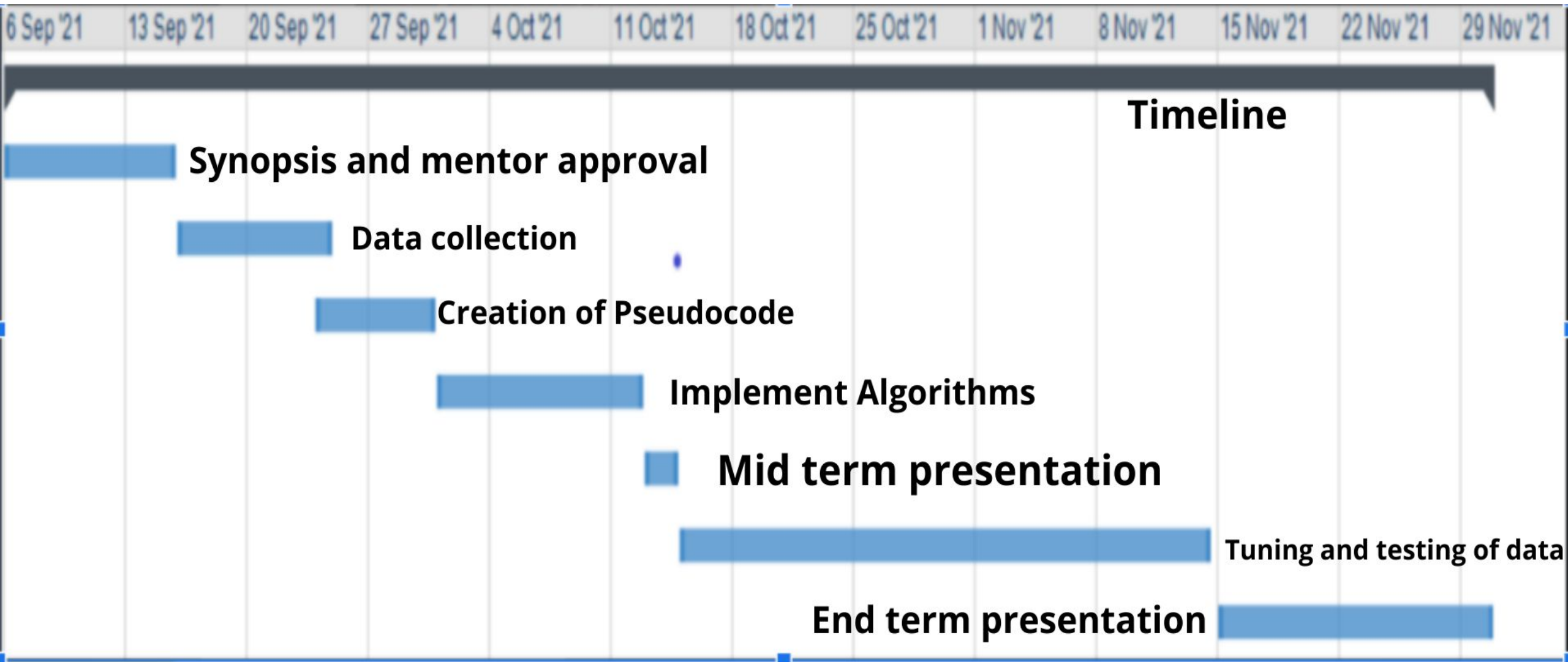
1. Problem understanding and definition
2. Data collection and preparation
3. Dataset understanding using various clustering algorithms
4. Data analysis
5. Data validation
6. Deployment



# Flow of work



# GANTT CHART



# Future Work to be done

1. Data Creation
2. Data Preprocessing
3. Creating Pseudo Code
4. Implementing Clustering Algorithms
5. Data Validation
6. Deployment

# References

- [1] A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu Institute for Computer Science, University of Munich Oettingenstr. 67, D-80538 Miinchen, German  
<https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>
- [2] International Journal of Enterprise Computing and Business Systems Analysis and Study of Incremental DBSCAN Clustering Algorithm SANJAY CHAKRABORTY Prof. N.K.NAGWANI National Institute of Technology National Institute of Technology (NIT) Raipur, CG, India.  
<https://arxiv.org/ftp/arxiv/papers/1406/1406.4754.pdf>
- [3] A Technical Survey on DBSCAN Clustering Algorithm Nidhi Suthar<sup>1</sup>, Prof. Indr jeet Rajput<sup>2</sup>, Prof. Vinit Kumar Gupta<sup>3</sup>  
<sup>1</sup> Department of Computer Engineering, Hashmukh Goswami college of Engineering, Vahelal, Gujarat.  
<https://www.ijser.org/researchpaper/A-Technical-Survey-on-DBSCAN-Clustering-Algorithm.pdf>
- [4] A Clustering Method Based on K-Means Algorithm Youguo Li, Haiyan Wu Department of Computer Science Xinyang Agriculture College Xinyang, Henan 464000, China  
[https://www.researchgate.net/publication/271616608\\_A\\_Clustering\\_Method\\_Based\\_on\\_K-Means\\_Algorithm/link/57da70fc08aeea1959316130/download](https://www.researchgate.net/publication/271616608_A_Clustering_Method_Based_on_K-Means_Algorithm/link/57da70fc08aeea1959316130/download)
- [5] Dynamic Incremental K-means Clustering Bryant Aaron, Dan E. Tamir Department of Computer Science, Texas State University, San Marcos, Texas, USA, Naphtali D. Rishe, and Abraham Kandel School of Computing and Information Sciences Florida International University Miami, Florida, USA  
[http://cake.fiu.edu/Publications/Aaron+al-14-DK.Dynamic\\_Incremental\\_K-means\\_Clustering\\_IEEE-downloaded.pdf](http://cake.fiu.edu/Publications/Aaron+al-14-DK.Dynamic_Incremental_K-means_Clustering_IEEE-downloaded.pdf)



THANK YOU

---



UNIVERSITY WITH A PURPOSE