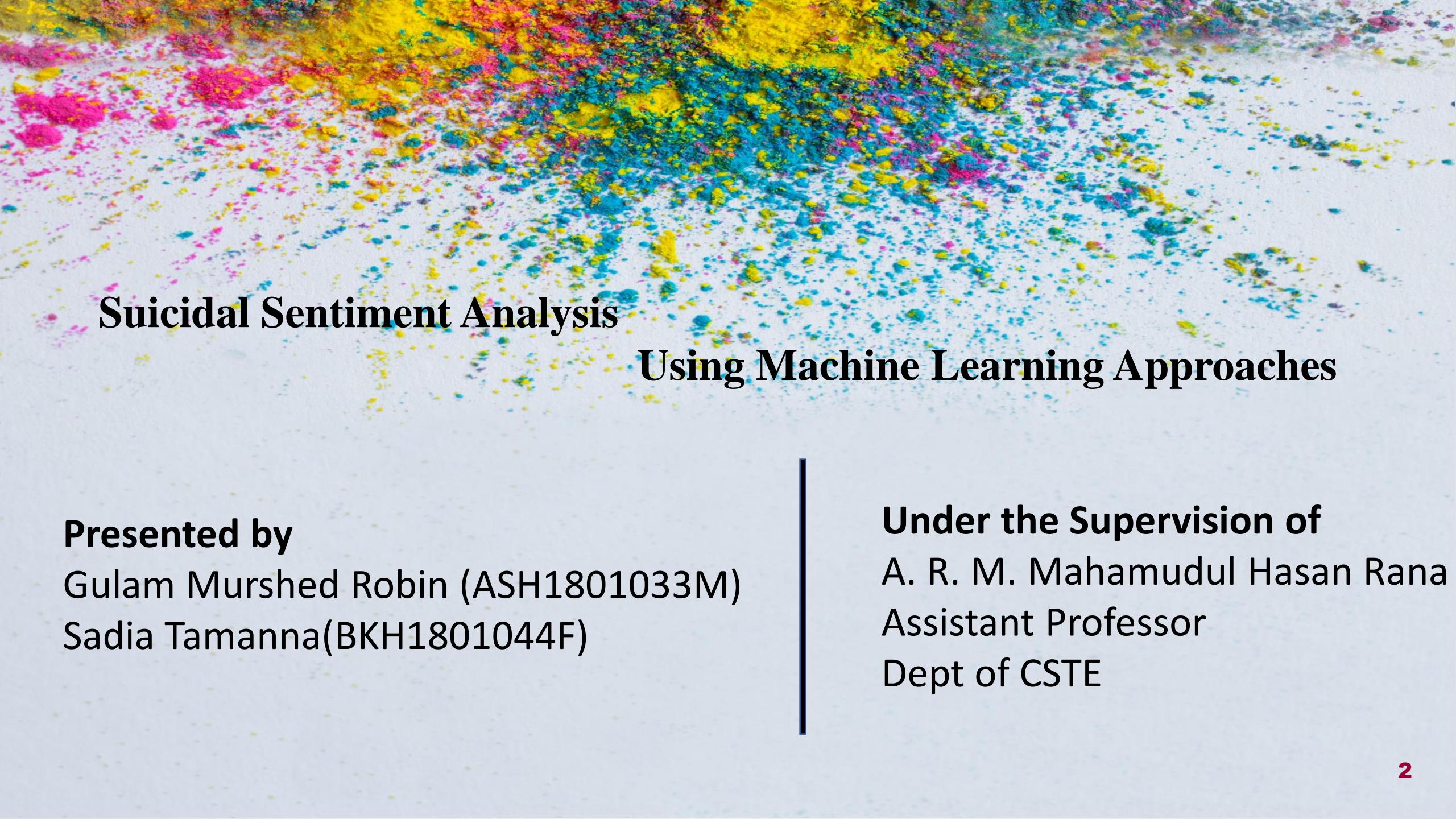


“Suicide is not a blot on anyone’s name; it is a tragedy.”

– Kay Redfield Jamison





Suicidal Sentiment Analysis

Using Machine Learning Approaches

Presented by

Gulam Murshed Robin (ASH1801033M)

Sadia Tamanna(BKH1801044F)

Under the Supervision of

A. R. M. Mahamudul Hasan Rana

Assistant Professor

Dept of CSTE



OUTLINE



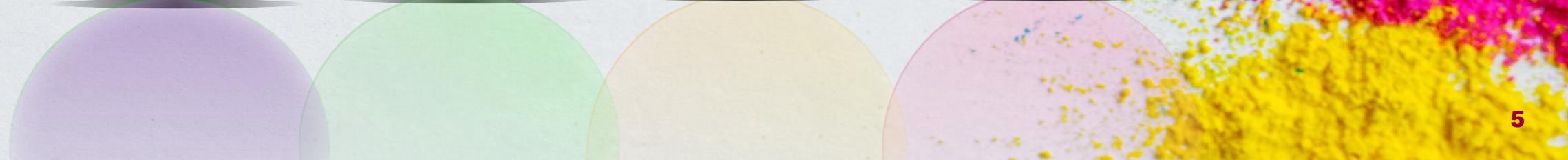
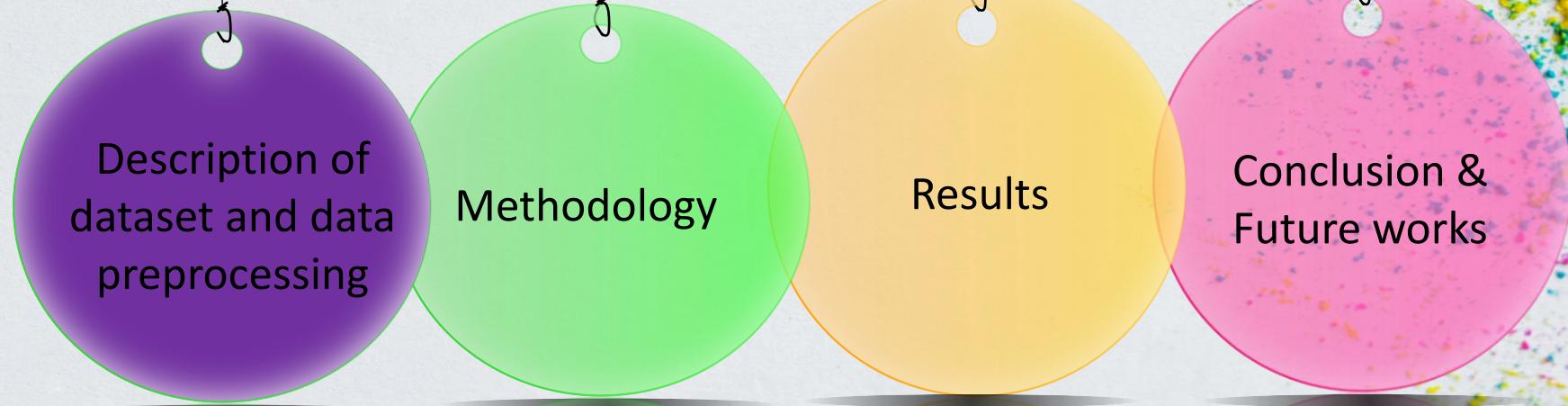
About Suicide
and reasons why
people attempt it.

Motivation and
Aims

Objectives

Related works

Procedure of
Suicidal Analysis

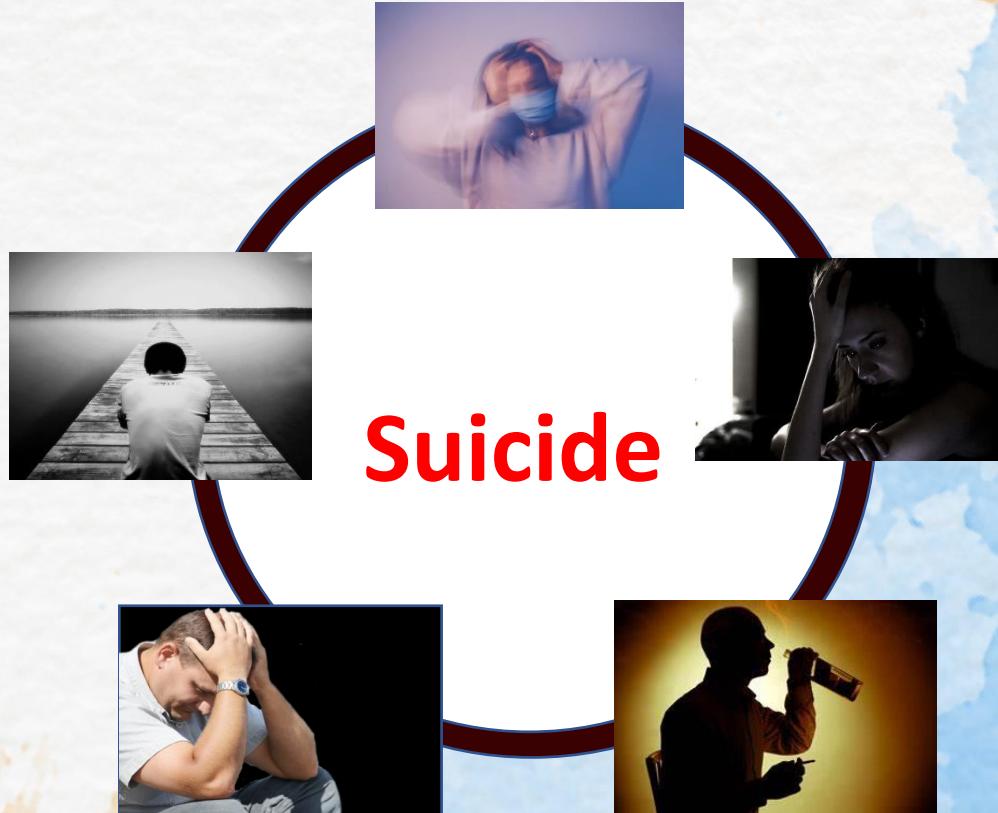


What is Suicide and why do people attempt it?

Suicide is the act of intentionally causing one's own death.

While there are many factors that can influence a person's decision to commit suicide, the most common factors are given below:

- 1) Mental Illness
- 2) Traumatic Stress
- 3) Substance Use and Impulsivity
- 4) Loss or Fear of Loss
- 5) Social Isolation



Motivation and Aims

The motivation and aims of suicidal sentiment analysis are to improve our ability to identify and intervene with individuals who may be at risk of suicide, and ultimately reduce the incidence of suicide.

The specific aims of suicidal sentiment analysis can vary depending on the application, but some common goals include:

- ✓ Identifying individuals at risk of suicide
- ✓ Developing suicide prevention tools
- ✓ Understanding the language of suicide

Objectives of our work

- To design an efficient Suicidal Analysis method.
- To Identify individuals who may be at risk of suicide.
- To Suicidal Sentiment Analysis by using Supervised and Unsupervised Learning .



Related Works

Suicide and Depression Identification with Unsupervised Label Correction[1]

Ensemble Classifier Model for Tweeter Sentiment Analysis[2]

Supervised Learning for Suicidal Ideation Detection in Online User Content[3]

An Unsupervised Learning Approach for Automatically to Categorize Potential Suicide Messages in Social Media[4]

Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm [5]

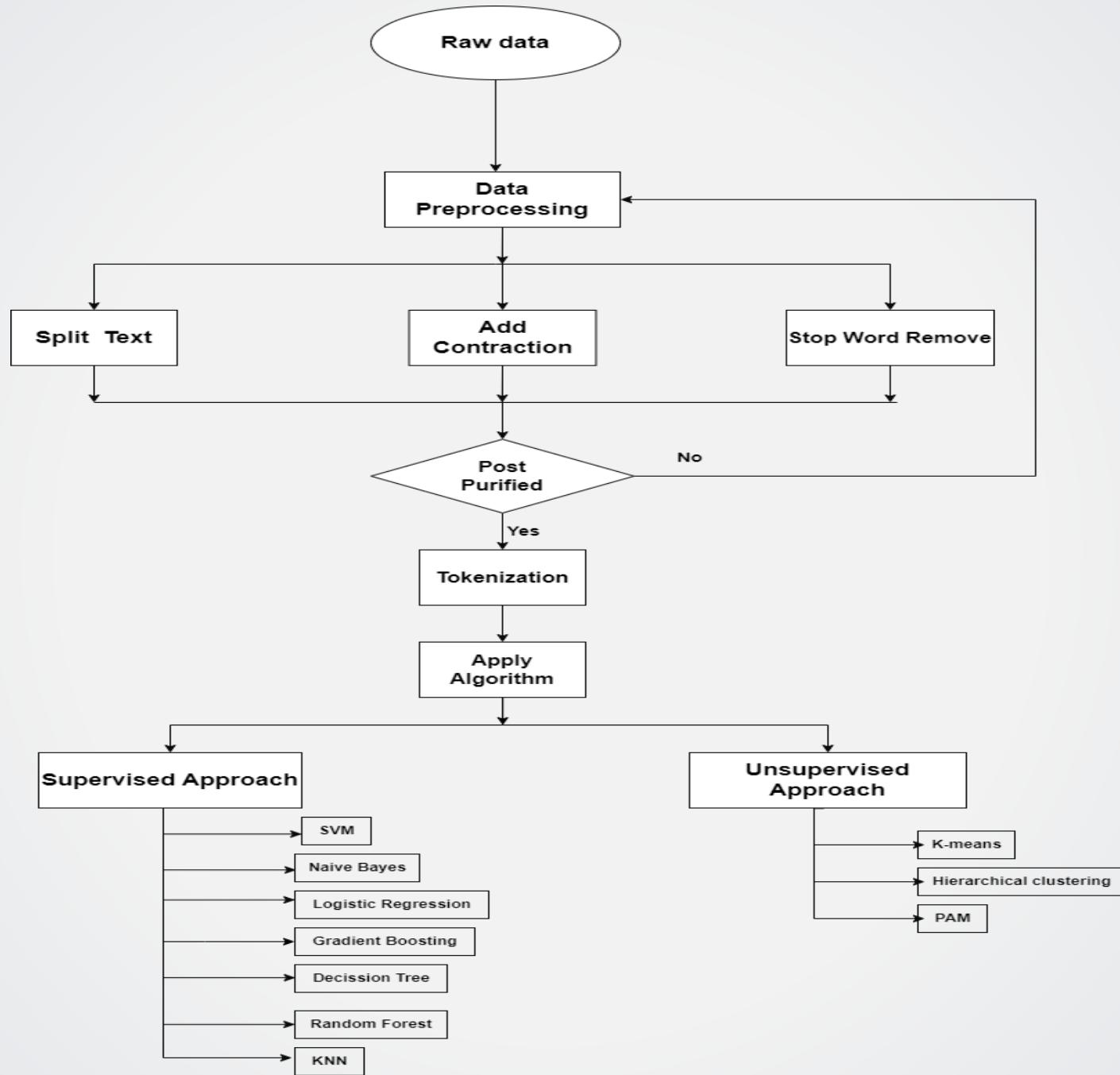
Suicidal ideation prediction in twitter data using machine learning techniques[6]

Probabilistic model to indicate depression[7]

Behavioral Analysis of Depressed Sentimental Over Twitter: Based on Supervised Machine Learning Approach[8]

Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks[9]

Procedure of Suicidal Analysis



Description of dataset

The dataset used in this research is collected from The University of Maryland Reddit Suicidality Dataset, Version 2. There are total 6,98,997 text data in it. Including, for each post, the post_id, anonymized user_id, timestamp, de-identified post title, de-identified post body, Combine post_title and post_body in text, labels and suicide_class.

In []: #first 10 row from dataset
suicide_data.head(10)

Out[5]:

	Unnamed: 0	post_id	user_id	timestamp	post_title	post_body	text	labels	suicide_class
0	0	wfimt	22002	1342075703	[real] motivation	This is my first post on reddit. Some time ago...	[real] motivation This is my first post on red...	0	Suicide
1	1	1bsqv3	22002	1365261010	simple question about transferring acc to anoth...	Hi.. What will happen with my ranked rating? I...	simple question about transferring acc to anoth...	1	Non_Suicide
2	2	1dr0xf	22002	1367787358	simple question: Did you get unbanned?	Hi. Simple question. Did you get unban from a ...	simple question: Did you get unbanned? Hi. Sim...	1	Non_Suicide
3	3	1e0noi	22002	1368125785	I can't win. Why... and it's noy my fault.	Hi... Am playing at Eu west... am diamond 5 ri...	I can't win. Why... and it's noy my fault. Hi....	1	Non_Suicide
4	4	1f0y6g	22002	1369483647	Diamond 5. Mrr rest	If i will switch server to EU and then go back...	Diamond 5. Mrr rest If i will switch server to...	1	Non_Suicide
5	5	1g4s52	22002	1370970723	Scared of next ban. What can i do?	Hi guys. I already got a one perm ban. I creat...	Scared of next ban. What can i do? Hi guys. I ...	1	Non_Suicide
6	6	1it5ts	22002	1374498763	Too all players who had huge impact on losing ...	Instead of "you are reported" you should say t...	Too all players who had huge impact on losing ...	1	Non_Suicide
7	7	1sge5s	22002	1386584670	Short question, short story. Pls help me decide.	Hi am LOL player.. I think i dont understand n...	Short question, short story. Pls help me decid...	1	Non_Suicide
8	8	1sjbf9	22002	1386664896	Maybe it's the right time to fa...	Maybe it's the right moment TO **FACE** to fu...	Maybe it's the right time to fa... Maybe it's ...	0	Suicide
9	9	226mcd	22002	1396611845	My conclusion after thinking about suicide in ...	Well, i want to share it with somebody... If i...	My conclusion after thinking about suicide in ...	0	Suicide

Figure : Sample Data from the Dataset

Data Preprocessing

In data preprocessing we use some of text preprocessing steps,

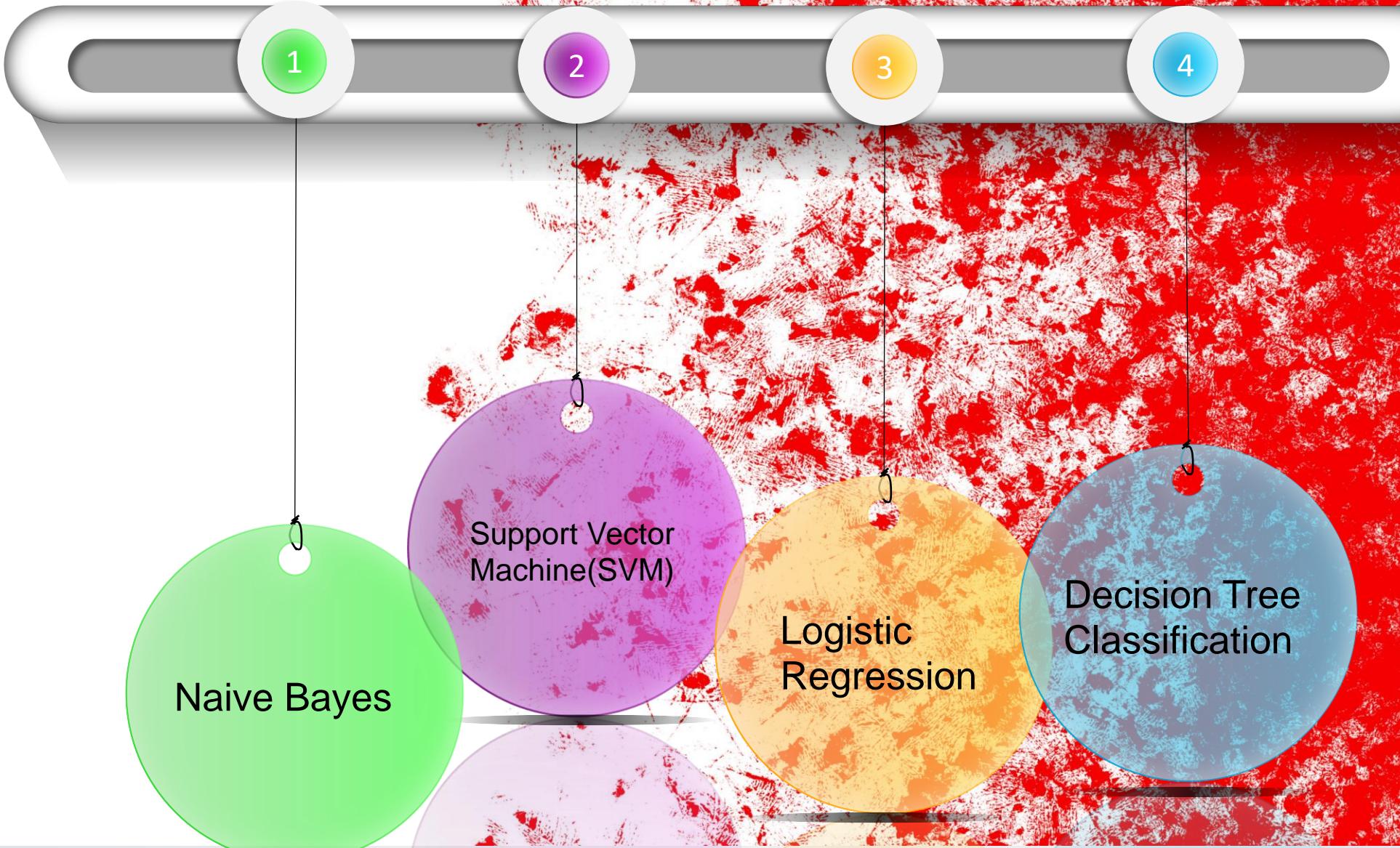
- ✓ Tokenization
- ✓ Lower casing
- ✓ Remove punctuation ,signs , number etc
- ✓ Stop words removal
- ✓ Stemming
- ✓ Lemmatization

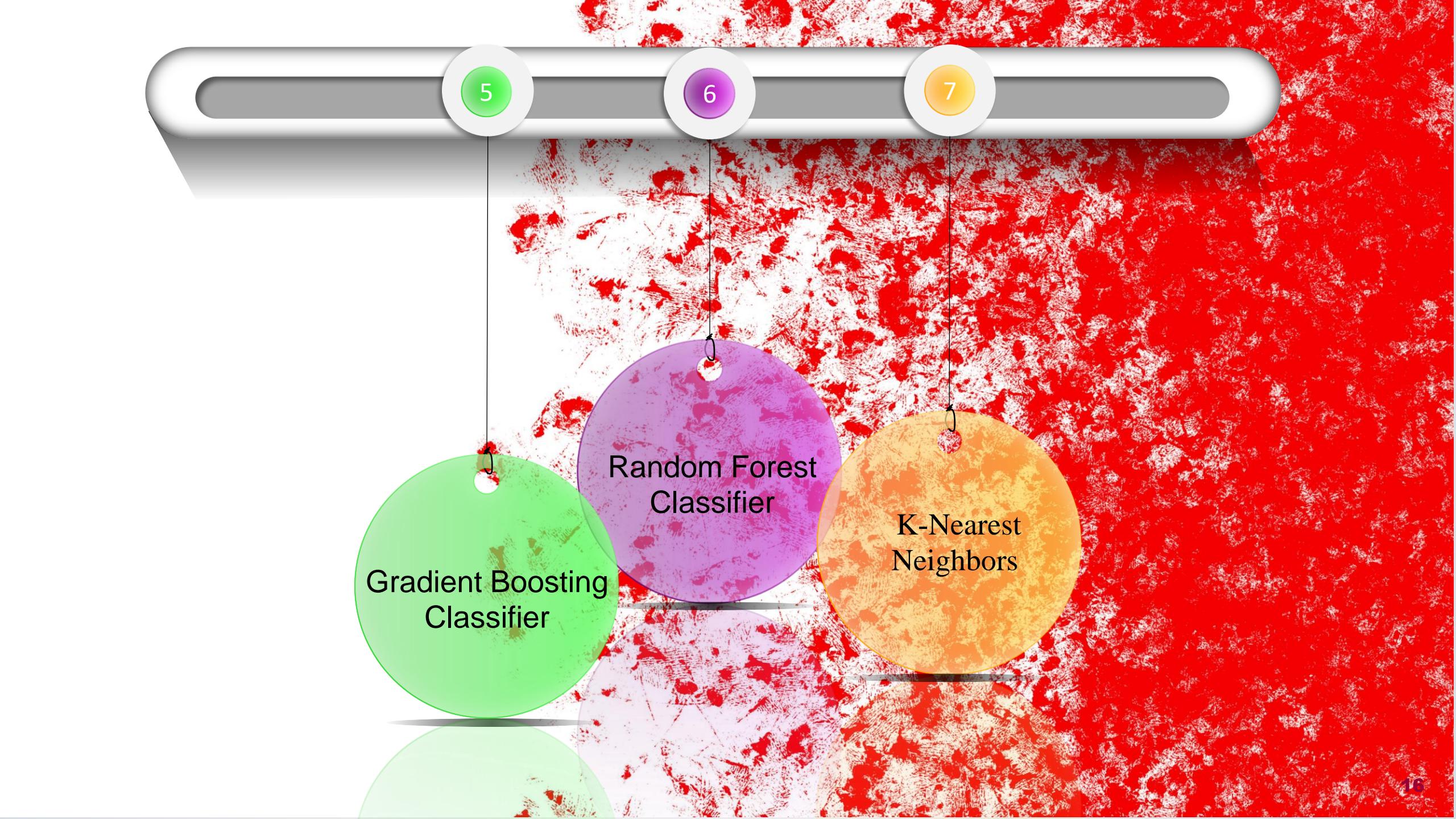
Unnamed: 0	post_id	user_id	timestamp	post_title	post_body	text	labels	suicide_class	
0	0	wfimt	22002	1342075703	[real] motivation	This is my first post on reddit. Some time ago...	motiv first post reddit time ago want chang li...	0	Suicide
1	1	1bsqv3	22002	1365261010	simple question about transferring acc to anoth...	Hi.. What will happen with my ranked rating? I...	simpl question transfer acc anoth server hi ha...	1	Non_Suicide
2	2	1dr0xf	22002	1367787358	simple question: Did you get unbanned?	Hi. Simple question. Did you get unban from a ...	simpl question get unban hi simpl question ge...	1	Non_Suicide
3	3	1e0noi	22002	1368125785	I can't win. Why... and it's noy my fault.	Hi... Am playing at Eu west... am diamond 5 ri...	win noy fault hi play eu west diamond right dr...	1	Non_Suicide
4	4	1f0y6g	22002	1369483647	Diamond 5. Mrr rest	If i will switch server to EU and then go back...	diamond mrr rest switch server eu go back west...	1	Non_Suicide

Fig: After Preprocessing

Methodology

Supervised Approaches





5

6

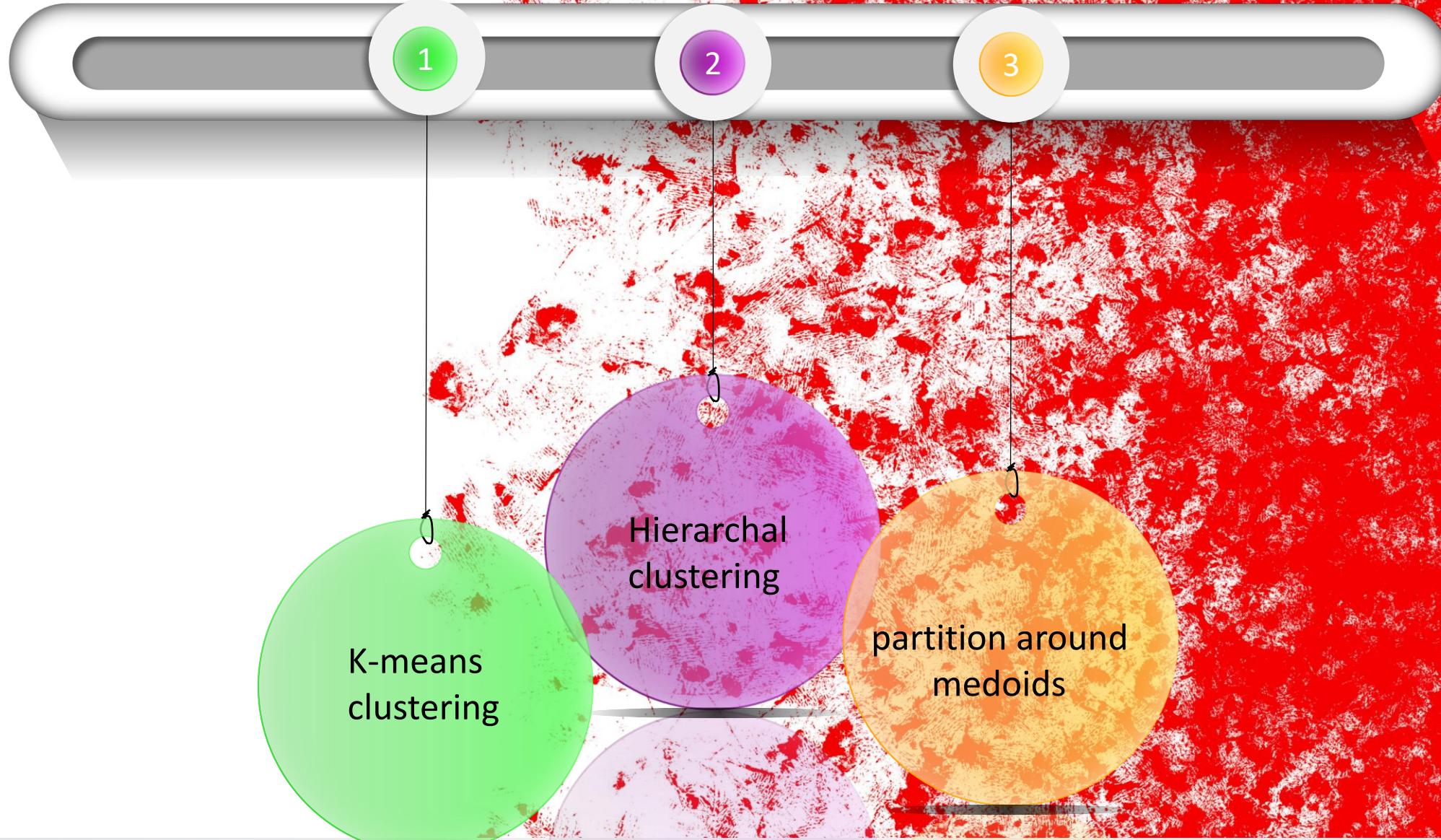
7

Gradient Boosting
Classifier

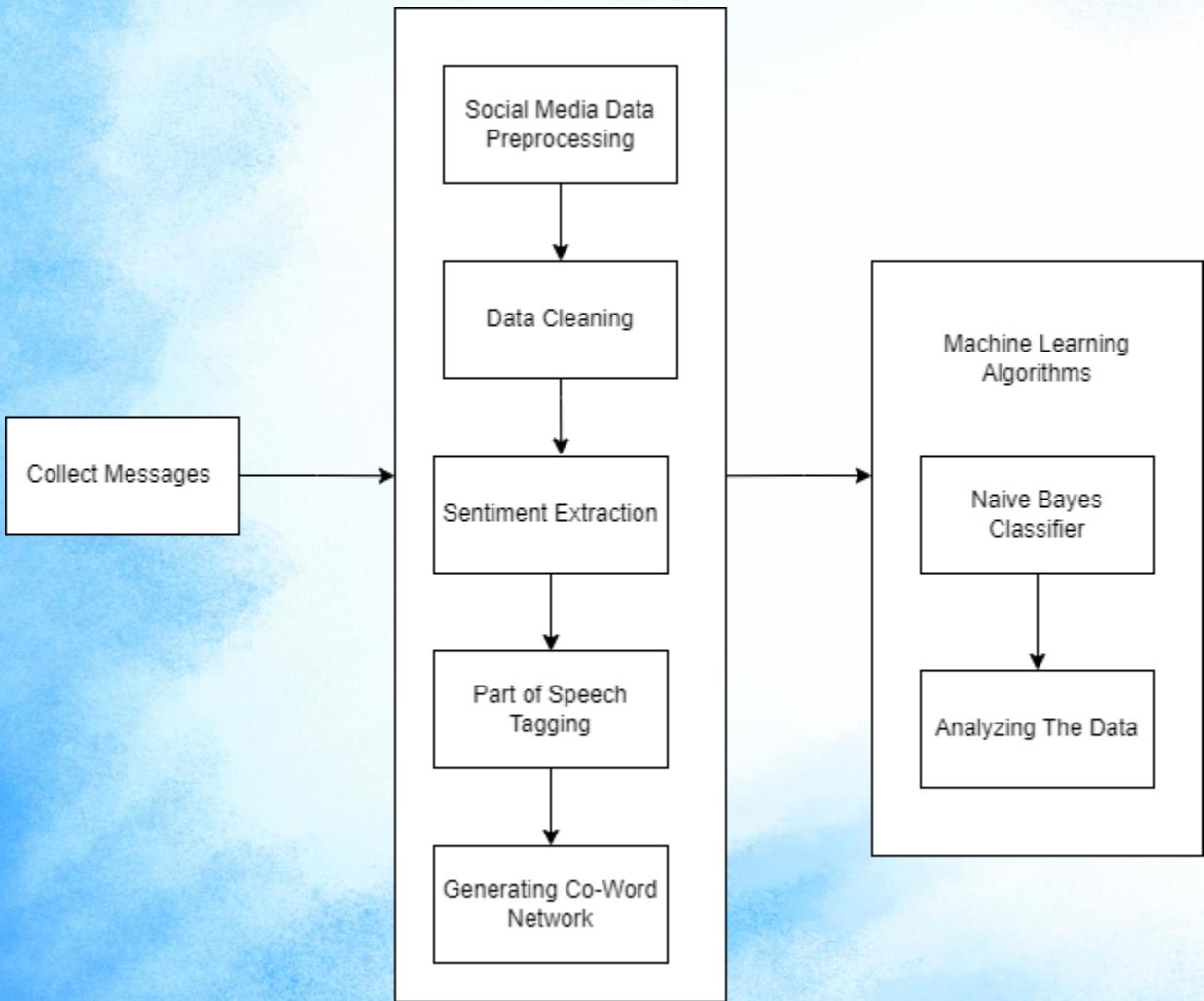
Random Forest
Classifier

K-Nearest
Neighbors

Unsupervised Approaches

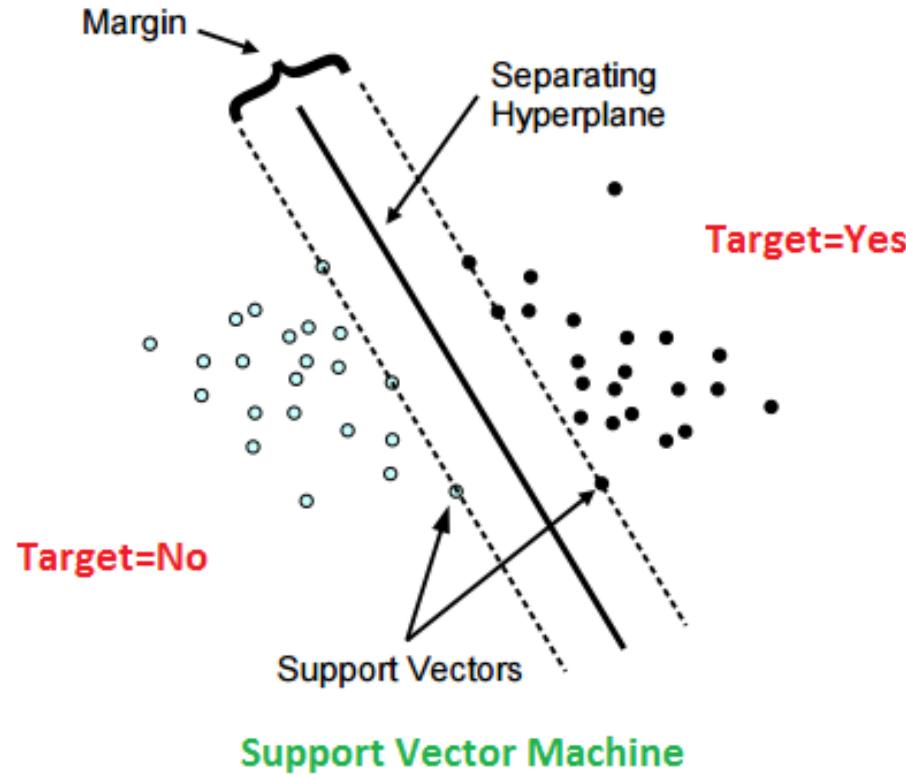


Naive Bayes

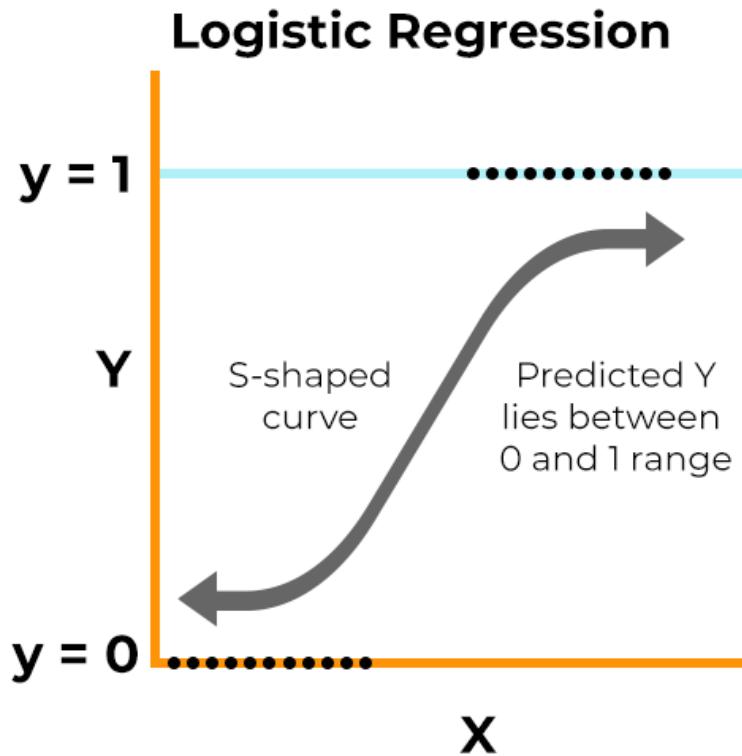


- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the dataset into training and testing sets
- ✓ Calculate the prior probability of each class
- ✓ Calculate the conditional probability of each feature for each class
- ✓ Use Bayes theorem to calculate the posterior probability of each class for a new instance
- ✓ Choose the class with the highest posterior probability as the predicted class
- ✓ Evaluate the performance of the classifier on the testing dataset.
- ✓ Iterate and improve the model by adjusting hyperparameters or adding new features to the model based on the evaluation results

Support Vector Machine

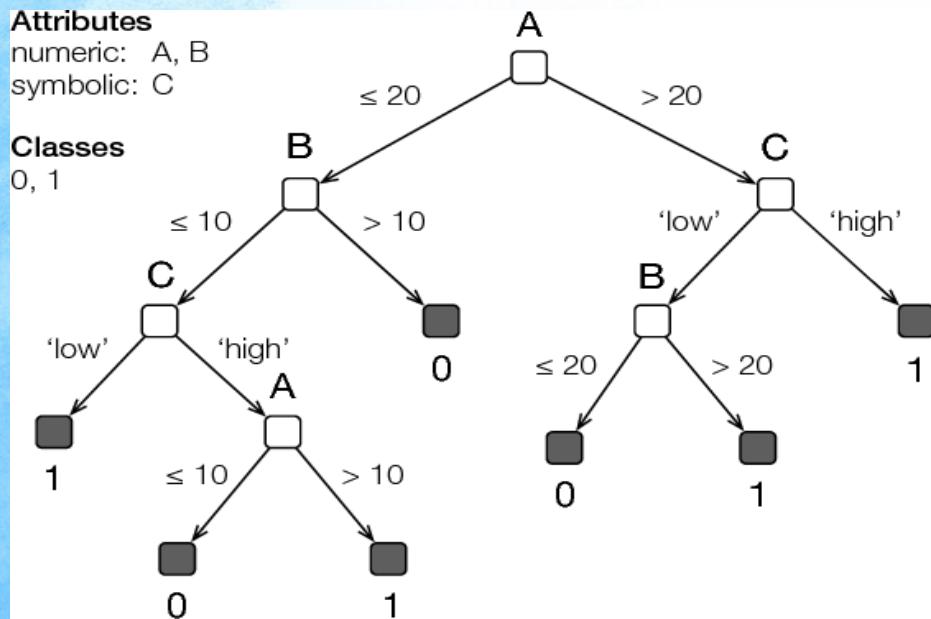


- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the dataset into training and testing sets
- ✓ Choose an appropriate kernel function
- ✓ Train the SVM model
- ✓ Evaluate the performance of the model on the testing dataset
- ✓ Iterate and improve the model



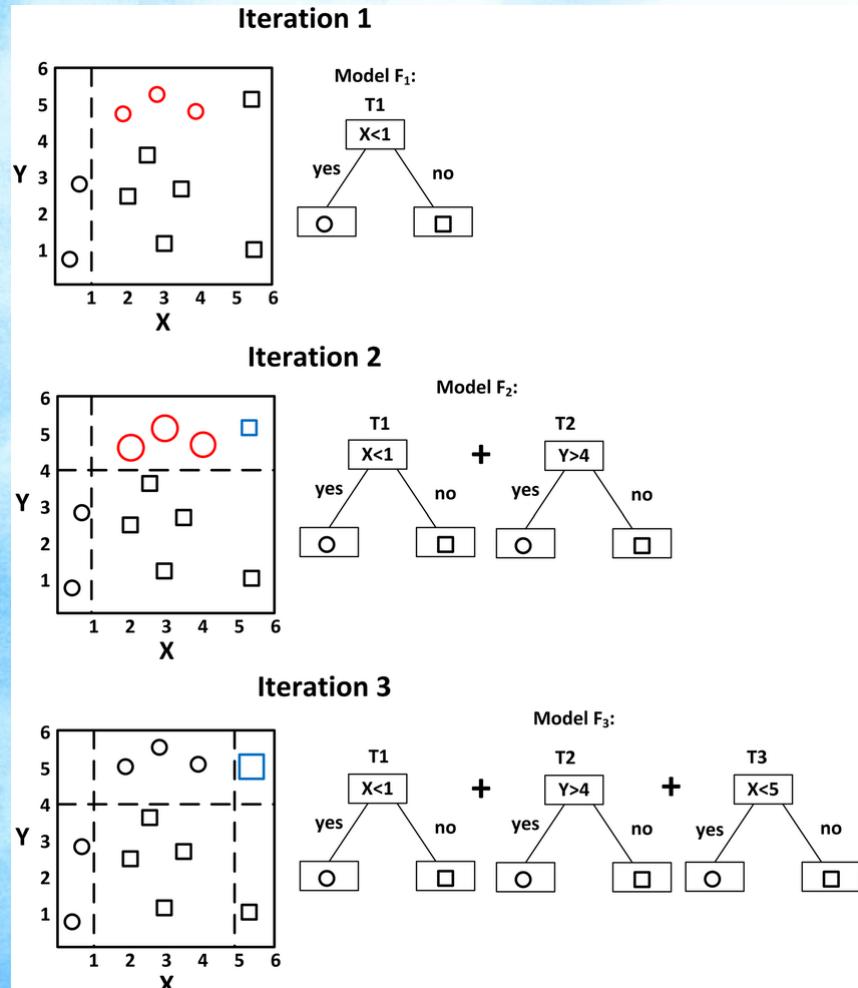
- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the dataset into training and testing sets
- ✓ Define the logistic regression model
- ✓ Train the logistic regression model
- ✓ Evaluate the performance of the model on the testing dataset
- ✓ Iterate and improve the model

Decision Tree Classification



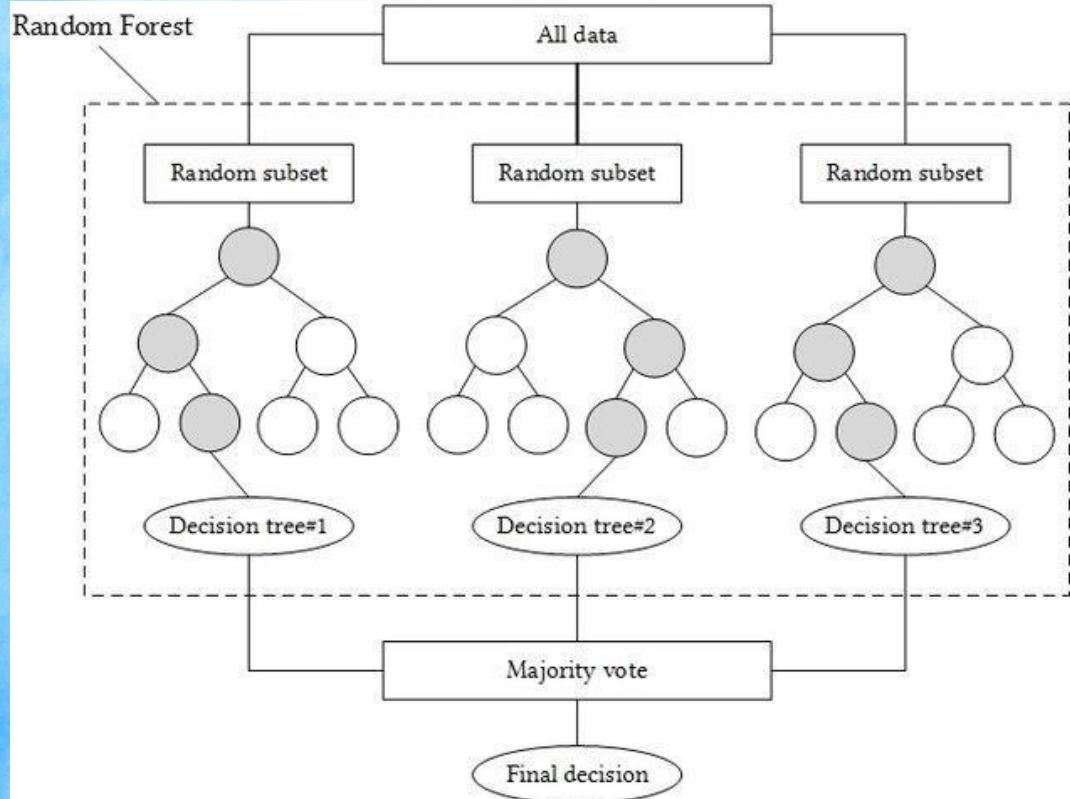
- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the dataset into training and testing sets
- ✓ Define the decision tree model
- ✓ Train the decision tree model
- ✓ Evaluate the performance of the model on the testing dataset
- ✓ Iterate and improve the model

Gradient Boosting Classifier



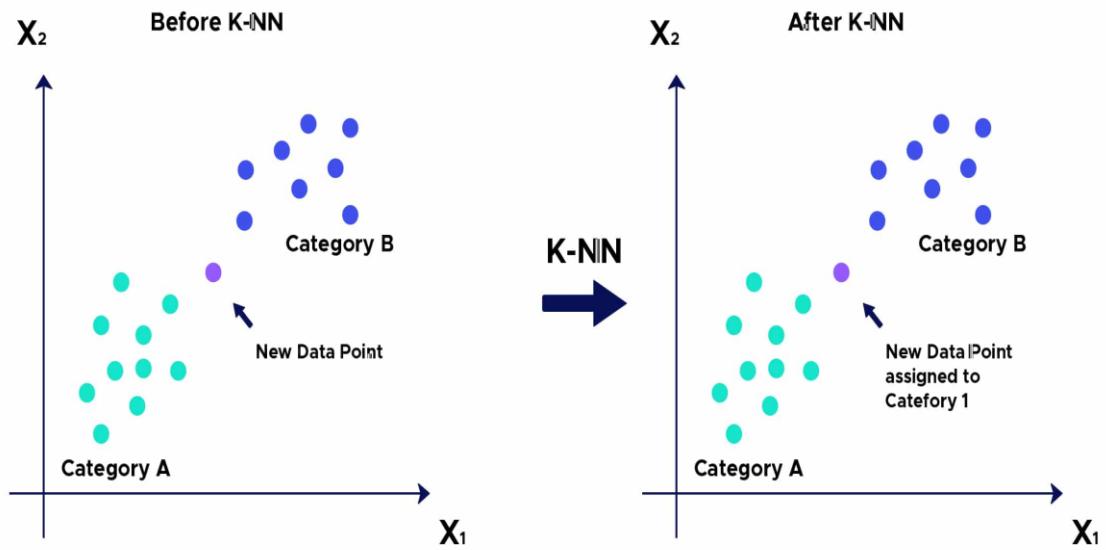
- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the dataset into training and testing sets
- ✓ Define the gradient boosting model
- ✓ Train the gradient boosting model
- ✓ Evaluate the performance of the model on the testing dataset
- ✓ Iterate and improve the model

Random Forest Classifier



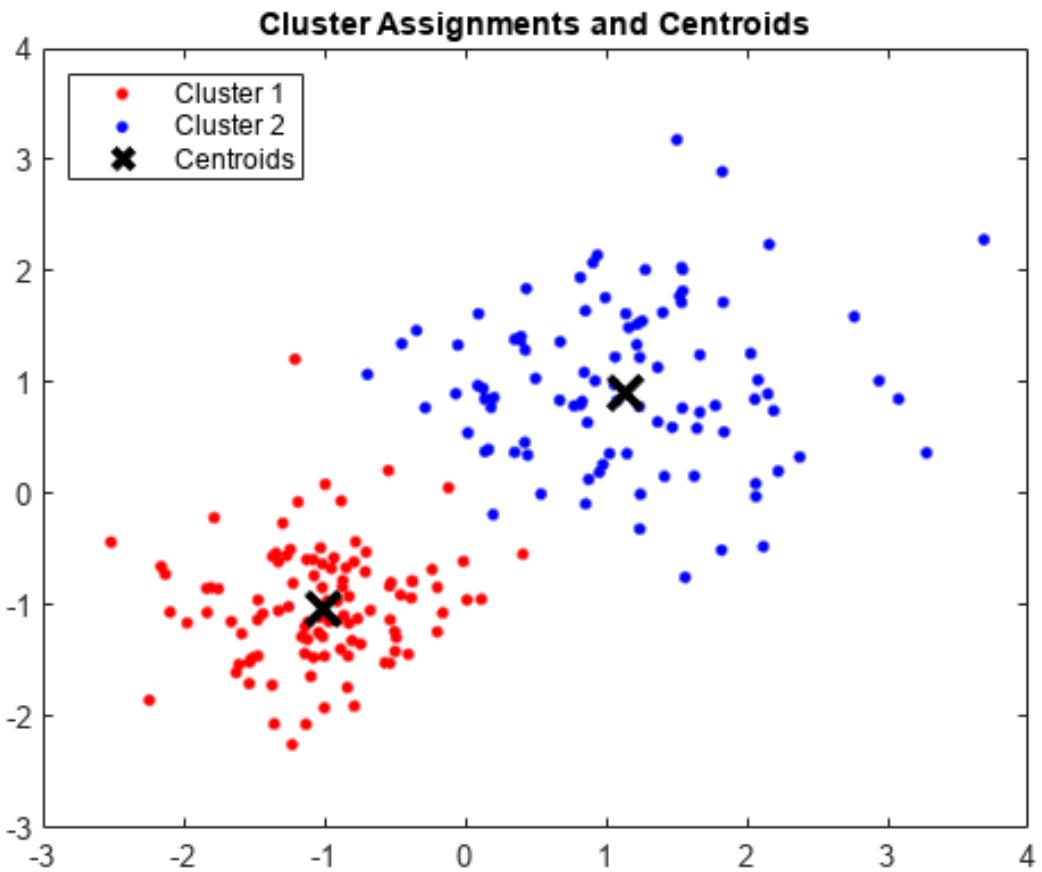
- ✓ Gather the training dataset
- ✓ Preprocess the dataset
- ✓ Split the input data into multiple random subsets (bootstrapping)
- ✓ Select a random subset of features
- ✓ Train a decision tree on the subset using the selected features
- ✓ Predict the class label of a new input by passing it through each of the decision trees and taking a majority vote

K-Nearest Neighbors



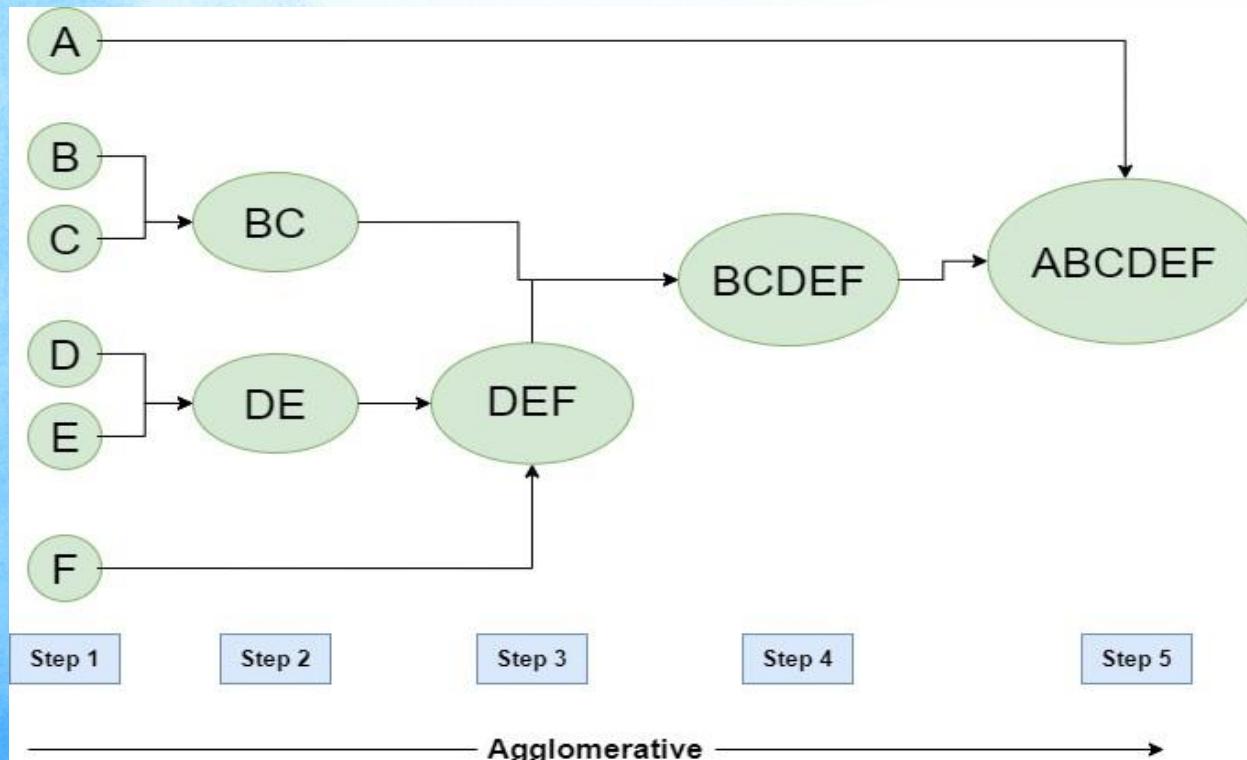
- ✓ Load the training dataset
- ✓ Preprocess the dataset
- ✓ Choose the value of k (number of neighbors)
- ✓ Compute the distance between the new input and each instance in the training dataset
- ✓ Select the k instances with the smallest distance to the new input
- ✓ Assign the class label of the new input based on the majority vote of its k nearest neighbors
- ✓ Output the predicted class label.

K-means Clustering:



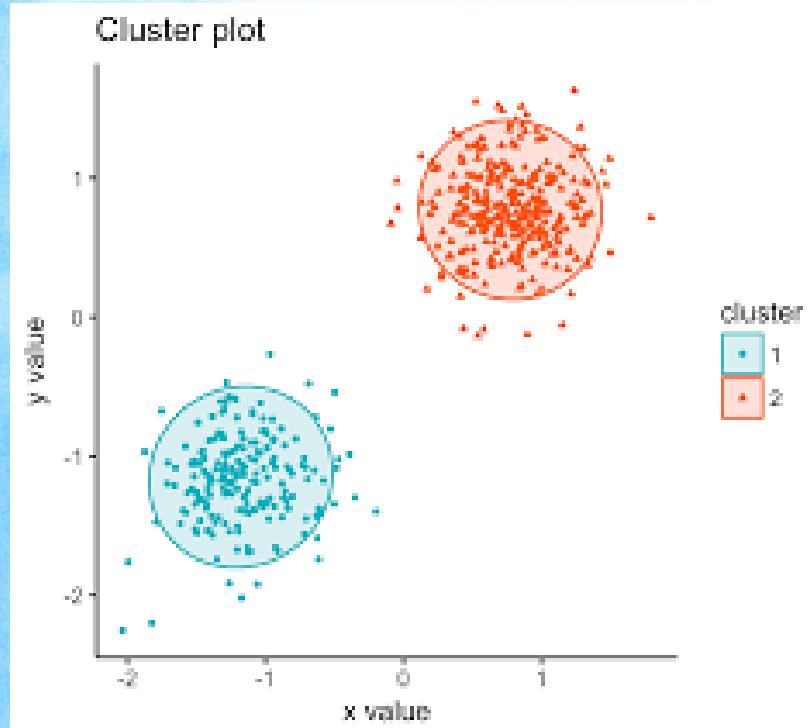
- ✓ Load the dataset
- ✓ Preprocess the dataset
- ✓ Initialize the number of clusters k and randomly choose k data points as the initial centroids.
- ✓ Assign each data point to the cluster with the nearest centroid based on the Euclidean distance.
- ✓ Recalculate the centroid of each cluster based on the mean of the data points in that cluster.
- ✓ Repeat steps 2 and 3 until convergence, i.e., the cluster assignments no longer change or a maximum number of iterations is reached.
- ✓ Output the final cluster assignments and centroids.

Hierarchical Clustering



- ✓ Load the dataset
- ✓ Preprocess the dataset
- ✓ Initialize each data point as its own cluster.
- ✓ Compute the distance matrix between each pair of clusters.
- ✓ Find the two closest clusters based on the distance matrix.
- ✓ Merge the two closest clusters into a single cluster.
- ✓ Update the distance matrix with the new distances between the merged cluster and the remaining clusters.
- ✓ Repeat steps 3-5 until all data points are in a single cluster.

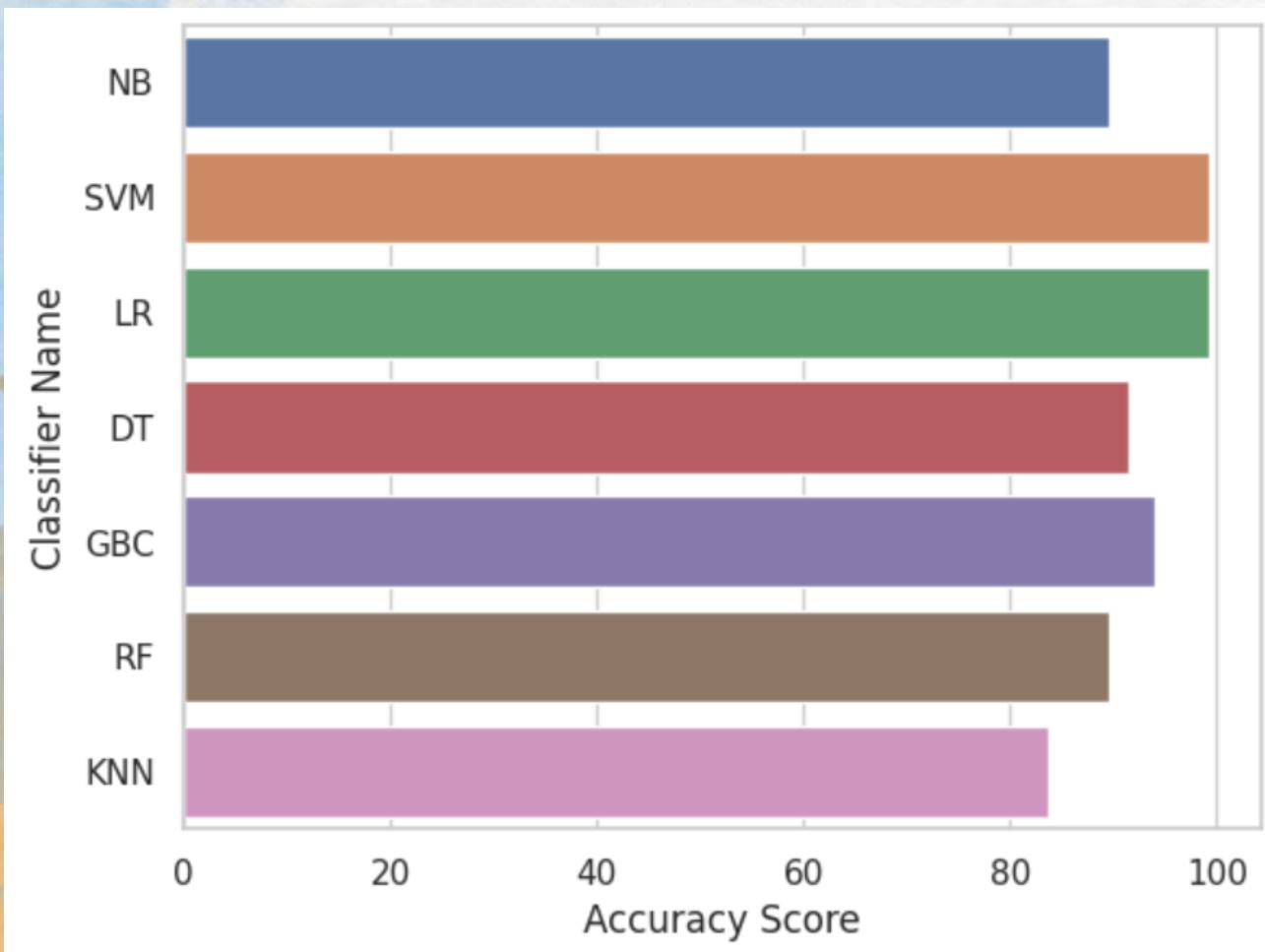
Partition Around medoids(PAM)



- ✓ Load the dataset
- ✓ Preprocess the dataset
- ✓ Initialize each data point as its own cluster.
- ✓ Compute the distance matrix between each pair of clusters.
- ✓ Find the two closest clusters based on the distance matrix.
- ✓ Merge the two closest clusters into a single cluster.
- ✓ Update the distance matrix with the new distances between the merged cluster and the remaining clusters.
- ✓ Repeat steps 3-5 until all data points are in a single cluster.

Results

This graph and the table shows result of my entire research using supervised models



	Classifier	Accuracy Score	Precision	Recall	F1-score
0	NB	89.66	88.84	71.67	76.62
1	SVM	99.37	98.95	98.8	98.88
2	LR	99.38	99.4	98.38	98.88
3	DT	91.54	84.94	84.68	84.81
4	GBC	94.04	94.02	83.55	87.71
5	RF	89.71	94.21	68.69	74.28
6	KNN	83.63	75.15	50.06	45.66

This table shows result of my entire research using Unsupervised models

Model Name	Accuracy
K-means clustering	96.91%
Hierarchical clustering	77.54%
partition Around Medoids	70%

Future Scopes

Suicidal sentiment analysis is an evolving field with many potential future applications. Some possible areas of development include:

- ✓ Improved accuracy
- ✓ Early intervention
- ✓ Going to apply audio, video and text analysis to identify depressed and suicidal people on the social network
- ✓ Integration with other technologies

References

- [1] Ayaan Haque, Viraaj Reddi, and Tyler Giallanza2(2021). Deep Learning for Suicide and Depression Identification with Unsupervised Label Correction. arXiv:2102.09427v2 [cs.LG]
- [2] Ankita & Nabizath Saleenaa (2018).An Ensemble Classification System for Twitter Sentiment Analysis .(ICCIDS 2018)
- [3] Shaoxiong Ji , Celina Ping Yu,Sai-fu Fung,Shirui Pan , and Guodong Long . Supervised Learning for Suicidal Ideation Detection in Online User Content(9 September 2018)
- [4] J. Parraga-Alava, R. A. Caicedo, J. M. Gómez and M. Inostroza-Ponta, "An Unsupervised Learning Approach for Automatically to Categorize Potential Suicide Messages in Social Media," 2019 38th International Conference of the Chilean Computer Science Society (SCCC), Concepcion, Chile, 2019, pp. 1-8, doi: 10.1109/SCCC49216.2019.8966443.
- [5] Orlando Iparraguirre-Villanuev & Victor Guevara-Ponce (2022). Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm . (*IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 6, 2022)
- [6] E. Rajesh Kumar, K.V.S.N. Rama Rao, Soumya Ranjan Nayak & Ramesh Chandra (2020) Suicidal ideation prediction in twitter data using machine learning techniques, *Journal of Interdisciplinary Mathematics*, 23:1, 117-125
- [7] Mandal S and Saha B. (2020). Forecasting Mental Disorders Through Aspect Identification from Social Media Posts. *Cyber Technologies and Emerging Sciences*. 10.1007/978-981-19-2538-2_13. (133-141).
- [8] Bhargava, Sandeep and Choudhary, Seema, Behavioral Analysis of Depressed Sentimental Over Twitter: Based on Supervised Machine Learning Approach (May 9, 2018). Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), 2018 held at Malaviya National Institute of Technology, Jaipur (India) on March 26-27, 2018
- [9] Marouane Birjali, Abderrahim Beni-Hssane, Mohammed Erritali. Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. (EUSPN 2017)



Thank You!