



Group-27

SOCIAL MEDIA COMMENT CLASSIFICATION AND CRIME FORECASTING



Team Members



DVS Sravan - AM.EN.U4CSE19117



Vidyacharan Varma - AM.EN.U4CSE19120



Ram Mohan Reddy S - AM.EN.U4CSE19150



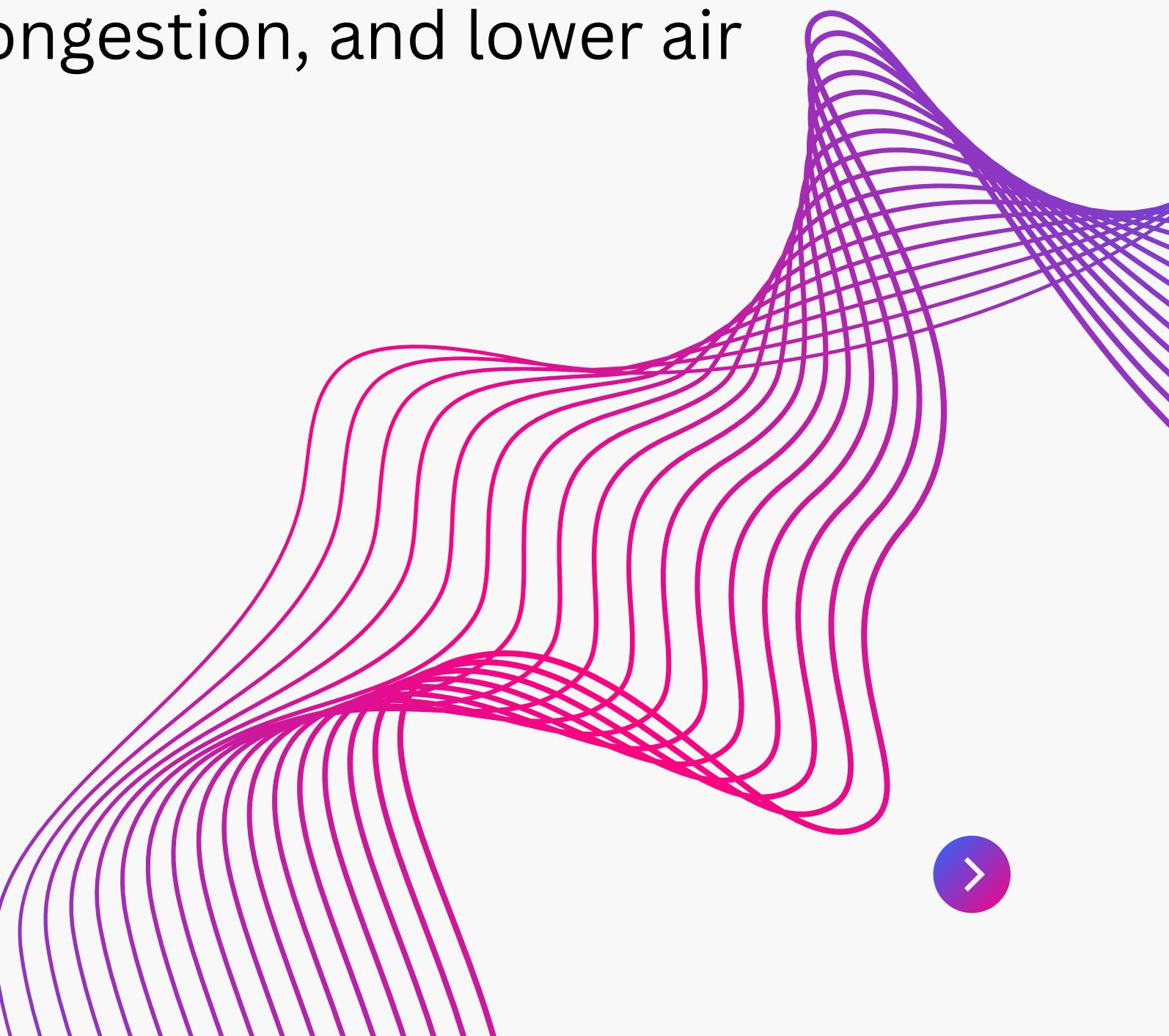
Sai Ganesh Chepuri - AM.EN.U4CSE19166

Mini Project Summary:

Carpooling is a form of transportation in which multiple people share a single vehicle to travel together to a common destination. This can help reduce the number of cars on the road, decrease traffic congestion, and lower air pollution.

Challenges and problems:

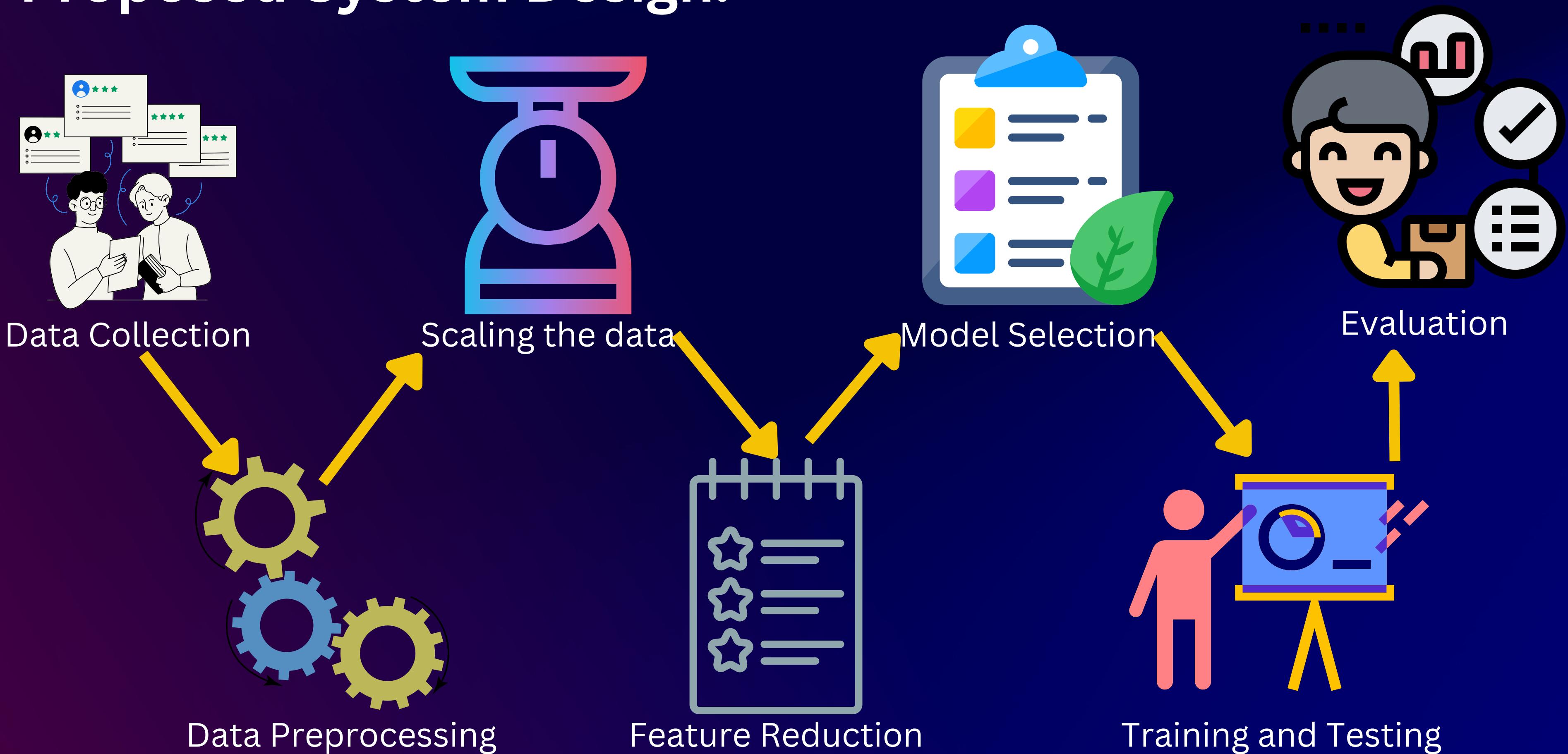
1. Scheduling Conflicts
2. Reliability
3. Safety and Privacy Concerns
4. Lack of flexibility
5. Limited Options



Social media comment classification and crime forecasting

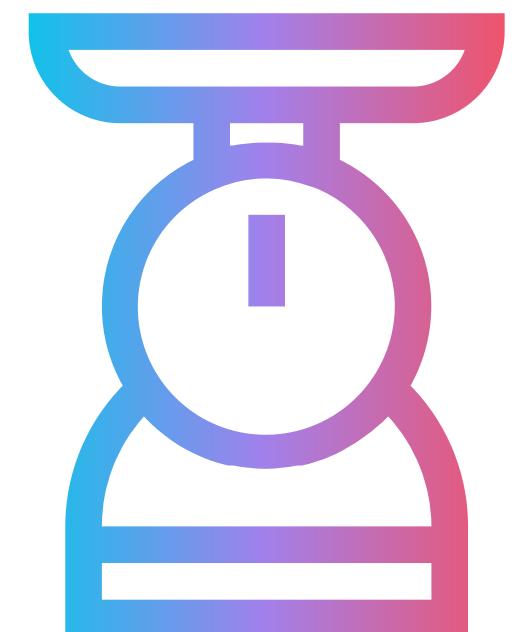
- Criminals often take advantage of social media users, leading to a variety of illegal activities and making cybercrime a significant global concern.
- In social media applications there is no restriction to talk abusive and humiliate people using text. With this benefit the criminals like Drug dealers, Terrorists, and normal users use opportunity to make threats to other people.

Proposed System Design:

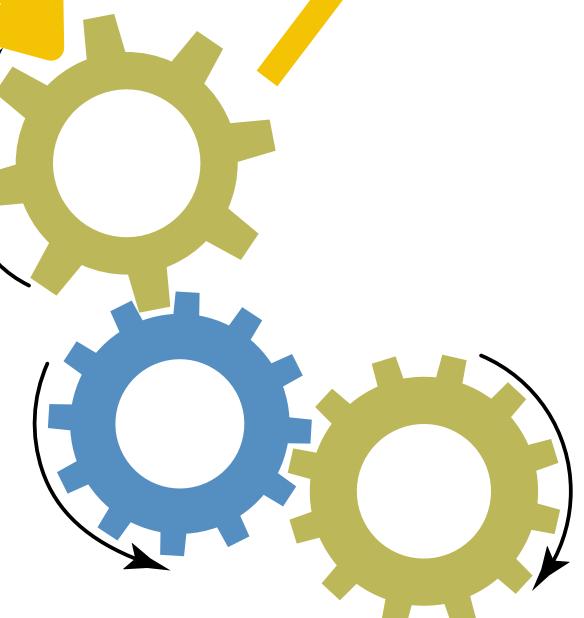




Data Collection



Scaling the data



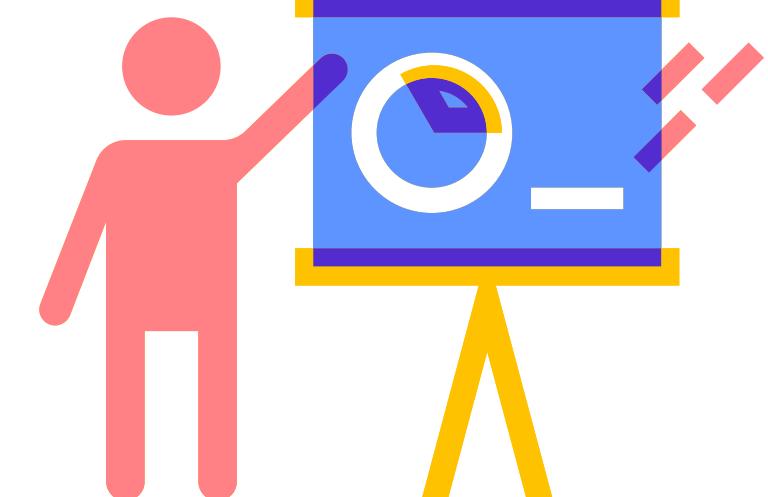
Data Preprocessing



Feature Reduction



Model Selection

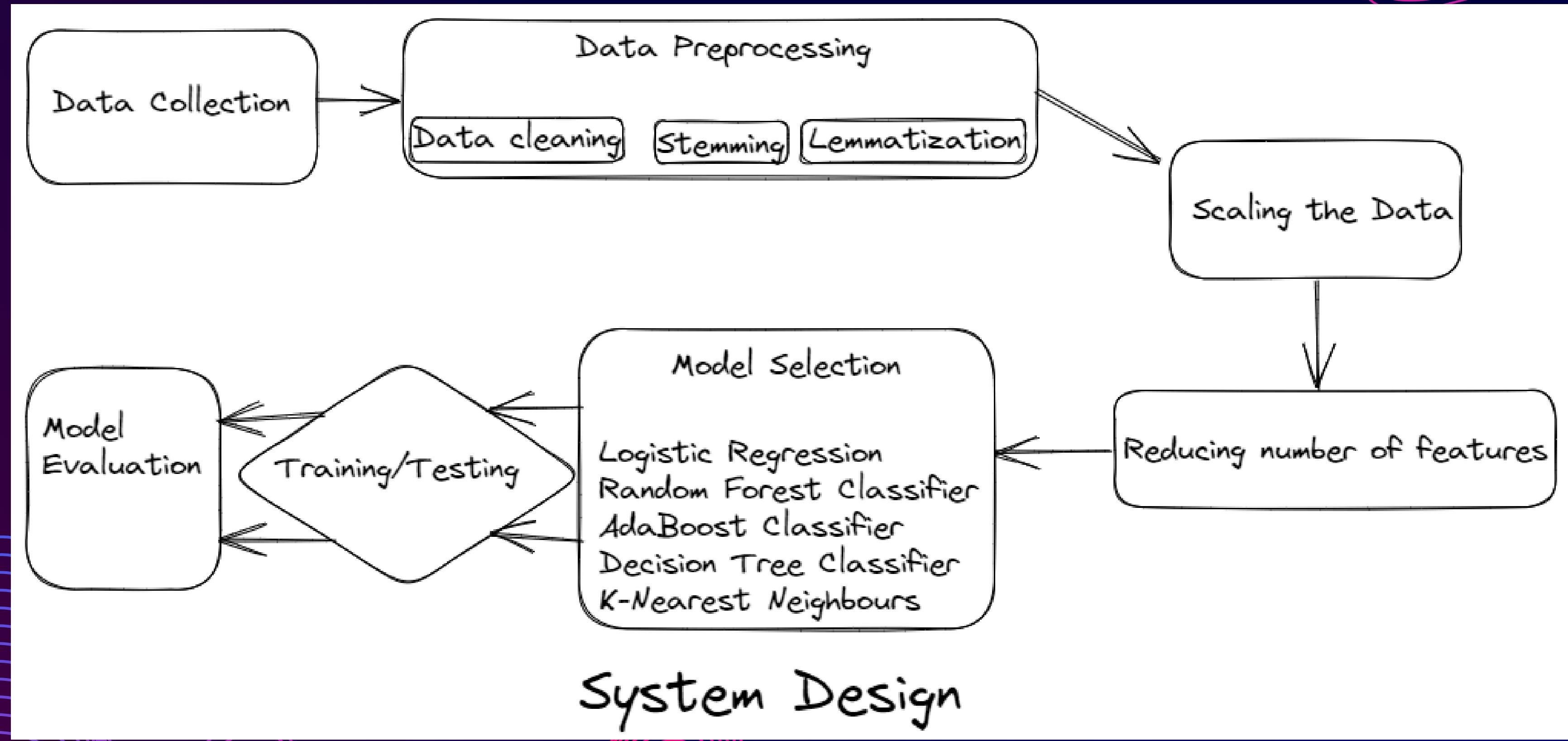


Training and Testing

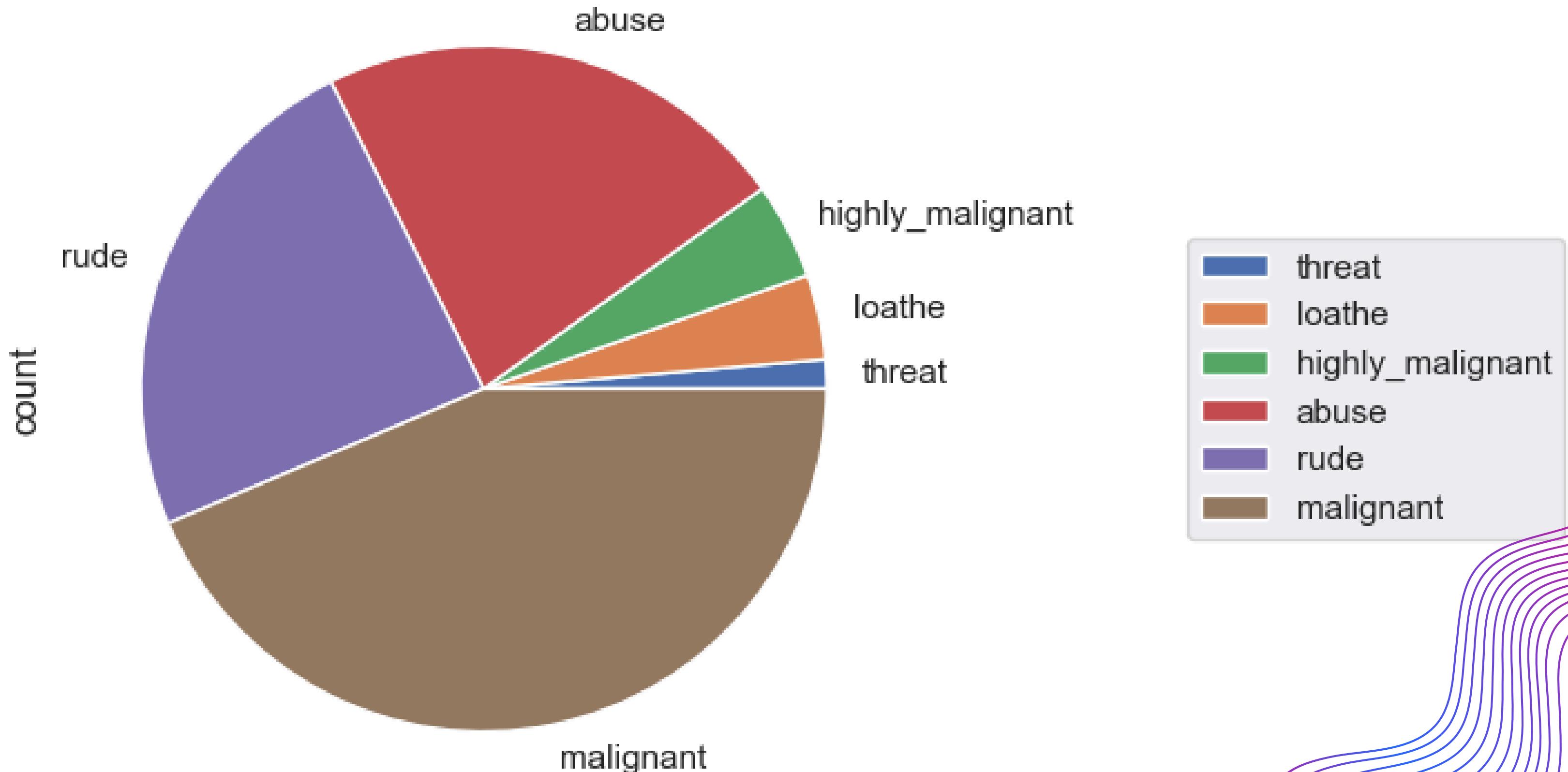


Evaluation

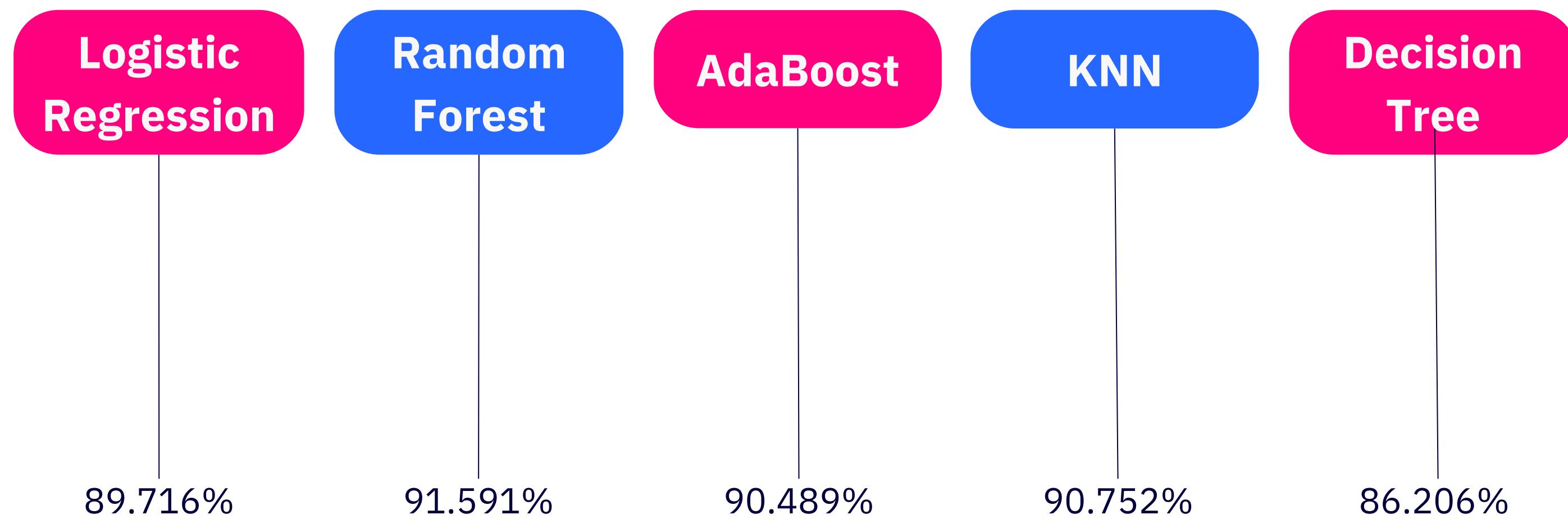
Proposed System Design:



Label distribution over comments



Algorithms



[Back to Agenda](#)



Logistic Regression:

- Logistic regression predicts binary outcomes.
- It models the relationship between predictor variables and the outcome.
- Effective use requires understanding of the data, preprocessing, and proper evaluation, as well as awareness of potential issues.

Decision Tree Classifier:

- Decision tree classifier is a popular machine learning algorithm for both classification and regression tasks. It creates a tree-like model of decisions and their possible consequences, based on the features of the data.
- It is easy to understand and interpret, making it useful for identifying important features and relationships in the data. It also allows for non-linear relationships between the features and the target variable.
- However, it is prone to overfitting, especially when the tree becomes too deep and complex. This can be mitigated through techniques such as pruning, limiting the depth of the tree, or using ensemble methods such as random forests.

Random Forest Classifier:

- Random forest classifier is a popular ensemble learning algorithm that combines multiple decision trees to improve the accuracy and robustness of the model.
- It creates a large number of decision trees and aggregates their predictions to make the final classification.
- Each tree is trained on a random subset of the data and a random subset of the features, which helps to reduce overfitting and increase generalization performance.
- It is known for its high accuracy, scalability, and ability to handle high-dimensional and noisy data.

AdaBoost Classifier:

- AdaBoost classifier is a popular boosting algorithm that combines multiple weak learners to create a strong classifier.
- AdaBoost classifier works by iteratively training a sequence of weak classifiers, each focusing on the samples that were misclassified by the previous classifiers.
- AdaBoost classifier is known for its high accuracy, ability to handle complex datasets, and ability to reduce overfitting. However, it can be sensitive to noisy data, and may be slower and more computationally intensive than other algorithms.

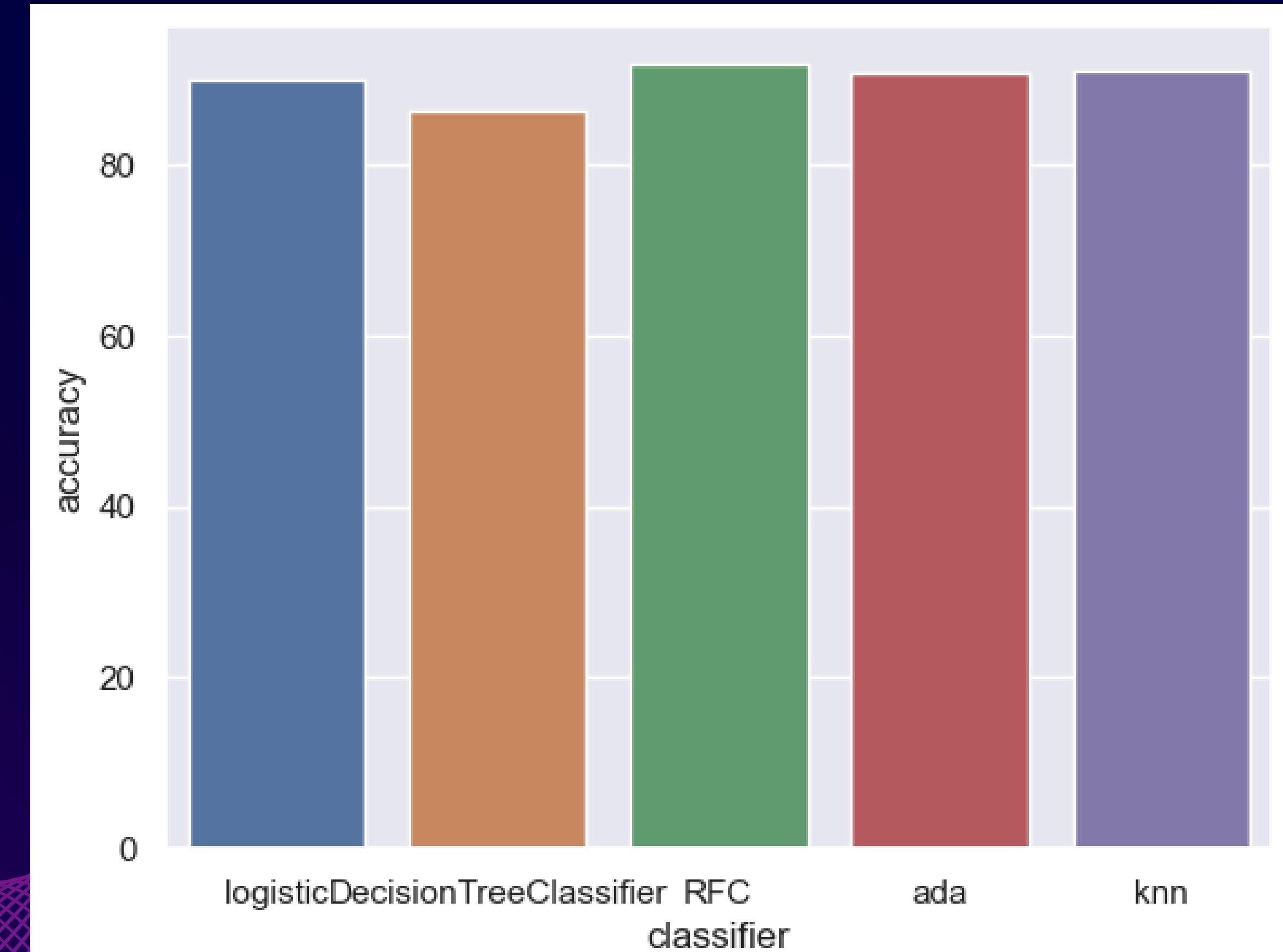
XGBoost Classifier:

- XGBoost is a popular machine learning algorithm that uses gradient boosting to improve the performance of decision trees.
- One of the key strengths of XGBoost is its scalability, which allows it to handle large datasets with millions of examples and thousands of features.
- XGBoost also includes several regularization techniques, such as L1 and L2 regularization, to prevent overfitting and improve the generalization performance of the model.

K-Nearest Neighbours:

- KNN is a simple and popular machine learning algorithm for both classification and regression tasks.
- KNN is a non-parametric algorithm, which means that it doesn't make any assumptions about the underlying distribution of the data. This makes it useful for handling non-linear relationships and complex data structures.
- It requires careful selection of the number of neighbors (k) and the distance metric used to measure similarity between data points.

Accuracy Scores for different algorithms:



Status of Project Completion:

Almost 95% of the coding part is done.

Some modifications may need to be done as part of increasing performance.

ALMOST
THERE!

Results and Discussion:

- The results showed that the system achieved high accuracy levels for all tested datasets, which suggests that it can be effectively used to classify large volumes of text data in various domains.
- Among the taken algorithms, Random Forest Classifier performs comparatively best.

Thank
you

