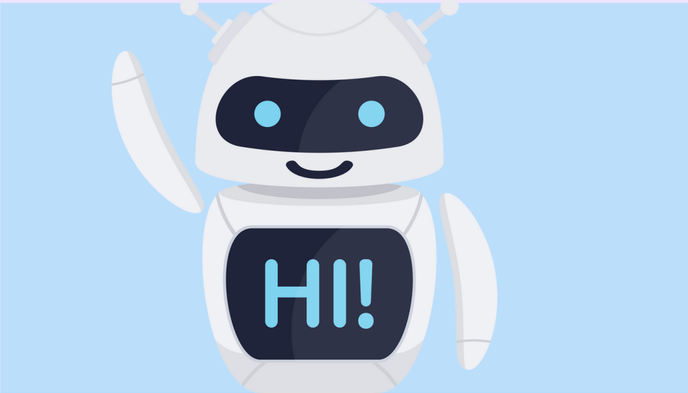
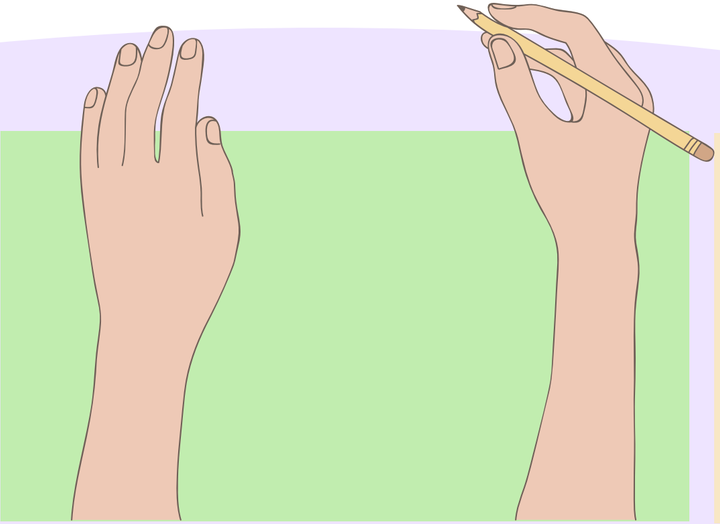
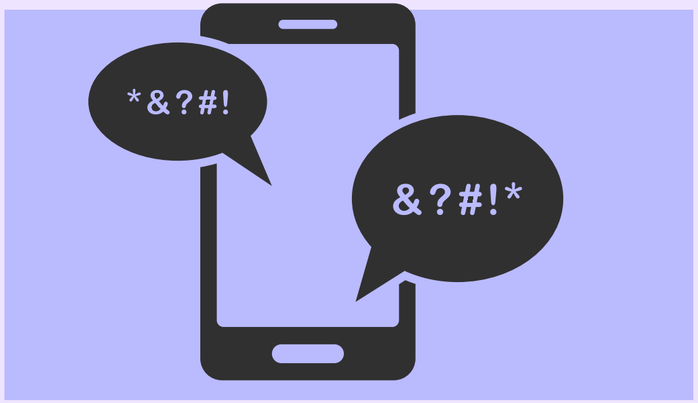
HATE SPEECH DETECTION CHATBOT



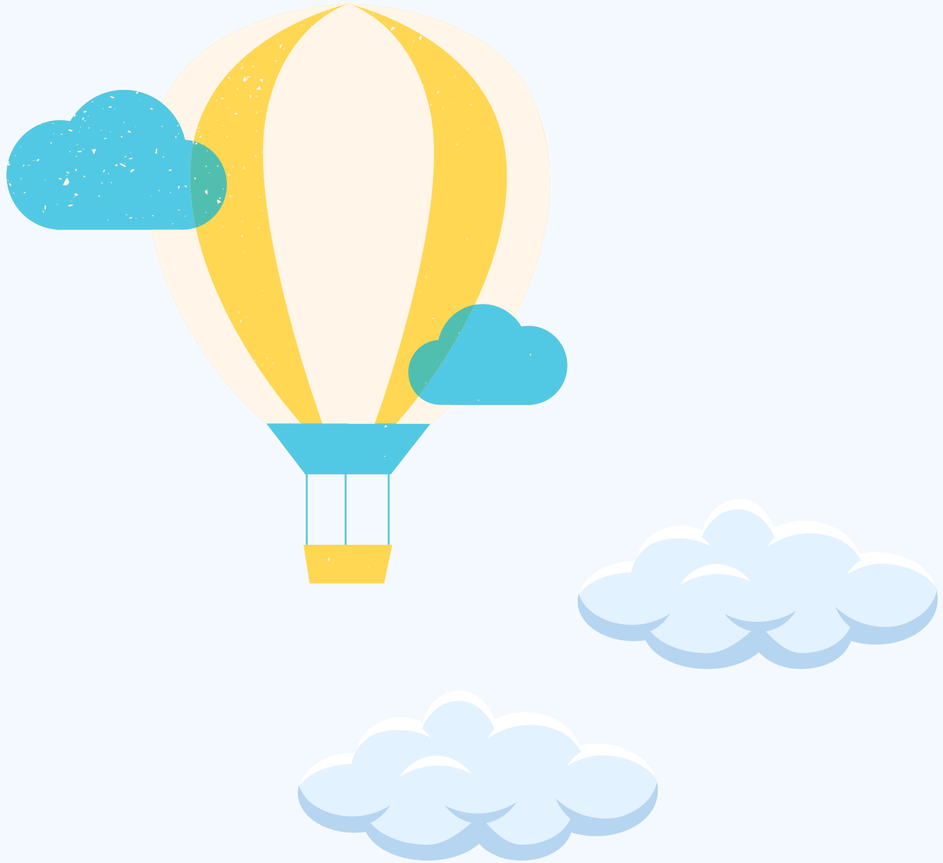




created by Team 8

Team Members

- Anoushka Srivastava
- Anish Borkar
- Ankur Kaushal
- Dyuti Dasmahapatra
- D Veera Harsha Vardhan Reddy
- Godavarthi Sai Nikhil
- Himanshu Sharma
- Nichenametla Karthik Raja
- Vandamasu Sai Sumanth



TODAY'S AGENDA

- DATA FLOW DIAGRAMS
- STORYBOARDS
- LOW FIDELITY PROTOTYPE
- BASIC WORKING SOLUTION
- ML (CRISP - DM)

1. DATA FLOW DIAGRAM

DATA FLOW DIAGRAM

Click on the link or scan the QR Code

https://colab.research.google.com/drive/1bP0kE7F5EYVAQNp5bZZ6rGUeBztU2zv4?usp=sharing

2. STORYBOARDS

Sam's Storyboard

Scenario: "Sam Uses App to Respond to Negative Feedback on New Product"

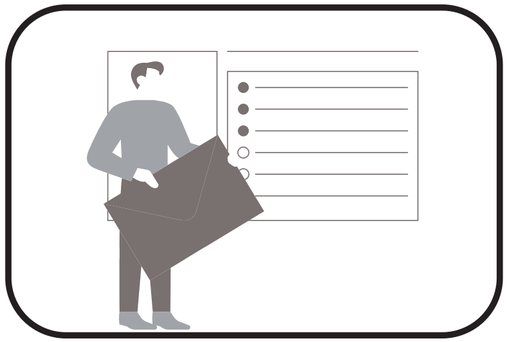
Sam launched a new product recently.

Sam received feedback from users after a certain number of days.

Sam received negative feedback, including hate comments from some users.

Sam initially felt upset and disappointed with the negative feedback, but he wanted to reply without causing offense.

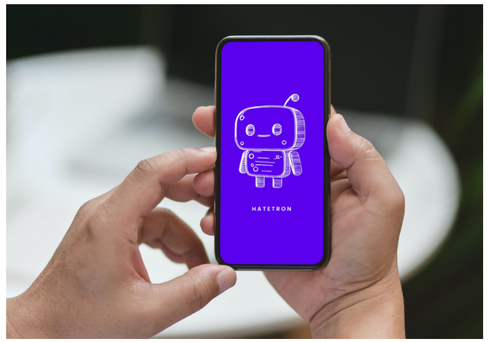
Sam checked for hate speech using an app, and edited his reply before sending to the user.



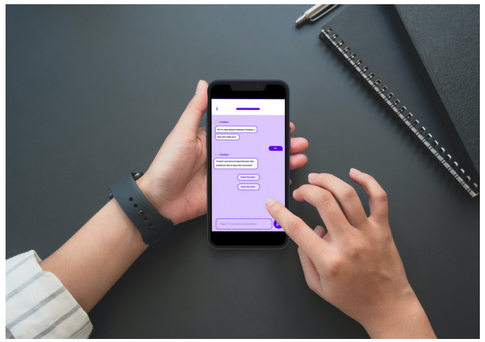
After using the app, he sent improved emails to his customers, showing his commitment to providing excellent customer service.

Storyboard on Sam's experience

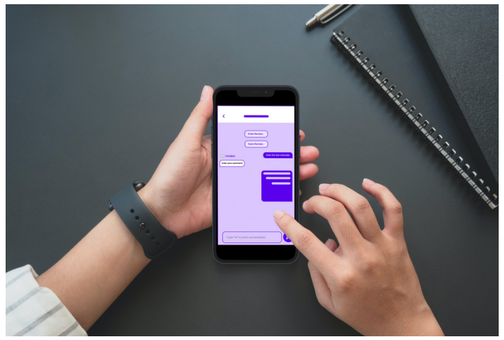
Scenario: "Sam Uses App to Respond to Negative Feedback on New Product"



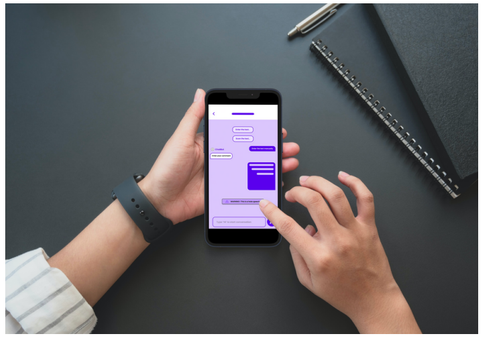
Sam opens the app and registers his account



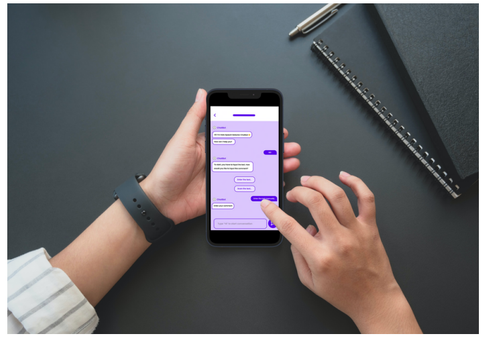
He selects "Enter the text Manually" from the options.



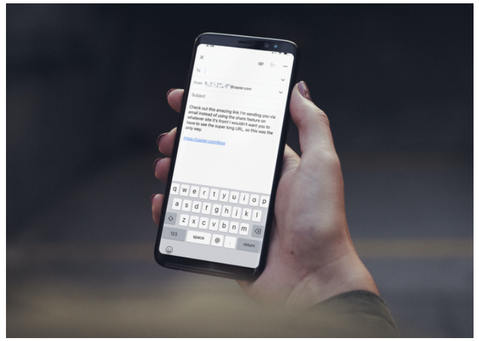
Once the 'text' option is selected, he enters the message into the app.



He received responses from the app, which helped him identify any inappropriate language in the text he uploaded.



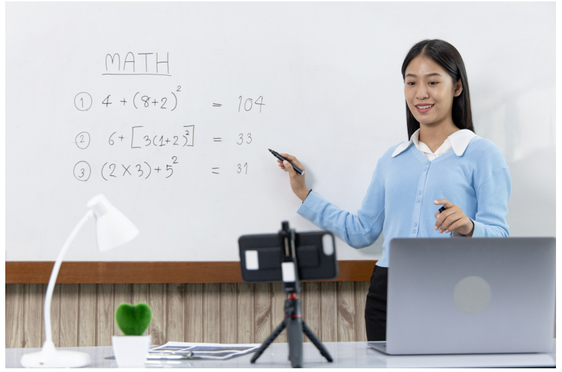
He used the app to avoid hate speech or offensive language in response.



After using the app, he sent improved emails to his customer.

Ms.Brown's Storyboard

Scenario: "Ms. Brown Implements Hate Speech Detection Chatbot for Safe Online Classroom Discussions"



Ms. Brown is an English teacher who is concerned about the use of hate speech in her online classroom discussions.

She adds a hate speech detection chatbot to her online classroom to ensure a safe and respectful learning environment.

She uploads a screenshot of the comments and clicks on the 'scan the text' option.

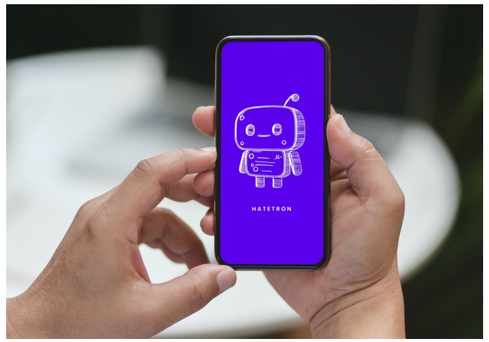
It scans the comments and discussion threads for any instances of hate or offensive speech

If the chatbot detects any harmful content, it gives a warning to review and address.

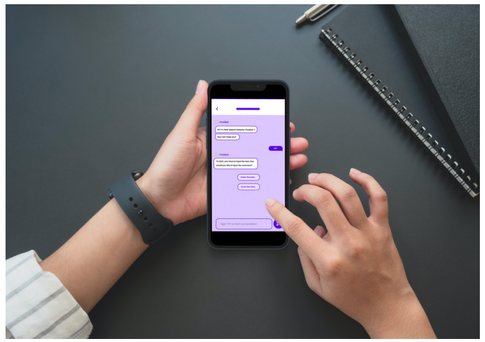
She uses it to educate on hate speech, provide feedback, and promote a positive environment.

Storyboard on Ms. Brown's experience

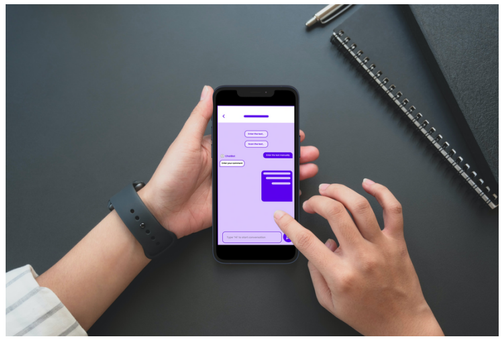
Scenario: "ms. brown Uses App to Respond to Negative Feedback on her online classroom discussions"



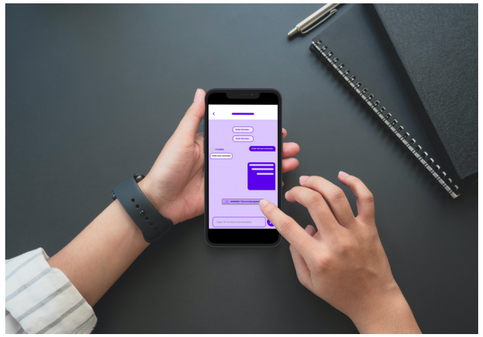
Ms.Brown opens the app and registers her account



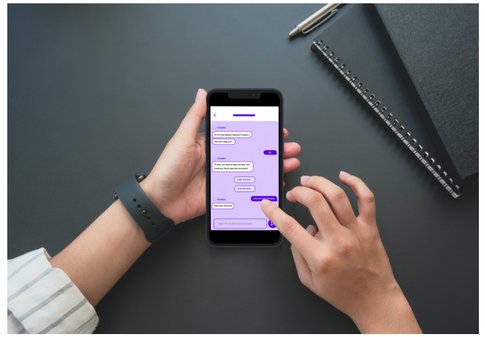
She selects "Scan the text" from the options.



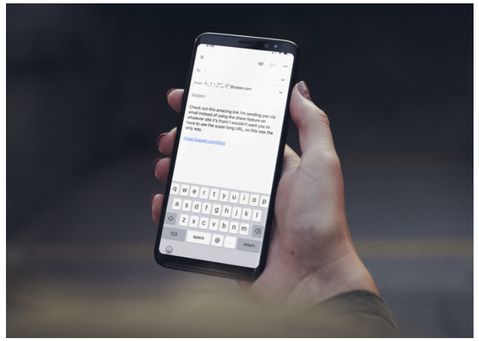
Once the option is selected, she scans the comments into the app.



It scans the comments for any instances of hate or offensive speech



The chatbot gives a warning regarding the hate / offensive comments



Ms. Brown uses the chatbot to provide immediate feedback to students and asks them to reconsider their words.

Yaseen's Storyboard

Scenario: "Mrs. Yaseen uses Hate Speech Detection Chatbot To Analyze The Comments on Her Youtube Channel "

Mrs. Yaseen, a housewife, has recently started her Youtube vlogging channel.

She can't read properly so she wants to understand if the comments she is getting are hateful or positive.

She decides to use the app and since she can't type them properly , she uploads a screenshot of the comments.

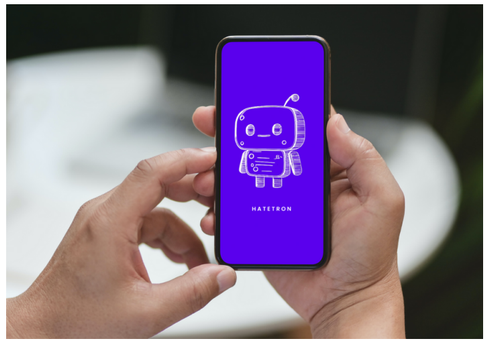
The app scans the comments to know if they are hateful/ offensive or positive

If the chatbot detects any harmful content, it gives a warning to review and address.

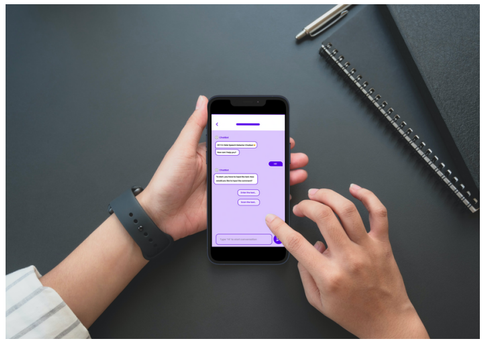
She uses this comment feedback to modify her channel content accordingly & also to give apt. reply to her haters.

Storyboard on Yaseen experience

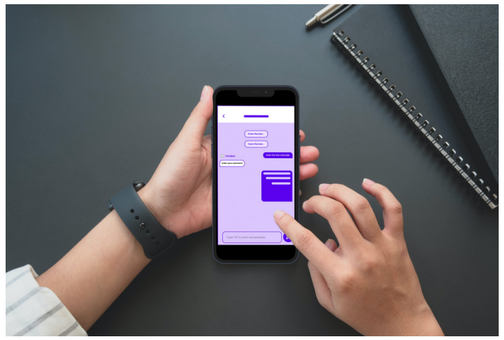
Scenario: "Mrs. Yaseen uses Hate Speech Detection Chatbot To Analyze The Comments on Her Youtube Channel



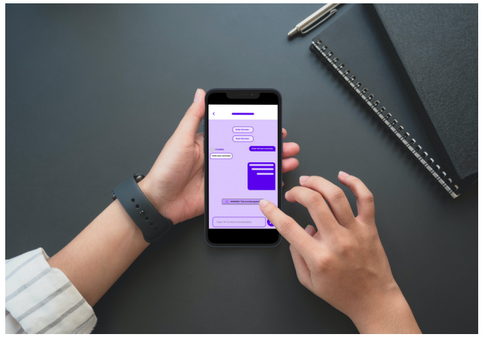
Mrs. Yaseen opens the app and signs in her account



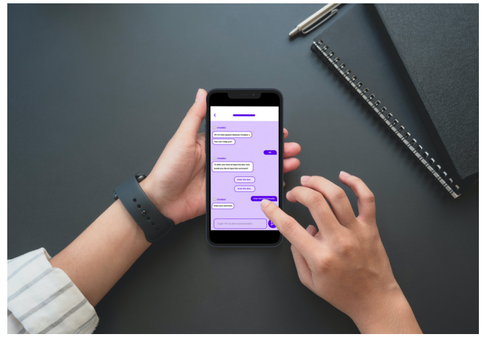
She goes through the 'demo' to understand the working of the app



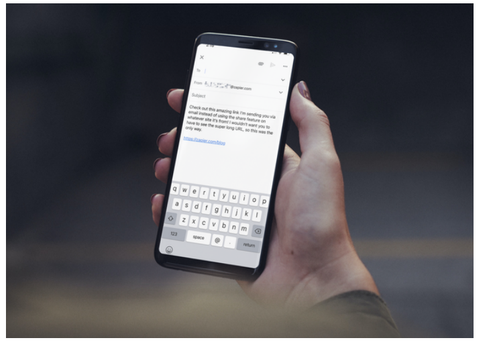
She selects "Scan the text" from the options.



Once the option is selected, she uploads the comments' pictures into the app.



The chatbot gives a warning regarding the hate / offensive comments



She uses this comment feedback to modify her channel content accordingly & also to give apt. reply to her haters.

3. LOW FIDELITY PROTOTYPE

LOW FIDELITY PROTOTYE

Click on the link or scan the QR Code

https://www.figma.com/file/U7qlExcJ8Qk4MybpAAxsRQ/Project-2-Figma?node-id=0%3A1&t=tiNJQGEKxTDCD8sj-1

4. BASIC WORKING SOLUTION

projet_2

5. ML (CRISP-DM)

CRISP-DM

Deployment

- UAT
- Automation

Business Understanding

- Project Objective
- Business Objective

Evaluation

- Blind Test
- Evaluation Metrics

Data

Data Understanding

- Data Exploration
- Data Selection

Modeling

- Predicting hate speech using ml models

Data Preparation

- Feature Engineering

BUSINESS UNDERSTANDTING

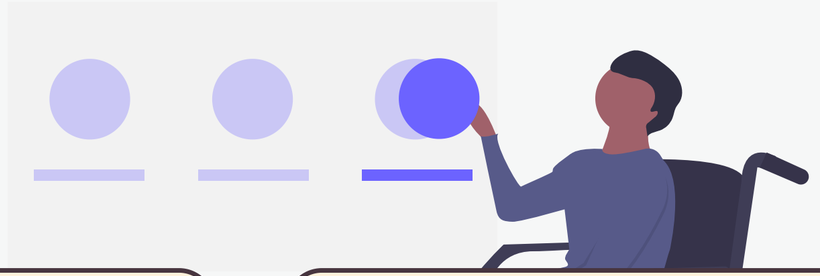
PROBLEM STATEMENT

- Hate speech and offensive language are common on social media platforms, hindering constructive conversations.
- The aim is to develop a real-time chatbot to detect hate speech.
- The chatbot should flag messages containing offensive language, slurs, or discriminatory remarks.
- The goal is to improve online conversations by reducing hate speech and promoting respect.

TARGET AUDIENCE

- Hate speech detection targets those who want to prevent it on their online platforms.
- The audience includes social media companies, news outlets, schools, government agencies, and non-profits.
- It is also relevant for influencers, students, and teachers.
- The goal is to monitor and prevent hate speech in online spaces.

DATA UNDERSTANDING



Social Media

- Social media platforms such as Twitter, Facebook, and Instagram are rich sources of text data containing instances of hate speech.

Human - Labeled

- Text data that involves having people read through texts and mark them as either containing hate speech or not.

User-generated

- User-generated content, such as comments on articles or blog posts, can also be used to train.

DATA PREPARATION

Click on the link or scan the QR Code

https://colab.research.google.com/drive/1E2hehc9oUEmjjh-j8WU3uEy51Q7J9bmZ?usp=sharing

DATA MODELING

Logistic Regression

- Statistical method used to analyze and model relationships between a binary response variable and one or more predictor variables.
- Used for predicting the probability of a certain outcome.
- The outcome variable in logistic regression is binary

Random Forest

- A popular ensemble learning algorithm used in machine learning predictor variables.
- Builds multiple decision trees using different random subsets of the input features and data.
- Algorithm aggregates the predictions of all the decision trees in the forest.

Support vector machines (SVMs)

- Popular supervised learning algorithm used in machine learning for classification and regression tasks.
- Useful for solving complex classification problems with non-linear decision boundaries.
- A set of training examples are used to construct a hyperplane that separates different classes of data.

Recurrent Neural Network (RNN)

- A type of artificial neural network designed to process sequential data, such as speech, text, and time series data.
- Used in a wide range of applications, such as speech recognition, natural language processing, and image captioning.

DATA EVALUATION

Split the dataset into training and test sets:

- Split the dataset into training and test sets.
- The training set will be used to train the chatbot, and the test set will be used to evaluate its performance.

Choose evaluation metrics:

- Select appropriate metrics to evaluate the performance of the chatbot.
- Metrics commonly used
  1. precision
  2. recall
  3. F1 score
  4. accuracy

Training & Testing the chatbot:

- Use the training set to train the chatbot on hate speech detection.
- Use the test set to evaluate the performance of the chatbot.

Evaluate the results:

- Calculate the metrics selected earlier to determine the performance of the chatbot.
- Analyse the results to identify areas for improvement and make necessary adjustments to the chatbot.

Iterate:

- Repeat the training, testing, and evaluation process until the chatbot's performance meets the desired level of accuracy and precision.

DEPLOYEMENT

- Deploy the model as a chatbot, and integrate it with the messaging platform.
- Test the chatbot in a live environment to ensure that it meets the desired performance standards.
- Depending on the deployment platform, there may be certain restrictions that need to be met.
- For example, if deploying as a mobile app, we may need to follow the guidelines set by the app store, such as size limits and content guidelines.
- Additionally, if the application is intended for enterprise use, there may be certain security and compliance requirements that need to be met.

THAT'S A WRAP!

THANK YOU FOR YOUR TIME!