TOXIC COMMENT CLASSIFICATION FOR SOCIAL MEDIA

Milestone 1: Project Initialization and Planning Phase

The Project Initialization and Planning Phase for Toxic Comment Classification on Social Media involves defining objectives, gathering diverse labeled datasets, determining technical and functional requirements, assigning team roles, setting timelines for phases like data collection, model development, and deployment, and addressing risks such as data bias and performance challenges, aiming to deliver a trained model integrated into a web or API interface for accurate and efficient toxic comment moderation**.**

Activity 1: Define Problem Statement

Problem Statement :. The problem statement for Toxic Comment Classification on Social Media is to develop an automated system capable of identifying and categorizing toxic comments, such as hate speech, threats, and harassment, to enhance moderation efficiency and foster a safer online community environment.

Toxic Comment Classification for Social Media Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

Project Proposal: The proposed solution for Toxic Comment Classification on Social Media involves leveraging machine learning and natural language processing techniques to build a robust model that detects and categorizes toxic comments, integrates seamlessly with social media platforms via APIs or web interfaces, and provides real-time insights to support automated content moderation and promote a positive online environment.

Toxic Comment Classification for Social Media Report : Click Here

Activity 3: Initial Project Planning

The initial project planning for Toxic Comment Classification on Social Media includes defining objectives, gathering labeled datasets, identifying technical and functional requirements, outlining team roles and responsibilities, creating a phased timeline for data collection, model development, and deployment, and addressing potential risks such as data bias and ethical concerns to ensure project success.

Toxic Comment Classification for Social Media Report : Click Here

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase for Toxic Comment Classification on Social Media involves gathering diverse, labeled datasets, cleaning and normalizing text, handling imbalances, removing noise, tokenizing, and preparing the data for training a machine learning model to accurately identify and categorize toxic comments.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The Data Collection Plan for Toxic Comment Classification on Social Media involves identifying raw data sources such as Kaggle's "Toxic Comment Classification" dataset, public APIs, and platform-specific comment data, ensuring data diversity across languages and contexts, and generating a Data Quality Report to assess completeness, consistency, and balance for effective model training.

Toxic Comment Classification for Social Media Report: Click Here

Activity 2: Data Quality Report

The Data Quality Report for Toxic Comment Classification on Social Media evaluates the collected data for completeness, consistency, accuracy, and balance, ensuring that the dataset contains sufficient labeled examples of various toxic comment types and non-toxic comments to support reliable and unbiased model training.

Toxic Comment Classification for Social Media Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data exploration and preprocessing for Toxic Comment Classification on Social Media involve analyzing the dataset to understand patterns, distributions, and imbalances, followed by cleaning text data, handling missing values, tokenization, removing stopwords, and normalizing text to prepare it for effective model training.

Toxic Comment Classification for Social Media Report: Click Here

Milestone 3: Model Development Phase

The Model Development Phase for Toxic Comment Classification on Social Media involves selecting appropriate machine learning algorithms, such as logistic regression, support vector machines, or deep learning models like LSTM or BERT, for training the model. This phase includes feature engineering, such as text vectorization using techniques like TF-IDF or word embeddings, followed by model training on the preprocessed data. Hyperparameter tuning is performed to optimize the model's performance, and techniques like cross-validation are used to evaluate model robustness. The goal is to develop a model that accurately classifies toxic comments while minimizing false positives and negatives.

The Feature Selection Report for Toxic Comment Classification on Social Media outlines the process of selecting the most relevant features from the dataset to improve model accuracy and reduce computational complexity. This includes evaluating various text features such as word frequency, n-grams, sentiment scores, and context-based embeddings. Feature selection methods, such as mutual information, chi-square tests, or recursive feature elimination, are applied to identify key indicators of toxicity, while irrelevant or redundant features are discarded. The report also assesses the impact of different feature sets on model performance, aiming to enhance prediction accuracy and efficiency.

Activity 1: Model Selection Report

The Model Selection Report for Toxic Comment Classification on Social Media details the evaluation and selection of the most suitable machine learning models for the task. Various algorithms, such as logistic regression, support vector machines (SVM), random forests, and deep learning models like LSTM or BERT, are compared based on performance metrics like accuracy, precision, recall, and F1-score. The report includes a rationale for model selection, considering factors such as dataset size, computational resources, interpretability, and real-time prediction requirements. The chosen model is validated using techniques like cross-validation and fine-tuned to ensure optimal classification of toxic and non-toxic comments.

Toxic Comment Classification for Social Media Report : Click Here

Activity 2: Initial Model Training Code, Model Validation and Evaluation Report:

The Initial Model Training Code for Toxic Comment Classification on Social Media involves the implementation of selected machine learning algorithms, such as Logistic Regression, SVM, or deep learning models like BERT, using preprocessed text data. This code covers the steps of data loading, feature extraction (e.g., TF-IDF or word embeddings), model training, and hyperparameter tuning. The Model Validation and Evaluation Report details the process of validating the model using techniques like cross-validation and a separate test set, along with evaluating performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The report includes analysis of overfitting or underfitting, compares model performance across different algorithms, and provides insights into areas for improvement, such as tuning or incorporating additional features for enhanced classification of toxic comments.

Toxic Comment Classification for Social Media Report: Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase for Toxic Comment Classification on Social Media involves refining the initial model to enhance its performance. This phase includes techniques such as hyperparameter tuning (e.g., adjusting learning rates, batch sizes, and regularization), optimizing feature selection, and employing advanced strategies like grid search or random search to find the best combination of parameters. Additionally, methods like cross-validation and early stopping are used to prevent overfitting and improve generalization. The goal is to maximize the model's accuracy, precision, recall, and F1-score while ensuring it performs efficiently for real-time classification of toxic comments across diverse social media platforms.

Activity 1: Hyperparameter Tuning Documentation

The Hyperparameter Tuning Documentation for Toxic Comment Classification on Social Media provides a comprehensive guide to the process of optimizing model performance by adjusting key hyperparameters. It includes a detailed list of hyperparameters considered, such as learning rate, batch size, number of layers, dropout rate, and regularization strength, along with the ranges or values explored during the tuning process. The documentation outlines the methods used for hyperparameter search, such as grid search, random search, or Bayesian optimization, and discusses the evaluation metrics (accuracy, precision, recall, F1-score) used to assess the performance of each configuration. Additionally, it includes insights on how hyperparameter choices impact model efficiency and accuracy, as well as recommendations for the best-performing parameter settings that improve the classification of toxic comments.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report for Toxic Comment Classification on Social Media compares the effectiveness of different machine learning models based on key performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The report includes a detailed analysis of how each model performs in detecting toxic comments, highlighting strengths and weaknesses, and providing insights into the trade-offs between false positives and false negatives. It also discusses model stability, overfitting or underfitting tendencies, and suggests optimizations based on the comparison results to enhance classification accuracy and robustness for real-world deployment.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification for Toxic Comment Classification on Social Media explains the rationale behind choosing the optimal model after thorough evaluation and comparison of various algorithms. Based on performance metrics like accuracy, precision, recall, and F1-score, the selected model, such as BERT or a fine-tuned deep learning model, offers the best balance of accuracy and generalization for identifying toxic comments. The justification also considers factors like model scalability, real-time classification requirements, and computational efficiency, ensuring the chosen model meets the project's goals of robust, accurate, and timely detection of toxic content.

Toxic Comment Classification for Social Media Report: Click Here

Milestone 5: Project Files Submission and Documentation

For project file submission in , Kindly click the link and refer to the flow.

Toxic Comment Classification for Social Media Report: Click Here

Milestone 6: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.