Project Initialization and Planning Phase

|  |  |
| --- | --- |
| Date | 15 AUGUST 2024 |
| Team ID | LTVIP2024TMID24818 |
| Project Title | Toxic Comments Classification For Social Media |
| Maximum Marks | 3 Marks |

**Project Proposal (Proposed Solution) report:**

Social media platforms are increasingly becoming spaces for both positive interactions and toxic behaviors. Toxic comments, such as harassment, hate speech, and personal attacks, can harm individuals and communities, negatively impacting user experience and overall platform engagement. This project aims to develop a Toxic Comments Classification System that can automatically detect and filter toxic comments to foster healthier conversations online.

|  |  |
| --- | --- |
| **Project Overview** | |
| Objective | Build a machine learning model capable of identifying toxic comments across different categories such as hate speech, insults, profanity, threats, etc.  Ensure the model can work in real-time or near real-time to classify incoming comments on social media platforms. |
| Scope | Develop a machine learning model to classify toxic comments on social media platforms. The model will detect various types of toxicity, including hate speech, threats, and harassment. It will process real-time user comments and provide accurate classifications, ensuring harmful content is flagged. The solution will be scalable, multilingual, and adaptable across different platforms. |
| **Problem Statement** | |
| Description | Toxic Comments Classification for Social Media aims to automatically detect and filter harmful comments like hate speech, insults, and threats. The project involves training a machine learning model on labeled datasets to classify comments as toxic or non-toxic. It helps maintain a healthy online environment by preventing the spread of harmful content. The system can be deployed across social media platforms to ensure safer interactions. |
| Impact | The Toxic Comments Classification project aims to enhance user experience on social media platforms by identifying and filtering harmful content, fostering a safer online environment. It supports mental well-being by reducing exposure to negativity and harassment, promoting healthier interactions among usersAdditionally, it empowers platform moderators and users with tools to effectively manage toxic behavior, ultimately contributing to a more inclusive and respectful digital community. |
| **Proposed Solution** | |
| Approach | The approach involves data collection, preprocessing, feature extraction, model training, and evaluation using datasets such as Kaggle's Toxic Comment Classification Challenge. The classification model will be deployed as a web application, enabling users to input comments and receive toxicity classifications with explanations. |
| Key Features | Automated Comment Classification, Toxicity Scoring (0-100%), and Categorization .The system utilizes Natural Language Processing (NLP) and Machine Learning (ML) algorithms for accurate classification. |

|  |  |
| --- | --- |
|  | * Real-time decision-making for quicker loan approvals. * Continuous learning to adapt to evolving financial landscapes. |

**Resource Requirements**

|  |  |  |
| --- | --- | --- |
| **Resource Type** | **Description** | **Specification/Allocation** |
| **Hardware** | | |
| Computing Resources | CPU/GPU specifications, number of cores | T4 GPU |
| Memory | RAM specifications | 8 GB |
| Storage | Disk space for data, models, and logs | 1 TB SSD |
| **Software** | | |
| Frameworks | Python frameworks | Flask |
| Libraries | Additional libraries | scikit-learn, pandas, numpy, matplotlib, seaborn |
| Development Environment | IDE | Google colab Notebook, vscode |
| **Data** | | |
| Data | Source, size, format | Kaggle dataset, 4269, csv |