Connected World.
Connected Experiences.

Tech Mahindra

SOCIAL CHAT TOXICITY ANALYZER

Tech Mahindra's social chat toxicity analyzer is an artificial intelligence (AI) based text classification system to identify toxicity and inappropriate content in chat messages in real time.

Many companies offer chat modules to all their users. Toxicity in terms of abuse, racial slurs, and cyberbullying is rampant and is a cause for concern while users use the chat option. Due to the nature of the domains, for example, gaming, the level of violence and nudity has a higher threshold than a general social domain. There is a need for a custom toxicity chat analyzer that is suited to such industries and provides a scoring mechanism for defaulters.

OUR SOLUTION

Our AI-based text classification engine and user rating system that masks or cleans inappropriate data, flags content that is racial, and has high toxic content. Response against the responsible party based on a points system, thus providing a more proactive approach to the current reactive approach. The solution is scalable and has low latency to ensure that users are not affected in terms of performance. It is configurable to accommodate various regions and languages.

Features of the Solution

- **Chat message clean up**: System to extract the various components of a chat message and clean up unwanted data
- **TOXDET**: AI-based system with localization settings to classify and identify toxic sentences and words. Mask inappropriate content
- **User Rating**: User rating system based on behaviour and page ranking. Enable a penalty system for frequent inappropriate behavior

Solution Technology

- Combination of regular expression dictionary-based and deep learning NLP based technologies to monitor toxicity in text, images, and videos (GIF)
- Tensor flow JS lightweight models that can be deployed in client systems to provide near real-time analysis

KEY CHALLENGES

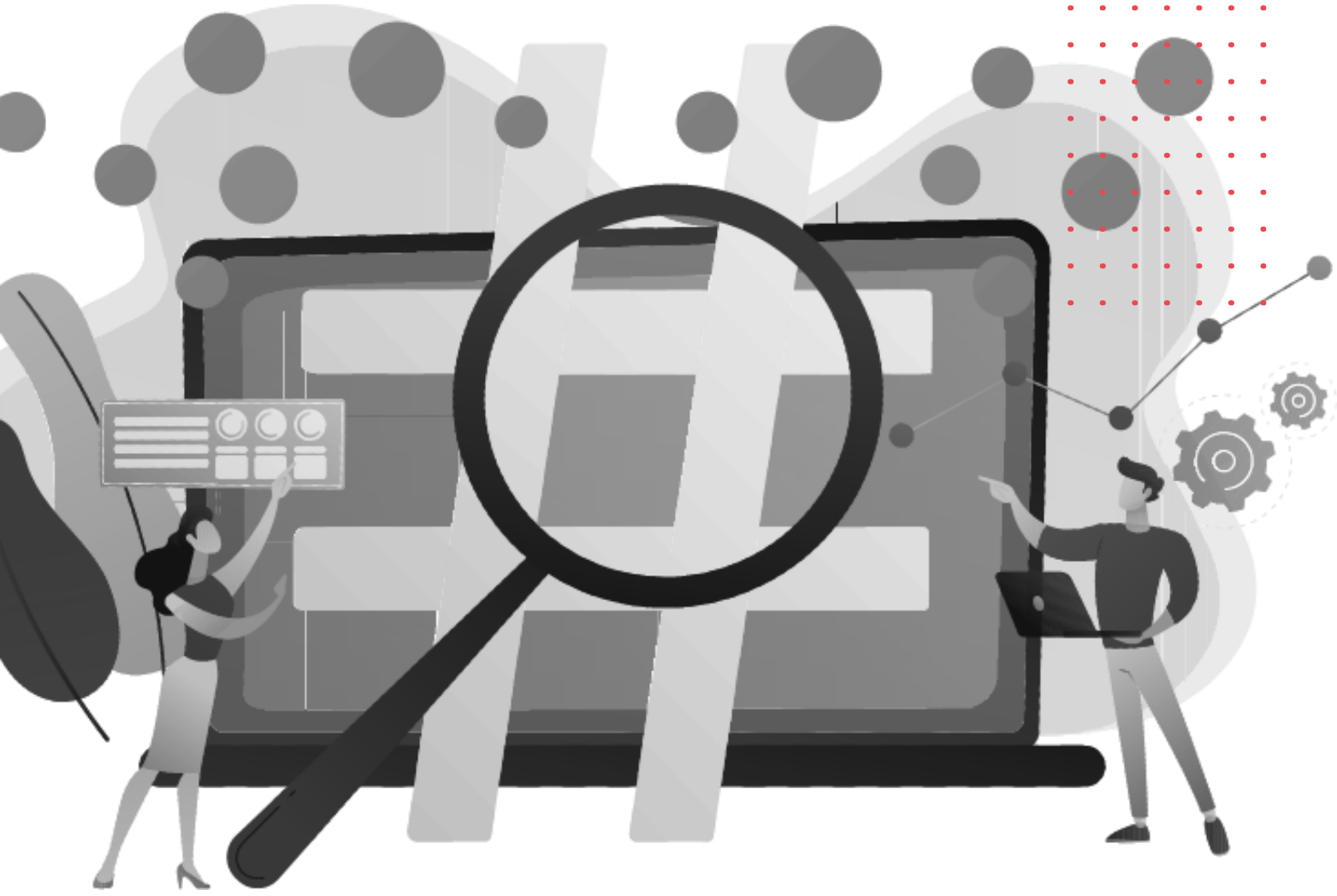
- No real time validation of toxicity in chat forums. A manual approach is followed on a reactive basis
- Loss in reputation and user experience due to inappropriate behaviour in chat forums resulting in a toxic environment for users
- Large volume of data causes the current manual process to be inefficient
- Global access needs customization and local language and regional nuances to be considered during the toxicity analysis

BENEFITS

- 98% hit ratio for identifying inappropriate content based on the given input
- Real-time capture of user behavior, creating a clean environment for communication and better user satisfaction.
- Region-based configuration for inclusion of regional nuances
- AI-powered real-time toxicity identifier for chat forums

TECHM NXT.NOW ADVANTAGE

- Diversified global player with 20+ years of data, analytics, and AI practice
- 11,000+ data, analytics, and AI associates with domain led consulting focus
- Global footprints across 55+ countries
- Partner-enabled ecosystem with COEs across all these technologies
- Comprehensive frameworks, processes and tools to move PoCs to production



To know more, reach us at
Wave2.Communications@techmahindra.com

Tech Mahindra
Connected World. Connected Experiences.