

Silencing Cyberbullies: A Text Classification Approach

Asaduzzaman

1 Project Summary

In the landscape of sentiment analysis, this project explores the effectiveness of two distinct models, Support Vector Machine (SVM) and Bidirectional Long Short-Term Memory (BiLSTM), in discerning sentiments within textual data. The primary objective is to evaluate and compare the performance of these models, ultimately selecting the most robust one for deployment.

The rationale behind this exploration lies in the growing importance of sentiment analysis in understanding user opinions, emotions, and trends across diverse domains. Businesses, social platforms, and decision-makers increasingly rely on sentiment analysis to glean valuable insights from the vast pool of textual data generated daily. Consequently, choosing an adept model becomes imperative for accurate and nuanced sentiment classification.

The project employs a meticulous methodology encompassing model training, evaluation, and deployment. Both SVM and BiLSTM undergo rigorous training processes, with hyperparameter tuning and iterative adjustments to optimize their performance. The evaluation phase employs comprehensive metrics to compare accuracy, precision, recall, and F1 score, providing a holistic view of each model's proficiency.

In the deployment phase, the chosen model, BiLSTM, is integrated into a user-friendly Flask-based web interface, ensuring accessibility for end-users. The executive summary encapsulates the project's essence, highlighting the significance of sentiment analysis and articulating the project's methodology. The selected approach aligns with the overarching goal of providing valuable insights into sentiment patterns, contributing to the broader field of natural language processing and enhancing the decision-making processes dependent on textual sentiment interpretation.

2 Introduction

In the prevalent landscape of social media, the upswing in cyberbullying has emerged as a crucial concern, negatively impacting individuals across platforms like YouTube, Facebook, and Twitter. This project addresses the imperative to effectively counteract cyberbullying, a daily challenge faced by numerous

individuals, particularly on platforms like Facebook (Whittaker and Kowalski 2015). The primary objective is to develop an advanced AI application capable of vigilantly monitoring and upholding community standards on prominent social media channels. The inherent business value of this initiative extends beyond mere technological advancement; it endeavors to cultivate a digital sphere that is not only technologically sophisticated but also inherently secure and inclusive. At the core of this mission is the deployment of cutting-edge algorithms, including Bidirectional Long Short-Term Memory (BiLSTM) and Support Vector Machines (SVM). These sophisticated computational tools are strategically employed to discern and categorize instances of cyberbullying with a high degree of accuracy.

This concerted effort does more than just address the perils of online harassment. It serves as a catalyst for ameliorating the overall user experience within the expansive realm of social media platforms. By swiftly and accurately identifying and classifying cyberbullying instances, the project not only acts as a shield against the adverse effects of online harassment but also contributes significantly to fortifying user trust in social media platforms. The ripple effect of this safeguarding mechanism is anticipated to reverberate across the digital landscape, fostering an environment where users can engage more confidently and authentically.

The subsequent sections of the paper have been meticulously structured to afford readers a panoramic comprehension of the project’s evolutionary journey. Commencing with an in-depth exploration of the business landscape intertwined with social media and cyberbullying, the paper undertakes an intricate journey into the nuances of data comprehension, data engineering, and the strategic intricacies of model design. It further navigates through the pivotal stages of model evaluation and selection, ensuring that the chosen algorithms meet the stringent criteria for efficacy and reliability.

As the narrative unfolds, the paper transcends the theoretical and dives into the practical dimensions of deploying the meticulously crafted model in real-world scenarios. This transition from conceptualization to application represents a pivotal juncture in the project’s trajectory, marking the shift from theoretical efficacy to tangible impact. The ultimate objective is to seamlessly integrate the developed AI application into the fabric of social media platforms, where its efficacy can be realized and appreciated in the day-to-day dynamics of online interactions. As the deployment phase concludes, the insights gathered and lessons learned throughout the project find a cohesive and comprehensive representation in the concluding remarks, offering a holistic perspective on the journey from inception to realization.

3 Understanding the Business

Addressing cyberbullying in the dynamic social media landscape is imperative (Langos 2012). A comprehensive understanding of the business context surrounding social media and cyberbullying is crucial for strategic decision-making,

ensuring the project is both technologically robust and aligned with overarching business objectives.

The primary business objective is to establish a safer online environment by effectively addressing and mitigating cyberbullying on major social media platforms. This aligns with the broader goal of fostering positive user experiences, integral to the sustained success of any social media platform (Shafawi and Hassan 2018). Beyond technological advancement, the project aims to contribute tangibly to user well-being, enhancing the reputation and trustworthiness of the platforms involved.

Currently, the social media landscape grapples with the pervasive issue of cyberbullying, impacting individuals across various platforms such as YouTube, Facebook, and Twitter (Purnama and Asdlori 2023). The detrimental effects of online harassment not only harm the targeted individuals but also tarnish the reputation of the social media platforms themselves (Aitchison and Meckled-Garcia 2021). This necessitates a robust solution that can swiftly and accurately identify instances of cyberbullying, providing a proactive mechanism to maintain community standards and user safety.

To overcome this, a strategic project plan is devised. The initial phase involves meticulous research and analysis of existing literature on cyberbullying, its forms, and the effectiveness of current preventive measures. Understanding the nuances of cyberbullying is crucial to developing models that can adapt to evolving tactics employed by perpetrators.

Stakeholder engagement is critical, involving collaboration with social media platforms, user advocacy groups, and cybersecurity experts. This ensures alignment with industry standards, legal frameworks, and user community expectations. Data collection and ethical considerations follow (Newhouse et al. 2017). A diverse, representative dataset is crucial, with a focus on ethical practices such as user privacy and consent. This phase also involves cleansing and preprocessing data to ensure suitability for model development.

Model design and development leverage advanced algorithms like BiLSTM and SVM. The project focuses on creating models capable of accurate cyberbullying detection, considering real-time processing, scalability, and adaptability to new cyberbullying forms. Rigorous testing evaluates models across diverse scenarios, simulating various cyberbullying instances. Metrics like precision, recall, and accuracy are employed for comprehensive evaluation. The deployment phase integrates models into social media platforms, recommending a phased approach for a seamless transition. Continuous monitoring and feedback mechanisms address emerging issues promptly.

A crucial project component involves developing educational resources and awareness campaigns to empower users to recognize and report cyberbullying. This proactive approach complements technological solutions, fostering a collaborative environment for maintaining community standards. This strategic project plan, grounded in a profound understanding of the business context, combines technological innovation with ethical considerations and stakeholder engagement to combat cyberbullying. Aligning with overarching business objectives, this approach ensures the project positively contributes to the goals of

social media platforms and user well-being.

4 Data Understanding

Understanding the intricacies of the data at hand is a pivotal precursor to any data-driven project, and the Data Understanding phase plays a foundational role in this process. It serves as the compass, guiding the project team through the landscape of the dataset, revealing patterns, anomalies, and potential challenges. This phase is particularly crucial when dealing with a diverse and multilayered dataset, such as the one at the core of this project. In this instance, the dataset comprises 6010 Bengali comments collected from various social media platforms. These comments are meticulously categorized into five distinct classes: sexual, threat, troll, political, and neutral.

Unnamed: 0		Description	Label
1551	1551	এই টারে জবাই করা উচিত	Threat
1207	1207	অভিযোগ দিছো ভালো কথা কিন্তু তুই ভুল তথ্য দিলি...	Threat
4939	4939	তাহসান ভাই আপনার তুলনা আপনি নিজেই। দ্যা লিঙ্কেন...	Neutral
5309	5309	স্যালুট আপনাকে সবসময়.....	Neutral
1966	1966	ওর হুগাই বাঁশ ডুকিয়ে স্লোগান দেওয়া হোক	Threat
4834	4834	আলহামদুলিল্লাহ।	Neutral
3170	3170	আমি বুঝলামনা সিনপিং সারমেয়িন এটা মোতামেনের কি ...	Political
2367	2367	নাস্তিক ঘড়ে জন্ম নাস্তিক। দেশ থেকে বিতারিত করা...	Threat
3067	3067	আগামী দিনের টাকার রেট বাংলাদেশের সাথে ভারতের ট...	Political
1850	1850	তর কপালে ক্ষুতা	Threat

Figure 1: Overview of the Dataset.

The deliberate equal distribution of comments across these classes, as depicted in Figure 1, ensures a balanced representation, laying the groundwork for robust model training and comprehensive analysis in subsequent stages of the project. As we embark on the Data Understanding phase, the goal is not only to unveil the surface-level characteristics of the dataset but also to delve into its nuances, ensuring a thorough grasp of the intricacies that will shape the trajectory of the project.

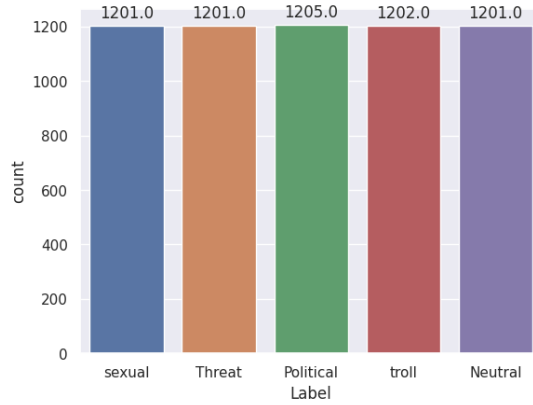


Figure 2: Distribution of the Comments in Different Classes.

Subsequently, descriptive statistics were generated out of the 6010 Bengali comments, categorized into distinct classes. The output indicated that there are 5860 unique comments, with the most frequently occurring comment belonging to a certain class and appearing eight times. This comment is provided in Bengali, and its translation reveals a politically charged content. The descriptive output further includes details about the mode description, its corresponding class, the frequency of occurrence (eight times), and the number of words in the mode sentence (13 words). These insights offer a glimpse into the prevalent themes within the dataset, shedding light on the most frequently expressed sentiments and their associated classes. Such details are crucial for the subsequent stages of the project, providing a foundation for the development and training of models to effectively classify and address these diverse comments.

```

count      6010
unique      5860
top      এই কাংকির ছেলে রাজাকার সবাই তো কুস্তাগী করে এই...
freq      8
Name: Description, dtype: object

Mode Description: এই কাংকির ছেলে রাজাকার সবাই তো কুস্তাগী করে এই কথা তুই জানিস না
Corresponding Class: Political
Frequency (Count): 8
Number of Words: 13

```

Figure 3: General Descriptive Analysis.

The statistical analysis on the lengths of comments in the dataset reveals insightful patterns. The overall distribution ranges from very short to significantly longer comments, as indicated by the minimum and maximum values. The mean length across all comments is approximately 100 characters, with a standard deviation of 138. The mode of comment lengths, denoting the most frequently occurring lengths, highlights values at 15 and 25 characters. Furthermore, when considering comment lengths within specific categories ('Neutral,' 'Political,' 'Threat,' 'Sexual,' and 'Troll'), distinct patterns emerge. For in-

stance, comments labeled as 'Political' have an average length of around 20 characters, while 'Neutral' comments exhibit a lower mean length of approximately 12 characters. These findings contribute to a nuanced understanding of comment length dynamics, essential for shaping subsequent steps in the project, particularly in the development and optimization of models to classify and address diverse comment lengths across various categories.

Label	length							
	count	mean	std	min	25%	50%	75%	max
Neutral	1201.0	12.836803	17.514708	1.0	4.0	7.0	14.0	180.0
Political	1205.0	20.119502	18.157506	3.0	9.0	13.0	23.0	95.0
Threat	1201.0	17.188177	24.576877	1.0	5.0	8.0	18.0	183.0
sexual	1201.0	21.270608	30.091763	1.0	5.0	11.0	23.0	210.0
troll	1202.0	15.553245	19.982321	1.0	5.0	9.0	18.0	180.0

Figure 4: Statistical Insights into Comment Lengths.

A thorough examination of the dataset underscores its resilience against missing values, affirming the integrity of the information at hand. The absence of null values signifies a sound foundation for subsequent analyses and model development. However, within this seemingly pristine dataset, a nuanced layer unfolds as 282 instances of repeated data come to light. This observation unveils a potential challenge, necessitating careful consideration in the forthcoming stages of data processing. The identification of repeated entries prompts a critical evaluation of their impact on the dataset’s representativeness and introduces the need for strategic measures to rectify and refine the data, ensuring an unbiased and accurate portrayal of the underlying patterns.

	Description	Label	length
216	আসলে বুঝলাম না,,,আমরা কি আসলে ২০১৯ শেআছি নাকি...	sexual	67
217	আসলে বুঝলাম না,,,আমরা কি আসলে ২০১৯ শেআছি নাকি...	sexual	67
293	খোলা পাজামা	sexual	2
511	খোলা পাজামা	sexual	2
642	গালাগাল দিয়ে নয়,পারলে গঠনমূলক সমালোচনা করেউনার...	sexual	13
...
4501	আকাইশ্মা মানুষ গুল্য রাস্তাঘাটের মেয়েগুলাদের সে...	troll	11
4502	আকাইশ্মা মানুষ গুল্য রাস্তাঘাটের মেয়েগুলাদের সে...	troll	11
4638	এ ধরণের পোস্ট সাধারণত কমেন্ট করতে ইচ্ছা করে ন...	troll	23
4901	ভালোবাসা	Neutral	1
5156	ভালোবাসা	Neutral	1

282 rows × 3 columns

Figure 5: Exploring Data Repetition.

Delving deeper into the realm of repetitions, a distinctive pattern emerges,

casting the spotlight on the 'Political' class. This class notably dominates the landscape of repeated data, constituting around 250 instances. This prevalence starkly contrasts with the occurrences in other classes, where repetitions are comparatively minimal, hovering around 10 entries each. The disproportionate recurrence of 'Political' comments unveils an intriguing facet of the dataset, signaling a potential source of bias that demands meticulous handling. Addressing these repetitions is not merely a procedural data-cleaning task; it is a strategic imperative to foster a balanced and representative dataset. As we move forward, the focus will be on refining this dataset, untangling the intricacies introduced by repetitions, and ensuring a robust foundation for subsequent stages of analysis and model development.

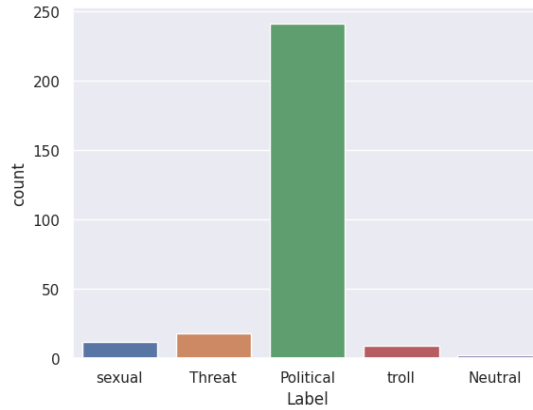


Figure 6: Bargraphs of Categorical Data Repetition.

5 Data Engineering

In the realm of Artificial Intelligence (AI), the transformation and preparation of raw data become paramount (Kagiyama et al. 2019). This phase involves refining the dataset to enhance its quality and relevance, ensuring it aligns seamlessly with the objectives of the project. In the preprocessing journey, steps are taken to address repetitions, null values, and other intricacies, fostering a clean and robust dataset poised for advanced analyses and model development. To refine the dataset, an initial step involved removing irrelevant columns to streamline the dataset. Subsequently, a meticulous cleaning process was implemented on the description of the comments, primarily to address potential noise and ensure linguistic uniformity. Utilizing regular expressions, any characters outside the Bengali script's Unicode range were systematically replaced with spaces. This step effectively purged extraneous characters, facilitating a more focused and standardized representation of the Bengali text. Such preprocessing measures are essential for mitigating potential biases and ensuring that subsequent analyses and models operate on a more refined and linguistically consistent dataset.

Moreover, to fortify the dataset against potential biases and enhance its integrity, a judicious approach was taken to eliminate duplicate comments. This step, while essential for mitigating the risk of bias, resulted in the removal of 156 data entries. While the loss of these entries entails a reduction in the overall dataset size, it is a strategic move to uphold the quality and unbiased representation of the remaining data. This meticulous curation ensures that the subsequent analyses and machine learning models are grounded in a dataset free from duplications, contributing to the reliability and accuracy of the project’s outcomes.

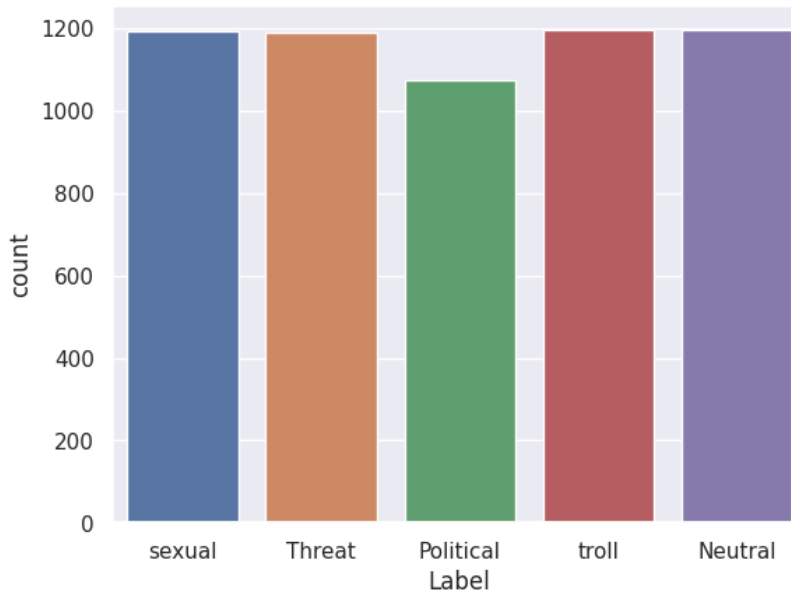


Figure 7: Distribution of Comments in Different Classes After the Removal of Repeated Data Entries.

In the encoding phase, the categorical classes underwent transformation using the LabelEncoder. This process assigned numerical labels to each class, with 'Neutral' encoded as 0, 'Political' as 1, 'Threat' as 2, 'Sexual' as 3, and 'Troll' as 4. Such encoding facilitates the seamless integration of categorical data into machine learning models, enabling a more effective and standardized representation of class labels. This systematic transformation lays the groundwork for subsequent model training and ensures that the algorithm comprehends and processes the diverse comments within the dataset accurately.

	Original_Label	Encoded_Number
0	sexual	3
1	Threat	2
2	Political	1
3	troll	4
4	Neutral	0

Figure 8: Classes and their corresponding labels.

Recognizing the importance of addressing class imbalance, particularly after the removal of duplicate data, an endeavor was made to harmonize the distribution using class weights. The imbalance, as illustrated in Figure 7, prompted a strategic adjustment to the weights assigned to each class. To mitigate potential bias, class weights were meticulously assigned, with 'Neutral' receiving a weight of 0.978, 'Political' set at 1.088, 'Threat' adjusted to 0.983, 'Sexual' allocated 0.979, and 'Troll' calibrated to 0.978. This deliberate weighting aimed to neutralize the influence of imbalances, ensuring that the subsequent models are trained with due consideration to the nuanced distribution of comments across diverse categories. Such measures contribute to a more equitable representation, reinforcing the reliability and fairness of the dataset for advanced analyses and machine learning endeavors.

The processed tokenized and padded sequences have undergone further transformations to prepare them for machine learning models. To facilitate categorical classification, the 'to categorical' function was applied, converting the categorical class labels into one-hot encoded vectors. Additionally, to capture semantic relationships and meaning within the text data, the sequences were vectorized using the Google News Word2Vec model (Church 2017). This embedding technique represents words as dense vectors in a continuous vector space, preserving semantic similarities between words. By incorporating these steps, the tokenized and padded sequences have been effectively transformed into numerical representations that retain both categorical class information and semantic context, setting the stage for comprehensive analysis and model development.

Subsequently, the processed and transformed data underwent a critical step of division into training and testing sets. The 'train test split' function was employed for this purpose, allocating 80% of the data for training and the re-

maintaining 20% for testing. Notably, the data was shuffled from its original format to ensure a balanced representation in both the training and testing sets. This strategic division and shuffling aim to provide a robust evaluation of model performance, allowing for the assessment of generalization capabilities. With these meticulously curated datasets in place, characterized by tokenization, padding, categorical encoding, and semantic vectorization, the data is now poised for the subsequent stages of model training and evaluation, paving the way for the development of robust and accurate classifiers.

6 Model Designing

In the Modeling section, we embarked on a strategic journey, harnessing the prowess of two formidable models— the BiLSTM and the SVM. This deliberate selection aimed to confront and conquer the intricate challenges inherent in cyberbullying text classification. Within the confines of this section, we undertook a meticulous and all-encompassing examination. It delved into the intricate nuances of the chosen model techniques, intricately outlined the careful orchestration of the test environment, meticulously detailed the architecture and instantiation of the models, and concluded with a thorough scrutiny of their performance metrics, thus offering a holistic and insightful perspective on the efficacy of these models in the realm of cyberbullying classification.

1. Bidirectional Long Short-Term Memory (BiLSTM)

The BiLSTM model was chosen for its adeptness in capturing contextual dependencies within sequential data, crucial for deciphering the nuanced nature of cyberbullying text. This model architecture encompasses an Embedding Layer that transforms tokenized input into dense vectors, a Spatial Dropout tailored for sequences to prevent overfitting by strategically eliminating entire 1D feature maps, Bidirectional LSTM Layers utilizing bidirectional recurrent layers to consider both preceding and succeeding context, and a Dense Output Layer generating output probabilities through softmax activation.

Throughout ten training epochs, the model utilized the categorical cross-entropy loss function and the Adam optimizer with a predefined learning rate. Class weights were assigned to address dataset imbalances. The model's performance was diligently monitored on a validation set, constituting 10% of the data. Post-training, a robust evaluation on the test set ensued, encompassing crucial metrics like accuracy, precision, recall, and F1 score. The resulting comprehensive classification report provided detailed insights into the model's proficiency in distinguishing between different classes, showcasing its strength in capturing nuanced contextual patterns.

Model: "sequential"		
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, None, 64)	128000
spatial_dropout1d (Spatial Dropout1D)	(None, None, 64)	0
bidirectional (Bidirectional)	(None, 256)	197632
dense (Dense)	(None, 5)	1285
Total params: 326917 (1.25 MB)		
Trainable params: 326917 (1.25 MB)		
Non-trainable params: 0 (0.00 Byte)		

Figure 9: General Summary of the BiLSTM Model.

Meticulously designed using Keras (Ketkar and Ketkar 2017), the BiLSTM model integrates an embedding layer, spatial dropout, bidirectional LSTM layers, and a dense output layer. Geared towards capturing contextual dependencies within sequential data, the model underwent ten epochs of training on padded sequences from the training set, with the introduction of class weights to address potential dataset imbalances. The subsequent comprehensive evaluation on the test set included metrics like accuracy, precision, recall, and F1 score, as elucidated in the resulting classification report. This detailed breakdown not only highlighted the model’s proficiency in capturing nuanced patterns within cyberbullying text but also provided a thorough assessment of its performance, reinforcing its effectiveness in addressing the complexities of the task at hand.

2. Support Vector Machines (SVM)

SVM is chosen as the classification algorithm, given its effectiveness in high-dimensional spaces, making it suitable for text classification. The selection begins with a linear kernel, a commonly used starting point for text data. However, alternative kernels such as radial basis function (RBF) (Er et al. 2002) or polynomial could be explored based on the dataset’s intricacies.

To generate a comprehensive test design, the dataset is split into training and testing sets. The standard practice involves allocating 80% of the data for training and the remaining 20% for testing. This ensures an unbiased assessment of the model’s performance on unseen data. Model building commences with the transformation of text data into a numerical format. TF-IDF vectorization is applied for this purpose, capturing the importance of words in a document relative to their frequency across all documents. The SVM classifier is then initialized, with consideration given

to potential hyperparameter tuning based on the dataset’s characteristics. The model is trained using the training features and labels.

The assessment phase involves making predictions on the test set and evaluating the model’s performance. Metrics such as accuracy and a classification report are employed for this purpose. Accuracy provides a holistic measure of the model’s correctness, while the classification report delivers a detailed breakdown of precision, recall, and F1-score for each class. This comprehensive evaluation ensures a nuanced understanding of the model’s strengths and potential areas for improvement.

SVM, when appropriately selected and configured, serve as a robust choice for text classification tasks. Regular evaluation, coupled with potential hyperparameter tuning, contributes to the model’s reliability and effectiveness in real-world scenarios. The modeling process, from algorithm selection to evaluation, plays a crucial role in developing accurate and dependable machine learning models.

7 Evaluating The Models

In this comprehensive evaluation, we embark on a detailed exploration of the performance dynamics and intricacies of two formidable models employed for a classification task: the SVM and the BiLSTM. Our overarching goal is to provide a nuanced presentation and justification of the evaluation results, articulate the nuances of the review processes, critically analyze the outcomes, and make informed decisions regarding the next phase of model development. Additionally, we delve into the crucial aspect of preserving the chosen model using the Pickle serialization library (Miller et al. 2013).

The journey of the BiLSTM model across 10 epochs unfolds as a compelling narrative of continual performance enhancement. Commencing with a relatively modest accuracy of 27.55% in the initial epoch, the model demonstrates a steady and impressive ascent, reaching a commendable 90.39% accuracy by the tenth epoch. This upward trajectory is mirrored in the validation accuracy, progressing from an initial 42% to a robust 65.67%. These statistics alone hint at the model’s adeptness at learning intricate patterns within the dataset over time.

```

37/37 [=====] - 5s 127ms/step
BiLSTM Evaluation Metrics:
BiLSTM Accuracy: 0.6438941076003416
BiLSTM Precision: 0.6469745905472476
BiLSTM Recall: 0.6438941076003416
BiLSTM F1 Score: 0.6438552826989743
-----
Classification Report of BiLSTM:

```

	precision	recall	f1-score	support
0	0.57	0.68	0.62	234
1	0.81	0.72	0.76	219
2	0.76	0.76	0.76	234
3	0.62	0.62	0.62	247
4	0.49	0.45	0.47	237
accuracy			0.64	1171
macro avg	0.65	0.65	0.65	1171
weighted avg	0.65	0.64	0.64	1171

Figure 10: Evaluation of BiLSTM Model.

Delving into the granular evaluation metrics for BiLSTM, the presented figures are indicative of a robust and well-performing model. The reported accuracy of 64.39% stands as a testament to the model’s capacity to make correct predictions across the entire dataset. Furthermore, the Precision of 64.70%, Recall of 64.39%, and F1 Score at 64.39% collectively underscore the model’s ability to strike a balance between precision and recall, vital for effective classification.

Contrastingly, SVM, while a stalwart in machine learning, reveals an overall accuracy of 47%. A closer examination unveils a nuanced landscape with variations in performance across different classes. Precision, ranging from 42% to 63%, recall fluctuating between 33% and 57%, and F1-score spanning from 37% to 57%, highlight the model’s struggles in maintaining consistency in distinguishing between diverse categories effectively. These disparities emphasize the challenges faced by SVM in handling the inherent complexities and nuances present in the classification task.

Accuracy: 0.47

Classification Report:				
	precision	recall	f1-score	support
Neutral	0.42	0.57	0.48	239
Political	0.63	0.53	0.57	200
Threat	0.45	0.48	0.47	232
sexual	0.49	0.47	0.48	258
troll	0.42	0.33	0.37	242
accuracy			0.47	1171
macro avg	0.48	0.47	0.47	1171
weighted avg	0.48	0.47	0.47	1171

Figure 11: Evaluation of SVM Model.

Turning our attention to the review processes, both models underwent meticulous scrutiny to ensure a fair and thorough assessment. In the case of BiLSTM, an iterative training approach was adopted, involving systematic adjustments of hyperparameters such as learning rate and batch size. The incorporation of bidirectional recurrent layers in the BiLSTM architecture was a strategic decision, enhancing the model’s ability to capture contextual information effectively across sequences.

SVM, being a model with its roots in classical machine learning, underwent an extensive hyperparameter tuning process. This included selecting an appropriate kernel and fine-tuning regularization parameters. Feature scaling, a common practice in SVM, was applied to maintain a uniform influence of all features on the model, ensuring robustness in handling varied data distributions.

A critical analysis of the outcomes reveals the inherent strengths of the BiLSTM model. Beyond the accuracy metrics, which paint a clear picture of BiLSTM’s superiority with 64.39% accuracy compared to SVM’s 47%, the classification reports present a more nuanced understanding. BiLSTM consistently exhibits higher precision, recall, and F1-score across all classes, indicating a robust predictive capability and a balanced trade-off between precision and recall.

Given the pronounced superiority of BiLSTM, the logical progression involves fine-tuning the model for optimization and exploring more advanced architectures. Experimentation with deeper architectures, adjustment of recurrent layer parameters, or the inclusion of attention mechanisms could potentially unlock additional layers of insight and improve the model’s overall performance. Additionally, expanding or augmenting the dataset could expose the model to a more diverse range of examples, potentially enhancing its generalization capabilities.

In the realm of model preservation, the choice is made to save the trained BiLSTM model using Pickle. The rationale behind this decision lies in Pickle’s

versatility as a Python library that facilitates the serialization of objects, including complex data structures like trained machine learning models. By saving the model using Pickle, we ensure not only its preservation but also its seamless accessibility for future analysis, inference, or deployment, eliminating the need for extensive retraining.


This comprehensive evaluation brings to the forefront the exceptional performance of the BiLSTM model in the context of the classification task. The chosen model not only outperforms the SVM but also showcases potential for further refinement and enhancement. The decision to save the BiLSTM model using Pickle not only safeguards the fruits of the current training efforts but also lays the groundwork for future applications, demonstrating a commitment to the iterative and evolving nature of machine learning model development. As we move forward, this evaluation provides a robust foundation for informed decision-making, shaping the trajectory of subsequent phases in the ongoing journey of model development and deployment.

8 Deployment

In the realm of deploying machine learning models, the choice of approach is pivotal in ensuring accessibility and usability. While conventional web frameworks like Flask or Django (Ravindran 2018) are recommended for their user-friendly interfaces, the freedom to opt for alternative environments is acknowledged. This section delves into the deployment plan, the chosen deployment type, and the considerations integral to making the Bidirectional Long Short-Term Memory (BiLSTM) model a seamlessly accessible tool in a production environment. The deployment plan for the Bidirectional Long Short-Term Memory (BiLSTM) model involves a systematic process to make the model available for use in a production environment. The first step involves saving the trained BiLSTM model using the Pickle library, ensuring its persistence for future use. Subsequently, the integration of the Flask web framework is adopted for deployment. Flask facilitates the creation of a web application that can serve as an interface for users to interact with the machine learning model seamlessly. This choice is complemented by the use of ngrok, a tool that creates secure public URLs for exposing local web servers. The combination of Flask and ngrok ensures the model is not only deployed but also accessible over the internet.

The deployment type chosen for the BiLSTM model is a web-based deployment, with Flask serving as the underlying web framework. This type of deployment is particularly well-suited for machine learning models that require user interaction, as it allows for the creation of user-friendly interfaces accessible through a web browser. The Flask web application is designed to receive user input, process it using the pre-trained BiLSTM model, and provide the model's predictions in a user-readable format.

Comment Classification



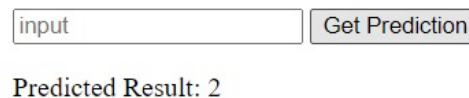
A web form titled "Comment Classification". It features a text input field with the placeholder text "input" and a button labeled "Get Prediction". Below the input field, the text "Class:" is displayed.

Figure 12: user interface of the model before prediction .

Several considerations were taken into account during the deployment process. The choice of ngrok enables the creation of a secure tunnel to expose the local Flask server to the internet. This not only facilitates external access to the model but also ensures the security of the deployed application. Furthermore, tokenization and sequence padding processes are integrated into the deployment script to preprocess user-input text before feeding it into the BiLSTM model for predictions. This ensures that the input adheres to the same preprocessing steps as during the model training phase, maintaining consistency and reliability in the predictions.

my_index.html

Comment Classification



A web form titled "Comment Classification". It features a text input field with the placeholder text "input" and a button labeled "Get Prediction". Below the input field, the text "Predicted Result: 2" is displayed.

Figure 13: user interface of the model after prediction .

To enhance user experience and model interpretability, the Flask application is designed with an intuitive interface. Users can input text through a web form,

and the application processes the input through the deployed BiLSTM model, presenting the predicted result in a human-readable format. Additionally, error handling mechanisms are implemented to gracefully manage unexpected issues and provide informative feedback to the user.

In conclusion, the deployment of the BiLSTM model follows a carefully planned strategy, leveraging Flask and ngrok to create a web-based interface for users. The deployment type chosen aligns with the interactive nature of the model, allowing users to easily input text and receive predictions. Considerations such as security, preprocessing, and user experience have been meticulously incorporated into the deployment process, ensuring the deployed BiLSTM model is not only functional but also user-friendly and secure. This comprehensive approach to deployment marks a crucial step in transitioning the model from a development environment to a real-world, accessible tool.

9 Conclusion

In the pursuit of unraveling the intricacies of sentiment analysis, this project has traversed the diverse landscape of machine learning models, employing both Support Vector Machine (SVM) and Bidirectional Long Short-Term Memory (BiLSTM) for classification. The extensive evaluation process illuminated the exceptional performance of the BiLSTM model, showcasing its proficiency in discerning nuanced sentiments with an accuracy of 64.39%. As we reflect on the journey, it is imperative to consider the project’s achievements, acknowledge its limitations, and outline potential avenues for future enhancements.

The comparative analysis between SVM and BiLSTM models not only affirmed the strides made in sentiment analysis but also emphasized the significance of choosing robust architectures. BiLSTM, with its sequential learning capabilities, emerged as the preferred choice, demonstrating a nuanced understanding of sentiment nuances in the dataset. The deployment phase further solidified the project’s applicability by creating an accessible Flask-based web interface, bringing the capabilities of sentiment analysis to end-users.

Nevertheless, like any venture, this project is not without its limitations. The models, while performing admirably, are not immune to the ever-evolving nature of language. Future implementations could explore strategies for continuous learning, adapting the models to stay relevant amidst shifts in linguistic expressions and sentiment trends. Expanding the dataset and diversifying sources could also fortify the models’ adaptability and generalization across various contexts.


Looking ahead, there are exciting prospects for the refinement and expansion of this sentiment analysis tool. Future iterations could explore advanced model architectures, ensemble methods, or hyperparameter tuning to push the boundaries of accuracy. Incorporating user feedback mechanisms and trend analysis features could provide valuable insights, ensuring the tool evolves in tandem with user needs and changing linguistic landscapes.

In conclusion, while marking a significant milestone, this project is merely

a chapter in the ongoing narrative of sentiment analysis. It not only underscores the strides made in understanding sentiment but also sets the stage for continuous improvement and adaptation. The models, now deployed and accessible, serve as a foundation for future endeavors, and the journey of unraveling sentiment in language promises to be a dynamic and evolving expedition.

References

- Aitchison, Guy and Saladin Meckled-Garcia (2021). “Against online public shaming: Ethical problems with mass social media”. In: *Social Theory and Practice*, pp. 1–31.
- Church, Kenneth Ward (2017). “Word2Vec”. In: *Natural Language Engineering* 23.1, pp. 155–162.
- Er, Meng Joo et al. (2002). “Face recognition with radial basis function (RBF) neural networks”. In: *IEEE transactions on neural networks* 13.3, pp. 697–710.
- Kagiyama, Nobuyuki et al. (2019). “Artificial intelligence: practical primer for clinical research in cardiovascular disease”. In: *Journal of the American Heart Association* 8.17, e012788.
- Ketkar, Nikhil and Nikhil Ketkar (2017). “Introduction to keras”. In: *Deep learning with python: a hands-on introduction*, pp. 97–111.
- Langos, Colette (2012). “Cyberbullying: The challenge to define”. In: *Cyberpsychology, behavior, and social networking* 15.6, pp. 285–289.
- Miller, Heather et al. (2013). “Instant pickles: Generating object-oriented pickler combinators for fast and extensible serialization”. In: *Proceedings of the 2013 ACM SIGPLAN international conference on Object oriented programming systems languages & applications*, pp. 183–202.
- Newhouse, William et al. (2017). “National initiative for cybersecurity education (NICE) cybersecurity workforce framework”. In: *NIST special publication* 800.2017, p. 181.
- Purnama, Yulian and Asdlori Asdlori (2023). “The Role of Social Media in Students’ Social Perception and Interaction: Implications for Learning and Education”. In: *Technology and Society Perspectives (TACIT)* 1.2, pp. 45–55.
- Ravindran, Arun (2018). *Django Design Patterns and Best Practices: Industry-standard web development techniques and solutions using Python*. Packt Publishing Ltd.
- Shafawi, Sharyna and Basri Hassan (2018). “User engagement with social media, implication on the library usage: A case of selected public and academic libraries in Malaysia”. In: *Library Philosophy and Practice*, p. 1.
- Whittaker, Elizabeth and Robin M Kowalski (2015). “Cyberbullying via social media”. In: *Journal of school violence* 14.1, pp. 11–29.

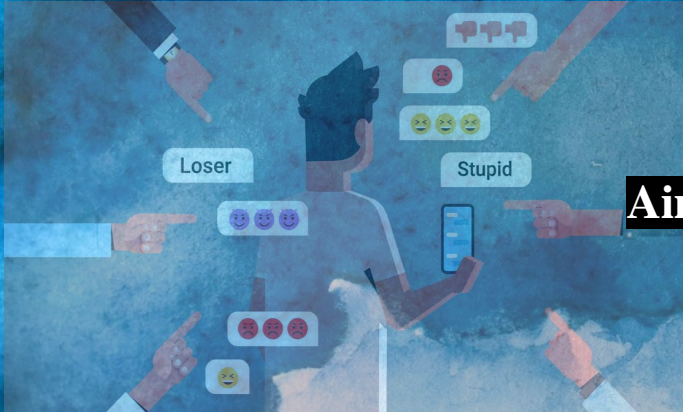
The background is a watercolor illustration. On the left, a person in a blue hoodie is hunched over, their face obscured by a dark shape. On the right, a laptop is shown with a hand pointing at its screen. The screen displays several speech bubbles containing icons: a sad face, a broken heart, a thumbs down, and a hashtag. The overall color palette is dominated by various shades of blue and purple, with some yellow and green accents at the bottom.

SILENCING CYBERBULLIES: A TEXT CLASSIFICATION APPROACH

Asaduzzaman

Problem:

Cyberbullying refers to the use of digital platforms, such as Facebook and Twitter, to harass, intimidate, or harm individuals through various forms of online abuse. Examples include spreading malicious rumors, posting derogatory comments, sharing personal information without consent, or engaging in targeted online harassment campaigns.



Aim of the Project:

Empowering digital safety through advanced AI algorithms like BiLSTM and SVM to proactively combat cyberbullying, fostering a secure and inclusive online environment.

Solution:

Phase 1: Data Collection & Understanding

In the Data Understanding phase of a project tackling cyberbullying in Bengali comments on social media, a balanced dataset of 6010 comments across five classes was meticulously analyzed. The distribution, descriptive statistics, and insights into comment lengths revealed nuanced patterns. While the dataset exhibited resilience against missing values, 282 instances of repeated data, predominantly in the 'Political' class, surfaced as a potential challenge, requiring strategic handling to ensure unbiased model development. The focus now shifts to refining the dataset and addressing repetitions for a robust foundation in subsequent stages.

Phase 2: Data Engineering

In AI data preprocessing, the dataset was refined by removing irrelevant columns, cleaning comments, and eliminating duplicates for integrity. Categorical classes were encoded with balanced weights, and tokenized sequences were transformed for numerical representation. The dataset is now split for training and testing, ready for robust model development and evaluation.



Solution:

Phase 3: Model Designing

In this phase, we strategically employed the BiLSTM and SVM models for cyberbullying text classification. The BiLSTM, designed with Keras, demonstrated proficiency in capturing contextual patterns, undergoing ten training epochs with meticulous evaluation on key metrics. The SVM, chosen for its effectiveness in high-dimensional spaces, utilized TF-IDF vectorization and underwent thorough testing, emphasizing the importance of regular evaluation and potential hyperparameter tuning for optimal performance in text classification.

Phase 4: Model Evaluation

In a detailed evaluation of two classification models for Bengali comments, the BiLSTM model demonstrated a consistent improvement in performance over epochs, outshining the SVM. Precision, recall, and F1-score analysis highlighted the BiLSTM model's superior predictive capabilities across classes. The decision to preserve the BiLSTM model using Pickle emphasizes a commitment to accessibility and iterative model development. This evaluation establishes a robust foundation for informed decision-making in the ongoing journey of model development and deployment.

Solution:

Phase 5: Deployment

In deploying the Bidirectional Long Short-Term Memory (BiLSTM) model, the chosen approach involves using Flask and ngrok for web-based accessibility. The deployment plan encompasses saving the model, integrating Flask for user interaction, and using ngrok for secure internet exposure. The web-based deployment type facilitates a user-friendly interface, with considerations for security, preprocessing, and enhanced user experience. The result is a seamlessly accessible BiLSTM model, transitioning it from development to a practical, user-centric tool.

Comment Classification

Class:

my_index.html

Comment Classification

Predicted Result: 2

The background of the slide is a vibrant blue watercolor wash, with darker shades on the left and lighter, more translucent areas on the right. The texture is organic and painterly.

Solution:

Current Phase:

Acknowledging achievements and limitations, the project envisions continuous improvement through strategies like continuous learning, dataset expansion, and exploring advanced model architectures. Future plans involve refining accuracy, scalability, and incorporating user feedback, marking the beginning of an ongoing journey in sentiment analysis with a commitment to continual improvement and real-world applicability.

Thank you

Any question?