ARTICLE

Check for updates

Decoding violence against women: analysing harassment in middle eastern literature with machine learning and sentiment analysis

Hui Qi Low[1], Pantea Keikhosrokiani [2✉] & Moussa Pourya Asl[3]

The rising prevalence of harassment in Middle Eastern countries is mirrored in literary works from the region. However, extracting data from these texts to understand the typology and frequency of the cases poses a significant challenge due to human cognitive limitations and potential biases. Thus, this study aims to use natural language processing (NLP) approaches to propose a machine learning framework for text mining of sexual harassment content in literary texts. The data source for this study consists of twelve Middle Eastern novels. The proposed framework involves the classification of physical and non-physical types of sexual harassment using a machine-learning model. Lexicon-based sentiment and emotion detection are applied to sentences containing instances of sexual harassment for data labelling and analysis. Finally, a long short-term memory-gated recurrent unit (LSTM-GRU) deep learning model is built to classify the sentiment characteristics that induce sexual harassment. The proposed model achieved an accuracy of 75.8% while outperforming five other algorithms. Additionally, a sentiment classification with three labels—negative, positive, and neutral— was developed using an LSTM-GRU RNN deep learning model. The accuracy of this model was 84.5%. Most statements, even those involving physical sexual harassment, which had greater levels of sexual harassment, had negative sentiments, according to lexicon-based sentiment analysis. This study contributes to the field of text mining by providing a novel approach to identifying instances of sexual harassment in literature in English from the Middle East. The use of machine learning models and sentiment analysis techniques allows for more accurate identification and classification of different types of sexual harassment. Furthermore, this study sheds light on the prevalence of sexual harassment in Middle Eastern countries and highlights the need for further research and action to address this issue.

[1] Cluster Engineering, Singapore Institute of Technology, Singapore, Singapore. [2] Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland. [3] Faculty of Humanities, University of Oulu, Oulu, Finland. ✉email: pantea.keikhosrokiani@oulu.fi

Introduction

The escalating prevalence of sexual harassment cases in Middle Eastern countries has emerged as a pressing concern for governments, policymakers, and human rights activists. In recent years, scholars have made significant strides in advancing our understanding of the typology and frequency of these cases through both empirical and theoretical contributions (Eltahawy, 2015; Ranganathan et al., 2021). Moreover, researchers have sought to supplement their findings by examining evidence from alternative sources such as literary texts and life writings. While literary representations from the region offer valuable insights into individual and collective experiences of sexual harassment, the analysis of these texts to extract relevant data presents considerable challenges due to inherent limitations in human cognitive processes and potential biases (Keikhosrokiani and Pourya Asl, 2023). Consequently, the task of extracting specific content from extensive texts like novels is arduous and time-consuming. The scholarly community has made substantial progress in comprehending the multifaceted nature of sexual harassment cases in the Middle East (Karami et al., 2021). Researchers have conducted rigorous empirical studies that shed light on various aspects of this issue, including its prevalence rates, underlying causes, and societal implications (Bouhlila, 2019). These studies have not only provided valuable statistical data but have also generated theoretical frameworks that enhance our understanding of the complex dynamics at play. In addition to empirical research, scholars have recognized the importance of exploring alternative sources to gain a more comprehensive understanding of sexual harassment in the region. Literary texts and life writings offer unique perspectives on individual experiences and collective narratives related to this issue (Asl, 2023). However, analysing these sources poses significant challenges due to limitations in human cognitive processes. Extracting specific content from large-scale literary works requires meticulous attention to detail and an extensive amount of time. Researchers must carefully navigate through vast amounts of text to identify relevant passages that provide insights into sexual harassment experiences (Ennaji and Sadiqi, 2011). This process is further complicated by potential biases that may influence researchers' interpretations or choices of which passages to include or exclude.

A hybrid computational method that combines interpretative social analysis and computational techniques has emerged as a powerful approach in digital social research. This method enables the establishment of statistical strategies and facilitates quick prediction, particularly when dealing with large and complex datasets (Lindgren, 2020). To conduct a comprehensive study of social situations, it is crucial to consider the interplay between individuals and their environment. In this regard, emotional experience can serve as a valuable unit of measurement (Lvova et al., 2018). One of the main challenges in traditional manual text analysis is the inconsistency in interpretations resulting from the abundance of information and individual emotional and cognitive biases. Human misinterpretation and subjective interpretation often lead to errors in data analysis (Keikhosrokiani and Asl, 2022; Keikhosrokiani and Pourya Asl, 2023; Ying et al., 2022). To address this issue, hybrid methods that combine manual annotation with computational strategies have been proposed to ensure accurate interpretations are made. However, it is important to acknowledge that computational methods have limitations due to the inherent variability of sociality. Sociality can vary across different dimensions, such as social interaction, social patterns, and social activities within different data ages. Consequently, there are no "general rules" or a universally applicable framework for analysing societies or defining a "general world" (Lindgren, 2020). In this context, text mining emerges as an invaluable tool for efficiently analysing large volumes of data. Its ability to quickly identify patterns and trends related to various phenomena makes it particularly well-suited for investigating issues such as sexual harassment.

Using natural language processing (NLP) approaches, this study proposes a machine learning framework for text mining of sexual harassment content in literary texts. The data source for this study consists of twelve Middle Eastern novels written in English. The proposed framework involves the classification of physical and non-physical types of sexual harassment using a machine-learning model. Additionally, lexicon-based sentiment and emotion detection are applied to sentences containing instances of sexual harassment for data labelling and analysis. Lexicon-based sentiment analysis involves analysing text for positive or negative sentiment using pre-defined lexicons or dictionaries. Emotion analysis involves identifying emotions expressed within text, such as anger or sadness. Finally, an LSTM-GRU deep learning model is built to classify the sentiment characteristics that induce sexual harassment. The neural network approach involves training a model using large datasets to recognize patterns and make predictions based on new data inputs.

The use of machine learning approaches can help to identify patterns within large datasets that may not be immediately apparent through manual analysis. This approach can also help reduce bias by removing human subjectivity from the process of analysis. The use of machine learning models and sentiment analysis techniques allows for more accurate identification and classification of different types of sexual harassment than traditional methods such as manual coding or human annotation. Lexicon-based sentiment and emotion allow for more nuanced analysis by taking into account the emotional context surrounding instances of sexual harassment. Finally, an LSTM-GRU deep learning model allows for a deeper understanding of the underlying factors that contribute to sexual harassment, which can inform future prevention and intervention efforts. The use of lexicon-based sentiment and emotion analysis, as well as a neural network, can help identify patterns and reduce bias in the analysis process. Overall, this study contributes to the field of text mining by providing a novel approach to identifying instances of sexual harassment in literary works from the Middle East. Furthermore, this study sheds light on the prevalence of sexual harassment in Middle Eastern countries, highlighting the need for continued efforts to address this issue.

Background

Sexual harassment types. Sexual harassment is a pervasive issue that can be categorized into three distinct forms: gender harassment, unwanted sexual attention, and sexual coercion. Each category represents a different manifestation of these harmful behaviours, highlighting the various ways in which individuals are subjected to harassment in their personal and professional lives. Gender harassment is a form of discrimination that aims to hinder women from attaining positions of power in traditionally male-dominated fields. It encompasses both verbal and non-verbal conduct that seeks to belittle, demean, or exclude women based on their gender (del Carmen Herrera et al., 2017). This type of harassment perpetuates gender inequality by creating hostile environments that discourage women from fully participating and advancing in their chosen careers. Unwanted sexual attention involves deliberate contact and repetitive requests for data that are intended to attract or express offensive sexual attraction (del Carmen Herrera et al., 2017). This form of harassment often includes unwelcome advances, explicit comments, or inappropriate gestures. It is characterized by the harasser's persistent

pursuit of sexual gratification at the expense of the victim's comfort and autonomy. Such behaviour not only violates personal boundaries but also creates an intimidating atmosphere that undermines the victim's sense of safety and well-being. Lastly, sexual coercion occurs when a harasser abuses their position of power to demand sexual favours from a victim in exchange for benefits within a Quid Pro Quo environment (Fateh, 2022). This insidious form of harassment involves leveraging promises of rewards or threats of punishment to manipulate the victim into complying with the harasser's demands. The power dynamics at play exacerbate the vulnerability of the victim, as they may fear negative consequences if they refuse or report the harassment. It is crucial to recognize that these three categories are not mutually exclusive; they often intersect and coexist within instances of sexual harassment. For instance, gender harassment may lay the foundation for unwanted sexual attention or coercion by perpetuating an environment where such behaviour is normalized or tolerated. Understanding the nuances and complexities of these categories is essential in addressing and combating sexual harassment effectively.

Sexual harassment in the Middle East. Sexual harassment is a pervasive and serious problem that affects the lives and well-being of many women and men in the Middle East. According to a UN Women survey, online harassment was the most common type of violence against women in nine countries in the region during the pandemic (Ranganathan et al., 2021). However, sexual harassment is not limited to the online sphere but also occurs in various forms, including gender harassment, unwanted sexual attention, and sexual coercion in different settings such as workplaces, educational institutions, public places, and homes. Throughout the region, gender harassment often manifests through verbal abuse, derogatory comments, or discriminary behaviour towards women (Asl, 2023; Hadi and Asl, 2022). Previous studies highlight how patriarchal norms and traditional gender roles contribute to gender harassment in this region. In particular, the cultural emphasis on modesty and honour perpetuates gender harassment by placing blame on women for their attire or behaviour. The concept of "honour" has become a tool for controlling women's actions and justifying harassment (Asl, 2022, 2020; Asl and Hanafiah, 2023; Chew and Asl, 2023; Yan and Asl, 2023). Gender harassment is perpetrated to reinforce power imbalances between men and women in Middle Eastern societies. Men often exert dominance over women through verbal abuse or by limiting their access to public spaces (Wei and Asl, 2023). Numerous studies have shown that gender harassment perpetuates a culture of silence where victims are discouraged from speaking out due to fear of social stigma or retaliation, hindering progress towards gender equality and reinforcing harmful stereotypes about women's roles in Middle Eastern societies (Asl, 2019).

Both unwanted sexual attention and sexual coercion are also influenced by cultural norms surrounding modesty and sexuality. Modesty is highly valued in many Middle Eastern cultures to preserve honour and maintain social order (Ennaji and Sadiqi, 2011). Unwanted sexual attention is often seen as a violation of these cultural norms, leading to victim-blaming and shaming (Eltahawy, 2015). It is argued that the prevalence of unwanted sexual attention perpetuates a culture of fear and insecurity for women in the Middle East. It restricts their freedom of movement and limits their opportunities for education and employment, hindering their overall empowerment (Bouhlila, 2019). In cases of sexual coercion, victims often face immense pressure to remain silent due to fears that their reputation or family's honour will be tarnished, which perpetuates a cycle of violence and oppression within Middle Eastern societies. Victims often find themselves trapped in abusive relationships without access to legal protection

or support systems, leading to long-term psychological trauma. Both types of sexual harassment are often justified or normalized by the harassers as a way of expressing their masculinity and asserting their dominance.

Text classification techniques. Sexual harassment can be investigated using computation literary studies that the activities and patterns disclosed from large textual data. Computational literary studies, a subfield of digital literary studies, utilizes computer science approaches and extensive databases to analyse and interpret literary texts. Through the application of quantitative methods and computational power, these studies aim to uncover insights regarding the structure, trends, and patterns within the literature. Computational literary studies encompass various disciplines such as computational linguistics, statistical methodology, natural language processing, machine learning, and text mining (Da, 2019; Elmi et al., 2023; Mohd Amram et al., 2023, Zhao and Keikhosrokiani, 2022). The field of digital humanities offers diverse and substantial perspectives on social situations. While it is important to note that predictions made in this field may not be applicable to the entire world, they hold significance for specific research objects. For example, in computational linguistics research, the lexicons used in emotion analysis are closely linked to relevant concepts and provide accurate results for interpreting context. However, it is important to acknowledge that embedded dictionaries and biases may introduce exceptions that cannot be completely avoided. Nonetheless, computational literary studies offer advantages such as quick interpretation, analysis, and prediction on extensive datasets (Kim and Klinger, 2018).

Natural language processing (NLP) techniques have been widely adopted for text classification, which assigns labels to sentences, paragraphs, or documents (Abadah et al., 2023; Asri et al., 2022; Chu et al., 2022; Fasha et al., 2022; Jafery et al., 2023; John and Keikhosrokiani, 2022; Al Mamun et al., 2022). This technique has been applied to a variety of fields, such as health, social science, business marketing and law. In particular, text classification has been used to uncover human activity in the past decades in the social sciences. Researchers have used text data such as chat messages, notes, and social media posts for analysis. Text classification can be applied at four different levels of text size: document level, paragraph level, sentence level and sub-sentence level. The process of text classification involves four phases: feature extraction, dimension reduction, classification selection and evaluation. Firstly, unstructured text data must be converted into structured data and then cleaned to retain important characters and words in feature extraction. Secondly, dimensionality reduction may be optionally applied to reduce time and memory complexity if the pre-processed data is large. Thirdly, machine learning models, deep learning models, and ensemble-based learning must be employed for text classification. Finally, the trained classification model must be evaluated to understand its performance. (Kowsari et al., 2019).

Machine learning-based text classification. Some machine classification technique was introduced and tabulated in Table 1. Rocchio classification uses the frequency of the words from a vector and compares the similarity of that vector and a predefined prototype vector. This classification is not general because it is limited to retrieving a few relevant documents. Boosting and Bagging are voting classification techniques used in text classification. Boosting is trained by ensemble learning, where the weight of the data point changes based on the previous performance. Bagging algorithm generated a sub-sample from the training set and trained different models, and the prediction was the most voted among the trained models. The limitations of

Table 1 Idea and limitation of machine learning-based text classification.

Machine learning technique	Idea	Limitation
Rocchio classification	Use the frequency of the words from the vector and compare the similarity of that vector and a predefined prototype vector.	Not general
Boosting	Trained by ensemble learning where the weight of data points changed based on the previous performance.	Computational expensive lack of interpretability
Bagging	Generate sub-sample from training set and trained different models and the prediction is the most voted among the trained models.	Computational expensive lack of interpretability
Logistic regression	A statistical model based on a decision boundary to predict the probability of labels.	Independent variables
Naïve Bayes	Used the frequency of the words in the document and based on Bayes theorem to predict the probability of the models	Large memory usage High data dependent
K-nearest neighbours	Predict the class based on the similarity of the test document and the k number of the nearest document	Large memory usage Required a variety of data point
Support vector machine (SVM)	Develop a features map for the frequency of the words and a hyperplane was found to create the boundary between the class of data	Time consumption high memory usage
Decision tree	a statistical model that categorized the data point past on the entropy of nodes to form a hierarchical decomposition of data spaces	sensitive to data perturbation
Random forest	an ensemble learning that parallel builds multiple random decision trees and the prediction is based on the most voted by the trees	Time consumption
Conditional random field	build an observation sequence and modelled based on conditional probability	Computational complexity Cannot work with unseen data
Semi-supervised	Use the clustering technique to find if there is more than one labelled cluster or to handle the data in labelled and unlabelled clustered.	Depends on clustering algorithm

Boosting and Bagging are the computational expensive and lack of interpretability. Logistic regression is a statistical model based on a decision boundary to predict the probability of labels. The data point must be independent to perform well in prediction. Naïve Bayes classification is popular in document categorization and information retrieval. This model used the frequency of the words in the document and based on Bayes theorem to predict the probability of the models. The limitation of Naïve Bayes models is the modal has a strong assumption on the distribution of data that must obey on Bayes theorem. K-nearest neighbours (KNN) algorithm predicts the class based on the similarity of the test document and the k number of the nearest document. KNN requires large memory to store the data points and it is dependent on the variety of trained data points. Support vector machine (SVM) developed a features map for the frequency of the words and a hyperplane was found to create the boundary between the class of data. The SVM model is time-complexible and has high memory usage. Decision tree model is a statistical model that categorizes the data point past on the entropy of nodes to form a hierarchical decomposition of data spaces. Decision trees are sensitive to small perturbations in the trained data. Random Forest is an ensemble learning that parallel builds multiple random decision trees, and the prediction is based on the most voted by the trees. Random forest required more training time compared to other machine learning techniques. Conditional random field (CRF) is an undirected graphical model, and it has high performance on text and high dimensional data. CRF builds an observation sequence and is modelled based on conditional probability. CRF is computationally complex in model training due to high data dimensionality, and the trained mode cannot work with unseen data. Semi-supervised is one type of supervised learning that leverages when there is a small portion of labelled with a large portion of unlabelled data. Clustering technique was used to find if there is more than one labelled cluster or to handle the data in labelled and unlabelled clusters (Kowsari et al., 2019).

Deep learning-based text classification. Deep learning-based models are more advanced than machine learning-based models in text classification. There are some limitations in using machine learning approaches which are dependency on the manual feature extraction and necessity of domain knowledge. By using deep learning, that is, neural approaches are able to embed machine learning models and map text into low-dimensional feature vectors without manual feature extraction (Minaee et al. 2021).

Some deep learning techniques are introduced and tabulated in Table 2. Feed-forward neural network converts the bag of words from the text to a vector representation of words and passes it through multiple feed-forward layers. The prediction is done by the final layer machine learning classifier. Recurrent neural networks (RNN) convert text into a sequence of words. It is designed to get the dependency between the word and the structure of the text. The most popular architecture of RNN is long short-term memory (LSTM) in tree structure, word relation and document topic. RNNs capture the pattern in the time dimension, while convolutional neural networks (CNN) capture the pattern in the space dimension. CNN works well in long-range semantic comprehension and detects the local and position-defined pattern. The model generates a feature map of sentences by using *k*-max-pooling to obtain the short and long relationship between words and phrases. Capsule neural network (CapsNets) view the capsule as a group of neurons that have different attributes of an entity. The vector has the magnitude to represent the probability of the entity and the director to represent the entity. Attention-based models can interpret the importance weights of each vector and predict the target based on the attention vector. Memory-augmented networks are extended from an attention model with external memory to maintain the understanding of input text by read, compose and write operation on it. Graph neural networks construct a graph structure of natural language, such as syntactic (Minaee et al., 2021).

Related works on text classification. There are some authors who have done text analysis and text classification on the topic of harassment. The comparison of the data source, feature extraction technique, modelling techniques, and the result is tabulated in Table 3.

Rezvan et al. (2020) conducted a comprehensive study to distinguish between five types of harassment: sexual, racial, appearance-related, intellectual, and political. To do this, they annotated a corpus of 24,000 posts from Twitter with the types of

Table 2 Idea and limitation of deep learning-based text classification.

Deep learning technique	Idea
Feed-forward neural network	Use a bag of words and core with multiple feed-forward layers and a machine learning model at the final layers
Recurrent neural networks (RNN)	Use a sequence of words to capture the dependency between the word and the structure of the text in the time dimension
Convolutional neural network (CNN)	Capture the relationship between the word and phrases in the space dimension
Capsule neural network (CapsNets)	Use a vector to capture the probability and the entity
Attention-based model	Prediction based on the importance of each weight
Memory-augmented networks	Extend the attention-based model with external memory to improve the understanding of text
Graph neural networks	Capture the relationship between text units in the node-edge graph
Siamese neural networks	Use text matching in question-answering
Unsupervised transformer-based pre-trained language models	Use an encoder to covert the sentences with similar semantic and syntactic properties

Table 3 Comparison of related works on text classification.

Title	Author	Data source	Methodology (algorithm)	Evaluation metric	Outcome/result
Analysing and learning the language for different types of harassment	Rezvan et al. (2020)	Total 24,189 posts from Twitter with 3119 harassing posts.	Classifier using SVM, KNN, GBM, NB. Focus on GBM with different feature settings (TFIDF, fastText and word2vec)	Precision, Recall, F-score, accuracy, specificity	GBM classifier outperformed
Understanding street harassment of children: identifying recurrent behaviours in a corpus of young people's accounts of harassment.	Wright et al. (2017)	115 reports from a web-based application	Part-of-speech	–	Analysis of victim's action before and after the street harassment
Detection of harassment on the web	Yin et al. (2009)	Text from online game platforms, social networking sites and technology forums. 167 harassing posts out of 11,051 posts	N-gram, foul language detection using representation dictionary, TFIDF	Precision, Recall, F-measure, hybrid	Hybrid model outperformed

harassment and then used linguistic analysis and statistical distribution of unigrams to analyse the annotated corpus. The 2D visualization showed the top 25 most frequent in each type. Subsequently, they built a type-aware classification model to identify the type-specific harassment by employing state-of-the-art methods to detect the harassing language and then constructing a multi-class classifier to predict the type of harassment.

Wright et al. (2017) also employed a corpus linguistic method to analyse patterns in children's descriptions of street harassment experienced. To do this, they collected children's reports of street harassment from web-based applications and extracted comments from these reports, which were stored in plain text (.txt) files. They focused on analysing behaviour and actions by identifying all verbs in the corpus using AntConc, a corpus analysis toolkit for text analysis. These 137 different verbs were manually categorized based on types of harassment such as verbal interaction, non-verbal interaction, physicality, etc.

Yin et al. (2009) proposed a supersized learning approach for detecting online harassment. To this end, they collected a dataset of 1946 posts from an online website and manually labelled them, with 65 posts being identified as harassment related. Three models were built to capture the content, sentiment, and contextual features of the data. Content features were extracted using Term Frequency/Inverse Document Frequency (TFIDF) to identify significant terms in each post. Sentiment features were derived from the grouping of second-person pronouns such as 'you',

which could be used to form a harassment format. Contextual features were also included to distinguish between posts that had a harassment-like quality. The similarity of these features was then computed to detect potential cases of online harassment. Finally, a hybrid model was constructed by combining the three models and its performance was compared against the individual models.

Sentiment and emotion analysis techniques. Sentiment analysis is an NLP technique to capture the positive, negative, or neutral attitude from a text such as a review of a product. Sentiment analysis is important in the analysis of public opinion related to certain topics in social media (Behera et al., 2021; Suhendra et al., 2022). Emotion recognition in text documents can be performed using NLP techniques. People convey different emotions to give responses and reactions according to different circumstances. Emotion detection has been proven to be beneficial in identifying criminal motivations and psychosocial interventions (Guo, 2022). Sentiment and emotions can be classified based on the domain knowledge and context using NLP techniques, including statistics, machine learning and deep learning approaches.

There are three types of procedures, which are supervised method, lexicon-based method, and semantic based method. Supervised method predicts the sentiment based on the sentiment-labelled dataset. Text classification techniques such as machine learning and deep learning approaches with suitable feature engineering can perform supervised sentiment

Table 4 Comparison of supervised and lexicon-based methods in sentiment analysis.

Method	Supervised	Lexicon
Idea	Prediction based on the sentiment-labelled dataset	Prediction based on the overall sentiment score of the words and phrases
Approaches	Machine learning with feature engineering technique	SentWordNet, MPQA subjectivity lexicon, LIWC lexicon
Advantages	Domain-specific due to trained based on a certain topic	Built-in dictionaries with its sentiment
Limitation	Required large amount of labelled data	Lack of consideration of semantic

Table 5 Comparison of related works on sentiment and emotion analysis.

Title	Author	Data source	Methodology (algorithm)	Evaluation metric	Outcome/result
Sentiment analysis-based sexual harassment detection using machine learning techniques	Alawneh et al. (2021)	Post form online reporting system	RF, MNB, SVC, Linear SVC, SGD, BNB, DT and KNN	Accuracy Precision Recall F1-Score	SGD and linear SVC are outperformed with around 80% accuracy
Sentiment analysis and emotion detection on cryptocurrency-related tweets using ensemble LSTM-GRU model	Aslam et al. (2022)	40,000 posts on Twitter	TextBlob, Text2Emotion. RF, DT, KNN, SVM, GNB, and LR; LSTM, GRU, LSTM-GRU	Accuracy Precision Recall F1-Score G mean	LSTM-GRU outperforms with 0.99 accuracy in sentiment analysis and 0.91 in emotion detection

classification. Lexicon-based sentiment method predicts the sentiment using a built-in dictionary that has been given sentiment orientation. The sematic-based method makes predictions based on the evaluation of conceptual semantic and contextual semantics by co-occurrence patterns of words in a text. The semantic network and word clustering are the external semantic knowledge that aids the prediction of sentiment by the captured semantic relationship. Semantic networks represent the words to convey sentiment, while WordNet exploits the ontological structure. (Behera et al., 2021). The comparison between supervised and lexicon-based procedures is tabulated in Table 4.

Related works on sentiment and emotion analysis techniques. There are some authors have done sentiment and emotion analysis on text using machine learning and deep learning techniques. The comparison of the data source, feature extraction technique, modelling techniques, and the result is tabulated in Table 5.

Alawneh et al. (2021) performed sentiment analysis-based sexual harassment detection using the Machine Learning technique. The collected data from the online reporting system had been labelled. The data was pre-processed using lower casing, removal of punctuation, stop-word removal, common word removal, rare word removal, spelling correction, emoji removal, emoticon's removal, removal of HTML tags, tokenization, stemming and lemmatization, and term frequency-inverse document frequency. They performed 8 classifiers which are Random Forest, Multinomial NB, SVC, Linear SVC, SGD, Bernoulli NB, Decision tree and *K* Neighbours.

Aslam et al. (2022) performed sentiment analysis and emotion detection on tweets related to cryptocurrency. TextBlob libraries are used to annotate sentiment, and Text2emotion is used to detect emotions such as angry, fear, happy, sad and surprise. They use different settings of feature extraction, which are Bag-of-word, TF-IDF and Word2Vec. They build several machine learning classifiers and deep learning classifiers using the neural network LSTM and GRU. The deep learning model of LSTM-GRU outperforms all other models.

Method
This study presents two models that have been developed to address the issue of sexual harassment. The first model is a machine learning model which is capable of accurately classifying different types of sexual harassment. The second model, which leverages a deep learning approach, is used to classify sentiment and emotion. To ensure the accuracy of the models, a comprehensive text pre-processing process was applied to the text data. Subsequently, data preparation, modelling, evaluation, and visualization phases were conducted for each model in order to assess their performance. The framework of this study is illustrated in Fig. 1 and provides an overview of the entire process, from data pre-processing to visualization. Furthermore, this framework can be used as a reference for future studies on sexual harassment classification.

Data source. The 12 novels are electronic books that are written in the context of the Middle East. The summary of the 12 novels is tabulated in Table 6

Text preparation. First, the e-pub and pdf e-books are converted and exported into text format. After that, some text pre-processing techniques, which are sentence tokenization, expanding contraction, POS tagging, word tokenization, lower case conversion, stop word removal, and lemmatization are performed to extract the meaningful data in text. The counts of the sentences, words, and vocabulary are summarized in Table 7. The text pre-processing steps are illustrated in Fig. 2.

Format conversion. Python libraries are used for format conversion. EbookLib is used for epub to txt format conversion, while PyPDF2 is used for pdf to txt format conversion. The ebook is converted into a separate txt file.

Sentences tokenization. Natural Language Toolkit (NLTK), a popular Python library for NLP, is used for text pre-processing. The separated txt files are imported, and the raw text is sentence tokenized. The sentences are compiled into a csv file and recorded with the filename. Table 8 shows the sample of the sentence after sentence tokenization.

Expanding contraction. A Python library named contractions is used to expand the shortened words in sentences. For example, 'you're' is fixed to you 'are'. Expanding contractions are done to aid the recognition of grammatical categories in POS tagging.
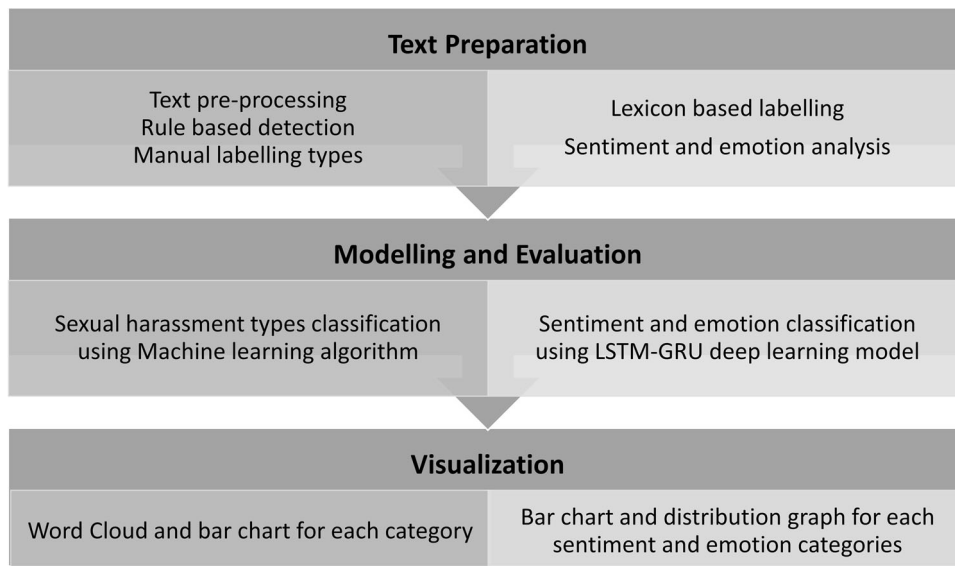
Text Preparation

| Text pre-processing Rule based detection Manual labelling types | Lexicon based labelling Sentiment and emotion analysis |

Modelling and Evaluation

| Sexual harassment types classification using Machine learning algorithm | Sentiment and emotion classification using LSTM-GRU deep learning model |

Visualization

| Word Cloud and bar chart for each category | Bar chart and distribution graph for each sentiment and emotion categories |

Fig. 1 Framework of study of sexual harassment in Anglophone literature.

Table 6 Details of data source, which are 12 Anglophone novels.

Filename	Title	Author	Published year
Balcony	Balcony on the Moon: Coming of Age in Palestine	Ibtisam Barakat	2016
Beekeeper	The Beekeeper_ Saving the Stolen Women of Iraq	Dunya Mikhail, Max Weiss	2018
City	City of Lies: Love, Sex, Death, and the Search for Truth in Tehran	Ramita Navai	2014
Diary	IraqiGirl: Diary of a Teenage Girl in Iraq	IraqiGirl	2009
Fundamentalist	The Reluctant Fundamentalist: From Book to Film	Mohsin Hamid	2013
Journey	Nujeen: One Girl's Incredible Journey from War-torn Syria in a Wheelchair	Nujeen Mustafa, Christina Lamb	2016
Memoir	Reading Lolita in Tehran: A Memoir in Books	Azar Nafísi	2003
Mountains	And the Mountains Echoed	Khaled Hosseini	2012
Psychoanalysis	Doing Psychoanalysis in Tehran	Gohar Homayounpour	2012
Rain	Rain Over Baghdad	Hala El Badry	2013
Shadow	The Shadow of the Crescent Moon	Fatima Bhutto	2013
Worlds	Between Two Worlds: Escape from Tyranny: Growing Up in the Shadow of Saddam	Zainab Salbi, Laurie Becklund	2005

Table 7 Count of sentences, words, and vocabulary in 12 novels.

Filename	Count of sentences	Count of words	Count of vocabulary
Balcony	3086	54,636	4929
Beekeeper	3828	62,454	4454
City	5418	93,857	9274
Diary	3163	55,501	4712
Fundamentalist	1980	43,495	5691
Journey	3521	69,351	5734
Memoir	7806	139,416	9293
Mountain	9634	124,405	9822
Psychoanalysis	1008	35,354	4484
Rain	9176	173,853	9120
Shadow	3914	62,071	6562
Worlds	5924	107,827	7300

Part-of-speech (POS) tagging. POS taggers process a sequence of words and attach a part of a speech tag to each word. For example, NN is a noun, VD is a verb, JJ is an adjective, and IN is a preposition. The meaningful words, which are verbs, nouns, and adjectives, are intended to be extracted to reduce the redundancy of words in the text. In this section, the word with a tag that starts with 'V', 'N', and 'J' is extracted.

Word tokenization. The word tokenization with defined regression expression is used to extract only word that only consists of alphabetical characters.

Lower case conversion. All the extracted word is converted into lowercase to reduce the duplicated vocab. This also aids the stop word removal when compared with a defined dictionary.

Stop word removal. The English stop words such as 'the', 'is', 'he' and 'has' do not give much meaning for text exploration and are removed from the text. Besides, some of the customized words removed, which are the high frequent verbs in the novel such as "said", "feel", "know", "come" and the name of the characters.

Lemmatization. This technique removes the affixes of words that are in its dictionary. Lemmatization results in a list of root words (lemmas) to remove redundant words in text. For example, lemmatize 'plays' to 'play'.

Text classification. The goal of text classification is to classify the types of sexual harassment. First, the sentences that contain
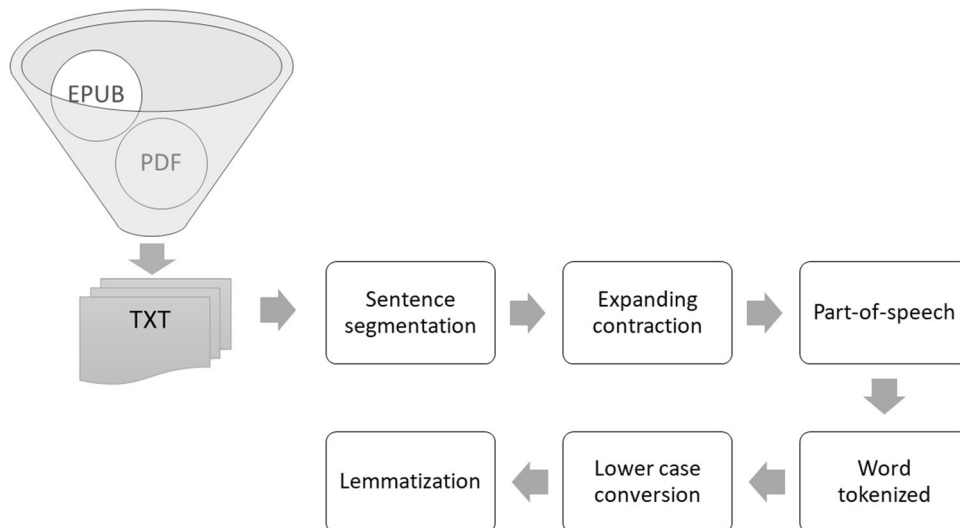
Fig. 2 Flow chart of text pre-processing.

Table 8 Sample of sentence after sentence tokenization.

Sentence	Filename
Mother had to fill the baboor with kerosene over and over and warned us repeatedly to stay away from it.	Balcony
But children playing in the house, especially our young relatives when visiting, sometimes knocked off a boiling pot of food.	Balcony
My right leg has a scar from a burn from that baboor.	Balcony
For the first time, we have hot and cold running water.	Balcony
This makes Mother dance as she turns the faucet on and off.	Balcony
Dancing is the way she celebrates.	Balcony

sexual harassment words are rule-based detected. A published harassment corpus created by Rezvan et al. (2020) has 452 words that related to sexual harassment are used to matching the words in the tokenized sentences. The matched sexual harassment-related words are extracted for each sentence. After that, the 570 sexual harassment-related words are reviewed to determine whether it is conceptually related to sexual harassment. From the sexual harassment sentences, the types of sexual harassment are manually labelled. The first type of label is the sexual harassment type, it has labels which are gender harassment, unwanted sexual attention, and sexual coercion. The second type of label is the sexual offence type, which has labels that are physical and non-physical. The manual label process is illustrated in Fig. 3.

There are only nearly 0.1% of sentences (570 out of 58,458) are detected as containing sexual harassment-related words. Of the 570 sentences, there is 23% which is 108 sentences that are conceptually related to sexual harassment. Besides, there are 65 and 43 sentences are physical and non-physical sexual harassment, respectively.

Table 9 presents the sentences that have been labelled as containing sexually harassing words, along with the corresponding keywords detected through a rule-based approach. For instance, in the first sentence, the word 'raped' is identified as a sexual word. This sentence describes a physical sexual offense involving coercion between the victim and the harasser, who demands sexual favours from the victim. As a result, this sentence is categorized as containing sexual harassment content. Similarly, the second and third sentences also describe instances of sexual harassment. In these cases, the harasser exposes the victim to

pornography and uses vulgar language to refer to them, resulting in unwanted sexual attention. On the other hand, the last three sentences contain sexual words but do not convey any sexual harassment content. For example, the keyword 'fear' is used to describe death, 'porn' refers to a career contextually unrelated to explicit material, and 'destroy' pertains to damaging dishes. Therefore, manual interpretation plays a crucial role in accurately identifying sentences that truly contain sexual harassment content and avoiding any exceptions.

Furthermore, while rule-based detection methods facilitate the identification of sentences containing sexual harassment words, they do not guarantee that these sentences conceptually convey instances of sexual harassment. Henceforth manual interpretation remains essential for accurately determining which sentences involve actual instances of sexual harassment. The distribution of sentences based on different types of sexual harassment and types of sexual offenses can be observed in Fig. 4.

Additionally noteworthy is that, on average, each sentence consists of ~12 words. A selection of cleaned sample sentences can be found in Table 10.

By using POS tagging, the nouns and verbs are extracted and shown in Figs. 5 and 6. From Fig. 5, the most frequent nouns in sexual harassment sentences are fear, Lolita, rape, women, family and so on. The fear of victims, especially women, can be noticed from the word cloud. From Fig. 6, the most frequent verbs are raped, fear, said, see, know, and so on. The sexual harassment behaviour such as rape, verbal and non-verbal activity, can be noticed in the word cloud.

For the second model, the dataset consists of 65 instances with the label 'Physical' and 43 instances with the label 'Non-physical. The feature engineering technique, the Term Frequency/ Inverse Document Frequency (TFIDF) is applied. After that, the Principal Component Analysis (PCA) is applied for dimensionality reduction. The 108 instances are then split into train dataset and test dataset, where 30% of the dataset is used for testing the performance of the model.

Six machine learning algorithms were utilized to construct the text classification models in this study. These algorithms include *K*-nearest neighbour (KNN), logistic regression (LR), random forest (RF), multinomial naïve Bayes (MNB), stochastic gradient descent (SGD), and support vector classification (SVC). Each algorithm was built with basic parameters to establish a baseline performance. To identify the most suitable models for predicting sexual harassment types in this context, various machine learning
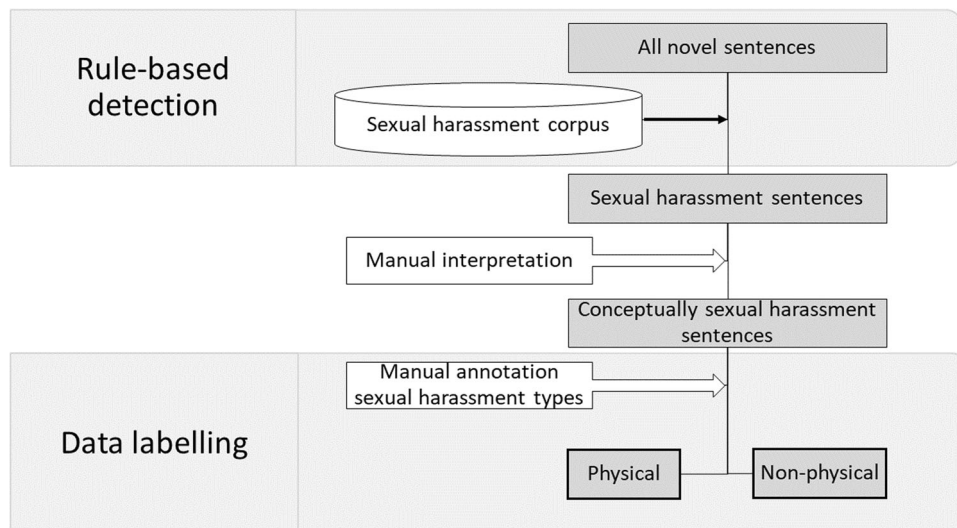
Fig. 3 Manual process of data labelling for sexual harassment types.

| Table 9 Sample sentences of sexual harassment with contain sexual harassment words.		
Sentences	**isHarassing**	**Sexual words**
He bound my hands and feet and then raped me.	Yes	raped
Taymour played his porn collection for Leyla.	Yes	porn
He describes Lolita as a vulgar vixen-"a disgustingly conventional little girl," he calls her.	Yes	lolita
And it is not because I fear death.	No	fear
And so began Leyla's career as a porn actress.	No	porn
The dish was destroyed a few hours later by a policeman and two young basijis, who warned the owner that next time they would destroy him too.	No	destroy
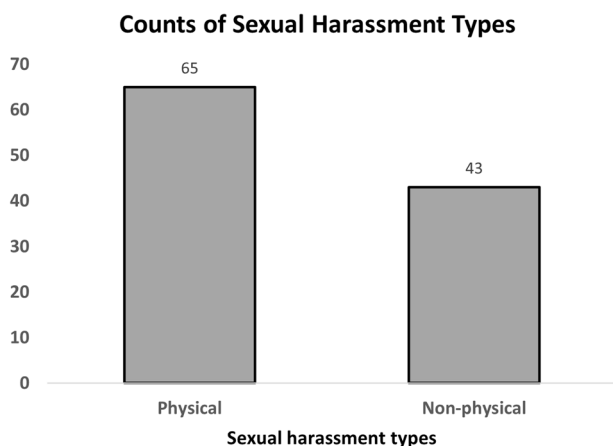
Fig. 4 Counts of physical and non-physical sexual harassment sentences.

techniques were employed. These techniques encompassed statistical models, optimization methods, and boosting approaches. For instance, the KNN algorithm predicted based on sentence similarity and the k number of nearest sentences. LR and MNB are statistical models that make predictions by considering the probability of class based on a decision boundary and the frequency of words in sentences, respectively. Similarly, LR and SVC employed a boundary to predict the class using a features map of words. SGD served as an optimization method that enhanced classifier performance for SVC and LR models. RF utilized a boosting technique by combining multiple decision trees and making predictions based on the voting results from

each tree. Following model construction, hyperparameters were fine-tuned using GridSearchCV. This method systematically searched for optimal hyperparameters within subsets of the hyperparameter space to achieve the best model performance. The specific subset of hyperparameters for each algorithm is presented in Table 11.

Sentiment and emotion analysis. The goal of the sentiment and emotion analysis is to explore and classify the sentiment characteristics that induce sexual harassment. The lexicon-based sentiment and emotion analysis are leveraged to explore the sentiment and emotion of the type of sexual offence. The data preparation to classify the sentiment is done by text pre-processing and label encoding.

Lexicon-based sentiment analysis. A NLTK's pre-trained sentiment analyser is applied to estimate the sentiment of the sexual harassment sentence. The result provides the sentiment of positive, negative, neutral, and compound. The compound sentiment is then encoded into 'negative' where the value is less than zero, 'positive' where the value is more than zero, and 'neutral' where the value is zero.

Lexicon-based emotion analysis. Text2emotion, a Python package, is used to extract the emotion of the sentences. The package analyses five types of emotion from the sentences which are happy, angry, surprise, sad, and fear. The value of each emotion is encoded to 'True', where the value is more than zero, and 'False', where the value is equal to zero. The highest score among the five emotions is recorded as the label of emotion in the sentences.

Table 10 Sample of cleaned sentences.

Sentences	Cleaned sentence
But the idea that they could leave each other, or one of them leave us, is beginning to destroy even the way I write my name.	idea beginning to destroy the way write name
But I think they are wrong: all they will have been a different life that is more of a prison with harder labour.	Wrong different life prison harder labour
My grandmother could never imagine that the merciless ones would destroy her own grave.	Grandmother imagines merciless one destroys grave
It would become much narrower when the offense is rape or murder, for example.	Become narrower offense rape murder example
When they told us it was time for "Quran lessons", this also meant that they were also going to rape us because they typically did that right after prayers.	Quran lesson meant rape right prayer

Fig. 5 Word cloud of nouns in sexual harassment sentences.

Fig. 6 Word cloud of verbs in sexual harassment sentences.

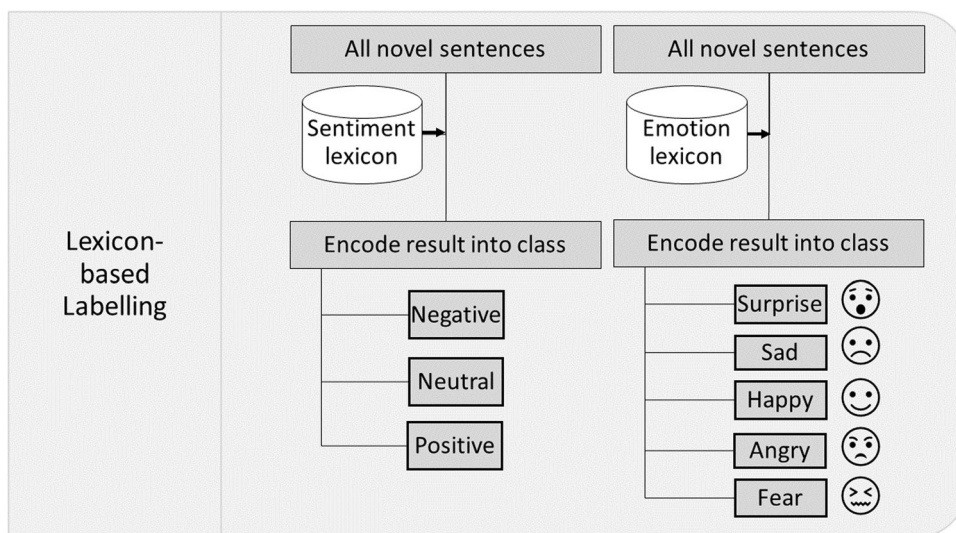The 58,458 sentences with the sentiment and emotion categories are prepared for sentiment classification and emotion detection. The flow of data preparation for sentiment and emotion classification is shown in Fig. 7. The data description of the data prepared for text classification to classify sentiment is tabulated in Table 12.

According to Aslam et al. (2022), a deep learning-based ensemble model will be constructed for sentiment analysis and emotion detection, utilizing LSTM-GRU, which is an ensemble of LSTM and GRU sequential recurrent neural networks (RNN). RNNs are a type of artificial neural network that excels in handling sequential or temporal data. In the case of text data, RNNs convert the text into a sequence, enabling them to capture the relationship between words and the structure of the text. The output of an RNN is dependent on its previous element, allowing it to consider context. LSTM, a widely used architecture for RNNs, is capable of capturing long-term dependencies and influencing current predictions. Additionally, GRU serves as an RNN layer that addresses the issue of short-term memory while utilizing fewer memory resources. By combining both LSTM and GRU in an ensemble model, the objective is to enhance long-term dependency modelling and improve accuracy. The ensemble

Table 11 Subset of hyperparameter for model 2 to classify the types of sexual harassment.

Algorithm	params
K-nearest neighbours (KNN)	{'n_neighbors': [3, 5, 7, 9, 11, 13], 'weights': ['uniform', 'distance'], 'p': [1, 2]}
Logistic regression (LR)	{'C': array([2.91545190e−03, 3.77964473e−01, 4.90000000e + 01]), 'penalty': ['l1', 'l2', 'elasticnet'], 'dual': [True, False], 'solver': ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga']}
Random forest (RF)	{'criterion': ['gini', 'entropy'], 'max_features': ['sqrt', 'log2'], 'n_estimators': [1, 10, 50, 100], 'max_depth': [2, 3, 5], 'random_state': [0, 12, 24]}
Multinomial naive Bayes (MNB)	{'clf__alpha': (0.01, 0.001)}
Stochastic gradient descent (SGD)	{'alpha': (0.01, 0.001), 'loss': ['hinge', 'log_loss', 'log'], 'penalty': ['l2', 'l1', 'elasticnet'], 'random_state': [0, 24, 42]}
Support vector classification (SVC)	{'clf__C': [1, 10], 'clf__kernel': ('linear', 'rbf')}

Fig. 7 Flow chart of data preparation for sentiment and emotion classification.

Table 12 Data description of the data prepared for text classification to classify sentiment.

Column	Level
Sentence	–
Sentiment	Negative (15,354)
	Positive (18,653)
	Neutral (24,451)
Emotion	happy (22,906)
	sad (9945)
	surprise (8699)
	fear (12,515)
	angry (4394)

model consists of an LSTM layer followed by a GRU layer, where the output from LSTM serves as input for GRU. The architecture of this LSTM-GRU ensemble model can be observed in Fig. 8.

The first layer of LSTM-GRU is an embedding layer with m number of vocab and n output dimension. The next layer is LSTM with 128 units, it produces a significant feature sequence as the input of the GRU layer. A dropout layer is followed by the LSTM to reduce the complexity of the ensemble model. The next layer is the GRU layer, with 64 units. A dense layer with 16 neurons is added to overcome the sparsity of GRU's output. An output layer which is the 3 neurons dense layer, is added for sentiment classification, and 5 neurons dense layer is added for emotion detection, respectively. The loss function of 'categorical_crossentropy' and the 'adam optimizer' is used for training. The model is then trained using 100 epochs. The optimal number of epochs is recorded with its performance.

Results and discussion

Text classification. To achieve the objective of classifying the types of sexual harassment within the corpus, two text classification models are built to achieve the goals respectively. For sexual harassment types of classification, the goal is to classify conceptually sexual harassment into physical and non-physical sexual offence.

Table 13 shows the sentences with physical and non-physical sexual harassment. For physical sexual harassment, the action taken by the sexual harasser is having physical contact with the victim's body, such as rape, push, and beat. For non-physical, the actions are unwanted sexual attention and verbal behaviour such as expressing sexual words such as "fuck" and "bastard".

There are six machine learning algorithms are leveraged to build the text classification models. *K*-nearest neighbour (KNN), logistic regression (LR), random forest (RF), multinomial naïve Bayes (MNB), stochastic gradient descent (SGD) and support vector classification (SVC) are built.

The models utilized in this study were constructed using various algorithms, incorporating the optimal parameters for each algorithm. The evaluation of model performance was based on several metrics, including accuracy, precision, recall, and F1. These metrics are presented in a tabular format in Table 14. Accuracy, precision, recall, and F1 are commonly employed to assess the performance of classification models. Accuracy serves as a measure of the proportion of correct predictions out of the total predictions made by the model. Precision and recall provide more nuanced evaluations of classification models. Precision represents the ratio of true positive predictions to all predicted positive instances, while recall denotes the ratio of true positive predictions to all actual positive instances. F1 is a composite

metric that combines precision and recall using their harmonic mean. In the context of classifying sexual harassment types, accuracy can be considered as the primary performance metric due to the balanced sample size and binary nature of this classification task. Additionally, precision, recall, and F1 can be utilized as supplementary metrics to support and provide further insights into model performance. As shown in Table 14, Logistic regression (LR) gained higher accuracy in compared to other algorithms.

The model using Logistic regression (LR) outperformed compared to the other five algorithms, where the accuracy is 75.8%. Stochastic gradient descent (SGD) and K-nearest neighbour (KNN) and had performed, followed by LR, which has 66.7% and 63.6% of accuracy. The other three models' performance accuracy is 60.6%.

However, its low recall for physical sexual harassment results in an F1 score of 60%, which represents the harmonic mean of precision and recall. Conversely, LR performs better in predicting non-physical sexual harassment ('No') compared to physical sexual harassment. This is evident from its high precision and recall values, leading to an F1 score of 82.6%.

Among the six models considered, both K-nearest neighbours (KNN) and stochastic gradient descent (SDG) exhibit superior performance. In contrast, random forest (RF), multinomial naive Bayes (MNB), and support vector classification (SVC) models are unable to effectively predict instances of physical sexual harassment ('Yes'), as indicated by their precision, recall, and F1 scores being zero. Comparing SDG and KNN, SDG outperforms KNN due to its higher accuracy and strong predictive capabilities for both physical and non-physical sexual harassment.

The training of machine learning models is crucial for their performance. However, the current train set consists of only 70 sentences, which is relatively small. This limited size can make the model sensitive and prone to overfitting, especially considering the presence of highly frequent words like 'rape' and 'fear' in both classes. Overfitting occurs when a model becomes too specialized in the training data and fails to generalize well to unseen data. To address these issues, it is recommended to increase the sample size by including more diverse and distinct samples in each class. A larger sample size helps to capture a wider range of patterns and reduces the risk of overfitting. Additionally, incorporating more varied samples can help mitigate the sensitivity caused by high-frequency words. Furthermore, it is important to consider the limitations of training models in a specific context, such as sexual harassment in Middle East countries. Models trained on such data may not perform as expected when applied to datasets from different contexts, such as anglophone literature from another region. This discrepancy arises due to cultural and social differences that influence language usage and interpretation. To enhance model performance across different contexts, it is advisable to train models on datasets that encompass a broader range of cultural backgrounds and social interactions. This approach ensures that the model learns more generalized patterns rather than being biased towards specific contexts.

Sentiment and emotion analysis
Lexicon-based sentiment analysis. Lexicon-based sentiment analysis was done on the 108 sentences that have sexual harassing content. The count of the category of sentiment is shown in Fig. 9. The histogram and the density plot of the numerical value of the compound sentiment by the sexual offense type are plotted in Fig. 10.

The sample of the sentences in each sentiment category is shown in Table 15. From the bar chart in Fig. 9, there is 78.7% of sentences had been categorized as negative sentiment. This shows that sexual harassment is mostly related to a negative sentiment. From the histogram in Fig. 10, the distribution of compound scores is different between the two types of sexual harassment. Most of the sentences with physical sexual harassment content has nearly a maximum of negative sentiment. For non-physical sentiment, the sentences are quite evenly distributed. This gives the insights that the physical sexual harassment may be impactfully to the effect the sentiment negatively compared to the non-physical sexual harassment.

Lexicon-based emotion analysis. Lexicon-based emotion analysis had been done on the 108 sentences. The sentences are categories multi-label with 5 emotions which are happy, angry, surprise, sad
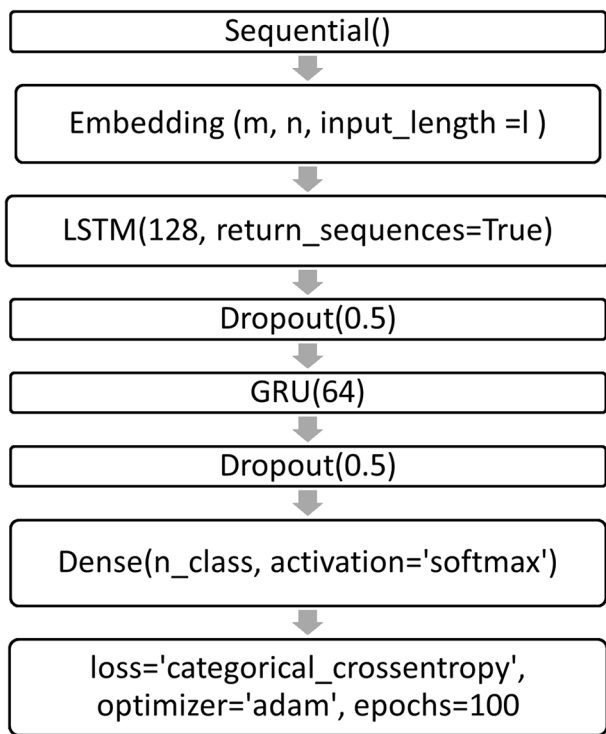
Fig. 8 Architecture of LSTM-GRU deep learning model.

The architecture diagram shows the following sequence:
- Sequential()
- Embedding (m, n, input_length =l)
- LSTM(128, return_sequences=True)
- Dropout(0.5)
- GRU(64)
- Dropout(0.5)
- Dense(n_class, activation='softmax')
- loss='categorical_crossentropy', optimizer='adam', epochs=100

Table 13 Sample sentences of sexual harassment with label sexual offence types.		
Sentences	**Sexual offence types**	**Sexual words**
Some Serbian soldiers were ordered to rape to prove their manhood and machismo.	Physical	rape
Instead of protecting the women, the police started to disperse us, pushing some with the butt of their guns and ordering the "sisters" to make no trouble and go home.	Physical	butt
They beat us, they force us to pray, and they force us to serve them food in porno outfits.	Physical	porno
'You mother-fucking sister-pimping bastard cunt!' Haj Agha blinked.	Non-physical	bastard
Humbert pins Lolita by first naming her, a name that becomes the echo of his desires.	Non-physical	Lolita
Fuck you, fuck you, he would say as I prayed for morning.	Non-physical	Fuck

Table 14 Performance metric of accuracy, precision, recall and F1 of sexual harassment types classification.

Algorithm	Accuracy	Class	Precision	Recall	F1
K-nearest neighbours (KNN)	0.636	Yes	0.667	0.154	0.250
		No	0.633	0.950	0.760
Logistic regression (LR)	**0.758**	Yes	0.857	0.462	0.600
		No	0.731	0.950	0.826
Random forest (RF)	0.606	Yes	0	0	0
		No	0.606	1.000	0.755
Multinomial naive Bayes (MNB)	0.606	Yes	0	0	0
		No	0.606	1.000	0.755
Stochastic gradient descent (SGD)	0.667	Yes	0.583	0.538	0.560
		No	0.714	0.750	0.732
Support vector classification (SVC)	0.606	Yes	0	0	0
		No	0.606	1.000	0.755

Fig. 9 Bar chart of count of sentiment categories.

Fig. 10 Histogram and density plot of the numeric value of the compound sentiment by sexual offence types.

and fear. The count of each emotion is shown in Fig. 11. The histogram and the density plot of the numerical value of each emotion by the sexual offence type are plotted in Fig. 12.

The sample of the sentences for each emotion category is shown in Table 16. From the bar chart in Fig. 11, there are nearly 50% of sentences have the emotion of Fear and Surprise. There are only 16.7% of sentences have angry emotion. From the histogram in Fig. 12, the distribution of the five emotion scores does not have much difference between the two types of sexual harassment. However, the most significant observation is the distribution of Fear emotion, where there is a higher distribution of physical sexual harassment than the non-physical sexual harassment sentences at the right side of the chart. This gives the insight that physical sexual harassment contributed to more fear emotion compared to non-physical sexual harassment.

The insights from the sentiment and emotion analysis are:

1. Most of sexual harassment is related to negative sentiment, whereas physical sexual harassment contributes to a higher intensity of negative sentiment compared to non-physical sexual harassment.
2. Fear and surprise emotion both presented in sexual harassment higher compared to sad, happy, and angry.
3. The physical sexual harassment has distributed in a higher score range of fear emotion compared to non-physical sexual harassment.

Sentiment classification. For the sentiment classification, a deep learning model LSTM-GRU, an LSTM ensemble with GRU Recurrent neural network (RNN) had been leveraged to classify the sentiment analysis. There are about 60,000 sentences in which the labels of positive, neutral, and negative are used to train the model. The count of the label is shown in Fig. 13.

The sequential model is built, and its architecture of the model is demonstrated in Fig. 14. The model starts with a Glove word embedding as the embedding layer and is followed by the LSTM and GRU layers. There is a dropout layer was added for LSTM and GRU, respectively, to reduce the complexity. Finally, the output is a 3-neuron dense layer to classify the 3 labels. The model had been trained using 20 epochs and the history of the accuracy and loss had been plotted and shown in Fig. 15. There is a first peak of validation accuracy at around 3 epochs. To avoid overfitting, the 3 epochs were chosen as the final model, where the prediction accuracy is 84.5%.
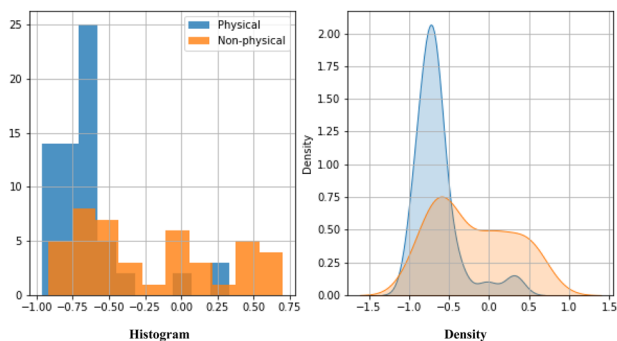
Emotion classification. The same dataset, which has about 60,000 sentences with the label of highest-scored emotion, is used to train the emotion classification. The count of the label of emotion is shown in Fig. 16.

A deep learning model is built where the architecture is like the sentiment classification, which is an LSTM-GRU model shown in Fig. 17. Instead of a 3 neurons-dense layer as the output layer, a 5 neurons-dense layer to classify the 5 emotions. The model had

Table 15 Sample sentences with labels sentiment category.

Sentences	Compound	Categories
Hawkers stood on the square and whispered super, the Persian word for a porn film, out of the side of their mouths at potential customers.	0.5994	Positive
Are you with the sazman?' 'No fucking way!' The gun-runner was laughing.	0.5411	Positive
Pictures of the goods on sale beside lists of services offered in the 'About me' section: threesomes, anal and lez sex for women.	0	Neutral
'Get that fucking jendeh out of my house.	0	Neutral
Women were being found half-naked and dazed in the middle of the street after being raped and released, and we talked about how vulnerable women were to rape in the midst of the chaos of war.	−0.9661	Negative
It had started after protesters were killed, beaten, and raped after the disputed elections in 2009.	−0.9451	Negative

been trained using 20 epochs, and the history of the accuracy and loss had been plotted and shown in Fig. 18. To avoid overfitting, the 3 epochs had been chosen as the final model where the prediction accuracy of 80.8%.

The accuracy of sentiment and emotion classification was evaluated, and the results are presented in Table 17. In this study, the training set consisted of approximately 60,000 sentences extracted from novels, all of which were labelled using a lexicon-based approach. It is important to acknowledge that there may be potential bias introduced during the data labelling process due to the nature of the dictionary used. Furthermore, it should be noted that the models developed in this study may not be specifically tailored to the topic of sexual harassment, as they were trained on sentences from various novels. The training process itself was time-consuming due to the large sample size involved. To enhance the performance of sentiment and emotion models, it is recommended to employ multiple lexicons or dictionaries for data labelling. Additionally, if enough sexual harassment-related sentences are available and suitable for input into a deep learning model, training solely on such data could potentially yield improved results. Similar to challenges encountered in machine learning models, computational literary studies face difficulties arising from societal diversity resulting from social interactions and activities. Consequently, the trained models developed in this study are expected to provide significant contextual advantages particularly within Middle Eastern countries.

Fig. 11 Bar chart of count of emotion categories.
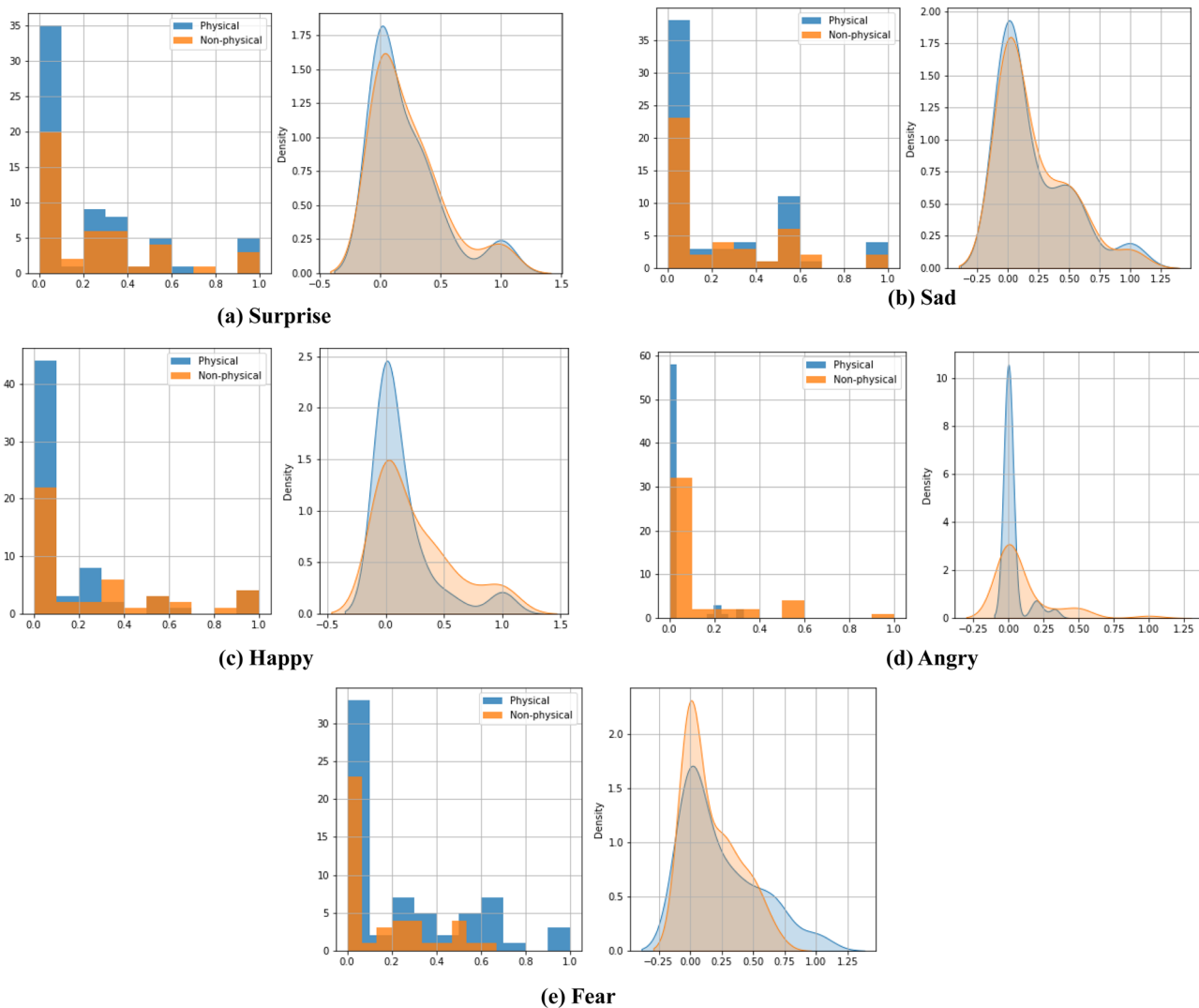
Fig. 12 Histogram and density plot of the numeric value for emotion classification by sexual offence types. a Surprise, **b** Sad, **c** Happy, **d** Angry, and **e** Fear.

Table 16 Sample sentences with labels emotion category.

Sentences	Emotion
If the rapist is a criminal, the woman is judged unmarriageable.	Sad
They ordered one of her sons to rape her and when he refused, they shot him.	Sad
Some Serbian soldiers were ordered to rape to prove their manhood and machismo.	Fear
They raped us there in that assembly hall however they wished.	Fear
I learned in one of my classes what I had never heard at a family gathering: Stalin ordered Russian soldiers to rape as many German women as possible during World War II.	Surprise
The main motivation for these Daesh men was sexual: they would kill anyone in order to rape women.	Surprise
His uncle was a part-time transvestite who wandered the streets in a dress and lipstick; some people thought he was mad and left him alone, others spat on him.	Angry
He describes Lolita as a vulgar vixen-"a disgustingly conventional little girl," he calls her.	Angry
'Get that fucking jendeh out of my house.	Happy
It was obvious really: they were all happy to fuck her and even to be seen with her, but no rich kid with high-society friends would marry a whore.	Happy

Counts of Sentiment Categories
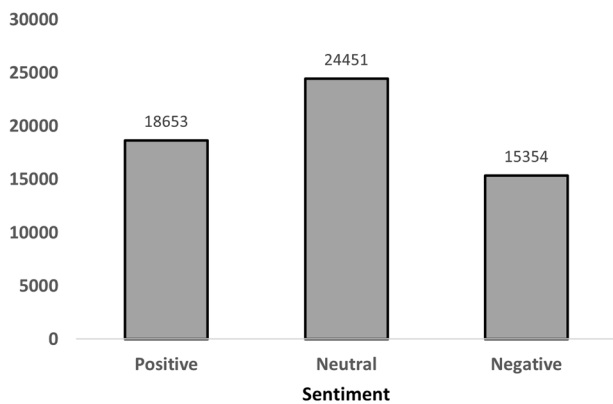
Fig. 13 Count of sentences with the label of sentiment.
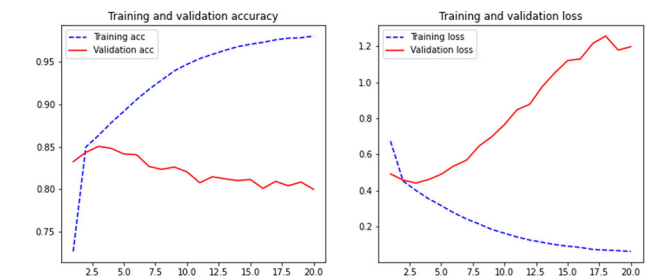
Fig. 15 History of the accuracy and loss of sentiment classification.

Fig. 14 Architecture of the sentiment classification model.

Counts of Emotion Categories

Fig. 16 Count of sentences with the label of emotion.

Conclusion

This study aimed to identify the types of sexual harassment within the text and the sentiment characteristics that induce it. To achieve this, a logistic regression machine learning model was trained to predict the types of sexual harassment, which were physical and non-physical. The model outperformed five other algorithms, achieving an accuracy of 75.8%. Additionally, an LSTM-GRU RNN deep learning model was leveraged to build a sentiment classification with three labels: negative, positive, and neutral. This model achieved an accuracy of 84.5%. Lexicon-based sentiment analysis revealed that most sentences had a negative sentiment, particularly physical sexual harassment, which had a higher intensity of sexual harassment. The results showed that most sexual harassment was related to negative sentiment, with physical sexual harassment contributing to a
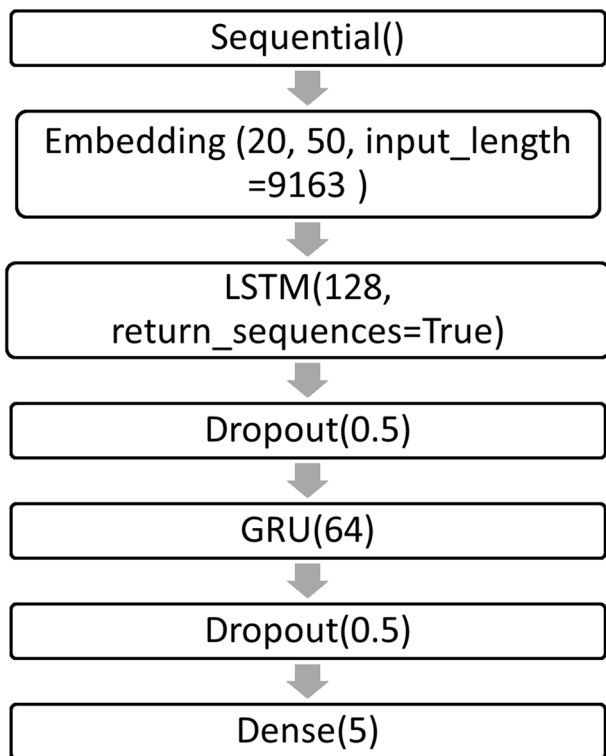
```
Sequential()
        ↓
Embedding (20, 50, input_length
=9163 )
        ↓
LSTM(128,
return_sequences=True)
        ↓
Dropout(0.5)
        ↓
GRU(64)
        ↓
Dropout(0.5)
        ↓
Dense(5)
```
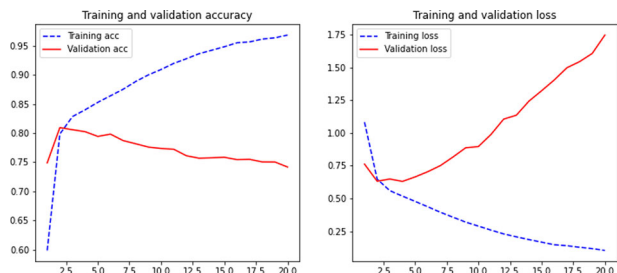
Fig. 17 Architecture of the emotion classification model.

Fig. 18 History of the accuracy and loss of emotion classification.

Table 17 Accuracy of the sentiment and emotion classification.		
	Sentiment classification	**Emotion classification**
Number of neurons in output layer	3	5
Epochs	3	3
Accuracy	84.5%	80.8%

higher intensity than non-physical forms. Fear and surprise emotions were both present in higher scores than sad, happy, and angry emotions in sexual harassment cases; physical sexual harassment had a higher score range for fear emotion compared to non-physical forms.

A comprehensive framework has been developed to facilitate text mining of sexual harassment in the Anglophone literature of the Middle East. The framework incorporates various data science approaches to accurately analyse and classify different types of sexual harassment, both physical and non-physical. A logistic regression model was trained to identify and categorize these types, while an LSTM-GRU RNN model was constructed to determine the sentiment associated with instances of sexual harassment. This domain-specific framework offers several advantages for analysing sexual harassment in the Middle East. It enables fast and precise analysis by leveraging data science techniques. However, it is important to acknowledge that the analysis of sexual harassment may vary when considering global contexts. For example, instances of sexual harassment occurring in religious places may differ from those in other countries outside the Middle East due to variations in religious practices and activities. Moreover, the cultural and patriarchal societies prevalent in the Middle East often result in women being disproportionately victimized by sexual harassment. This framework considers these societal factors, making it particularly suitable for studying sexual harassment cases specific to this region. It is worth noting that this framework is primarily designed for analysing English-language texts. While this is advantageous for studying anglophone literature related to sexual harassment in the Middle East, it may not be applicable to non-English documents commonly found in this region. However, with appropriate modifications and training using documents in other languages, this framework can be adapted for analysing non-English texts as well.

This framework, despite its merits, is not without flaws. One notable drawback lies in the subjective bias introduced by the algorithm. The bias of machine learning models stems from the data preparation phase, where a rule-based algorithm is employed to identify instances of sexual harassment. The accuracy of this process heavily relies on the collection of sexual harassment words used to detect such sentences, thereby influencing the final outcome. Consequently, it becomes imperative to incorporate manual interpretation in order to review and validate the selection of sexual harassment sentences. However, it is important to acknowledge that both manual annotation and computational modelling introduce systematic errors that can lead to bias. To mitigate these defects, a few domain experts should be involved in the manual interpretation process to ensure a more reliable result. Additionally, implementing boosting techniques that combine multiple machine learning models can yield a more robust and accurate outcome by considering the majority vote among these models. Furthermore, enhancing this framework can be achieved by incorporating emotion and sentiment labelling using established dictionaries. This additional layer of analysis can provide deeper insights into the context and tone of the text being analysed. Finally, expanding the size of the datasets used for training these models can significantly improve their performance and accuracy. By exposing them to larger and more diverse datasets, these models can better generalize patterns and nuances present in real-world data.

Data availability
The dataset can be available upon request to any of the authors or the corresponding author Pantea Keikhosrokiani.

References
Abadah MSK, Keikhosrokiani P, Zhao X (2023) Analytics of public reactions to the COVID-19 vaccine on twitter using sentiment analysis and topic modelling. In: Valle-Cruz D, Plata-Cesar N, González-Ruíz JL (eds) Handbook of

Research on Applied Artificial Intelligence and Robotics for Government Processes. IGI Global, pp. 156–188

Al Mamun MH, Keikhosrokiani P, Asl MP, Anuar NA, Hadi NH, Humida T (2022) Sentiment analysis of the Harry Potter series using a lexicon-based approach. In: Keikhosrokiani P, Pourya Asl M (eds) Handbook of research on opinion mining and text analytics on literary works and social media. IGI Global, pp. 263–291

Alawneh E, Al-Fawa'reh M, Jafar MT, Al Fayoumi M (2021) Sentiment analysis-based sexual harassment detection using machine learning techniques. In: 2021 international symposium on electronics and smart devices (ISESD). IEEE, Bandung, Indonesia p 1–6. https://doi.org/10.1109/ISESD53023.2021.9501725

Asl MP (2019) Foucauldian rituals of justice and conduct in Zainab Salbi's between two worlds. J Contemp Iraq Arab World 13(2-3):227–242. https://doi.org/10.1386/jciaw_00010_1

Asl MP (2020) Spaces of change: Arab women's reconfigurations of selfhood through heterotopias in Manal al-Sharif's Daring to Drive. KEMANU-SIAAN Asian J Humanit 27(2):123–143. https://doi.org/10.21315/kajh2020.27.2.7

Asl MP (2022) Truth, space, and resistance: Iranian women's practices of freedom in Ramita Navai's City of Lies. Women's Stud 51(3):287–306. https://doi.org/10.1080/00497878.2022.2030342

Asl MP (ed) (2023) Urban poetics and politics in contemporary South Asia and the Middle East. IGI Global

Asl MP, Hanafiah SABM (2023) Mapping violence against women in Pakistan: a Galtungian reading of Fatima Bhutto's The Shadow of the Crescent Moon. South Asian Rev. https://doi.org/10.1080/02759527.2024.2304994

Aslam N, Rustam F, Lee E, Washington PB, Ashraf I (2022) Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model. IEEE Access 10:39313–39324

Asri MAZBM, Keikhosrokiani P, Asl MP (2022) Opinion mining using topic modeling: a case study of Firoozeh Dumas's Funny in Farsi in Goo dreads. In: Saeed F, Mohammed F, Ghaleb F (eds) Advances on intelligent informatics and computing. Springer Cham, p 219–230

Behera RK, Jena M, Rath SK, Misra S (2021) Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. Inf Process Manag 58(1):102435

Bouhlila DS (2019) Sexual harassment and domestic violence in the Middle East and North Africa. https://www.arabbarometer.org/?report=sexual-harassment-and-domestic-violence-in-the-middle-east-and-north-africa

del Carmen Herrera M, Herrera A, Expósito F (2017) To confront versus not to confront: women's perception of sexual harassment. Eur J Psychol Appl Leg context 10(1):1–7

Chew XY, Asl MP (2023) The poetics of identity making: precarity and agency in Tahmima Anam's The Good Muslim. J Cult Res https://doi.org/10.1080/14797585.2023.2291711

Chu KE, Keikhosrokiani P, Asl MP (2022) A topic modeling and sentiment analysis model for detection and visualization of themes in literary texts. Pertanika J Sci Technol 30(4):2535–2561. https://doi.org/10.47836/pjst.30.4.14

Da NZ (2019) The computational case against computational literary studies. Crit Inq 45(3):601–639

Elmi AH, Keikhosrokiani P, Asl MP (2023) A machine learning approach to the analytics of representations of violence in Khaled Hosseini's novels. In: Keikhosrokiani P, Asl MP (eds) Handbook of research on artificial intelligence applications in literary works and social media. IGI Global, pp. 36–67

Eltahawy M (2015) Headscarves and hymens: Why the Middle East needs a sexual revolution. Farrar, Straus and Giroux

Ennaji M, Sadiqi F (eds) (2011) Gender and violence in the Middle East, 1st edn. Routledge

Fasha EFBK, Keikhosrokiani P, Asl MP (2022) Opinion mining using sentiment analysis: a case study of readers' response on Long Litt Woon's the way through the woods in goodreads. Advances on Intelligent Informatics and Computing, Cham

Fateh N (2022) A narrative literature review on sexual harassment in higher education in the Middle East from a feminist framework perspective. American University of Beirut

Guo J (2022) Deep learning approach to text analysis for human emotion detection from big data. J Intell Syst 31.1:113–126

Hadi NHA, Asl MP (2022) The real, the imaginary, and the symbolic: a Lacanian reading of Ramita Navai's City of Lies. GEMA Online J Lang Stud 22(1):145–158. https://doi.org/10.17576/gema-2022-2201-08

Jafery NN, Keikhosrokiani P, Asl MP (2023) An artificial intelligence application of theme and space in life writings of middle eastern women: A topic modelling and sentiment analysis approach. In: Handbook of Research on Artificial Intelligence Applications in Literary Works and

Social Media. IGI Global, p 19–35. https://doi.org/10.4018/978-1-6684-6242-3.ch002

John SA, Keikhosrokiani P (2022) COVID-19 fake news analytics from social media using topic modeling and clustering. In: Big Data Analytics for Healthcare: Datasets, Techniques, Life Cycles, Management, and Applications. Academic Press, p 221–232. https://doi.org/10.1016/B978-0-323-91907-4.00003-0

Karami A, Spinel MY, White CN, Ford K, Swan S (2021) A systematic literature review of sexual harassment studies with text mining. Sustainability 13(12):6589, https://www.mdpi.com/2071-1050/13/12/6589

Keikhosrokiani P, Asl MP (eds) (2022) Handbook of research on opinion mining and text analytics on literary works and social media. IGI Global. https://doi.org/10.4018/978-1-7998-9594-7

Keikhosrokiani P, Pourya Asl M (2023) Handbook of research on artificial intelligence applications in literary works and social media. IGI Global. https://doi.org/10.4018/978-1-6684-6242-3

Kim E, Klinger R (2018) A survey on sentiment and emotion analysis for computational literary studies. arXiv preprint arXiv:1808.03137

Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D (2019) Text classification algorithms: a survey. Information 10(4):150

Lindgren S (2020) Data theory: interpretive sociology and computational methods. John Wiley & Sons

Lvova EN, Khammatova R, Klepach YV, Kazakova SN, Shukshina L, Dzhafarov NK (2018) Approaches to the study of social situation and its subjective interpretation. Espacios 39(20):17

Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J (2021) Deep learning-based text classification: a comprehensive review. ACM Comput Surv 54(3):1–40

Mohd Amram NAL, Keikhosrokiani P, Asl MP (2023) Artificial intelligence approach for detection and classification of depression among refugees in selected diasporic novels. Soc Sci Humanit Open 8(1):100558. https://doi.org/10.1016/j.ssaho.2023.100558

Ranganathan M, Wamoyi J, Pearson I, Stöckl H (2021) Measurement and prevalence of sexual harassment in low- and middle-income countries: a systematic review and meta-analysis. BMJ Open 11(6):e047473. https://doi.org/10.1136/bmjopen-2020-047473

Rezvan M, Shekarpour S, Alshargi F, Thirunarayan K, Shalin VL, Sheth A (2020) Analyzing and learning the language for different types of harassment. Plos one 15(3):e0227330

Suhendra NH, Keikhosrokiani P, Asl MP, Zhao X (2022) Opinion mining and text analytics of literary reader responses: a case study of reader responses to KL Noir volumes in goodreads using sentiment analysis and topic. In: Keikhosrokiani P, Pourya Asl M (eds) Handbook of research on opinion mining and text analytics on literary works and social media. IGI Global, pp. 191–239

Wei JTJ, Asl MP (2023) Dystopia and heterotopia: Poetics and politics of space in Khaled Hosseini's A Thousand Splendid Suns. In: Asl MP (ed) Urban poetics and politics in contemporary South Asia and the Middle East. IGI Global, pp. 18–36

Wright D, Betts L, Harding R, Knight CS, Peart S, Newbold K (2017) Understanding street harassment of children: identifying recurrent behaviours in a corpus of young people's accounts of harassment. Nottingham Trent University, University of Birmingham, UK

Yan CX, Asl MP (2023) Precarious lives and resisting women: a Butlerian reading of Tahmima Anam's A Golden Age. Women's Stud 52(8):888–906. https://doi.org/10.1080/00497878.2023.2251165

Yin D, Xue Z, Hong L, Davison BD, Kontostathis A, Edwards L (2009) Detection of harassment on web 2.0. Proc Content Anal WEB 2:1–7

Ying SY, Keikhosrokiani P, Asl MP (2022) Opinion mining on Viet Thanh Nguyen's the sympathizer using topic modelling and sentiment analysis. J Inf Technol Manag 14(5th Int Conf Reliable Inf Commun Technol (IRICT)) 2020:163–183. https://doi.org/10.22059/jitm.2022.84895

Zhao X, Keikhosrokiani P (2022) Sales prediction and product recommendation model through user behavior analytics. Comput Mater Contin 70(2). https://doi.org/10.32604/cmc.2022.019750

Author contributions

LHQ: Conducted this study as part of her studies. She implemented the methodology, performed thorough data pre-processing and analytics. PK: Provided guidance in defining the study objectives, proposed the methodology, assisted with data analytics and interpretation, and provided input in writing the paper. Additionally, she revised the paper for further improvements. MPA: Provided the necessary datasets for the study. He proposed and specified the problem statement and objectives. Furthermore, he provided assistance in interpreting the results, contributed to the writing process, and conducted a thorough revision of the paper.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors; therefore, ethical approval was not required.

Informed consent

This article does not contain any studies with human participants performed by any of the authors; therefore, Informed consent was not required.

Additional information

Correspondence and requests for materials should be addressed to Pantea Keikhosrokiani.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.