Machine Learning Algorithm Accuracy for Code-Switching Analytics in Detecting Mood

Latifah Abd Latib[1*], Hema Subramaniam[2*], Siti Khadijah Ramli[1], Affezah Ali[3], Astri Yulia[4], Tengku Shahrom Tengku Shahdan[5], and Nor Sheereen Zulkefly[6]

[1]Faculty of Communication, Visual Art and Computing, Universiti Selangor
[2]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[3]School of Liberal Arts & Sciences, Taylor's University
[4]Department of Language Education, Faculty of Education and Social Sciences, University Selangor.
[5]School Of Education & Human Sciences, Albukhary International University.
[6]Faculty of Medicine and Health Sciences, Universiti Putra Malaysia
Corresponding author

Abstract
Nowadays, as we can notice on social media, most users choose to use more than one language in their online postings. Thus, social media analytics needs reviewing as code-switching analytics instead of traditional analytics. This paper aims to present evidence comparable to the accuracy of code-switching analytics techniques in analysing the mood state of social media users. We conducted a systematic literature review (SLR) to study the social media analytics that examined the effectiveness of code-switching analytics techniques. One primary question and three sub-questions have been raised for this purpose. The study investigates the computational models used to detect and measures emotional well-being. The study primarily focuses on online postings text, including the extended text analysis, analysing and predicting using past experiences, and classifying the mood upon analysis. We used thirty-two (32) papers for our evidence synthesis and identified four main task classifications that can be used potentially in code-switching analytics. The tasks include determining analytics algorithms, classification techniques, mood classes, and analytics flow. Results showed that CNN-BiLSTM was the machine learning algorithm that affected code-switching analytics accuracy the most with 83.21%. In addition, the analytics accuracy when using the code-mixing emotion corpus could enhance by about 20% compared to when performing with one language. Our meta-analyses showed that code-mixing emotion corpus was effective in improving the mood analytics accuracy level. This SLR result has pointed to two apparent gaps in the research field: i) lack of studies that focus on Malay-English code-mixing analytics and ii) lack of studies investigating various mood classes via the code-mixing approach.

Keywords:
machine learning, mental health prediction, code switching analytics, systematic review, accuracy measurement.

1. Introduction

Communication mediums are undergoing a revolution in recent years with the remarkable advancement of the virtual environment. Social networking sites (SNS) are the most used communication medium (Fedric & Saumya, 2017). Traditionally, social media users always focus on posting messages, commenting on postings, and reacting. Indeed, SNS is embedded in our lives until the younger generation cannot imagine their lives without it. Data analytics through SNS has become a central investigation point to detect various social issues. As a result, the research studies proposed various intervention techniques for solving those social issues.

On the other hand, mental health issues among young adults worldwide with rising mental health problems. Many solutions regarding mental health for this age group appear to be questionable due to the unpredictable mood state of young adults. For the past few years many studies have discovered a relation between time spent using social media and mental health problems, such as depression and anxiety (Berryman et al., 2018) (Cain, 2018) (Karim et al., 2020). One study exposes that access for more than 3 hours significantly impacted body comparison tendencies among female university students (Abdalqader & Joseph, 2020). Generally, most research has focused on social media's effect on mental health problems rather than finding diagnosis solutions. Severe criticism of the diagnosis solution is through identifying moods via social media postings.

Notably, most data analytics embarks on human behaviour understanding. Recent social media analytics have directed those understandings into various purposes such as promotional purposes, security identification, and more. Furthermore, the researchers have paid considerable attention to detecting the seriousness of social media postings and how they affect their mental wellness (Hattingh et al., 2021). For several years, significant effort has been devoted to studying SNS in detecting youngsters' mental wellness. Several recent publications have documented mood analytics techniques using SNS postings

Manuscript received September 5, 2022
Manuscript revised September 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.9.44

(Zucco et al., 2020). Since youngsters' SNS postings may predominantly affect their mental well-being, examining the strength and weaknesses of those automated mood analytics techniques and algorithms seems relevant and applicable.

Language analytics has developed to become a vital factor in social media analytics. Most social media users combine their language while posting on social media (Hattingh et al., 2021). This activity is known as the code-switching technique. The code-switching writing format in the single posting becomes the major challenge when analysing social media posts. The unavailability of code-switching corpus is the reason for it. Nevertheless, there is research that has focused on code-switching analytics.

In realising how existing mood analytic techniques can significantly contribute to the code-switching analytics tool's effectiveness, we conducted a proper investigation based on its implementation requirements. Thus, this study aims to contribute to the body of knowledge of SNS-based mood analytics and improve the existing techniques. The critical contribution of this paper is to produce the overall view of the mood analytic model using the systematic literature review. The result serves as a pathway to achieving a comprehensive mood analytics outcome. It can achieve by comprehending mood analytics' current skill condition.

This paper's structures are as follows; Section 2 highlights the other researchers' related works on language code-switching analytics and mood detection, while Section 3 clarifies the incredible detail of the method undertaken in this study. Section 4 explains and discusses the findings from the compilations of this study. Finally, Section 5 is the conclusion of this paper

2. Related Works

Social media users tend to use informal and conversat ionaltext in their postings. Previous research indicates that social media users prefer English to other languages (Azmi n & Dhar, 2019). Mainly, youngsters prefer to use multiple languages while posting messages (Singh et al., 2018). Re cently, there has been a considerable demand for regional l anguage usage in social media. On the other hand, social m edia postings now are inundated with mixing multiple lang uages in a sentence. A survey among 338 students from Al Buraimi University College in Oman resulted in 86.40% o f students practising code-switching on SNS (Al-Emran & Al-Qaysi, 2017). Due to that, several researchers have paid attention to multi-language analysis, especially on languag e mixing in SNS postings. Nevertheless, the analytics purp ose's mixed-language corpus is still not adequately provide d, even though it could significantly advance sentiment ana lytics (Jose et al., 2020) (Ho et al., 2020). For instance, cod e-switching posts are believed to be uniformly efficient in detecting emotions (Wang et al., 2017).

To more comprehend social media analytics, it is esse ntial to review the SLR from mood/emotion analytics, cod e-switching analytics, and lexical analytics. For instance, a n SLR work is conducted on personality traits classificatio n using SNS textual content. The study has classified the w orks according to machine learning types, such as deep, su pervised, and unsupervised learning (H. Ahmad et al., 202 0). Another literature investigation shows a significant rela tionship between SNS and users' romantic relationships (R us & Tiemensma, 2017). Furthermore, another major SLR work reviews the influence of social media among adolesc ents. The predictors categorise four domains: addiction, act ivities, duration consumed, and investments on social platf orms. Those classification domains were associated with d epression, anxiety, and psychological distress (Keles et al., 2020). Apart from that, another SLR study concentrated o n the effects of SNS on users' psychological well-being. Ac cording to social media usage, the study has classified the r elated works positively and negatively (Erfani, 2018).

The review of language analytics towards mood detec tion via SNS is still limited. Most of the current SLR work s focused deliberately on the result of SNS analysis toward s society. The embarkment into technical aspects is yet to b e explored. Hence, we aimed to analyse the techniques, ele ments, and theories of mood analytics using SNS from the perspective of language usage.

3. Methodology

We have performed an investigation using literature related to language code-switching analytics. The work was inspired by Barbara Kitchenham's (Kitchenham & Charters, 2007) structured, systematic literature review method. Figure 1 shows the flow of research activities in 4 phases. Initially, we identified the PICOC (Population, Intervention, Context, Outcome and Comparison) elements. Next, we formulate the research review questions to locate the associated research articles. The review aims to gather and classify the machine learning algorithms used during SNS analytics. Another primary concern is to understand the analytics procedure for language mixing postings.

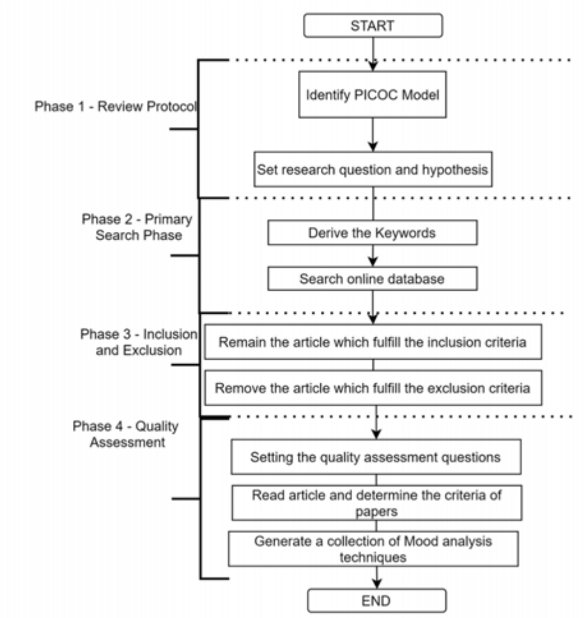

Figure 1. Review Methodology

As a first step, we have developed the PICOC model. Table 1 explains the PICOC model elements.

Table 1 PICOC model

PICOC Elements	Explanation
Population	Mood analytics using social media
Intervention	Using multiple languages
Comparison	Types of machine learning algorithms.
Outcome	Model accuracy
Context	Review all studies of SNS-based mood analytics using code-switching.

The study aims to identify the various machine learning algorithms used in analysing mood status. It includes studies on mood identification, analysis, and prediction. We have focused on studies related to young social media users. We have focused our review on the multiple language preferences. The comparison in terms of model accuracy was taken into consideration as well. We have tried for answering the following review research question:
Primary Question: What evidence of mood analytics studies on social media analysed using code-switching?

Meanwhile, this SLR also seeks to answer the following secondary sub-questions:
Sub-Question 1: What evidence regarding which machine learning algorithms affect the accuracy of mood analytics via code-switching?
Sub-Question 2: What evidence of code-switching corpus would affect mood analytics accuracy?
Sub-Question 3: What evidence of code-switching analytics would affect the classification of moods?
The PICOC model has identified keywords for the article during the search process and uses logical operators to combine the synonym words. Table 2 lists the search strings used during the searching process.

Table 2 List of Keywords with logical operator

No	Keywords
1	mood identification OR mood analysis OR mood detection
2	corpus OR lexical
3	social networking sites OR social media OR Twitter OR YouTube OR Facebook OR Instagram OR online
4	Code-switching OR code-mixing OR multiple languages

The next step is to identify the relevant article and remove the duplicates. The inclusion and exclusion criteria were determined to accomplish this task (Nassif et al., 2021). The review will include a study that explains the mood analytics techniques within code-switching. Meanwhile, investigation for mood identification studies without social media platforms will exclude. The final analysis for the studies covering platforms such as blogs and websites also will exclude.
Finally, the quality of the selected articles was assessed using predefined questions. It consists of twelve general questions with the ratio scale: Yes = 1 point; No = 0 points. Each study's quality score ranged between 0 (very poor) and 7 (very good). The suitability of the proposed models further checked their relation to mood detection. It ensures that the chosen article relates to the objective of this SLR work. One of our researchers (Siti Khadijah) was responsible for reading and completing the extraction form for each primary study. In order to avoid any bias in the quality assessment process, a random sample was extracted and read by other authors. The review discussion initiated if the quality score differed by more than 20% among the three authors.

4. RESULT ANALYSIS AND DISCUSSION

4.1 Article Selection

The SLR investigation uses online databases such as Google Scholar, IEEE Xplore, Science Direct, and Springer Link. We had about 32 articles out of 1414 articles for final analysis. From 1414 articles, we have selected 107 articles for the following process. Many studies were excluded due to inadequate format, the irrelevancy of the article, and not comprehensive content. Then, after imposing the inclusion and exclusion criteria, we had about 53 articles. The further exclusion was performed by reviewing the full-text article and eliminating any study that failed to elaborate on the methodology and techniques used for mood detection via SNS. Finally, a total of 32 articles will select for review purposes. Fig 2 illustrates the filtration process and the number of articles.

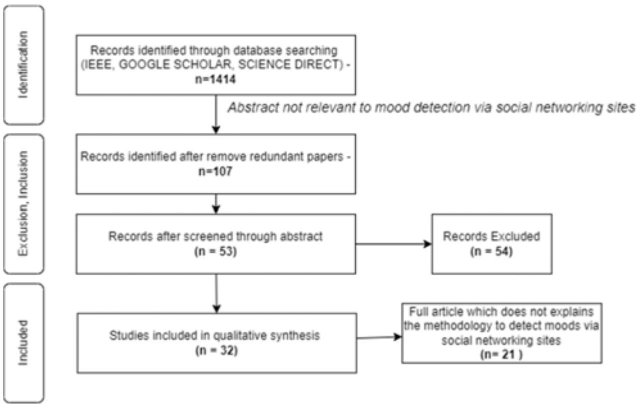

Fig 2. Article numbers after literature Investigation

4.2 Quality Assessment

Table 3 shows the quality scores obtained by all primary studies. Most articles achieved above-average quality: 15 studies (47%) and 10 studies (31%) were deemed excellent. None of the studies scored inferior quality. Due to the high quality, we include all the studies in the final analysis discussion. However, some detailed analysis of the score indicates that about 36.59% of 32 studies had not comprehensively explained the mood analysis technique. For instance, many authors used the method of Decision Trees Classifier widely, yet its usage explanation is not comprehensive (Joseph, 2019).

Table 3. Quality Assessment

Quality scale	Very Poor (<7)	Poor (8 marks)	Fair (9 marks)	Good (10 marks)	Very Good (>10)	Total
Number of studies	0	2	5	10	15	32
Percentage (%)	0	6	16	31	47	100

4.3 Research Questions

Primary questions:
What evidence of mood analytics studies on social media investigated using code-switching?

Social media users have prominently used English to post their opinions, thoughts, and more. Significantly, it creates a strong interconnection between mood analytics and lexical analytics. The SLR activity reveals that mood analytics' effectiveness depends on integrating knowledge from different affective lexical resources. Indeed, those resources believe to be enhancing the mood analytics baseline numbers. The SLR reveals that 32 studied the code-switching-based mood analytics research. The context of language switching varies according to the mood classes. Figure 3 shows the breakdown of the language of code-switching that has famously been used in mood analytics.

Generally, most code-switching-based mood analytics has kept English as the second language. Among those studies, the code-switching analysis on Hindi + English scored about 40% higher than other code-switching languages in the comprehensive research studies. Referring (Sasidhar et al., 2020), Hindi is the numerous verbal speech in India. Above 45% of 1.3 billion people generally use code-mixed language at most minuscule two on social media. However, more researchers focus only on emotion detection and code-mixed social media text for Hindi and English separately. Apart from that, China researchers also show a considerable interest score of about 13% in code-switching-based mood analytics and (Wang et al., 2017) is the first study that has attempted to analyse emotion in Chinese-English code-switching text.

Meanwhile, the Bengali + English language code-switching score was 6% in the comprehensive research study. This language is also used among Indian citizens, but less researcher uses this language to study code-switching-based mood analytics. The same goes for Malay + English; researchers focus only on a code-switching subjectivity corpus for Malay and English.

The graph also shows an Urdu + English score of 7%, the same as other languages such as Kannada + English, Singlish, Spanish + English and Arabic + English. Based on (M. Z. Ali et al., 2021), the study used Urdu and English because they use this multi-language to study sentiment analysis and emotion classification in Pakistan. However, in the Arabic + English code-switching language, in the study (A. Ali et al., 2021), there have been initiatives in building and analysing code-switching in Arabic + English speech using automatic speech recognition (ASR).

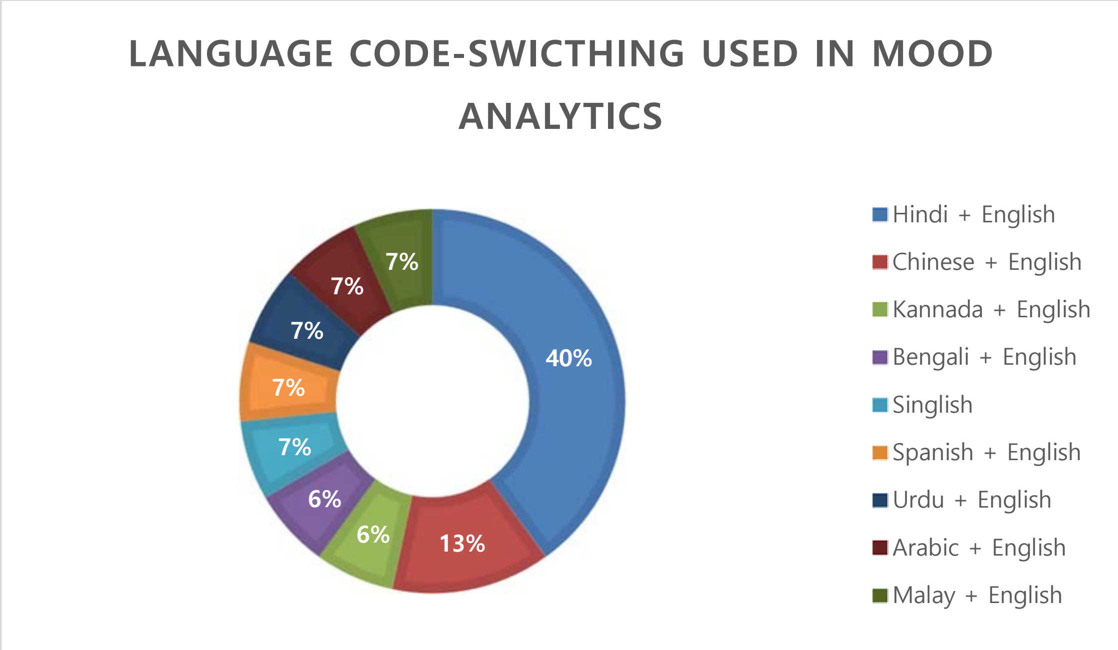

Fig 3 Breakdown of Language Code-Switching used in Mood Analytics

Sub-Question 1: What evidence regarding which machine learning algorithms affect the accuracy of mood analytics via code-switching?

The overall review categorises the mood analytics model according to machine learning algorithms. Among the three categories of machine learning, almost all studies focused on the supervised learning algorithm instead of unsupervised or reinforcement learning. Among the machine learning algorithm, Support Vector Machine (SVM) technique and Naive Bayes classifier (NBC), and deep learning network, the Convolutional Neural Network (CNN) technique was frequently used by the researchers when dealing with the code-switching postings. The hybrid technique performs better as compared to the single classification technique. Different machine learning and deep learning techniques reveal different accuracy measurements. Figure 4 shows the model accuracy for different techniques when dealing with code-switching postings.

Convolutional Neural Network (CNN) can learn local responses from temporal or spatial data but cannot learn sequential correlations (Zhou et al., 2015). CNN also is a solid performance performed on the practically important task of sentence classification. Current studies have also suggested its incredible performance for text classification tasks (Collobert; et al., 2011) (Wallace, 2014). CNN operates into two phases; the first step is feature representations, which encode the target information into feature representation vectors. The second step is

classification layers, where the representation vectors from the first stage are input into the classification part (H. Ahmad et al., 2020).

Table 4 Accuracy of Machine Learning Algorithm

Language	Machine Learning Algorithm	Accuracy	Referenc e
Hindi + English	1D-CNN	78%	(Sasidhar et al., 2020)
	LSTM	81%	
	Bi-LSTM	81%	
	CNN-LSTM	82.85%	
	CNN-BiLSTM	83%	
Hindi + English	SVM	58%	(Vijay et al., 2018)
Hindi + English	CNN	66.63%	(Gupta et al., 2016)
Hindi + English	LSTM	69.70%	(Joshi et al., 2016)
Canada + English	CNN	73%	
Hindi + English	Bi-LSTM	73.60%	
Hindi + English	CNN	82.62%	

Chinese + English	Factor Graph Model Belief Propagation Algorithm	69.30%	(Wang et al., 2017)
Bengali + English	Multilayer Perceptron	69%	(Ghosh et al., 2017)
Singlish	SVM	NA	(Lo et al., 2016)
Spanish + English	Supervised Approaches		(Vilares et al., 2017)
Malay + English	NA	NA	(Kasmuri & Basiron, 2019)
Chinese + English	Joint Learning	NA	(Gao et al., 2013)
Urdu + English	1d-CNN - 74%	74%	(M. Z. Ali et al., 2021)
	LSTM	75.60%	
	BiLSTM	77.6	
	Attention Mechanism	76%	
Arabic + English	CNN-LSTM	81.80%	(Abdullah et al., 2019)

Long Short-Term Memory (LSTM) is the most popular Recurrent Neural Network (RNN) model, consisting of an input gate, a memory cell, an output gate, and a forget gate (Nassif et al., 2021). LTSM and CNN operate to categorise character features from online content. All the deep learning models have a particular way of encoding the target information into a feature when given an input (H. Ahmad et al., 2020).

A bidirectional LSTM (BiLSTM) network operates to learn past features (forward layer) and future features (backward layer) from the tokens on both sides and is required when dealing with unstructured language analytics at a specific time (Fan et al., 2018). Meanwhile, LSTM and BiLSTM can remember sequential patterns which can have important implications while analysing the text. Hence, they are also operated by bypassing the CNN layer. At last, the CNN-LSTM and CNN- BiLSTM models use because CNN can abstract features and reduce the complexity of training LSTM or BiLSTM. Typically, researchers used LSTM and BiLSTM algorithms to classify sentiment in code-mixed datasets (G. I. Ahmad et al., 2022).

The results above show that CNN as a classifier has done well in learning the new importance of the prototype of a pre-trained model. The word embeddings developed from a vast corpus reasonably captured semantics achieved 82.62% accuracy. However, the CNN-BiLSTM model achieved 83.21% classification accuracy and shows that deep neural networks mainly associated with CNN as the primary layer have provided better results (Sasidhar et al., 2020).

Sub-Question 2: What evidence of code-switching corpus would affect mood analytics accuracy?

In single language analytics, the corpus availability determines the accuracy of the prediction model. For instance, the Spanish language has used four different lexical such as Affect in Tweets Dataset (AIT), Spanish Emotion Lexicon (SEL), improved Spanish Opinion Lexicon (iSOL), NRC Word-Emotion Association Lexicon (EmoLex). In recognising the mood, Vietnamese researchers have designed a corpus called UIT-VSMEC. For the Arabic language, there is neither an Arabic corpus with models labelled for emotions nor studies to detect emotions from Arabic microblogs content. Even though there is no development in the Arabic corpus with models labelled for emotions, there have been few trials for Arabic content. An emotion-labelled tweets corpus has been created (TEC) as a word-emotion lexicon. The data preprocessing took place using five different techniques: basic preprocessing, basic preprocessing in addition to removing a list of stopwords, Lucene light Arabic stemmer, Shereen Khoja Arabic stemmer, and modified Khoja Arabic stemmer (Rabie & Sturm, 2014). On the other hand, the collection of terms under code-switching posting had also gained considerable interest.

Most corpus descriptions are textual and based on researchers' opinions. Such impressions are highly subjective, not a corpus similarity or proper complexity measurement. In contrast, questions about the limitations and benefits of using that corpus will arise whenever we work on a new corpus. The data's size and homogeneity are used as some of the factors intensively. However, such approaches are mainly word-based and apply to monolingual texts. Corpus similarity measure has a vast collection of applications. It has theoretical research applications where one can judge the complexity of the dataset before performing technical analysis, and one may want to substitute a dataset with another. It is worthwhile only if there is some way to determine whether the two datasets are equivalent and similar in terms of their complexity and usage. It would also help in the inter-domain portability of Natural Learning Processing (NLP) systems.

For example, Code Mixing Index (CMI) is the index that tries to assess the level of code-switching in a phrase. The measure aimed at comparing one code-switched corpus with another. Nonetheless, CMI only considered the word fraction in the corpus, which is code-switched, as an initial parameter and suggested some improvements, considering the number of languages and the number of code-switching points present in the corpus (Ghosh et al., 2017).

Despite the important implications of codeswitching for emotion analysis, existing emotion analysis techniques fail to accommodate the code-switching content. Thus, there is an essential requirement to analyse emotions in code-switching texts. This paper provides a well-defined and efficient method for constructing and analysing a large-scale code-switching corpus from social media. We believe the corpus provides a valuable resource for linguistic analyses and natural language processing of emotion and code-switching texts (Lee & Wang, 2015).

Sub-Question 3: What evidence of code-switching analytics would affect the classification of moods?

The authors claimed that this classifier is fast and straightforward when training and classifying words. In addition to CNN and SVM, the NBC technique creates a prediction based on experience and decides the possibility classification (Sunarya et al., 2019). Even though many pieces of research had focused on Twitter data analyses, it seems like only the NBC technique was applied extensively in analysing Facebook data. The Facebook data analyses applied the most significant mood classifiers, such as happy, sad, and angry (Azmin & Dhar, 2019). Although positive, negative, and neutral are the standard classification used in mood analyses, specific authors apply different classification criteria. The authors analysed Twitter content, assigning different colours to different moods. For example, the neutral mood is grey, the happy mood is yellow, and the sad mood is blue (Harvey et al., 2018).

5. Conclusion and Future Works

The study aims to find and collect related research regarding mood analytics for multi-language postings. We conducted a systematic literature review by gathering information from various journals and conference articles. After executing the extraction process, we have included 32 journal articles that focus on mood analytics for code-switching social media postings. In analysing the systematic literature review evidence, we found that code-switching analytics performed better using three primary data analytics techniques. The techniques include Convolutional Neural Network (CNN) technique, Support Vector Machine (SVM) technique, and Naive Bayes classifier (NBC). Among those three techniques, CNN used prominently to analyse social media postings. In addition, the hybrid technique CNN-BiLSTM performs better with 83.21% accuracy than other algorithms.

The SLR also reveals the importance of mood classification while performing code-switching analytics. The classification either can be done in the form of mood expressed or polarity of the mood expressed. Especially for the mood expressed classification, we found out that a set of moods that are used commonly is "Happy", "Sad", and "Angry". Instead, social media users tend to express more than these moods, such as anger, frustration, calm, and more. Performing the code-switching analytics would tend to employ slightly different steps than standard social media postings analytics. Instead of just concentrating on mood analytics, it shall also concentrate on lexical analytics. Lexical analytics could be executed successfully by creating the corpus for mixed-language instead of one language corpus.

In terms of data profiling, an ideal data duration is about 12 weeks to ensure model accuracy. Since we deal with abstract output (mood/emotions), at least 12 weeks of data is necessary (Larsen et al., 2015) (Nimeshika & Ahangama, 2019). Also, continuously collecting the social media data within 24 hours is essential in ensuring the outcome's comprehensibility. It shows the importance of considering duration while collecting data for mood analytics. The data's availability had helped classify tweets data into emotions according to geographic location and gender (Larsen et al., 2015). It uses a weighted estimation average to award a score for mood.

Acknowledgement

We want to thank the Ministry of Higher Education Malaysia for the Fundamental Research Grant Scheme (FRGS/1/2020/SS0/UNISEL/02/2) for the study's financial support.

References

[1] Abdalqader, M. A., & Joseph, S. A. (2020). The Impact of Social Media on Body Comparison Tendency, Body-Esteem and Sleep Quality Among Female Students in a Private University in Shah Alam/ Malaysia. Global Journal of Public Health Medicine, 2(2), 229–234. https://doi.org/10.37557/gjphm.v2i2.66

[2] Abdullah, M., Hadzikadicy, M., & Shaikhz, S. (2019). SEDAT: Sentiment and Emotion Detection in Arabic Text Using CNN-LSTM Deep Learning. Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, January 2019, 835–840. https://doi.org/10.1109/ICMLA.2018.00134

[3] Ahmad, G. I., Singla, J., Ali, A., Reshi, A. A., & Salameh, A. A. (2022). Machine Learning Techniques for Sentiment Analysis of Code-Mixed and Switched Indian Social Media Text Corpus: A Comprehensive Review. International Journal of Advanced Computer Science and Applications, 13(2), 455–467. https://doi.org/10.14569/IJACSA.2022.0130254

[4] Ahmad, H., Asghar, M. Z., Khan, A. S., & Habib, A. (2020). A systematic literature review of personality trait classification from the textual content. Open Computer Science, 10(1), 175–193. https://doi.org/10.1515/comp-2020-0188

[5] Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment Analysis of Tweets using SVM. International Journal of Computer Applications, 177(5), 25–29. https://doi.org/10.5120/ijca2017915758

[6] Ali, M. Z., Javed, K., Haq, E. ul, & Tariq, A. (2021). Sentiment and Emotion Classification of Epidemic Related Bilingual data from Social Media. http://arxiv.org/abs/2105.01468

[7] Azmin, S., & Dhar, K. (2019). Emotion Detection from Bangla Text Corpus Using Naïve Bayes Classifier. 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 1–5. https://doi.org/10.1109/EICT48899.2019.9068797

[8] Berryman, C., Ferguson, C. J., & Negy, C. (2018). Social Media Use and Mental Health among Young Adults. Psychiatric Quarterly, 89(2), 307–314. https://doi.org/10.1007/s11126-017-9535-6

[9] Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016). Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. JMIR Mental Health, 3(2), e21. https://doi.org/10.2196/mental.4822

[10] Cain, J. (2018). It is time to confront student mental health issues associated with smartphones and social media. American Journal of Pharmaceutical Education, 82(7), 738–741. https://doi.org/10.5688/AJPE6862

[11] Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., Ajilore, O., Ryan, K., & Leow, A. D. (2017). DeepMood: Modeling mobile phone typing dynamics for mood detection. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F1296, 747–755. https://doi.org/10.1145/3097983.3098086

[12] Collobert;, R., Weston, J., Bottou;, L., Karlen, M., Kavukcuoglu;, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. 12, 2493–2537.

[13] Erfani, S. S. B. A. (2018). Impacts of the use of social network sites on users' psychological well-being: A systematic review. https://doi.org/https://doi.org/10.1002/asi.24015

[14] Fedric, K., & Saumya, S. (2017). Engaging customers through online participation in social networking sites.

[15] Gao, W., Li, S., Lee, S. Y. M., Zhou, G., & Huang, C. R. (2013). Joint learning on sentiment and emotion classification. International Conference on Information and Knowledge Management, Proceedings, August 2015, 1505–1508. https://doi.org/10.1145/2505515.2507830

[16] Ghosh, S., Ghosh, S., & Das, D. (2017). Sentiment Identification in Code-Mixed Social Media Text. http://arxiv.org/abs/1707.01184

[17] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences, 18, 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

[18] Gupta, D., Lamba, A., Ekbal, A., & Bhattacharyya, P. (2016). Opinion Mining in a Code-Mixed Environment: A Case Study with Government Portals. Proc. of the 13th Intl. Conference on Natural Language Processing, 249–258. http://ltrc.iiit.ac.in/icon2016/proceedings/icon2016/pdf/W16-6331.pdf

[19] Harvey, R., Muncey, A., & Vaughan, N. (2018). Associating colours with emotions detected in social media tweets. Proceedings of AISB Annual Convention 2018, 5–8.

[20] Hasan, M., Rundensteiner, E., & Agu, E. (2019). Automatic emotion detection in text streams by analysing Twitter data. International Journal of Data Science and Analytics, 7(1), 35–51. https://doi.org/10.1007/s41060-018-0096-z

[21] Hattingh, M., Matthee, M., Smuts, H., Pappas, I., Yogesh, K., Mäntymäki, M., Hattingh, M., Matthee, M., Smuts, H., Pappas, I., & Dwivedi, Y. K. (2021). Responsible Design, Implementation and Use of Information and Communication Technology. https://doi.org/10.1007/978-3-030-44999-5

[22] Ho, V. A., Nguyen, D. H. C., Nguyen, D. H., Pham, L. T. Van, Nguyen, D. V., Nguyen, K. Van, & Nguyen, N. L. T. (2020). Emotion Recognition for Vietnamese Social Media Text. Communications in Computer and Information Science, 1215 CCIS, 319–333. https://doi.org/10.1007/978-981-15-6168-9_27

[23] Jose, N., Chakravarthi, B. R., Suryawanshi, S., Sherly, E., & McCrae, J. P. (2020). A Survey of Current Datasets for Code-Switching Research. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 136–141. https://doi.org/10.1109/ICACCS48705.2020.9074205

[24] Joseph, F. J. J. (2019). Twitter Based Outcome Predictions of 2019 Indian General Elections Using Decision Tree. 2019 4th International Conference on Information Technology (InCIT), 50–53. https://doi.org/10.1109/INCIT.2019.8911975

[25] Joshi, A., Prabhu, A., Shrivastava, M., & Varma, V. (2016). Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text. COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, 2012, 2482–2491.

[26] Karim, F., Oyewande, A., Abdalla, L. F., Chaudhry Ehsanullah, R., & Khan, S. (2020). Social Media Use and Its Connection to Mental Health: A Systematic Review. Cureus, 12(6). https://doi.org/10.7759/cureus.8627

[27] Kasmuri, E., & Basiron, H. (2019). Building a Malay-English code-switching subjectivity corpus for sentiment analysis. International Journal of Advances in Soft Computing and Its Applications, 11(1), 112–130

[28] Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth, 25(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

[29] Khan, A., Shahid Husain, M., & Khan, A. (2018). Analysis of Mental State of Users Using Social Media to Predict Depression! A Survey. International Journal of Advanced Research in Computer Science, 9(2).

[30] Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. January 1–57.

[31] Larsen, M. E., Boonstra, T. W., Batterham, P. J., O'Dea, B., Paris, C., & Christensen, H. (2015). We Feel: Mapping Emotion on Twitter. IEEE Journal of Biomedical and Health Informatics, 19(4), 1246–1252. https://doi.org/10.1109/JBHI.2015.2403839

[32] Lee, S. Y. M., & Wang, Z. (2015). Emotion in code-switching texts: Corpus construction and analysis. Proceedings of the 8th SIGHAN Workshop on Chinese Language Processing, SIGHAN 2015 - Co-Located with 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, ACL IJCNLP 2015, 91–99. https://doi.org/10.18653/v1/w15-3116

[33] Lo, S. L., Cambria, E., Cornforth, D., Ling, S., & David, A. (2016). Institutional Knowledge at Singapore Management University A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection Knowledge-Based Systems A multilingual semi-supervised approach in deriving Singlish sentic pa. 236–247.

[34] Malaysian Communication and Multimedia Commission (MCMC). (2020). Internet users survey 2020: Infographic. Statistics and Data Intelligence Department, Malaysian Communications and Multimedia Commission, 1–6.

[35] Malaysian Communications and Multimedia Commission. (2020). Internet Users Survey 2020. The Internet Users Survey, 76. https://doi.org/ISSN 1823-2523

[36] Nalinde, P. B., & Shinde, A. (2019). Machine learning framework for detection of psychological disorders at OSN. Int J Innov Technol Explor Eng (IJITEE), 8(11).

[37] Napitu, F., Bijaksana, M. A., Trisetyarso, A., & Heryadi, Y. (2017). Twitter opinion mining predicts broadband internet's customer churn rate. 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), 141–146. https://doi.org/10.1109/CYBERNETICSCOM.2017.8311699

[38] Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. (2021). Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. Applied Soft Computing, 98(November). https://doi.org/10.1016/j.asoc.2020.106836

[39] Nimeshika, S., & Ahangama, S. (2019). A Method to Identify the Current Mood of Social Media Users. 2019 14th Conference on Industrial and Information Systems (ICIIS), 356–359. https://doi.org/10.1109/ICIIS47346.2019.9063291

[40] Plaza-del-Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., & Mitkov, R. (2020). Improved emotion recognition in Spanish social media through incorporation of lexical knowledge. Future Generation Computer Systems, 110, 1000–1008. https://doi.org/https://doi.org/10.1016/j.future.2019.09.034

[41] Rabie, O., & Sturm, C. (2014). Feel the heat: Emotion detection in Arabic social media content. The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014), 37–49.

[42] Rus, H. M., & Tiemensma, J. (2017). "It's complicated." A systematic review of associations between social network site use and romantic relationships. Computers in Human Behavior, 75, 684–703. https://doi.org/https://doi.org/10.1016/j.chb.2017.06.004

[43] Sasidhar, T. T., B, P., & P, S. K. (2020). Emotion Detection in Hinglish (Hindi+English) Code-Mixed Social Media Text. Procedia Computer Science, 171, 1346–1352. https://doi.org/https://doi.org/10.1016/j.procs.2020.04.144

[44] Sunarya, P. O. A., Refianti, R., Mutiara, A. B., & Octaviani, W. (2019). Comparison of Accuracy between Convolutional Neural Networks and Naïve Bayes Classifiers in Sentiment Analysis on Twitter. 10(5), 77–86.

[45] Vijay, D., Bohra, A., Singh, V., Akhtar, S. S., & Shrivastava, M. (2018). Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text. 128–135.

[46] Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. Information Processing and Management, 53(3), 595–607. https://doi.org/10.1016/j.ipm.2017.01.004

[47] Wang, Z., Lee, S. Y. M., Li, S., & Zhou, G. (2017). Emotion Analysis in Code-Switching Text With Joint Factor Graph Model. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(3), 469–480. https://doi.org/10.1109/TASLP.2016.2637280

[48] Wallace, B. C. (2014). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification.

[49] Wang, Z., Lee, S. Y. M., Li, S., & Zhou, G. (2017). Emotion Analysis in Code-Switching Text With Joint Factor Graph Model. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(3), 469–480. https://doi.org/10.1109/TASLP.2016.2637280

[50] Wanniarachchi, V. U., Mathrani, A., Susnjak, T., & Scogings, C. (2020). A systematic literature review: What is the current stance towards weight stigmatisation in social media platforms? International Journal of Human-Computer Studies, 135, 102371. https://doi.org/https://doi.org/10.1016/j.ijhcs.2019.102371

[51] Yusop, F. D., & Sumari, M. (2013). The Use of Social Media Technologies among Malaysian Youth. Procedia - Social and Behavioral Sciences, 103, 1204–1209. https://doi.org/10.1016/j.sbspro.2013.10.448

[52] Zhou, C., Sun, C., Liu, Z., & Lau, F. C. M. (2015). A C-LSTM Neural Network for Text Classification. http://arxiv.org/abs/1511.08630

[53] Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. WIREs Data Mining and Knowledge Discovery, 10(1), e1333. https://doi.org/https://doi.org/10.1002/widm.1333