Tech Science Press

MELex: The Construction of Malay-English Sentiment Lexicon

Nurul Husna Mahadzir[1], Mohd Faizal Omar[2], Mohd Nasrun Mohd Nawi[3,*], Anas A. Salameh[4], Kasmaruddin Che Hussin[5] and Abid Sohail[6]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Kedah, 08400, Merbok, Kedah, Malaysia
[2]Department of Decision Sciences, School of Quantitative Sciences, Universiti Utara Malaysia, 06010, Kedah, Malaysia
[3]Disaster Management of Institute, School of Technology Management and Logistic, Universiti Utara Malaysia, 06010, Kedah, Malaysia
[4]Department of Management Information System, College of Business Administration, Prince Sattam Bin Abdulaziz University, 165, Al-Kharj, Saudi Arabia
[5]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Malaysia
[6]Department of Computer Science, COMSATS University Islamabad, Lahore, Pakistan
*Corresponding Author: Mohd Nasrun Mohd Nawi. Email: nasrun@uum.edu.my

Abstract: Currently, the sentiment analysis research in the Malaysian context lacks in terms of the availability of the sentiment lexicon. Thus, this issue is addressed in this paper in order to enhance the accuracy of sentiment analysis. In this study, a new lexicon for sentiment analysis is constructed. A detailed review of existing approaches has been conducted, and a new bilingual sentiment lexicon known as MELex (Malay-English Lexicon) has been generated. Constructing MELex involves three activities: seed words selection, polarity assignment, and synonym expansions. Our approach differs from previous works in that MELex can analyze text for the two most widely used languages in Malaysia, Malay, and English, with the accuracy achieved, is 90%. It is evaluated based on the experimentation and case study approaches where the affordable housing projects in Malaysia are selected as case projects. This finding has given an implication on the ability of MELex to analyze public sentiments in the Malaysian context. The novel aspects of this paper are two-fold. Firstly, it introduces the new technique in assigning the polarity score, and second, it improves the performance over the classification of mixed language content.

Keywords: Machine learning; data sciences; artificial intelligence; opinion mining; sentiment analysis; sentiment lexicon; lexicon-based; bilingual lexicon

1 Introduction

Social network sites like Facebook, Twitter, and blogs have changed the way people communicate and the way they connect. The rapid growth of the social network has led to an explosion of user-generated content on the Internet. The new challenge is how to process and interpret this massive

amount of information available in social media. This challenge is the object of research in the discipline called "sentiment analysis."

Sentiment analysis is one of the most active research areas in Natural language processing (NLP) since early 2000 [1]. To date, sentiment analysis has been applied to various domains such as product [2], movie [3], and political reviews [4]. Most of the previous work in this field has focused on analyzing only a single language, especially English. However, with the need for globalization, it is common to see the post written in multiple languages, making the sentiment analysis process even more complex and challenging. Besides, in unstructured content like Twitter posts, people tend to mix languages in one sentence. According to Lo, Cambria, Chiong, and Cornforth [5], specific information in other languages might miss out if the analysis is done for a single language only.

The amount of textual data produced in different languages is so massive that it introduces many challenges for these services to perform sentiment analysis on the available data. The practice of using more than one language in a single sentence has arisen. Such mixed language has rarely been a subject of sentiment analysis before [6].

Since the existing NLP technique was designed for processing monolingual text, multilingual mixed texts cannot be well processed by these approaches [7,8]. Thus, the highlighted issues motivate us to develop a sentiment analysis model that could accurately categorize Tweets written in mixed languages. The focus is to effectively mine the subjective information from the user-generated content in mixed languages, Malay and English. The expected contributions of our work are two-fold:

- The development of Malay-English sentiment lexicon
- The accuracy improvement of sentiment classification using the developed Lexicon.

This paper presented a new bilingual sentiment lexicon known as MELex (Malay-English lexicon), covering two main languages in Malaysia: Malay and English, in mining public opinion. In demonstrating the performance of the proposed approach, we have conducted the experimental studies, and the results were compared and benchmark with the state-of-the-art. We used Twitter as a corpus and WordNet as a resource to obtain the synonyms. Furthermore, this study improves the traditional approach in assigning the polarity score by utilizing the word vector model to determine the polarity and term frequency to determine the weight of sentiment words.

The paper is organized as follows. In Section 2, the construction of MELex is proposed. Then, experimental setup and results are reported in Sections 3 and 4. Finally, we conclude the paper in Section 5 with future research.

2 The Construction of MELex

This research recognizes that words used in different domains may associate with different sentiment values or sentiment orientations. Due to that matter, this study proposes the use of Twitter data with specific keywords for a specific domain as seed words in generating the Lexicon.

In this section, we explain in detail a process of constructing a bilingual lexical resource specifically for the property domain, which covers the two most spoken languages in Malaysia: Malay and English. This lexical resource is named as MELex (**M**alay-**E**nglish **lex**icon). The process of generating MELex is illustrated in Fig. 1, and several definitions were given as follows.
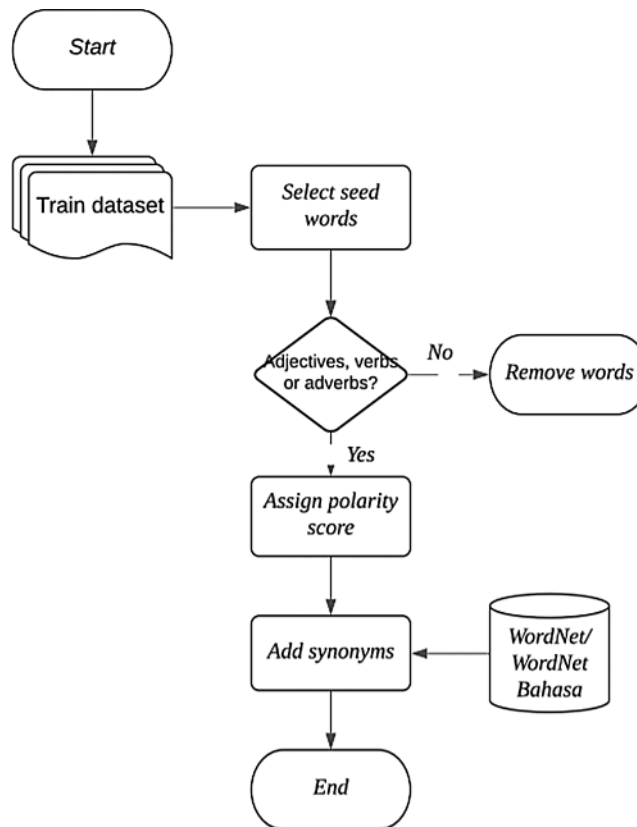
Figure 1: Flow diagram of the construction of MELex

2.1 Definitions

A training data D contains tweets t_i, and their manually annotated polarity score $s(t)_i$) was taken as input. Each t_i was annotated as +1 (positive) or −1 (negative). The output from this process is a bilingual sentiment lexicon *MELex* containing entries of *O: P* where the opinion indicator *O* is each word tagged as adjectives, adverbs, or verbs in the training data D and *P* is its polarity value.

The algorithm for the construction of *MELex* is shown in Algorithm 1.

Algorithm 1: Construction of MELex

Input: Training data $D = \{t_i, \ s(t)_i)\}$
Output: Sentiment lexicon $MELex = \{S : P\}$ where *S*: sentiment word, *P*: polarity value
Initialization: $TF = 0$, where *TF*: total number of *S* in *D*
Begin
1: For each word w in t_i do
2:     #seed word selection
3:         If $w = $ [adj, adv, verb] then
4:             $w = S$
5:         End If

(Continued)

6:    #polarity assignment
7:        For each S in t_i do
8:             assign P, TF
9:        #calculate polarity score P
10:            $P = (s(t)_i)/TF$
11:     End For
12:     For each S in D do
13:        If $P > 0$ then
14:           If $(TF < 5)$ then
15:              $P = 1$
16:           Else if $(5 >= TF < 10)$ then
17:              $P = 2$
18:           Else
19:              $P = 3$
20:           End if
21:        Else if $P < 0$ then
22:           If $(TF < 5)$ then
23:              $P = -1$
24:           Else if $(5 >= TF < 10)$ then
25:              $P = -2$
26:           Else
27:              $P = -3$
28:           End if
29:           Remove TF
30:      End if
31:     End For
32:     #synonym expansion
33:     For each S in L do
34:           Search synonym Y in WordNet
35:           Add Y in L
36:           Assign P of S to Y
37:     End For
38:   End For
39: End For
End

There are three main activities in constructing MELex; seed words selection, polarity assignment, and synonym expansion elaborated further in the following sections.

2.2 Seed Words Selection

The first step is to detect subjective words in each t_i by their corresponding PoS Tagging. Following the work presented by Gupta and Yadav [9] as well as Shamsudin, Basiron, & Sa'aya [10], a combination of adjectives, verbs, and adverbs were used as an indicator of subjective contents, and hence those words were chosen as a candidate for seed word O in the Lexicon *MELex*. The words tagged other than adjectives, verbs or adverbs were excluded from the lexicon list as it does not carry

any sentiments. Tab. 1 shows a sample of seed word selection where the word '*mahal*' and '*jauh*' become the only O candidates for t_i '*rumah ini mahal dan jauh*'.

Table 1: Sample seed word O

e.g: rumah ni mahal dan jauh *(this house is expensive and far away)*					
t_i	rumah	ni	Mahal	dan	Jauh
PoS Tag	NOUN		ADJ		ADJ
O	-	-	Mahal	-	Jauh

2.3 Polarity Assignment

The next step is to assign polarity value P for each seed word O. This research employed the word vector space model [11] in calculating the score. The word vector space model returns a vector of length equal to the number of words of the sentence. Thus, as shown in Tab. 2, t refers to each tweet in the training data, and w indicates the word.

Table 2: Representation of the word vector model

	w_1	w_2	w_3	. . .	w_j
t_1	1	1	0	. . .	1
t_2	0	1	1	. . .	0
. . .	. . .	. . .	. . .	. . .	. . .
t_i	0	0	1	. . .	1

The score generated by the word vector model is used to determine the polarity of the sentence, while the *term frequency TF decides the weightage of each word O* within the training data D. The score of each seed word O is calculated according to Eq. (1).

$$S = \sum_{i=1}^{T} \left(\frac{1}{n} . s(t_i) \right) \tag{1}$$

where n is the number of tokens consists of adjectives, adverbs, and verbs in t_i. T indicates the total number of t_i in training data D, while$s(t)_i)$ is the polarity value manually assigned by the annotators for each t_i. The score S will return a value in the range of -1.0 to $+1.0$. Tab. 3 shows an example of how polarity score S is calculated.

As shown in Tab. 3, the S score for the word '*mahal*' and '*jauh*' is -0.5. Next, the score of seed word O will be summed up based on its total appearance within document D to get the final value. The v value equals the total summation of S divide by its term frequency TF as defined in Eq. (2)

$$v = \frac{\sum S}{TF} \tag{2}$$

As mention earlier in this section, the final polarity of P, either positive or negative, is determined by v. While the weight of the polarity, which is scaled from -3 to $+3$, is determined by the frequency TF

of the seed word O in the training data D. The polarity score P is defined as in eq. (3):

$$P = +1 \quad or \quad -1, \quad if \quad TF < 5 + 2 \quad or \quad -2, \quad if \quad 5 >= TF < 10 \tag{3}$$

Table 3: Polarity score v calculation

e.g: rumah ni mahal dan jauh (this house is expensive and far away)						$[s(t]_1)$	Total
t_1	rumah	Ni	mahal	Dan	Jauh	−1	
PoS Tag	noun		adjectives		adjectives		
n	0	0	1	0	1	2	2
S			−0.5		−0.5		

From Eq. (3), if *the TF* of an O is less than 5, the final value P will be assigned as +1 or −1. If *TF* is between 5 and 9, then the P-value will be + 2 or −2, and for *TF* above 9, the P-value is +3 or −3. Tab. 4 presents a few examples of seed word O with its v score, frequency *TF*, and P's final value.

Table 4: Samples of seed word, TF, and polarity value

1	O	v	TF	$\rightarrow$	1	O	P
2	mahal	−0.32623	79		2	mahal	−3
3	murah	0.142955	57		3	murah	3
4	mampu	0.000985	34		4	mampu	3
5	naik	−0.08276	29		5	naik	−3
6	lulus	0.14707	26		6	lulus	3
7	dapat	0.101789	26		7	dapat	3
8	besar	0.090385	26		8	besar	3
9	senang	0.007804	18		9	senang	3
10	jauh	−0.03235	17		10	jauh	−3
11	diluluskan	0.000486	9		11	diluluskan	2

2.4 Synonym Expansion

In order to add more candidate words, the WordNet interface from the NLTK package was used to extract synonyms for each seed word in the Lexicon *MELex*. Each synonym word was then assigned the same P score as the original seed word O.

As shown in Tab. 5, the word 'inexpensive' is the synonym of the seed word 'cheap'. Hence, the Lexicon *MELex* will be updated with the word 'inexpensive' and a positive score; +2, a similar score as its seed word 'cheap' is assigned.

Table 5: Sentiment words and their synonyms

Language	O	Synonyms
English	cheap	Inexpensive
Malay	mampu	berdaya, berkaliber, berkebolehan, berkemampuan

3 Experimental Setup

To evaluate the bilingual sentiment lexicon developed using our approach, we design an experimental setup that compares with a baseline lexicon and two state-of-the-art machine learning classifiers. In the following subsections, the details of the experiments are described.

3.1 Datasets

We have extracted Twitter data concerning the affordable housing projects in Malaysia. Overall, a total of 23 837 raw data have been collected. This raw data will go through the pre-processing stage, which will be explained further in the following section.

3.2 Data Preprocessing

The pre-processing phase is meant to clean and remove the unnecessary symbols from the data. The raw data requires some initial pre-processing before implementing sentiment analysis to avoid incorrect and misleading results.

Pre-processing task performs various activities such as the removal of repetitive tweets (re-tweets), URLs, symbols, and hashtags (#), handling words with repeated letters, identify languages either Malay, English, or *Bahasa Rojak* and Part-of-Speech (PoS) tagging. Various activities of data pre-processing applied in this research are shown in Fig. 2.

i) Removal of RTs - The first step in cleaning the raw data was the removal of repetitive tweets.
ii) Removal of URLs, symbols, and hashtags – All the unnecessary HTML tags was removed from the data using the Python module, Regex.
iii) Language Identification – Each tweet is classified into one of the following tags: My (Malay), En (English), Mix (*Bahasa Rojak*/Other languages). For this task, the Python library, langdetect, was used.
iv) Part-of-speech (PoS) Tagging – Once the language for each tweet is identified, each sentence in the datasets is tokenized and PoS tagged. Since there is no available Python package to handle mixed language, tweets identified other than My or En will be manually tagged as objective/subjective tweets. If it is subjective, the annotators will further annotate those tweets as either positive or negative sentiments.
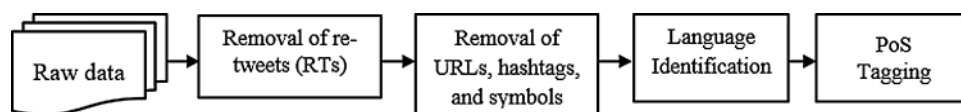
Figure 2: Pre-processing activities

After assigning each word with its PoS tags, sentences that match the adjectives, adverb, and verb rules are added to the candidate list of the sentiment analysis process. Finally, a total of 6666 tweets were used for the following process, and the data is divided into two sets which are:

  i) **Training data –** A total of 4666 tweets were used to construct a sentiment lexicon.
 ii) **Testing data** – A total of 2000 data were used for sentiment classification purposes.

3.3 Data Annotation

After pre-processing activities, the whole datasets were annotated by two native speakers of the Malay language, proficient in English. They are well aware of the case studies used in this research. The annotated datasets were used as a baseline for evaluation purposes during the testing phase. The annotators were asked to assign each tweet/sentence in the datasets with its polarity. Two polarity types, namely positive and negative, were used in this study.

In order to assess the reliability of agreement between both annotators, Cohen's Kappa score has been used to calculate the inter-annotator agreement rate. The agreement of both assessors based on Cohen's kappa is 0.84 for the positive sampled dataset and 0.89 for the negative sampled dataset.

3.4 Sentiment Classification

In order to examine the performance of MELex, classification on the testing data was applied. Algorithm 2 shows how the classification process was conducted.

Algorithm 2: Sentiment classification process

Input: Testing data T, Lexicon L, Negation file N
Output: $S = \{Pos, Neg$ or $Neut\}$, where *Pos*: Positive, *Neg*: Negative, *Neut*: Neutral
Initialization: Total_Pos and Total_Neg $= 0$, where
Total_Pos: accumulates the positive polarity t_i in T
Total_Neg: accumulates the negative polarity t_i in T
Begin
1:   For each t_i in T do
2:       For each word w in t_i do
3:          Search for w in L
4:       If w exist in L then
5:          assign $P\ w$ $\longleftarrow$
6:          If word before w exist in N then
7:             reverse P
8:          End If
9:       End If
10:      End For
11:      If $P > 0$ then
12:             Total_Pos Total_Pos $+ t_i$ $\longleftarrow$
13:          Else if $P < 0$ then
14:             Total_Neg Total_Neg $+ t_i$ $\longleftarrow$
15:      End If
16:   End For

(Continued)

17:	If Total_Pos > | Total_Neg| then
18:	$S = Pos$
19:	Else if Total_Pos < | Total_Neg| then
20:	$S = Neg$
21:	Else
22:	$S = Neut$
23:	End If
End	

Negation handling – In this work, negation was also handled in the classification process. The negation terms in Malay and English language added to a negation file N. Here, the polarity of sentiment term under the influence of negation is inverted. For instance, "*not bad jugak rumah ni*" would return positive sentiment due to the presence of the negation term "*not*" before the sentiment word "*bad*" where the negative polarity for the word "*bad*" as stated in MELex was reversed to positive polarity.

In order to calculate the final score of each tweet/sentence, the Term Counting method is applied. It is a method to classify positive and negative sentiments by counting the positive and negative polarity found in a text. The sentiment score of each review is calculated and categorized based on the score as mentioned in Tab. 6.

Table 6: Polarity criteria

Condition	Polarity
Sentiment score > 0	Positive
Sentiment score = 0	Neutral
Sentiment score < 0	Negative

Based on Tab. 6, a tweet is categorized as positive if the highest polarity score is positive and vice versa. A tweet will turn neutral if the sentiment score is equal to 0. The highest polarity score determines the final polarity.

3.5 Performance Evaluation

Tab. 7 shows the confusion matrices for sentiment classification.

Table 7: Confusion matrix

	True (Predicted)	False (Predicted)
Positive (Actual)	True Positive (TP)	False Positive (FP)
Negative (Actual)	True Negative (TN)	False Negative (FN)

Based on the confusion matrix,

Accuracy A is defined as:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Precision P is defined as:

$$P = \frac{TP}{TP + FP} \tag{5}$$

Recall R is defined as:

$$R = \frac{TP}{TP + FN} \tag{6}$$

Finally, F-measure F is defined as:

$$F = 2.\frac{P.R}{P + R} \tag{7}$$

3.5 Baseline Experiments

In order to evaluate the performance of MELex, a comparison with other classifiers from both sentiment analysis approaches, lexicon-based and machine learning were employed.

i. General Purpose Lexicon

AFINN-111 – AFINN Lexicon is an English-based sentiment lexicon developed by Nielsen [12]. This sentiment dictionary used a polarity scale ranging from -5 (extremely negative) to $+5$ (extremely positive), and it includes 3382 English words. Following the work presented by Tan et al. [13] in creating the SentiLexM lexicon, all the English words in AFINN are translated into Malay, and the sentiment score is assigned similar to its original English words. Overall, the total number of words, including English and its Malay translated word, is 6764 sentiment words.

ii. Machine Learning Classifiers

This research relied on an open-source Python library, the scikit-learn library, to perform the machine learning classification [14,15]. In addition, two commonly used machine learning classification algorithms were considered, NB and SVM, and described further below for performance evaluation.

Feature extraction – As machine learning classification requires mathematical formats to train the models, the textual data must be transformed into numeric form. Feature extraction is meant to perform the task of converting the textual data into a numeric form.

Train the Model – Once the data is split into training and testing, the machine learning algorithms are applied to learn from the training data. In this study, the NB and SVM algorithms are chosen due to their ability to acquire promising results in the previous research.

1. **Naïve Bayes (NB) –** The NB classifier relies on Bayesian probability, and it is a well-known machine learning technique for sentiment analysis due to its simplicity and effectiveness.
2. **Support Vector Machine (SVM)** – SVM classifier has shown to be highly effective for the classification task, and it generally outperforms other machine learning classifiers. Unlike the NB classifier, SVM utilizes a hyperplane to separate classes and is represented by a support vector that distinguishes the positive and negative training vectors.

4 Results and Discussion

This section presents the results obtained to develop sentiment lexicon, MELex, and sentiment classification for the test set.

4.1 MELex

As explained in the previous chapter, MELex is a lexical resource constructed using tweets data as a seed word. Therefore, the number of words generated in the Lexicon is presented below.

Seed word – MELex contains 2220 entries as a seed word with 1069 positive and 1151 negative words. Fig. 3 illustrates that the Word cloud represents the 50 most frequent sentiment words extracted from Twitter and stored in MELex.

Figure 3: Word cloud: Most frequent sentiment words

Synonym expansion – Through the synonym expansion process, a total no of sentiment words in MELex achieve 6132 words, as shown in Tab. 8.

Table 8: MELex dictionary

Class	No of words
Positive	3521
Negative	2611
Total	6132

4.2 Sentiment Classification

The sentiment classification has been implemented on the datasets, and Tab. 9 presents the output samples.

In order to evaluate the performance of MELex, the results obtained were compared with the manually annotated testing data. Using the evaluation measures of accuracy, precision, recall, and F-measure, the results are calculated and compared. The sentiment classification was performed for mixed-language and single language content. The results from both mixed and single language were then summed up to get the overall results of the test set. The division between mixed and single language is to examine the influence of mixed language's accuracy on the overall performance.

Table 9: Sample of the output: Tweets and their polarity

Tweets	Polarity
mohonlah sekarang, cerah masa depan. seremban pun dah dekat ngan kuala lumpur *(apply now for a bright future. Seremban is near to Kuala Lumpur)*	Positive
padan muka group of women cheated greedy vips in housing scam *(serve you right, group of women cheated greedy vips in housing scam)*	Negative

A total of 2000 tweets were used as testing data. Out of 2000 tweets, 250 tweets consist of mixed language data. Tab. 10 shows the type of data that indicates that mixed language content contributed 12.5% to the total number of test data.

Table 10: Type of data

	No of data
Mixed language	250
Single language	1750
Total	**2000**

Mixed Language – Tab. 11 presents the classification results of 250 mixed language tweets extracted from the test set, which show that 85 positive reviews and 136 negative reviews have been classified correctly. Seventeen reviews have been classified as neutral. The evaluation metrics are calculated based on the overall confusion matrix and presented in Tab. 12.

Table 11: Confusion matrix: Mixed language

Polarity (Manually annotated)	No of sentences	Predicted positives	Predicted negatives	Predicted neutral
Positive	98	89	7	6
Negative	152	23	125	

The classification using MELex has yielded an accuracy of 87.7% for 250 mixed-language sentences, with high values of recall, as shown in Tab. 12. However, the low precision achieved indicates that the classification using MELex has misclassified negative reviews more than the positive ones.

Single language – 1750 tweets consist of Malay or English, showing that single language tweets contributed 87.5% towards the overall results. Tab. 13 shows the results obtained from the classification using MELex.

From the result shown in Tab. 14, the accuracy achieved using MELex is 90.35%, with a precision and recall of 84 and 92, respectively.

Table 12: Evaluation metrics

Metric	(%)
Accuracy	87.7
Precision	79.5
Recall	92.7
F-measure	85.6

Table 13: Confusion matrix: Single language

Polarity	No of sentences	Predicted positives	Predicted negatives	Predicted neutral
Positive	697	607	51	40
Negative	1053	114	938	

Table 14: Evaluation metrics

Metric	(%)
Accuracy	90.35
Precision	84
Recall	92
F-measure	88

Overall classification – The overall classification result for the dataset with regards to the accuracy, precision, recall, and F-measure are detailed out as follows:

As can be seen in Tab. 15, it is clear that the results obtained are promising, with an overall accuracy of 90.01% using the MELex lexicon. The classification of mixed language alone has contributed approximately 11% to the overall accuracy. Even though the classification suffers from a low precision value, the results achieved relatively high values in both recall and F-measure.

Table 15: Overall performance

	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Mixed language	10.96	9.9	11.59	10.7
Single language	79.05	73.5	80.5	77
Total	**90.01**	**83.4**	**92.09**	**87.7**

4.3 Performance Comparison

The main objective of this evaluation is to determine whether the proposed sentiment analysis approach using MELex can improve sentiment classification accuracy. Tab. 16 presents the performance's results for the first experiment regarding overall accuracy, average precision, average recall, and average F-measure.

Table 16: Performance comparison: Mixed language

Technique	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
MELex	**87.7**	**79.5**	**92.7**	**85.6**
AFINN	82.68	75	83.5	79
SVM	76.4	77	76	76
NB	78	78	78	78

Tab. 16 summarizes the performance measures for mixed-language contents, which shows MELex achieved the highest accuracy with 87.7%. As for the SVM classifier, it performs the worst as compared to the others. It can be concluded that classification using MELex is significantly better than other classifiers.

As shown in Tab. 17, the most accurate result was achieved by MELex for the dataset with an accuracy of 88.79%. MELex gained a high score on recall which indicates a low false negative. The classification using the NB technique has produced the worst accuracy in the test set. It can be seen that both lexicons, MELex and AFINN outweigh machine learning techniques except on average precision. The accuracy, recall, and F-measure using the AFINN lexicon are reasonable and close to the accuracy obtained using MELex. The classification performed using MELex in this study achieved a reasonable classification accuracy comparable to previous sentiment analysis research.

Table 17: Performance comparison: Overall classification

Technique	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
MELex	**90.01**	**83.4**	**92.09**	**87.7**
AFINN	85.29	75.4	90.8	82
SVM	77.88	80	80	80
NB	77.68	78	78	78

5 Discussions

The previous section offers empirical results of sentiment classification tasks for Malaysia's affordable housing reviews. The experiments were conducted to classify mixed language and single language content. The performance of the developed sentiment lexicon known as MELex was compared with AFINN, which is one of the general-purpose sentiment lexicons widely used in other research and two well-known machine learning methods; NB and SVM classifiers. It was found that the classification using MELex remarkably outperforms the baseline approaches either through lexicon-based or machine learning approaches. The main reason for the success of MELex is that it

relies on the bilingual and domain-specific Lexicon. In addition, the use of Twitter data that is specific for the Malaysian property domain as the seed words have a massive impact on boosting the results.

Furthermore, it is noticeable that the accuracy obtained using MELex is even higher than the accuracy reported by the previous works on *Bahasa Rojak*, such as RojakLex, which reported 71.9% accuracy and SentiLexM with 78.5% accuracy. Besides, to examine whether the accuracy of mixed language classification will improve the overall performance, an experiment solely on mixed language sentences has been carried out. Out of 2000 test data, 227 data are coming from a mixed language with approximately 11%. Therefore, it was concluded that the classification of mixed language using MELex had contributed about 10% towards the overall classification accuracy.

In addition, the incorrect classification of mixed language has only contributed 9.8% towards the misclassification results, which consider as a small portion. While in terms of the overall performance, the accurate classification of mixed language has contributed 11.7% towards the total accuracy of the test set. One crucial observation is the performance of MELex in classifying mixed-language sentences, which were reasonable and comparable with the other baseline methods. In addition, this research is in line with several works reported previously where machine learning techniques are less effective in classifying low resource language [16–18]. One of the advantages of the lexicon-based approach is that it works even in situations with no available labeled data. In terms of lexical resources, only the sentiment lexicon is needed to calculate the overall sentiment of a property review. Another advantage of obtaining a domain-specific sentiment lexicon is that it can be reused 'as-is' in other property projects. Furthermore, it shares similar domains since the domain-specific lexicons are similar, and the calculation of the overall sentiment is straightforward.

6 Conclusion and Future Work

In this paper, a new approach in constructing a bilingual sentiment lexicon using the combination of term frequency and word vector representation has been proposed. We prove the effectiveness of the classification via experiments using the newly constructed bilingual and domain-specific sentiment lexicon known as MELex. The result obtained from the experiments is improved using the proposed approach. Besides, few experiments involving the general Lexicon and machine learning approaches were conducted for comparison purposes. In comparison to sentiment classification using AFINN, SVM, and NB, MELex has obtained better results, which indicates the performance of the proposed sentiment analysis approach is effective in this experiment. Furthermore, the performance of MELex in analyzing mixed language content has been evaluated as well. Additionally, we explained the misclassifications that lead to inaccurate results in this research. In summary, the research has shown promising results in property domains during the evaluation phase and achieved better results than previous research in similar areas of study.

Future research will focus on handling more complex sentence calculations such as bi-grams and trigrams using the generated sentiment lexicon. Another direction is to include more sentiment words in the lexical resource to enhance the capability of MELex. In order to expand the Lexicon, slang, and dialects, words commonly used by Malaysians will be considered as Lexicon's candidate because these types of words might provide helpful information in determining the sentiments. Furthermore, antonyms of sentiment words in MELex that can be extracted from WordNet should be considered to expand the Lexicon.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies,* vol. 5, no. 1, pp. 1–167, 2012.

[2]   J. Feng, C. Gong, X. Li and R. Y. K. Lau, "Automatic approach of sentiment lexicon generation for mobile shopping reviews," *Wireless Communications and Mobile Computing,* vol. 2018, Article ID 9839432, 13 pages, 2018. https://doi.org/10.1155/2018/9839432.

[3]   H. Cho and S. -H. Choi, "Automatic construction of movie domain Korean sentiment dictionary using online movie reviews," *International Journal of Software Engineering and its Applications,* vol. 9, no. 2, pp. 251–260, 2015.

[4]   K. Jaidka, S. Ahmed, M. Skoric and M. Hilbert, "Predicting elections from social media: A three-country, three-method comparative study," *Asian Journal of Communication,* vol. 29, no. 3, pp. 252–273, 2019.

[5]   S. L. Lo, E. Cambria, R. Chiong and D. Cornforth, "Multilingual sentiment analysis: From formal to informal and scarce resource languages," *Artificial Intelligence Review,* vol. 48, no. 4, pp. 499–527, 2017.

[6]   K. Becker, V. P. Moreira and A. G. L. dos Santos, "Multilingual emotion classification using supervised learning: Comparative experiments," *Information Processing & Management,* vol. 53, no. 3, pp. 684–704, 2017.

[7]   K. Dashtipour, P. Soujanya, H. Amir, C. Erik, H. YA. Ahmad *et al.*, "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognitive Computation,* vol. 8, no. 4, pp. 757–771, 2016.

[8]   D. Vilares, M. A. Alonso and C. Gómez-Rodríguez, "Supervised sentiment analysis in multilingual environments," *Information Processing & Management,* vol. 53, no. 3, pp. 595–607, 2017.

[9]   A. Gupta and D. K. Yadav, "Semantic similarity measure using information content approach with depth for similarity calculation," in *Int. Journal of Scientific & Technology Research*, vol. 3, no 2, pp. 165–169, 2014.

[10] N. F. Shamsudin, H. Basiron and Z. Sa'aya, "Lexical based sentiment analysis-verb, adverb & negation," *Journal of Telecommunication, Electronic and Computer Engineering,* vol. 8, no. 2, pp. 161–166, 2016.

[11] L. Gui, Y. Zhou, R. Xu, Y. He and Q. Lu, "Learning representations from heterogeneous network for sentiment classification of product reviews," *Knowledge-Based Systems,* vol. 124, pp. 34–45, 2017.

[12] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," in *Proceedings of the ESWC Workshop on 'Making Sense of Microposts': Big things come in small packages*, Heraklion, Crete., pp. 93–98, 2011.

[13] Y. F. Tan, A. Azlan, H. S. Lam and W. K. Soo, "Sentiment analysis for telco popularity on twitter big data using a novel Malaysian dictionary," *Advances in Digital Technologies. Frontiers in Artificial Intelligence and Applications*, vol. 282, pp. 112–125, 2016.

[14] H. Alshalabi, S. Tiun, N. Omar and M. Albared, "Experiments on the use of feature selection and machine learning methods in automatic malay text categorization," *Procedia Technology,* vol. 11, pp. 748–754, 2013.

[15] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in *Seventh Int. Conf. on Contemporary Computing (IC3)*, 7–9 August, Noida, India, pp. 437–442, 2014.

[16] N. I. Zabha, Z. Ayop, S. Anawar, E. Hamid and Z. Z. Abidin, "Developing cross-lingual sentiment analysis of malay twitter data using lexicon-based approach," *International Journal of Advanced Computer Science and Applications,* vol. 10, no. 1, pp. 346–351, 2019.

[17] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM,* vol. 56, no. 4, pp. 82–89, 2013.

[18] R. Wijayanti, and A. Andria, "Automatic Indonesian sentiment lexicon curation with sentiment valence tuning for social media sentiment analysis," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP),* vol. 20, no. 1, pp. 1–16, 2021.