ISSN 2278-3091
Volume 9 No.2, March - April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse128922020.pdf
https://doi.org/10.30534/ijatcse/2020/128922020

SENTIPUBLIKO: Sentiment Analysis of Repost Jejemon Messages using Hybrid Approach Algorithm

Adomar L. Ilao[1], Arnel C. Fajardo[2]
[1]Technological Institute of the Philippines, Quezon City-Campus, Philippines, alilao@mcl.edu.ph
[2]Technological Institute of the Philippines, Quezon City-Campus, Philippines, acfajardo@gmail.com

ABSTRACT
Jejemon language becomes a form of communication dialect. It was a form of expression used by a particular social group unknown as Jejemon. However, the Jejemon expression has different formats ranging from basic form of changing letter to number, lowercase letter to uppercase letter, inserting shortcut texts into more complicated format. This paper aims to classify Jejemon tweet whether it is a positive, negative or neutral sentiment through sentiment analysis techniques. Experiment included translation of Jejemon formatted tweet, reduction of sentiment scores on repost tweets and sentiment classification. Analysis of experiment results involves Paired T-Test, confusion matrix, precision, recall, f-score and accuracy. Evidently, translated Jejemon tweet resulted 78.5% similar from the actual message using cosine similarity algorithm. Furthermore, Paired T-Test shows no significant difference between new sentiment scores from translated expression and actual sentiment scores using Hybrid Algorithm. Sentiment analysis metrics such as precision, recall, f-score and accuracy show acceptable values of 71%, 76%, 71% and 73% respectively.

Key words: Sentiment Analysis, Social Media, Jejemon, Hybrid Algorithm, Cosine Similarity, Dictionary Substitution Approach, Tweet

1. INTRODUCTION

Human language is constantly changing. It happens across time and social groups. These changes on human language yields negative perception from people who are unable to cope with new vocabulary or new visual representation. The downside of the new trends might result possible miscommunication between social groups.

In the Philippines, several languages were invented to cater specific social groups. Predominantly, millennial and member of the third sex are the most socially active group in terms of modern language expression [1]. They developed new language semantics such as Jejemon (p30pL3, o+h3r, pl4c3s) [2] and bikemon (Aglipay, Chiquito, Churchill). However, the most controversial language trend in 2016 is Jejemon [3].

Nevertheless, Jejemon language users suffered from weak speaking ability and mangled word spelling as observed by English teachers. They average 12 text messages everyday. Each jejemon message normally is composed of symbols and phonetics. However, the advent of technological communication medium focuses on social connection such as social media created new Jejemon followers in Twitter aside from JEJETYPING, wearing a jeje-hat and jeje-photos online [4].

The overwhelming success of Jejemon language lead to an award as Word of the Year in 2010 by the Filipinas Institute of Translation Incorporated based on significant impact on Filipino life in terms of socio-cultural, political, social and economic [5].

As an influential millennial language, Jejemon expression is a result of self-expression which designed to resolve concerns on limited available space provided by text messages and Twitter [5]. However, it is significantly important beyond translation to understand the actual feeling behind the person's message or opinion.

Classification of one's opinion can be done through sentiment analysis. It can classify opinion whether positive, negative or neutral via polarity score. Unfortunately, other factor to consider is the impact of repost known as retweeting in Twitter towards sentiment classification in a document level.

This study is designed to evaluate the accuracy and precision Hybrid Algorithm developed by Ilao and Fajardo [6] applied to Jejemon Tweets. Furthermore, integration of string similarity algorithm in reducing sentiment polarity score for repost or retweet messages will provide better understanding of the side effect of repost messages in a document level sentiment evaluation.

2. REVIEW OF RELATED LITERATURES

2.1 Jejemon Word Structure

Jejemon language becomes popular 2010 based on the limited space available for text messages and tweets [5]. It is primarily composed of alphabet known as Jejebet.

Jejebet uses Roman alphabet, Arabic numerals and other special characters namely 4, b, c, D, 3, f, 6, h, 1, j, k, 7, m, N, 0, p, Q, r, 5, t, u, V, w, x, Y, or z. Jejemon word is arranged in alternating capitalization, over-usage of the letters H, X or Z and mixture of numeric characters and English alphabet [5] as shown in Table 1.

Table 1: Example of Jejemon Expression [5]

Characteristics of Jejemon	Example
Insertion of unnecessary numbers and letters.	phfue or p0w
Unique orthography based on how the words sound	eHyUoeW fPuoEh
Unconventional use of punctuations	psenxa na ha!!
Numbers to substitute letters	bzt4h
Alternate use of lower and upper case	WE wnT 2 BE~ P0wh.
Use of onomatopoiec lexis/emotional language	tnx pfowh jejeje
Lengthening of vowels and consonants	TAMAAA!
Substitution of spelling	Maq

2.2 Jejemon Translation

Jejemon translation is available online. One particular online translator is Jejemon Translator found in http://173.254.110.65/jejeschool/index.php. It is capable to translate English language to Jejemon expression as shown in Figure 1.

Figure 1: English Language to Jejemon Language Translation.

It was applied to transform the English thesis document version [7] to jejenese or Jejemon language document version available in iskwiki.upd.edu.ph. It shows the reliability of the Jejemon translator as an effective technical tool. The jejemon document revision demonstrated combination of three Jejemon techniques namely numbers to substitute letters, symbols to substitute letters and alternate use of lowercase and uppercase letters.

An alternative online Jejemon translator can be found in http://akosijairah.blogspot.com/2010/04/Jejemon-translator-v3_28.html as shown in Figure 2. The translated expression is comprised of case conversion, p0wh insertion and modification of word to a totally different spelling.

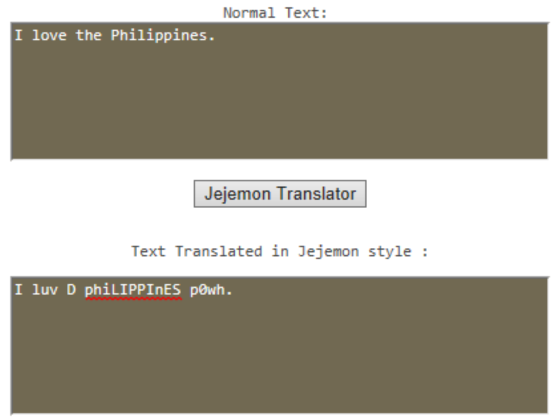

Figure 2: English Language to Jejemon Language Translation.

2.3 Dictionary Substitution Approach

Dictionary Substitution Approach is a technique commonly identified as search and replace approach. The key characteristic is to match word from the corpus. However, when multiple entries are found from the corpus, randomize word will be selected from list possible alternative [8].

Jejemon language does not follow specific pattern. Through Dictionary substitution approach it can replace non-standard Jejemon token into a meaningful context of English or Filipino word.

English or Filipino sentence found in Table 2 was translated to Jejemon equivalence from three sources. It shows three different techniques of Jejemon translation.

Table 2: Sample Online English Expression translated into Jejemon Expression

English Sentence	URL Source of Jejemon Translator	Jejemon Expression
I would like to know more about you, care to tell me your name? Hehehehe!	https://pinoychronicle.wordpress.com/Jejemon/	i wuD LLyK tO knOw moR3 bOut u. crE 2 t3ll mE yur N@me? jejejejeje!
	Online Jejemon Translator (http://173.254.110.65/jejeschool/index.php)	1 wUD 77yk +0 kn0w mUhr3 4b0U+ U', cr +0 +377 m3 U'r nm3?" j3j3j3h3!
	Online Jejemon Translator (http://akosijairah.blogspot.com/2010/04/Jejemon-translator-v3_28.html)	i WOUlD Lyk To KNow more aboutz u, CeyRTo Tell me uR nMe, n0H? JEjEJEjE LOLz!

2.4 Lexicon-Based Algorithm

Lexicon-based Algorithm works by defining rules to classify the opinion which is created by tokenizing every sentence in each document and testing if the token or word is present in the database [8]. It is based on rule which is composed of antecedent and consequent. An antecedent defines a condition and consists of either a token or a sequence of tokens. This process provides a technique to single out positive, negative or neutral about the subjective opinion [9][10][11].

It uses sentiment lexicon to assign a polarity value. A lexicon is comprised of words or phrase where each label is categorized based on polarity value whether positive or negative orientation [12]. In building a sentiment lexicon have three strategies namely hand-craft elaboration, automatic expansion from an initial list of seed words and corpus-based approach.

A comparative study [13] on Lexicon-based review involving AFINN, General Inquirer, Micro-WNOP, Opinion Lexicon, SentiSense, SentiWordNet, Subjectivity Lexicon and WordNet-Affect. The investigation resulted 78% accuracy towards SentiWordNet which utilizes WordNet corpus.

2.5 String Similarity Algorithm

String based similarity measurement defines the similarity of strings in terms of the longest prefix common to both strings. It is applied to several fields namely data cleaning, data integration, error checking or pattern recognition [14]. It uses either character-based or term-based technique [15]. The commonly used string based similarity measurement. It is a term-based technique known as edit distance algorithm which performs minimum number of insertions, deletions or substitutions to string1 to string 2 [14].

The comparative study on edit distance algorithms namely Q-gram similarity, cosine similarity and dice coefficient similarity. The study resulted in favor of cosine similarity algorithm with an average accuracy of 63% [16]. Cosine similarity algorithm measures two-finite-dimensional vectors of the same dimension [15]. Furthermore, cosine similarity was applied to analyze the similarity of sentiment scores from SentiWordNet and the similarity of each sentiment score contributed to product review rating prediction [17].

2.6 Hybrid Polarity Score Algorithm

Each synset polarity score is derived by computing the average of SentiWordNet algorithm and VADER algorithm as elaborated in Equation 1, Equation 2 and Equation 3 [18]. Normally, VADER polarity score use compound score to determine the sentiment classification. However, Ilao's Hybrid algorithm used VADER's positive and negative scores to derived sentiment score.

Hybrid positive score = (SentiWordNet Positive Score + VADER Positive Score)/2 (1)
Hybrid negative score =(SentiWordNet Negative Score + VADER Positive Score))/2 (2)
Over-all Hybrid Score (OHS) = (Hybrid positive score – Hybrid negative Score) (3)

Where:
If OHS >0, then sentiment is "Positive".
If OHS<0, then sentiment is "Negative".
If OHS=0, then sentiment is "Neutral".

The number of occurrences of positive, negative and neutral tweets will determine the over-all sentiment of the entire population of collected political tweets. The hybrid approach was experimented to different political datasets. The experiment yielded 88.33% accuracy better than SentiWordNet and VADER algorithm.

3. METHODOLOGY

The study was designed to implement sentiment analysis approach as illustrated in Figure 3.

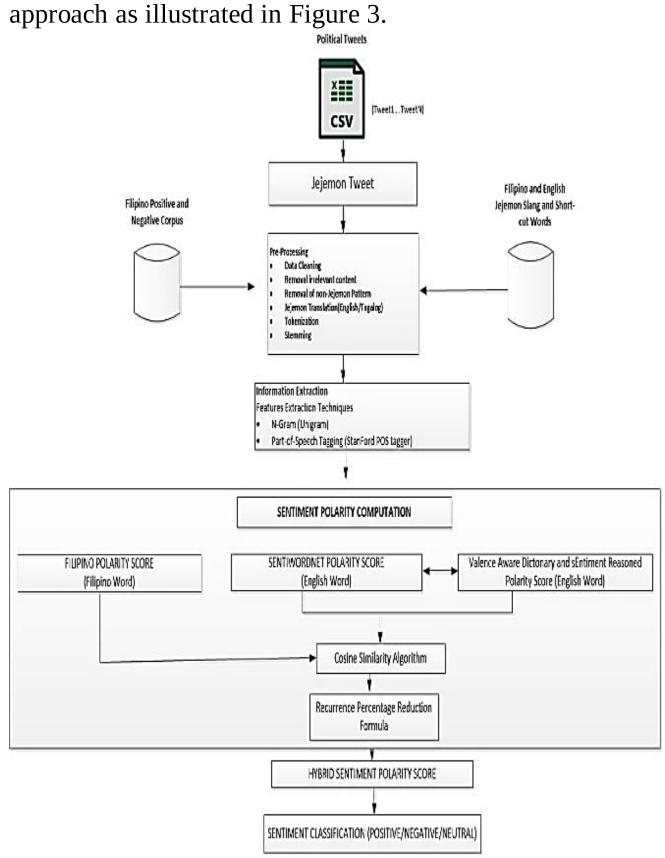

Figure 3: Jejemon Repost Hybird Polarity Score Algorithm Conceptual Framework

It involves several stages such as data cleaning, removal of irrelevant words, dictionary substitution and loan translation approach, tokenization, stemming, feature extraction, hybrid sentiment polarity score approach (SentiWordNet Score, VADER score, Filipino Score, Recurrence Percentage Reduction Score), and sentiment classification.

3.1 Pre-Processing

3.1.1 Data Cleaning

It involves elimination of unwanted expression such as removal of punctuations, website URL, emoticons, special characters and apostrophes.

3.1.2 Removal of Irrelevant Words

It is intended to remove irrelevant contents such as slang words [19], stop words, Jejemon expression (example: poWH, jejeje, xD) and non-Jejemon formatted text.

3.1.3 Dictionary Substitution and Loan Translation Approach

Jejemon lexeme might contains abbreviation or shortcut text in English or Filipino expression. A CSV file and Text file will stored shortcut English text, shortcut Filipino text and Jejemon most sentimental words. When translated lexeme is found from the seedlist, detected text will be replaced into its actual value. It also involves translation of Jejemon expression to purely English expression, purely Filipino expression and combination of both languages.

Translation utilized enchant libraries [20] and collection of Filipino words from Tagalog dictionary [21] to cross-checked spelling or possible word suggestion in transforming complicated Jejemon expression.

Furthermore, customized Tagalog dictionary was constructed from collection of rubbish words of English or Filipino words will resolve some ambiguity coming from Jejemon expression translation.

3.1.4. Tokenization

The process of tokenization segments where Jejemon expression might be in the form of paragraph, sentences and word into a lexeme (single word).

3.1.5 Stemming

In this stage, Jejemon lexeme will subjected to extraction of base word by simplifying plural form to singular or by removing prefix, suffix and infix.

3.2 Feature Extraction

3.2.1 N-gram Approach

The study will implement a hybrid approach in classifying the given Jejemon expression whether positive, negative and neutral sentiment. The two lexicon-based algorithms that was proven effective by Ilao's study in 2019 combined SentiWordNet and VADER algorithm polarity scores as shown in Equation 1, Equation 2 and Equation 3 in which uni-gram and tri-gram approach were implemented by the said algorithms respectively.

3.2.2 Part-of-Speech (POS) Tagging

Each English lexeme will be tagged via StanFord POS Tagger. Each tag will determine whether the lexeme can be a source of sentiment of SentiWordNet.

However, Filipino lexeme will not undergo POS Tagging procedure. The study will used Tagalog Corpus in identifying positive and negative political words.

3.3 Hybrid Approach

3.3.1 SentiWordNet Polarity Score

SentiWordNet polarity score features positive, negative and neutral scores. Nonetheless, Hybrid approach requires positive and negative elements of its derived polarity score.

3.3.2 VADER Polarity Score

Classification of sentiment using VADER normally uses compound score. However, hybrid approach will apply positive score and negative score as shown in Equation 1.0 and Equation 2.0.

3.3.3 Filipino Sentiment Polarity Score

Filipino sentiment polarity score will be applied for Filipino expression based on Equation 4.0.

For every detected positive or negative word from collected list of commonly used tagalog political sentiment word will be scored as 1 point.

Filipino Sentiment Polarity Score (FOSS) = (Number of Positive Words-Number of Negative Words)/Number of Words in the Tweet (4)

Where:
If FOSS >0, then sentiment is "Positive".
If FOSS<0, then sentiment is "Negative".
If FOSS =0, then sentiment is "Neutral".

3.3.4 Cosine Similarity

Cosine Similarity algorithm will be used to determine if there are some similarities between posts. If cosine similarity returns a string similarity value of 70%, there exists a significantly identical tweet between the lists of collected tweets which is considered as a repost message.

3.3.5 Recurrence Percentage Reduction Score

After string similarity algorithm identified similarity between tweets, Equation 5.0 and Equation 6.0 will define reduction score from previously derived value and generate a new polarity score for repost message.

Relative Frequency Rate = (number of item repost /total number of post messages)x100 (5)

Recurrence Percentage Reduction Score= Over-all Hybrid Score - (Relative Frequency Rate x Over-all Hybrid Score) (6)

3.4 Sentiment Classification

The study was designed to perform sentiment analysis in a document level. Each Jejemon expression will be classified whether positive, negative or neutral. Afterward classification, each identified sentiment will be counted individually; the highest number of frequencies between positive, negative or neutral will generally classify the over-all sentiment of the collected instances.

4. DATA SOURCE

The study prepared dataset collected from Twitters accounts namely Jejemonilao and Jejemonkami from January 15,2020 up to February 29,2020 as stated on Table 3, Table 4, Table 5 and Table 6 with 131 instances.

Dataset is consist of instances characterized of maximum number of 37 tokens, minimum number of 3 tokens and an average of 12 tokens per instance.

Jejemonilao tweets followed Jejemon format available in http://akosijairah.blogspot.com/2010/04/Jejemon-translator-v3_28.html considered as controlled environment. However, uncontrolled environment instances were extracted from Jejemonkami tweets. Tweets were formatted based on preference of participants from any available Jejemon patterns. Moreover, collected dataset is comprised of purely English Jejemon tweet, purely Filipino Jejemon tweet and combination of both.

Table 3: Sample of Jejemon Format Domain

Jejemon Format Domain	Expresion
Online Resource (Controlled Environment)	GOveRnmEnt'Z Slow respOnSe TO d cOvID-19 hz Led tO D ViRuz entErinG d pHilIpPinEs p0wh.
Personal Preference (Uncontrolled Environment)	i h8 tht d govERNmeNt prioRiTizE themsELVez ratheR THn D PeopLe

Table 4: Tweet Post Frequencies

Tweet Post Tweets	Frequencies
Single Post	116
Repost	15

Table 5: Jejemon Collected Tweet used Language

Jejemon Tweet used Languages	Frequencies
English	113
Filipino	14
Combination	4

Table 6: Sentiment Classified Frequencies

Sentiment Classification	Frequencies
Positive	29
Negative	76
Neutral	26

5. MEASUREMENT OF ALGORITHM PERFORMANCE

Algorithm performance will be validated in terms of paired T-Test, accuracy, recall, f-score and precision using Equation 7. Equation 8, Equation 9 and Equation 10 via confusion matrix values as shown in Table 7.

Table 7: Confusion Matrix

		Predicted	
		Negative (F)	Positive (T)
Actual	Negative (F)	FF	FT
	Positive (T)	TF	TT

$$\text{Accuracy} = \frac{FF+TT}{FF+FT+TF+TT} \quad (7)$$

$$\text{Precision} = \frac{TT}{FT+TT} \quad (8)$$

$$\text{Recall}=\frac{TT}{TT+TF} \qquad (9)$$

$$\text{F-Score}=\frac{2*Precision*Recall}{Precision+Recall} \qquad (10)$$

Where:
FF: the frequency of correctly predicted negative emotion.
FT: the frequency of incorrect predicted negative emotion.
TF: the frequency of incorrect predicted positive emotion
TT: the frequency of correct predicted positive emotion.

6. EXPERIMENT

The study evaluated the translation success rate of Jejemon instances to counterpart language via string similarity algorithm presented in Table 8.

Similarity Percentage is based cosine similarity value of translated expression against actual expression. Similarity values falls from 70% and above has “Strong Similarity” while values falls below 70% has “Weak Similarity”.

Table 8: Similarity Percentage of Translated Expression and Actual Expression

Jejemon Format Domain	No. of Strong Similarity (70% and Above)	No. of Weak Similarity (Below 70%)	Average Cosine Similarity Percentage
Online Resource (Controlled Environment)	92%	7%	81%
Personal Preference (Uncontrolled Environment)	73%	27%	76%

The controlled group’s translation similarity percentage ranges from 23% up to 100%. The factors which affect translation reflected Table 9 were word case, missing letters that were omitted by Jejemon expression when translated lead to a different English or Filipino expression due to ambiguity, proper space between Jejemon expression and single/plural forms. However, uncontrolled group’s similarity percentage ranges 25% up to 100%. Significant difference from controlled group, uncontrolled group’s factors are special characters such as Ë, ë, $, @, or ü ; repetition of letters namely “z”,”s”; two representation of 8 either “te” or “ate” , “i” either “!” or “1”, “a” either “4” or “@” and Filipino word interchangeably used “po” or “poh”.

Table 9: Cosine Similarity Percentages per Language

Languages	Lowest Similarity Percentages	Highest Similarity Percentages
English	23%	100%
Filipino	25%	100%
Combination	40%	70%

Table 10 shown the lowest similarity percentage of 23% belongs to English expression. Several words of an English instance were translated wrongly brought by ambiguity from the Jejemon expression namely missing letters were replaced by different letters and capitalization of words. While similar scenario happened to English and Filipino expression (combination) where it gained the lowest value under highest similarity percentage of 70% from the series of languages. Experiment encountered difficulties specially from missing letters. List of suggested words that might be suitable to replace rubbish word brought by missing letters comes from English or Tagalog Dictionary. There are some instances where an English word was taken as a Filipino word or vice versa.

After Jejemon translation, the new dataset comprised of English, Filipino or combination underwent pre-selection. As stated on previous studies on machine translation on several Filipino dialects, accuracy rates are 70.67% [22] and 69.5% [23]. Pre-selection of instances will be based on similarity percentage between 70% up to 100%. Qualified instances are revealed on Table 10.

Table 10: New Instances after Pre-Selection

Languages	Controlled Environment	Uncontrolled Environment
English	33	64
Filipino	0	2
Combination	1	1
Total	34	67

Table 10 shows 12% decreased of instances under the controlled environment from 39 instances down to 34 instances. Similarly, 27% decreased on uncontrolled environment instances from 92 instances down to 67 instances.

Table 11: Paired T-Test of Computed Sentiment Scores

		Mean	N	Mean Difference	p-value	Interpretation
Uncontrolled	New Score	-.07382433796	67	-.002619660970	.821	Not Significant
	Original Score	-.07120467699	67			
Controlled	New Score	-.06255770562	34	.016663253824	.257	Not Significant
	Original Score	-.07922095944	34			
Similarity Reduction	New Score	-.16391217813	32	-.017796546469	.246	Not Significant
	Original Score	-.14611563166	32			

Table 11 described the relationship between different sentiments' scores namely new score is derived from translated expression, original score is derived from the actual expression express either English, Filipino or combination provided by participants. Furthermore, cosine similarity found instances with 70% similarity considered as repost messages. All repost sentiments' scores underwent either 1% or 3% reduction. As stated on Table 11, p-values resulted 0.821, 0.257 and 0.246 are greater than 0.05 implied there is no significant difference between new scores and original scores under controlled environment, uncontrolled environment and after application of recurrence percentage reduction scores.

Table 12: Confusion Matrix of Controlled Environment

		Predicted	
		Negative (F)	Positive (T)
Actual	Negative (F)	11	1
	Positive (T)	7	6

Table 13: Confusion Matrix of Uncontrolled Environment

		Predicted	
		Negative (F)	Positive (T)
Actual	Negative (F)	31	3
	Positive (T)	9	11

Table 12 and Table 13 shown number of positive and negatives instances classified correctly or incorrectly. It appears out of 79 instances, 59 instances were classified correctly and 20 instances were classified incorrectly.

Table 14: Sentiment Analysis Metric Summary

Jejemon Format Classification	Label	Precision	Recall	F-Score	Accuracy
Online Resource (Controlled Environment)	Positive	46%	86%	60%	68%
	Negative	92%	61%	73%	
Personal Preference (Uncontrolled Environment)	Positive	55%	79%	65%	78%
	Negative	91%	78%	84%	
Average		71%	76%	71%	73%

Both negative sentiments resulted high precision with at least 91% and lowest precision value is 46% under Positive Controlled Environment. While, recall highest value is 86% falls under positive sentiment of Controlled Environment Domain; however the rest of the recall values are 61% up to 79%. Lastly, the highest f-score value is 84% under the negative sentiments of Uncontrolled Environment Domain whereas the remaining values are within 60% up to 84%.

The improved Ilao's hybrid model provided accuracies of 68% (Controlled Environment Domain), 79% (Uncontrolled Environment Domain) and an average of 74% from collected Jejemon format domain.

Generally, average scores of precision, recall, f-score and accuracy resulted 70% and above in term of classification performance towards the different Jejemon format domain.

7. CONCLUSION

This paper experimented on sentiment analysis solely focus on Filipino fascination in expressing their idea through Jejemon language.

However, Jejemon language raised issues on millennial English proficiency. Jejemon language as a form of communication is normally expressed through short-text and special representation of word or expression. Furthermore, Jejemon expression offered several language structure ranging simple to complicated techniques.

In general, study of machine translation from Jejemon expression into counterpart language namely English, Filipino or combination yielded successful conversion based on the cosine similarity done. The similarity rates of translated expression against actual expression gained 76% and 81% for controlled and uncontrolled environment respectively.

Even though some instances were not converted exactly as compared to actual expression, paired T-Test shows no significant difference between original polarity score and new polarity score derived from the two dataset domain. Furthermore, repost messages that underwent reduction through recurrence percentage reduction technique do not shows significant difference from the two domains based on the available data.

Moreover, sentiment classification through hybrid approach resulted positively in terms of consistency and accuracy based on the average precision, recall and accuracy such as 71%, 76% and 73% respectively. Average F-score of 71% signifies the acceptable classification performance since there is an uneven representation of sentiment instances from collected domains.

ACKNOWLEDGEMENT

Authors would like to express their greatest gratitude to Graduate School of Technological Institute of the Philippines, Quezon City for the support of realization of this paper publication.

REFERENCES

1. Ramos, J. J. R. (2018). **Bekimon in Social Networking Site: Modernized Idiolect and Sociolect**. *International Journal of Social Sciences & Humanities*, *3*(2), 16-25.
2. Nocon, N., Cuevas, G., Magat, D., Suministrado, P., & Cheng, C. (2014, October). **NormAPI: An API for normalizing Filipino shortcut texts**. In *2014 International Conference on Asian Language Processing (IALP)* (pp. 207-210). IEEE. https://doi.org/10.1109/IALP.2014.6973494
3. Mongaya, K.M. (2010). (**http://karlomongaya.wordpress.com/2010/07/09/the-internet-as -corrective-of-individualist-culture/**)
4. **The Social and Educational Influences of Jejemon Texting Style**.(2010) Date Retrieved: November 10, 2019
5. Tubac, Angelo. (2017). **LINGUISTIC INNOVATIONS IN THE JEJEMON PHENOMENON**. 10.13140/RG.2.2.18434.07360.
6. Ilao, A. & Fajardo, A. **Sentiment Analysis of Tweet Messages using Hybrid Approach Algorithm**. In 2019, The 17th International Conference on ICT and Knowledge Engineering (ICT-KE). IEEE. https://doi.org/10.1109/ICTKE47035.2019.8966887
7. Cataan, J. C. (2011). ***A 'World' Within the World: The Jejemons as the 'Other' Culture***, Unpublished Undergraduate Thesis, University of the Philippines College of Mass Communication.
8. Raghunathan, K., & Krawczyk, S. 2009. **CS224N: Investigating SMS Text Normalization using Statistical Machine Translation**. Department of Computer Science, Stanford University.
9. Devika, M. D., Sunitha, C., & Ganesh, A. (2016). **Sentiment Analysis: A Comparative Study on Different Approaches**. *Procedia Computer Science*, *87*, 44-49. https://doi.org/10.1016/j.procs.2016.05.124
10. Romanyshyn, M. (2013). **Rule-Based Sentiment Analysis of Ukrainian Reviews**. *International Journal of Artificial Intelligence & Applications*, *4*(4), 103
11. Kawathekar, S. A., & Kshirsagar, M. M. (2012). **Sentiments Analysis Using Hybrid Approach Involving Rule-Based & Support Vector Machines Methods**. *IOSRJEN*, *2*(1), 55-58. https://doi.org/10.9790/3021-0215558
12. Almatarneh S, Gamallo P (2018) **A Lexicon Based Method to Search for Extreme Opinions**. Plos One 13(5): E0197816. Https://DOI.Org/ 10.1371/Journal.Pone.0197816
13. Cho, H., Lee, J. S., & Kim, S. (2013). **Enhancing Lexicon-Based Review Classification by Merging and Revising Sentiment Dictionaries**. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (pp. 463-470).
14. Lu, W., Du, X., Hadjieleftheriou, M., & Ooi, B. C. (2014). **Efficiently Supporting Edit Distance Based String Similarity Search Using B+-Trees**. *IEEE Transactions on Knowledge and Data Engineering*, *26*(12), 2983-2996.
15. Khuat, Tung & Duc Hung, Nguyen & Thi My Hanh, Le. (2015). **A Comparison of Algorithms used to Measure the Similarity between Two Documents**. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 4. 1117-1121.
16. Chen, H. (2012). **String Metric and Word Similarity Applied to Information Retrieval** (Doctoral Dissertation, Master's Thesis, School Of Computing. University Of Eastern Findland).
17. Prasanna Lakshmi, K., Shraddha, V., Abhinava, V., Kavya, K., & Gayathri, R. (2017). **Sentiment Analysis And Prediction Using Text Mining**. Indian Journal Of Science And Technology, 10 (28). Doi: 10.17485/Ijst/2017/V10i28/113441
18. Ilao, A. & Fajardo, A (2019). **Sentiment Analysis of Tweet Messages using Hybrid Approach Algorithm**. 17th International IEEE Conference of ICT and Knowledge Engineering 2019 https://doi.org/10.1109/ICTKE47035.2019.8966887
19. **https://www.internetslang.com/trending.asp**. Accessed Date : March 10,2020
20. **https://abiword.github.io/enchant/**. Accessed Date: March 12, 2020.
21. **https://github.com/raymelon/tagalog-dictionary-scraper/blob/master/tagalog_dict.txt**. Accessed Date: March 12, 2020.
22. Domingo and R. Roxas. 2006. **Utilizing Clues in Syntactic Relationships for Automatic Target Word Sense Disambiguation**. Journal of Research for Science, Computing and Engineering. 3(3), 18-24. https://doi.org/10.3860/jrsce.v3i3.99
23. Alcantara and A. Borra. 2008. **Constituent Structure for Filipino: Induction through Probabilistic Approaches**. Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation (PACLIC). 113-122.