An Exploration of Criminality Linguistics on Social Networks Through Personality Traits

Saravanan Sagadevan[1], Nurul Hashimah Ahamed Hassain Malim, Muhammad Baqir Hakim Mohammad Bashir[1], Nurul Izzati Ridzuwan[1]
[1]Universiti Sains Malaysia, Penang, Malaysia
[1]nurulhashimah@usm.my

INTRODUCTION

Francis Galton hypothesized that natural language terms might represent the personality differences in humankind [1]. Personality is referred to "combination of emotion, behavior, motivation and thinking patterns of human that often mirror the true characteristics of them through their activities that are conducted intentionally or unintentionally" [2].

Therefore, by embracing the correlation with psychology-personality-linguistics, our earlier study successfully demonstrated the approaches to study the criminality writing in social network English messages using personality model, statistics and machine learning approaches [3]. Therefore, the current study embraced the similar approaches described in [3] and evaluated the representation of criminality element in Malay tweets messages. In this studies, we employed four main classifiers namely Sequential Minimal Optimization (SMO), Naive Bayes (NB), K- Nearest Neighbor (KNN) and J48 with ZeroR as baseline to measure the classification performance as well as presented some linguistic findings that associated to the criminality contents.

LITERATURE REVIEW

As pioneers of personality detection study using digital texts, Argamon and his colleagues [12] investigated the personalities of undergraduate students in the context of Neuroticism and Extraversion using Sequential Minimal Optimization (SMO) classifier. Since that, many studies had been conducted in heterogenous perspectives such as Saxena [13] employed Big 5 traits, tagging and Support Vector Machine (SVM) classifier to classify blog data, Firoj [5] investigated the personality of Facebook users using the Bag of the Word (BOW), Farnadi [7] achieved significant precision using SVM and psycholinguistic features, Golbeck [9] trained the ZeroR and Gaussian Processes learning classifiers on Tweets and eventually found the score of the five Big Personality traits to be within 11% - 18% of their actual values.

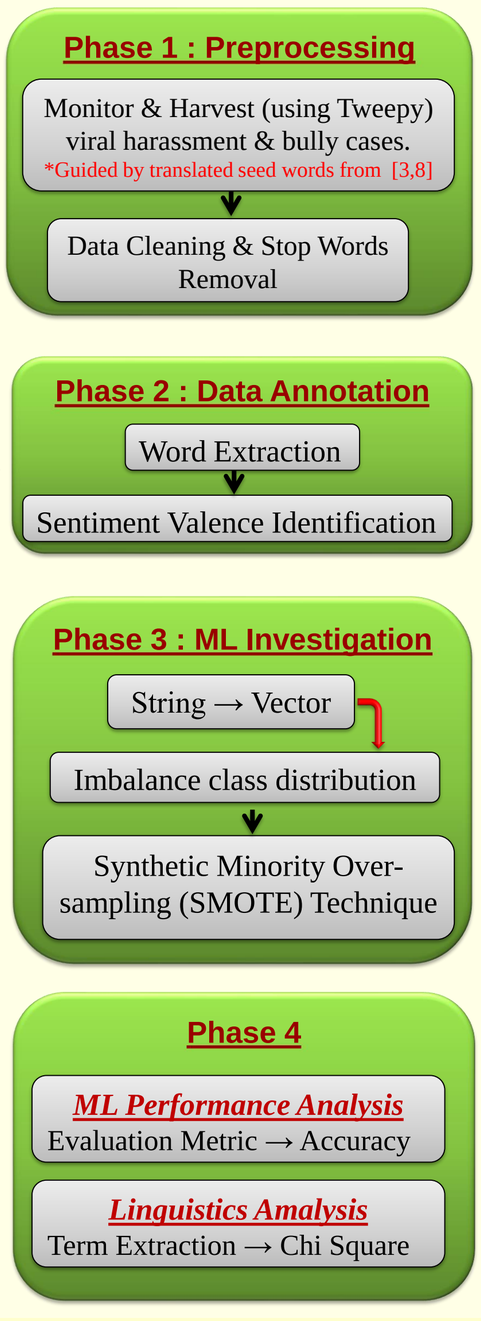
Our previous study used the foundation of PEN Model together with statistical analysis and machine learning classifiers to automatically classify the English Facebook & Twitter messages [3]. Three Factor Model consisted three main traits as illustrated in Table 1. The data annotation was conducted using the seed words labeled with sentiment valences that gathered through our preliminary study [3] and AFINN [8]. We have identified dozen of seed words that correlated with PEN model traits and our machine learning classification showed that SMO outperformed other classifiers as well as extracted some linguistics terms strongly correlated with criminality writings.

Table 1 : PEN Model Traits

Trait	Description
Extraversion	**Sociable, lively, active, assertive, sensation seeking, carefree, dominant, surgent and venturesome.**
Neuroticism	**Anxious, depressed, guilt feelings, low self-esteem, tense, irrational, shy, moody and emotional.**
Psychoticism	**Aggressive, cold, egocentric, impersonal, impulsive, antisocial, unempathetic, creative and tough-minded.**

METHODOLOGY

Fig 1 : Research Methodology



Language-Personality Bridge

The possible Malay sentiment words manually identified based on Noun (*kata nama*), Verb (*kata kerja*) and Adjective (*kata adjektif*). The relationship between the candidate words and personality traits determined through translating the extracted sentiment words (by using Kamus DBP) and identifying the sentiment valences labeled by previous studies [3] or assigned the valence through examining the semantical aspects of the words. In other words, the Malay terms that having similar semantical meaning directly translated whereas lexical that does not have similar semantical meaning were measure through the rules stated in Table 2. Afterwards, we used the term frequency method to calculate the presence of higher negative words (Psychoticism) followed by lower negative and positive sentiment words.

Table 2: Valence Categorization

Sentiment	Valence	Trait
Positive	Positive (1 to 5)	Extraversion
Negative	Lower Negative (-1 to -3)	Neuroticism
	Higher Negative (-4 and -5)	Psychoticism

Next, we used average Term Frequency (TF) method to calculate the intensity and tendency of the user towards the relevant PEN model traits using the formula stated in (1). We trace the terms starting from higher negative words to positive words due to the nature of the online writings typically constructed using more positive terms than negative.

(1) Average TF = $\frac{\text{Total number of traits based term}}{\text{Total number of users in each category of trait.}}$

*** Our analysis on Malay Tweets showed that users that categorized as Psychoticism used at least **7 higher negative words**.

RESULTS

Fig 2: Malay Words Categorized Based on Valence

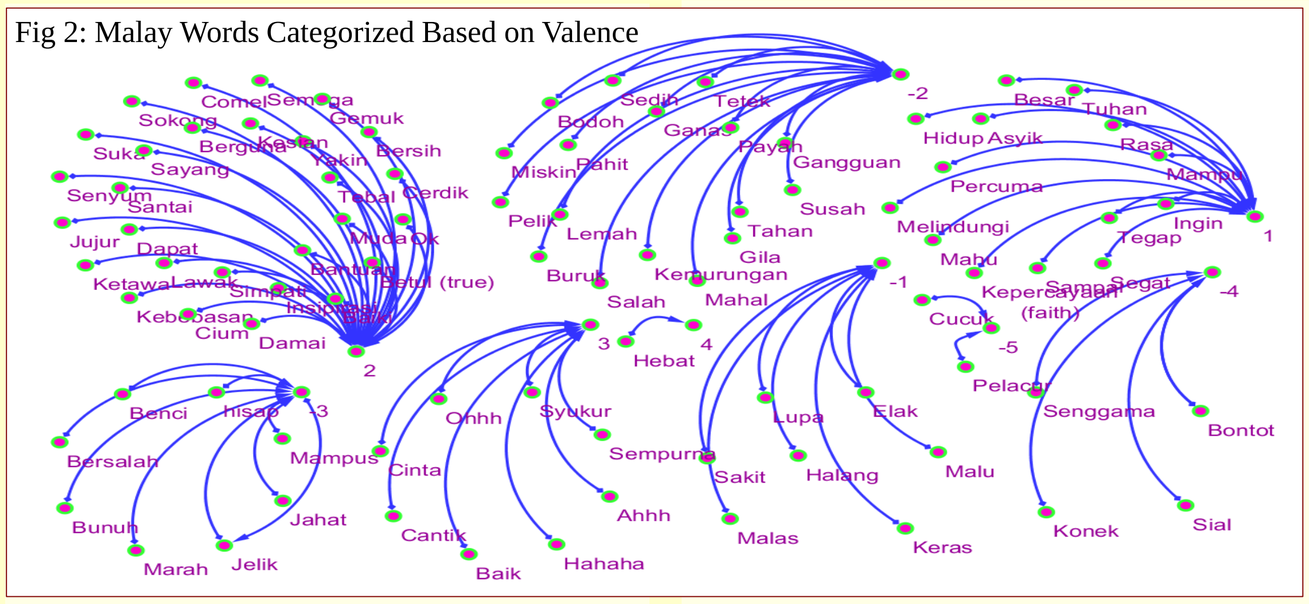


Table 3: Performance measurement based on Accuracy (%) [11].

Machine Learning Classification on Malay Tweets			
Cross Validation	**3**	**5**	**10**
Classifiers	**(%)**	**(%)**	**(%)**
ZeroR (Baseline)	**53.3**	**53.3**	**53.3**
NB	**80.0**	**90.0**	**90.0**
KNN	**63.3**	**56.7**	**56.7**
SMO	**73.3**	**70.0**	**86.3**
J48	**50.0**	**63.3**	**70.0**

Table 4: Terms that highly associated with Psychoticism trait.

Extracted Twitter Terms using x^2		
Unigram	**Bigram**	**Trigram**
Anjing	**Gila Babi**	**Nak Main *Game***
Babi	**Anak / Makan Babi**	***Fuck Off* la**
Tetek	**Tetek Kecil**	**Pakai Kondom tak**
Bontot	***Fuck* Orang**	***Shit* kalau aku**
Melancap	**Pakai Kondom**	**Tetek Size A/kecil**
Kote	**Lubang Buntut**	
Cilaka	**Nak Main / Sex**	
Lancau	**Pancut Dalam**	
Senggama		

Table 5: Additional findings

Additional Findings	
Language Models	**Unigram provide better classification results.**
Dwilingual Cursing Expression	**Malay Psychoticism sentences tended to be developed using English Cursing terms.**
Cursing Intensity Between English & Malay Terms	**Based on the perception as local citizen, we felt that Malay cursing words connoting more negative intensity or abusive (*** sounds awakward) compare to English terms. (*Refer Table 6*).**

Table 6: Example Malay Cursing terms and Its English Translation.

Cursing Word Differences		
Malay Cursing Word	**Translation in English**	**Standard/Polite Utterance in Malay**
Kote / Lancau	**Penis / Cock**	**Zakar**
Bontot	**Buttock / Ass**	**Punggung**
Tetek	**Breast**	**Payu Dara**
Kimak	**Mother's Cunt**	****Abusive**

Discussion

To the extent of our knowledge, this is a first study that measured the Malay online criminality writings through the concept of personality. Although, we initially used the sentiment valences based seed words as a main cue to identify the criminality writings, consequently, the markers assisted us to identify further characteristics of such postings. Our observation on Psychoticism sentences showed that the criminals may construct their sentences using the mix sentiment of words. For instance, the sentence "*Muka baik tapi sayang, pelacur*" that present as one of the instance in Psychoticism class revealed that the word *baik* and *sayang* semantically refferring to positive sentiment. In the other hand, our observation showed that Malay swear words looks like more cursing and taboo compare to the such words in English. Perhaps, it is due to the cursing words in English became a commonly used words or local people more prefer and practicising unofficial terms/language (Bahasa Selanga/Pasar).

CONCLUSION

Eventually, our investigation showed that NB performed better than other classifiers in all three cross validation approaches. This investigation also produced a list of the sentiment words identified through Chi Square (X^2) feature selection analysis on language models that could be used by cyber criminals. Furthermore, the sexiest remark and inflammatory words become essential cues to identify criminality related postings.

REFERENCES

1. Sapir, Edward. Language: An Introduction to the Study of Speech. New York: Harcourt, Brace, 1921.
2.Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. Deep learning based document modeling for personality detection from text. *IEEE Intelligent Systems* 32(2):74–79, 2017.
3. Saravanan Sagadevan. Thesis : Comparison Of Machine Learning Algorithms For Personality Detection In Online Social Networking, 2017.
4..Matthews, G, Ian, J. D., & Martha, C. W. Personality Traits (2nd edition). *Cambridge University Press*, 2003.
5. Firoj, A., Evgeny, S., & Giuseppe, R. Personality Traits Recognition on Social Network-Facebook. *AAAI Technical Report WS-13-01*. Computational Personality Recognition (Shared Task), 2013.
6.Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. Workshop on Computational Personality Recognition (Shared Task). *In Proceedings of WCPR13, in conjunction with ICWSM-2013*.
7. Farnadi et al. Recognising Personality Traits Using Facebook Status Updates. *AAAI Technical Report WS-13-01*. Computational Personality Recognition (Shared Task), 2013.
8. Finn Arup Nielsen. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings* : 93-98. 2011 May. http://arxiv.org/abs/1103.2903, 2011.
9. Golbeck, J., Robles, C., Edmondson, M., & Turner, K. Predicting Personality from Twitter. *In Proceeding of International Conference on Social Computing*, 2011.
10. Muhd, Baqir. Profiling Online Social Network (OSN) User Using PEN Model and Dark Triad Based on English Text Using Machine Learning Algorithm, 2017 (In Review).
11.Nurul Izzati Binti Ridzuwan. Online Social Network User-Level Personality Profiling Using Pen Model Based On Malay Text (In Review), 2017.
12.Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James W. Pennebaker. Lexical predictors of personality type. *In Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*, 2005.
13. Saxena, S., & Mani, S., C. Machine learning based approach for human trait identification from Blog data. International Journal of Computer Applications, 48(10), 17–22, 2012. doi:10.5120/7384-0005.
14. Farkhund Iqbal, H. B., Benjamin C.M. Fung, Mourad Debbabi. Mining writeprints from anonymous e-mails for forensic investigation. *Digital Investigation*, 7(1-2), 56-64, 2010.