Using Swear Words Increases the Irritability - a Study Using AI Algorithms

Costin Andrei BRATAN[1, 2], Ana Voichita TEBEANU[1], Gabriela BOBES[3],
Ionut POPESCU[4], Gabriela IORGULESCU[5, 6], Liliana NEAGU[6], Adriana
APOSTOL[1], Razvan GOIAN[7], Monica DASCALU[1, 2], Eduard FRANTI[2, 7], and
Ana Maria OPROIU[6]

[1]*Politehnica* University of Bucharest, Bucharest, Romania

[2]Research Institute for Artificial Intelligence *Mihai Drgnescu*, Bucharest, Romania

[3]Okaua Theater Company, Bucharest, Romania

[4]Pink Stil SRL, Bucharest, Romania

[5]Academic Society of Behavioral Sciences from Romania, Bucharest, Romania

[6]University of Medicine and Farmacy *Carol Davila*, Bucharest, Romania

[7]National Institute for Research and Development in Microtechnology, Bucharest, Romania

E-mails: costin.bratan@upb.ro*, anatebeanu@yahoo.com,
gabriella_iorgulescu@yahoo.com, monica.dascalu@upb.ro,
eduard.franti@imt.ro

* Corresponding author

Abstract. This paper presents the effects' analysis produced by the frequent use of swearing from the perspective of irritability. The analysis was carried out with the help of two psychological questionnaires that were completed by the volunteers before and after the inducement of the negative emotions and automatic recognition functions implemented by *Convolutional Neural Networks* (CNN), applied for the speech signals of two volunteer groups for whom negative emotions were induced. The CNN architecture uses *Mel-frequency cepstral coefficients* (MFCCs), obtained from the speech signal, and has 87,944 trainable parameters, the outputs of the network being the 8 main classes of emotion detected by the algorithm (1 neutral, 3 positive, and 4 negative). The CNN also gives information about the negative emotion and irritability level. For the volunteers who swore during the experiment, there is an increase of 14% in negative emotion intensity and of 21% for the irritability level than for the volunteers who didn't swear during the trials. The use of this current research is the understanding that cursing causes a higher level of irritability.

Key-words: Convolutional Neural Networks; Emotion Recognition; Irritability Level; Speech Recognition; Swearing.

1. Introduction

Swearing and its historical, legal, and cultural aspects have been subjects of resistance throughout history. Although extreme measures such as imprisonment and death penalties are no longer prevalent, societal resistance to swearing still exists to varying degrees across cultures. The frequent use of swearing has seemingly diminished its power over time. Swearing has been a part of human language since its emergence and can be defined as language use that refers to taboo or stigmatized subjects, expresses strong emotions or attitudes, and is not meant to be taken literally. Swearing possesses a great expressive power and serves functional purposes in specific situations. It can involve various forms and types, with religious, bodily, sexual, and disease-related taboos being commonly used across different cultures.

From the linguistic point of view, swear words differ from other verbal expressions by their vulgar, threatening, and insulting content. Some linguists have classified swear words into four categories: 1) swear words related to sex; 2) swear words related to excrement; 3) swear words related to religion; 4) swear words related to mental illness [1]. Montagu, in his book on the history of swearing, identifies fourteen categories of swear words that have been utilized throughout history, including references to supernatural powers, sacred religious matters, saints, sacred places, future life, ancestors, rulers or authorities, oaths by natural objects, vulgar or obscene words, blood-related terms, miscellaneous phrases with strong impact, classical divinities, animals, plants, products, and valued personal attributes [2]. Additionally, a fifteenth category encompasses social insults, including racial and ethnic slurs. Swearing can be categorized based on its function and degree of conscious control and can range from automatic and unconscious to consciously controlled forms. Different classifications of swearing include descriptive, idiomatic, abusive, emphatic, and cathartic forms.

The topic of swearing can be approached from various perspectives, and research has provided insights into its relationship with emotional regulation, frequency, impact on negative emotions, self-reported anxiety levels, emotional state, and historical/cultural aspects. However, further research is still needed to gain a comprehensive understanding of swearing and its multifaceted effects on individuals and society.

Many psychological studies have shown that swearing helps those who use it to relax nervously and causes intense negative emotions in those to whom they are intended [3]-[4]. When curses are spoken with certain tones of voice, they can cause the person to whom they are intended to have much more intense negative effects than if they were addressed to him in writing. If swear words are addressed verbally with certain tones and repeatedly to a child, they can cause him major psychological imbalances that will then affect him for the rest of his life [5]-[13]. These negative psychological effects can occur in children who are repeatedly cursed by adults, since children have not yet formed the emotional regulation mechanisms to help them drain the intense negative emotions induced in their psychic system by adults with the help of swearing. Swears words and expressions trigger at the same time nervous discharge effects on the one who utters them and negative psychological effects on the one to whom they are intended. In this study we will analyze the effects that swearing has on those who utter it.

The importance and interest of this study come from the fact that irritability often interferes

in the daily activities of any person, many people that face an intense feeling of irritability are predisposed at depression, anxiety, panic, anger management issues, substance abuse and other conditions. The use of this experiment is to show the evidence that swearing can increase the irritability level and so can slowly damage the quality of life.

2. State of the Art

Numerous studies have yielded mixed findings regarding the relationship between swearing and emotional regulation. Philipp and Lombardo (2017) focused on the alleviation of social distress through swearing. Their study revealed that swearing can reduce the experience of hurt feelings caused by social distress. Participants who were allowed to swear after recalling a socially distressing event reported lower levels of hurt feelings compared to those who were not given the opportunity to swear. These findings indicate that swearing may serve as a coping mechanism to regulate negative emotions and mitigate the impact of social distress [14]. Bushman, Baumeister, and Phillips (2001) explored the catharsis hypothesis, which suggests that aggressive behavior can improve mood. Their study investigated the relationship between catharsis beliefs, affect regulation opportunities, and aggressive responding. The results indicated that individuals with stronger catharsis beliefs are more likely to engage in aggressive behavior when given the opportunity to regulate their negative affect [15]. Stephens (2017) found that the frequency of swearing is related to emotional regulation. Participants who played a video game that raised their emotional arousal exhibited greater swearing fluency compared to those who played a control game [16]. Additionally, Stephens (2011) discovered that people who frequently swear in their daily lives show a lesser pain tolerance effect of swearing when compared to those who swear less frequently [17]. These findings suggest that individuals who swear more frequently may have a different relationship with emotional regulation. Robbins (2011) investigated naturalistic swearing and its impact on emotional support and depressive symptoms [18]. The study revealed that swearing in the presence of others, but not when alone, was related to decreases in reported emotional support and increases in depressive symptoms over time. Swearing's potential to produce a catharsis-effect, relieving stress or pain, and its interpersonal consequences, including emotional pain to others, were discussed by Vingerhoets [19]. Swearing's impact on people's emotional state after experiencing negative emotions has been examined by several psychologists. Popusoi discovered that swearing had a cathartic effect on drivers who experienced anger in certain traffic situations, leading to more positive affect and lower physical activation [20].

Some studies show that irritability and emotional arousal can increase the use of swear words. Lancker found that neurological disorders that increase the use of swear words share linguistic features, suggesting a common neurobiological foundation for swearing behaviors [21]. Understanding the link between emotional arousal and swearing is crucial both theoretically and practically. The concept of venting suggests that individuals may deliberately exaggerate emotional responses and swearing may serve as a form of emotional regulation. However, research on the direct relationship between swearing and emotional experience is limited. Instances of swearing can be categorized as propositional or non-propositional. Propositional swearing is consciously planned, and it overlaps with research on linguistic impoliteness and rudeness. It can be polite, impolite, or neither, depending on the speaker's intent and its impact on social harmony or face-threatening acts. Non-propositional swearing is unintentional and uncontrollable, arising from automatic emotional responses or brain damage. It is not considered polite or impolite, except when the content offends an uninformed listener. Swearing is influenced by

neurological, psychological, and sociocultural factors. These factors develop and evolve over time, relying on maturation and experience.

From the automatic and technical point of view, many experiments and research trials were made on how curse words can be detected from various types of publications, papers, journals, magazines, or internet sites, with the help of artificial intelligence algorithms and classifiers. In [22], the authors describe an automatic detection method that provides features at

different conceptual levels and detects offensive language, often represented by swear words, with a multi-level classificatory. Furthermore, in [23], the authors scraped 14,000 Twitter tweet samples and labeled them into two categories, abusive and non-abusive, then used a Multi-class F1-score, Precision, Recall, *Matthew's Correlation Coefficient* (MCC), and a Confusion matrix to evaluate the results. This recent study showed that many internet insults are made of curse words or phrases, and they indeed have an emotional impact on human social life on the internet. Also, in [24] and [25], machine learning and deep learning approaches were used to detect swearing in Arabic online social networks and Filipino language, both hitting a recognition rate of over 98%, meaning that curse words are very easy to detect and identify in cultural texts.

All the studies mentioned before used artificial intelligence only to recognize swearing and curse words. Deep learning techniques and algorithms were also used to establish the impact of swearing on human's emotions. Studies show that, over time, the level of negative emotions and the upset emotional context on online messages that contains curse words has increased significantly [26], and that swearing is corelated with the increased level of toxicity on human's behavior. *Natural Language Processing* (NLP) and *Machine Learning* (ML) methods to detect and use the features indicative of toxicity can be a required and mandatory tool to strengthen the supposition [27].

3. Description of the Experiment

A total of 28 volunteers, divided into two groups, participated in the experiment. The first group consisted of individuals who frequently used swear words, while the second group comprised individuals who did not regularly use vulgar language. The volunteers were selected from diverse backgrounds and were not acquainted with each other, minimizing communication among them and potentially affecting the authenticity of the study results. The recruitment process involved distributing flyers on university campuses in Bucharest, inviting volunteers to participate in an analysis of the effects of frequent swearing on emotional state.

From the outset, all volunteers were fully informed about the study's topic and its various stages. They were explicitly warned that the study would expose them to negative emotions through the use of swearing. Additionally, they were informed of their right to withdraw from the study at any time, should they feel adversely affected by the negative emotions. All 28 volunteers who enrolled in the study participated in all stages of the experiment.

The study proceeded as follows: Each volunteer entered a room where they anticipated an argument that would evoke negative emotions. Inside the room, two actors had prepared a skit, the specifics of which were unknown to the volunteers. As the volunteer entered, the opening of the door triggered the breaking of a vase, producing loud sounds that had a negative impact on the volunteer. Consequently, the scene implied that the volunteer was responsible for the vase's destruction and received reprimands from the two actors. During the heated exchange, the actors repeatedly directed curses at the volunteer and provoked aggressive responses. The entire scene was recorded on video and audio, which were later analyzed by psychologists to examine

the progression of the volunteer's negative emotions. Following the conclusion of the skit, the volunteer moved to another room, where their level of irritability was assessed.

Throughout the skit, volunteers from the first group defended themselves using swear words, whereas those from the second group refrained from using such language. Prior to the study, a team of psychologists conducted individual interviews with each volunteer to establish their psychological profiles and assess their baseline level of irritability. The degree of irritability for each volunteer was also evaluated at the end of the study.

4. Measurements

The audio recordings of the volunteers' reactions throughout the skit were analyzed with different types of AI algorithms (the best one with the best results being chosen in the end) in order to make the following measurements:

- measuring of the negative emotions intensity impregnated in the volunteers' voices.
- the evolution of the negative emotions impregnated in the volunteers' voices during the skit.
- measuring the irritability level of the volunteers at the beginning and at the end of the skit.

The irritability level of the volunteers was evaluated at the beginning and at the end of the study by a team of psychologists with the help of specific questionnaires (SF 12 and The State-Trait Irritability Inventory Spielberger- STAI) [28–29].

The psychologists employed the aforementioned clinical scales to verify that the volunteers involved in the study were emotionally stable individuals without any recent history of psychiatric conditions or evident clinical symptoms (especially from the area of anxiety) during the experiment.

The AI classifier used to detect the intensity of the negative emotions adopts a deep learning architecture, a convolutional neural network with fully-connected layers trained on the *Mel-frequency cepstral coefficients* (MFCCs). The input audio files are divided into segments of 2 s length with the help of a fixed window size; then the DFT is applied, extracting 40 features for each file [30].

The convolutional neural network proposed for this application has a total of 87,944 trainable parameters, with a ReLu (Rectified Linear Unit) activation function, a dropout of 20% and a fully-connected layer that reduces the output shape from 640 components to 8 components. The final 8 elements represent the classes of emotions that this algorithm can identify, more specifically: Neutral, Calm, Happy, Sad, Irritability, Fearful, Disgust, Surprised. In this experiment, the irritability emotion and the level of it is the most important aspect and for the negative emotion detection we considered the Sad, Irritability, Fearful and Disgust as the main negative emotions. The range for the irritability level is between 0 and 3, with a step of 0.5.

The CNN was trained using 1,440 speech files from the *Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS) dataset and 2,800 files from the *Toronto emotional speech set* (TESS) dataset, audio-only (.wav) with 16 bits per sample and a sample rate of 48kHz. This dataset contains files of 24 professional actors (12 female, 12 male) vocalizing two lexically matched statements in a neutral North American accent. The speech includes calm, happy, sad, angry, and fearful emotions. Each expression is produced in two levels of emotional intensity (normal, strong), plus there is a neutral expression [30].

The data files were randomly split into a training dataset consisting of 3315 MFCC vectors with 40 features and a testing dataset consisting of 1/3 of the full existing dataset. It was used the

F1 score as a solid index of determining of the CNN, this method being used just as an evaluating method for the AI classifier. The training process of the architecture was realized with the sparse categorical cross-entropy loss function. Correlating with this information, the best architecture was used for the classification step, and for optimization arguments, it was set a number of 16 batches [30].

5. Results

As we mentioned before, the pre-trained CNN gives information about the type of emotion that is present in the audio files and the negative emotion and irritability levels. Some audio files didn't include negative emotions or irritability, so those values were labeled as '0'.
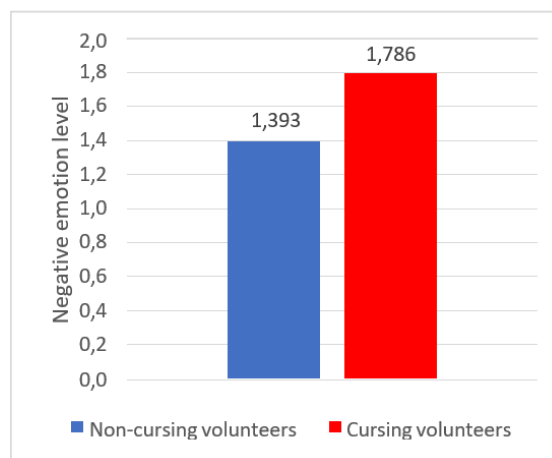

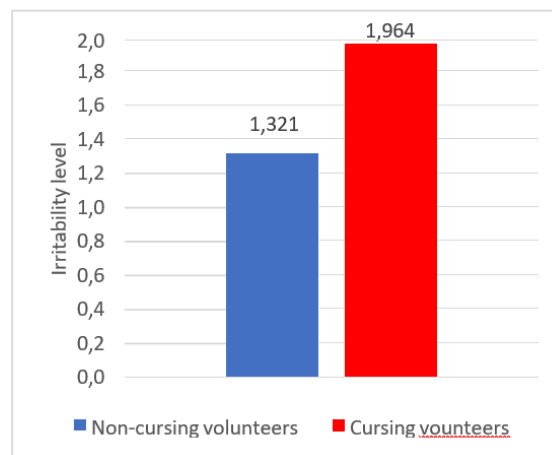**Fig. 1.** The average values of the negative emotion level.

Fig. 2. The average values of the irritability level.

The mean values of the negative emotion levels from both groups can be seen in Fig.1 and the ones for the irritability levels can be seen in Fig. 2. On both graphs the intensity of the negative emotion and irritability is higher for the volunteers who cursed during the skit than from the group that didn't curse. If we convert the results into a percentage, then the increase of the negative emotion for the cursing volunteers is up to 14% and the increase of irritability is at 21%.

Also, the histograms for the negative emotion and irritability levels for the two groups were generated. Fig.3a shows that the histogram has the highest value (4 volunteers) for the (0,5-1] interval, and Fig.3b shows that the negative emotion level for the volunteers who swore during the skit almost follows a binomial distribution, with the peak value (5 volunteers) for the (1,5-2] interval.
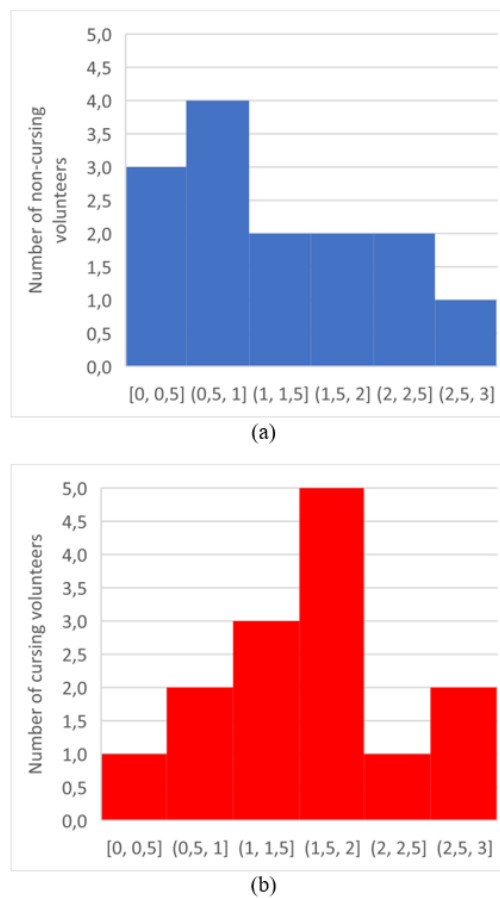

(a)

(b)

Fig. 3. The histograms for the negative emotion level (a) for the volunteers who didn't swear; and (b) for the volunteers who swore.

The histograms for the irritability level are shown on Fig.4a and 4b. The highest value (4) of the non-cursing volunteers' histogram is found at the (1-1,5] interval and we can also observe that, for the first half of the histogram, the values almost follow a uniform distribution. In Fig.4b, the interval (1,5-2,5] has more than half of the total number of cursing volunteers (57% more exactly), these values approving the major difference for the mean irritability level values seen in Figure 6.
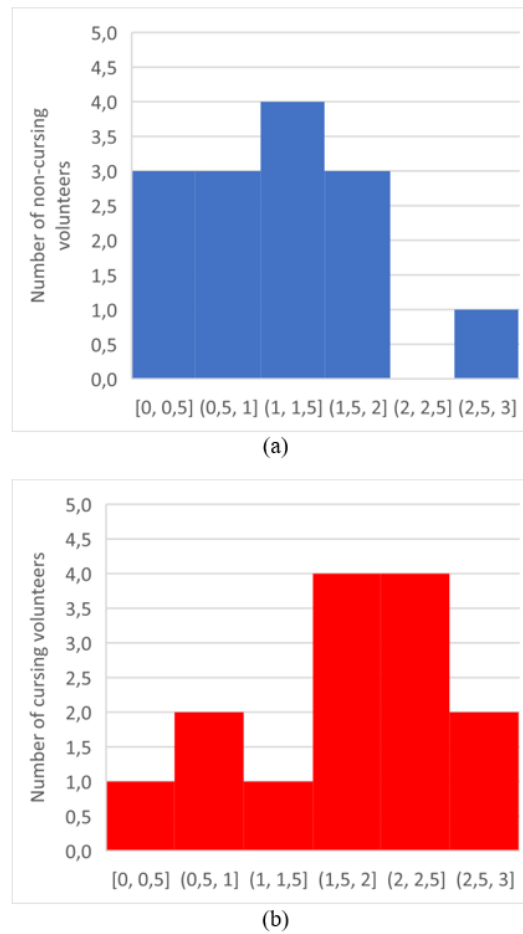
(a)

(b)

Fig. 4. The histograms for the irritability level (a) for the volunteers who didn't swear; and (b) for the volunteers who swore.

6.   Conclusions

The results obtained in this study showed that the subjects from the first group of volunteers managed to destress with the swearing's help, but their irritability level increased at the end of the study (it doubled). The subjects from the second group of volunteers managed to release their nervousness with the help of screams and some non-insulting verbal lines spoken with a high tone. The subjects' irritability level in the second group of volunteers increased compared to the beginning of the study, but much less compared to the subjects in the first group of volunteers who used swear words. Thus, swearing can have both positive and negative interpersonal consequences. While it may provide a cathartic effect in certain situations, swearing increases the level of irritability. Based on the conclusions of this study, a further investigation, using AI algorithms, could explore the long-term effects of swearing and alternative forms of emotional release on interpersonal consequences and emotional regulation. This type of research could in-

volve longitudinal assessments and follow-up sessions for volunteers from both groups to assess the sustained effects on irritability and emotional well-being.

References

[1] C. SARNIKA, *Types and Functions of Swear Words Used in American Sitcom How I Met Your Mother Season One*, Journal of Language and Literature **6**(2), 2018.

[2] A. MONTAGU, *On the physiology and psychology of swearing*, Psychiatry **5**, 1942, 189-201.

[3] R. PUDIR, *People Who Swear May Make Better Friends*, Science Says, Health, Wellness & Productivity, https://www.lifehack.org/532720/people-who-swear-may-make-better-friends-science-says.

[4] N. BURTON, *Hell Yes: The 7 Best Reasons for Swearing*, Psychology Today, May 2012.

[5] N. B. WASHMUTH, R. STEPHENS, *Frankly, we do give a damn: improving patient outcomes with swearing*, Arch Physiother, 2022.

[6] K. C. LAFRENIERE, S. G. MOORE, R. J. FISHER, *The Power of Profanity: The Meaning and Impact of Swear Words in Word of Mouth*, Journal of Marketing Research **59**(5), January 24, 2022.

[7] T. JAY, *Do offensive words harm peoPLE?*, Psychology, Public Policy and Law 15(2), 2009, 81–101.

[8] Y. MORIMOTO, A. SHARMA, *Long term outcomes of verbal aggression: The role of protective factors*, Journal of Emotional Abuse, **4**(2), 2004, 71–99.

[9] I. L. SPERTUS, R. YEHUDA, C. M. WONG, S. L. HALLIGA, *Childhood emotional abuse and neglect as predictor of psychological and physical symptoms in women presenting to primary care practice*, Child Abuse & Neglect **27**(11):1247–58, November 2003.

[10] C. L. HOGLUND, K. B. NICHOLAS, *Shame, Guilt, and Anger in College Students Exposed to Abusive Family Environments*, Journal of Family Violence **10**(2), June 1995, pp. 141–157.

[11] V. S. LAMPHEAR, *The impact of maltreatment on children's psychosocial adjustment: A review of the research*, Child Abuse & Neglect **9**(2), 1985, 251–263.

[12] [12] L. DI GIUNTA, W. A. ROTHENBERG, C. LUNETTI, J. E. LANSFORD, C. PASTORELLI, N. EISENBERG, L. M. URIBE TIRADO, Longitudinal associations between mothers' and fathers' anger/irritability expressiveness, harsh parenting, and adolescents' socioemotional functioning in nine countries, Developmental psychology **56**(3), 458, 2020.

[13] G. PICKERT, H. PALS, *Parental Anger and Trajectories of Emotional WellBeing from Adolescence to Young Adulthood*, Journal of Research on Adolescence **30**(2), 2020, pp. 440–457.

[14] M. C. PHILIPP, L. LOMBARDO, *Hurt feelings and four letters' words: Swearing alleviates the pain of social distress*, European Journal of Social Psychology **47**, 2017, pp. 517–523.

[15] B. J. BUSHMAN, R. F. BAUMEISTER, C. M. PHILLIPS, *Do people aggress to improve their mood? Catharsis beliefs, affect regulation opportunity, and aggressive responding*, Journal of personality and social psychology **81**(1), 2001, pp. 17-32.

[16] R. STEPHENS, A. ZILE, *Does Emotional Arousal Influence Swearing Fluency?*, Journal of Psycholinguistic Research **46**, 2017, 983–995.

[17] R. STEPHENS, C. UMLAND, *Swearing as a response to pain-effect of daily swearing frequency*, The journal of pain: official journal of the American Pain Society 12(12), 2001, 1274–81.

[18] M. L. ROBBINS, E. S. FOCELLA, S. KASLE, A. M. LOPEZ, K. L. WEIHS, M. R. MEHL, *Naturalistically observed swearing, emotional support, and depressive symptoms in women coping with illness*, Health psychology: official journal of the Division of Health Psychology, American Psychological Association **30**(6), 2011, 789–92.

[19] A. VINGERHOETS, L. M. BLYSMA, C. D. VLAM, *Swearing: A biopsychosocial perspective*, Psychological topics **22**, 2013, 287–304.

[20] S. A. POPUSOI, G. M. HAVAMEANU, C. E. HAVAMEANU, *Get the f#∗k out of my way! Exploring the cathartic effect of swear words in coping with driving anger*, Transportation Research Part F: Traffic Psychology and Behaviour, 2018.

[21] D. R. LANCKER, J. L. CUMMINGS, Expletives: neurolinguistic and neurobehavioral perspectives on swearing, Brain Research Reviews **31**, 1999, 83-104.

[22] A. H. RAZAVI, D. INKPEN, S. URISKY, S. MATWIN, *Offensive Language Detection Using Multi-level Classification*, Canadian AI 2010: Advances in Artificial Intelligence, 2010, pp 16–27.

[23] Galinato, Valfrid and Amores, Lawrence and Magsino, Gino Ben and Sumawang, David Rafael. Context-Based Profanity Detection and Censorship Using Bidirectional Encoder Representations from Transformers. Available at SSRN: https://ssrn.com/abstract=4341604, 2023.

[24] A. OMAR, T.M. MAHMOUD, T. A-E-HAFEEZ, *Comparative Performance of Machine Learning and Deep Learning Algorithms for Arabic Hate Speech Detection in OSNs*, AICV 2020: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), 2020, pp 247–257.

[25] B. FERRER, C. TOMILAS, L. MALLARE, B. PINEDA, A. DE GUZMAN, F. SIAPNO, K. PAYANG, M. YU, A. BALANOS, Z. LEE, *A machine learning model for the profanity detection in the filipino language*, Journal of Engineering Science and Technology Special Issue on ICITE2021, 2021, pp. 37–46.

[26] H. DU, W. X DONG, K. LOPEZ, *Detecting the Emotion Dynamics of Profanity Language In Social Media Through Time*, http://cs230.stanford.edu/projects_fall_2020/reports/55770462.pdf.

[27] A. SHETH, V. L. SHALIN, U. KURSUNCU, *Defining and detecting toxicity on social media: context and knowledge are key*, Elsevier Neurocomputing **490**, 14 June 2022, 2022, pp. 312–318.

[28] J. E. WAVE JR., M. KOSINSKI, S. D. KELLER, *A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity*, Medical care 34.3 (1996), 220-233 – The SF-12v2 PRO Health Survey (qualitymetric.com).

[29] C. D. SPIELBERGER, R. L. GORSUCH, R. LUSHERNE, P. R. VAGG, G. A. JACOBS, *The State-Trait Anxiety Inventory (STAI), form Y (40 items for state anxiety and trait anxiety)*, Manual for the State- Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press, 1983.

[30] M. G. DE PINTO, M. POLIGNANO, P. LOPS, G. SEMERANO, *Emotions Understanding Model from Spoken Language using Deep Neural Networks and Mel-Frequency Cepstral Coefficients*, 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2020.