

## HATE SPEECH IN THE COMMENTS' COLUMN INSTAGRAM: A DISCOURSE ANALYSIS

<sup>1,2\*</sup>Yasir Mubarak, <sup>1</sup>Dadang Sudana, <sup>1</sup>Wawan Gunawan

<sup>1</sup>Universitas Pendidikan Indonesia, Bandung, Indonesia

<sup>2</sup>Universitas Pamulang, Tangerang, Indonesia

\*Corresponding Author Email: [yasir.mubarak@upi.edu](mailto:yasir.mubarak@upi.edu)

Article Info	Abstract
<b>Article History</b> Received: October 2023 Revised: December 2023 Published: January 2024	<i>The phenomenon of social media has undergone a transformation, serving as a platform for facilitating online commerce, exchanging ideas, disseminating information, and unfortunately, facilitating illicit activities such as fraud, intimidation, defamation, hate speech, and other related behaviors. The anonymity provided by social media platforms renders it a potent tool for disseminating harmful or discriminatory material. The purpose of this research is to describe hate speech directed at Putri Delina (PD) in her personal Instagram comment column. The descriptive qualitative research methods are used in this study. The main theory in this study is Searle's pragmatic speech act theory, and the data source is a screenshot of netizen comments on PD's Instagram account in 2022. The results of the study show that PD received hate speech in the comments section of her Instagram account. First, the expressive illocutionary, category-criticizing, found four data. Second, assertive illocutionary type, the state category, found one data. Third, the type of illocutionary commissive, swear category, found one data. Fourth, the illocutionary type of directive, the advising category, found three data. Last, the illocutionary type of declarative, the category of prohibit, found one data.</i>
<b>Keywords</b> Hate Speech; Pragmatics; Instagram;	
<b>How to cite:</b> Mubarak, Y., Sudana, D., & Gunawan, W. (2024). Hate Speech in the Comments' Column Instagram: A Discourse Analysis, <i>JOLLT Journal of Languages and Language Teaching</i> , 12(1), pp. 439-450. DOI: <a href="https://dx.doi.org/10.33394/joltt.v12i1.9050">https://dx.doi.org/10.33394/joltt.v12i1.9050</a>	

### INTRODUCTION

Current public usage of social media is widespread, particularly among young people. According to Ningrum (2018) claims the role of social media in the lives of Indonesians is currently astounding. One method of communication on social media is done in public. Social media can also readily and swiftly disseminate information to all levels of society without having to select who and where it is disseminated to (Yulistiyono, 2021). Indonesia has 144.2 million internet users by 2021. 129.2 million of the above-mentioned internet users have active social media profiles (Mawarti, 2018). The progression of time necessitates an improvement in the realms of science and technology. The Internet also allows us to express our opinions anonymously, or at least creates the illusion of anonymity (Kocoń et al., 2021). The use of social media such as YouTube, Facebook, Twitter, Instagram, and WhatsApp is one of these advancements. Social media is used by people for a variety of purposes, like staying in touch with friends, making their interests known, getting useful information, asking for help, and relieving stress (Zhao and Rosson in Celli et al., 2021).

Nonetheless, some users exploit the proliferation of online social networks to spread extremist and discriminatory ideas, resulting in the spread of hate speech and hate crimes (Aljarah et al., 2021). Mawarti (2018) claims that social media has evolved into a tool for conducting online business, sharing ideas, spreading information, and even engaging in various fraudulent acts, intimidation, slander, inciting hatred, and similar activities. Hinduja & Patchin (as cited in Plaza-del-Arco et al., 2021) states that social media users may

occasionally experience negative psychological effects from the messages or content shared, and in some cases, these effects may even cause them to commit suicide. Its anonymity (media social) makes it an effective medium for spreading offensive or hateful content (Kovács et al., 2021). Ningrum (2018). added that the nature of media openness or information disclosure on social media is what causes people to have a greater proclivity to make hate speech. In short, social media can now be used for anything and is difficult to control. On social media, hoaxes, slander, hate speech, and blasphemy are common (Isral & Heryandy, 2022; Septanto, 2018).

Oftentimes, the misuse of information technology to disseminate hate speech is done for personal reasons, such as fostering enmity towards certain individuals or groups in the form of SARA and diminishing a person's electability for certain positions. Typically, the dissemination of hate speech through the use of information technology involves the dissemination of negative information about a person or organization (Sepima et al., 2021).

The National Police have found a lot of hate speech, hoaxes, and ethnic, religious, racial, and intergroup (SARA) speech in 2018. There have been 642 pieces of provocative content (Brigadier General Mohammad Iqbal as cited in Widayati, 2018). The proliferation of hate speech has fueled a climate in which people are more likely to use profanity and openly express their hatred of one another. Eriyanto (in Juditha, 2017) states hate speech, in which an individual or an entire group of people is portrayed in a negative light, is another form of marginalization. Hate speech, according to Subagyo (2020), serves short-term emotional objectives. People who are found guilty of using hate speech frequently have to pay fines or even go to jail (Kocoń et al., 2021).

Social media firms are crucial in this sense since some US companies, like Facebook, exacerbate both traditional and contemporary forms of harassment, abuse, and discrimination. Racist behaviors, both new and old, are becoming increasingly prevalent on social media as it has taken over the socio-political landscape in practically every country in the world (Matamoros-Fernández & Farkas, 2021). Fangen and Holter says in (Anni'mah Nurul et al., 2020) the United States, disdain for other groups is expressed in the comments section of the official news media. Literature study shows that hate is present in Twitter (Anni'mah Nurul et al., 2020; Lyrawati, 2019; Mahardhika & Zuliarso, 2018), Facebook (Permatasari & Subyantoro, 2020), Instagram (Astuti, 2019; Himawan & Zamzani, 2022; Pertiwi, 2020; Salutfiyanti, 2018; Suryani et al., 2021), WhatsApp (Suryani et al., 2022), and YouTube (Ayudya et al., 2019; Jamilah & Wahyuni, 2020; Pangaila et al., 2021).

Fortuna et al. (2021) claim some social media sites, like Facebook and Instagram, automatically delete messages that look too much like messages in their database of hate speech. This is done to stop hate speech. But removing hate speech isn't always enough from a legal standpoint. For example, law enforcement agencies can go after people who say things like are aggressive, make threats, are racist, etc. So far, only human moderation has made it possible to spot this kind of hate speech. Besides that, only 3% of people who send harmful messages are charged, owing to the pervasiveness of online communications (Vashistha & Zubiaga, 2020).

Hate speech is rude to the person who hears or recipient it, and the person who says it has certain goals in mind when they say it (Suryani et al., 2022). Hate speech is also defined as communication that is "hateful," contentious, fosters intolerance, and/or is in some other way polarizing and demeaning (Vashistha & Zubiaga, 2020). Nockleby (as cited in Rangel et al., 2021) claims hate speech is generally understood as language that disparages an individual or group because of one or more of their racial, ethnic, gender, sexual orientation, nationality, religious, or other characteristics. Hate speech's targets range from kids to the general public, public figures, and government authorities (Permatasari & Subyantoro, 2020). Hate speech also is speech that contains hatred, attacks, and outbursts that is intended to have

a certain effect, either directly (actually) or indirectly (stopping on the intention), which is to incite others to commit acts of violence or harm other individuals or groups (Widayati, 2018).

The newly created datasets on offensive content, hate speech, and cyberbullying cover a wide range of non-English languages such as Arabic (Aldjanabi et al., 2021; Aljarah et al., 2021; Alshalan et al., 2020; Awane et al., 2021), Polish (Adamczak-Krysztofowicz & Szczepaniak-Kozak, 2017; Habrat, 2021), Spanish (Arango et al., 2020; García-Díaz et al., 2022; Uzan & HaCohen-Kerner, 2021), Italian (Caiani et al., 2021; Celli et al., 2021; Florio et al., 2021), and Chinese (Chen, 2022; Uyheng et al., 2022; Wang et al., 2022). Therefore, it would be fruitful to examine social media as a forum for the propagation of hate speech. In addition, it is imperative to accurately assess hate speech and other types of online harassment in order to ensure, among other things, non-discriminatory access on digital forums.

In addition, online hate speech presents unprecedented challenges to the protection of equality rights when compared to traditional legal approaches that have been taken in the past to address discriminatory practices. First, pertinent legislation may lag behind technological advancement. Second, the characteristics of digital media containing hate speech pose substantial obstacles to direct government oversight (Chen, 2022).

In previous linguistic forensic research, data on hate speech and defamation originated from social media platforms such as Facebook, Instagram, and Twitter. In this study, hate speech is a topic of interest for researchers and the data focused on the comments section of PD's Instagram media and used Searle's theory. Speech acts are classified into five types by Searle. The five types are assertive, directive, expressive, commissive, and declaration speech acts. This study was designed to identify the hate speech in the comments' column Instagram: a discourse analysis.

## **RESEARCH METHOD**

### **Research Design**

This study adopted a qualitative descriptive research design, with a theoretical framework grounded in forensic linguistics. The research focused on the analysis of illocutionary utterances suspected of harboring hate speech within the Instagram comments section of PD's account. The primary data source was derived from speech instances in the comments column of the Instagram account @PD during the year 2022. The application of purposive sampling in this investigation involved the intentional selection of data that specifically pertained to the manifestation of hate speech phenomena. To ensure a nuanced exploration of the subject matter, a sequential exploratory technique was employed, which encompasses the systematic collection and analysis of qualitative data. This methodological approach allowed for a detailed examination of the identified illocutionary utterances within the context of hate speech, contributing to a nuanced understanding of linguistic patterns and communicative strategies associated with such expressions in the digital realm.

### **Instruments**

The primary instrument utilized in this study is the researchers themselves, who actively engaged in the data collection process through a combination of observation, note-taking, and substitution approaches. In employing the observation technique, particular attention was directed towards scrutinizing various types of postings on social media platforms. Subsequently, instances of hate speech were meticulously documented through the systematic method of note-taking. To comprehensively analyze and identify hate speech expressions within individual posts, a detailed examination was conducted. Additionally, the researcher applied the substitution strategy, strategically replacing informal language with formal terms, thereby facilitating the categorization of words or phrases into their respective types. The examination of hate speech within PD's Instagram comments column unfolded in three

distinct phases. Firstly, the study sought to determine the quantity of hate speech present in the comments column. Following this, the analysis extended to categorizing the identified hate speech into different types. Finally, conclusions were drawn based on the outcomes of the categorization process. This structured approach not only elucidates the multifaceted nature of the study's methodology but also underscores the meticulousness involved in the researchers' efforts to comprehensively understand and interpret hate speech within the specified social media context.

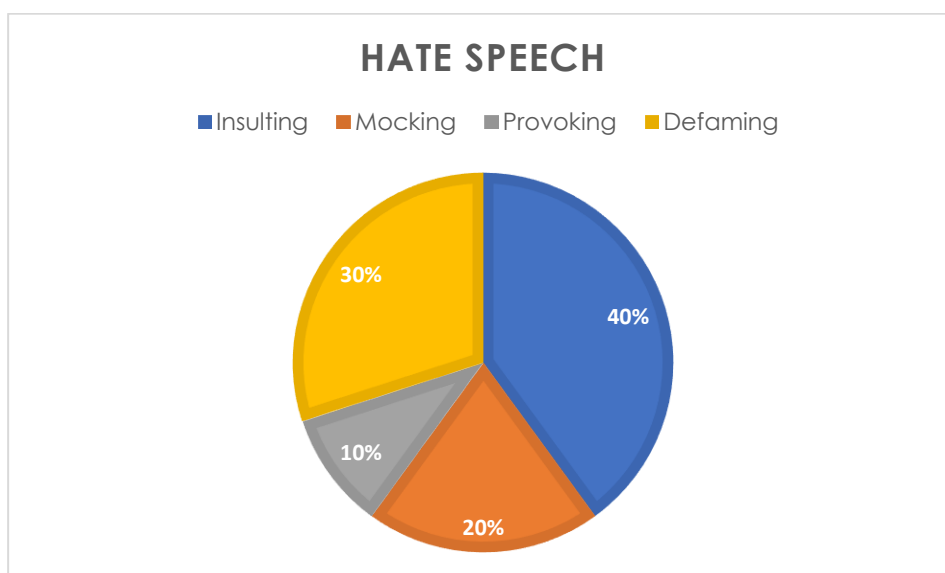
### Data Analysis

To facilitate a thorough and systematic data analysis, the initial phase of this study involved the meticulous organization of the collected data. All posts gathered during the research were methodically arranged, and the information within them was transcribed through a process of transforming note-taking into textual data. This transcription process enhanced the manageability of the data, allowing for more convenient classification based on distinct categories of hate speech. The analytical framework employed in this study utilized Microsoft Office 2019 software to quantify the prevalence of various forms of hate speech. The outcomes of this quantitative analysis were then visually represented through graphics, providing clear depictions of the proportional or percentage distribution of the identified hate speech categories. This research adopted a mixed-method approach, incorporating both qualitative and quantitative methodologies. Consequently, the study's findings will be presented through a descriptive qualitative analysis, offering nuanced insights into the nature of hate speech. The graphical data visualization will play a pivotal role in illustrating the observed rise in hate speech percentages, enhancing the accessibility and interpretability of the study's outcomes.

## RESEARCH FINDINGS AND DISCUSSION

### Research Findings

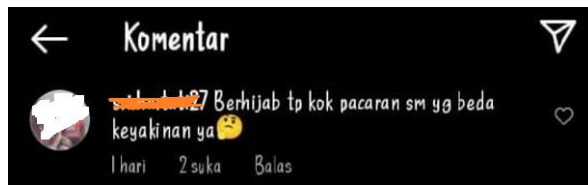
Based on the aforementioned research, the form of hate speech in the PD comment column can be determined as follows:



According to the data in the table above, 40% of hate speech takes the form of insults, followed by 30% in the form of defamation, 20% in the form of mockery, and 10% in the form of provocation. The analysis is presented in the form of social media utterances that are

thought to contain hate speech and is based on the data gathered. The following is a presentation of the results and analysis of the speech data.

### Data 1



*"Berhijab tp kok pacaran sm yg beda keyakinan ya (emotikon tangan di dagu, bingung)"*

The speech above is a form of hate speech spoken by the account @s\*\*\*. The statement has the expressive illocutionary, and category-criticizing. This speech serves as an expression of outrage at PD's actions. Furthermore, PD was shown the statement as a Muslim woman wearing a hijab because she does not reflect a pious woman who deviates from religious teachings. The utterance has legal implications because it is defamatory and is posted in public or on social media.

Furthermore, the S\*\*\* statement in the preceding speech is a performative speech. The speech arose as a result of feeling unhappy to PD. S\*\*\* utterance's (locutionary) has the intention (illocutionary) of being an insult to the PD personally.

### Data 2



*"Reseh lu muka lu 2 wajah lu sok lembut tpi ternyata kya serigala"*

The speech in data 2 is a form of hate speech spoken by @m\*\*\* because it contains insults. The utterance is also included in the illocutionary form of expressive type, an insulting form. By using words or expressions like "*mukanya dua* [two-faced], "*sok lembut*," [like gentle] and "*seperti serigala*," [like a wolf] this speech act expresses rage toward PD. The utterance "M\*\*\*" are intentionally used in a way that can be interpreted as an insult to the PD personally.

In the preceding context, the term "wolf" refers to a person who appears decent and obedient but is truly malevolent. The remarks propagated in the comments column were merely intended to demonstrate to the public that PD was not a good person who was associated with a wolf (a wicked character).

### Data 3



*“Kalo gak cantik minimal tau diri dan bersyukur (emotikon tertawa menangis, mengejek)”*

This statement by @L\*\*\* constitutes hate speech because it contains an insult to PD. The utterance is also included in the illocutionary of expressive type, insulting form. The humiliation was shown to PD's physical '*tidak cantik*' [not beautiful], and non-physical '*tau diri & bersyukur*' [self-aware & grateful].

The above comment also includes an unpleasant act that is both insulting in public and has a legal dimension. The use of a laugh emoticon, which has the meaning of mocking, clarified the insult.

#### Data 4

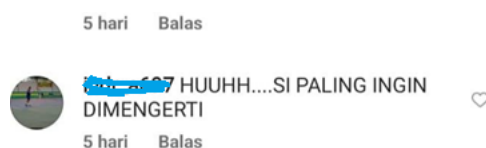


*“Sangat luar biasa orangtua terutama emakmu mendidikmu seperti ini harus disebar grup wa keluarga mu nih (emotikon tertawa lebar)”*

The utterance by @R\*\*\* constitutes hate speech because it contains provocative actions. The expression belongs to the expressive illocutionary category of blame. Through her speech on Instagram, the @R\*\*\* account placed the blame on PD's parents, specifically her mother, for failing to educate her so that she behaved in such a manner. The phrase '*sangat luar biasa*' [extremely extraordinary] in the preceding sentence has a negative connotation or vice versa.

In addition, @R\*\*\* mentions in the statement "*seperti ini harus disebar grup wa keluarga mu nih*", implying that PD's extended family must know of PD's embarrassing behavior. This is also confirmed by the use of a big laugh emoticon which means mocking and laughing at PD's actions.

#### Data 5



*“HUUHH....SI PALING INGIN DIMENGERTI”*

The above statement by @I\*\*\* is a form of hate speech. The utterance expresses anger and irritation at PD's behavior. The utterance is also included in the assertive illocutionary type, the state category. The statement claims that PD desires to be understood by her family, particularly her stepmother.

Furthermore, the utterance '*SI PALING INGIN DIMENGERTI*' implies selfishness and seeking attention through viral problems between herself and her stepmother. The @I\*\*\* account spread hate speech (mocking) in response to PD's actions, which were marked in the comments with 'HUUHH' and capital letters or Caps Lok.

**Data 6**

*“Tak doain putri dapet duda anak 7 biar tau rasanya jadi ibu sambung yang susah banget ngatur anak!”*

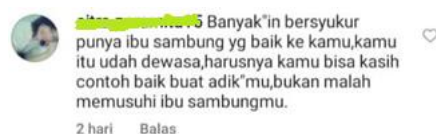
The above speech by @T\*\*\* is a form of hate speech (provoking). The utterance is a type of illocutionary commissive, swear category. @T\*\*\* swore that PD would marry a widower with seven children in response to PD's statement. The utterance contained an act of anger and irritation towards PD.

In addition, the speech was intended to make PD know how it feels to be a stepmother who has to live in a foreign place by taking care of children who are not her biological children. The utterance was motivated by intense hatred and outbreaks of rage due to irritation with PD. This is indicated by the presence of an exclamation mark, which signifies a strong emotional response.

**Data 7**

*“Kalo punya sifat gituh jangan coba” untuk menikah put intropeksi dulu yah (emotikon tersenyum)”*

The preceding statement by @I\*\*\* is an example of hate speech because it involves a warn to PD. The utterance is also classified as the illocutionary type of directive, the advising category. The utterance ‘intropeksi dulu’ & ‘kalo punya sifat gituh’ implies that no one would want to marry a woman with negative characteristics such as PD, and it is then further reinforced using emoticons. The humiliation was shown to PD's non-physical attribute ‘sifat gituh’ that means she has a bad character.

**Data 8**

*“Banyak”in bersyukur punya ibu sambung yg baik ke kamu, kamu itu udah dewasa, harusnya kamu bisa kasih contoh baik buat adik”mu, bukan malah memusuhi ibu sambungmu”*

The utterance above is a form of hate speech by @C\*\*\*. This utterance is included in the illocutionary type of directive, the category of advising. The utterance was suspected of being hate speech because it contained the act of giving a warning to PD. In addition, @C\*\*\* utterance's (locutionary) has the intention (illocutionary) of being a mocking to the PD



personally by mentioning '*kamu itu udah dewasa, harusnya kamu bisa kasih contoh baik buat adik"mu, bukan malah memusuhi ibu sambungmu*' [you're an adult, you should be able to set a good example for your younger siblings]. It implies that PD is immature and instead allegedly incited her younger siblings to hate her stepmother. The statement was thought to be hate speech because it was posted on the Internet and hurt PD in both a material and an emotional way. Also, it can make people think bad things about her. This utterance can be interpreted to defame in public.

#### Data 9



*"Aneh orang tua suruh minta maaf, pikirin tu bpkmu sm adek2mu kshan suruh ngikutin kemauanmu yg gak jls itu"*

The utterance above is a form of speech spoken by @R\*\*\*. The utterance is included in the illocutionary type of directive, the category of advising. Furthermore, @R\*\*\* utterance's (locutionary) has the intention (illocutionary) of being a mocking to the PD personally by mentioning '*Aneh*' & '*kemauanmu yg gak jls itu*'. The utterance was suspected of constituting hate speech. It damages PD both materially and immaterially because it uploaded to the Internet. Besides, it can incite others to think negatively about her. This utterance can be interpreted to defame in public.

#### Data 10



*"Mba tolong ya next day kalo ada konten gausah merasa paling tersakiti, yang punya masalah bukan Cuma mba doang, masih banyak loh orang diluar sana yang berjuang lebih dr km, jangan caper plis"*

The preceding statement by @N\*\*\* is an example of hate speech. This utterance is included in the illocutionary type of declarative, the category of prohibit. Furthermore, @N\*\*\*(locutionary) \*'s statement bears the goal (illocutionary) of mocking the PD personally by referencing utterances '*gausah merasa paling tersakiti*', '*jangan caper plis*'. The remark was suspected of being hate speech. It causes material and immaterial harm to PD because it was uploaded to the Internet. Furthermore, it may incite others to have bad opinions about her and will eventually attack PD on her Instagram. This utterance can be interpreted to defame in public.

#### Discussion

The findings are dominated by hate speech in the form of insults. This indicates that hate speech is caused by disharmony from the interlocutor's (netizens') point of view due to dislike for the speaker's (PD) behavior, and the utterance tends to be open or expressive. According to the findings of Paz et al. (2020), they claim that in the case of hate speech, some people seek to discredit those they consider to be enemies.



Furthermore, Soesilo (as cited in Permatasari & Subyantoro, 2020) says that to insult someone is to attack their honor and reputation. Most people who are attacked feel ashamed. Insults, such as physical and non-physical attacks, are meant to hurt someone's reputation and honor such as in data 1, 2, 3, and 7.

Defamation is the second most common type of hate speech. This type of hate speech has indirect or indirect characteristics. The results show that the utterance hurts PD in both a material and an immaterial way. Also, the utterance is posted online, so it can harm PD's reputation as an artist and may also have a negative impact on the public's perception of her. In line with Smola (Widayati, 2018), the damage to a person's reputation comes from what people think when they hear or read a statement of contempt. According to the definition of defamation in the Criminal Code (KUHP) (in Permatasari & Subyantoro, 2020), it is the act of defaming someone's name or honor through oral or written statements and it is reflected in data 9, 8, and 10.

The third finding is mockery-type hate speech. The characteristics of this hate speech are usually deliberate and have implicit meanings, such as the phrase '*sangat luar biasa*' [extremely extraordinary] having a negative connotation or vice versa. Another example, the utterance "*seperti ini harus disebar grup wa keluarga mu nih*", implies that PD's extended family must know of PD's embarrassing behavior. In line with Ningrum, et. al. (2018), this hate speech is to mock the interlocutor indirectly. Therefore, this speech is an implied reprimand or a compliment with a negative connotation or the opposite of what is expressed. It is reflected in data 4, and 5.

Finally, hate speech is a form of provocation. This utterance aims to enrage the interlocutor (PD), diminish PD's image, and increase support from other netizens, as shown in data 6. The speaker prays for PD to marry a widower with seven children. That has the potential to not only bring down PD's image, but also indirectly for other people to join and support her statement. In line with Suryani's (2022) findings that the person who says it has certain goals in mind when she or he says it. According to Bachari (2010), statements like this are very likely to result in legal repercussions if the speaker perceives the speech as an act that hurts the other person's feelings, especially PD who became the target. Criminal threats are also aimed at anyone who intentionally and without rights disseminates information aimed at instilling hatred or hostility toward specific individuals and/or community groups in violation of SARA (Article 28 paragraph (1) in conjunction with Article 45 of the ITE Law). According to Hinduja & Patchin (as cited in Plaza-del-Arco et al., 2021), users of social media may occasionally suffer adverse psychological effects due to the messages or content shared, and in extreme cases, this may lead to suicide. This ability, along with the fact that people can remain anonymous, makes it a good way to spread hateful or offensive content (Kovács et al., 2021). It has been shown that the presence of hate speech on social media platforms correlates with actual hate crimes (García-Díaz et al., 2022). Wang et al. (2022) claim, some social media users tend to say mean things about people who don't agree with them and use hate speech when they don't like something. They also claim that users may issue hate speech based on personal characteristics as well as characteristics of an ethnic group or country, in addition to using rude language. Because of anonymity, people are more likely to speak freely on an online platform, which is reflected in the data.

The findings of this study may have ramifications for the Indonesian government in terms of developing legislation or rules to detect and prevent the spread of hate speech on social media. Furthermore, it is imperative for the government to expeditiously develop an educational program aimed at mitigating the proliferation of hate speech in subsequent generations. Indonesian individuals ought to exhibit heightened awareness regarding the prospective perils associated with disseminating hate speech on various social media platforms (Elfrida & Pasaribu, 2023).

In accordance with the present results, previous studies (Af'al, 2022; Halid, 2022; Oktiawan, 2021; Ramadani, 2021; Suryani et al., 2021) have demonstrated that the phenomenon of hate speech directed towards artists or public figures in the column comments that is disseminated through social media platforms possesses the capacity to contravene the ITE Law and the Criminal Code, thereby potentially incurring legal consequences.

## CONCLUSION

Based on research data, the characteristics of hate speech on social media are labeled with the use of insults, defamation, mockery, and provocation. According to the data in the table above, 40% of hate speech takes the form of insults, followed by 30% in the form of defamation, 20% in the form of mockery, and 10% in the form of provocation. The findings are dominated by hate speech in the form of insults. The most common form of the hate speech is explicit. One of the triggering factors for hate speech, such as ignoring ethical or legal social norms in society. Another factor is the dislike for someone, particularly PD. There are various limitations inherent in this study. In order to get accurate and valid findings, it is imperative to augment the corpus data in this study. Furthermore, this study utilizes manual techniques for both data collection and processing. To enhance the efficiency of data analysis and improve the accuracy and comprehensiveness of data interpretation, it is advisable to utilize software tools like AntCont for future research purposes.

## REFERENCES

- Adamczak-Krysztofowicz, S., & Szczepaniak-Kozak, A. (2017). A disturbing view of intercultural communication: Findings of a study into hate speech in Polish. *Lisguistica Silesiana*, 38.
- Af'al, W. (2022). Ujaran Kebencian Terhadap Aktor Arya Saloka di Media Sosial Twitter: Kajian Linguistik Forensik. *Jurnal Sinestesia*, 12(2), 435–444.
- Aldjanabi, W., Dahou, A., Al-qaness, M. A. A., Elaziz, M. A., Helmi, A. M., & Damaševičius, R. (2021). Arabic Offensive and Hate Speech Detection Using a Cross-Corpora Multi-Task Learning Model. *Informatics*, 8(4), 69.
- Aljarah, I., Habib, M., Hijazi, N., Faris, H., Qaddoura, R., Hammo, B., Abushariah, M., & Alfawareh, M. (2021). Intelligent detection of hate speech in Arabic social network: A machine learning approach. *Journal of Information Science*, 47(4), 483–501.
- Alshalan, R., Al-Khalifa, H., Alsaeed, D., Al-Baity, H., & Alshalan, S. (2020). Detection of hate speech in covid-19–related tweets in the arab region: Deep learning and topic modeling approach. *Journal of Medical Internet Research*, 22(12), e22609.
- Anni'mah Nurul, F., Nurhadi, N., & Pranawa, S. (2020). Konflik dan Ujaran Kebencian di Twitter (Studi Tentang Hashtag# 2019TetapJokowi and# 2019GantiPresiden Periode Januari-Februari 2019). *JUPIIS: Jurnal Pendidikan Ilmu-Ilmu Sosial*, 12(1), 132–142.
- Arango, A., Pérez, J., & Poblete, B. (2020). Hate speech detection is not as easy as you may think: A closer look at model validation (extended version). *Information Systems*, 101584.
- Astuti, F. (2019). *Perilaku Hate Speech pada Remaja di Media Sosial Instagram*. Universitas Muhammadiyah Surakarta.
- Awane, W., Lahmar, E. H. Ben, & El Falaki, A. (2021). Hate Speech in the Arab Electronic Press and Social Networks. *Rev. d'Intelligence Artif.*, 35(6), 457–465.
- Ayudya, T. D., Aritonang, B. M., & Krisnawati, E. (2019). Analisis Wacana Hate Speech dalam Live Streaming Youtube Ligagame E-Sports Tv. *Jurnal Komunikasi Dan Bisnis*, 7(2).
- Bachari, A. D. (2010). Analisis Pragmatik terhadap Tindak Tutur yang Berdampak Hukum. *Jurnal KIMLI*.

- Caiani, M., Carlotti, B., & Padoan, E. (2021). Online hate speech and the radical right in times of pandemic: The Italian and English cases. *Javnost-The Public*, 28(2), 202–218.
- Celli, F., Lai, M., Duzha, A., Bosco, C., & Patti, V. (2021). Polycorpus XL: An Italian Corpus for the Detection of Hate Speech Against Politics. *CLiC-It*.
- Chen, G. (2022). How equalitarian regulation of online hate speech turns authoritarian: a Chinese perspective. *Journal of Media Law*, 1–21.
- Elfrida, R., & Pasaribu, A. N. (2023). Hate speech on social media: A case study of blasphemy in Indonesian context. *English Review: Journal of English Education*, 11(2), 433–440.
- Florio, K., Basile, V., & Patti, V. (2021). Hate speech and topic shift in the covid-19 public discourse on social media in Italy. *8th Italian Conference on Computational Linguistics, CLiC-It 2021*, 3033, 1–7.
- Fortuna, P., Soler-Company, J., & Wanner, L. (2021). How well do hate speech, toxicity, abusive and offensive language classification models generalize across datasets? *Information Processing & Management*, 58(3), 102524.
- García-Díaz, J. A., Jiménez-Zafra, S. M., García-Cumbreras, M. A., & Valencia-García, R. (2022). Evaluating feature combination strategies for hate-speech detection in spanish using linguistic features and transformers. *Complex & Intelligent Systems*, 1–22.
- Habrat, D. (2021). Protection of Human Dignity as a Basis for Penalization of Hate Speech against People with Disabilities in Polish Criminal Law. *Studia Iuridica Lublinensia*, 30(4), 259–279.
- Halid, R. (2022). Tindak tutur pelaku pecemaran nama baik di media sosial kajian linguistik forensik. *Kredo: Jurnal Ilmiah Bahasa Dan Sastra*, 5(2), 441–458.
- Himawan, R., & Zamzani, Z. (2022). Analisis Bahasa Pelaku Ujaran Kebencian Berpotensi Hukum Terhadap Lesty Kejora Pada Laman Instagram @Lambe\_Turah: Kajian Linguistik Forensik. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(2), 23–31.
- Isral, I., & Heryandy, S. (2022). Iklan Layanan Masyarakat UU ITE Khususnya Ujaran Kebencian Berbasis Motion Graphic. *Academic Journal of Computer Science Research*, 4(1).
- Jamilah, F., & Wahyuni, P. (2020). Ujaran kebencian dalam kolom komentar YouTube pada tahun politik pemilihan presiden 2019. *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, Dan Asing*, 3(2), 325–341.
- Juditha, C. (2017). Hatespeech in Online Media: Jakarta on Election 2017. *Jurnal Penelitian Komunikasi Dan Opini Publik*, 21(2), 137–151.
- Kocoń, J., Figas, A., Gruza, M., Puchalska, D., Kajdanowicz, T., & Kazienko, P. (2021). Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. *Information Processing & Management*, 58(5), 102643.
- Kovács, G., Alonso, P., & Saini, R. (2021). Challenges of hate speech detection in social media. *SN Computer Science*, 2(2), 1–15.
- Lyrwati, D. P. N. (2019). Deteksi Ujaran Kebencian Pada Twitter Menjelang Pilpres 2019 dengan Machine Learning. *Mathunesa: Jurnal Ilmiah Matematika*, 7(3), 206–211.
- Mahardhika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier.
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, 22(2), 205–224.
- Mawarti, S. (2018). Fenomena Hate Speech Dampak Ujaran Kebencian. *TOLERANSI: Media Ilmiah Komunikasi Umat Beragama*, 10(1), 83–95.
- Ningrum, D. J. et. al. (2018). Kajian Ujaran Kebencian di Media Sosial. *Jurnal Ilmiah Korpus*, 2(3), 241–252.
- Oktiawan, C. (2021). Analisis Yuridis Tindak Pidana Ujaran Kebencian Dalam Media Sosial.

- Al-Adl: Jurnal Hukum*, 13(1), 168–188.
- Pangaila, G. E. V, Tangkudung, J. P. M., & Tulung, L. E. (2021). Sikap Siswa SMA Negeri 9 Binsus Manado Tentang Pesan Ujaran Kebencian di Media Sosial Youtube. *ACTA DIURNA KOMUNIKASI*, 3(4).
- Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *Sage Open*, 10(4), 1–12.
- Permatasari, D. I., & Subyantoro, S. (2020). Ujaran Kebencian Facebook Tahun 2017-2019. *Jurnal Sastra Indonesia*, 9(1), 62-70.
- Pertiwi, A. W. E. (2020). *Analisis Ujaran Kebencian dalam Akun Instagram Lambe Turah*. Universitas Muhammadiyah Malang.
- Plaza-del-Arco, F. M., Molina-González, M. D., Urena-López, L. A., & Martín-Valdivia, M. T. (2021). Comparing pre-trained language models for Spanish hate speech detection. *Expert Systems with Applications*, 166, 114120.
- Ramadani, F. (2021). Ujaran kebencian netizen Indonesia dalam kolom komentar Instagram selebgram Indonesia: Sebuah kajian linguistik forensik. *Aksara*, 22(1), 1–19.
- Rangel, F., De la Peña Sarracén, G. L., Chulvi, Bert., Fersini, E., & Rosso, P. (2021). Profiling Hate Speech Spreaders on Twitter Task at PAN 2021. *CLEF (Working Notes)*, 1772–1789.
- Salutfiyanti, D. A. (2018). *Analisis Ujaran Kebencian dalam Komentar Warganet pada Akun Instagram Obrolan Politik*. University of Muhammadiyah Malang.
- Sepima, A., Siregar, G. T. P., & Siregar, S. A. (2021). Penegakan Hukum Ujaran Kebencian Di Republik Indonesia. *Jurnal Retentum*, 2(1), 108–116.
- Septanto, H. (2018). Pengaruh hoax dan ujaran kebencian sebuah cyber crime dengan teknologi sederhana di kehidupan sosial masyarakat. *Jurnal Sains Dan Teknologi*, 5(2), 157–162.
- Subagyo, A. (2020). Implementasi Pancasila Dalam Menangkal Intoleransi, Radikalisme Dan Terorisme. *Jurnal Rontal Keilmuan Pancasila Dan Kewarganegaraan*, 6(1), 10–24.
- Suryani, Y., Istianingrum, R., & Fatin, I. (2022). Tindak Tutur Kebencian dalam Status Whatsapp. *SUAR BETANG*, 17(1), 113–122.
- Suryani, Y., Istianingrum, R., & Hanik, S. U. (2021). Linguistik Forensik Ujaran Kebencian terhadap Artis Aurel Hermansyah di Media Sosial Instagram. *BELAJAR BAHASA: Jurnal Ilmiah Program Studi Pendidikan Bahasa Dan Sastra Indonesia*, 6(1), 107–118.
- Uyheng, J., Bellutta, D., & Carley, K. M. (2022). Bots Amplify and Redirect Hate Speech in Online Discourse About Racism During the COVID-19 Pandemic. *Social Media+ Society*, 8(3), 20563051221104748.
- Uzan, M., & HaCohen-Kerner, Y. (2021). Detecting Hate Speech Spreaders on Twitter using LSTM and BERT in English and Spanish. *CLEF (Working Notes)*, 2178–2185.
- Vashistha, N., & Zubiaga, A. (2020). Online multilingual hate speech detection: experimenting with Hindi and English social media. *Information*, 12(1), 5.
- Wang, C.-C., Day, M.-Y., & Wu, C.-L. (2022). Political Hate Speech Detection and Lexicon Building: A Study in Taiwan. *IEEE Access*, 10, 44337–44346.
- Widayati, L. S. (2018). Ujaran Kebencian: Batasan Pengertian dan Larangannya. *Info Singkat: Kajian Singkat Terhadap Isu Aktual Dan Strategis*.
- Yulistiyono, A. (2021). Media Sosial. In *Etika Komunikasi Dalam Media Sosial: Saring Sebelum Sharing* (1st ed.). Penerbit Insania.