Master Thesis

Master of Science (M.sc)

Department of Tech and Software

Major: Data Sciences (DS)

Topic:

Analyzing Public Perception of South Korea's Low Birth Rate Policies using NLP-based Sentiment Analysis

Author:

Sungeun Lee

Matrikel-Number:

First supervisor:

Second supervisor:

Submitted on:

Bibliographic Description and Presentation

Master thesis 2022, 81 pages, 20 figures, 4 tables, 101 references.

University of Europe for applied sciences, Department of Tech and Software

.

Content Overview of this paper:

•

Statutory Declaration:

I hereby declare that I have developed and written the enclosed Bachelor Thesis completely by myself and have not used sources or means without declaration in the text. I clearly marked and separately listed all the literature and all the other sources which I employed when producing this academic work, either literally or in content. I am aware that the violation of this regulation will lead to failure of the thesis.

Berlin, ...

Place, Date Signature

Abstract

This paper investigates the public's perception of South Korea's low birth rate policy using natural language processing (NLP)-based sentiment analysis of YouTube comments. South

Korea's low birth rate problem presents a significant socio-economic challenge for the country's future, necessitating an understanding of public sentiment for effective government intervention. The study utilizes NLP techniques and sentiment analysis to explore the potential of YouTube comments as a valuable source of public opinion.

The Theoretical Background section offers an extensive literature review, examining South Korea's low birth rate, its socioeconomic implications, and the government policies implemented across sectors to address the issue. It also explores the application of NLP and sentiment analysis in policy analysis and opinion research, shedding light on the relationship between NLP and sentiment analysis in capturing public sentiment.

The research question guiding this study is: "What is the public's perception of South Korea's low birth rate policy based on NLP-based sentiment analysis of YouTube comments?" To address this question, the study employs a robust methodology. The YouTube API is utilized to collect comments from videos related to the low birth rate policy, followed by NLP preprocessing steps to organize and prepare the data. Subsequently, a logistic regression model is used to perform sentiment analysis and assign sentiment scores to the comments.

The Results section presents the research findings, showcasing outcomes from data preprocessing, sentiment analysis, and the performance analysis of the logistic regression model. Detailed analyses are provided, including sentiment comparisons across policy areas, potential correlations between public sentiment and policy implementation or effectiveness, and implications of the findings for policymakers.

The Discussion section reflects on study limitations and potential sources of bias, providing insights into their impact on result interpretation. It also discusses broader implications for policymaking and opinion research. The Future Research section suggests improvements to the methodology and proposes new directions for NLP-based policy analysis.

In conclusion, this study contributes to a deeper understanding of the public's perception of South Korea's low birth rate policy through NLP-based sentiment analysis. The findings reveal a diverse range of opinions and sentiment expressed by the public, reflecting the complexity of the low birth rate issue. The study underscores the importance of considering public sentiment in policy formulation and implementation and highlights the need for open dialogue and public engagement. Policymakers can utilize these insights to develop targeted strategies that address concerns and aspirations of the population, fostering sustainable solutions to South Korea's low birth rate problem. However, it is crucial to acknowledge the limitations of the study, such as reliance on YouTube comments and potential biases. Future research should explore additional data sources and employ a more comprehensive approach to capture public sentiment accurately.

Table of Contents:

Glossary

List of tables:

List of Figures:

Chapter .1 Introduction

1.1 South Korea's Demographic Challenge and the Role of Public Perception

1.1.1 Explanation of Korea's low birth rate problem, implications for Korea's future

The human step was a journey of constant development and growth. The world is striving for faster growth and development, and as a result, we live in an industrial age where there is an unprecedented level of technology. As the times progressed, more people had fewer or no children for various reasons, such as a job-oriented society, rising living costs, and a society centered on personal growth. Moreover, this problem has given us a common problem and challenge of low birth rates.

Currently, Korea faces a significant challenge of low birth rates. Korea's birth rate has declined over the past few years, reaching surprisingly low levels. Korea's birth rate dropped sharply in the 2000s to 0.92 in 2019, and the net population fell for the first time in 2020. As of 2022, the total fertility rate was 0.78, the most serious among OECD countries. Recorded birth rate. [2] Currently, Korea's low birth rate is the subject of global research.

These demographic trends have severe implications for the country's future. With population decline and aging, the country may face an economic slowdown, reduced productivity, and increased burdens on health and social welfare systems, and many regions are already beginning to face these challenges. In Korea, the application rate for pediatrics and adolescents in 2021 has already fallen by half compared to 2019. As a result, the Department of Pediatrics in Korea is starting to collapse, and there are cases where children die while going to and from the emergency room because there are no hospitals that can accommodate pediatric patients in emergencies. The effects of the severe low birth rate problem are already being revealed throughout society.

Korea's low birth rate depends on several factors. Fast urbanization, changing social values, education, and job emphasis have delayed marriage and childbirth. Economic uncertainty and the high cost of living also contribute to couples' reluctance to have children. In addition, gender inequality and lack of good work-life balance policies make it difficult for individuals to balance family responsibilities and professional aspirations.

As the low birth rate continues, Korea faces significant challenges in maintaining economic competitiveness and ensuring social welfare. Population decline means a decline in the workforce, which could hamper economic growth and pressure public finances. This means that more and more companies are suffering from labor problems due to the lack of people to work and that Korean society is already facing a big crisis. The aging population is burdening the health and social security system as the demand for elderly care and pension support increases, and the poverty rate for the elderly is also very high because the system is not yet adequately prepared.

Chapter .2 Literature Review

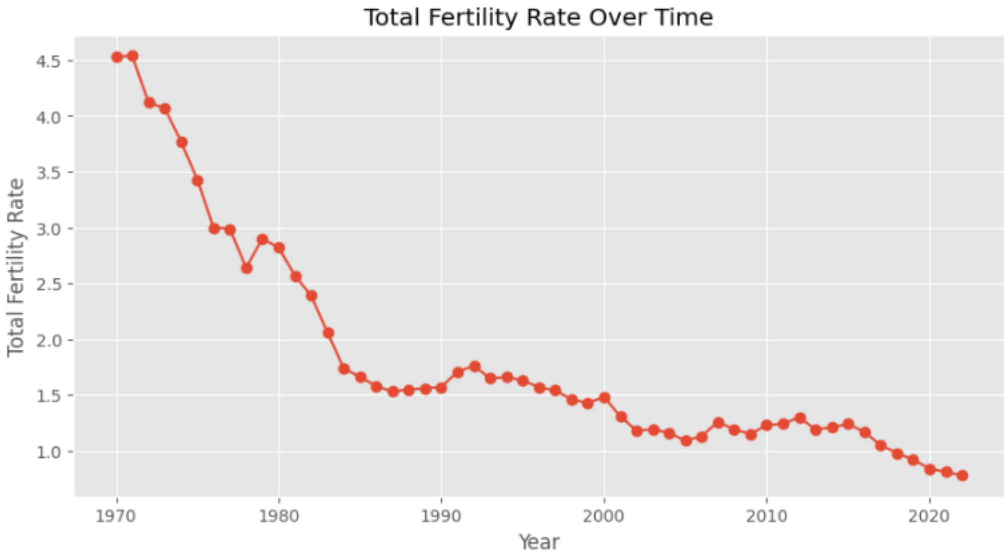
2.1 South Korea's Low Birth Rate

2.1.1. Introduction to South Korea's low birth rate problem

Population change, emphasized as Korea's low birth rate and aging population, is becoming a phenomenon that is difficult to find worldwide. The low birth rate, with a combined fertility rate of less than 2.1, has been going on for more than 40 years (So-young and Shin-hwi, 2022) and is the biggest threat to the Korean economy and growth.

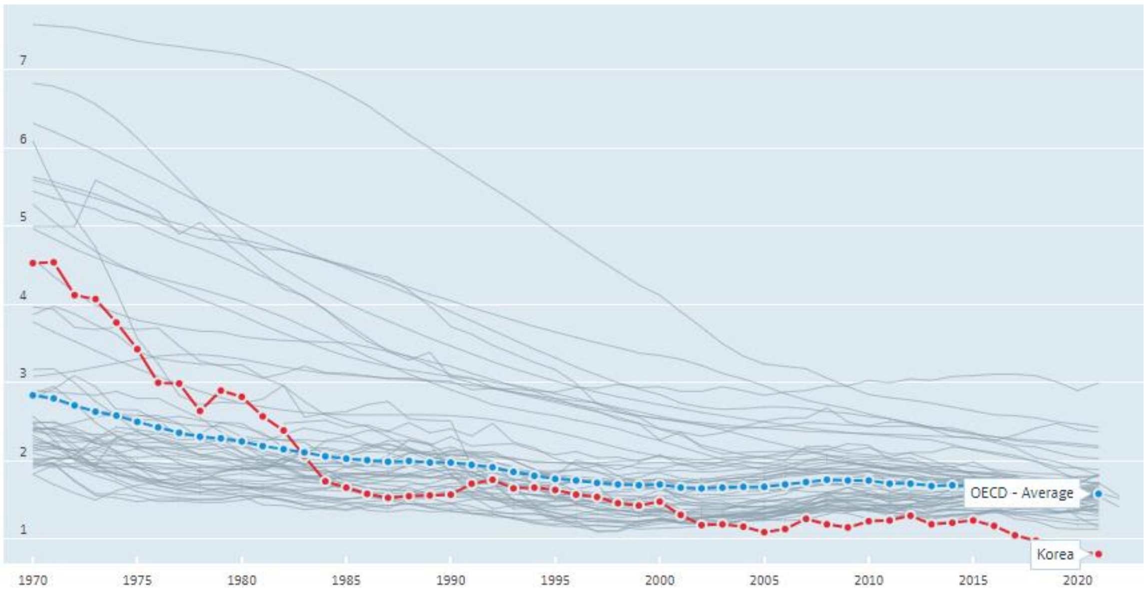
Since 2001, Korea has been unable to get out of the total fertility rate below 1.3. However, Korea's total fertility rate was not so low from the beginning. Based on the total fertility rate, the domestic fertility rate recorded a high figure of 6.0 in the early 1960s. However, the birth rate began to decline with the implementation of the government-led family planning project in the early 1960s. The fertility rate, which had been continuously declining since then, declined further through the IMF financial crisis, approaching 1.1 from 2002 to 2004.(Sam-Sik Lee, 2005)

As the low birth rate worsened, the government enacted the Basic Act on Low Fertility and Aging Society in 2005 and established the Presidential Commission on Low Fertility and Aging Society. Since 2006, the central government has responded earnestly to the low birth rate. Then, starting with the first basic plan for low birth rate and population aging, the second and third basic plans were promoted in 2011 and 2016 (Roh and Yang, 2019). Despite these efforts, the fertility rate recorded zero range in 2019 and 0.78 in 2022.



< Fig 1 : Total Fertility Rate Over Time >

Figure 1 shows trends in birth rates by year provided by the National Statistical Office of Korea. As the graph shows, the birth rate declined sharply from the 1970s to the 1980s and by a smaller margin from the 1990s, but steadily declined without rebounding until it reached 0.78 in 2022. As a result, South Korea's Moon Jae-in government acknowledged its failure to respond to the low birth rate and announced the 'third low birth rate and aging society roadmap (Kim, 2019). Attention is focused on whether the government's efforts can solve the problem.

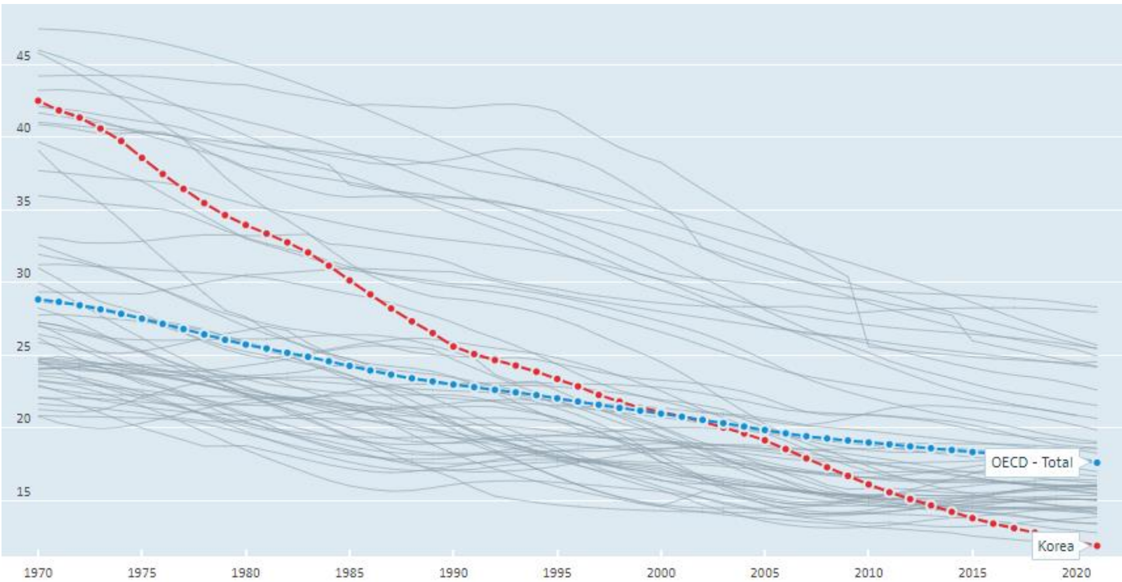


Source : OECD Data

< Fig 2 : Fertility Rates >

Figure 2 sheds light on the seriousness of South Korea's fertility rate crisis when benchmarked against the OECD average. As the graph illustrates, South Korea once had a fertility rate significantly higher than the OECD average, reaching a peak in 1970. However, by 1983, the fertility rate had dropped to 2.1, aligning with the OECD average of the same year.

Since then, South Korea's fertility rate has remained below the OECD average and has been on a steady downward trajectory. The situation has reached a critical point since 2015 when South Korea's fertility rate plummeted to become the lowest among all OECD countries and, strikingly, the entire world. By 2021, the fertility rate in South Korea had reached a new low of 0.81, raising serious concerns about the country's demographic and economic future.

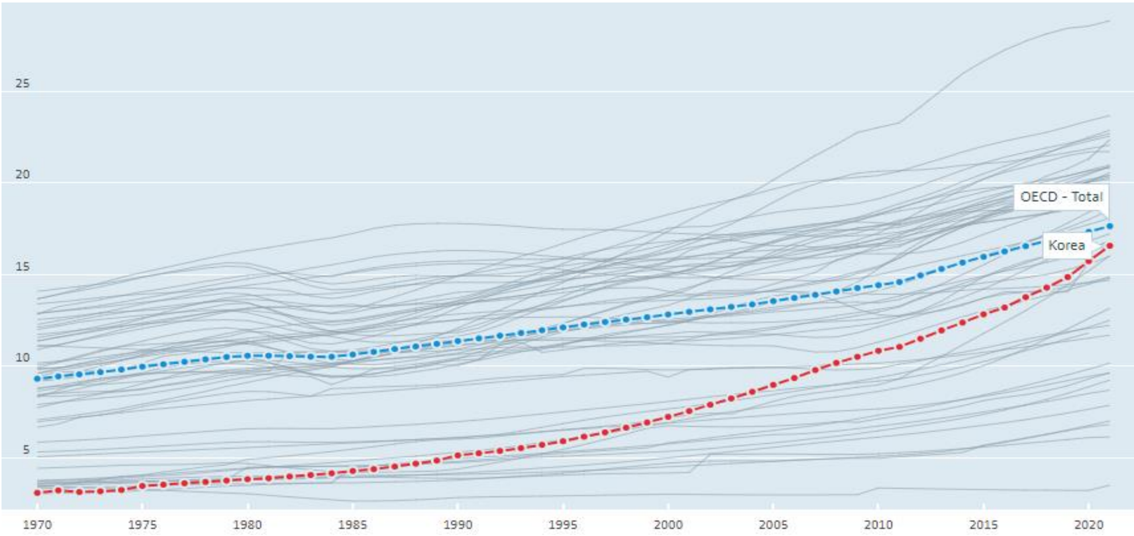


Source : OECD Data

< Fig 3 : Young Population >

Figure 3 illustrates a troubling trend regarding the ratio of children and adolescents under 15 in South Korea. There has been a substantial decrease in this age group's ratio within the population in starting 2000. By 2021, the percentage of children and adolescents had shrunk to 11.9% of the total population.

This figure is not just lower than the OECD average, which stands at a more robust 17.6%, but it is alarmingly the lowest ratio of children and adolescents in the world for 2021. This sharp decrease in the younger generation poses substantial challenges to the country's future workforce and socio-economic stability. It is expected to impact significantly the future workforce and socio-economic status.



Source : OECD Data

< Fig 4 : Elderly population >

In contrast to the shrinking youth population, figure 4 shows a significant rise in the proportion of the elderly population in South Korea, defined as individuals aged 65 or older. While South Korea's proportion of the elderly population remains lower than the OECD average, the upward trend has been rapid and consistent since 2000. This increase coincides with the sharp decrease in the proportion of the child population.

The growth in the elderly population brings a unique set of socio-economic issues, including increased healthcare costs, pension requirements, and a higher dependency ratio. If the current trend continues unabated, South Korea's proportion of the elderly population is projected to surpass the OECD average soon, intensifying the country's socio-economic challenges.

The impact of persistent low birth rates threatens Korea's future and poses significant challenges to the country's sustainability and growth potential. Recognizing governments' ability to respond to these demographic shifts and assessing public sentiment on these policies is critical to formulating a sustainable strategy for fertility recovery.

Therefore, this paper is expected to help improve and readjust future strategies to alleviate this by analyzing public sentiment about Korea's policies related to low birth rate measures. It is also expected to lead to the successful implementation of the government's policy to reverse the declining birth rate. This approach will contribute to understanding and developing effective measures to address these national challenges.

2.1.2. Socio-economic impact

statistical overview and historical trends above, the decline in the birth rate causes apparent changes in the population structure. Korea's fertility rate has been steadily declining since the 1970s, but especially since 2000, it has shown a steady and rapid decline. Comparing the deceleration of the fertility rate after 2000 and the trends of the young and elderly populations, it can be seen that as the fertility rate has declined since 2000, the young population has been rapidly decreasing, and the elderly population has been rapidly increasing. This is unmistakable evidence that the declining fertility rate is affecting the structure of the population..

South Korea's persistently low fertility rate has significantly impacted the country's demographic structure. Since 2000, this proportion has declined rapidly, leading to a significant decline in the young population and an increase in the elderly population. These demographic changes, shaped by falling fertility rates, significantly impact the country's economic and social structure (Kyung-Soo Choi et al., 2003).

Economic impacts are reflected in the labor market. A low fertility rate hinders the growth of the productive labor force, leading to a reduction in the size of the working-age population. Also, as the labor force itself is aging, the supply of the labor force is decreasing.(Seong, 2009) These changes profoundly impact the country's economic capacity and growth prospects.

As explained by the life cycle hypothesis (Ando and Modigliani, 1963), changes in population structure also affect capital markets. According to this hypothesis, consumers maintain a constant level of consumption throughout their lives, and their income level determines the level of savings. An increase in low-income seniors shifts the balance of saving and spending patterns relative to a decline in the high-income young population. These changes lower the overall savings rate and affect financial market dynamics.(Seong, 2009)

In the long run, the decrease in labor supply and the decrease in capital accumulation (due to the reduction in the saving rate) affect the factors of production. Economic growth potential is only possible with concurrent productivity growth to offset these changes.(Seong, 2009a) The social ramifications of these changes could be far-reaching, leading to increased unemployment and socio-political instability.

Finally, these demographic shifts negatively impact national fiscal health. As the aging population increases, expenditures on pensions, medical expenses, and education expenses for retraining the retired population are also increasing. At the same time, a decrease in the number of young and middle-aged contributors to the tax and social security pools could reduce national revenues. The double blow of rising spending and falling incomes can negatively impact a country's financial stability.(Seong, 2009a)

2.1.3. Government Policies

The government established the Aging Society Committee in September 2005 to address the issue of low birth rates and has been making efforts to respond effectively and realistically. Since 2005, a comprehensive plan that integrates the measures of each ministry has been established and implemented every five years. The 4th Basic Plan for Low Fertility and Aging (2021-2025) is in progress.(Youngmi, 2023)

Looking at the impact of policies since the establishment of the Aging Society Committee in 2005 until now, the low birth rate strategy first created a legal and organizational structure to strengthen social responsibility for childbirth and childcare. The government gradually expanded the policy's scope and intensity and implemented universal free education in 2013. Since then, efforts have been made to ease the financial burden of raising children through the implementation of various cash benefits systems, such as child allowance (2018), first meeting rights (2022), and infant allowance (2022). In addition, while improving parental leave provisions and introducing a supplemented parental leave system, the government supports the balance between work and family. Nevertheless, the investment of about 280 trillion won over 15 years; these attempts failed to reverse the ultra-low birth rate trend (Youngmi, 2023), leading to the lowest fertility rate in the OECD of 0.78 in 2022.

Along with the low birth rate policy, the government is implementing an aging policy, a legal and institutional device to establish a foundation for financial stability and elderly care for older adults (Youngmi, 2023). Various elderly support and welfare policies, such as job expansion for the elderly and long-term care insurance, have been implemented and tend to focus more on welfare for older adults than systems that support 2030. Despite these efforts, some strategies have been effective. However, given the speed and trend of aging, more is needed to meet the older generation's needs and accommodate and realize the necessary parts.

Looking at the government's policies for low birth rate and aging population, there has been a strong tendency to focus on older adults and support welfare policies rather than systems that support and satisfy the values and needs of those in their 20s and 30s, which is the key to solving the low birth rate. However, the problem of low birth rate is a life-cycle problem, and issues in various fields such as politics, economy, education, and welfare are intertwined. Therefore, it is essential to identify and understand the public opinion on whether the current government response is reasonable and adequate.

2.1.4. Assessment of Policy Effectiveness.

The government launched the Aging Society Committee in September 2005 to attempt a comprehensive and practical solution to the low birth rate and old-age welfare issues. Since then, the government has promoted a multifaceted strategy by integrating proposals from each ministry every five years. The current system, the 4th Basic Plan for Low Fertility and Population Aging (2021-2025), is being implemented.(Youngmi, 2023)

The first low birth rate policy in 2005 aimed to establish a legal and institutional basis to strengthen social responsibility for childrearing. As a result, the government has expanded the scope of this policy to provide free education since 2013. To offset the economic burden of childcare, we introduced an infant and toddler allowance in 2022. At the same time, the government has improved parental leave rules and unveiled a more comprehensive scheme to promote work-life balance. However, despite these efforts and an investment of about 280 trillion won over 15 years(Youngmi, 2023), the ultra-low birth rate trend did not reverse, recording the lowest fertility rate (0.78 in 2022) among OECD countries.

Along with the low birth rate policy, the government introduced an aging policy to prepare a legal and institutional basis for income guarantee and care for the elderly. Several measures were proposed, such as job expansion for the elderly and long-term care insurance. However, responding to the needs of the younger generation was neglected, and welfare for the elderly was a major concern. While some of these initiatives have been effective, they pose challenges in addressing the multifaceted needs of an aging population in a rapidly aging society.(Youngmi, 2023)

A 2022 study assessing the government's policy effectiveness presented three key findings. First, programs targeting middle-aged people need to be more budget-friendly and efficient than other demographics. The middle-aged, which is called the 'double dependent generation' in Korea, has the dual responsibility of caring for their children and caring for their parents (Oh and Tak, 2022). Lack of adequate support for the middle-aged weakens the propensity for childbirth and marriage of individuals in their 20s and 30s, which plays a decisive role in the birth rate. Those in their 20s and 30s who are looking at their parents' generation, who are burdened with excessive double support, are reluctant to marry and give birth while experiencing the future indirectly through their parents' generation. Second, policy interest in children and youth must still increase. Policies for children and adolescents often overlap with or are included in the policies of the parents' generation. This is a testament to the lack of interest in the children's generation. As interest in and support for the children's generation are sluggish, the responsibility for childrearing is being implicitly passed on to the parents' generation (Oh and Tak, 2022). This eventually reduces interest and motivation for childbirth in the 20s and 30s. Finally, in the case of the cash policy, the budget allocation was small, but the achievement rate was high (Oh and Tak, 2022). This suggests that the government prefers cash policies that can be easily presented to the public without practical and systematic consideration. However, in order to alleviate the problem of low birth rate, a service-oriented approach such as job creation and advanced child care system is urgently needed rather than cash policy.

In conclusion, the government's low birth rate and aging policy seem biased toward welfare for older adults, which does not correctly reflect the values and needs of the younger generation, which is essential in correcting the low birth rate problem. However, low fertility goes beyond a single demographic and interacts with various fields such as politics, economy, education, and welfare. Therefore, it is essential to identify and understand the public sentiment toward the government's response. A pivotal step towards fertility recovery is the ability of governments to understand and respond to the needs of their citizens.

2.1.5. The Role of Public Opinion

Modern researchers have extended the concept of public opinion as a collective communication process that forms opinions on issues prevalent in society (Jara-Amézaga, 2023). Public opinion, now understood as a multifaceted outcome of social communication, has considerable potential to shape agenda-setting processes and policy development (Yang and Zhang, 2023). It can also determine the ultimate success or failure of implemented policies, as seen in various sectors such as healthcare and service delivery (Bust et al., 2023).

The importance of public opinion in agenda-setting is increasingly recognized in contemporary research. Public sentiment and perspectives significantly impact policymakers, politicians, and other decision-makers as they can significantly shape public perceptions of support and competence (Dibattista, 2022). This influence is particularly pronounced in democratic societies where governance is derived from the people's will (Große, 2022). As a result, the importance of an issue in public consciousness can change the setting of the policy agenda (Liu et al., 2022). This is corroborated by agenda-setting theory, which posits that public opinion can shape policymakers' importance to particular issues (Große, 2022).

In addition, public opinion and sentiment play a decisive role in the execution and success of policies. Policies that empathize with the public view are often successfully implemented, with less resistance and more public support (Anderson et al., 2022). On the contrary, policies that run counter to popular sentiment can ultimately threaten success in the face of real obstacles, from peaceful protests to widespread non-compliance. Anderson reinforces this concept, explaining the dynamics between popular sentiment and policy implementation success in his book "Public Policymaking: Introduction."(Anderson et al., 2022)

Public opinion has traditionally been formed through direct dialogue and discussion, but the emergence of high-tech and the Internet has changed the landscape of collective communication. Today, digital platforms, including social media networks like YouTube and Facebook, are at the forefront of public expression and analysis. Individuals exchange ideas and opinions on these platforms, potentially affecting or creating public opinion on various issues.

Given the theoretical importance of public sentiment in policy implementation and success, leveraging the power of digital platforms for sentiment analysis can provide valuable insights to policymakers. For example, in this paper, we try to analyze public sentiment on the low birth rate policy through YouTube video comments. Policymakers can leverage the digital footprint of millions of users to measure public opinion in real time to address issues and formulate policies that align with public sentiment. This increases the chances of policy success and strengthens the democratic process by encouraging public participation in policy debates.

In conclusion, public sentiment is decisive in making policy decisions and forming results. This affects the success or failure of agenda settings, implementation processes, and policies. Leveraging the power of the Internet, especially social media platforms, for public sentiment analysis in the modern digital age can provide valuable insights to policymakers. However, more research is needed to understand the relationship between popular sentiment and policymaking in the digital age.

2.2 Natural Language Processing (NLP)

2.2.1 Introduction to NLP

The onset of the Fourth Industrial Revolution has catalyzed the rapid development of Internet technologies, spurring the digital transformation of societies worldwide. This digital age has witnessed the rise and refinement of myriad data collection and analysis technologies that provide profound insights into human behavior. Natural language processing (NLP) is at the forefront of these innovative technologies.

NLP, a key subfield of artificial intelligence, emphasizes the interaction between humans and computers using natural language. The overarching goal of NLP is to understand, interpret, and use human language meaningfully. Although its roots can be traced back to the mid-20th century, recent developments, such as the development of BERT for morphologically rich languages such as Turkish, have shown the potential of NLP for various applications (Özçift et al., 2021).

This complex technology converges computer science, artificial intelligence, and computational linguistics. It bridges human communication and digital data processing, allowing computers to understand, interpret and produce human language.

In recent years, NLP has transitioned from primarily rule-based systems by incorporating advanced machine learning and deep learning techniques. Recent developments in NLP have contributed to successfully implementing machine translation, language models, speech recognition, and automatic text generation applications. (Vedantam, 2021)

Today, NLP is indispensable for processing and analyzing the vast amounts of text data born of the digital age, spanning web content, social media discourse, and academic publications. It helps researchers and data scientists run compute-intensive tasks, manage massive text data sets, and easily extract valuable insight (Yang and Huang, 2023).

NLP offers a variety of functions, including syntactic analysis, sentiment analysis, keyword extraction, translation, and linguistic text generation. This allows machines to understand and respond to text or voice commands and subtle nuances such as emotion and sarcasm. The importance of NLP spans various applications, from enhancing chatbot and voice assistant interactions to sentiment analysis in social media and optimizing search engine performance.

Recent advances such as the development of the Generative Pretrained Transformer (GPT) model and its conversational variant, ChatGPT, have demonstrated the potential of NLP in a variety of applications, including language translation, question answering, and text summarization (Thakkar and Jagdishbhai, 2023). The application of NLP in domains such as smart cities is also recognized, playing an essential role in innovative healthcare, smart business, intelligent community, savvy media, intelligent research, and innovative education (Tyagi and Bhushan, 2023).

In particular, NLP at the beginning of the study was made based on English, so there was a limitation that it was difficult to use in Korean data, but from 1998 to 2007, through the "21st Century Sejong Plan," which the government aimed at building Korean digital language resources at the national level, the Sejong corpus, which is currently the largest among Korean language resources, began to be made (Man-hwi et al., 2015) and active Korean NLP research has been conducted. This study will be another study using Korean NLP and an excellent example of applying NLP to the Korean public sector.

2.2.2 Importance of NLP in Korean

NLP has become an integral part of data analytics and machine learning, and has a significant impact on interactions with digital systems. Since the research first began in the 1950s, its importance has grown exponentially. The research aimed to teach computers to understand, analyze and replicate human language patterns. However, as research has expanded, it has been clearly demonstrated that, while techniques developed for English are effective in a linguistic context, they are handicapped when applied to languages with different structural characteristics, such as Korean.

One of the critical issues is related to the preprocessing techniques used in NLP research. Preprocessing in NLP often removes non-terms that do not contain essential meanings and are usually removed from the text. Rajaraman argued that no stopwords would be familiar to all languages (Rajaraman and Ullman, 2011). This works well for English with a relatively simple morpheme structure but needs fixing with the complex structure of Korean as it happens to be a severe problem with Korean. Korean is a deadlock language much more comprehensive than inflectional languages such as German. Its words are often formed by combining morphemes with their respective meanings. Hangeul also consists of more than 10,000 syllables, each of which can express a concept or sentiment (Sangah et al., 2020). This linguistic complexity and stark contrast with English has resulted in inherent limitations in applying English-focused NLP techniques to Korean.

Recognizing these problems, the focus has shifted to developing NLP technologies explicitly designed for the Korean language, aiming to explain Korean-specific characteristics and provide a more vital analysis of Korean text. The pivotal measure for this was the "21st Century Sejong Plan," from 1998 to 2007. This national project aimed to build digital language resources that fit our language. As a result, the Sejong Corpus, currently the largest Korean language resource (Man-hwi et al., 2015a), was created. The development of this corpus marked an important milestone in studying Korean NLP by providing a powerful tool for solving the linguistic complexity of Korean. Since then, further progress has been made in generating specific vocabulary and tokenizers designed to efficiently handle complex word formats unique to the Korean language. These tools have significantly improved the capabilities of NLP techniques for processing Korean text, increasing the accuracy and depth of the analysis.

The rise of Korean NLP research not only represents the advancement of technology in processing specific languages but also represents the broader drive in the field of NLP to embrace the rich diversity of human languages. We highlight the importance of better developing language-specific methodologies within NLP to capture each language's complexity and nuance, allowing more accurate and contextual analysis of language data. Continuous research in this area is critical for NLPs, their applications, their broad understanding of human languages, and their ability to create more sophisticated, subtle, and effective digital tools. The development of Korean NLP research and devices continues to face challenges, but it is an ongoing process that opens up new opportunities and potential for the future.

2.2.3 NLP Techniques – Tokenization, Stemming and Lemmatization

Tokenization is an essential procedure in NLP and is necessary for decomposing raw text into separate analytically available units called 'tokens'(Grefenstette, 1999). This is essential for converting simple text strings into manageable data chunks so that computer systems can understand human language.

With its complex structure, the Korean language presents a unique challenge to tokenization. The unique structure of the Korean language complicates the tokenization process. It has led to in-depth exploration by prominent researchers such as Seo Hyun-Jae, Jeon Tae-hee, and Kim Byeong-chang.

Their extensive research has contributed to the development of numerous tokenization methodologies tailored to the Korean language. For example, when tokenizing M words into morphemes, the problem of excessive decomposition of words was supplemented by simplifying the issue of part-time tagging using only two types of tokens: word morphemes and grammatical morphemes (Seo Hyun-jae, 2022). Jeon also continued experimenting with tokenization with word embeddings and tried to derive better performance (Jeon, 2022) , providing positive insights into the Korean language.

Despite these developments, Korean tokenization continues to present complex challenges. The complex Korean language continues to show more significant challenges. These problems provide a rich pool of language data for NLP tasks while presenting tremendous obstacles to researchers. Despite these obstacles and challenges, researchers are constantly striving for the future of Korean NLPs. The collective efforts of researchers are shaping the future of NLP, paving the way for language understanding and human-machine interaction.

Therefore, research on Korean tokenization has the potential to contribute significantly to the development of NLPs as an evolving and dynamic field, and further research and exploration are required.

2.2.4 NLP Techniques – Stemming and Lemmatization

Stemming and title extraction are essential in natural language processing (NLP) to reduce words to their base or root form. For example, 'take,' 'takes,' and 'took' are all shortened to 'take.' Stemming and heading extraction aims to increase computational efficiency by reducing the complexity of language models and the amount of data that needs to be processed (Cho et al., 2023). However, different technologies approach this goal differently, and the choice between them may differ depending on the specific requirements of the task at hand.

Stemming, the more radical of the two, usually involves removing prefixes and suffixes from words to obtain 'stems.' While this effectively reduces word inflection, it can also lead to errors because the stem of a word may not always be a valid root in a language (Shin, 2022). For example, the popular stemming algorithm Porter Stemmer reduces 'arguing' and 'argues' to 'argu', which is not a valid English word.

On the other hand, heading extraction is a more sophisticated process of determining a word's primary or dictionary form based on its intended meaning in context. The algorithm uses a comprehensive dictionary that can be consulted to link the form of a word to its auxiliary form. For example, 'better' and 'good' are both lemmatized as 'good', while 'better' remains 'better' in stemming. However, lemmatization requires a more complex analysis of words and context, which is more computationally intensive (Cho et al., 2021).

Applying morpheme analysis and title extraction in the Korean language processing context presents unique challenges. Introducing a process applicable to Korean, an agglutinative language, and access to appropriate procedures will be the key to successful Korean NLP. Recently, studies have been conducted to improve the application of morpheme analysis and heading extraction techniques in Korean NLP. For example, the work of Jinwoo et al. has shown promising results that will improve the accuracy and efficiency of these techniques. They analyzed morphemes and tagged parts of speech using BERT (Shin and Jung, 2021).

In conclusion, stemming and subject extraction are essential components of NLP and contribute significantly to the accuracy and efficiency of language models. However, the choice between these technologies is highly dependent on the specific requirements of the task and the nature of the language being processed. As this field continues to evolve, further research is needed to optimize these techniques for different languages and contexts.

2.2.5 NLP Techniques – NER and Sentiment Analysis

Named Entity Recognition (NER) and Sentiment Analysis are fundamental NLP techniques. It is essential in understanding, interpreting, and organizing text data and contributes significantly to various NLP applications (Lee and Kim, 2023).

NER identifies and classifies critical elements of text, which forms an integral part of extracting structured information from unstructured data sources (Asudani et al., 2023). NER is particularly useful in various applications, such as query-answering systems, machine translation, and information retrieval, improving the system's understanding of complex semantic relationships between entities (Han et al., 2022).

On the other hand, Sentiment analysis is concerned with identifying and understanding the sentiment behind text data. It determines the tone of emotion expressed in words, making it possible to understand attitudes, opinions, and sentiments expressed in the text (Alonso del Barrio and Gatica-Perez, 2023). This technique is widely used in marketing and customer service to understand customer attitudes and market trends.

2.2.5 Applications in Public Policy Analysis

NLP, a technology that digitizes and utilizes human language, has significant potential in public policy analysis and public opinion research because of its intrinsic focus on language interpretation. This potential is evident in two key areas.

First, NLP enables large-scale text data analysis, including public comments, social media posts, and news articles. This analysis makes it possible to detect new trends and gauge public sentiment about policies. This fine-grained understanding can help policymakers identify and classify public interests and concerns, facilitating more effective and responsive policymaking (Liddy, 2001).

For example, Ha Sang-hyun's research shows the potential of NLP in public opinion analysis. His research is related to real-time analysis of Twitter data collected during Korea's 20th National Assembly election campaign. This study confirmed that the sentiment analysis method can be applied to the poll by distinguishing positive and negative sentiment and predicting the direction thereof. (Sang Hyun and Tae Hyup, 2020)

Second, NLP provides a mechanism for real-time policy monitoring to track changes in public sentiment as policies are implemented and modified. This feature allows policymakers to make data-driven decisions and proactively adjust policies to change public opinion.

In this context, Heo Se-young's research presents a remarkable case. His research includes time series analysis through NLP according to the implementation of the extramarital childbirth policy. This analysis demonstrated the usefulness of NLP in measuring policy effectiveness by detecting temporal changes in public opinion (Seyoung et al., 2022).

NLP modeling is valuable for understanding public opinion and evaluating public policy effects. The potential in various fields requires more adoption and further development and highlights the need for more research and research to leverage the capabilities for public policy analysis fully.

2.3 Sentiment Analysis

2.3.1 Introduction to Sentiment Analysis

Sentiment analysis or opinion mining is a computational process that uses NLP, text analysis, and computational linguistics to identify, extract, and quantify subjective information from various sources (Purba and Yadi, 2023). These sources range from social media platforms and blogs to forums and news articles. The basic premise of sentiment analysis is to analyze a given text and classify it as positive, negative, or neutral depending on the emotion expressed. (Cebral-Loureda et al., 2023)

Sentiment analysis can be used in various fields. In marketing, for example, real-time sentiment analysis monitors consumer reactions and opinions about products and services, helping companies quickly resolve customer complaints, improve product offerings, and make strategic decisions based on customer feedback (Kyaw et al., 2023). In finance, sentiment analysis is used to predict stock market trends by analyzing news articles and social media posts, providing valuable information for traders and investors looking to make data-driven decisions. (Barone and Barone, 2022)

Machine learning approaches, including deep learning and transfer learning techniques, have been widely used in sentiment analysis, especially for analyzing customer reviews in the hotel and restaurant industries and predicting stock market trends in the financial sector (Monika, 2022). Transformer-based models and text augmentation have also been studied to reduce computational costs and improve performance in sentiment analysis tasks (Gong et al., 2022).

In the healthcare sector, sentiment analysis was applied to explore public sentiment and opinions about COVID-19 vaccination, providing insight into motivations and barriers to vaccination (Al-Garaady and Mahyoob, 2022). Additionally, in a recent study by Mensah, Sun, and Aletras (2021), the authors explored how positional embeddings can be used to improve target-oriented opinion word extraction. They experimented with a variety of text encoders, including BiLSTM-based models, and utilized parts of speech and positional embeddings to capture the position of words relative to their target. Their results showed that BiLSTM-based models can effectively encode location information into word representations, contributing to the continued development of sentiment analysis techniques (Mensah et al., 2021).

In an information tsunami era flooded with vast amounts of data, sentiment analysis provides an efficient and effective means of extracting valuable insights from this vast amount of unstructured data. This approach will be a valuable step toward growth and development in each field through a comprehensive and diverse approach regardless of politics, economy, culture, and society.

2.3.2 Tools and Techniques

Sentiment analysis utilizes various tools and techniques, each with its strengths and limitations. At the most basic level, a vocabulary-based method uses a list of predefined words with associated sentiment tags, called a sentiment vocabulary. Although simple, this approach can suffer from complex linguistic structures such as sarcasm and contextual emotion.

Machine learning techniques represent advanced approaches including naive Bayes, logistic regression, random forests, support vector machines, stochastic gradient descent, and extreme gradient boosting. These techniques include training models on labeled data, allowing the system to perform supervised learning to learn the complex language patterns associated with different emotions. However, these methods can require large amounts of labeled data and are resource-intensive. A study conducted on COVID-19 Twitter data in April 2021 utilized other machine learning models and BERT, which achieved an accuracy of 84.2% (Darad and Krishnan, 2023).

Recently, a combination of a transformer-based models such as BERT and BERT, CNN, and BiLSTM has been used for sentiment analysis, showing high accuracy. Combining these techniques, a model called BCBL-Att has been proposed and shown to have more advantages in text sentiment classification tasks (Yang, 2022). The Korean model supports various models such as KoBERT and KcBERT, and improves the accuracy of emotion detection by understanding the context of Korean words and sentences.

Deep learning and machine learning-based sentiment analysis is applied to Bitcoin price prediction concerning specific applications, demonstrating the success of a deep learning architecture using the Globe word embedding approach (SARIKAYA and Aslan, 2022). In the context of the COVID-19 pandemic, a BERT-based transfer learning model was used for sentiment analysis on Turkish Instagram comments, achieving high classification success. (Karayiğit et al., 2022)

2.3.3 Sentiment Analysis Challenges and Future

Sentiment analysis has been recognized as an essential tool for understanding public opinion, but it has difficulties and limitations. One of the major concerns is the correct interpretation of complex and multidimensional language. Satire, irony, and cultural idioms pose significant challenges to sentiment analysis algorithms, as these linguistic features can invert or significantly alter the sentiment of a text (Ng et al., 2023). Also, the context in which a statement is made can have a significant impact on the sentiment of the statement, adding to its complexity.

In multilingual environments, sentiment analysis faces additional challenges. Translating emotions between languages is sometimes complicated given that the emotions conveyed by a given word can vary from culture to culture (Shanmugavadivel et al., 2022a). Words that convey positive emotions in one language may be neutral or even negative in another. Sentiment analysis tools must therefore be resource-intensive tools.

Sentiment analysis also deals with subjective information, which is usually highly variable and inconsistent. Many factors can influence feelings, including personal circumstances, current mood, and broad social or cultural tendencies. This makes it difficult to fully capture the depth and breadth of emotion through automated analysis. Biases inherent in data sources can distort sentiment analysis results, as viewpoints expressed on social media or other data sources may not accurately represent the sentiment of a larger population (Zhou et al., 2022)

Despite these obstacles, the future of sentiment analysis in public policy looks promising. As models and algorithms continue to improve, the accuracy and applicability of sentiment analysis will undoubtedly expand. One of the future directions is to integrate sentiment analysis with various data sources such as demographic or geographic data. For example, people will be able to understand the nuances of public opinion, such as how public opinion differs by demographic group or region (Che et al., 2023). The future of such sentiment analysis will contribute to a deeper and broader understanding of the people's thoughts and public opinion, and contemplation and discovery of necessary policies for the people.

In addition, as the utilization of deep learning models increases, the boundaries of sentiment analysis will be further strengthened. These models are suitable for understanding the complexity of human language as they can capture complex patterns and dependencies in the data (Shanmugavadivel et al., 2022a). Deep learning algorithms, particularly those based on transformer architectures such as BERT and RoBERTa, have shown promise in accurately capturing contextual emotions, handling negativity, and identifying more subtle emotions. The use of different types of deep learning models will be diverse and widespread, such as incorporating individual BERT models to create ensemble models.

And as our society becomes increasingly digital, new data sources for sentiment analysis will emerge. Sentiment analysis on data that could not be collected due to technical limitations, such as video content or virtual reality environments, or data that had not been attempted before would be an appropriate example (Ng et al., 2023a). Innovative applications of sentiment analysis in the field of public policy may include sentiment analysis to track sentiment in real time, predict policy impact based on public sentiment, and promote more effective public counseling. With these potential future developments, sentiment analysis will play an even more important role in shaping and evaluating public policy. These developments are expected to provide a more comprehensive understanding of public sentiment.

2.3.4 Actual application

The practical application and effectiveness of sentiment analysis in shaping and evaluating public policy have been demonstrated in numerous studies.

Sajwan analyzed Twitter data regarding the Department of Defense's agnipath plan using sentiment analysis to provide insights that can be used for decision-making or policy related to the plan (Sajwan et al., 2023). Similarly, research on healthcare workers' mental health and burnout during the COVID-19 pandemic has used natural language processing and sentiment analysis to characterize policymakers' attitudes and perspectives (Abrams et al., 2023). Analyzing SNS posts made it possible to extract public sentiment on various aspects of the policy. Through this, it was possible to obtain essential insights on areas that had been well received and criticized and ultimately helped to establish future medical policies. Additionally, in Indonesia, public reaction to billboard placement by politicians was investigated during the pandemic to categorize the reaction and provide advice for decision-making related to billboard construction (Bastian et al., 2023).

As such, the application of sentiment analysis to public policy, politics, economy, society, and culture has been studied for a long time in various fields at home and abroad. It is expected that this study will play a helpful role in finding a breakthrough by understanding public opinion on government policies for Korea's low severe birth rate problem and promoting policies in a more productive and rational direction.

2.4 The Power of YouTube

2.4.1 Introducing YouTube's Influence

Since its inception in 2005, YouTube has evolved from a simple video-sharing platform to a pivotal media force, significantly impacting public stories and perspectives (Burgess and Green, 2018). The number of YouTube users worldwide is expected to continue growing between 2023 and 2028, totaling 263 million. After five consecutive years of growth, the number of YouTube users is estimated to reach 1.1 billion, and is expected to reach a new peak in 2028.(Statista, 2021)

It started as entertainment, but the scope of YouTube now goes beyond that. It serves as an essential information channel, activism platform, educational hub, and forum for cross-cultural exchange. In this digital age, YouTube is not a passive viewing platform. It empowers media as powerful tools by providing users with authentication and voice to highlight pressing issues, critique policies, or provide educational insights (Dai et al., 2023). Direct consumer communication channels such as comments should recognize their role in directly or indirectly shaping and influencing the public agenda. The vast amount of content coupled with its expansive platform reverberates globally, influencing decision-makers and ordinary citizens, without limiting YouTube's influence.

2.4.2 Social Media in Public Policy Analysis

Entering the 21st century, traditional public policy analysis still yields essential statistics and findings. However, the insights and opinions often provided by social media now serve as a critical influence channel for public policy analysis. Among these platforms, YouTube stands out as a dynamic indicator of public sentiment by aggregating highly engaged visual content (Hendriyadi et al., 2023)

YouTube's unique global reach ensures representation by giving affluent people a voice, from large institutions to individuals broadcasting from their bedrooms. This democratization has led to a content revolution that presents a richer and more comprehensive understanding of social issues and diverse perspectives (Yang, 2023). These inclusions reflect diverse opinions and provide tangible data about public opinion through the structure of YouTube, including engagement metrics such as likes, shares, and comments.

The unique design of YouTube's algorithm is primarily to keep users interested but with unexpected results. It amplifies people's interest in social issues and pushes them to the forefront (Batcha et al., 2023). Thanks to these algorithms, people are becoming accustomed to understanding and reading other people's thoughts and opinions or encountering social issues they did not know about through YouTube. These YouTube-created trends accelerate policy response and underscore the need for policymakers to align policy direction with public sentiment and opinion continually.

2.4.3 Utilization of YouTube from a Social Problem Perspective

The 2010-2012 Arab Spring movement demonstrated the essential role of YouTube in global policy. Footage on a smartphone described ground reality challenged official stories, and provided a filter-free view of the uprising. This user-generated content is gaining worldwide attention by eliciting international responses, influencing foreign policy decisions, and determining the nature and scope of international intervention. (Howard and Hussain, 2011)

During the COVID-19 pandemic, YouTube became a hub for information dissemination. Governments, healthcare organizations, and independent professionals are all leveraging this platform to educate the public. Nevertheless, it has also become a hotbed of misinformation, sparking debate over content regulation, the role of tech giants in public policy, and responsibility in a crisis (Cinelli et al., 2020).

A representative example of using YouTube exists in Korea as well. In late 2016, South Korea was embroiled in a political scandal involving then-President Park Geun-hye and her aide, Choi Soon-sil. Corruption, influence, and suspicion of manipulation of state affairs aroused tremendous national anger (Sang-Hun, 2006). This led to large-scale protests called the 'candlelight revolution' in Korea. They were differentiated by using social media platforms such as YouTube to support nationwide protests, share information, and coordinate them. Mainstream Korean media is often criticized for being influenced by political affiliation, so many Koreans have turned to alternative media outlets like YouTube and Instagram for news.

Independent journalists and ordinary citizens used YouTube to share real-time updates, findings, and personal perspectives on scandals, and these videos went viral and racked up millions of views, revealing the depth of public sentiment and displeasure. YouTube and other social media also played an essential role in organizing and promoting candlelight rallies and demonstrations in Seoul and nationwide. The accumulated pressure from the public, amplified and coordinated through platforms such as YouTube, played a significant role in the impeachment and removal of President Park Geun-hye in March 2017. It proved a safe and powerful force in a peaceful protest where violence was not used. This is a significant moment in the history of democracy in South Korea, powerfully demonstrating the power of public sentiment to influence national policy decisions through social media platforms.

2.5 Comparative Study of NLP Models

2.5.1 Introduction to machine learning and deep learning models

With the start of the 4th industrial revolution, big data technology and analysis technology centered on artificial intelligence (AI) are rapidly developing (Theodosiou and Read, 2023a). Machine learning (ML), a sub-field of AI, has developed spectacularly over the past decades and is emerging in the digital age (Sahu et al., 2023b).In essence, ML refers to the ability of a system to learn and improve from experience without being explicitly programmed. The massive explosion of digital data has fueled the adoption of ML technologies in fields ranging from finance to healthcare, as it facilitates the automation of building analytical models. (Vinod and Prabaharan, 2023a)

The importance of ML is further emphasized in the current era of data, which offers significant benefits. First, ML explores the complexity of large data sets to ensure timely and efficient analysis (Mah, 2022b). Predictive capabilities can also identify patterns and anomalies essential for decision-making processes in many fields, such as diagnosing COVID-19 pneumonia using artificial intelligence technology (Vinod and Prabaharan, 2023a). For example, investors are using deep learning models to predict and evaluate stock and foreign exchange markets due to the advantages of artificial intelligence (Sahu et al., 2023b).

Deep learning, a subfield of ML, models the human brain with algorithms called artificial neural networks (Theodosiou and Read, 2023a). Unlike traditional ML models, deep learning can automatically extract features from raw data, making it suitable for complex tasks such as image and speech recognition (Mah, 2022b). Deep learning models, particularly those that utilize neural network layers (referred to as 'deep' architectures), have demonstrated superior performance in various applications.

NLP is a domain that has benefited significantly from advances in deep learning. Historically, NLP has relied heavily on hand-crafted features and rule-based systems. Incorporating deep learning models into NLP is a paradigm shift, enabling more data-driven approaches and achieving higher levels of accuracy in tasks ranging from machine translation to sentiment analysis (Mah, 2022b). Deep learning's strength in NLP is often due to its ability to capture contextual information, and it is adept at understanding human language's nuances and meanings.

BERT (Bidirectional Encoder Representations from Transformers) introduced by Devlin et al. (2018) demonstrates the transformative power of deep learning in NLP (Devlin et al., 2018). As a pre-trained model, BERT captures the bidirectional context of the text, setting new benchmarks in several NLP tasks and paving the way for numerous transformations and adaptations suitable for different languages and functions.

The transformative impact of machine learning and deep learning is not limited to technological advances. It has far-reaching social implications. The power of ML to analyze massive data sets has fueled innovation in many sectors. For example, in finance, ML algorithms are now fundamental to fraud detection by exploiting patterns and anomalies that are insurmountable by human analysts (Sahu et al., 2023b). In the creative arts, AI-powered tools that support music creation, art creation, and even scriptwriting are emerging, challenging traditional paradigms of creativity.

The dominance of deep learning, particularly in computer vision and NLP, has resulted in smart devices and applications with enhanced user experiences. Virtual assistants such as Siri, Alexa, and Google Assistant, which rely heavily on NLP and deep learning models, are seamlessly integrated into our daily lives, demonstrating the ubiquity and impact of these technologies.

The future potential of machine learning and deep learning seems limitless. As computational power expands and algorithms become more sophisticated, these systems are expected to come closer, if not exceed, to human capabilities in various complex tasks. The realm of quantum computing offers tantalizing possibilities for ML with the prospect of exponentially faster data processing and analysis.

However, the rapid proliferation of these technologies also requires caution. In particular, ethical considerations around data privacy, bias, and transparency of algorithms will play a pivotal role in shaping the future trajectory of ML and deep learning. Ensuring these technologies are utilized in the public interest while mitigating potential risks is paramount.

2.5.2 Individual Overview of Models

a. Logistic Regression

Originating from the pioneering work of Pierre François Verhulst in the 19th century, logistic regression has become an established mainstay within statistical modeling, primarily when binary outcomes are used. Its elegance lies in its ability to express a linear relationship between an independent function and the log probability of its dependent binary outcome.

In the multifaceted domain of NLP, logistic regression's adaptability and interpretability make it the preferred choice for various tasks. An important application is sentiment analysis, where the model efficiently classifies text data into predefined sentiments (positive, negative, or neutral). Its power is emphasized when considering word frequency, the presence or absence of a particular word, the integration of n-grams, and features ranging from contiguous sequences of 'n' items in a given text or speech sample (Musleh et al., 2023d).

Over the years, innovations such as L1 and L2 regularization techniques have been combined with logistic regression to optimize the performance of NLP. These advances prevent overfitting and facilitate identifying and selecting salient features in high-dimensional text data sets (Zhang, 2022)

A retrospective review of NLP methodologies highlights the continuing importance of logistic regression. Despite the rapid rise of neural networks and complex deep learning architectures, logistic regression is highly regarded, especially in applications that require fast, interpretable, and reliable results. Its intuitive nature and computational efficiency make it an indispensable tool for sentiment analysis, such as game review (Tan et al., 2022) and YouTube comments analysis. (Musleh et al., 2023d)

In conclusion, the resilience and versatility of logistic regression in adapting to the evolving needs of NLP and sentiment analysis confirm its fundamental importance in the field by linking coherent insights with the nuances of raw texts.

b. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) have become exemplary classifiers for their outstanding efficiency, especially in high-dimensional spaces. The essence of SVM lies in achieving the broadest possible margin between classes by determining hyperplanes that unambiguously separate categories in the feature space.

The influence of SVMs is felt prominently within NLP. The high dimensionality of text data, potentially with thousands of individual words, makes SVMs attractive for text classification and sentiment analysis. For example, SVM has been applied to analyze tourist reviews in Indonesia and has been shown to outperform other algorithms, such as K-Nearest Neighbors (Sari et al., 2023) Another study examined public sentiment on social media Twitter about childlessness using SVMs, achieving an F1 weighted average of 95.639% for education data and 60.45% for test data (Siregar et al., 2023).

One of the characteristics of SVM is the integration of kernel tricks. This allows for proper handling of the non-linearly separable data frequently occurring in the text, allowing the data to be projected into a high-dimensional space (Schölkopf and Smola, 2002). Kernels ranging from linear and polynomial to radial basis functions (RBF) provide versatility, allowing SVMs to model various complex language structures and complexities.

SVMs also exhibit resilience to overfitting, which is particularly advantageous when dealing with high-dimensional text data with a limited sample. This property is often attributed to the margin-maximization principle, which ensures that models are as general as possible. In a study on sentiment analysis of public responses to the Indonesian government, SVM achieved 85% accuracy, 95% recall, and 83% accuracy, showing greater accuracy than other methods. (Setyawan et al., 2023)

A recent application of SVM in the analysis of flood disaster management in Jakarta on Twitter achieved an accuracy of 88.6%, precision of 88.6%, and recall of 89.4%, further highlighting the power of SVM in sentiment analysis. (Saddam et al., 2023)

In summary, the SVM's blend of mathematical rigor and flexibility reflects its ability to transform raw textual information into nuanced, actionable knowledge, cementing its reputation as an essential tool for sentiment analysis in NLP environments.

c. Naïve Bayes

The naive Bayes classifier, based on Bayes' theorem, is a stochastic approach in the domain of supervised machine learning known for its simplicity and efficiency. These methods often exploit the stochastic dependence of a feature variable on a particular class label to make a blanket "naive" assumption of feature independence.

Naive Bayes has gained considerable recognition in NLP, particularly in text classification and sentiment analysis. Given the basic premise of the model, text data, where words or phrases are independent features, provides a suitable playground for the Naive Bayes approach. An important application is sentiment analysis for product reviews, where Naive Bayes classifiers have shown impressive efficiency in classifying text sentiment into positive, negative, or neutral buckets (Sathya and Mythili, 2023).

The Naive Bayes methodology exhibits different characteristics depending on the underlying data distribution. Gaussian Naive Bayes is good for continuous data, but the polynomial version takes discrete counts into account. Therefore, it is ideal for text data and suitable for word count-based features in NLP (M, 2023).

Despite naive assumptions and the advent of complex models, Naive Bayes classifiers remain attractive due to their computational ease and consistent performance, especially on large text corpora. Its power is amplified when data sparsity occurs because models can be created even with limited training data (Waworundeng et al., 2022a).

Naive Bayes classifiers deftly decode linguistic complexity into identifiable sentiment indicators, leaving them as instrumental mechanisms for sentiment analysis with both stochastic bases and empirical success. Its applications range from sentiment analysis to online lectures to text-based detection of signs of depression (Waworundeng et al., 2022a, Erfina and Nurul, 2023)

d. Long short-term memory (LSTM)

Long short-term memory (LSTM) networks, a specific type of recurrent neural network (RNNs), have been instrumental in solving the problem of learning long-term dependencies from sequence data (Altameemi and Altamimi, 2023). These networks have been particularly effective at overcoming challenges such as gradient collapse and bursting problems that have been significant obstacles in conventional RNNs (Kaur and Sharma, 2023).

LSTMs are famous for their unique architecture that includes a gating mechanism to control the flow of information. This allows the network to remember or forget information about extended sequences, making LSTMs suitable for various NLP tasks where context and line are essential (Kothuri and RajaLakshmi, 2022).

In sentiment analysis, it is essential to understand the context, as the sentiment of a word can vary greatly depending on the surrounding terms. LSTMs with memory cells efficiently capture these contextual relationships, outperforming traditional methods whose complexity is often overlooked (Alsayat, 2022).

A bidirectional LSTM (BiLSTM) that handles past and future state data further enhances this capability, providing a more comprehensive view of the textual context (Sadanand et al., 2022). However, LSTMs have their problems. Computational complexity has led to simpler variants, such as Gated Recurrent Units (GRUs), which perform similarly on many NLP tasks.(Kothuri and RajaLakshmi, 2022a)

Recent advances, such as the attention mechanism, have further improved the efficiency of LSTMs, allowing the model to focus on specific parts of the input sequence. This has resulted in improved results in sentiment analysis and other NLP applications (Lasri et al., 2023).

In summary, LSTMs have left a lasting mark in NLP, especially in sentiment analysis. Its ability to capture long-term dependencies and complex contexts affirms its importance in modern research and applications.

e. Bidirectional Encoder Representations from Transformers (BERT)

BERT's real strength lies in its pre-training methodology. Instead of being narrowly specialized, BERT models are trained on a broad corpus to get a general picture of a language before fine-tuning it for a specific task. A recent study by George and Sumathy (2023) proposed a hybrid model of BERT and Latent Dirichlet Allocation (LDA) in topic modeling, demonstrating the effectiveness of BERT in extracting meaningful themes from large text corpus. (George and Sumathy, 2023)

This universal applicability has made BERT a mainstream NLP tool. Ruder argued that BERT is a powerful tool but that a one-size-fits-all approach can only partially express a language with unique characteristics (Ruder, 2019). The BERT multilingual model attempted to address this problem, but as discussed by Goldberg, a dedicated model may be required to achieve true linguistic segmentation (Goldberg, 2022). For example, applying BERT in low-resource languages such as Kurdish has outperformed conventional machine learning classifiers in sentiment analysis, highlighting its adaptability and efficiency. (Badawi, 2023)

In this context, BERT variant models such as KoBERT and KcBERT have emerged as Korean-specific counterparts. Recognizing the inherent complexity of the Korean language, SK Telecom introduced KoBERT in 2019, specifically designed to capture the complexities of Korean syntax, sentiment, and cultural meaning (SKTelecom, 2019). KcBERT was designed with a focus on capturing the changing characteristics of the Korean language under the influence of the digital age. Lee emphasized the importance of specialized models by emphasizing that general models may miss some subtleties(Lee, 2020). Deepening the discourse, Yeonji extensively explored Korean NLP tools, advocating a transition to professional models (Yeonji et al., 2022). Their findings revealed scenarios where even advanced tools like BERT faltered for the unique construction of Korean.

In the broader context of NLP, BERT's applications have extended to various domains, including network intrusion detection where it has been used to improve the field adaptability of intrusion detection systems (Nguyen and Watabe, 2022). Additionally, the BERT model is evaluated for machine-assisted indexing of digital library collections, demonstrating its versatility (Chou and Chu, 2022)

BERT has been applied to electronic health records (EHRs) for various purposes in the medical domain. For example, BuHamra developed an NLP tool to extract COVID-19 mortality and co-morbidity from EHRs using BERT, demonstrating the potential of BERT in handling complex and messy medical data (BuHamra et al., 2022). Another study by Zhou utilized BERT to standardize diagnostic statements for pituitary adenoma and highlighted the need for a structured pattern in medical diagnosis. (Zhou et al., 2022a)

Looking back at the journey from BERT to Korean derivatives and applications in the medical field, it becomes clear that NLP is moving from a generalized model stage to a specialized language tuning solution. As NLP advances at an unprecedented pace fueled by technological innovation and language recognition, models such as KoBERT and KcBERT highlight the convergence of technological and linguistic heritage, setting a precedent for future developments in the field.

2.5.3 Comparative Analysis

NLP has become the point of convergence of algorithms and models, each evolving to decipher the nuances of human language. Explore the complexities of logistic regression, SVM, Naive Bayes, LSTM, and BERT to gain a panorama of features, challenges, and innovations. Therefore, comparative analysis by model is an essential part of this study. A comparative analysis follows.

Accuracy is still an essential benchmark in the NLP domain. Logistic regression, SVM, Naive Bayes, and other models have been used for sentiment analysis with varying degrees of success.

For example, a study on Arabic sentiment analysis of YouTube comments achieved 94.62% accuracy with Naive Bayes (Musleh et al., 2023d). Another research of sentence-level sentiment analysis showed that SVM achieved 90.43% accuracy. (Maree et al., 2023a)

The spectrum of computational needs across models is vast. BERTs require high-end GPUs and significant memory, often pushing current hardware's boundaries. Conversely, traditional algorithms such as SVM and logistic regression can be more resource efficient. A study on aspect-based sentiment analysis in Indonesia utilized SVM, Logistic Regression, and BERT, with BERT achieving the best performance (Bahri and Suadaa, 2023a)

Balancing model complexity and interpretability is a recurring theme. In deep learning, the layered architectures of LSTMs and BERTs provide state-of-the-art performance but can be perceived as opaque. However, the inherent transparency of Naive Bayes and the mathematical elegance of logistic regression often allow more direct insight into the decision-making process.

Models like BERT offer a wide range of fine-tuning possibilities, allowing them to adapt across tasks and languages. Conversely, traditional models often rely on domain-specific feature engineering for optimal results. Studies using the GRA methodology for sentiment analysis have highlighted the adaptability of these techniques using various models including Random Forests, SVMs, and Naive Bayes. (Inder, 2022a)

All models in the NLP repertoire have constraints. Logistic regression may need improvement in handling non-linearity, SVMs may be sensitive to kernel choice, and Naive Bayes' naive assumptions may fall short of capturing complex relationships. LSTMs can face vanishing gradient problems, and BERTs require substantial computational resources.

The potential applications for the various models are vast. Logistic regression and SVM have been pivotal in sentiment analysis and text classification tasks. Naive Bayes has been central to BERT for complex tasks such as spam filtering, LSTMs for sequence-related tasks, and question-answering systems.

Over the years, NLP models have made significant progress. From statistical models to deep learning innovations, there is a clear trajectory of increasing complexity and capabilities. Introducing various kernel and optimization techniques, attention mechanisms, and transformer architectures such as BERT has driven this evolution. The dynamic tapestry of NLP models opens up a field of profound challenges and transformative possibilities through the complex interplay of neural structures with traditional methods. Model selection depends on specific goals, computational constraints, and explainability. NLP is at the cusp of continuous innovation, and a comparative analysis of these models provides valuable insights into their strengths, weaknesses, and applicability.

2.5.4 Model Selection Criteria for NLP

NLP has various modeling options, making selection criteria a complex combination of task details, data constraints, and computational capabilities. There are several things to consider when choosing a data model.

Every NLP task has its own set of challenges. For example, sentiment analysis can benefit from models optimized for sharp-boundary decision-making, such as SVMs (Maree et al., 2023a). Inter-order tasks such as translation require the memory and processing power of models such as LSTMs that can effectively handle the interdependencies of sequential data (Inder, 2022a). NER or other token classification tasks may require an architecture that leverages contextual information from surrounding tokens to make transformer-based models such as BERT more appropriate (Bahri and Suadaa, 2023a).

Data serves as the foundation for all machine learning tasks. Normalize deep neural networks with richly labeled data to achieve state-of-the-art performance. In scenarios with sparsely labeled data, models like BERT shine by leveraging extensive pre-training on large corpora and fine-tuning capabilities on smaller, task-specific datasets (Mensah et al., 2021). Conversely, simpler models such as Naive Bayes, SVM, or Logistic Regression may be more suitable in data-poor situations because they reduce the risk of overfitting (Musleh et al., 2023d).

Every model presents unique computational requirements. Deep learning architectures, especially those like BERT, require powerful GPUs for training and inference. These requirements can sometimes be prohibitive for real-time applications or edge device deployments. Conversely, algorithms such as Logistic Regression or Naive Bayes may be more tractable in scenarios with limited computational resources due to their relative simplicity and linear complexity (Musleh et al., 2023d).

Another criterion governing model selection is interpretability. Deep learning models deliver excellent performance, but their "black box" nature can hinder them, especially in healthcare or finance, where model decisions must be transparent. Simpler models, such as decision trees or logistic regression, provide a more transparent decision-making framework.

In summary, NLP environments with numerous models require a careful selection process based on multiple considerations to achieve optimal results. The choice of model depends on specific goals, computational constraints, and explainability. The dynamic nature of NLP tasks and the availability of various models require a tailored approach to model selection.

2.5.5 Performance Metrics in NLP

Evaluating models in NLP requires a multi-faceted approach that considers different domain problems and specific data characteristics.

Pure accuracy can sometimes paint an imperfect picture. This is especially true when the data set exhibits class imbalance. Combining both precision and recall, the F1 score bridges this gap. It is used as an indispensable indicator in various fields, such as disease prediction and deep learning model selection. For example, in hepatitis C prediction, these metrics have been used to compare the performance of different machine learning techniques (Alizargar et al., 2023a). Similarly, deep learning models for selecting appropriate requirements extraction techniques have been evaluated using these metrics (Dafaalla et al., 2022a).

ROC and AUC emerge as valuable tools for model performance at different thresholds. Their applications are particularly relevant in sentiment analysis, text classification, and other binary NLP tasks. Recent studies on hepatitis C prediction and computer network security have utilized these metrics to evaluate the effectiveness of different machine-learning models (Alizargar et al., 2023a, BaniMustafa et al., 2022a).

Multiclass NLP workspaces with problems such as subject classification and sentiment spectrum analysis require a detailed understanding of error patterns. The confusion matrix provides detailed insight into the type and frequency of misclassifications, enabling more targeted model improvement. This metric selected an appropriate requirements extraction technique in a deep learning model (Dafaalla et al., 2022a)

Perplexity is the primary metric for probabilistic model evaluation, especially in language modeling and machine translation. It quantifies a model's predictive ability for a given sample, serving as a reliable indicator of its general quality and predictive power. The BLEU score is most important in the machine translation domain. Machine-generated translations are compared against a set of reference translations to evaluate the quality and fluency of the output, making it a widely used standard for assessing translation models.

In addition to the above metrics, other metrics such as accuracy, specificity, sensitivity, and calibration have been used in various contexts. For example, in traffic security in computer networks, these metrics have been used to evaluate machine learning models (BaniMustafa et al., 2022a).

Taken together, these metrics highlight the depth and breadth of consideration in NLP model evaluation. Their diversity is a testament to language's complex and diverse nature and the models designed to understand and generate language. Due to ongoing advances in NLP and related fields, these metrics are continually being improved and extended to address the unique challenges of language processing.

2.5.6 Trade-offs

Each criterion for selecting a model has its pros and cons. Exploring the vast landscape of NLP models often requires making deliberate trade-offs based on the needs of a particular task, computational constraints, and domain-specific requirements. We will discuss the potential pros and cons of choosing an NLP model. Regarding the potential pros and cons, there are several considerations.

First, speed versus accuracy. There is a constant tension between a model's run-time efficiency and accuracy. Fast response times are paramount for real-time applications such as chatbots or instant translations. We lean towards less complex and faster models even at the slightest loss of accuracy The second is interpretability versus complexity. Specific sectors, especially those subject to stringent regulations such as healthcare, finance, or law, need models that predict and explain the 'why' behind decisions. In this context, the transparency and interpretability of models such as Logistic Regression or Naive Bayes can be considered more valuable than the predictive potential of complex deep learning architectures (Caruana et al., 2015, Doshi-Velez and Kim, 2017). Third is performance versus computing requirements. For state-of-the-art performance on complex tasks such as sentiment analysis, question answering, or language translation, sophisticated models like BERT often emerge as frontrunners. However, the computational requirements regarding processing power and memory can be significant, making them potentially unsuitable for deployment on edge devices or in environments with limited computing resources (Wu et al., 2016, Tan and Le, 2019). Finally, data availability versus model capacity. Despite their capabilities, deep learning models often require vast amounts of labeled data to reach their full potential. In scenarios with limited data, simpler models or techniques, such as transfer learning, may be more effective (Ruder, 2019).

By nature, choosing an NLP model is not a one-size-fits-all solution. It exemplifies the adage that context is king in machine learning, as it requires a careful understanding of the task at hand, the constraints under which it runs, and the desired outcome.

Chapter 3. Research Question

In the digital age, public sentiment is increasingly being expressed on online platforms. Websites like YouTube are invaluable to policymakers as they provide a rich data source for measuring public opinion on various topics. This study addresses the question: How do different NLP models work to interpret public sentiment in YouTube comments?

a. Key research questions:

- How effectively can leading NLP models (logistic regression, naive Bayes, SVM, LSTM, and BERT) evaluate and contrast the sentiment of YouTube comments on a demographic issue in South Korea?

b. Sub-question:

- How do traditional machine learning models such as logistic regression and SVM compare to advanced deep learning models such as BERT and Transformers for sentiment analysis tasks?

- What are each model's unique strengths and limitations for discerning public sentiment in text data?
- Which model provides policymakers with the most accurate and actionable insights into the public's view of Korea's demographic challenges?

c. Rationale

Sentiment analysis has long played a pivotal role in determining public perception. As digital data grows exponentially with advances in technology, pinpointing a model that excels at identifying these emotions becomes an even more important variable. The decision to compare these particular models is driven by their current reputation and the different methodologies they implement within NLP. Existing models such as Logistic Regression and SVM are favored for their clarity and simplicity, while innovative models such as BERT and Transformers are redefining the limits of sentiment analysis by understanding the context in ways previously thought out of reach.

The purpose of this study is to provide useful insights for public policy evaluation by identifying the optimal model or combination of models for sentiment analysis on YouTube comments. By studying in depth the efficacy, strengths, and limitations of these models, this study will not stop at revealing public sentiment about Korea's demographic problems. This study will guide future scholars to wisely apply NLP tools to similar studies.

d. Expected Outcomes

The comparative analysis is designed to:

- Highlight models that outperform others in terms of accuracy, processing speed, and knack for detecting subtle emotions.
- Provides a holistic understanding of potential biases or shortcomings inherent in each model.
- Provides policymakers with actionable information about public sentiment and the consequences for demographic strategies.

Chapter 4. Methodology

Recently, as comments on Naver, Korea's leading portal site, were suspended due to political influence, public attention was focused on platforms such as YouTube. Recognized for its extensive user-generated content in the Korean language, YouTube serves as a lens through which to look into public sentiment , providing an essential repository for understanding prevailing opinions. The platform's real-time distribution of various news stories and its interactive nature allows many users to express their feelings. This study is rooted in sentiment analysis of YouTube comments, focusing on Korea's low birth rate policy.

Given YouTube's vast amount of content, it is paramount that researchers adopt a systematic approach to extract relevant data from its wealth. The YouTube Data API plays a key role in this effort. Tailored for efficient data extraction, this API is particularly adept at mining video comments, making it an indispensable tool for sentiment analysis for the Korean public.

The table below is a literature review table of previous studies that performed machine learning and deep learning using NLP dataset. Sentiment analysis using NLP data is being actively implemented, and in line with this trend, this study is expected to contribute to the identification and analysis of public policy public opinion using NLP dataset.

< Table 1 : Literature Survey on Previous Studies on Sentiment Analysis Using NLP Data>

Authors	Used Model	Speicific Features	Data Source	Dataset Size	Accuracy	Key Findings
(Hossain et al., 2021)	Logistic Regression, Naive Bayes, SVM, SGD	Unigram, Bigram, Trigram features	Bengali Book Reviews	2000 reviews	84%	Multinomial Naive Bayes with Unigram outperforms other techniques
(Raza et al., 2019)	Naive-Bayes, SVM, Logistic Regression, Decision Tree, K-Nearest Neighbor, Random Forest	Lemmatization, N-graming, Tokenization, Stop word removal	Scientific Articles	8736 citation sentences	Improved by 9%	Improved accuracy with additional features selection techniques
(Gamal et al., 2018)	Naïve Bayes, SGD, SVM,	Unigram	IMDB, Cornell		87% to 99.96%	Performance of PA with a unigram is

	PA, ME, AdaBoost, MNB, BNB, RR, LR		Movies, Amazon, Twitter			the best among other algorithms
(Ali, 2021)	Naïve Bayes, MNB, KNN, Logistic Regression, SVM	Information Gain (IG) as a filtering technique	Arabic Tweets about COVID-19		89.6%	Proposed model performs well in analyzing perception about coronavirus
(Bahri and Suadaa, 2023)	SVM, Complement Naïve Bayes, Logistic Regression, BERT, IndoBERT, mBERT	Aspect-based analysis (attractions, facilities, access, price)	Bromo Tengger Semeru National Park, Indonesia (Google Maps Reviews)		91.48% (IndoBERT), 89.16% (SVM)	IndoBERT achieved the best performance among the models
(Musleh et al., 2023)	SVM, Naïve Bayes, Logistic Regression, KNN, Decision Tree, Random Forest	Arabic comments classification	YouTube Comments	4212 labeled comments	94.62% (NB)	Naïve Bayes achieved the highest accuracy among the models
(Jain et al., 2023)	Logistic Regression, Naïve Bayes, SVM, LSTM, BERT	Bag-Of-Words, TF-IDF	COVID-19 Vaccine Tweets		90.42% (BERT), 88.7989% (SVM)	SVM and BERT achieved the highest accuracies among the models
(Herdiyani and Zailani, 2022)	Random Forest, Naïve Bayes, Lexicon, SVM, NBC, Neighbor Weighted K-NN, Logistic Regression	TF-IDF	Tweets about Indonesia's Capital Relocation	1639 tweets	76%	Random Forest method used to classify tweets into positive, negative, and neutral
(Maree et al., 2023a)	Support Vector Machines, Naive Bayes, Logistic Regression, Random Forest	Lemmatization, knowledge-based n-gram features, WordNet	Movie Reviews	50,000 movie reviews, 10,662 sentences, 300 generic movie	90.43% (SVM)	Coupling lemmatization and knowledge-based n-gram features produced higher accuracy results

				reviews		
(Eom et al., 2022)	SVM, RNNs, LSTMs, BERT, KoBERT	Sentiment analysis regarding vaccination after COVID-19 Omicron variant	South Korean Twitter data		71% (KoBERT)	KoBERT showed the best performance in all predictive performance indicators
(Kastrati et al., 2021)	NLP, Deep Learning	Sentiment analysis in education	Students' feedback in learning platforms	92 studies from 2015-2020		Deep Learning is the most recent trend in sentiment analysis in education
(Ali et al., 2019)	Self-Organizing Map, Principal Component Analysis, Adam Deep Learning	Sentiment classification in social networks	Social network data	Various sizes		Proposed approach is efficient and feasible for Big Data
(Ghorbani et al., 2020)	Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM)	Sentiment analysis in cloud computing			89.02%	ConvLSTMConv network provides an appropriate solution for sentiment analysis
(Prottasha et al., 2022)	BERT, CNN-BiLSTM	Transfer learning for sentiment analysis	Bangla NLP domain			BERT-based supervised fine-tuning outperforms other embeddings and algorithms

4.1.1. Collecting Public Voice

Detailed methodology for data collection:

- **Setting:**
- Platform registration: To extract data, the project should be registered in the Google Cloud Console, a precursor to accessing the Google API suite.

- API activation: After registration, the YouTube Data API v3 was activated, providing a robust feature set for needs.
- Generate API key: An API key was generated after. This key is pivotal for two reasons. Authenticates data requests and monitors usage to ensure compliance with specified quotas.

- **Video identification:**

- Formulating criteria: Due to the vast repository of Korean language content on YouTube, meticulous criteria were essential to ensure the accuracy of the study. We focused on videos that match Korean research topics with significant user interaction.
- Content source: To secure accurate information, we selected news from famous domestic broadcasters such as JTBC, SBS, and MBC, focusing on policy news related to the low birth rate over the past three years.
- Keyword Implementation: Search parameters were created around the country's approved low birth rate policy. Videos with significant comment interaction were targeted for analysis according to criteria set by public broadcasters.

- **Extract comments:**

- Tool of choice: PYTHON and COLAB, the primary programming language for these tasks, facilitated annotation extraction. A batch request was made to mine comments using the generated API key and video ID.
- Data retrieval: Each request often gets several comments and a pagination token to ensure comprehensive comment navigation. Expressions, including emoticons, were captured to provide nuanced sentiment analysis. Extraction was not limited to top comments to maintain overall sentiment capture.

- **Data structure and storage:**

- Standardization: Homogenize comments into a unified format before storage.
- Storage Separation: Comments categorized by corresponding video ID were partitioned into separate datasets but maintained a consistent, structured format. This way of working ensures clarity during analysis by ensuring each comment is contextually linked to the

original video.

4.1.2. NLP Preprocessing

It is essential to start with clean, structured data for practical sentiment analysis, especially for data sets as varied and complex as YouTube comments on politically charged issues such as South Korea's low birth rate policy. This chapter looks at the preprocessing steps to refine the data set and ensure preparation for subsequent analysis steps.

The raw data set culled directly from YouTube had some problems. First of all, comments sometimes mixed Hangul with Chinese characters or Latin scripts due to the global nature of the platform. In addition, there was a problem that the complexity inherent in Korean, such as regional dialects and deadlock structures, could confuse analysis. In addition to this, comments contained non-essential information such as URLs, user comments, and repetitive phrases that could distort the analysis. Many cases were included, and there were unnecessary elements in constructing and analyzing sentences, such as using emoticons. Given these challenges, pretreatment was not only desirable but essential.

a. Data preprocessing methodology

- Standardize scripts: Before going into deeper language processes, ensuring homogeneous scripts across the data set was crucial. All comments were converted to use Korean as the primary language to standardize scripts and ensure linguistic consistency. In addition, we went through the process of removing emoticons for efficient tokenization.
- Tokenization: The hierarchical morpheme structure of the Korean language required a unique tokenization approach. Utilizing advanced NLP tools tailored for Korean, we subdivided the data set into meaningful words or morphemes, ensuring each unit retains its contextual meaning.
- Eliminate stopwords: We used a list of stopwords curated for the context of the study and the Korean language. We have removed unnecessary expressions and terms that do not contain the primary investigation or meaning used in Korean. By removing these standard terms and particles that lack sentiment analysis value, we ensure that the main content of each comment is the focus of our analysis.

- Headwordization: A circular restoration tool was used to counteract word variations that could introduce redundancy. By standardizing words into root forms, the data set was simplified so that different forms of words were recognized and treated as a single entity.

Performing the preprocessing steps as described above ensures that the data set is representative and clean, laying the foundation for accurate sentiment analysis. By applying these improvements, the subsequent stages of modeling and analysis are established on a solid foundation, ensuring the reliability and robustness of the results.

4.1.2 Process and Scoring

In text analysis, sentiment analysis is essential, especially when measuring public reaction to policies. Sentiment analysis is complex in all languages due to linguistic and cultural nuances, but the complexity of Korean amplifies these challenges. This chapter outlines the approach adopted and highlights the sentiment scoring and data labeling process.

Sentiment analysis in South Korea is unique and complex for several reasons. With its long history, Korean has an agglutinative structure and uses honorifics that can subtly change the sentiment of a sentence. Also, many emotions in the Korean language are deeply rooted in a culture or historical context, and can only be deciphered with an intrinsic understanding of that culture. Also, Korean is characterized by the fact that a sentence can convey both positive and negative meanings depending on the context. Given these difficulties, this study opted for sentiment analysis, which adopts a binary method to classify opinions as positive or negative to ensure clarity and accuracy.

a. Sentiment scoring using 'KNU Korean Emotional Dictionary'

There are various tools and vocabularies for analyzing Korean sentiment, but we chose the 'KNU Korean Sentiment Dictionary,' which comprehensively covers Korean sentiment terms. The sentiment of each comment was scored based on the prevalence and weight of the terms listed in this dictionary.

- Word weight: Words in comments were mapped into a dictionary, assigning each word a predefined sentiment weight.

- Aggregated Sentiment Score: The scores for individual words in a comment are aggregated to provide an overall sentiment score for all comments.

b. **Data labeling process for sentiment analysis**

- Labeling training data: Comments were initially labeled as 'positive' or 'negative' using the emotion score derived from the 'KNU Korean Emotional Dictionary.' This automated process provided preliminary labeling but required manual intervention to ensure accuracy.
- Manual review: Each comment was carefully reviewed along with an initial sentiment label.
- Fix and Validate: Fixed a miscategorized comment where the automated sentiment did not match the actual sentiment.

c. Creation of labeled data sets:

- After validation, a comprehensive labeled data set was curated. This dataset, full of positive and negative sentiment tags, served as the basis for subsequent predictive modeling.
- Consolidate Labels: Ensure each comment has a single, accurate sentiment label.
- Structuring the dataset: Structure the dataset to be optimally configured for training the machine learning model in subsequent steps.

c. Conclusion

Sentiment analysis, by its very nature, is a very challenging study, but it forms the basis for understanding public sentiment. This study tried to accurately capture and express Koreans' feelings about the low birth rate policy through a meticulous combination of automated sentiment scores and manual verification.

4.1.2 Comparing Models

Uncovering the underpinnings of public sentiment requires a discerning eye and the most accurate and efficient computational models. In the age of digital expression, emotions are no longer confined to surveys or interviews. They resonate in the vast sea of online content. Therefore, this study aims to discover the most powerful model for sentiment analysis, focusing on YouTube comments, which are a treasure trove of opinions related to Korea's low birth rate policy. A careful evaluation was performed on seven models: Naive Bayes, Logistic

Regression, SVMs, LSTM, KoBERT, KcBERT, and BERT_Multilingual.

a. Traditional model:

- Naive Bayes: Derived from the fundamental principles of the Bayes theorem, Naive Bayes classifiers have historically been admired for their agility and efficiency, especially on large data sets. Its probabilistic basis provides a clear view of affectability.
- Logistic Regression: With a rich statistical base, it was speculated that this model would provide transparent and insightful probability estimates for different emotion classes.
- SVM: Boasting a unique ability to manage high-dimensional data sets, SVMs have been considered valuable tools for fine-grained analysis of text data.

b. Deep learning model:

- LSTM: A subset of recurrent neural networks, LSTMs have been praised for their proficiency in sequence-based tasks, particularly relevant to sentence-level sentiment analysis.

c. BERT variants:

In this study, we also aimed to compare BERT_Multilingual, designed for application to various languages, and a model specialized for Korean, out of the framework of the existing BERT.

- KoBERT & KcBERT: Given the linguistic complexity inherent in Korean, these specialized models were expected to be more adept at identifying subtle emotional expressions.
- BERT_Multilingual: The ability to identify emotions in multiple languages predicted that this model would be indispensable given YouTube's massive global user base.

The YouTube API served as a conduit for sourcing YouTube comments during data collection. These comments are raw and unfiltered but represent pure public opinion. They were subsequently preprocessed to ensure uniformity and relevance. Tokenization, stemming, and stopword removal were pivotal steps in this preprocessing step. This collection and preprocessing process establishes data collection and methodological accuracy.

This study adopted a systematic approach for each model to ensure accurate analysis.

- Training: A broad subset of YouTube comments previously annotated with sentiment labels were used. The key to this step is to immerse the model into the data set to unravel underlying patterns and complexities.

- Validation: It was essential to ensure that the model was neither overfitting nor underfitting. A separate subset of data is reserved for this purpose, allowing hyperparameter fine-tuning.

- Testing: This step was pivotal in evaluating practical applicability. A series of untouched opinions served as a testing ground for each model's efficiency and accuracy.

In addition, a multifaceted evaluation method has been developed for accurate research performance. Models were evaluated on various aspects such as accuracy, precision, recall, and F1 score. Each metric provided unique insights, collectively painting a comprehensive picture of model effectiveness. It focused on detecting negative sentiment, which is essential for understanding public concerns and concerns.

Every robust methodology has its problems, and this study is no exception.

- Data Imbalance: An imbalance in emotional expression was observed. Some emotions were exaggerated while others were underrepresented. To mitigate this, a properly balanced dataset was adopted and used as labeling data, which became the basis for future emotion prediction of unlabeled data.
- Decoding Complex Language Structures: Comments were sometimes straightforward. There were many examples of satire, metaphor. Existing models in particular suffered from this complexity.
- Dealing with Computational Requirements: Deep learning models, while powerful, were resource intensive. Managing the computational requirements, especially for the BERT variant, required careful planning and infrastructure.

In essence, methodology serves as the basis for all empirical investigations. The models chosen for this exploration represent a harmonious combination of time-tested techniques and newly developed deep-learning approaches. This study aims to provide valuable insights into sentiment analysis by evaluating each model's strengths, limitations, and unique suggestions. Our overarching goal is to provide policymakers and analysts with a robust, stable, and scalable framework to decipher accurately and leverage sentiment expressed on platforms like YouTube.

Chapter 5. Result

5.1 Disclosure of Public Opinion: Results of Data Preprocessing

Sentiment analysis is emerging as a pioneering tool in the modern digital realm. It gives stakeholders, policymakers, and organizations a privileged look into the vast expanse of the public's collective psyche. Capturing genuine emotion on platforms like YouTube that provide unfiltered voices has become both a challenge and a necessity. This effort allowed us to accumulate a data set with opinions on seven policies. Raw, unfiltered comments are grains of gold in the sand. This comprehensive report describes the meticulous preprocessing steps taken to capitalize on these treasures and ultimately sets the stage for in-depth sentiment analysis.

As a melting pot of global narratives, YouTube receives daily comments. A vast commenting environment reflects myriad viewpoints, cultures, and opinions. Seven data sets were carefully extracted from this dynamic digital realm, each reflecting a unique policy. Looking at the overall characteristics related to the entire data set, the data set contains an impressive 45,960 comments. In its initial state, analyses averaged 13.96 words. Repeated terms such as '69 hours', 'more,' and 'possible' were emphasized in the preliminaries. While these phrases reflect popular trends, they can also mask the subtle emotions lurking beneath them and act as potential noise amongst the precious data. The entire data preprocessing process is as follows.

5.1.1 Preprocessing steps and rationale

Tokenization: A fundamental yet pivotal step in text analytics, tokenization is akin to unraveling a fabric by exposing each unique thread or token. These threads often make it possible to identify an underlying emotion, suggesting that sentiment is always in a particular word or phrase and not the sum of opinions. In this study, emotion labeling was prepared based on the KNU emotion dictionary through data organization through tokenization.

Stemming: Stemming acts as a language equalizer. The etymology of words is traced to ensure some level of unity. Considering the complex nature of the Korean language, we avoided this in our analysis. Stemming usually acts as a sieve to filter out linguistic redundancies while preserving essence. In order to apply the KNU Sentimental Dictionary, the basic morphemes in the dictionary and words in the dataset must match. Therefore, this study's words in the emotion dictionary and dataset were homogenized through morpheme analysis.

Eliminate stopwords: Every language has essential words that are structurally important but often dilute the rich meaning of a text. A preliminary scan showed "더" (more) and "진짜" (really) as possible stopwords. Their extraction paved the way for fine-tuned datasets for emotion extraction. Since Korean contains various postpositions such as "의, 을, 를, 한." these search terms are a significant obstacle to NLP data analysis. Therefore, a process was implemented to remove these stop words.

5.1.2. data after preprocessing

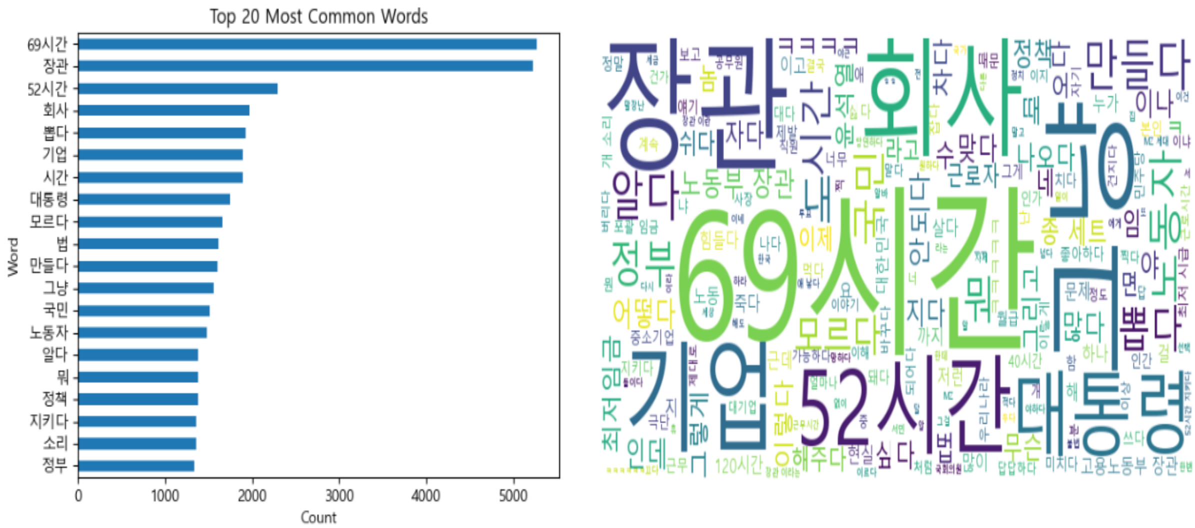
The ramifications of the pretreatment manipulation were evident in the resulting data set. Lighter comments now represent a data set that averages 13.68 words and is ready for scrutiny. Most notably, the linguistic landscape has changed. Words and phrases have gone from generic or contextually ambiguous to emotionally charged. New terms such as 'Walwalmeongmeong', which reminds us of meaningless words and sounds like '다' (all) and '그냥' (just), heralded this change.

A visual representation of data often acts as a bridge, transforming data metrics into understandable insights. In this study, visualization was performed using word clouds and bar graphs. The visualization allowed us to see the underlying themes and interests of the data set. The primary analysis of each of the seven policies was made through visualization.

5.1.3 Detailed review of pre-processing data for specific policies

Public sentiment on significant topics and policies is illuminated in the preprocessing and post-review. Extracted from extensive YouTube commentary, each dataset provides a rich tapestry of public opinion and attitudes toward the policy. A closer look at these subjects helps reveal more than just emotions. It provides a narrative, an ongoing debate, and a collective pulse that resonates with each policy.

a. Office Hours Policy



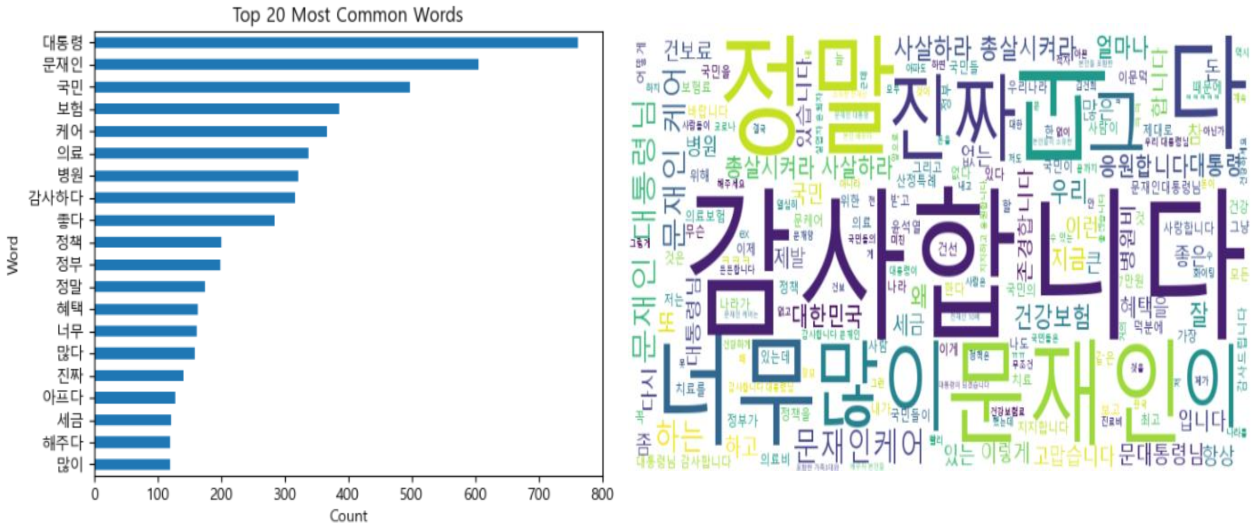
<fig5. Word Cloud & bar chart for Office Hours Policies>

Among the dominant terms in the working hour policy postprocessing dataset, '69 시간' and '52 시간' stand out. South Korea's recent policy of 52-hour work is now considering a 69-hour work week under President Seok-Yeol Yoon, indicating a lively discourse surrounding the details of potentially fixed working hours in the process of shifting the work paradigm.

The frequency of 'ministers' proposing discussions involving ministers or cabinets may indicate government interventions or decisions that have attracted public attention. Interestingly, '기업' (Enterprise) and '회사' (Company) suggest exploring the impact of these policies on businesses and their employees, emphasizing the central role of the corporate sphere in this conversation.

The existence of a '법' (law) means that legal effect or change is expected or actively discussed. These word frequencies suggest that stories of legal, corporate, and policy change intertwine to create a multi-layered discourse.

b. Medical Policy

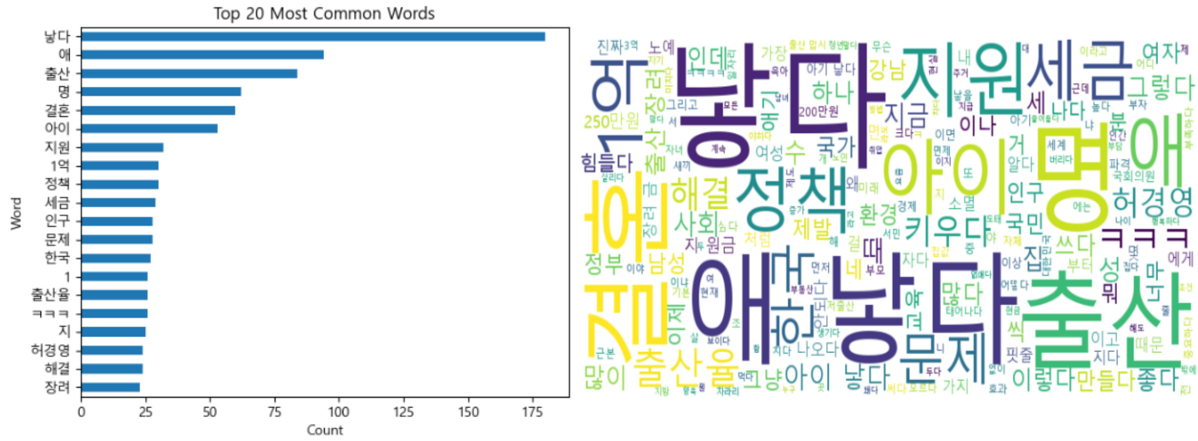


<fig6. Word Cloud & bar chart for Medical Policy>

Health care, universally of significant public interest, gravitated toward a few central themes in the data set. The term 'president' stands out. Political decisions, implications, and potential reforms start at the presidential level. The intersection of '보험' (insurance) and '의료' (medical services) within the discourse highlights the public's interest in the accessibility and quality of health care.

Interestingly, words like '감사합니다' (thank you) and '좋다' (good) in this conversation imply a favorable response to some healthcare reform or initiative, indicating areas where the policy might resonate well with the public.

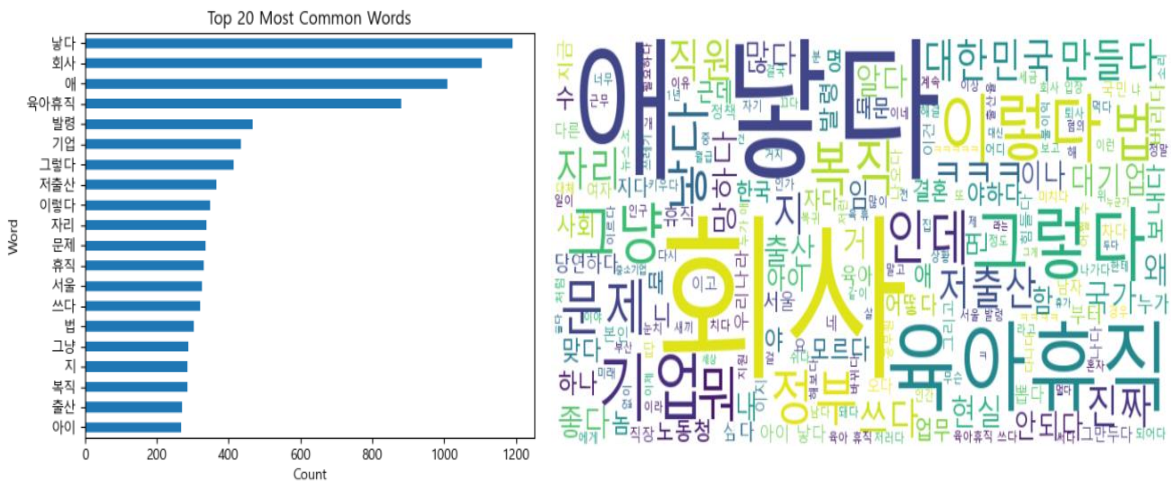
c. Maternity Policy



<fig 7. Word Cloud & bar chart for Maternity Policy >

Terms emerging in infertility policy datasets leverage social culture and policy-driven conversations. '낳다,' '출산,' and '아이' imply a narrative centered on childbirth and parenting. The frequent appearance of '정책' (policy) alongside '100 million' could suggest a financial incentive or support plan to address Korea's well-known low birth rate problem. The term' 결혼' (marriage) adds another layer of pointing to the discussion of the role of marriage in procreation discourse. Are policies framed in such a way as to encourage marriage as a precursor to procreation? Or is there a deeper cultural conversation going on in the commentary?

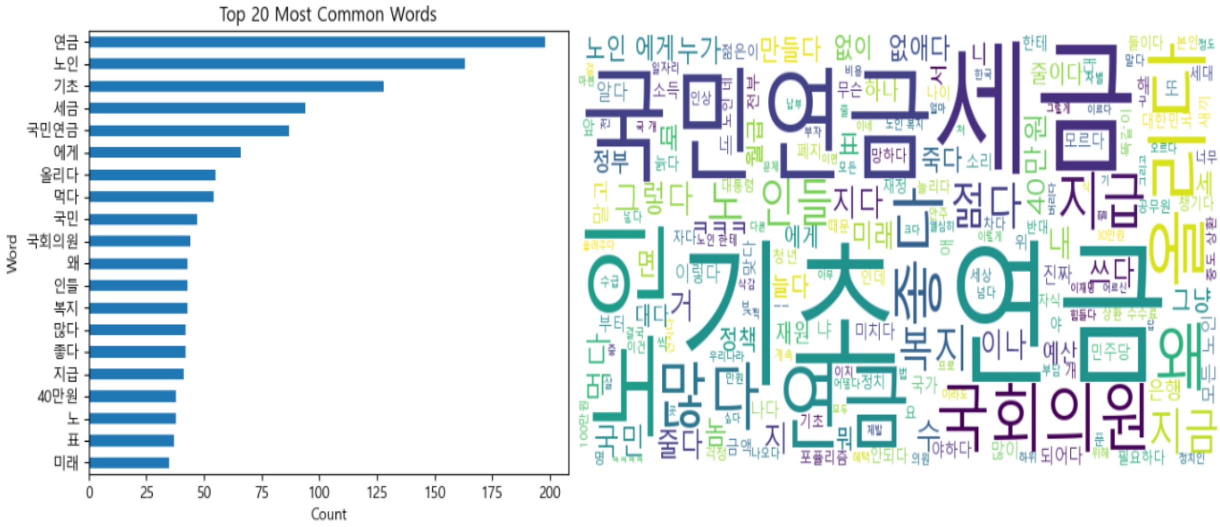
d. Parental leave



<fig 8. Word Cloud & bar chart for Parental leave>

The parental leave system, which is essentially connected to parental leave, directly refers to parental leave by putting terms such as '육아휴직' (Parental leave) and '휴직' (Leave) in the foreground. The prominence of '회사' (Company) indicates that the role of the corporate sector, or perhaps the response to these policies, is a central theme. Reappearing here, '저출산' (low birth) emphasizes the interconnectedness of these policy discussions, weaving narratives that encapsulate personal and professional realms.

e. Senior benefits

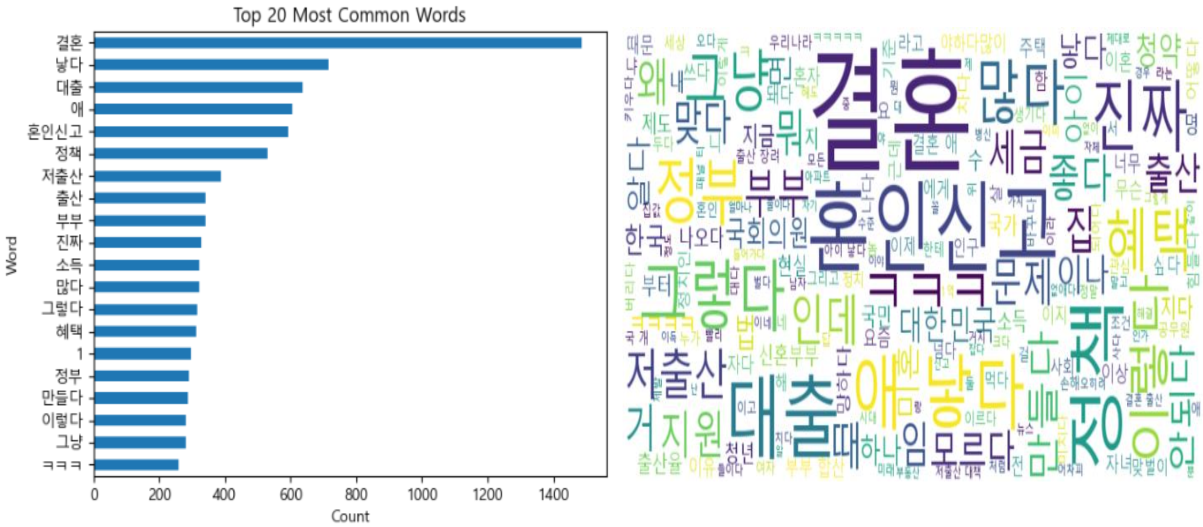


<fig 9. Word Cloud & bar chart for Senior benefits >

In the datasets related to welfare for the elderly, terms such as "연금" (pension) and "노인" (elderly) stand out. There is a clear emphasis on" 국민연금" (National Pension), implying concerns or discussions about the National Pension system.

Phrases such as '기초' may be associated with fundamental support for seniors and describe essential services and benefits. Interestingly, terms like '국회의원' (member of parliament) and "표" (vote) have a political connotation, implying a potential policy change or future decision. Specific references to "400,000 won" may refer to discussions surrounding certain monetary benefits or benefits.

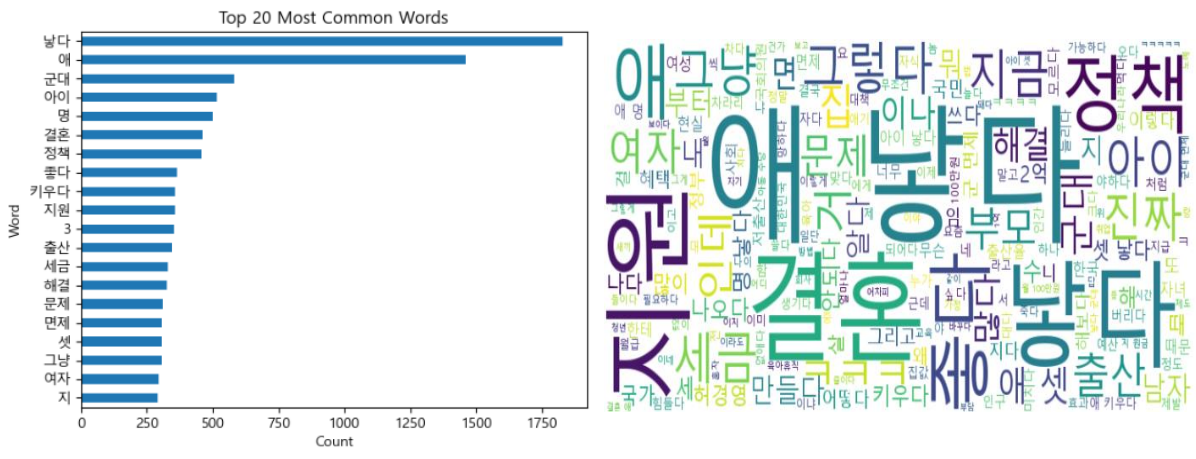
f. Loans for new couples



<fig 10. Word Cloud & bar chart for Loans for Newlyweds >

In the field of support for newlyweds, of course, '결혼' (marriage) stands out, followed by '대출' (loan) and 'marriage registration.' These terms emphasize the government's efforts to support young couples in the housing sector, as '집' (house) implies. Interestingly, '저출산' (low birth) reappears here, showing that marriage, housing, and fertility policies are intertwined. Terms like '소득' (income) and '혜택' (benefits) suggest an evaluation conversation about the financial thresholds or incentives associated with these loans. Recurring '문제' (problems) suggest potential issues or controversies related to this policy.

g. Child Care Assistance



<fig 11. Word Cloud & bar chart for Child Care Assistance >

The discourse surrounding childcare support is vividly expressed in terms such as '군대' (Army) and '키우다' (raise). It suggests stories about raising children and social expectations or roles related to children. '지원' (support) is evidence of the support provided to the family or the discussion surrounding support.

Also, terms like '면제' (exemption) can refer to certain privileges or exemptions available to families with multiple children or those who meet specific criteria. The frequency of terms such as '남자' (male) and '여자' (female) indicate gender discussions about the roles, responsibilities, or issues each gender faces in the context of raising children.

5.2 Public Sentiment Revealed: Sentiment Analysis Results

5.2.1 Sentiment Distribution: Deep

In the era of digital proliferation, public opinion is not confined to cafes and parlors. They permeate the vast expanse of the internet, empathizing with tweets, blogs, reviews, and comments. Like a digital footprint, these emotions provide a rich tapestry of human thoughts and feelings. Using sentiment analysis, we explore the core of these expressions and decipher the comments that make up the global or national psyche on various topics—the following results from classifying the collected comments into positive or negative comments using the KcBERT model.

a. Office Hours Policy

The positive response to the Office Hours Policy remained at only 5.8%, indicating low satisfaction with working hours. Responding to this response can be attributed to the industry embracing a "work from home" culture, allowing employees to schedule and provide a work-life balance. In addition, a plan to reduce the current 52 hours of work can be discussed. The national opposition and response to the Yoon Seok-yeol government's 69-hour working hour policy reflect public opinion on the working environment. The negative sentiment towards the Office Hours Policy was 94.2%, which is interpreted as a sign of great dissatisfaction.

These negative emotions stem from the current prevalence of chronic fatigue, reduced personal time, and encroachment on family time. This sentiment warns the industry to rethink operating hours and adopt a more holistic approach to employee well-being.

b. Medical Policy

In the healthcare sector, positive emotions are confirmed at 81%. The overwhelming majority of satisfaction with health care reflects the recent developments that our society has witnessed, and in particular, Moon Jae-in Care, which was introduced during the Moon Jae-in administration, has had a significant impact on this. A representative positive foundation could be contributing factors: improved medical research, AI integration in diagnostics, more comprehensive medical coverage, and more responsive emergency services. Negative emotions are a small but not negligible part, at around 19%. This negative sentiment can strongly affect areas with poor healthcare infrastructure, exorbitant medical bills, malpractice stories, and bureaucratic red tape delaying essential services.

c. Maternity and child support:

The positive response to childbirth and child support is low at 1.9%.
On the other hand, 98.1% of negative responses are analyzed as areas requiring immediate attention. Comprehensive maternal care, postnatal counseling, parenting support, and, where possible, educational support can help with this policy. In addition, high medical costs associated with childbirth, insufficient maternity leave, lack of childcare facilities, and possibly social pressures are pressing issues.

d. Parental leave:

The positive response to parental leave was 3.2%, showing little satisfaction. In order to improve this, it is necessary to create an environment in which parental leave can be implemented, such as an organizational culture that provides ample childcare benefits such as extended leave, childcare facilities, and flexible working hours. It is analyzed that the concern is very high.

In addition to the length of leave, factors leading to these outcomes may include underlying fears such as stigma over extended leave, concerns about job security, or the financial impact of unpaid leave. It is expected that these areas can be improved through the environment and national consensus in which the state can lead parental leave and through political, economic, social, and cultural support.

e. Senior Support:

10.7% of positive comments about senior support suggest that some districts or communities may be doing it right. These feelings can arise from areas with excellent geriatric care, communities that promote intergenerational interaction, or policies that provide financial and medical support to the elderly. However, with 89.3% of negative responses, supporting the elderly seems to be an urgent issue. When caring for the elderly, inadequate nursing homes, a lack of skilled aged care facilities, social apathy, and financial insecurity face numerous challenges for older people. Governments and society need to overcome these difficulties and approach the problem from a broader low birthrate system perspective.

f. Loans for Newlyweds:

Responses to loans for newlyweds are faintly positive at 1.4%. These may be the beneficiaries of a simple loan approval process, low-interest rates, or the counseling sessions some banks offer to new couples. However, negative emotions reached 98.6%, so improvement is needed. Strict documentation requirements, high collateral requirements, fluctuating interest rates, and a lack of awareness of available schemes can be several problems, and a response to address them is urgent.

5.2.2 Qualitative Insights: The Soul Beyond Numbers

All abilities of sentiment analysis on Korea's low birth rate policy based on the NLP dataset are accompanied by a sea of emotions. The anguish of a mother who cannot spend quality time with her child due to hard labor, the relief of a patient receiving treatment at a hospital, and the quiet despair of an abandoned old man, these are the stories created. Looking at the results of this sentiment analysis, the current low birth rate phenomenon in Korea is not a problem in one field, but all parts of society are closely related, and it should be seen as a result of organic relationships.

This characteristic of the sensibility of the low fertility policy further emphasizes the power of information. Knowledge is power, but lack of knowledge can lead to negative emotions. Many people need to be made aware of policies or support systems, which can lead to misinformation and frustration. When conducting sentiment analysis, it is also confirmed that there needs to be more information on the current policy or a negative perception due to a lack of accuracy. A robust information dissemination mechanism accessible to all levels of society is important. The government's intervention and efforts are essential to provide a channel for disseminating such information.

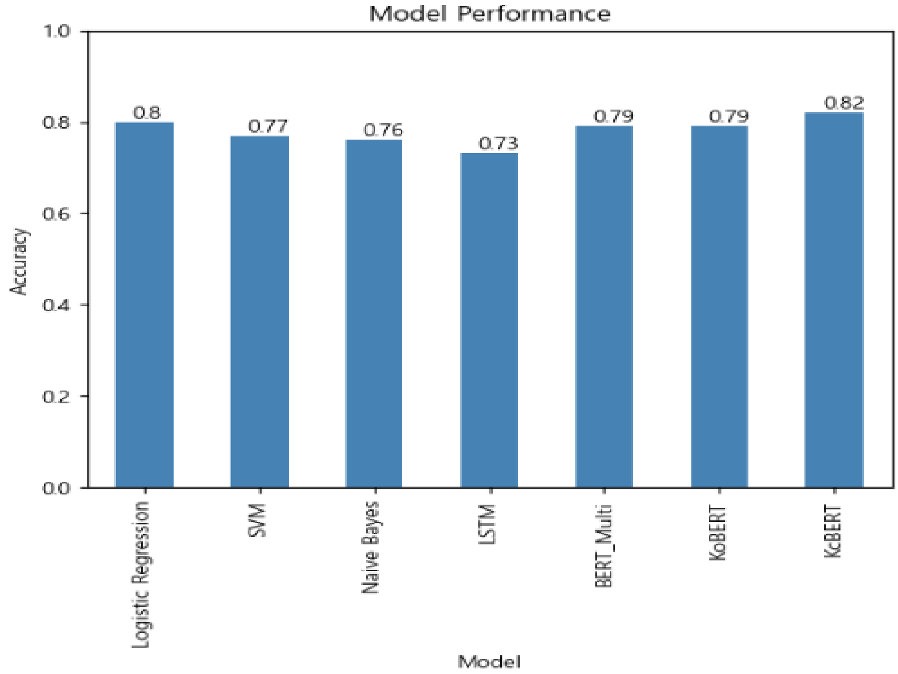
Culture is an important variable that influences people's perceptions and emotions. Culture, often the invisible hand, shapes perceptions. In a society that regards overwork as a mark of honor, negative feelings about working long hours are a clear call for cultural reflection. Likewise, in societies with strong family values, we can witness fervent cries for better parents and support for the elderly. Such cultural remarks of the people appearing in YouTube comments will be an essential criterion in determining the government's direction and focus on policy, and various channels must be prepared to confirm the public's perception through these cultural remarks.

As we weave this complex web of emotions, we find opinions, stories, hopes, dreams, and sometimes silent cries for change. When flowing in the right direction, these insights can be catalysts for transformative social change.

5.3 Performance Comparison of NLP Models for Sentiment Analysis of Korea's Low Fertility Policy: A Comprehensive Review

5.3.1 Overview of comparison training models

As we journey into the sophisticated realm of natural language processing (NLP), the myriad techniques available to decode emotions are a testament to the rapid progress in this field. Our primary goal in this study is to juxtapose deep learning paradigms with traditional machine learning methods while focusing on sentiment analysis of Korea's low birth rate policy. A selected dataset from YouTube comments provides the spontaneity and candor typical of public discourse.



<Fig 12. Comparison of Training Models>

An overview of model accuracy is introduced before looking at each model in detail. A comparison of model accuracies reveals varying performance levels across the seven models tested for sentiment analysis. Accuracy ranged from about 73% to 82%, highlighting the difference in accurately classifying emotions.

KcBERT leads the way with the highest accuracy of 82% and exhibits an excellent ability to capture complex text patterns, especially in language-specific contexts such as Korean.
BERT_Multi and KoBERT follow closely with 79% accuracy each, demonstrating the strength of transformer-based models in handling complex language tasks.
Naive Bayes was surprised with an impressive 75.9% accuracy. This is commendable for existing algorithms and indicates its adaptability to text classification tasks.
Logistic regression provides a respectable accuracy of 79.9%, highlighting its usefulness as a robust baseline model for binary classification tasks.
SVM has an accuracy of 76.8%, which is a testament to its ability to find optimal boundaries in high-dimensional data, even if slightly better than other models.
Despite its sequence modeling ability, LSTM gives an accuracy of 73.0%, which may suggest the need for model optimization or problems inherent to the data set.
Visually, the bar chart effectively shows the marginal difference in performance, with the

highest bar for KcBERT and the bars for BERT_Multi, KoBERT, and Naive Bayes highlighting competing accuracies.

a. The Traditional Algorithmic Model: Revisiting the Basic Skills

- Logistic regression:

<Table 3 : Classification Report – Logistics Regression>

	precision	recall	F1-score	Support
0	0.83	0.55	0.66	375
1	0.79	0.94	0.86	670
Accuracy			0.80	
Macro avg	0.81	0.74	0.76	1045
Weight avg	0.80	0.80	0.79	1045

The logistic regression model's achievement of 80% accuracy is a testament to the resilience and reliability of this veteran model. The balance of precision and recall inferred from the f1 score indicates balanced classification power. Although logistic regression may seem rudimentary to some, its performance highlights its ability to be a reliable criterion. However, we must recognize the potential downside of not being able to capture the more complex linguistic nuances embedded in YouTube comments. Laudable accuracy on negative sentiment means potentially harmful content can be quickly flagged on large data sets where rapid evaluation is essential. However, the model's decline in recall of negative emotions is a caveat. This means it can underperform in applications such as content alerts, where identifying all negative emotions is critical. The model's efficiency and computational lightness make it an invaluable tool for fast sentiment decisions on massive data sets.

- Support Vector Machine (SVM):

<Table 4 : Classification Report – Support Vector Machine>

	precision	recall	F1-score	Support
0	0.88	0.41	0.56	375
1	0.75	0.97	0.84	670

Accuracy			0.77	
Macro avg	0.81	0.69	0.70	1045
Weight avg	0.79	0.77	0.74	1045

SVM shows a remarkable strength in discriminating positive sentiment with an accuracy of 77%, especially the f1 score. What is interesting about SVMs is their inherent flexibility. The choice of kernel and hyperparameters can significantly impact performance, mainly when applied to multifaceted topics such as low birthrate policies. The guiding principle of maximizing margins between classes is adept at separating complex data clusters, especially when sentiment is polarizing. However, recall of negative emotions is a concern. Missing complaint detection in applications such as automated customer support can lead to significant customer complaints, indicating that relying solely on SVMs can be a gamble.

- Naive Bayes:

<Table 5 : Classification Report – Naïve Bayes>

	precision	recall	F1-score	Support
0	0.65	0.73	0.69	375
1	0.84	0.78	0.81	670
Accuracy			0.77	
Macro avg	0.74	0.75	0.75	1045
Weight avg	0.77	0.76	0.76	1045

With a respectable accuracy of 76%, Naive Bayes reminds us of the beauty of simplicity. Performance is good despite the default hypothesis of functional independence. However, there is a fundamental question. Can these assumptions be applied consistently for interconnected and context-rich YouTube comments? If emotion is deeply embedded within the linguistic nuances of a text, such as poetry or complex prose, Naive Bayes may not be the best choice.

b. Deep learning models: venture into advanced paradigms

- LSTM (Long Short Term Memory):

<Table 6 : Classification Report – LSTM>

	precision	recall	F1-score	Support
0	0.63	0.63	0.63	375
1	0.79	0.79	0.79	670
Accuracy			0.73	
Macro avg	0.71	0.71	0.71	1045
Weight avg	0.74	0.73	0.74	1045

The LSTM architecture showed the ability to learn over multiple iterations, reaching a final test accuracy of 73%. This indicates the ability to learn complex patterns and dependencies within the text when juxtaposed with an initial accuracy of 65.33%. LSTMs with memory cells are designed to process sequences in nature, making them suitable for annotation analysis. However, a closer look at the validation loss and accuracy metrics over different epochs reveals potential overfitting. The model may be too tuned to the training data, reducing its capacity to generalize to unseen data. Considering that this model is a deep learning model and is suitable for a much more complex and large data set than conventional machine learning models, the potential overfitting shows limitations in choosing an LSTM model.

c. BERT variant

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a Google invention tailored to understand the context of words in a sentence. Multilingual variants provide a uniform solution for multiple languages, but for specific languages such as Korean, custom models such as KoBERT and KcBERT have been trained. This study used a BERT multilingual variant model and three BERT variant models, KoBERT and KcBERT. These models' training dynamics and performance metrics provide insight into their efficiency and nuance. BERT_Multilingual Training Dynamics, KoBERT and KcBERT, say a lot about the consistency of BERT-based architectures with over 79% parallel performance. Details for each model are as follows.

- BERT_Multilingual

<Table 7 : Classification Report – BERT_Multilingual >

	precision	recall	F1-score	Support
0	0.69	0.72	0.71	375
1	0.84	0.82	0.83	670
Accuracy			0.79	
Macro avg	0.77	0.77	0.77	1045
Weight avg	0.79	0.79	0.79	1045

The BERT_Multilingual Training Dynamics model performance showed significant loss reduction during early epochs. It started at a loss of 0.6867 in the first epoch and experienced a sharp decline, with a loss of 0.1862 at the start of the second epoch. This behavior demonstrates the model's proficiency in quickly adapting to the complexity of the data set. However, the loss values oscillated as training continued, suggesting continuous model recalibration to fine-tune the emotion classification function. The eighth epoch witnessed a meager loss value of 0.0017, a testament to the model's profound understanding of the data set. However, subsequent epochs, especially the ninth, show an increase in loss, which may indicate overfitting. Early dismissal was introduced in response, and the final accuracy was 79%.

- KoBERT

<Table 8 : Classification Report – KoBERT >

	precision	recall	F1-score	Support
0	0.71	0.71	0.71	375
1	0.84	0.83	0.84	670
Accuracy			0.79	
Macro avg	0.77	0.77	0.77	1045
Weight avg	0.79	0.79	0.79	1045

Designed explicitly for Korean text, KoBERT consistently performed well throughout the learning process. In the early stages of training, the ability of the model to quickly adapt to the structure and nuances of the data set was evident. This may be due to the Korean-optimized architecture, which can help capture complex language patterns more effectively than typical BERT models. A significant decrease in loss values was observed in early epochs. This fast adaptability is a model that can benefit from prior knowledge and context by fine-tuning pre-trained weights to the Korean corpus. The stability of loss reduction, especially when compared to BERT_Multilingual, suggests a more focused and streamlined training trajectory. However, as with most deep-learning models, later epochs began to show minimal signs of oscillation in the loss curve. This behavior is not uncommon, especially when the model reaches an optimal state, and can sometimes be exceeded in gradient updates. A slight increase in the loss by the ninth epoch could be an early indicator of overfitting. This is where the model starts memorizing the training data rather than generalizing it. Implementing strategies such as early dismissal can be an effective countermeasure against such pitfalls. An accuracy of 0.79% is

It is not different from BERT_Multilingual Training Dynamics, but given the characteristics of a Korean-specific model, it is evidence of solid performance and the effectiveness of the applied measures.

- KcBERT

<Table 9 : Classification Report – Naïve Bayes>

	precision	recall	F1-score	Support
0	0.89	0.66	0.76	100
1	0.78	0.94	0.86	132
Accuracy			0.82	232
Macro avg	0.84	0.80	0.81	232
Weight avg	0.83	0.82	0.81	232

The training dynamics of KcBERT presented an exciting trajectory. The model starts with a loss of 0.3939 on the first batch of the first epoch and rises slightly to 0.4384 on the 200th batch. This initial performance suggests that the model understands the structure and complexity of the data set. In the second epoch, the model started with a loss close to the second half of the first epoch (0.4266) but managed a significant decrease, reducing the loss to 0.1895 by the end. The third epoch showed a much more pronounced improvement, with the loss dropping to 0.0937 but increasing to 0.2674. The trend of significant loss reduction continued, reaching the lowest values at the fifth and sixth epochs, where the loss decreased to values such as 0.0056 and 0.0013, respectively. However, spikes such as 0.3875 in the sixth epoch are placed between these troughs, demonstrating the oscillatory behavior of the loss value during training. Since the seventh epoch, the model has maintained low loss figures with intermittent slight spikes. At the 9th and 10th epochs, the loss values consistently remained below 0.004, indicating that the model achieved an optimal understanding of the data distribution.

Regarding classification performance, KcBERT provided an accuracy of 0.82. Precision, recall, and f1-score show that the model does an excellent job at detecting 'negative' sentiment (f1-score of 0.86), while it lags slightly behind at identifying 'positive' sentiment, with an f1-score of 0.76. Nevertheless, the accuracy of the 'positive' class is surprisingly high at 0.89, but the recall of 0.66 suggests that there are certain positive instances that the model missed.
In summary, KcBERT's training showed fast adaptability, scattered loss variability, and commendable performance in the second half. The classifier's high precision and slightly lower recall for the 'positive' class suggests that future tuning can focus on increasing the model's sensitivity to positive sentiment without compromising specificity.

5.4 Final Conclusion:

Selecting the ideal model for sentiment analysis requires an evaluation that combines training dynamics, adaptability, performance metrics, and specific use-case relevance.

a. Training mechanics:

- BERT_Multilingual : Experienced fluctuations in loss values, which may suggest potential overfitting in later epochs.
- KoBERT: More reliable than BERT_Multilingual, but hints at potential overfitting in later steps.
- KcBERT: shows rapid adaptability with significant loss reduction, especially in late epochs, indicating learning ability.
- LSTM: In general, LSTMs tend to learn sequential patterns well, but can be computationally intensive and struggle with long sequences unless combined with other mechanisms.
- SVM, Logistic Regression, and Naive Bayes: As traditional algorithms, these models may not capture complex patterns in text data as effectively as deep learning models, but they

can serve as solid baselines.

b. Performance metrics:

- BERT_Multilingual and KoBERT: Both rendered an accuracy of 0.79%. However, given the nature of deep learning, the computational cost may not justify the performance.
- KcBERT: achieved the highest accuracy of 0.82%. Balanced performance across classes (precision, recall and F1 score) stands out.
- LSTM: LSTMs can be robust, but without details they can generally perform well on sequence data, but can outperform transformer-based models such as BERT derivatives.
- SVM and Logistic Regression: Deep learning models may have provided significant accuracy, but they have demonstrated the ability to outperform these models on large and complex data sets.
- Naive Bayes: Generally good for text classification, but may make limited assumptions about feature independence.

c. Model specificity:

- KcBERT: Specialized KcBERT is more likely to capture nuances specific to a dataset than more generalized models such as BERT_Multilingual.
- LSTM: Provides sequential modeling, but lacks the depth and complexity of transformer architecture.
- SVM, Logistic Regression, and Naive Bayes: May not capture the contextual relevance of text data because they are non-sequential.

d. Computational considerations:

- Deep learning models (BERT variants, LSTMs): require significant computing resources, especially during training. However, the ability to process large data sets and capture complex patterns can justify the computational cost.
- Traditional models (SVM, Logistic Regression, Naive Bayes): Not computationally intensive and fast to learn. Excellent for setting baselines or when computational resources are limited.

youWhile traditional models such as LSTM and SVM, Logistic Regression and Naive Bayes have their advantages, KcBERT's consistent performance combined with its specialization provides the best combination of accuracy and model complexity for that dataset. This comprehensive analysis supports the decision to select KcBERT for the study.

Chapter 6. Discussion

6.1 Comparative analysis with existing work

In the rapidly evolving field of natural language processing (NLP), sentiment analysis has become an essential tool for interpreting public sentiment. This study set out to explore various models of traditional learning and deep learning to analyze YouTube comments on Korea's low birth rate policy. Comparative analysis with existing studies reveals interesting insights into how these models match or differ from existing studies.

KcBERT, which performed best in our study, achieved an accuracy of 82%, demonstrating the potential of a specialized model in language-specific contexts. These results are slightly different from the trends observed in the COVID-19 vaccine tweet analysis, where the BERT technique achieved 90.42% accuracy, but are consistent in terms of overall analysis (Jain et al., 2023). However, KcBERT's specialized Korean characteristics contribute to KcBERT's unique differentiation and outstanding performance. While these specializations highlight the value of language customization, they may also limit the applicability of the model to other languages, suggesting that more extensive exploration is warranted.

In this study, traditional machine learning models such as logistic regression, naive Bayes, and SVM (Support Vector Machine) have been established, and the 80% accuracy of logistic regression is elastic by reflecting the Moroccan dialect and Arabic Twitter comment analysis results. This reaffirmed its status as a reliable model (Errami et al., 2023). However, the focus of this study on a unique dataset related to Korea's low birth rate policy adds a new dimension, demonstrating the model's adaptability to complex topics.

The research performance of SVM, which achieved 77% accuracy, is consistent with previous results such as analysis of COVID-19 vaccine tweets, where SVM achieved the highest accuracy among machine learning models (Jain et al., 2023). However, this study further explores the model's flexibility and kernel choice, providing a more comprehensive understanding of its applicability. Naive Bayes demonstrates the effect of simplicity, consistent with previous work (Erami et al., 2023) on the Moroccan dialect with 76% accuracy. However, our study presents potential limitations by questioning the model's assumptions in context-rich YouTube comments, a perspective that has not been deeply explored in previous studies. The 73% accuracy of LSTM in our study highlights its ability to learn complex patterns, a result consistent with LSTM topology optimization work (Bataineh & Kaur, 2021). However, a detailed analysis of the validation loss and accuracy metrics reveals potential overfitting, providing a nuanced understanding of the limitations of LSTMs that may not have been fully addressed in previous studies.

A comparative analysis also sheds light on the broader implications of the study. By juxtaposing traditional machine learning methods and deep learning paradigms, and focusing on a unique dataset relevant to Korea's low birth rate policy, our study not only confirms the findings of previous studies, but also provides insights into model adaptability, training dynamics, overfitting, and domains. do. We anticipate that certain applications and new insights will be discovered through our research. This research makes an important contribution to the field through its alignment with global trends, its innovative approach to traditional models, and its nuanced understanding of the challenges and opportunities of sentiment analysis. By studying the effectiveness, strengths and limitations of these models, this study not only revealed popular sentiment on demographic issues in Korea, but also guided future scholars to wisely apply NLP tools to similar studies.

In conclusion, the comparative analysis conducted in this study provides a multifaceted view of the existing models and various deep learning models. By juxtaposing the results with existing research, we not only validated previous insights, but also uncovered a new level of understanding. The subtle insights derived from our study contribute to a broader understanding of sentiment analysis, particularly in the context of social issues such as low birth rate policy. The comparative analysis highlights the strengths and weaknesses of the various models, providing guidance for future researchers and practitioners in choosing the approach that best suits their specific needs. This study serves as a valuable addition to the growth of research on sentiment analysis by providing a comprehensive and in-depth analysis that is consistent with and expands on existing research.

6.2 Policy Implications

The power of NLP and sentiment analysis, demonstrated through various models, provides policy makers with a rich avenue to explore public sentiment. This is especially relevant to important issues such as Korea's low birth rate and related government policies. The complexity of public opinion on such topics is difficult to decipher by traditional methods and this is where the application of advanced NLP models such as KcBERT comes to the fore.

Understanding the nuances of public opinion provides a more in-depth look at the prevailing sentiments and emotions within a population. By accurately measuring these sentiments, policymakers can gain insight into not just what the public thinks, but why they think the way they do. This information is translated into the ability of models to process vast amounts of textual data to draw more comprehensive and accurate public opinion, along with the ability to adopt a data-driven approach to policy development. This alignment of policies with citizens' needs and concerns, harnessed through actionable insights, brings important changes to how policies are designed and evaluated.

Beyond simple policy formation, continuous monitoring of public sentiment through NLP can serve as a continuous feedback mechanism. Policymakers can track public reaction to low fertility policies over time, evaluate their effectiveness, and identify areas for improvement. This real-time feedback not only enables faster governance, but also empowers to adjust as sentiment and circumstances change.

The adaptability and success of models such as KcBERT extends the methodology explored in this study far beyond the realm of birth rate policy. The versatile nature of NLP in policy analysis allows these methods to be utilized in a wide range of policy domains, from healthcare to education to environmental policy. This flexibility and customization enables more nuanced and contextual insights, further enhancing public engagement by leveraging opinions expressed online.

By taking citizen voices into account, policy makers promote more participatory forms of governance that build trust and support democratic decision-making. This inclusivity is complemented by the collaboration opportunities that may emerge between various government agencies and research institutions. A cohesive approach to policy analysis and development is enhanced by cross-sector collaboration to increase the efficiency and effectiveness of governance.

In conclusion, the insights gained from the comparative analysis of NLP models have profound implications for policy makers. The ability to understand and interpret public sentiment on complex issues such as Korea's low birth rate not only enriches policy development, but fosters adaptability, inclusion and cooperation. The transformative and versatile nature of NLP in policy analysis extends its potential far beyond birth rate issues. Policymakers and researchers alike must continue to explore and leverage these innovative technologies to pave the way towards more responsive, evidence-based, and citizen-centric policies.

6.2 Challenges and Limitations

The results derived from this study based on a comparative analysis of various NLP models should be interpreted with caution due to the specificity of the YouTube dataset utilized. Although the insights are powerful in the context of this particular data, there are a series of issues and limitations that need to be carefully considered.

First, the need for cross-validation across different sources, platforms and topics is essential to confirm the generalizability of the model. The exclusive focus on YouTube comments may not capture the full spectrum of public sentiment, raising concerns about the broad applicability of the insights. As the success of the model in this study can be linked to the unique characteristics of the data, confirming that the model is equally effective across different data types and sources is an important step towards wider relevance.

Another important issue arises from the complexity of transformer-like models. Their complex architecture, while powerful, can hinder interpretability, which is essential for policy makers and researchers. Understanding why a model reached a particular conclusion is just as important as the conclusion itself. This is especially true when insights are meant to inform public policy. A delicate balance is needed here. Simpler models can provide more transparent reasoning but may not capture all the subtleties, while complex models can provide deeper insights but at the cost of understanding how these insights were arrived at. This balance between complexity, accuracy and interpretability adds another layer of complexity to the research findings.

Additionally, the computational requirements of models such as BERT and LSTM introduce constraints that can affect applications. These models require significant computational resources, which is a potential barrier in resource-constrained environments. These constraints can hinder widespread use of these powerful tools, especially in smaller organizations or underfunded research institutions, making the tools inaccessible to those who might benefit from them.

In conclusion, the comparative analysis provided in this discussion demonstrates both opportunities and challenges in the field of NLP as applied to policy analysis. Through the exploration of models such as KcBERT, LSTM and existing algorithms, this study highlights key considerations for future research and practical applications in public policy, from the need for cross-validation and thoughtful model selection to the careful balance between complexity and interpretability. emphasizes.

The success in decoding public sentiment of specialized NLP models such as KcBERT is promising, but there are caveats that need to be addressed. Reflection on challenges and limitations deepens our understanding, guiding researchers and policymakers alike to more responsibly and effectively utilize this powerful tool.

The implications of this study go far beyond South Korea's demographic challenges and shed light on a broader area of public policy analysis. By recognizing and addressing these challenges, the field can continue to innovate and explore by unlocking the full potential of NLP, mindful of its constraints and complexities. This thoughtful approach lays out a clear path for future exploration and application, highlighting the exciting potential and continuing relevance of NLP in the ever-evolving public policy landscape.

Chapter 7. Implications and Future Research

7.1 Beyond Research

a. Wider Policy Implications and Role of NLP Models

The ramifications of this study go far beyond the specific context of YouTube sentiment analysis, providing broader insight into the dynamic interface between NLP models and public policy. As governments and organizations around the world grapple with complex social problems, NLP models have emerged as important tools for deciphering public sentiment, preferences, and needs.

The success of models such as KcBERT in capturing the nuanced expression of emotion highlights the potential of NLP techniques to inform and enhance policy decisions. These models provide policy makers with a detailed understanding of public opinion, facilitating more responsive and inclusive policy development. By accurately capturing the multifaceted nature of public sentiment, policymakers can devise strategies that resonate with a broad range of constituents, strengthening trust and cooperation between government and citizens.

However, this potential is not without its problems. This study highlighted important considerations in model selection, including the delicate balance between complexity, interpretability and computational constraints. The impact on policy analysis is profound. This is because model selection is no longer a technical decision, but a complex interaction of factors that must be carefully evaluated.

Resource limitations, particularly in smaller organizations or underfunded research institutes, may hinder widespread application of more complex models. This raises broader questions about equity and access, potentially disproportionate our ability to utilize these powerful tools.

Additionally, the specificity of the data used in the study highlights the need for cross-validation and consideration of different data types and sources. Ensuring that NLP models are effective in a variety of contexts is essential for their broad applicability to public policy and adds another layer of complexity to model selection and implementation.

7.2 Outlook

a. Future Directions of NLP-Based Public Policy Research and Emerging Models

As the landscape of NLP continues to evolve, new opportunities and challenges arise when applying NLP to public policy research. The possibilities of NLP models go beyond simple sentiment analysis to include a range of tasks that can inform and shape public policy decisions.

New models and technologies are pushing the boundaries of what is possible for understanding complex social dynamics. For example, transformer-based models and reinforcement learning provide new ways to explore public opinion, policy effects, and decision-making processes. These advances hold great promise for creating more informed, adaptive and resilient policy frameworks.

At the same time, the advent of interpretable machine learning provides a pathway to bridge the gap between complexity and transparency. Interpretability is becoming central to responsible AI deployment, especially in public policy where it is important to understand why models reach certain conclusions.

Future research must also face ethical considerations including issues of bias, privacy and liability. Potential bias in data and algorithms represents a critical issue that requires constant vigilance to ensure that models accurately represent different points of view and do not inadvertently reinforce existing inequalities.

Additionally, collaboration between researchers, policy makers, data scientists and the public is essential to leveraging NLP for policy analysis. An interdisciplinary approach fosters a richer understanding of societal needs, enabling more innovative and empathetic policy solutions.

The road ahead is full of potential, but requires careful exploration. By embracing complexity, prioritizing interpretability, and building interdisciplinary collaborations, the future of NLP-based public policy research can flourish and contribute to more responsive, inclusive and effective governance. This research laid the groundwork, but the journey did not end with continued exploration, innovation and application that paved the way towards a more nuanced and dynamic understanding of public sentiment and policy.

References :

ABRAMS, M. P., PELULLO, A. P., MEISEL, Z. F., MERCHANT, R. M., PURTLE, J. & AGARWAL, A. K. 2023. State and Federal Legislators' Responses on Social Media to the Mental Health and Burnout of Health Care Workers Throughout the COVID-19 Pandemic: Natural Language Processing and Sentiment Analysis. *JMIR infodemiology,* 3**,** e38676.

AL-GARAADY, J. & MAHYOOB, M. 2022. Public sentiment analysis in social media on the SARS-CoV-2 vaccination using VADER lexicon polarity. *Humanities and Educational Sciences Journal***,** 591-609.

ALI, M. M. 2021. Arabic sentiment analysis about online learning to mitigate covid-19. *Journal of Intelligent Systems,* 30**,** 524-540.

ALI, M. N. Y., SAROWAR, M. G., RAHMAN, M. L., CHAKI, J., DEY, N. & TAVARES, J. M. R. 2019. Adam deep learning with SOM for human sentiment classification. *International Journal of Ambient Computing and Intelligence (IJACI),* 10**,** 92-116.

ALIZARGAR, A., CHANG, Y.-L. & TAN, T.-H. 2023a. Performance comparison of machine learning approaches on Hepatitis C prediction employing data mining techniques. *Bioengineering,* 10**,** 481.

ALONSO DEL BARRIO, D. & GATICA-PEREZ, D. Examining European Press Coverage of the Covid-19 No-Vax Movement: An NLP Framework. Proceedings of the 2nd ACM International Workshop on Multimedia AI against Disinformation, 2023. 52-59.

ALSAYAT, A. 2022. Improving sentiment analysis for social media applications using an ensemble deep learning language model. *Arabian Journal for Science and Engineering,* 47**,** 2499-2511.

ALTAMEEMI, Y. & ALTAMIMI, M. 2023. Thematic Analysis: A Corpus-Based Method for Understanding Themes/Topics of a Corpus through a Classification Process Using Long Short-Term Memory (LSTM). *Applied Sciences,* 13**,** 3308.

ANDERSON, J. E., MOYER, J. & CHICHIRAU, G. 2022. *Public policymaking,* Cengage Learning.

ANDO, A. & MODIGLIANI, F. 1963. The" life cycle" hypothesis of saving: Aggregate implications and tests. *The American economic review,* 53**,** 55-84.

ASUDANI, D. S., NAGWANI, N. K. & SINGH, P. 2023. Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review***,** 1-81.

BADAWI, S. S. 2023. Using Multilingual Bidirectional Encoder Representations from Transformers on Medical Corpus for Kurdish Text Classification. *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY,* 11**,** 10-15.

BAHRI, C. A. & SUADAA, L. H. 2023. Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems),* 17.

BAHRI, C. A. & SUADAA, L. H. 2023a. Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems),* 17.

BANIMUSTAFA, A., BAKLIZI, M. & KHATATNEH, K. 2022a. Machine learning for securing traffic in computer networks. *International Journal of Advanced Computer Science and Applications,* 13.

BARONE, E. & BARONE, G. 2022. Unscrambling Codes: From Hieroglyphs to Market News. *International Journal on Natural Language Computing (IJNLC) Vol,* 11.

BASTIAN, A., MARDIANA, A. & WULANSARI, D. S. 2023. Sentiment Analysis Against Political Figure's Billboard During Pandemic Using Naïve Bayes Algorithm. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* 7**,** 138-145.

BATCHA, B. B. C., SINGARAVELU, R., RAMACHANDRAN, M., MUTHUSAMY, S., PANCHAL, H., THANGARAJ, K. & RAVINDARAN, A. 2023. A novel security algorithm RPBB31 for securing the social media analyzed data using machine learning algorithms. *Wireless Personal Communications***,** 1-28.

BUHAMRA, S. S., ALMUTAIRI, A. N., BUHAMRAH, A. K., ALMADANI, S. H. & ALIBRAHIM, Y. A. 2022. An NLP tool for data extraction from electronic health records: COVID-19 mortalities and comorbidities. *Frontiers in Public Health,* 10**,** 1070870.

BURGESS, J. & GREEN, J. 2018. *YouTube: Online video and participatory culture,* John Wiley & Sons.

BUST, L., WHYLE, E. & OLIVIER, J. 2023. A discourse and content analysis of representation in the mainstream media of the South African National Health Insurance policy from 2011 to 2019. *BMC Public Health,* 23**,** 1-19.

CARUANA, R., LOU, Y., GEHRKE, J., KOCH, P., STURM, M. & ELHADAD, N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015. 1721-1730.

CEBRAL-LOUREDA, M., HERNãNDEZ-BAQUEIRO, A. & TAMëS-MUóOZ, E. 2023. A text mining analysis of human flourishing on Twitter. *Scientific Reports,* 13**,** 3403.

CHE, S., WANG, X., ZHANG, S. & KIM, J. H. 2023. Effect of daily new cases of COVID-19 on public sentiment and concern: Deep learning-based sentiment classification and semantic network analysis. *Journal of Public Health***,** 1-20.

CHO, D., LEE, H. & KANG, S. 2021. An empirical study of Korean sentence representation with various tokenizations. *Electronics,* 10**,** 845.

CHO, S., MOON, J., BAE, J., KANG, J. & LEE, S. 2023. A Framework for Understanding Unstructured Financial Documents Using RPA and Multimodal Approach. *Electronics,* 12**,** 939.

CHOU, C. & CHU, T. 2022. An Analysis of BERT (NLP) for Assisted Subject Indexing for Project Gutenberg. *Cataloging & Classification Quarterly,* 60**,** 807-835.

CINELLI, M., QUATTROCIOCCHI, W., GALEAZZI, A., VALENSISE, C. M., BRUGNOLI, E., SCHMIDT, A. L., ZOLA, P., ZOLLO, F. & SCALA, A. 2020. The COVID-19 social media infodemic. *Scientific reports,* 10**,** 1-10.

DAFAALLA, H., ABAKER, M., ABDELMABOUD, A., ALGHOBIRI, M., ABDELMOTLAB, A., AHMAD, N., ELDAW, H. & HASABELRSOUL, A. 2022a. Deep Learning Model for Selecting Suitable Requirements Elicitation Techniques. *Applied Sciences,* 12**,** 9060.

DAI, Y., CHENG, X. & LIU, Y. 2023. Information Alienation and Circle Fracture: Policy Communication and Opinion-Generating Networks on Social Media in China from the Perspective of COVID-19 Policy. *Systems,* 11**,** 340.

DARAD, S. & KRISHNAN, S. 2023. Sentimental analysis of COVID-19 twitter data using deep learning and machine learning models. *Ingenius. Revista de Ciencia y Tecnología***,** 108-117.

DEVLIN, J., CHANG, M.-W., LEE, K. & TOUTANOVA, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

DIBATTISTA, I. 2022. Building a collective ecological identity: A multidisciplinary reflection. *European Journal of Cultural Management and Policy,* 12**,** 11106.

DOSHI-VELEZ, F. & KIM, B. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

EOM, G., YUN, S. & BYEON, H. 2022. Predicting the sentiment of South Korean Twitter users toward vaccination after the emergence of COVID-19 Omicron variant using deep learning-

based natural language processing. *Frontiers in Medicine,* 9**,** 948917.
ERFINA, A. & NURUL, M. R. 2023. Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. *Data & Metadata,* 2**,** 45-45.
GAMAL, D., ALFONSE, M., M. EL-HORBATY, E.-S. & M. SALEM, A.-B. 2018. Analysis of machine learning algorithms for opinion mining in different domains. *Machine Learning and Knowledge Extraction,* 1**,** 224-234.
GEORGE, L. & SUMATHY, P. 2023. An integrated clustering and BERT framework for improved topic modeling. *International Journal of Information Technology***,** 1-9.
GHORBANI, M., BAHAGHIGHAT, M., XIN, Q. & ÖZEN, F. 2020. ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing. *Journal of Cloud Computing,* 9**,** 1-12.
GOLDBERG, Y. 2022. *Neural network methods for natural language processing*, Springer Nature.
GONG, X., YING, W., ZHONG, S. & GONG, S. 2022. Text sentiment analysis based on transformer and augmentation. *Frontiers in Psychology,* 13**,** 906061.
GREFENSTETTE, G. 1999. Tokenization in Syntactic Wordclass Tagging", Hans van Halteren. Dordrecht, Kluwer Academic Publishers.
GROáE, C. 2022. Towards a Holistic Perspective on Future Transportation Systems: A Swedish Case and a Conceptual Framework. *Future Transportation,* 2**,** 846-867.
HAN, P., LI, X., WANG, X., WANG, S., GAO, C. & CHEN, W. 2022. Exploring the effects of drug, disease, and protein dependencies on biomedical named entity recognition: A comparative analysis. *Frontiers in Pharmacology,* 13**,** 1020759.
HENDRIYADI, I., PUTRI, A. F., RAHMAWATI, R. & SAPUTRA, D. D. 2023. Sentiment Analysis of Rising Fuel Prices on Social Media Twitter using the Naïve Algorithm Bayes Classifiers And AdaBoost. *Informatics and Software Engineering,* 1**,** 14-23.
HERDIYANI, T. C. & ZAILANI, A. U. 2022. Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia. *Jurnal Teknologi Sistem Informasi,* 3**,** 154-165.
HOSSAIN, E., SHARIF, O. & MOSHIUL HOQUE, M. 2021. Sentiment polarity detection on bengali book reviews using multinomial naive bayes. *Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2020.* Springer.
HOWARD, P. N. & HUSSAIN, M. M. 2011. The role of digital media. *J. Democracy,* 22**,** 35.
INDER, M. D. 2022a. Sentiment Analysis Techniques in Recent Works Using GRA Methodology. *REST Journal on Banking, Accounting and Business***,** 1(3)57-63.
JAIN, T., VERMA, V. K., SHARMA, A. K., SAINI, B., PUROHIT, N., MAHDIN, H., AHMAD, M., DARMAN, R., SU-CHENG, H. & SHAHARUDIN, S. M. 2023. Sentiment Analysis on COVID-19 Vaccine Tweets using Machine Learning and Deep Learning Algorithms. *International Journal of Advanced Computer Science and Applications,* 14.
JARA-AMëZAGA, C. 2023. The Impact of YouTube in Tourism Destinations: A Methodological Proposal to Qualitatively Measure Image Positioning—Case: Saudi Arabia. *Sustainability,* 15**,** 9879.
JEON, T. 2022. A linguistic study on tokenization methods for Korean text. *Language Facts and Perspectives,* 55**,** 309-354.
KARAYIĞIT, H., AKDAGLI, A. & ACı, Ç. İ. 2022. BERT-based Transfer Learning Model for COVID-19 Sentiment Analysis on Turkish Instagram Comments. *Information Technology and Control,* 51**,** 409-428.
KASTRATI, Z., DALIPI, F., IMRAN, A. S., PIREVA NUCI, K. & WANI, M. A. 2021. Sentiment analysis of students’ feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences,* 11**,** 3986.
KAUR, G. & SHARMA, A. 2023. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data,* 10**,** 5.
KIM, Y. 2019. An Analysis of the Newspaper Discourse on Issue of the Low Fertility. *Civic Humanities***,** 44.
KOTHURI, S. R. & RAJALAKSHMI, N. 2022. MALO-LSTM: Multimodal Sentiment Analysis Using Modified Ant Lion Optimization with Long Short Term Memory Network.

International Journal of Intelligent Engineering & Systems, 15.

KOTHURI, S. R. & RAJALAKSHMI, N. 2022a. MALO-LSTM: Multimodal Sentiment Analysis Using Modified Ant Lion Optimization with Long Short Term Memory Network. *International Journal of Intelligent Engineering & Systems,* 15.

KYAW, K. S., TEPSONGKROH, P., THONGKAMKAEW, C. & SASHA, F. 2023. Business Intelligent Framework Using Sentiment Analysis for Smart Digital Marketing in the E-Commerce Era. *Asia Social Issues,* 16**,** e252965-e252965.

KYUNG-SOO CHOI, KI-SEOK HONG, HYUNG-PYO MOON, IN-SEOK SHIN, JIN-HEE HAN & KYUNG-MOOK LIM, E. A. 2003. Economic Impact of Population Aging and Countermeasures (I).

LASRI, I., RIADSOLH, A. & ELBELKACEMI, M. 2023. Self-Attention-Based Bi-LSTM Model for Sentiment Analysis on Tweets about Distance Learning in Higher Education. *International Journal of Emerging Technologies in Learning,* 18.

LEE, J. 2020. *Proceedings of the 32nd Annual Conference on Human and Cognitive Language Technology ; KcBERT: Korean Comments BERT.*

LEE, S.-K. & KIM, J.-H. SENER: Sentiment Element Named Entity Recognition for Aspect-Based Sentiment Analysis. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023. IEEE, 1-5.

LIDDY, E. D. 2001. Natural language processing.

LIU, L., PERLAVICIUTE, G. & SQUINTANI, L. 2022. Opposing out loud versus supporting in silence: who wants to participate in decision-making about energy projects? *Environmental Research Letters,* 17**,** 114053.

M, I. Y. 2023. Text based Tweet Classification using Ensemble Classifier. *Journal of Trends in Computer Science and Smart Technology***,** 5(2), 136-145.

MAH, P. 2022b. Analysis of artificial intelligence and natural language processing significance as expert systems support for e-health using pre-train deep learning models. *Acadlore Trans. Mach. Learn,* 1**,** 68-80.

MAN-HWI, H., PARK, S.-C., LEE, H., JONG-HUM, Y. & SANG-GOO, L. 2015. Investigation of the current state of natural language processing technology in Korean language resources. *Journal of the Korean Society of Information Science and Technology***,** 681-683.

MAN-HWI, H., PARK, S.-C., LEE, H., JONG-HUM, Y. & SANG-GOO, L. 2015a. Investigation of the current state of natural language processing technology in Korean language resources. *Journal of the Korean Society of Information Science and Technology***,** 681-683.

MAREE, M., ELEYAT, M., RABAYAH, S. & BELKHATIR, M. 2023a. A hybrid composite features based sentence level sentiment analyzer.

MENSAH, S., SUN, K. & ALETRAS, N. 2021. An empirical study on leveraging position embeddings for target-oriented opinion words extraction. *arXiv preprint arXiv:2109.01238*.

MONIKA, P., KULKARNI, C., HARISH KUMAR, N., SHRUTHI, S. AND VANI, V. 2022. *Machine learning approaches for sentiment analysis: A survey", International journal of health sciences ,.*

MUSLEH, D. A., ALKHWAJA, I., ALKHWAJA, A., ALGHAMDI, M., ABAHUSSAIN, H., ALFAWAZ, F., MIN-ALLAH, N. & ABDULQADER, M. M. 2023. Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation. *Big Data and Cognitive Computing,* 7**,** 127.

MUSLEH, D. A., ALKHWAJA, I., ALKHWAJA, A., ALGHAMDI, M., ABAHUSSAIN, H., ALFAWAZ, F., MIN-ALLAH, N. & ABDULQADER, M. M. 2023d. Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation. *Big Data and Cognitive Computing,* 7**,** 127.

NG, Q. X., LEE, D. Y. X., NG, C. X., YAU, C. E., LIM, Y. L. & LIEW, T. M. 2023. Examining the negative sentiments related to influenza vaccination from 2017 to 2022: An unsupervised deep learning analysis of 261,613 Twitter posts. *Vaccines,* 11**,** 1018.

NG, Q. X., LEE, D. Y. X., NG, C. X., YAU, C. E., LIM, Y. L. & LIEW, T. M. 2023a. Examining the negative sentiments related to influenza vaccination from 2017 to 2022: An unsupervised deep learning analysis of 261,613 Twitter posts. *Vaccines,* 11**,** 1018.

NGUYEN, L. G. & WATABE, K. Flow-based network intrusion detection based on BERT masked language model. Proceedings of the 3rd International CoNEXT Student Workshop, 2022. 7-8.

OH, M. & TAK, H. 2022. Meta-evaluation through categorization of budgetary programs in a low fertility and aging society: A diachronic approach to the performance evaluation of the 1st~3rd basic plan of low fertility and aging society. *KOREAN SOCIETY AND PUBLIC ADMINISTRACTION*, 33**,** 211: 189-215.

ÖZéIFT, A., AKARSU, K., YUMUK, F. & SøYLEMEZ, C. 2021. Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije,* 62**,** 226-238.

PROTTASHA, N. J., SAMI, A. A., KOWSHER, M., MURAD, S. A., BAIRAGI, A. K., MASUD, M. & BAZ, M. 2022. Transfer learning for sentiment analysis using BERT based supervised fine-tuning. *Sensors*, 22**,** 4157.

PURBA, M. & YADI, Y. 2023. Implementation Opinion Mining For Extraction Of Opinion Learning In University. *Sinkron: jurnal dan penelitian teknik informatika,* 8**,** 694-699.

RAJARAMAN, A. & ULLMAN, J. D. 2011. *Mining of massive datasets*, Cambridge University Press.

RAZA, H., FAIZAN, M., HAMZA, A., AHMED, M. & AKHTAR, N. 2019. Scientific text sentiment analysis using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 10.

ROH, B.-R. & YANG, K.-E. 2019. Text Mining Analysis of South Korea's Birth-rate Decline Issue in Newspaper Articles **-** Transition Patterns over 18 Years. *Korean Journal of Social Welfare,* 71**,** 155 : 153-175.

RUDER, S. 2019. *Neural transfer learning for natural language processing.* NUI Galway.

SADANAND, V. S., GURUVYAS, K. R. R., PATIL, P. P., ACHARYA, J. J. & SURYAKANTH, S. G. 2022. An automated essay evaluation system using natural language processing and sentiment analysis. *International Journal of Electrical & Computer Engineering (2088-8708),* 12.

SADDAM, M. A., DEWANTARA, E. K. & SOLICHIN, A. 2023. Sentiment Analysis of Flood Disaster Management in Jakarta on Twitter Using Support Vector Machines. *Sinkron: jurnal dan penelitian teknik informatika,* 8**,** 470-479.

SAHU, S. K., MOKHADE, A. & BOKDE, N. D. 2023b. An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. *Applied Sciences*, 13**,** 1956.

SAJWAN, V., AWASTHI, M., GOEL, A. & SHARMA, P. 2023. Sentiment analysis of Twitter data regarding the agnipath scheme of the defense forces. *Indonesian Journal of Electrical Engineering and Computer Science,* 30**,** 1643.

SAM-SIK LEE, I.-C. S., NAM-HOON CHO, HEE-KYUNG KIM, YUN-SUN JEONG, EUN-YOUNG CHOI, ET AL. 2005. Research on causes of low fertility and comprehensive measures. Korea Institute for Health and Social Affairs, 2005.

SANG-HUN, C. 2006. Choi Soon-sil, at Center of Political Scandal in South Korea, Is Jailed. *The New York Times, 3*.

SANG HYUN, H. A. & TAE HYUP, R. O. H. 2020. Sentiment Analysis for Public Opinion in the Social Network Service. *The Journal of the Convergence on Culture Technology (JCCT),* 6**,** 111-120.

SANGAH, L., HANSOL, J., YUNMEE, B., SUZI, P. & HYOPIL, S. 2020. A Small-Scale Korean-Specific BERT Language Model. *Journal of KIISE,* 47**,** 682-692.

SARI, A. W., HERMANTO, T. I. & DEFRIANI, M. 2023. Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine. *Sinkron: jurnal dan penelitian teknik informatika,* 8**,** 1366-1378.

SARIKAYA, A. & ASLAN, S. 2022. Deep Learning and Machine Learning Based Sentiment Analysis on BitCoin (BTC) Price Prediction. *NATURENGS,* 3**,** 1-17.

SATHYA, A. & MYTHILI, M. 2023. Evaluating Sentiment Classification to Specify Polarity by Lexicon-Based and Machine Learning Approaches for COVID-19 Twitter Data Sets. *JOURNAL OF ADVANCED APPLIED SCIENTIFIC RESEARCH,* 5**,** 12-27.

SCHøLKOPF, B. & SMOLA, A. J. 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press.

SEO HYUN-JAE, K. J.-M. A. K. S.-S. 2022. Korean parts of speech tagging using Head-Tail tokenization technique. *Smart Media Journal 2022 Vol.,* 11**,** 17-25.

SEONG, M. 2009. Impacts and Implications of Low Fertility and Population Population on the Economy. *Seoul: National Assembly Budget Office***,** 27.

SEONG, M. 2009a. Impacts and Implications of Low Fertility and Population Population on the Economy. *Seoul: National Assembly Budget Office***,** 28.

SETYAWAN, H., AZIZAH, L. M. R. & PRADANI, A. Y. 2023. Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine. *Emerging Information Science and Technology,* 4**,** 1-7.

SEYOUNG, H., CHO-WON, K., ANYONG, C. & SAE BOM, L. 2022. Sentiment Analysis on 'Non-maritalism Childbirth' Using Naver News Comments. *JOURNAL OF THE KOREA CONTENTS ASSOCIATION,* 22**,** 74-85.

SHANMUGAVADIVEL, K., SATHISHKUMAR, V., RAJA, S., LINGAIAH, T. B., NEELAKANDAN, S. & SUBRAMANIAN, M. 2022a. Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports,* 12**,** 21557.

SHIN, G.-H. 2022. Automatic analysis of caregiver input and child production: Insight into corpus-based research on child language development in Korean. *Korean Linguistics,* 18**,** 125-158.

SHIN, G.-H. & JUNG, B. K. 2021. Automatic analysis of passive constructions in Korean: Written production by Mandarin-speaking learners of Korean. *International Journal of Learner Corpus Research,* 7**,** 53-82.

SIREGAR, D., LADAYYA, F., ALBAQI, N. Z. & WARDANA, B. M. 2023. Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter. *Jurnal Statistika dan Aplikasinya,* 7**,** 93-104.

SKTELECOM. 2019. *KoBERT* [Online]. https://github.com/SKTBrain/KoBERT. [Accessed].

SO-YOUNG, L. & SHIN-HWI, O. 2022. Low birth rate and aging society policy monitoring and tasks. Korea Institute for Health and Social Affairs.

STATISTA 2021. YouTube users in the world. Accessed [07/08/2023 ed.

TAN, J. Y., CHOW, A. S. K. & TAN, C. W. 2022. A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS OF GAME REVIEWS.

TAN, M. & LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning, 2019. PMLR, 6105-6114.

THAKKAR, K. & JAGDISHBHAI, N. 2023. Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing. *Journal of Management Research and Analysis,* 10**,** 18-20.

THEODOSIOU, A. A. & READ, R. C. 2023a. Artificial Intelligence, Machine Learning and Deep Learning: Potential Resources for the Infection Clinician. *Journal of Infection*.

TYAGI, N. & BHUSHAN, B. 2023. Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions. *Wireless Personal Communications,* 130**,** 857-908.

VEDANTAM, V. K. 2021. the survey: Advances in natural language processing using deep learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT),* 12**,** 1035-1040.

VINOD, D. N. & PRABAHARAN, S. 2023a. COVID-19-The Role of Artificial Intelligence, Machine Learning, and Deep Learning: A Newfangled. *Archives of Computational Methods in Engineering***,** 1-16.

WAWORUNDENG, J. M. S., SANDAG, G. A., SAHULATA, R. A. & RELLELY, G. D. 2022a. Sentiment Analysis of Online Lectures Tweets using Naïve Bayes Classifier. *CogITo Smart Journal,* 8**,** 371-384.

WU, Y., SCHUSTER, M., CHEN, Z., LE, Q. V., NOROUZI, M., MACHEREY, W., KRIKUN, M., CAO, Y., GAO, Q. & MACHEREY, K. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

YANG, C. & HUANG, C. 2023. Natural Language Processing (NLP) in Aviation Safety: Systematic

Review of Research and Outlook into the Future. *Aerospace*, 10**,** 600.
YANG, H. 2022. Network public opinion risk prediction and judgment based on deep learning: A model of text sentiment analysis. *Computational Intelligence and Neuroscience,* 2022.
YANG, R. 2023. Current Situation and Frontier Evolution of Online Public Opinion on Public Emergencies: Knowledge Graph Analysis Based on Citespace. *Advances in Engineering Technology Research,* 5**,** 332-332.
YANG, X. & ZHANG, Z. 2023. Agenda Setting in Media Coverage of Natural Disaster Exogenous Events: The 9.5 Sichuan Ganzi Luding Earthquake as an Example. *Journal of Humanities and Social Sciences Studies,* 5**,** 35-39.
YEONJI, J., JISEON, C. & HANSAEM, K. 2022. KcBert-based Movie Review Corpus Emotion Analysis Using Emotion Vocabulary Dictionary. *Journal of KIISE,* 49**,** 608-616.
YOUNGMI, K. 2023. Evaluation and Development Direction of Low Fertility and Aging Society Policies. *Health and Welfare Forum 2023 Vol,* 2023**,** 3 : 2-5.
ZHANG, H. 2022. Model Comparison in Sentiment Analysis: A Case Study of Amazon Product Reviews. *Highlights in Science, Engineering and Technology,* 16**,** 23-31.
ZHOU, J., GUO, X., DUAN, L., YAO, Y., SHANG, Y., WANG, Y. & XING, B. 2022a. Moving toward a standardized diagnostic statement of pituitary adenoma using an information extraction model: a real-world study based on electronic medical records. *BMC Medical Informatics and Decision Making,* 22**,** 319.
ZHOU, X., ZHANG, X., LARSON, H. J., DE FIGUEIREDO, A., JIT, M., FODEH, S., VERMUND, S. H., ZANG, S., LIN, L. & HOU, Z. 2022. Global spatiotemporal trends and determinants of COVID-19 vaccine acceptance on Twitter: a multilingual deep learning study in 135 countries and territories. *medRxiv***,** 2022.11. 14.22282300.