Belieview

NLP based blog viral marketing detection service

19102083 Hanseung Kim
20102122 Hyoan Jeong

서울과학기술대학교
SEOULTECH SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

Table of Contents

1. Motivation
2. Related Work
3. Collecting Data
4. Belieview
5. Conclusion & Limitation

Motivation

- **78%** of domestic consumers **base their purchasing decisions on reviews** from other buyers

- Of the viral marketing that consumers see most, **stealth marketing** that doesn't identify itself as advertising is **on the rise** with 21,037 cases caught by the Korea Fair Trade Commission in 2022

- Recently, **consumer distrust has deepened due to social media posts that focus on buzz marketing**. This can negatively impact consumers' ability to make rational purchasing decisions

- But **these reviews** use sophisticated techniques that **are clearly illegal but difficult to identify**

Motivation

According to the Korea Fair Trade Commission's 『Guidelines for examination on labeling and advertising of recommendation, guarantee, etc ("The Guidelines")』, advertising post must adhere to the following three principles:

(1) Disclosures indicating economic interests (hereinafter referred to as 'advertisement phrase') must be placed at the beginning or end of each post so that consumers can easily find them. → *proper disclosure placement*

If the advertisement phrase is written in the middle of the text without distinction, making it difficult to recognize
If the advertisement phrase is written in a comment
If you have to scroll down a lot after the text ends to confirm

(2) It should be expressed in a form that consumers can easily recognize. In the case of text, it should be clearly distinguished from the background and expressed by selecting an appropriate text size, font, color, etc., that consumers can easily recognize. → *clear expression method*

If the text size is too small to be found
If the text color is similar to the background, making it difficult to recognize the text
If the advertisement phrase is posted among numerous hashtags

(3) It should be clearly indicated in content. The content of economic interests such as financial support, sponsorships, etc., should be clearly indicated so that consumers can understand it easily. → *clear indication content*

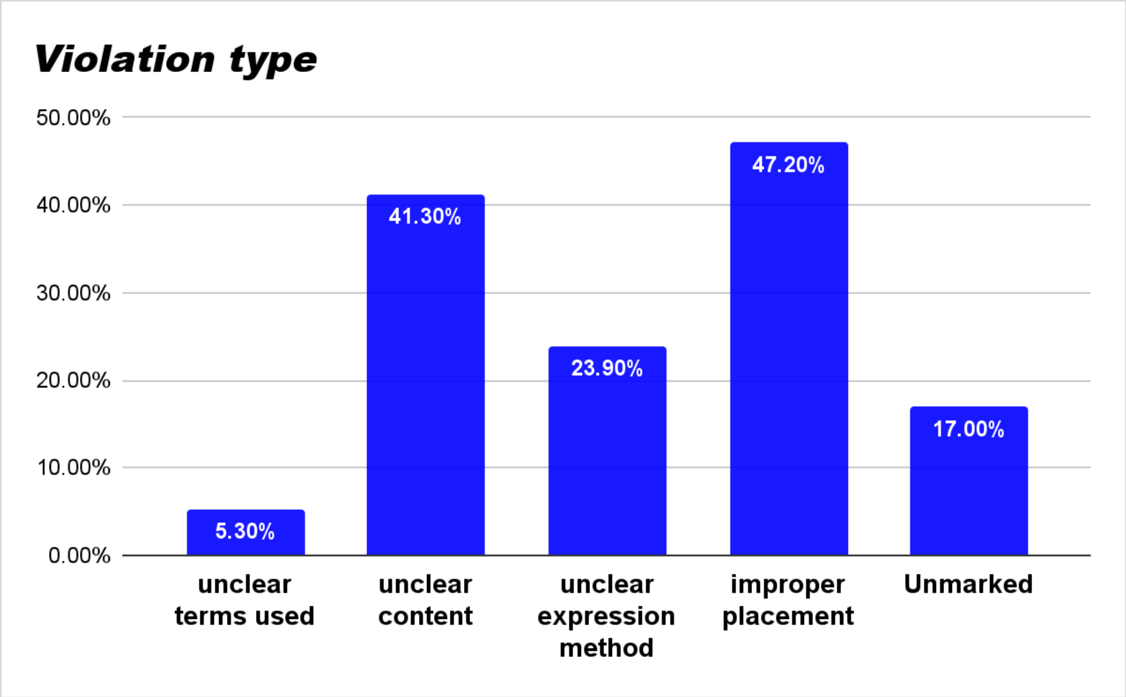
ex> 'I received an economic reward from ◇◇ Company while recommending (guaranteeing, introducing, promoting, etc.) the above ○○ product.'

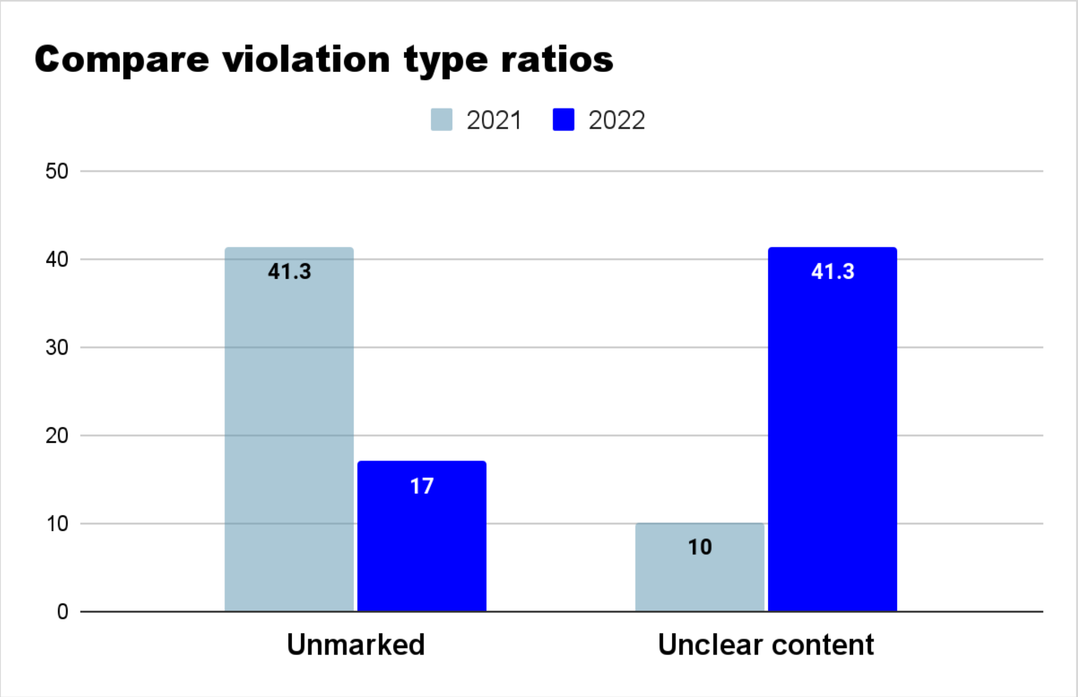
'Honest reviews written without any additional compensation besides the provided product/service', 'Experience review', 'Used for a week', 'Test Panel', 'This post is informational', 'This post includes a promotional phrase', 'Gift', 'Thank you, CEO ○○.', 'Sent from ~.'

*Advertisement phrase must be disclosed in a **way that is easily and clearly understood** by consumers*

Motivation

- According to the KFTC, a notable drop in unmarked ads in South Korea since 2021. However, illegal posts persist, using subtle tactics like inappropriate placement and unclear content to evade detection.





Motivation

- **Lack of former research**: No exact papers found on detecting buzz marketing (fake reviews) in Korean within clarivate JCR-listed academic journals.

- **Limited accuracy**: In related topics like fake news or short review comment spam detection, the highest accuracy reached only 84%.

- Researched **similar services** and found some that remove ad banners and analyze reviews, but **none that inform people if a review is fake.**

- We surveyed 114 people to determine the need for the service and found that

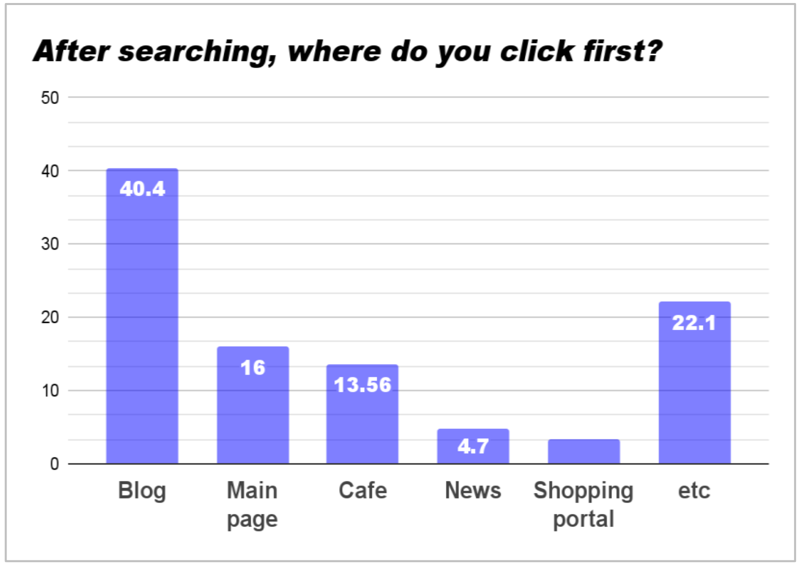
  1) **68% of people didn't recognize a post as an advertising even if it had advertisement phrase in the body**
  2) **85% of people said the presence of an advertisement indicator made the post less trustworthy**

→ ***The need for a service that lets consumers know if a review is a real review or fake review***

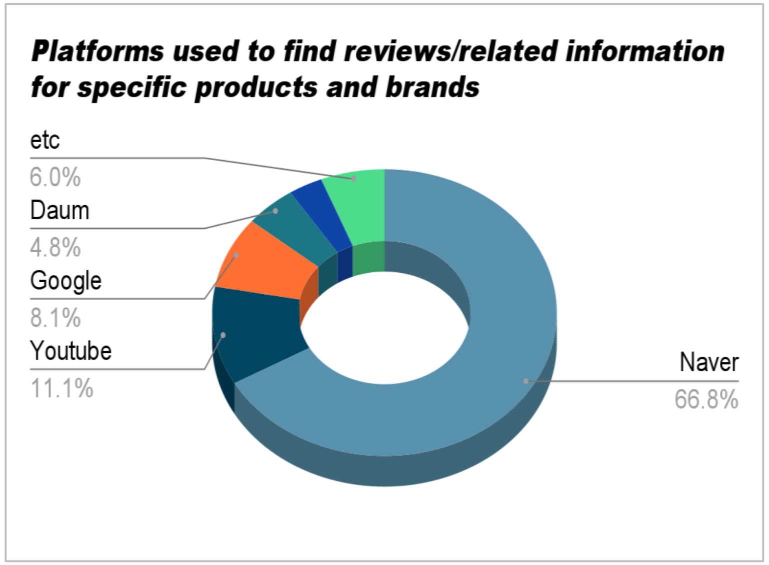
Related Work - Detection channel & category

- **Naver Blog**: **most popular review platform in Korea** and has a large number of viral marketing posts
  - Illegal stealth ads: among detections, the service sector related to restaurants had the highest proportion
  - Top search topic: The most searched topic was restaurant-related content

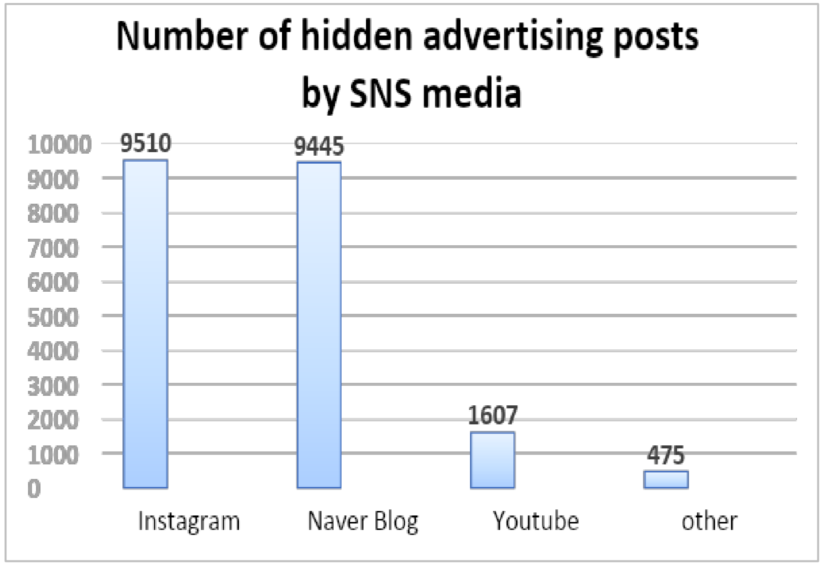
→ *So we decided to detect **Naver Blog restaurant related posts***



After searching, where do you click first?



Platforms used to find reviews/related information for specific products and brands



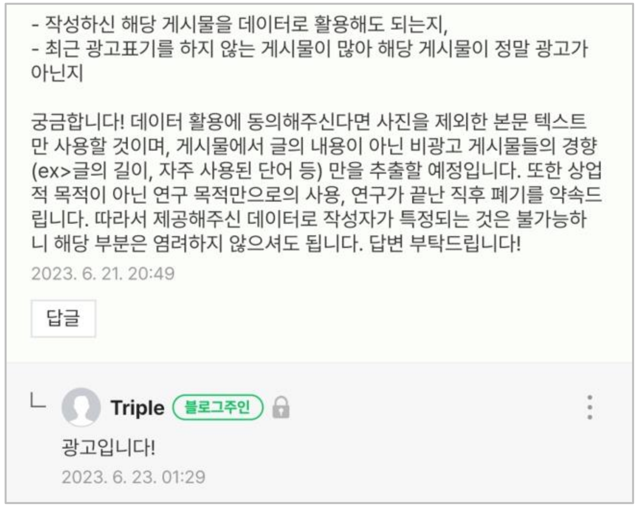
Number of hidden advertising posts by SNS media

Collecting Data

1. **Crawled nearly 30,000 blog posts** using Naver API and labeled them ad/non-ad
→ *Criteria: presence of a advertisement phrase*

2. To ensure the integrity of our data, we **commented on posts** and **emailed authors to confirm that they were indeed non-advertising**.

→ *Found **12 posts were 'unmarked' hidden advertisement posts** that did not indicate that they were ads.*
→ ***10073 ads, 17285 non-ads, 12 hidden (unmarked) ads***



- 작성하신 해당 게시물을 데이터로 활용해도 되는지,
- 최근 광고표기를 하지 않는 게시물이 많아 해당 게시물이 정말 광고가 아닌지

궁금합니다! 데이터 활용에 동의해주신다면 사진을 제외한 본문 텍스트만 사용할 것이며, 게시물에서 글의 내용이 아닌 비광고 게시물들의 경향(ex>글의 길이, 자주 사용된 단어 등) 만을 추출할 예정입니다. 또한 상업적 목적이 아닌 연구 목적만으로의 사용, 연구가 끝난 직후 폐기를 약속드립니다. 따라서 제공해주신 데이터로 작성자가 특정되는 것은 불가능하니 해당 부분은 염려하지 않으셔도 됩니다. 답변 부탁드립니다!

2023. 6. 21. 20:49

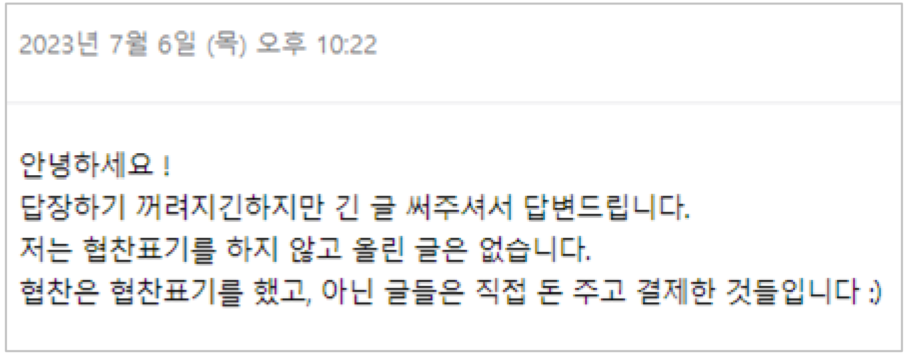
답글

Triple 블로그주인

광고입니다!

2023. 6. 23. 01:29

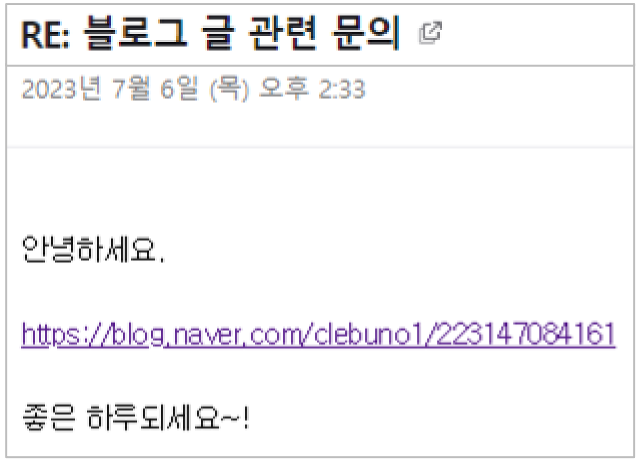
A reply that the post with the comment is actually a hidden advertisement



2023년 7월 6일 (목) 오후 10:22

안녕하세요 !
답장하기 꺼려지긴하지만 긴 글 써주셔서 답변드립니다.
저는 협찬표기를 하지 않고 올린 글은 없습니다.
협찬은 협찬표기를 했고, 아닌 글들은 직접 돈 주고 결제한 것들입니다 :)

Email reply that the writer have never posted a hidden advertising post



RE: 블로그 글 관련 문의

2023년 7월 6일 (목) 오후 2:33

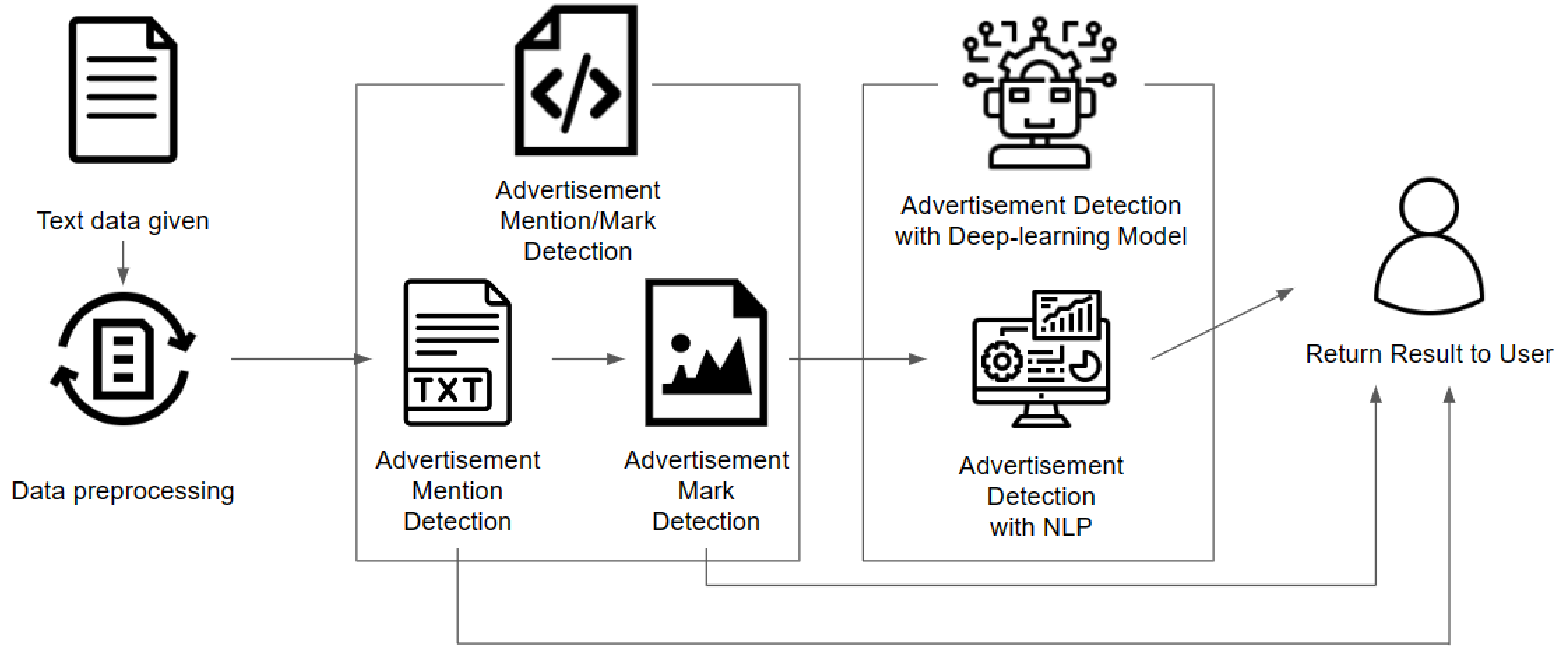
안녕하세요.

https://blog.naver.com/clebuno1/223147084161

좋은 하루되세요~!

Mail replies with links to hidden advertising posts

Belieview – Proposed System Architecture


Text data given
Data preprocessing
Advertisement Mention/Mark Detection
Advertisement Mention Detection
Advertisement Mark Detection
Advertisement Detection with Deep-learning Model
Advertisement Detection with NLP
Return Result to User

Belieview - Proposed System Architecture

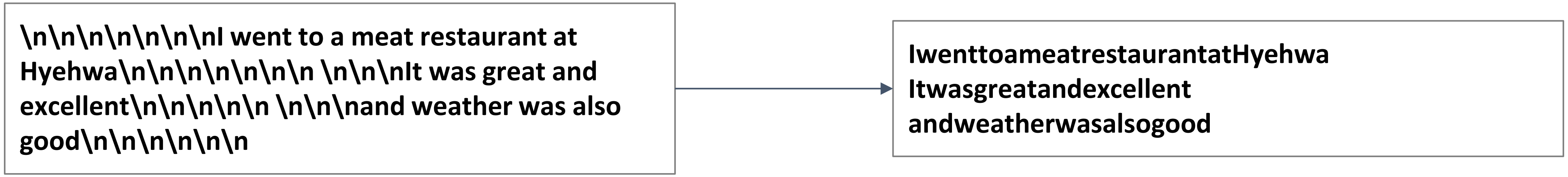
- Data preprocessing

  - **Step 1: replace multiple line changes into one line change**

    ex) \n\n\n\n\n\n\n\n\n → \n
  - **Step 2: split it into paragraphs**
  - **Step 3: if there are many hashtags and one hashtag consists one paragraph,**
    integrate all the hashtags in one paragraph

    ex) #A
    #B : Three paragraphs → #A#B#C : One paragraph
    #C
  - **Step 4: Remove blanks and marks for easy comparison with keywords**

Example)



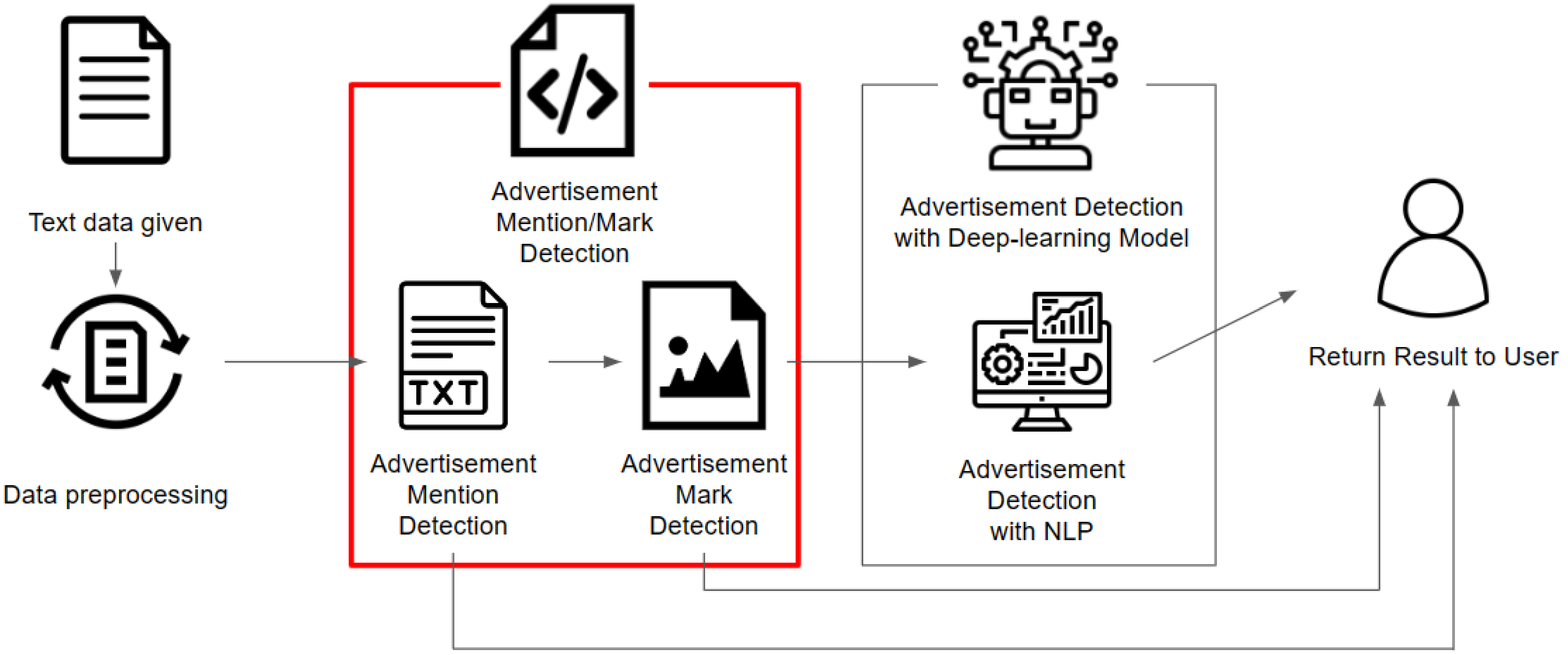
\n\n\n\n\n\n\nI went to a meat restaurant at Hyehwa\n\n\n\n\n\n\n \n\n\nIt was great and excellent\n\n\n\n\n \n\n\nand weather was also good\n\n\n\n\n\n

IwenttoameatrestaurantatHyehwa
Itwasgreatandexcellent
andweatherwasalsogood

→ *Process text data to the right form for training*

Belieview - Proposed System Architecture

- Advertisement Phrase (Mention/Mark) Detection

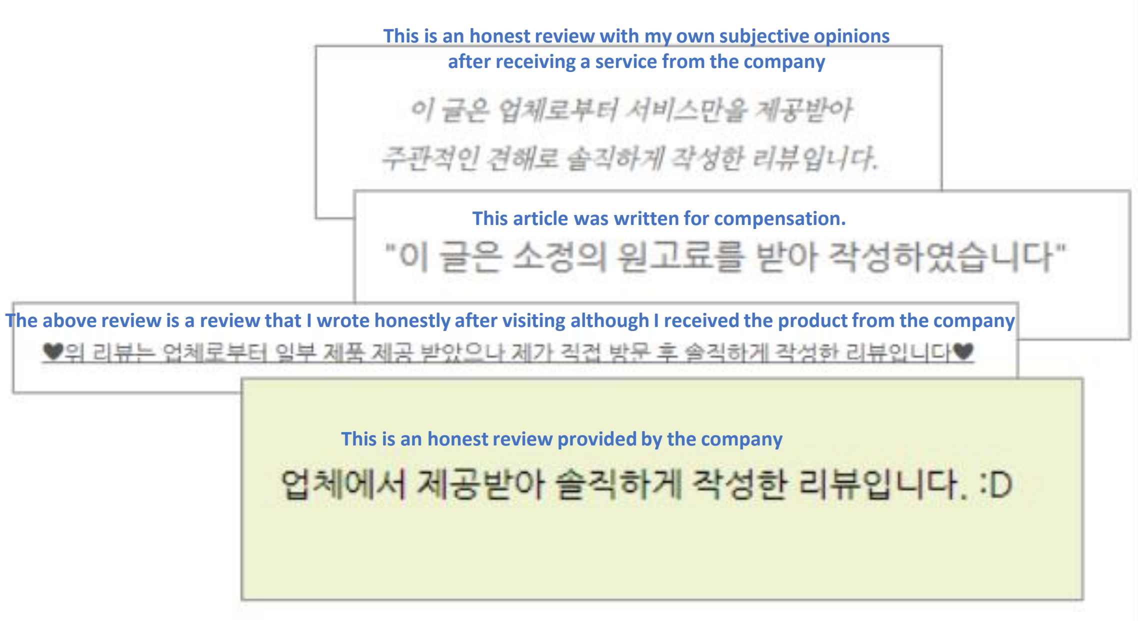

Text data given
Data preprocessing
Advertisement Mention/Mark Detection
TXT
Advertisement Mention Detection
Advertisement Mark Detection
Advertisement Detection with Deep-learning Model
Advertisement Detection with NLP
Return Result to User

Belieview - Proposed System Architecture

- Advertisement Phrase (Mention/Mark) Detection

  : The bloggers inform whether the post is an advertisement or not in the form of **text** or **image**.

  - The examples of advertisement mention(text)/mark(image)
    - Advertisement mention
    - Advertisement mark



This is an honest review with my own subjective opinions after receiving a service from the company

이 글은 업체로부터 서비스만을 제공받아 주관적인 견해로 솔직하게 작성한 리뷰입니다.

This article was written for compensation.

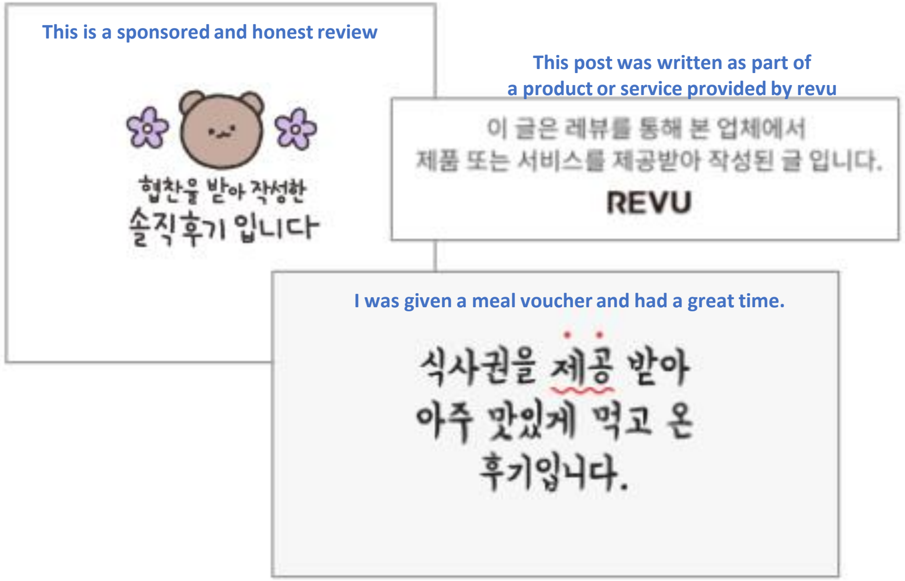
"이 글은 소정의 원고료를 받아 작성하였습니다"

The above review is a review that I wrote honestly after visiting although I received the product from the company

♥위 리뷰는 업체로부터 일부 제품 제공 받았으나 제가 직접 방문 후 솔직하게 작성한 리뷰입니다♥

This is an honest review provided by the company

업체에서 제공받아 솔직하게 작성한 리뷰입니다. :D



This is a sponsored and honest review

협찬을 받아 작성한 솔직후기 입니다

This post was written as part of a product or service provided by revu

이 글은 레뷰를 통해 본 업체에서 제품 또는 서비스를 제공받아 작성된 글 입니다.

REVU

I was given a meal voucher and had a great time.

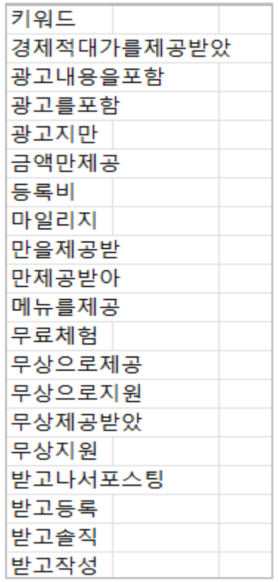
식사권을 제공 받아 아주 맛있게 먹고 온 후기입니다.

: **If such cases, detection with deep-learning model is not needed**

→ so, we decided to detect such indicators : **need to establish some classification standard**

Belieview - Proposed System Architecture

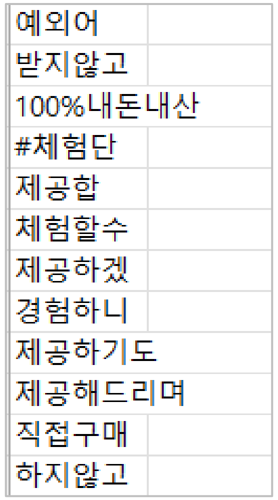
- Advertisement Phrase (Mention/Mark) Detection
  - Standard establishment result:
    226 Keywords, 11 Exception-words
    : **If keywords are included in text/image, it can be seen as advertisement indicator**



키워드
경제적대가를제공받았
광고내용을포함
광고를포함
광고지만
금액만제공
등록비
마일리지
만을제공받
만제공받아
메뉴를제공
무료체험
무상으로제공
무상으로지원
무상제공받았
무상지원
받고나서포스팅
받고등록
받고솔직
받고작성

Keywords

Received economic aid
Including advertising content
Including advertisements
Although it's an advertisement
Provided only the amount
Registration fee
Mileage
Received only
Received only
Provided menus
Free trial
Provided for free
Supported for free
Received for free
Free support
Posting after receiving
Receiving and registering
Receiving and honestly
Receiving and writing



예외어
받지않고
100%내돈내산
#체험단
제공합
체험할수
제공하겠
경험하니
제공하기도
제공해드리며
직접구매
하지않고

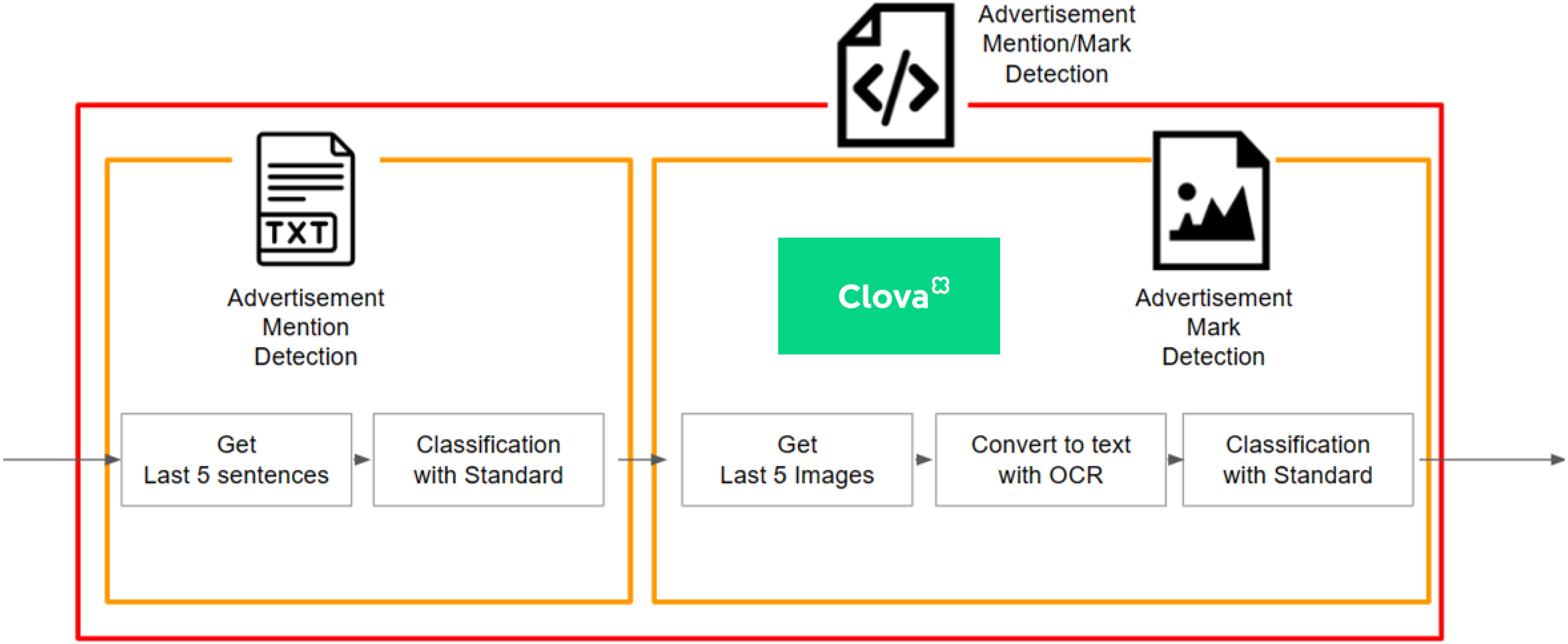
Exception-words

Without receiving
100% self-funded
#Reviewer
Providing
Available for experience
Will provide
Having experienced
Also providing
Providing to you
Direct purchase
Without doing

Examples)
‘Sponsorship’ → Case 1: Advertisement
‘Without sponsorship’ → Case 2: Non Advertisement
‘Was delicious’ → Case 3: Not Defined (Target for additional stages)

Belieview - Proposed System Architecture

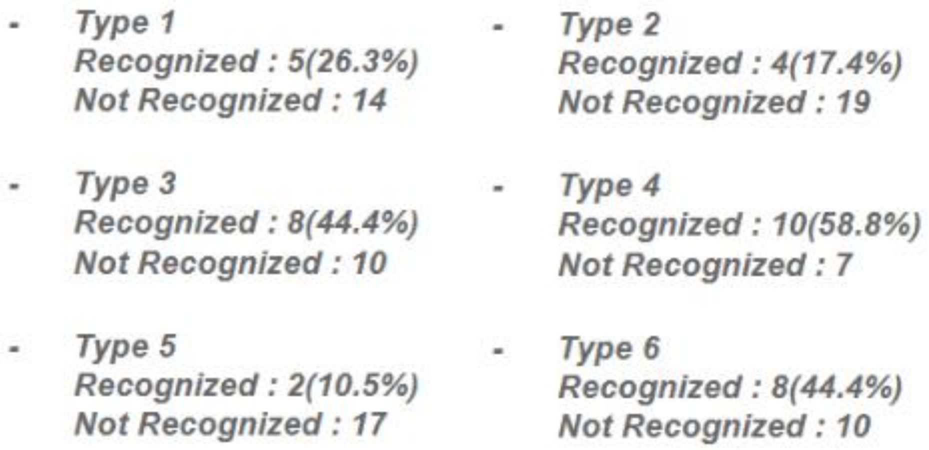
- Advertisement Phrase (Mention/Mark) Detection – more detailed view



Belieview - Proposed System Architecture

- Advertisement Phrase (Mention/Mark) Detection
  - Detection rate test
    - Test 1 : random 500 posts from dataset
      Result : 0.99
    - Test 2 : newly crawled 100 posts
      Result : 0.99

- Detection rate test for human



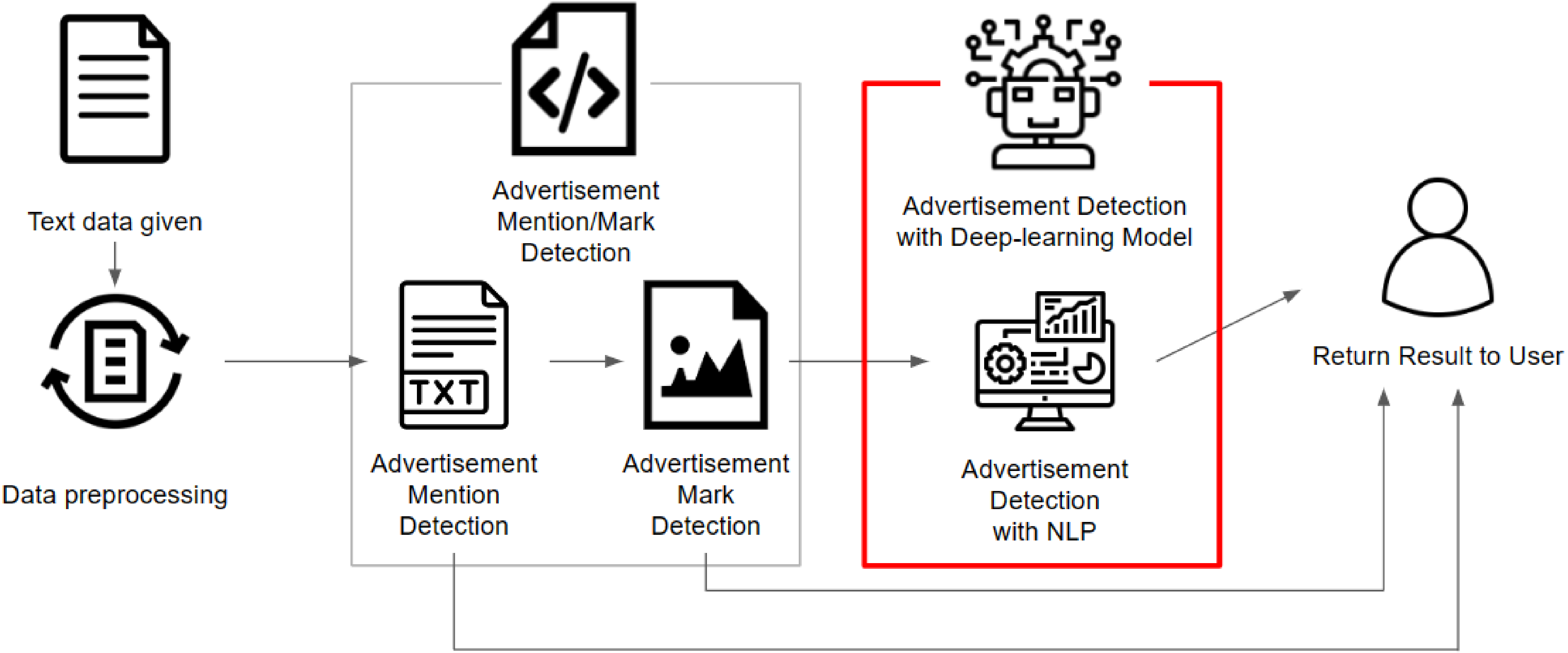
- *Type 1*
  Recognized : 5(26.3%)
  Not Recognized : 14
- *Type 2*
  Recognized : 4(17.4%)
  Not Recognized : 19
- *Type 3*
  Recognized : 8(44.4%)
  Not Recognized : 10
- *Type 4*
  Recognized : 10(58.8%)
  Not Recognized : 7
- *Type 5*
  Recognized : 2(10.5%)
  Not Recognized : 17
- *Type 6*
  Recognized : 8(44.4%)
  Not Recognized : 10

: our service can detect **99%** of advertisement mentions/marks while overall detection rate of human is **32%**

→ our service can help users cope with maliciously hided/hard-recognizing advertisement mark

Belieview - Proposed System Architecture

- Advertisement Detection with Deep-learning Model


Text data given
Data preprocessing
Advertisement Mention/Mark Detection
Advertisement Mention Detection
Advertisement Mark Detection
Advertisement Detection with Deep-learning Model
Advertisement Detection with NLP
Return Result to User

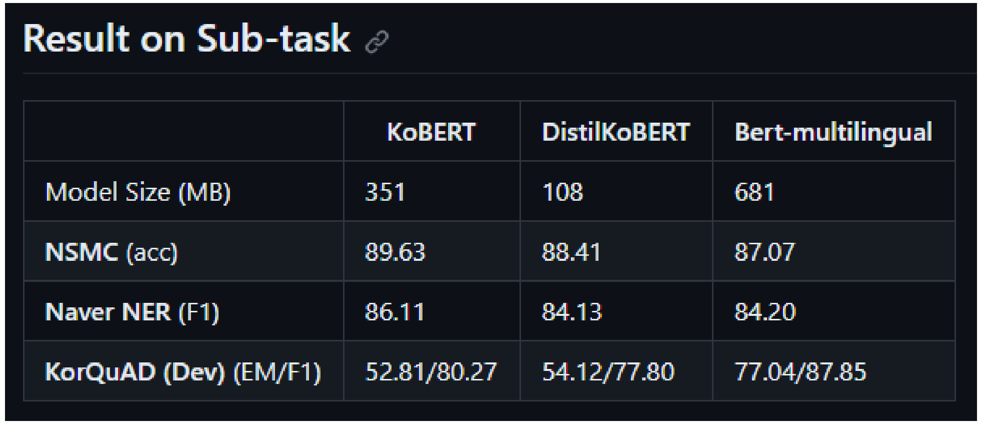
Belieview - Proposed System Architecture

•Advertisement Detection with Deep-learning Model
→ *we fine-tuned **DistilKoBERT** as our model since it shows reasonable performance with fewer resources*

DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter

Victor SANH, Lysandre DEBUT, Julien CHAUMOND, Thomas WOLF
Hugging Face
{victor,lysandre,julien,thomas}@huggingface.co

- **DistilKoBERT**
  - Korean version of DistilBERT
  - Light version of KoBERT
  - 40% smaller, 60% faster in training
  - Recorded stable performance with less resources in many benchmarks



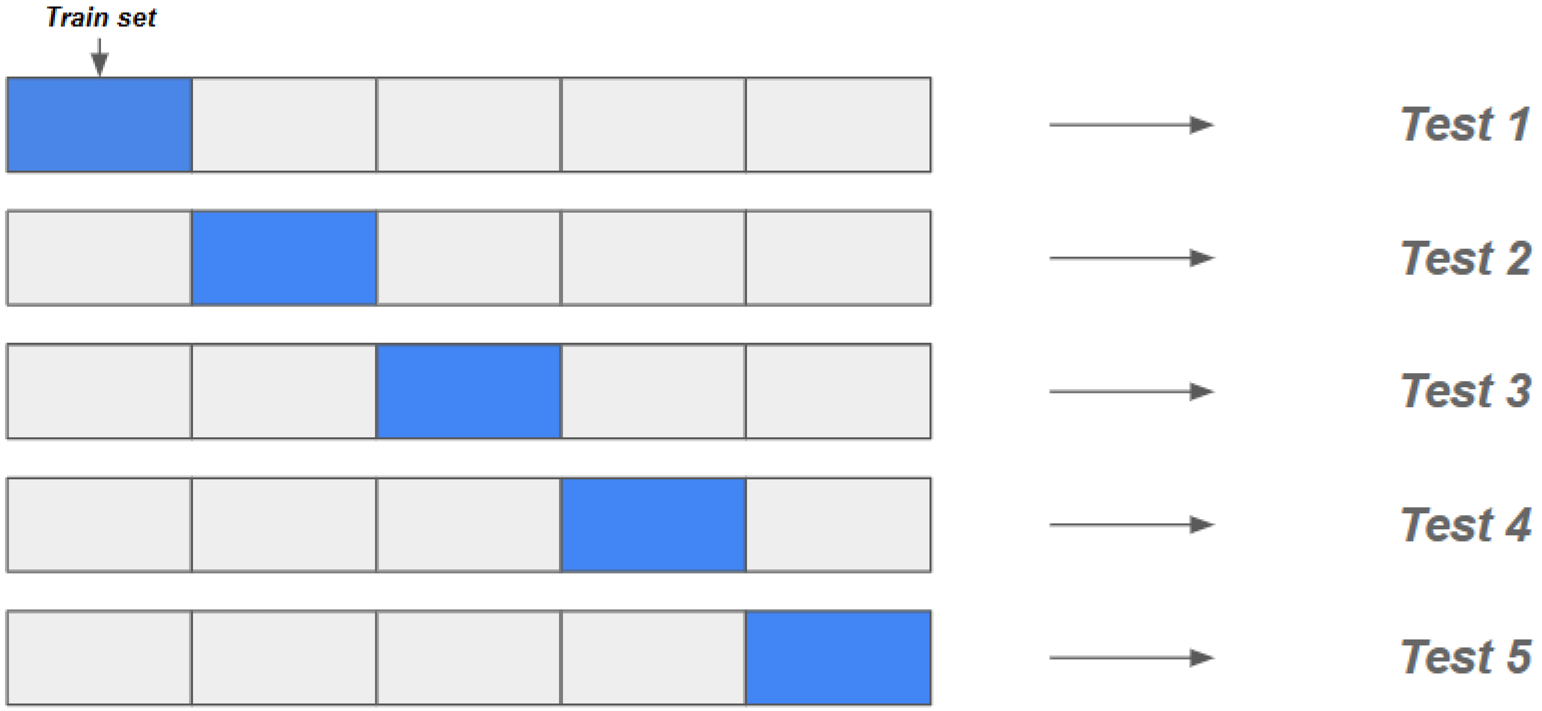
Result on Sub-task

	KoBERT	DistilKoBERT	Bert-multilingual
Model Size (MB)	351	108	681
NSMC (acc)	89.63	88.41	87.07
Naver NER (F1)	86.11	84.13	84.20
KorQuAD (Dev) (EM/F1)	52.81/80.27	54.12/77.80	77.04/87.85

- Also, we conducted **Data Augmentation**
  → As a result, dataset become five times larger

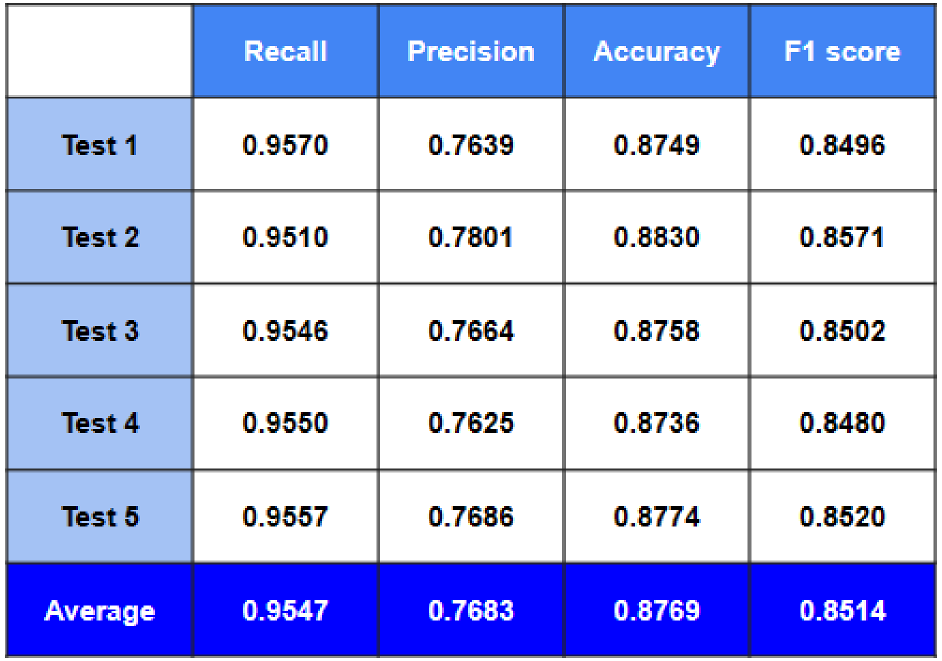
Belieview - Proposed System Architecture

- Advertisement Detection with Deep-learning Model
  - DistilKoBERT
    - Performance
      : Referring to the concept of Stratified 5-fold validation, tested whether it produces stable performance



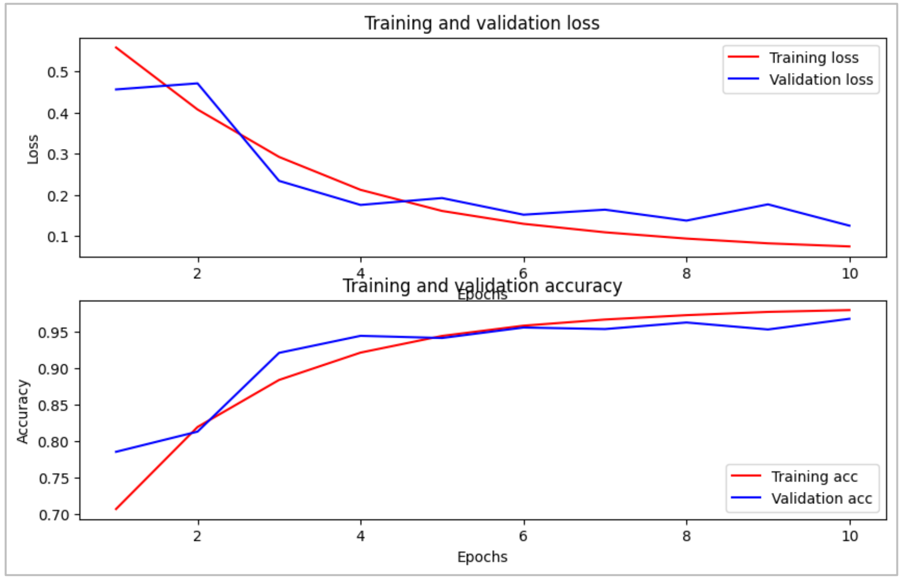
Belieview - Proposed System Architecture

- Advertisement Detection with Deep-learning Model
    - DistilKoBERT
        - Performance
          : **Scored accuracy of 0.8769**

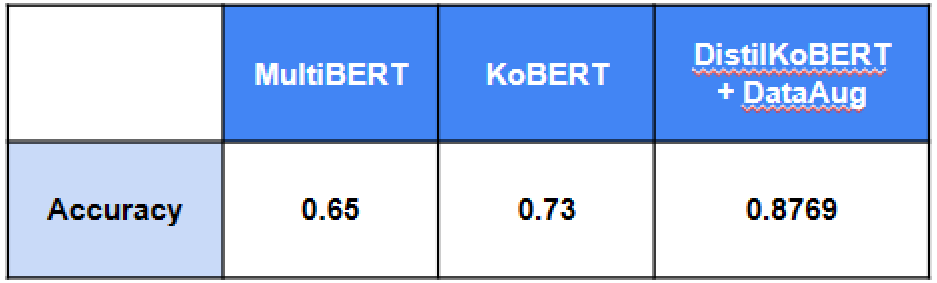


	Recall	Precision	Accuracy	F1 score
Test 1	0.9570	0.7639	0.8749	0.8496
Test 2	0.9510	0.7801	0.8830	0.8571
Test 3	0.9546	0.7664	0.8758	0.8502
Test 4	0.9550	0.7625	0.8736	0.8480
Test 5	0.9557	0.7686	0.8774	0.8520
Average	0.9547	0.7683	0.8769	0.8514

Performances of all 5 tests



Well trained, Epoch: 5

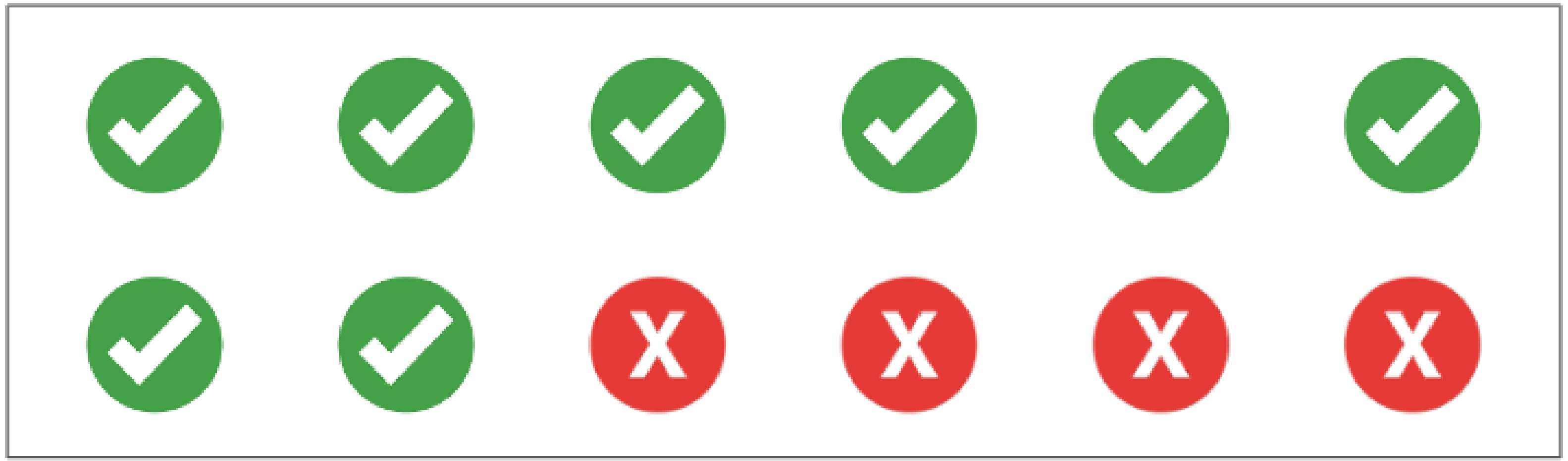


	MultiBERT	KoBERT	DistilKoBERT + DataAug
Accuracy	0.65	0.73	0.8769

Performances compared to other models

Belieview - Proposed System Architecture

- Advertisement Detection with Deep-learning Model
  - Test on hidden advertisement posts
    → captured 8 hidden advertisements within 12 candidates


✓ ✓ ✓ ✓ ✓ ✓

✓ ✓ X X X X

: **Our service can be used in not only detecting advertisement indicators, but also hidden advertisements** (advertisements that indicators are not revealed)

Belieview - Proposed System Architecture

- Advertisement Detection with Deep-learning Model
  - **Comparison with human**
    : Asked to classify 6 blog posts, 39 people participated
    (Label is deleted from given data)

  - Time consumed
    - ***Our service***
      Overall time consumed : 1.53 sec

    - ***Human***
      Overall time consumed : 13.96 sec

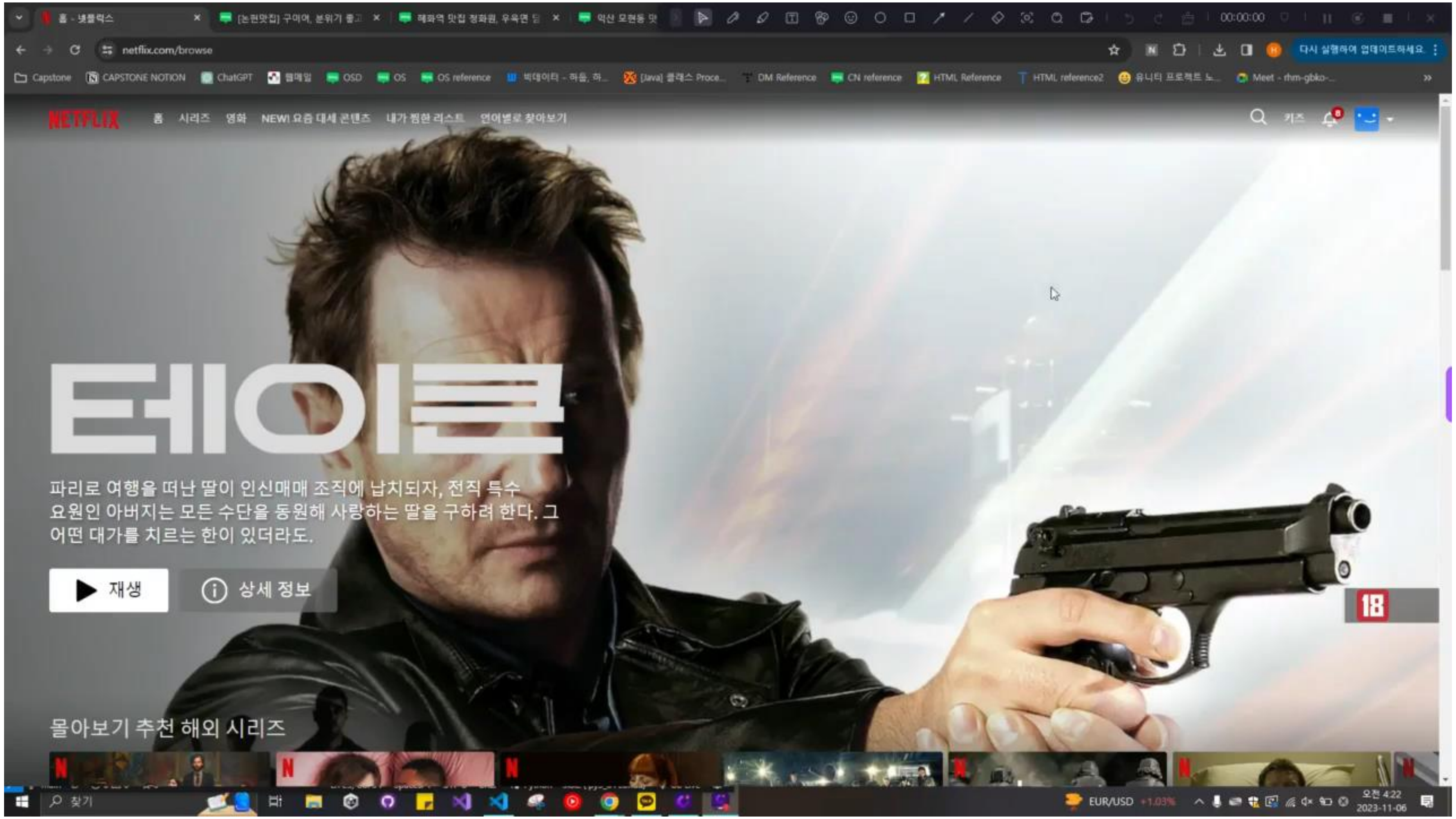
  - Classification accuracy
    - ***Our model***
      - ***Overall accuracy : 0.88***

    - ***Human***

      - ***Overall accuracy : 0.52***

Belieview - Demonstration

- Demo video



파리로 여행을 떠난 딸이 인신매매 조직에 납치되자, 전직 특수 요원인 아버지는 모든 수단을 동원해 사랑하는 딸을 구하려 한다. 그 어떤 대가를 치르는 한이 있더라도.

Conclusion & Limitation

•Can effectively cope with maliciously hided/hard-recognizing advertisement marks, and can help to make reasonable reasoning about the stealth advertisements

•Our model recorded accuracy of **0.88**, **69%** more accurate and **9 times** faster than people

•Got through thorough consideration on how to make genuine dataset, which is a chronic problem of spam detection research

•This did not gone well due to lack of motivation on hidden-advertised bloggers to tell the truth

If cooperation with the public authorities is possible, better results can be expected

References

1. https://github.com/monologg/DistilKoBERT

2. Sanh, Victor, et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." *arXiv preprint arXiv:1910.01108* (2019).

3. Wei, Jason, and Kai Zou. "Eda: Easy data augmentation techniques for boosting performance on text classification tasks." *arXiv preprint arXiv:1901.11196* (2019).

4. https://github.com/catSirup/KorEDA/tree/master

5. https://www.sejungilbo.com/news/articleView.html?idxno=41461

6. http://www.neobizsys.co.kr/?page_id=101

7.https://www.ftc.go.kr/www/selectReportUserView.do?key=10&rpttype=1&report_data_no=9936