[1] Yung Gi Wu

[2] Jwu Jenq Chen

[3] Wen Fun Su

[4] Shu Hui Lin

[5] E Cheng Chen

[6] Yun Jei Cheng

Intelligent Alerting System Development by Network Media Semantic Analysis and Face Recognition

Journal of Electrical Systems

Abstract: - Nowadays, the network has become the major data provider to everyone. Various providers feed data to the end users to achieve the flow of information for their purposes. In this research, we develop an intelligent alerting system that can automatically sematic analyze the contents from the network and extract the altered information including event description and suspect's photo. We crawl the PTT, Dcard, and Google News in Taiwan as our sources to process due to the popularity of the three social network systems. These are the three very popular social media in Taiwan and real-time events, news is always posted to the three media very soon, therefore, we adopt them as our research source. Video monitors installed on campus or in the community after receiving alerts and photos will enter alert mode. Real-time face recognition will be issued for the people captured by the camera. Once the system recognizes the suspect, it will notify the security guard to take necessary actions to protect the people. There are two major parts to this research, the first one is the media semantic analysis and the second one is the real-time face recognition. It is an intelligent altering system that is capable to find out the suspected assault or rapist rapidly once he is captured by a monitor and recognized after people post the crime events to the media.

Keywords: altering; face recognition; intelligent system; media semantic analysis; real-time monitoring.

1) Introduction

The following figure showed the partial paragraphs from the news: On 22 August 2022, two police officers chased a wanted criminal for theft, but they were killed instead in Tainan, Taiwan. According to the latest footage, the police officer was seen with blood all over his face and neck. The suspect smiled weirdly at the police car's dash cam with a gun. The weird smile makes people shudder when they see it. The police traced the dark red ordinary motorcycle that the suspect rode, with the license plate "369-PGB". They found the motorcycle at Tu-cheng High School in the An-Nan district in the afternoon and found that there were obvious bloodstains on the motorcycle, but the criminals had fled. Figure 1 shows the photos of this event in the news.

Figure 1 The photos posted in the news.

This is an emergency event that all people should be notified of and help to find the suspect. In this research, an intelligent alerting system was developed by crawling the popular social network system and conducting a semantic analysis to determine the emergency of the news.

People like to post to social network systems for any purpose. Once a dangerous incident occurs, people are now taking pictures with their mobile devices and posting what they see on social networks in a fast and easy way.

[1,2,3,4,5,6] Department of Computer Science & Information Engineering, Chang Jung Christian University/Tainan, Taiwan.

Email: 1wgy@mail.cjcu.edu.tw , 2jchen@mail.cjcu.edu.tw , 3fangfangsusu@gmail.com , 4a86890502w@gmail.com , 5s0951211@gmail.com , 6 fa506lian @gmail.com

Corresponding Author*: Yung Gi Wu.

News channels on the network broadcast various news to end users' mobile devices. Therefore, crawling the posts and news on social network systems and News channels can make people aware of dangerous events. If a dangerous event occurs nearby, he will be more careful to keep an eye out for suspicious people.
Based on the aforementioned background and motivation, this research will implement a warning development system based on the semantic analysis of online media combined with face recognition. It mainly captures the search trend keywords of Google Trend and conducts trend analysis on Internet forums PTT and Dcard [1] which are the most popular Network media. The biggest commonalities of choosing these two platforms are that they are updated quickly, can obtain information, are diverse, and the age distribution is not too large. Similarly, the published positions are also different, so the data from different sources of the two parties are collected and compiled. Mainly use BERT to classify articles and analyze positive and negative comments, combine face recognition to detect pictures in the text, and use article classification to judge whether the article needs corresponding warning effects. These functions will enable network users of news media or social platforms to reduce time costs when sorting out and viewing data, and the results of positive and negative comments can be used for reference and comparison, finally, the compiled data will be visualized in the form of charts presented on the web page, making reading and query easier.
There are many available systems like Google Alerts[2], Daily view[3], Meltwater[4], etc. The relevant functions include keywords and semantic analysis in hot news posts. Most of the captured information comes from well-known social media. The layout configuration is also exquisite, and the operation is simple and easy to understand. However, the functions are not merged, so when using the software, you must switch between each other to obtain complete information, In different systems, the data sources are not consistent, resulting in a gap in the amount of data between each other when operating, and classification In terms of the type of article or news, it is also not detailed. Therefore, we intend to realize or optimize the above functions when making this topic, integrate the currently popular online media, and make further subdivisions in the classified part so that users can search more quickly and conveniently to achieve the goal of fast alerting in emerging events. There are some issues to be solved in the following: a. The way of building the corpus and the added definition of vocabulary classification. b. Scheduling technology, automatic transmission of warning messages. c. Trend analysis, data evaluation, and methods of judging positive and negative reviews. d. Database planning, big data access.
There are four advantages of this system. a. Save time and cost: You don't need to spend too much time checking the news or social media platforms to know the ranking of popular keywords and related basic information. b. News label: When ranking keywords, users can use this label to know which news categories are related to this keyword. c. Data visualization: The data results of discussions and positive and negative comments are presented in the form of line charts and pie charts in the visual chart, which is convenient for users to browse and view, and improves the readability of data. d. Facial recognition early warning: Use the Dlib package of Python to select pictures for comparison connect a camera for detection, and identify the person who needs to be warned. e. Alert setting: Through the classified labels and setting the desired recipients, the alert mail can be sent to different recipients through the APScheduler scheduler.

2) Literature Review

In this paper, we compare our system to Google Alerts, Daily View, and Meltwater. We give brief descriptions of the three systems.
Google Alerts Search for news by keyword and send the search results to the email address provided by the user, which provides greater convenience, it is also faster for companies or operators to learn about their brand news, but the source of information is limited to the Google platform.
Daily view: Obtain a huge database through Internet crawler technology, covering more than 90% of Taiwan's Internet news channels, social platforms, FB, PTT, discussion boards, and blogs, and through semantic sentiment analysis, exclusive Keyword thesaurus, analyze hot topics and produce professional network big data survey reports.
Meltwater It is a set of community monitoring software that captures the content of global media and social platforms, which can increase monitoring keyword groups, identify opinion leaders, manage multiple fan groups, export or automatically generate reports after data analysis, and provide customized modules.
In the following table, we summarize the comparison of functions:

Table 1. Function Comparison

	Proposed system	Google Alerts	Daily View	Meltwater
Data Source	Multiple	Single	Multiple	Multiple
Data Classification	YES	NO	NO	NO
Positive-Negative Comments	YES	NO	NO	YES
Face Recognition	YES	NO	NO	NO

Alerting system	YES	NO	NO	YES
Charge	NO	NO	NO	YES

From Table 1, we can observe the differences between the proposed system and others, and the most special function is face recognition. Many semantic analyzes aim to focus on public opinion analysis, the reason why we hope to combine face recognition is not only to respond to the trend of technology but also to achieve the effect of increasing warning through the function of face recognition. If we can use the recognition system to help find urgently missing family members or combine face recognition with community cameras to locate suspects or rescue victims will effectively improve community safety.

The technology to develop the proposed system includes Django to design the web framework, MTV(Model Template Views) [5]. It decides which data to process to correspond to the Template screen we designed, and because the core of the project architecture is "function" as the main axis, it increases the flexibility of system expansion and development speed and provides an ORM (Object-Relation Mapping) framework to Develop databases and allow developers to more easily observe and track changes in data access through the background management interface. [6]

MySQL provides different program (c++, PHP, Java, etc.) interfaces, with the characteristics of small size, fast speed, relatively low cost, and open source. Most of the small and medium-sized websites (perfect combination of MySQL and PHP) choose MySQL as their Website database [7]. Through the previously mentioned Django website framework, its built-in project framework can provide a method to connect to the database and use ORM framework construction, which can increase the development speed [8]. In addition, considering the huge amount of data captured after a news search will cause a great burden on the memory space, so in the associative database distinguished by the data structure, follows the ACID principles (Atomicity (primitive), Consistency (Consistency), Isolation (isolation), Durability (continuity)), can ensure that the data captured in our database conforms to the specifications we set, and sort out the relationship between each association to facilitate query actions.

Automatic update information: Advanced Python Scheduler (APScheduler) [9] is used. It has three different schedulers. Here, Interval-based execution will be used to run jobs at even intervals, and the start and end times can be set.

BERT(Bidirectional Encoder Representations from Transformers) is the model based on Transformers' Encoder can reflect the semantic meaning of sentences more comprehensively than bidirectional RNN and LSTM and can be parallelized to improve training speed [10][11]. Google uses the two tasks of the Masked Language Model and Next Sentences Prediction, with a large amount of text, to train the model in an unsupervised learning manner. According to specific tasks, we choose to use the Simple Transformer package extended from Transformers. The Transformers suite itself provides thousands of pre-trained models, which can perform natural language processing tasks such as text classification, information extraction, question answering, summarization, translation, and text generation [12], while Simple Transformers can train and test more quickly. And evaluate the Transformers model [13], for the task of news classification and positive and negative review analysis, because it is the processing of Chinese data, so choose BERT's bert-base-Chinese pre-training model, and the former is multi-category, the latter is two Meta tags, both belong to classification tasks, so in the fine-tuning (Fine Tune) stage with the collected data sets, use the MultiLabelClassificationModel that uses multi-label classification in text classification and the ClassificationModel model that provides binary and multi-class text classification for training [14]

As to the corpus, which can be divided into two parts. One is to use the open data provided by UdicOpenData at National Chung Hsing University, Taiwan. Popular Data and Smart Computing Laboratory [15], which is used in the judgment of positive and negative evaluation analysis. A large number of villagers leave messages, including a large number of analyzed and organized texts, with established formats and tags, and it is a database that can be continuously and dynamically supplemented. By using the corpus, users can quickly understand the use of a word in all contexts.

The other part is the event judgment applied to the warning system of social news, and in the process of collecting semantic analysis data, it is found that most of the current news platforms will give a general direction to the published content, such as social, domestic, sports, travel, etc. They also provide users with the ability to search for subdivided items by entering specific keywords. However, "social events" cover a wide range, including current problems facing society, local realistic records, and large and small Knowledge content, etc., if only expressed in a single general direction, it is impossible to quickly understand its classification attributes before reading the article. Here we use the recently emerging deep learning BERT model to deal with it because the fine-tuning stage is a supervised learning method, so we built a label set for news edition classification, Figure 2 shows the content of the label set, and the source of the data is from China Times [16], Liberty Times [17], ETtoday [18] and UDN News [19], crawled with input keywords (firearms, sexual assault, fight, traffic accident, missing, murder) as tags to query news, a total of 7,072 entries, this original label also provides us with one of the

benchmarks for model prediction evaluation. Among them, it is considered that a piece of news may meet more than two classification properties at the same time, and it is also observed that because the data is obtained from the input keyword search, the label may not match the actual content. For example, the missing news example shown in Figure 3, this news has a high degree of correlation with "missing" based on human judgment, but the original tag crawled is the classification of "murder", so we will manually label the collected information one by one, and finally subdivide it into 13 sub-tags (firearm, fight, sexual assault, missing, traffic accident, drug, homicide, steal, kidnap, robbery, swindle, intimidate, other).

content	raw_label	multi_label	single_label
阮哲夫面對刑警，神色	3	Sexual assault	Sexual assault
隨著塔利班政權上台，	0	homicide,steal	homicide
台大醫院整形外科醫師	4	other	other
（記者張瑞楨／台中報	1	fight,accident	fight
北市聯合醫院和平院區	6	other	other
心疼好友亞藍遭逢不幸	3	Sexual assault,other	Sexual assault
歷時近半個月的東京奧	0	other	other
一度面臨瓦解的四海幫	0	other	other
東京奧運今（23日）開	0	other	other
桃園市一名媽媽搶愛蔥	4	other	other
民進黨台北市黨部前請	0	drug	drug
相關新聞連結：獨家》	0	firearm,rointimidateery	firearm
2020年7月3日，任職於	2	Sexual assault,homicide	Sexual assault

Figure2 Labeled items from the left contents

〔記者萬于甄／台南報導〕雲林一名58歲蘇姓婦人因失蹤多時，家住台南永康的女兒聯繫不上，趕緊向永康警分局報案，警方受理後隨即通報緊急協尋，女兒當初報案時，也曾懷疑媽媽遭同居人藏匿，心急如焚，孰料，最後仍傳來不幸的消息。警方說明，蘇婦的女兒8月10日向派出所報案，指稱媽媽住在雲林，平時都會與媽媽通話聯繫，但媽媽已一週沒有消息，打電話也都聯繫不上，懷疑媽媽遭同居人藏匿，深怕會有危險，趕緊報案請求警方協尋。請繼續往下閱讀...警方表示，因蘇婦的同居人鄭男被列為家庭暴力高關懷對象，因此，受理蘇婦女兒報案後，即向雲林西螺警分局通報緊急協尋。

Figure3 Partial news content concerning a missing daughter.

Dlib is a set of original libraries that include machine learning, image processing, image recognition, etc., and comes with image encoding and decoding and provides a function library with Python API. It is developed in C++ and is currently widely used in industry and academics. It is also used in robots, embedded systems, mobile phones, and even large-scale computing architectures, and most importantly, it is not only open source and completely free but also cross-platform (Linux, Mac OS, Windows) [20]. In this system, two trained models are also used. First, the face recognition model (shape_predictor_68_face_landmarks) is used to obtain 68 key points through the rectangular frame of the face, and the face key point detector (Dlib_face_recognition_resnet_model_v1) is used to acquire them. The newly obtained key points are used to obtain face descriptors for face recognition.
OpenCV (Open Source Computer Vision) Library is a cross-platform computer vision library that can handle many fields, such as developing real-time image processing, augmented reality, computer vision, and graphic recognition programs. Its core is written by C++, there are also development interfaces for other languages, such as Python, Java, etc., which is a very convenient and free tool. This system will use OpenCV with Dlib to implement the face recognition function [21].

3) System Configuration

Figure 4 is the architecture diagram of the system for this research, including Django, Google News API, Database using Mysql, Dlib, crawlers, BERT news classification labels, positive and negative evaluation analysis, face recognition, and warning systems.
Crawl the search trend on Google Trend for keyword ranking, and Google News API is the medium for crawling network content, grabbing the content and related pictures of each news article. BERT is used for article classification and conducting positive and negative analysis. Keyword extraction uses tf-idf to calculate the content of the article to obtain the keyword of the article. Dlib is used to train image detection and recognition. Mysql is the main database for accessing data. Django is the presentation method of this system.

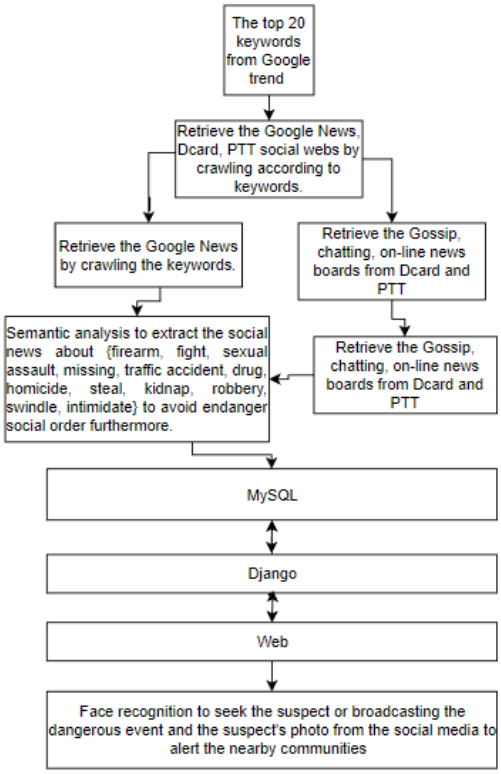

Figure4 System configuration

Figure 5 is the hot keywords ranking. You can observe the keyword ranking on that day, WordClouds, and news tags. Figure6 shows the article overview will display keywords and some basic information including title, time, and source of information. Figure7 shows the line chart of trend analysis, calculated daily, mainly showing the discussion degree of PTT and Dcard, including the number of likes, posts, and comments. Figure8 is the analysis of positive and negative comments. It will calculate the title or content related to the keywords of PTT and Dcard, output the comment prediction results as a pie chart through statistics, and display it on the webpage. Note that, the reason why "Maserati" was so hot on that day was due to a car accident occurred because a young man drove a Maserati to hit another young man. Maserati collided with a male college student. The male loudly was beaten by three opponents until he was seriously injured. This social event raises a lot of concerns for the people in Taiwan.

Figure 5 Keywords ranking

Figure 6 News about Maserati

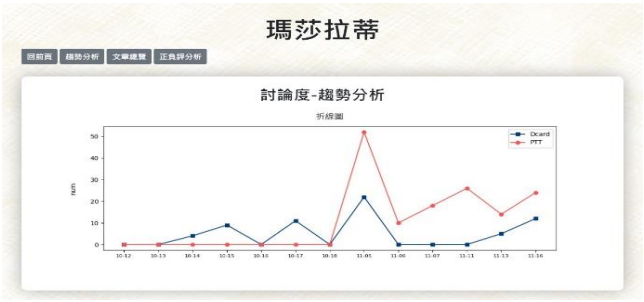

Figure 7 Discussion trend

Figure 8 Positive and negative pie chart about the Maserati event.

Crawl the search trends on Google Trends as keywords, update them regularly, store them in the database, calculate the number of news media in the database at the same time, sort the keywords to sort them accordingly, and display them on the webpage. Sorting words by keywords, crawling URLs through Google News API crawling article content and pictures from relevant URLs, and collecting basic information of topic content, such as URL, time, title source of data, etc., displayed on the web page.

Since social events cover a wide range of content, sub-tabs are further classified for the social version, which is convenient for users to understand the direction of news more quickly. Figure9 shows the flow chart of news classification. In the beginning, the keyword-related news is crawled first, and multi-label labeling is performed manually. It is divided into 13 sub-label items. In the pre-processing, the label item represented by the news is set to 1 at the same time. Otherwise, set it to 0. We use an array to represent, and then divide the data into three parts, including training data, test data, and evaluation data, through the simple transformers package, use BERT's Bert-base-Chinese pre-training model, change the parameters and perform Train and test, and finally store the trained model [22].

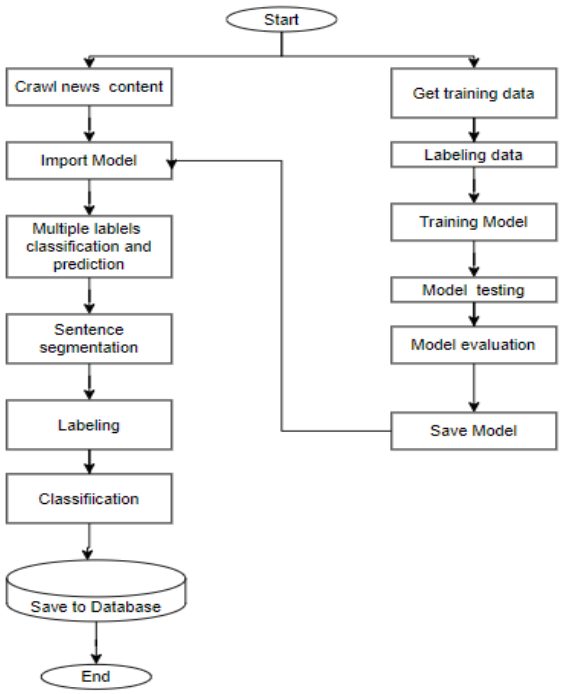

Figure 9 Classification of News

When predicting data, first load the trained multi-label classification model, and perform multi-label prediction on the content of the article. At the same time, use the ckiptagger package developed by Academia Sinica, Taiwan to segment the article and get the text. The collection of words, followed by part-of-speech labeling and named entity identification, to obtain the names, place names, dates, and times mentioned in the post or news, and finally stored in the database together with the multi-label classification results for subsequent use. Figure 10 shows the result after classification which was News content about the Maserati fight event.

content	multi_label
台中市議員黃健豪今天檢舉，瑪莎拉蒂擦撞毆人案涉案者以及助陣者，亂停車、未戴口罩、亂丟煙蒂等。台中市政府表示，蒐證依法裁處。 台中市宋姓男大生開車擦撞瑪莎拉蒂遭毆傷案，舉國矚目。 國民黨籍台中市議員黃健豪今天在台中市議會召開警消環衛業務質詢具名檢舉，他表示，調閱影片，發現打人者以及助陣的人，亂停車、亂丟煙蒂，還有人沒戴口罩未走人行道等。\r\n黃健豪表示，這些違規罰責加起來，所有人遭開罰的金額可能超過新台幣12多萬元，要求警察局、衛生局、環保局依法開單。\r\n警察局表示，交通違規事實舉發以及亂丟菸蒂、未戴口罩的事實認定及裁罰，第六警分局會蒐集相關違規事證，分別函請環保局及衛生局依法裁處。\r\n環保局說明，瑪莎拉蒂行車糾紛案件當事人隨意棄置煙蒂，經檢視影片確有違規行為，16日已函請警察局提供行為人的基本資料，將依違反廢棄物清理法告發處分，可處新台幣1200元以上6000元以下罰鍰。\r\n衛生局補充指出，未依規定佩戴口罩者，將處新台幣3000元至1萬5000元。該案件是否違反傳染病防治法，將依案件發生的場域及事實，依照市府今年5月21日公告，續由場域權責主管機關釐清後進行裁處。	fight

Figure 10 An example of labeling

Figure 11 is a flowchart of positive and negative evaluation analysis. In the process of model building, the collected data is first processed and then divided into three data sets: training data, test data, and evaluation data. BERT's bert-base-Chinese pre-training model is used through the simple transformers package, train and test after changing its parameters, and finally stores the trained model [23][24]. Next is data prediction. After normalizing the crawled comments, special keywords are removed, and then the previously trained model is used to predict whether the comments are positive or negative and finally written into the database for storage.

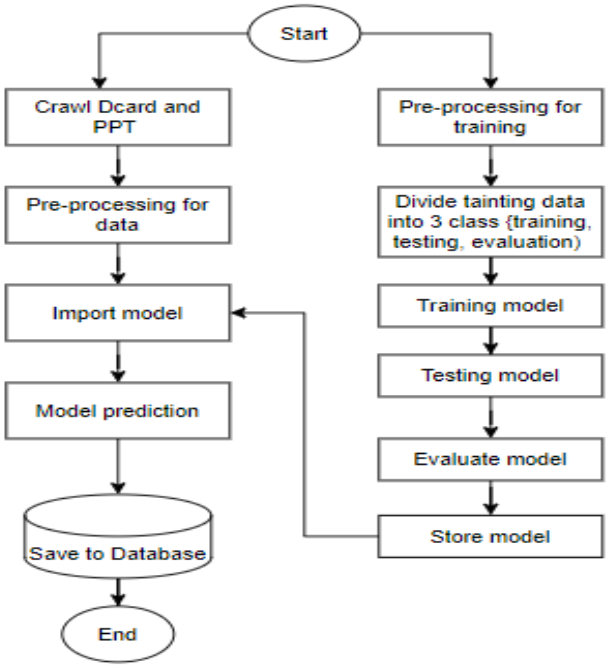

Figure 11 Positive and Negative analysis flow chart

When training the model, the model will also be evaluated and the current evaluation results will be saved. To prevent the model from entering a state of overfitting, early stopping technology is used. The performance of the model will be regularly evaluated against the test data set and the model will stop improving. Training will be terminated when testing data, and weights and biases are used to visualize model training, and hyper-parameter values, losses, evaluation indicators, etc. will be recorded [25]. In the parameter adjustment process, common

parameters such as train batch size, test batch size, epochs, etc., are adjusted after certain training numbers [26]. Model evaluation indicators: Table 2 shows the four elements of the confusion matrix: True Positive (TP): That is, the actual value is True and the prediction is Positive. The predicted result is the same as the actual situation. True Negative (TN): That is, the actual value is True and the prediction is Negative. The predicted result is the same as the actual situation. False Positive (FP) "False positive": that is, the actual value is False and the prediction is Positive. The predicted results are different from the actual situation. False Negative (FN) "False negative": that is, the actual value is False and the prediction is Negative.

Table 2. Confusion matrix

		True Condition	
	Total Population(T)	Positive	Negative
Predicted outcome	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

Here, we adopt the accuracy, recall, precision, F1-score and Matthews's correlation coefficient to evaluate the performance in a mathematical sense.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{F1} - \text{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

$$\text{Matthews correlation coefficient} = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

From the above equations, we can see that the accuracy rate is the percentage of correct prediction results in the total sample. Although it can be used to judge the overall accuracy rate, it cannot be used as a good indicator to measure the results when the sample is unbalanced; accuracy rate is the probability of a sample that is actually positive among all the samples that are predicted to be positive, representing the accuracy of the prediction of the positive sample result; the recall rate is the probability of being predicted to be a positive sample among all the samples that are positive; F1-Score is the harmonic average of precision and recall. It is affected by Precision and Recall at the same time. Simply pursuing the improvement of Precision and Recall does not have much effect, but some information can still be seen from Precision and Recall. The former the higher, the latter the lower it can be used as a more cautious model, and the opposite is a loose model [27]; the Matthews correlation coefficient (Matthews Correlation Coefficient, MCC) takes into account true and false positives and false negatives, and is usually used Treated as a balance, returns a value between -1 and +1. A coefficient of +1 indicates a perfect prediction, 0 indicates no better than the random prediction, and -1 indicates complete inconsistency between prediction and observation [28]. Figure 12 shows the parameters of the model adjusted, and the best model is used for evaluation. Figure 13 shows the results obtained by actually predicting text. Notice that 1 denotes positive and 0 denotes negative in the predicted column.

```
('mcc', 0.9709388865849985)
('tp', 3172)
('tn', 3135)
('fp', 50)
('fn', 43)
('accuracy', 0.98546875)
('precision', 0.9844816883923029)
('recall', 0.9866251944012442)
('f1', 0.9855522759049247)
```

Figure 12 Parameters of model evaluation

Table 3. Prediction results

Text	Prediction Results
今天天氣真好	1

(The weather is so nice today.)	
感覺好累,運氣好差 (Feeling so tired and so unlucky.)	0
希望能夠順利 (Hope it goes well.)	1
好想放棄一切,但是這樣是不行的 (I really want to give up everything, but this won't work.)	0
如果可以,希望能重來,可是過去的就過去了 (If possible, I hope I can do it again, but the past is the past)	1

Use keywords to search the database, count the number of posts, likes, and comments of PTT and Dcard, and use them as discussion points and display them in a line chart. The part of the line chart is displayed on the trend analysis webpage, in daily units, and can be compared and observed, respectively daily to find discussion trends from the two data sources. That allows users to compare which forum has a higher proportion currently it will also show whether there is a gap in the tags and topics that different age groups pay attention to.
The following Figure 13 was clipped from the “UNC police” account on Twitter. Eyes were masked in this paper. We can see the post content and the photo of a suspected man. The man shown in this post was armed and dangerous so everyone should be careful to keep their safety. However, not everyone has been following the UNC police Twitter account. This research is to crawl the data from source webs to find the emergency event and then, send this event information including crime type and photo to related agencies like schools, community, transportation, and street monitor guarding systems. Here, the guarding system should equipped with the function of face recognition to achieve the intelligent alerting goal.

Figure 13 A post on Twitter

When the guarding systems receive emergency information from our system, such as a person missing, gun shooting, rape, etc., it can remind everyone to pay attention when the alerting system issues the dangerous people found by the face recognition function in the guarding system. The setting alarm event labels include firearm, fight, sexual assault, missing, traffic accident, drug, homicide, stealing, kidnap, robbery, fraud, swindle, and intimidation.
As to face recognition, using dlib to first extract 68 feature points of the face, then calculate the 128-dimensional vector of the face, and use OpenCv technology to mark the user's face by turning on the camera or uploading the picture, and comparing it with the face file in the database. For comparison, Euclidean distance will be used as shown in equation (6). x_1 and y_1 are the face coding values to be compared, x_2 and y_2 are the face coding values in the database, and the shortest distance is obtained, the value will be near to 0 when it is close, and if it is very close to 1, it is not similar at all. The test results are more in line with expectations, and it is not easy to make recognition errors.

$$p = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (6)$$

Figure 14 is a photo of a man used for image face recognition testing. Figure 15 is the first placed in the database as a standard image file and named "dy". Figure 16 is the result after calculating the Euclidean distance. If it is less than 0.46, mark "dy" and display a red box, as shown in Figure17. Figure 18 is the output result of testing using photos of multiple people. Since these people do not have images stored in the database, the calculated Euclidean distance will be greater than 0.46. Figure19 shows the values of the Euclidean distance. Notice that, faces will be marked with a green frame once they are not in the database.

The function will be used in this system. When the crawled posts on the social network sites and detected as alerting events, the photos in the posts will be uploaded to the distributed guarding systems to remind the staff when the cameras discover the suspicious people passing by.

Figure14 Testing photo

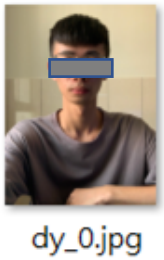

Figure 15 The image in the Database

```
[('dy', 0.35975999835322464), ('yg', 0.
4824569383959438), ('yz', 0.60320839164
89924), ('yj', 0.6076847144831533), ('s
h', 0.6141244538231302), ('cxr', 0.6707
487913536123), ('wf', 0.719392465060369
6), ('yl', 0.7334593909816222), ('syh',
 0.7370615363147626), ('cl', 0.74034981
33650678), ('xyc', 0.7455176607549932),
 ('yx', 0.7471106953528801), ('AngelaBa
by', 0.7819697031053551)]
```

Figure16 Calculation p result of Figure 15

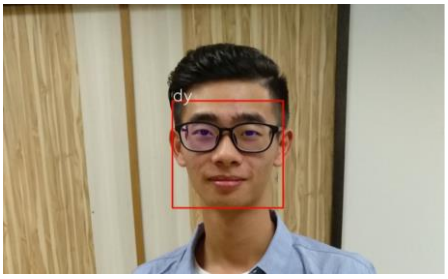

Figure17 Marked red frame

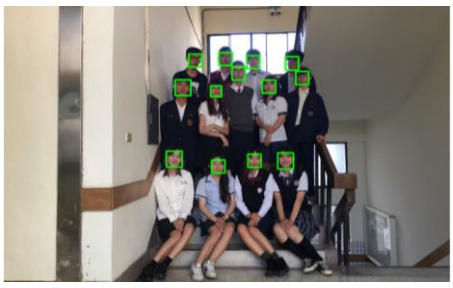

Figure 18 Face recognition.

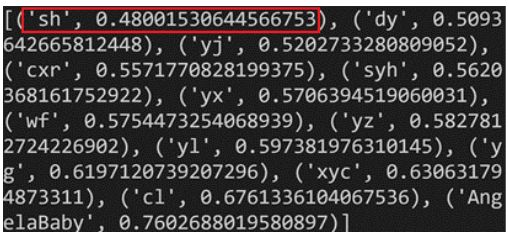

Figure 19 Calculation of p values in Figure 18.

4) Conclusions

The internet is a vast source of information, but not all of it is reliable or relevant. Sometimes, important events are overlooked or distorted by biased media outlets or online posts. To help users avoid being influenced by one-sided perspectives, this system aims to integrate news and forum content with AI deep learning technology. This way, users can gain a deeper understanding of the issues and the public opinions on them. Moreover, the system can also alert users to dangerous incidents in their vicinity through a warning system that uses named entity recognition to identify locations and events. This can enhance public safety and awareness. The system can also assist police agencies in tracking suspects or finding missing persons by using community cameras or warning systems. However, the system respects the personal privacy of the users and does not collect or disclose any personal information. The system's performance depends on the quality and quantity of the data in the corpus. If there are not enough emotional words, or if the data does not match the news content or the recognized text words, the prediction results may be inaccurate.

References.

1.輿情、網絡輿情、輿情分析
https：//www.itread01.com/content/1534149490.html
2.Google alerts-用關鍵字訂閱，打造專屬自己的報紙
https：//blog.dcplus.com.tw/market/google-alerts
3.網路溫度計-時事網路大數據分析
https：//dailyview.tw/
4.Meltwater-全球媒體及社群監測平台
https：//blog.dcplus.com.tw/market/meltwatertool
5.傳統 MVC 模式與 Django MTV 模式介紹與比較
https://mropengate.blogspot.com/2015/08/mvcdjangomtv.html
6.Django 介紹
https://developer.mozilla.org/zh-TW/docs/Learn/Server-side/Django/Introduction
7.科普：常見的資料庫管理系統有哪些?MySQL有哪些特點？
https://kknews.cc/zh-tw/code/xpxajj9.html
8.「不是工程師」後端服務的根基，淺談SQL關連式資料庫 RDBMS
https://progressbar.tw/posts/263
9.APScheduler 3.7.0
https://pypi.org/project/APScheduler/
10.BERT Github
https://github.com/google-research/bert
11.[Notes] BERT / BERT 架構理解語料庫
https://haren.medium.com/paper-notes-bert-bert-%E6%9E%B6%E6%A7%8B%E7%90%86%E8%A7%A3-31c014d7dd63
12.Transformers Github
https://huggingface.co/docs/transformers/index
13.Simple Transformers Github
https://github.com/ThilinaRajapakse/simpletransformers
14.Simple Transformers Classification
https://simpletransformers.ai/docs/classification-specifics/
15.Open Sentiment Training Data
https://github.com/UDICatNCHU/UdicOpenData
16.中時新聞網
https://www.chinatimes.com/search/%E9%AC%A5%E6%AF%86?&chdtv

17.自由時報
https://search.ltn.com.tw/list?keyword=%E6%AE%BA%E4%BA%BA
18.ettoday新聞雲
https://www.ettoday.net/news/focus/%E7%A4%BE%E6%9C%83/
19.聯合新聞網
https://udn.com/news/archive?
20.Dlib介紹
https://www.itread01.com/content/1547287507.html
21.OpenCV 介紹
https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/541606/
22.Multi-Label Classification using BERT, RoBERTa, XLNet, XLM, and DistilBERT with Simple Transformers
https://towardsdatascience.com/multi-label-classification-using-bert-roberta-xlnet-xlm-and-distilbert-with-simple-transformers-b3e0cda12ce5
23.Simple Transformers — Introducing The Easiest Way To Use BERT, RoBERTa, XLNet, and XLM
https://towardsdatascience.com/simple-transformers-introducing-the-easiest-bert-roberta-xlnet-and-xlm-library-58bf8c59b2a3
24.[Python機器學習]-自動判斷留言正負評(運用BERT model）with Colab pro GPU
https://medium.com/@ethan.chen927/python%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-google%E6%88%91%E7%9A%84%E5%95%86%E5%AE%B6%E6%AD%A3%E8%B2%A0%E8%A9%95%E7%8C%9C%E6%B8%AC-%E9%81%8B%E7%94%A8bert-model%E7%90%86%E8%A7%A3%E4%B8%8A%E4%B8%8B%E6%96%87%E7%9A%84%E8%AA%9E%E8%A8%80%E6%A8%A1%E5%9E%8B-with-colab-pro-gpu-ec8ef8cb8a25
25.Simple Transformers Using Early Stopping and Visualization Support
https://simpletransformers.ai/docs/tips-and-tricks/
26.Simple Transformers General Usage
https://simpletransformers.ai/docs/usage/
27.機器學習模型評估指標-confusion matrix, precision, recall, and ,F1-score
https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E6%A8%A1%E5%9E%8B%E8%A9%95%E4%BC%B0%E6%8C%87%E6%A8%99-confusion-matrix-precision-and-recall-e9d64ff14d81
28.馬修斯相關係數（Matthews correlation coefficient）
https://www.twblogs.net/a/5bbce7b12b71776bd30bb39c