SARCASTIC AND NON-SARCASTIC TWEET CLASSIFICATION USING DEEP LEARNING

Dilli Babu. M
Associate Professor
Department of Information Technology,
Panimalar Engineering College, Chennai.
deenshadilli@gmail.com

Benny Richards. R
Student
Department of Information Technology,
Panimalar Engineering College, Chennai.
richards12102@gmail.com

Allen Harris. A
Student
Department of Information Technology,
Panimalar Engineering College, Chennai.
allensushi24@gmail.com

Magesh. S
Student
Department of Information Technology,
Panimalar Engineering College, Chennai.
mageshchitra121@gmail.com

***Abstract*— Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text. While speaking, people often use heavy tonal stress and certain gestural clues like rolling of the eyes, hand movement, etc. to reveal sarcastic. In the textual data, these tonal and gestural clues are missing, making sarcasm detection very difficult for an average human. Due to these challenges, researchers show interest in sarcasm detection of social media text, especially in tweets. Rapid growth of tweets in volume and its analysis pose major challenges. In this project, we proposed a machine learning based framework that captures real time tweets and processes it with a set of algorithms which identifies sarcastic sentiment effectively. We observe that the elapse time for analyzing and processing under ML based framework significantly outperforms the conventional methods and is more suited for real time streaming tweets.**
In our project we will be using Convolution Neural Network (CNN) + Long Short Term Memory (LSTM) as existing and Tree Convolution Neural Network (TCNN) as proposed system. From the result its proved that proposed Tree Convolution Neural Network (TCNN) works better than existing Convolution Neural Network (CNN) + Long Short Term Memory (LSTM) in terms of accuracy.

***Index Terms*— Accuracy, Precision, Recall, TCNN, LSTM, Tweets**

I. INTRODUCTION

In recent years, hate speech has been increasing in-person and online communication. The social media as well as other online platforms are playing an extensive role in the breeding and spread of hateful content – eventually which leads to hate crime. For example, according to recent surveys, the rise in online hate speech content has resulted in hate crimes including Trump's election in the US, the Manchester and London attacks in the UK, and terror attacks in New Zealand. To tackle these harmful consequences of hate speech, different steps including legislation have been taken by the European Union Commission. Recently, the European Union Commission also enforced social media networks to sign an EU hate speech code to remove hate speech content within 24 hours.

However, the manual process to identify and remove hate speech content is labor-intensive and time-consuming. Due to these concerns and widespread hate speech content on the internet, there is a strong motivation for automatic hate speech detection. The automatic detection of hate speech is a challenging task due to disagreements on different hate speech definitions. Therefore, some content might be hateful to some individuals and not to others, based on their concerned definitions. According hate speech is: "the content that promotes violence against individuals or groups based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity". Despite these different definitions, some recent studies claimed favorable results to detect automatic hate speech in the text. The proposed solutions employed the different feature engineering techniques and ML algorithms to classify content as hate speech. Regardless of this extensive amount of work, it remains difficult to compare the performance of these approaches to classify hate speech content. To the best of our knowledge, the existing studies lack the comparative analysis of different feature engineering techniques and ML algorithms.

II. LITERATURE SURVEY

Social media has been an important way for people to get news. It is designed to make the sharing of messages very fast and easy. It also attracted the attention of a large number of researchers. There has been research concerning predicting what messages will be popular. But it lacks of in-depth study of what features play an important role in the prediction task. In the work, we systematically and comprehensively study three types of features: user features, text features and time features. Multiple comparison experiments are carried out on big data platform. Experimental results show that time features are the most valuable features, almost close to the effect of all the features, and the popularity of messages is predicted with a satisfactory accuracy.

Since social media is one of the most likable products of technpeople get easier to express their opinions. Anyone be able to tell their opinion freely there. Unfortunately, its convenience has also become a boomerang for us, the easier every opinion conveyed the easier hate speech is expressed. This matter become the dark side of social media. Hate speech

face us with a lot of dangers, such as violence, social conflict, even homicide. Therefore, preventing all of those dangers that might be occur because of hate speech is one of the prior things we need to do. This research was done as an attempt to take care of the dangers that could be done by hate speech. The attempt we tried to do is using multi-label text classification to predict hate speech with the Bidirectional Long Short-term Memory (BiLSTM) method. This multi-label text classification labelled every tweet in the dataset crawled from Twitter with 12 labels about hate speech. From this experiment, we obtained the best hyperparameter value that could achieve great performance with 82.31% accuracy, 83.41% precision, 87.28% recall, and 85.30% F1-score.

Social media has long been a popular resource for sentiment analysis and data mining. In this paper, we learn to predict reader interest after article reading using social interaction content in social media. The abundant interaction content (e.g., reader feedback) aims to replace typically private reader profile and browse history. Our method involves estimating interest preferences with respect to article topics and identifying quality social content concerning informativity. During interest analysis, we combine and transform articles and their reader responses into PageRank word graph to balance author- and reader-end influence. Semantic features of words, such as their content sources (authors vs. readers), syntactic parts-of-speech, and degrees of references (i.e., significances) among authors and readers, are used to weight PageRank word graph. We present the prototype system, Interest Finder, that applies the method to reader interest prediction by calculating word interestingness scores. Two sets of evaluation show that traditional, local Page Rank can more accurately cover more span of reader interest with the help of topical interest preferences learned globally, word nodes' semantic information, and, most important of all, quality social interaction content such as reader feedback.

Computational methods to model political bias in social media involve several challenges due to heterogeneity, high-dimensionality, multiple modalities, and the scale of the data. Most of the current political bias detection methods rely heavily on the manually-labeled ground-truth data for the underlying political bias prediction tasks. Such methods are human-intensive labeling, labels related to only a specific problem, and the inability to determine the near future bias state of a social media conversation. In this work, we address such problems and give Deep learning approaches to study political bias in two ideologically diverse social media forums: Gab and Twitter without the availability of human-annotated data. We propose a method to exploit the features of entities on transcripts collected from political speeches in US congress to label political bias of social media posts automatically without any human intervention. With existing Deep learning algorithms we achieve the highest accuracy of 70.5% and 65.1% to predict posts on Twitter and Gab data respectively. We also present a Deep learning approach that combines features from cascades and text to forecast cascade's political bias with an accuracy of about 85%.

Given a tweet, predicting the discussions that unfold around it is convoluted, to say the least. Most if not all of the discernibly benign tweets which seem innocuous may very well attract inflammatory posts (hate speech) from people who find them non-congenial. Therefore, building upon the aforementioned task and predicting if a tweet will incite hate speech is of critical importance. To stifle the dissemination of online hate speech is the need of the hour. Thus, there have been a handful of models for the detection of hate speech. Classical models work retrospectively by leveraging a reactive strategy – detection after the postage of hate speech, i.e., a backward trace after detection. Therefore, a benign post that may act as a surrogate to invoke toxicity in the near future, may not be flagged by the existing hate speech detection models. In this paper, we address this problem through a proactive strategy initiated to avert hate crime. We propose DRAGNET, a deep stratified learning framework which predicts the intensity of hatred that a root tweet can fetch through its subsequent replies. We extend the collection of social media discourse from our earlier work [1], comprising the entire reply chains up to ∼5k root tweets catalogued into four controversial topics Similar to [1], we notice a handful of cases where despite the root tweets being non-hateful, the succeeding replies inject an enormous amount of toxicity into the discussions. DRAGNET turns out to be highly effective, significantly outperforming six state-of-the-art baselines. It beats the best baseline with an increase of 9.4% in the Pearson correlation coefficient and a decrease of 19% in Root Mean Square Error. Further, DRAGNET'S deployment in Logically's advanced AI platform designed to monitor real-world problematic and hateful narratives has improved the aggregated insights extracted for understanding their spread, influence and thereby offering actionable intelligence to counter them

Automatic detection of abusive online content such as hate speech, offensive language, threats, etc. has become prevalent in social media, with multiple efforts dedicated to detecting this phenomenon in English. However, detecting hatred and abuse in low-resource languages is a non-trivial challenge. The lack of sufficient labeled data in low-resource languages and inconsistent generalization ability of transformer-based multilingual pre-trained language models for typologically diverse languages make these models inefficient in some cases. We propose a meta learning-based approach to study the problem of few-shot hate speech and offensive language detection in low-resource languages that will allow hateful or offensive content to be predicted by only observing a few labeled data items in a specific target language. We investigate the feasibility of applying a meta learning approach in cross-lingual few-shot hate speech detection by leveraging two meta learning models based on optimization-based and metric-based (MAML and Proto-MAML) methods. To the best of our knowledge, this is the first effort of this kind. To evaluate the performance of our approach, we consider hate speech and offensive language detection as two separate tasks and make two diverse collections of different publicly available datasets comprising 15 datasets across 8 languages for hate speech and 6 datasets across 6 languages for offensive language. Our experiments show that meta learning-based models

outperform transfer learning-based models in a majority of cases, and that Proto-MAML is the best performing model, as it can quickly generalize and adapt to new languages with only a few labeled data points (generally, 16 samples per class yields an effective performance) to identify hateful or offensive content.

As social media grew in popularity among the general public, content and opinion sharing has become rapid and convenient. The majority of users rely on social media for news and trust the content shared by the network. Some individuals, both purposefully and unintentionally, disseminate hate content and instill hatred in their readers. Even in Sri Lanka, the spread of hate propaganda on social media has resulted in communal discord and a variety of concerns. Only a limited number of research studies have been conducted to analyze the hate content written in Sinhala. This work investigates a mechanism for detecting hate content typed in Sinhala language and posted on Twitter. The proposed supervised mechanism is an ensemble method that selects the most accurate result from different models. 63% of accuracy, 58% of F1 Score, 61% of Precision and 58% of Recall were achieved when predicting hate content.

There is an enormous growth of social media which fully promotes freedom of expression through its anonymity feature. Freedom of expression is a human right but hate speech towards a person or group based on race, caste, religion, ethnic or national origin, sex, disability, gender identity, etc. is an abuse of this sovereignty. It seriously promotes violence or hate crimes and creates an imbalance in society by damaging peace, credibility, and human rights, etc. Detecting hate speech in social media discourse is quite essential but a complex task. There are different challenges related to appropriate and social media-specific dataset availability and its high-performing supervised classifier for text-based hate speech detection. These issues are addressed in this study, which includes the availability of social media-specific broad and balanced dataset, with multi-class labels and its respective automatic classifier, a dataset with language subtleties, dataset labeled under a comprehensive definition and well-defined rules, dataset labeled with the strong agreement of annotators, etc. Addressing different categories of hate separately, this paper aims to accurately predict their different forms, by exploring a group of text mining features. Two distinct groups of features are explored for problem suitability. These are baseline features and self-discovered/new features. Baseline features include the most commonly used effective features of related studies. Exploration found a few of them, like character and word n-grams, dependency tuples, sentiment scores, and count of 1st, 2nd person pronouns are more efficient than others. Due to the application of latent semantic analysis (LSA) for dimensionality reduction, this problem is benefited from the utilization of many complex and non-linear models and CAT Boost performed best. The proposed model is compared with related studies in addition to system baseline models. The results produced by the proposed model were much appreciating.

III EXISTING SYSTEM

Convolution Neural Network (CNN):

In deep learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as **shift invariant** or **space invariant artificial neural networks** (**SIANN**), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition, recommender systems, image classification, Image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.

CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage.

Long Short Term Memory (LSTM):

Long short-term memory (**LSTM**) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feed forward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDSs (intrusion detection systems).

A common LSTM unit is composed of a **cell**, an **input gate**, an **output gate** and a **forget gate**. The cell remembers values over arbitrary time intervals and the three *gates* regulate the flow of information into and out of the cell.

LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.

IV. METHODOLOGY

Tree Convolution Neural Network (Tree CNN):
A Convolution Neural Network (CNN) is a class of artificial neural networks where connections between nodes form a graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from feed forward neural networks, CNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. The term "Enhanced neural network" is used indiscriminately to refer to two broad classes of networks with a similar general structure, where one is finite impulse and the other is infinite impulse. Both classes of networks exhibit temporal dynamic behavior. A finite impulse recurrent network is a directed acyclic graph that can be unrolled and replaced with a strictly feed forward neural network, while an infinite impulse recurrent network is a directed cyclic graph that cannot be unrolled.

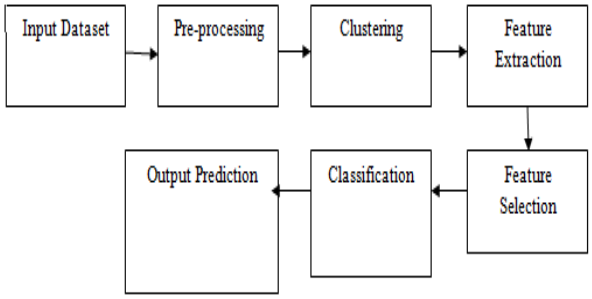

Fig 1: Architecture Diagram

MODULE DESCRIPTION

A. Twitter API
User needs to register first by giving his/her own information. While registering user should give their exact current location. If he/she is giving wrong location means, he is not supposed to register and login. And that user will be considered as a blocked user. So user needs to give only the current location. After registration user will login with username and password. Then he/she can see their profile and can view all user tweets. Admin needs to login with username and password. If both match, he/she will be considered as a valid person. After login, admin can view all blocked user who gave wrong location while registration. Admin can able to see all users profile and tweets.

B. Post contents
In this module registered user can post tweets in instagram. If the user tries to post any content which contains bad words means, it will not get posted in the instagram account. So the algorithm will restrict the user not to post bad words. The general tweets can be posted in application and as well as in twitter.

C. Search Query
Here in this module, user can search for any query in the application. The query has been processed and extracted live tweets from the real time twitter. The Keywords related 100 tweets are extracted from the live twitter.

D. Preprocessing
In this step all the tweets are extracted from twitter are processed and the noise data are removed.

1) **Stop words Removal:** A dictionary based approach is been utilized to remove stop words from tweets. A generic stop word list containing 75 stop words created using hybrid approach is used. The algorithm is implemented as below given steps. The target text is tokenized and individual words are stored in array. A single stop word is read from stop word list. The stop word is compared to target text in form of array using sequential search technique. If it matches, the word in array is removed, and the comparison is continued till length of array. After removal of stop word completely, another stop word is read from stop word list and again algorithm runs continuously until all the stop words are compared. Resultant text devoid of stop words is displayed, also required statistics like stop word removed, no. of stop words removed from target text, total count of words in target text, count of words in resultant text, individual stop word count found in target text is displayed.

2) **Stemming Technique:** After removing the unwanted words from the tweet, stemming technique is processed. Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

E. Classification
After stemming process, all the tweet terms containing the keyword are classified into positive, negative and neutral tweets. Tree CNN is used for classification. Here we are having good words and bad words datasets. By comparing with this, we can classify the tweets into positive, negative and neutral tweets.

V. TOOLS USED

OpenCV is a storage of programming operations for actual time computer scope actually created by Intel and now assisted by Willogarage. It is liberately used under the free source BSD license. This has greater than five hundred effective set of rules to be followed. It is widely employed around the world, with forty thousand users in the user community. Used in wide limit ranging from communicating resource, to fine audit, and upcoming robotics. The package is developed in C, which does it movable to few particular surface such as Digital Signal

Processor. Packaging for languages such as C, Python, Ruby and Java (using JavaCV)

A. Python

Python is a remarkably powerful dynamic, object-oriented programming language that is used in a wide variety of application domains. It offers strong support for integration with other languages and tools, and comes with extensive standard libraries. To be precise, the following are some distinguishing features of Python:

- Very clear, readable syntax.
- Strong introspection capabilities.
- Full modularity.
- Exception-based error handling.
- High level dynamic data types.
- Supports object oriented, imperative and functional programming styles.
- Embeddable.
- Scalable
- Mature

With so much of freedom, Python helps the user to think problem centric rather than language centric as in other cases. These features makes Python a best option for scientific computing.

B. Open CV

Open CV is a library of programming functions for real time computer vision originally developed by Intel and now supported by Willogarage. It is free for use under the open source BSD license. The library has more than five hundred optimized algorithms. It is used around the world, with forty thousand people in the user group. Uses range from interactive art, to mine inspection, and advanced robotics. The library is mainly written in C, which makes it portable to some specific platforms such as Digital Signal Processor. Wrappers for languages such as C, Python, Ruby and Java (using Java CV) have been developed to encourage adoption by a wider audience. The recent releases have interfaces for C++. It focuses mainly on real-time image processing. Open CV is a cross-platform library, which can run on Linux, Mac OS and Windows. To date, Open CV is the best open source computer vision library that developers and researchers can think of.

C. Tesseract

Tesseract is a free software OCR engine that was developed at HP between 1984 and 1994. HP released it to the community in 2005. Tesseract was introduced at the 1995 UNLV Annual Test OCR Accuracy and is currently developed by Google released under the Apache License. It can now recognize 6 languages, and is fully UTF8 capable. Developers can train Tesseract with their own fonts and character mapping to obtain perfect efficiency.

VI SIMULATION RESULT

The Accuracy graph is plotted between CNN+LSTM and TCNN. Accuracy graph is given in the following figure 2. The model performed very well on both on the training and test dataset. There by showing an accuracy of 95.8% as shown by figure 3. The novel model also exhibited a negligible loss of less than one percent on both on the know and the unknown data as shown in Fig 4.

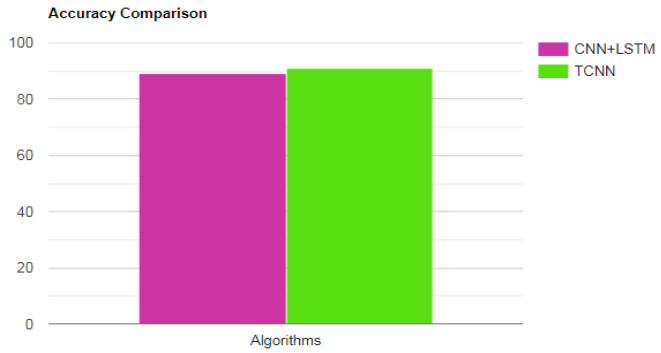

Fig 2: Accuracy Analysis

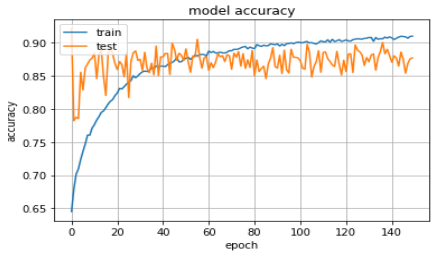

Fig 3. Model Accuracy on train and test data

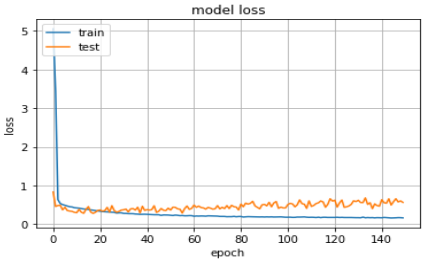

Fig 4. Model loss on train and test data

VII CONCLUSION

The experiments conducted on the Twitter dataset demonstrate the effectiveness of the two proposed models. The prediction ability of the proposed model is further verified on the opinion word prediction task. Based on the learned influence, we explore the expression styles of users with different influence powers, which provide the valuable information for companies to manage their accounts and design marketing plans. Proposed approach can be used to extract both facts and opinions from social media content. It uses two different cross domain datasets to analyze the sentiment of another domain. Most of the time the rating of specific reviews plays an important role for sentiment analysis. Hence the accuracy of the sentiment polarity will be very high.

REFERENCES

[1] Daniele Cenni, Paolo Nesi, Gianni Pantaleo, Imad Zaza., "Twitter vigilance: A multi-user platform for cross-domain Twitter data analytics, NLP and sentiment analysis", In Proceedings of the IEEE International Conference.

[2] H. Sankar, V. Subramaniyaswamy, "Investigating Sentiment Analysis Using Deep Learning Approach",

International Conference on Intelligent Sustainable Systems (ICISS) (2017).
[3] Lavika Goel, Anurag Prakash, “Sentiment Analysis of Online Communities Using Swarm Intelligence Algorithms”, 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN).
[4] Lu Ma, Dan Zhang, Jian-wu Yang, Xiong Luo., “Sentiment Orientation Analysis Of Short Text Based On Background And Domain Sentiment Lexicon Expansion”, 2016 5th International Conference on Computer Science and Network Technology (ICCSNT).
[5] Shokoufeh Salem Minab, Mehrdad Jalali, Mohammad Hossein Moattar, “Online Analysis Of Sentiment On Twitter”, 2015 International Congress on Technology, Communication and Knowledge (ICTCK).
[6] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen Xiaofei He, "Interpreting The Public Sentiment Variations On Twitter", IEEE Transactions on Knowledge and Data Engineering (Volume: 26 , Issue: 5 , May 2014).
[7] Desheng Dash Wu, Lijuan Zheng, David L. Olson, “A Decision Support Approach For Online Stock Forum Sentiment Analysis”, IEEE Transactions on Systems, Man, and Cybernetics: Systems (Volume: 44 , Issue: 8 , Aug. 2014).
[8] Oussalah M, Bhat F, Challis K, Schnier T. A software architecture for Twitter collection, search and geolocation services, In Knowledge-Based Systems. Vol. 37, pp.105-120, 2013.
[9] Alexandre Trilla, Francesc Alias, "Sentence- Based Sentiment Analysis For Expressive Text-To-Speech, IEEE Transactions on Audio, Speech, and Language Processing (Volume: 21 , Issue: 2 , Feb. 2013).
[10] Ruhi U., Social Media Analytics as a Business Intelligence Practice: Current Landscape & Future Prospects, In Journal of Internet Social Networking & Virtual Communities, 2014.
[11] K. Young, M. Pistner, J. O’Mara, and J. Buchanan. Cyber-disorders: The mental health concern for the new millennium. Cyberpsychol. Behav., 2019..

[12] J. Block. Issues of DSM-V: internet addiction. American Journal of Psychiatry, 2019.

[13] K. Young. Internet addiction: the emergence of a new clinical disorder, Cyberpsychol. Behav., 2019.

[14] I.-H. Lin, C.-H. Ko, Y.-P. Chang, T.-L. Liu, P.-W. Wang, H.-C. Lin, M.-F. Huang, Y.-C. Yeh, W.-J. Chou, and C.-F. Yen. The associa- tion between suicidality and Internet addiction and activities in Taiwanese adolescents. Compr. Psychiat., 2019.

[15] Y. Baek, Y. Bae, and H. Jang. Social and parasocial relationships on social network sites and their differential relationships with users’ psychological well-being. Cyberpsychol. Behav. Soc. Netw., 2019.

[16] D. La Barbera, F. La Paglia, and R. Valsavoia. Social network and addiction. Cyberpsychol. Behav., 2019.

[17] K. Chak and L. Leung. Shyness and locus of control as predictors of internet addiction and internet use. Cyberpsychol. Behav., 2019.

[18] K. Caballero and R. Akella. Dynamically modeling patients health state from electronic medical records: a time series approach. KDD, 2019.

[19] L. Zhao and J. Ye and F. Chen and C.-T. Lu and N. Ramakr- ishnan. Hierarchical Incomplete multi-source feature learning for Spatiotemporal Event Forecasting. KDD, 2019.

[20] E. Baumer, P. Adams, V. Khovanskaya, T. Liao, M. Smith, V. Sosik, and K. Williams. Limiting, leaving, and (re)lapsing: an exploration of Facebook non-use practices and experiences. CHI, 2019.