Popularity Prediction of Social Media based on Multi-Modal Feature Mining

Dhruv Kundu
Jay Rawal
Krit Verma
Rishabh Sharma
{dhruv17146, jay17240, krit17348, rishabh17087}@iiitd.ac.in
Indraprastha Institute of Information Technology
New Delhi, Delhi

ABSTRACT

Content posted on social media platforms are highly driven by the popularity achievability factor associated with it. The solution, to the problem of popularity prediction, can be highly useful in understanding commonalities of user interest on online platforms and their content consumption patterns.
Modern day social media platforms lets it's user to post textual as well as multimedia content. Thus, it's imperative to consider all this multimodal information,in order to accurately predict the popularity of any social media. In this paper, we propose to introduce few novel features along with Flickr, SMD2019 datasets to achieve better results. Also, we compared the performance of Light-GBM and DeepGBM frameworks and have presented the AUC and MSE (mean square error) values.
Finally we reported the result, which were better than the current state-of-the-art model for popularity prediction of online social media.

KEYWORDS

Multi-modal computing, Deep Learning, Ensemble Learning, Social Media

ACM Reference Format:
Dhruv Kundu, Jay Rawal, Krit Verma, and Rishabh Sharma. 2020. Popularity Prediction of Social Media based on Multi-Modal Feature Mining. In *Proceedings of* . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

An immediate increase in the use of social media in the previous decade on various platforms, has given rise to various new interesting problems and challenges. One of the most common concern, which has developed a lot of interest, is being able to predict the popularity of the post. As social media platforms vary, the format of these posts across the domains also varies, as each post consists of multiple data types such as categories, titles, tag, geo-location, meta-data, etc. With this heterogeneous data, the complexity of this task also increases, since it becomes challenging to apply the available machine learning models directly. Therefore, this complex problem requires a new practical approach which incorporates the multi-modal data of social media for popularity prediction.

Numerous machine learning methods have been tried to tackle the social media prediction problem, as [28] proposed that Local Polynomial Regression and Decision Tree algorithms outperform the other methods tested in their research. Various other studies[25] and [27] which used regression models also showed significant results; therefore, this problem can be considered as a regression problem. To further improve the results, we can use ensemble learning approaches rather than the traditional machine learning regression models because of their superior performance in regression. For example, an ensemble regressor approach was used in [8] and [12] for the prediction of social media headline. The use of sentiment features and context features was proposed in [6] to predict the number of views of images which was implemented using Support Vector Machine (SVM) and Convolutional Neural Networks (CNN). A novel prediction framework was called Deep Temporal Context Networks (DTCN) proposed by [26] , for prediction of popularity and their results showed that their DTCN method outperforms all methods used in the TPIC17 dataset previously.

, ,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Various other works have also been done to predict popularity on social media, covering different aspects of popularity such as a study[12] used meta-data and image features, social features based on the random forest to predict the number of views of a post. Results of this approach can further be improved by applying a random forest regressor on meta-data only [8]. A caffe deep learning framework was implemented on Flickr by [13] and they used tag features and visual features to predict popularity and concluded that tag features also outperform all other unimodal approaches. Moreover, a text-based feature engineering approach was proposed in[19], which gave better insights to information. However, all these approaches still don't cover the heterogeneous data comprehensively.

After exploring these approaches for social media prediction, the main issue we observed was that there is a need of an approach which significantly uses all the available multimodal data as a whole single unit , which when taken into account holistically, would be able to predict the popularity accurately of a given post.

2 RELATED WORK

To do this challenge we looked at various multi-modal approaches as well as papers which helped us step forward with each mode of data. So, we carefully looked into the way in which we can boost the features from each mode namely images, text, numerical and categorical data.

In multimodal domain [22], showed that the engagement of user should be an important feature. The features which were given importance were sentiments in the text, brands used in the captions, facial features, filters applied on images etc. A lot of stress has been given not just to the data present, but also to the metadata which help in deriving the contextual information. [20] is one such work which uses user-item context for making a multimodal context-aware tool using factorization machines.

Khosla et al. investigated the importance of color, gradients, deep learning features for image popularity on Flickr [15]. In [5] Gelli et al. proposed a solution to derive 'sentiments' from an image by classifying it into Adjective-Noun-Pairs using a pretrained model [16] and then deriving the sentiments from them using another network. Cappallo el al. [3] introduced the concept of latent visual features which essentially was dealt as an information retrieval problem, which ranked the images. Convolutional neural networks and other deep learning architectures like SAT(Show, Attend and tell) models, ResNets are being used for getting high level features out of images as shown in [9, 23]. Apart from image data, social media also consists of video data which can also be analysed for predicting popularity as done in [2] using attentional networks.

[1] uses textual features to increase the accuracy of prediction. In [18] Li et al. used text features engineering to solve Social Media Headline Prediction problem, where apart from textual content they looked at the metadata and used categorical encoding & Doc2Vec(Document to vector encoder) to improve the accuracy. [26] takes a look at the temporal pattern which helps to predict virality for instagram posts. This is user-centric as it takes a look at a users previous posts as well. [21] emphasizes on training using categorizing social media posts in various categories. Works in [17] uses Natural Language processing techniques such as Parts of Speech, section prediction, Lexical-word embedding in the context of predicting the popularity of a headline to get hidden meaning out of a headline.

One more modality which has come up in recent social media is micro-video formats. These were made popular by snapchat and vine. The dataset is limited to vine, but some works such as [4] has crawled the web to create such a dataset and achieve a low mean squared error.There are some approaches which have applied the state of the art deep learning methods such as residual learning on a pretrained model [11] with random forest regression. Whereas [24] demonstrates the use of a custom made neural network for social media predication.
All of the efforts are being done in order to extract a better and comprehensive representation of knowledge.

3 DETAILS OF BASELINE PAPER

We decided to use paper titled "Popularity Prediction of Social Media based on Multi-Modal Feature Mining" [10] as our baseline paper. The flowchart of the proposed regression model is depicted in Figure 1 above. It uses mutimodal features of an image post to predict popularity prediction.Dataset used was SMHP2019. We are trying to implement similar approach for our own project to get comparable if not better results for the prediction problem.
The image data was passed into a CNN-RNN captioning model with attention. The output of the CNN was compressed using techniques like PCA and the captions generated were passed through a trained word2vec model. The model detected objects using a pretrained VGG16 or ResNet model on MS COCO.
The word2vec model consists of a two-layer shallow neural network which was trained on the vocabulary file of the image captioning model.
The rest of the categorical data was made in a one-hot encoding format and then was eventually concatenated with normalized numerical data. All these features were fed into

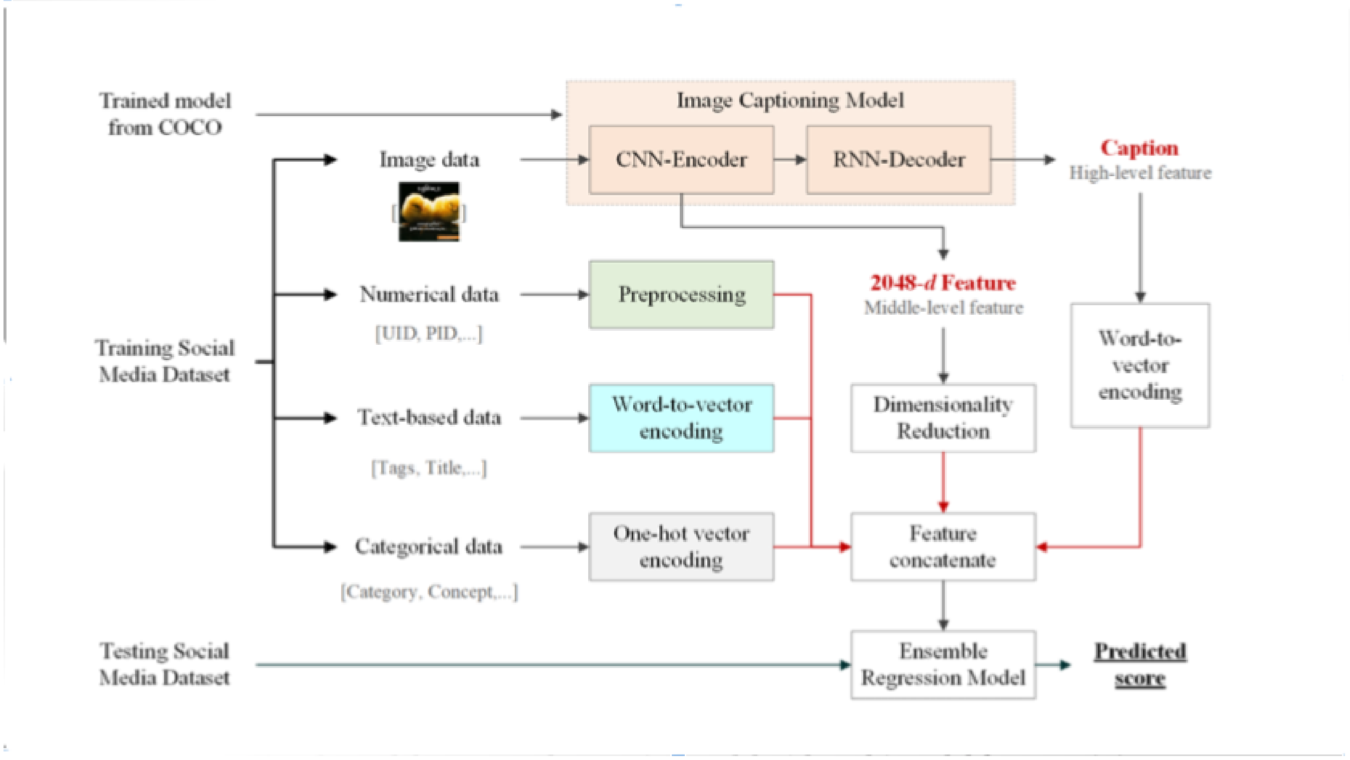

Figure 1: Baseline of the paper

a LightGBM random forest regressor to predict metrics such as MAE, MSE and SRC score on the test set (80 - 20 split). The state-of-the-art results achieved by the above model were given by a Mean Squared error of 3.561, Mean Absolute error of 1.497 and Spearman's Rho rank correlation of 0.656.

4 PROPOSED SOLUTION

After exploring various approaches and extensively studying multiple models, we came up with a proposal to introduce three new factors along with the baseline of the chosen paper. The flowchart of our proposed model is depicted in Figure 2 .

Firstly, we propose to use the feature mining techniques used in the baseline paper. Some of the previous work done in this domain suggested that using information about the user, i.e., the people posting the content would add to the accuracy of prediction[7]. We tried this approach and got the information about the users:

(1) If the user is a PRO user
(2) If the user can buy PRO
(3) The first date when the user posted
(4) Number of posts by the user
(5) User description

We appended these new features to our dataset generated by following the baseline paper.

Secondly, we included the image data from the image EXIF (Exchangeable image file format), which can give us the various insights about the image like, image location, camera model, camera aperture, focal length, etc. This will in-turn gives us understanding about the quality of the multimedia attached with our post, thus making the prediction accuracy higher.

In [9] Chih-Chung Hsu et al. proposed to apply LightGBM to learn the popularity prediction for social media based on multi-modal feature mining techniques. Finally and most importantly, we also propose to compare the performance of LightGBM and DeepGBM[14] . DeepGBM enables us to use sparse categorical, dense numerical and online large-scale data for prediction in our work which yields higher AUC and lower MSE on test data.

5 TIMELINE

We have successfully collected SMHP dataset as our base Flickr dataset. Due to the number of images being so large, we had to use parallel scripts with fragmented dataset, to download all the images.
We have implementedd Image Captioning, to generate the same for all images in our dataset. We used git repository

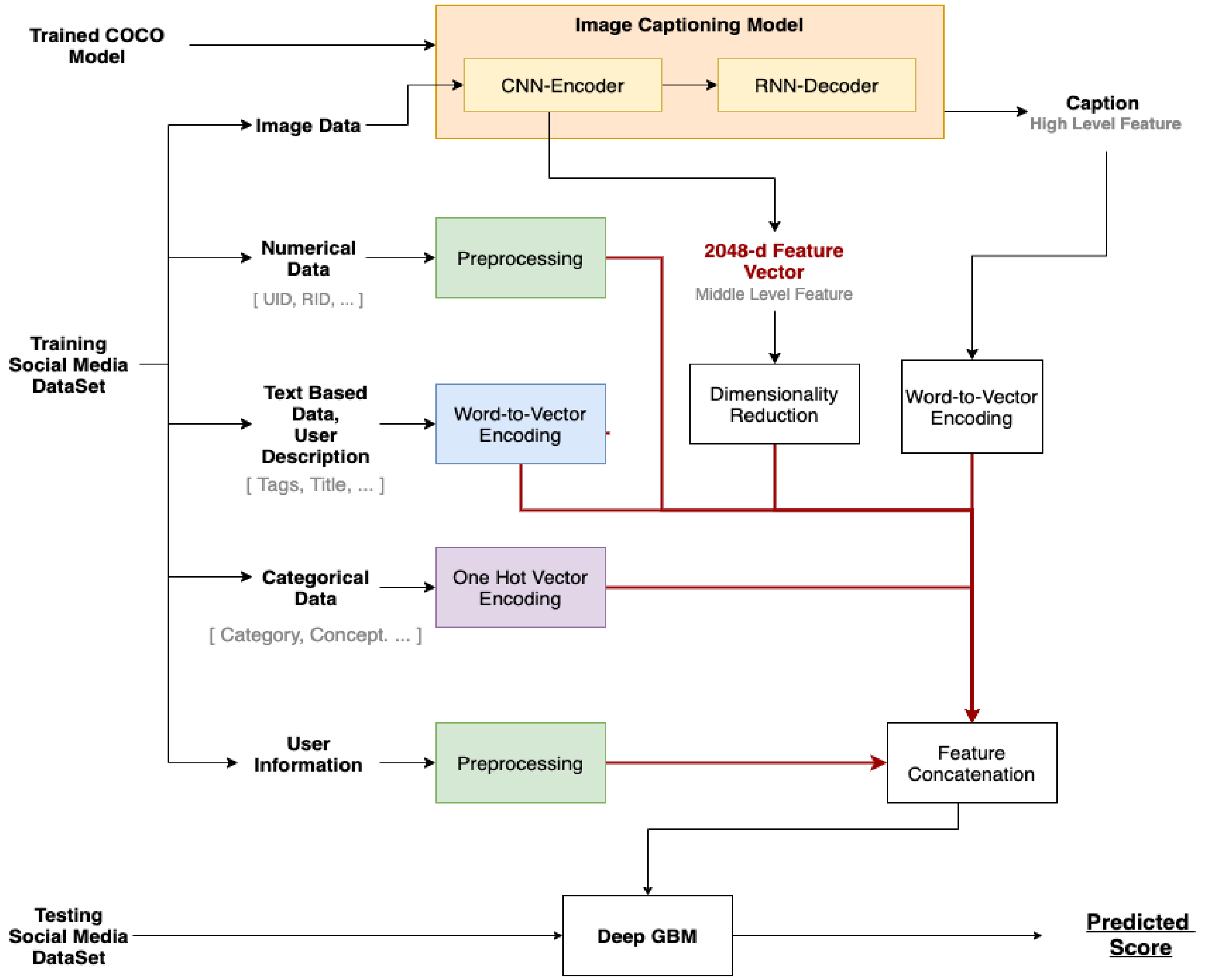

Figure 2: Baseline of Our Approach

titled "image_captioning" by DeepRNN, to implement the same. We generated the captions, post which we have also completed Word2vec model for using the captions in our prediction framework and implementation of the Light GBM to validate our baseline paper's results.
We requested the code source of our baseline paper, titled "Popularity Prediction of Social Media based on Multi-Modal Feature Mining" [10] but due lack of any response from the authors, we had to implement baseline model from scratch and have provided it in our github repository, along with all the previous mentioned work. The link for the same have been submitted along with this report. Validation of the baseline results was completed and the results were reported as well. Finally we have reported our findings in our paper. Then we also added our novel features along with textual and categorical feature mining to the Flickr, SMD2019 dataset. Finnally, we have also trained our DeepGBM model with this newly created database and have reported the results.

6 BASELINE REPRODUCTION AND RESULTS

We were able to use SMHP dataset which includes PID, UID, URL of images, metadata, timeflag etc, as shown in the Figure ??. Further, the entire image data was passed into a CNN-RNN model with attention which generated the captions of the images as shown in Figure 3.

After getting the results from the caption model, we generated the vector representation of the captions by using Word2Vec model. After that we concatenated all the features and passed them into the LightGBM regressor. The LightGBM configuration consists of 27 leaves with bagging fraction of 0.9 to reduce overfitting. Standard gradient boosted decision trees is used to train with ’l1’ and ’l2’ loss on 20 iterations.

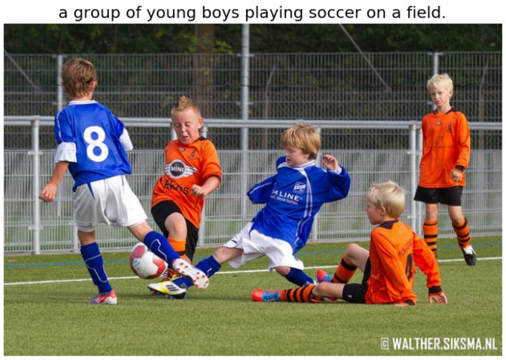

Figure 3: Images with captions Generated from the Model

The whole data comprises of around 300k+ images with various features as mentioned above. To effectively verify the results of our model, we further split our dataset into train-set and test-set of 300k and 5k, respectively. The batch size used by them in the image captioning SAT model is 24, and the size of the features that are passed into CNN-encoder is 2048, whereas in our approach we increased the batch size to 32 for the same image captioning SAT model and decreased the size of the feature to 512 to further improve the results.

We also increased the length of the encoded vectors from 50 (proposed by them) to 100, which we received from the Word-2-Vec model to capture more features from various text-based data of the images. In contrast to the paper, we used the beam search of the Neural Architecture Search (NAS) to improve the performance of the SAT model further. The results are shown in the Table 1 below.

As shown in the results below, we have a MSE(2.729) and MAE(1.196) compared to their proposed results of MSE(3.56) and MAE(1.49). And we have approximately better SRC(0.756) to theirs which is SRC(0.65). We will try to further improve the results by tuning the hyper parameters and incorporating our approach in their baseline.

7 INTRODUCING DEEPGBM INTO THE MODEL

The whole DeepGBM framework, as shown in Figure 5 , consists of two major components: A Neural Network (NN) structure with the input of categorical features and GBDT2NN being another Neural Network structure distilled from GBDT with focusing on learning over dense numerical features.

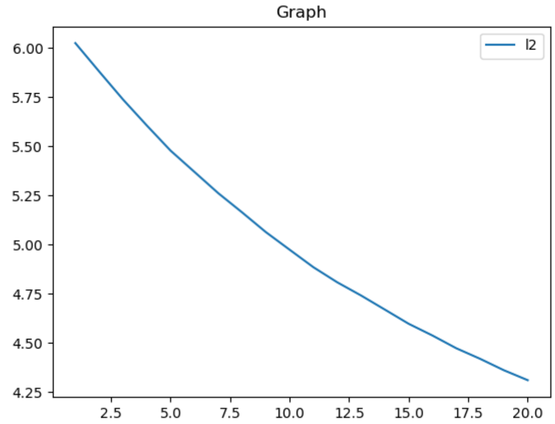

Figure 4: Training Curve Graph

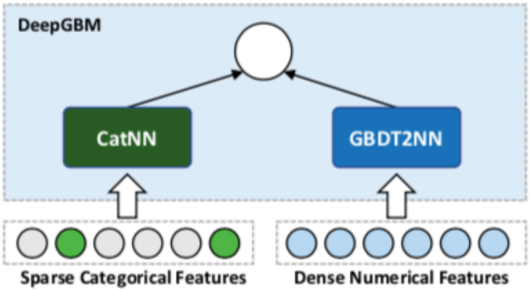

Figure 5: The framework of DeepGBM, which consists of two components, CatNN and GBDT2NN, to handle the sparse categorical and dense numerical features, respectively.

Both GBDT and NN have been widely used for the task of online prediction. Nonetheless, either of them does not perform very significantly as both of them have some disadvantages. Firstly, in GBDT the non-differentiable nature of the trees makes it hard to update the model and additionally, it fails to effectively use the sparse categorical features of the images to grow the trees. Whereas in case of NN’s it is unable to effectively learn over the dense numerical features of the images.

Due to the respective pros and cons of the NN and GDBT, there have been proposal to use them combinely and making most out of the advantages provided by both of them.

8 OBSERVATION AND FUTURE WORK

Firstly, our idea of using EXIF information for better analysis didn’t work out as the dataset avaiblable didn’t have that information available. Thus we propose it as a future work

Methods	SRC	MSE	MAE
Model Baseline	0.656	3.561	1.497
LightGBM	0.756	2.729	1.196
DeepGBM	**0.760**	**2.624**	**1.151**
CatBoost	0.704	3.293	1.360
LightGBM with user information	0.49	6.287	2.07
DeepGBM with user information	0.443	5.438	1.864
CatBoost with user information	0.480	6.346	2.072
LightGBM with only user info	0.478	6.435	2.071

Table 1: Performance comparison among the different methods evaluated on the testing set with different hyperparameters

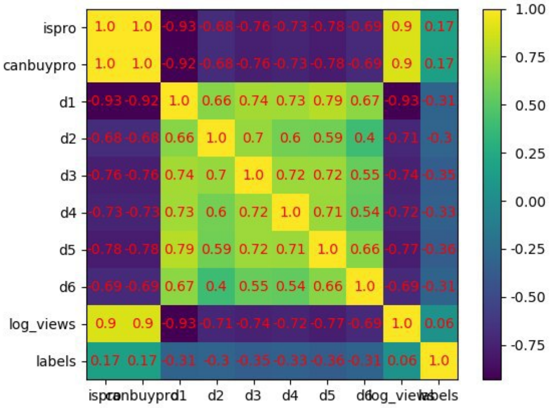

Figure 6: Correlation heatmap of user features

extension to our proposed solution.

Using user information should have yielded better results but the results were against our expectations. The performance of the model dropped by adding these features. We then analysed the results and found that apart from the "is PRO" feature, all the other user features had negative correlations with the results as can be seen from the correlation heat-map in Figure 6. This leads to the conclusion that number of posts, user description and since when the user has been posting have no direct relation with the popularity of user and hence degrade the learning.

We observed that the implementation of our LightGBM model beat the baseline results. This could be due to the hyper parameters tuning defined above.

Also, our DeepGBM model had even lesser MSE, MAE and a higher SRC compared to both the baseline results and our LightGBM model implementation, thus suggesting an overall better social media analysis prediction by our framework. However, adding the user data to the model resulted in worsening of the result with either of the three models. Thus suggesting an underline non-linear complex relation between the user popularity and post popularity on social media platforms. This complex relation study suggested by obtained results along with EXIF data and other curated features could be key to achieving even better results for the prediction.

9 AUTHORS WORK CONTRIBUTION

J. Rawal and R. Sharma worked upon exploring the problem statement and past work done (Literature Review) in the same domain by various researchers. A thorough discussion and brain storming by all group members was followed by formalization of proposed solution which was done by D. Kundu and K. Verma. Inspection of various usable Datasets and collection of TPIC2017 and SMHP was done by J. Rawal along with Dhruv Kundu. Image Captioning model to generate captions of collected image set and Word2vec based model for the captions was implemented by R. Sharma and K. Verma. The collection of dataset has been done by D. Kundu and R. Sharma. Whereas the dataset matching and pre-processing has been done by K. Verma and J. Rawal. Implementation of code and training of the data has been done by all of us. Each member contributed to fullest of his capability and the current progress is equally attributed to all four members.

ACKNOWLEDGMENTS

We would like to thank our instructor and mentor Dr. Rajiv Ratn Shah who guided us with choosing the problem statement and helped us in initialization of our research path. We would also like to thank Mr. Hitkul and Mr. Mohit, our Teaching Assistants for the course.

REFERENCES

[1] Younggue Bae and Hongchul Lee. [n.d.]. Sentiment Analysis of Twitter Audiences:Measuring the Positive or Negative Influence of Popular Twitterers.J. *Am. Soc.Inf. Sci. Technol.63* 12 ([n. d.]), 2521–2535.

[2] A. Bielski and T. Trzcinski. [n.d.]. Pay Attention to Virality: Understanding Popularity of Social Media Videos with the Attention Mechanism. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW* (2018). https://doi.org/10.1109/cvprw.2018.00309

[3] S. Cappallo, T. Mensink, and C.G. Snoek. [n.d.]. Latent Factors of Visual Popularity Prediction. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval - ICMR 15* (2015). https://doi.org/10.1145/2671188.2749405

[4] J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T.-S. Chua. [n.d.]. Micro Tells Macro. In *Proceedings of the 2016 ACM on Multimedia Conference* (2016), Vol. MM 16. https://doi.org/10.1145/2964284.2964314

[5] F. Gelli, T. Uricchio, M. Bertini, A.D. Bimbo, and S.-F. Chang. [n.d.]. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In *Proceedings of the 23rd ACM International Conference on Multimedia - MM 15* (2015). https://doi.org/10.1145/2733373.2806361

[6] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. 2015. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In *Proceedings of the 23rd ACM international conference on Multimedia - MM '15*. ACM Press, Brisbane, Australia, 907–910. https://doi.org/10.1145/2733373.2806361

[7] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. 2015. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. Association for Computing Machinery, New York, NY, USA, 907–910. https://doi.org/10.1145/2733373.2806361

[8] Chih-Chung Hsu, Hsiang-Chin Chien, Chia-Yen Lee, Ting-Xuan Liao, Jun-Yi Lee, Tsai-Yne Hou, Ying-Chu Kuo, Jing-Wen Lin, Ching-Yi Hsueh, and Zhong-Xuan Zhang. 2018. An Iterative Refinement Approach for Social Media Headline Prediction. In *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*. ACM Press, Seoul, Republic of Korea, 2008–2012. https://doi.org/10.1145/3240508.3266443

[9] C.-C. Hsu, L.-W. Kang, C.-Y. Lee, J.-Y. Lee, Z.-X. Zhang, and S.-M. Wu. [n.d.]. Popularity Prediction of Social Media based on Multi-Modal Feature Mining. In *Proceedings of the 27th ACM International Conference on Multimedia - MM 19* (2019). https://doi.org/10.1145/3343031.3356064

[10] Chih-Chung Hsu, Li-Wei Kang, Chia-Yen Lee, Jun-Yi Lee, Zhong-Xuan Zhang, and Shao-Min Wu. 2019. Popularity Prediction of Social Media Based on Multi-Modal Feature Mining. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2687–2691. https://doi.org/10.1145/3343031.3356064

[11] C.-C. Hsu, Y.-C. Lee, P.-E. Lu, S.-S. Lu, H.-T. Lai, C.-C. Huang, and W.-T. Su. [n.d.]. Social Media Prediction Based on Residual Learning and Random Forest. In *Proceedings of the 2017 ACM on Multimedia Conference - MM 17* (2017). https://doi.org/10.1145/3123266.3127894

[12] Chih-Chung Hsu, Ying-Chin Lee, Ping-En Lu, Shian-Shin Lu, Hsiao-Ting Lai, Chihg-Chu Huang, Chun Wang, Yang-Jiun Lin, and Weng-Tai Su. 2017. Social Media Prediction Based on Residual Learning and Random Forest. In *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*. ACM Press, Mountain View, California, USA, 1865–1870. https://doi.org/10.1145/3123266.3127894

[13] Jiani Hu, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2016. Multimodal learning for image popularity prediction on social media. In *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. 1–2. https://doi.org/10.1109/ICCE-TW.2016.7521017 ISSN: null.

[14] Guolin Ke, Zhenhui Xu, Jia Zhang, Jiang Bian, and Tie-Yan Liu. 2019. DeepGBM: A Deep Learning Framework Distilled by GBDT for Online Prediction Tasks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 384–394. https://doi.org/10.1145/3292500.3330858

[15] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. [n.d.]. What makes an image popular?InWWW.

[16] A. Krizhevsky, I. Sutskever, and G.E. Hinton. [n.d.]. Imagenet classification with deep convolutional neural networks. In *Proc. of NIPS* (2012).

[17] S. Lamprinidis, D. Hardt, and D. Hovy. [n.d.]. Predicting News Headline Popularity with Syntactic and Semantic Knowledge Using Multi-Task Learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (2018). https://doi.org/10.18653/v1/d18-1068

[18] L. Li, S. Huang, Z. He, and W. Liu. [n.d.]. An Effective Text-based Characterization Combined with Numerical Features for Social Media Headline Prediction. In *2018 ACM Multimedia Conference on Multimedia Conference - MM 18* (2018). https://doi.org/10.1145/3240508.3266438

[19] Liuwu Li, Sihong Huang, Ziliang He, and Wenyin Liu. 2018. An Effective Text-based Characterization Combined with Numerical Features for Social Media Headline Prediction. In *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*. ACM Press, Seoul, Republic of Korea, 2003–2007. https://doi.org/10.1145/3240508.3266438

[20] M. Mazloom, B. Hendriks, and M. Worring. [n.d.]. Multimodal Context-Aware Recommender for Post Popularity Prediction in Social Media. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017 - Thematic Workshops 17* (2017). https://doi.org/10.1145/3126686.3126731

[21] M. Mazloom, I. Pappi, and M. Worring. [n.d.]. Category Specific Post Popularity Prediction. *MultiMedia Modeling Lecture Notes in Computer Science* ([n. d.]), 594–607. https://doi.org/10.1007/978-3-319-73603-7_48

[22] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, and W.V. Dolen. [n.d.]. Multimodal Popularity Prediction of Brand-related Social Media Posts. In *Proceedings of the 2016 ACM on Multimedia Conference* (2016), Vol. MM 16. https://doi.org/10.1145/2964284.2967210

[23] M. Meghawat, S. Yadav, D. Mahata, Y. Yin, R.R. Shah, and R. Zimmermann. [n.d.]. A Multimodal Approach to Predict Social Media Popularity. https://doi.org/10.31219/osf.io/z8t2b

[24] A.M. Ramadhani and H.S. Goo. [n.d.]. Twitter sentiment analysis using deep learning methods. In *7th International Annual Engineering Seminar (InAES*. https://doi.org/10.1109/inaes.2017.8068556

[25] Alexandru Tatar, Marcelo Dias de Amorim, Serge Fdida, and Panayotis Antoniadis. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications* 5, 1 (Dec. 2014), 8. https://doi.org/10.1186/s13174-014-0008-y

[26] B. Wu, W.-H. Cheng, Y. Zhang, Q. Huang, J. Li, and T. Mei. [n.d.]. Sequential Prediction of Social Media Popularity with Deep Temporal Context Networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (2017). https://doi.org/10.24963/ijcai.2017/427

[27] Sheng Yu and Subhash Kak. 2012. A Survey of Prediction Using Social Media. *arXiv:1203.1647 [physics]* (March 2012). http://arxiv.org/abs/1203.1647 arXiv: 1203.1647.

[28] Alireza Zohourian, Hedieh Sajedi, and Arefeh Yavary. 2018. Popularity prediction of images and videos on Instagram. In *2018 4th International Conference on Web Research (ICWR)*. 111–117. https://doi.org/10.1109/ICWR.2018.8387246 ISSN: null.