

Fig. 2: top 10 TF-IDF-based features plus readability related features; feature ids from 100 and onward refer to readability features

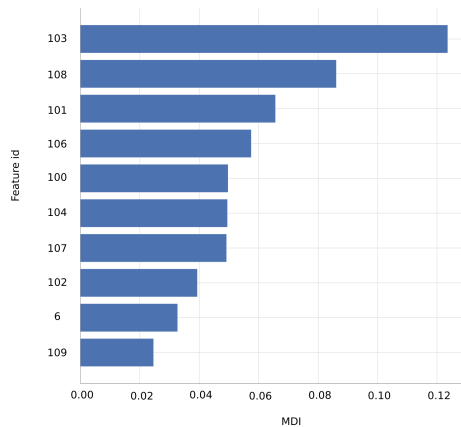
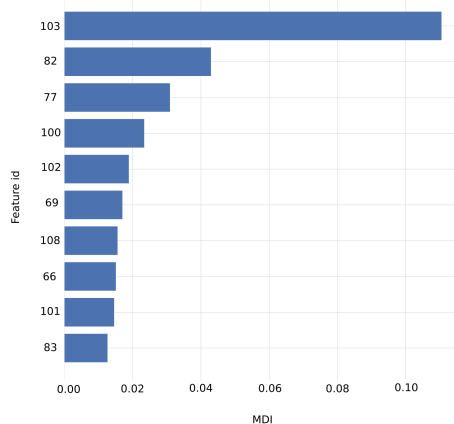


Fig. 3: top 10 word2vec-based features plus readability related features; feature ids from 100 and onward refer to readability features



VII. CONCLUSION

In this work, for the first time, we investigated the possible effect of readability of political tweets posted by a large number of Twitter accounts associated with European political actors, on the (positive) user engagement received from users and followers of these accounts. To do so, we formalised the task of measuring user engagement

as a regression problem. Then we used a set of features which could proxy the readability of the text in the tweets and compared the performance of four regressors with two baselines and two augmented sets of features. Our results indicate that the inclusion of readability related features in training the prediction models used in our work significantly improves the accuracy as well as the robustness of the predictions. The results indicate the critical role of readability in user engagement of political tweets.

For future work, we consider three directions. First, effective measuring of the readability of political tweets may need further investigation, because currently, there is no consensus on which textual and linguistic features of a tweet are mainly affecting the readability. In addition, further research is necessary to extend readability measures to other languages and re-evaluate the robustness of readability as a predictor. Second, the application of deep neural networks techniques can be a promising direction of research because these models can find complex linguistic and textual interconnections between terms that can bypass the possible current limitations such as the need to extensive feature engineering. Lastly, future work can focus on the interconnections between visual materials provided in the tweets such as memes, readability and engagement levels. It is conceivable that visual materials can significantly influence ease-of-read and cognitive demands for comprehension by presenting the information via more comprehensible mediums. This, however, is beyond the scope of the current work.

ACKNOWLEDGEMENT

This research is funded by the Norwegian University of Science and Technology (NTNU) through the Trondheim Analytica project.

REFERENCES

- [1] ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46, 3 (1992), 175–185.
- [2] AZPIAZU, I. M., AND PERA, M. S. Is readability a valuable signal for hashtag recommendations? *CEUR Workshop Proceedings 1688* (2016), 1–2.
- [3] BAEZA-YATES, R. A., AND RIBEIRO-NETO, B. *Modern Information Retrieval*. ACM press, 1999.

- [4] BOBBA, G. Social media populism: features and ‘likeability’ of Lega Nord communication on Facebook. *European Political Science* 18, 1 (2019), 11–23.
- [5] BUCHER, T., AND HELMOND, A. The Affordances of Social Media Platforms. In *The SAGE Handbook of Social Media*. SAGE Publications, 2018, pp. 233–253.
- [6] CHA, M., HADDADI, H., BENEVENUTO, F., GUMMADI, P. K., ET AL. Measuring user influence in twitter: The million follower fallacy. *Icwm* 10, 10-17 (2010), 30.
- [7] COLEMAN, M., AND LIAU, T. L. A computer readability formula designed for machine scoring. *Journal of Applied Psychology* 60, 2 (1975), 283–284.
- [8] CROSSLEY, S. A., ALLEN, D. B., AND MCNAMARA, D. S. Text Readability and Intuitive Simplification: A Comparison of Readability Formulas. *Reading in a Foreign Language* 23, 1 (2011), 84–101.
- [9] DALE, E., AND CHALL, J. S. A formula for predicting readability: Instructions. *Educational research bulletin* (1948), 37–54.
- [10] DAVENPORT, J. R., AND DELINE, R. The readability of tweets and their geographic correlation with education. *arXiv:1401.6058* (2014).
- [11] DAVIS, S. W., HORVÁTH, C., GRETRY, A., AND BELEI, N. Say what? How the interplay of tweet readability and brand hedonism affects consumer engagement. *Journal of Business Research* 100 (2019), 150–164.
- [12] DUBAY, W. The principles of readability, 2004.
- [13] EBERL, J.-M., TOLOCHKO, P., JOST, P., HEIDENREICH, T., AND BOOMGAARDEN, H. G. What’s in a post? How sentiment and issue salience affect users’ emotional reactions on Facebook. *Journal of Information Technology & Politics* 17, 1 (2020), 48–65.
- [14] ELTANTAWY, N., AND WIEST, J. B. The Arab spring—Social media in the Egyptian revolution: reconsidering resource mobilization theory. *International journal of communication* 5 (2011), 18.
- [15] FLESCH, R. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.
- [16] GIMPEL, K., SCHNEIDER, N., O’CONNOR, B., DAS, D., MILLS, D., EISENSTEIN, J., HEILMAN, M., YOGATAMA, D., FLANIGAN, J., AND SMITH, N. A. Part-of-speech tagging for twitter: Annotation, features, and experiments. Tech. rep., Paper presented at the 49th Annual Meeting of the association for computational linguistics: Human language technologies. Portland, Oregon: Association for Computational Linguistics, 2011.
- [17] GRUZZD, A., AND MAI, P. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7, 2 (2020), 2053951720938405.
- [18] GUMP, J. E. The Readability of Typefaces and the Subsequent Mood or Emotion Created in the Reader. *Journal of Education for Business* 76, 5 (2001), 270–273.
- [19] HEISS, R., SCHMUCK, D., AND MATTHES, J. What drives interaction in political actors’ Facebook posts? Profile and content predictors of user engagement and political actors’ reactions. *Information, Communication & Society* 22, 10 (2019), 1497–1513.
- [20] HO, T. K. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* 20, 8 (1998), 832–844.
- [21] JENDERS, M., KASNECI, G., AND NAUMANN, F. Analyzing and predicting viral tweets. In *WWW* (2013), pp. 657–664.
- [22] JOST, P., MAURER, M., AND HASSLER, J. Populism Fuels Love and Anger: The Impact of Message Features on Users’ Reactions on Facebook. *International Journal of Communication* 14, 0 (2020), 22.
- [23] KALSNES, B., AND LARSSON, A. O. Understanding News Sharing Across Social Media. *Journalism Studies* 19, 11 (2018), 1669–1688.
- [24] KE, G., MENG, Q., FINLEY, T., WANG, T., CHEN, W., MA, W., YE, Q., AND LIU, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (2017), pp. 3146–3154.
- [25] KWAK, H., LEE, C., PARK, H., AND MOON, S. What is Twitter, a Social Network or a News Media? In *WWW* (2010), pp. 591–600.
- [26] LOUPPE, G., WEHENKEL, L., SUTERA, A., AND GEURTS, P. Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems* (2013), pp. 431–439.
- [27] MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S., AND DEAN, J. Distributed representations of words and phrases and their compositionality. In *NIPS* (2013), pp. 3111–3119.
- [28] NOUH, M., NURSE, R. J., AND GOLDSMITH, M. Understanding the radical mind: Identifying signals to detect extremist content on twitter. In *International Conference on Intelligence and Security Informatics* (2019), pp. 98–103.
- [29] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., ET AL. Scikit-learn: Machine learning in python. *the Journal of machine Learning research* 12 (2011), 2825–2830.
- [30] REHUREK, R., AND SOJKA, P. Gensim—statistical semantics in python. *Retrieved from genism.org* (2011).
- [31] ROWE, M., ANGELETOU, S., AND ALANI, H. Predicting Discussions on the Social Semantic Web. In *The Semantic Web: Research and Applications*, vol. 6644. Springer Berlin Heidelberg, 2011, pp. 405–420. Series Title: Lecture Notes in Computer Science.
- [32] SAMUEL, J., GARVEY, M., AND KASHYAP, R. That Message Went Viral?! Exploratory Analytics and Sentiment Analysis into the Propagation of Tweets. *arXiv:2004.09718* (2020).
- [33] SUH, B., HONG, L., PIROLLO, P., AND CHI, E. H. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In *International Conference on Social Computing* (2010), pp. 177–184.
- [34] VENTURI, G., BELLANDI, T., DELL ORLETTA, F., AND MONTEMAGNI, S. NLP-Based Readability Assessment of Health-Related Texts: a Case Study on Italian Informed Consent Forms. In *International Workshop on Health Text Mining and Information Analysis* (2015), pp. 131–141.
- [35] XIE, S., TANG, J., AND WANG, T. Resonance Elicits Diffusion: Modeling Subjectivity for Retweeting Behavior Analysis. *Cognitive Computation* 7, 2 (2015), 198–210.
- [36] XU, J., AND CHAWLA, N. V. Mining Features Associated with Effective Tweets. In *ASONAM* (2017), pp. 525–532.