DEEP LEARNING APPROACH FOR IDENTIFICATION OF POVERTY THROUGH SENTIMENT ANALYSIS

Sri Redjeki[1], Setyawan Widyarto[2], and Anang Suwasto[1]
[1]STMIK AKAKOM, Indonesia
[2]Faculty of Communication, Visual Art and Computing, Universiti Selangor, 45600 Bestari Jaya, Selangor
E-mail : dzeky@akakom.ac.id

Abstract – This research aims to identify poverty in Indonesia through sentiment analysis using a deep learning approach to the Long Short-Term Memory (LSTM) method. Poverty is one of the main problems that the Indonesian government has become aware of over the years. Many policies have been implemented and created by the government, either with their efforts or assistance from other countries or the World Bank. The dataset used is 10288 twitter data that is crawled using poverty-related keywords. Training uses 80% and testing uses 20 datasets. The training data is further divided into 2, namely the training set and the validation set. The LSTM model produces a training accuracy of 88% with a validation-accuracy of 72%.

Keywords : Deep Learning; poverty; LSTM; sentiment analysis

1. INTRODUCTION

Many countries, organizations, and individuals pay attention to the problem of poverty, how to accurately and precisely measure poverty, or to how to determine whether their efforts have an impact on poverty reduction. This issue is essential because there are different approaches to poverty measurement, and none of them is perfect and can become an overall standard. It is not certain that the existing standards are suitable for every region, where household economic conditions and cultures are quite diverse [1-2]. Poverty is one of the main problems that the Indonesian government has become aware of over the years. Many policies have been implemented and created by the government, either with their efforts or assistance from other countries or the World Bank, to overcome poverty's issues. Therefore, there is a need to approach the poverty prediction models using other variables to obtain rights. Poverty prediction is based on previous data using the Bayesian regulation method to predict poverty data in Indonesia [3].

Online social networks such as Facebook, Twitter, and Weibo have become an integral part of everyday life in recent years and provide a platform for information exchange with one another. Because large amounts of social network data have exact features such as high quality, big data, semi-structured and direct flow to the real human community, social networks are more attentive to many researchers from existing fields or disciplines. Mining and analysing social network information, however, is a daunting task, facing two challenges: incomplete and dynamic [4-5]. Research related to social media is experiencing rapid development and evolution due to commercial pressures and the potential use of social media data for the benefit of social science research [6]. In particular, for computational social science research using quantitative techniques (e.g., computer statistics, machine learning, and complexity) called big data, social media is critical [7-8]. Social media connects many individuals around the world, thereby increasing the number of prediction participants. In social media, the diversity of individuals has the potential to improve the quality of the results of the forecast. Empirically, there are many prediction markets for economic, social, and political products in the social media era [9].

The purpose of sentiment analysis or opinion mining is to determine the attitude of a speaker, writer, or other topics concerning a specific subject or event. In different fields, sentiment analysis has many trending applications. It allows companies to automatically collect their customer's views on their products or services [10-12]. In order to investigate data and to gain business insights, business organizations need to process and study these feelings. The demand for sentiment analysis is increased by the need to analyse and structure hidden information from social media in the form of unstructured data [13]. Deep learning has become an effective mechanism for the production of high precision. In recent years, with comparatively remarkable results, deep learning models such as convolutionary neural networks (CNNs) and recurrent neural networks (RNNs) have been applied to text sentiment analysis

[14]. In a number of NLP tasks, such as machine translation, neural networks have achieved state-of-the-art performance. Aspect-level Sentiment Classification [15], Long Short-Term Memory (LSTM) Recurrent Neural Networks to Speech Enhancement, LSTM for Aspect-level Sentiment Classification. This study aims to identify poverty in Indonesia through the Sentiment Analysis approach with a deep learning LSTM algorithm. We explore the potential of the LSTM model to obtain the highest accuracy in identifying public opinion through sentiment analysis.

2. METHODOLOGY

With a total of 10228 data text, the data used for this research is Twitter data. The information is divided into three classes, namely positive, negative, and neutral classes.

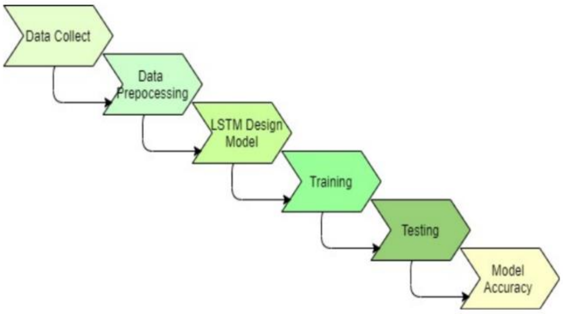

Figure 1 Methodology

The most class data are neutral class as many as 4252, positive class 3410, and negative class as much as 2566. The research dataset was divided into two, with a ratio of 80:20 for training and testing. The training dataset is divided into training and validation. The research stages consisted of six steps: data collection, data preprocessing, designing LSTM models, training, testing, and model accuracy (Fig.1). The LSTM algorithm data processing in this study has stages of text-preprocessing, lemmatization, stemming, word-embedding, tokenizer.

3. RESULTS AND DISCUSSION

In this study, the training carried out using a maximum epoch of 100. The outcomes of the study are shown in the graph in the form of precision and loss. The initial accuracy of this epoch is relatively very large, namely 0.7542, and the final accuracy of 0.8758, which means that the accuracy of the model training is 87.58%. Meanwhile, the validation accuracy at epoch 75 also produced the largest value among the other epochs, amounting to 0.7218, which means predicting new data is 72.18%. The final results of the training set and the validation set in the form of the final accuracy number, final validation accuracy, final loss, and final validation are shown in Table 1.

Table I Table Accuracy, Loss, and Validation

No	Epoch	Start Accuracy	Final Accuracy	Final Val_Accuracy	Final Loss	Final Val_Loss
1	20	0.5587	0.7542	0,6435	0.6749	2.3936
2	30	0.6661	0.7582	0,6741	0.6012	2.8793
3	40	0.7578	0.8042	0,6792	0.6127	3.4888
4	50	0.8048	0.8297	0,6540	0.6456	3.3008
5	60	0.4091	0.7769	0,7194	0,5613	3.0731
6	75	0.8402	0.8758	0,7146	0,5672	3.5658
7	85	0.3543	0.8249	0,7218	0,4801	3.3643
8	90	0.3471	0.8689	0,7141	0,4725	4.2014
9	100	0.8398	0.8646	0,7074	0,4912	4.4213

4. CONCLUSION

Analysis of the results of the study, one of the deep learning algorithms, namely LSTM, on sentiment analysis for poverty identification in Indonesia, provides training accuracy of 88%. These results indicate that LSTM can provide good model accuracy, but the model still has weaknesses, as the accuracy validation value is 72%. The results found that there were always the same text data during the training. The training accuracy is higher for a large number of epochs. The results of training losses are very small compared to validation-loss; this shows that the model can identify poverty through social media, namely Twitter. LSTM can learn the words in the training dataset so that it can use these words to classify sentiment analysis with training accuracy of 88%

REFERENCES

[1] S. Alkire et al., Multidimensional Poverty Measurement and Analysis : Chapter 3 – Overview of Methods for Multidimensional Poverty Assessment years . We are also grateful for direct comments on this working paper from Tony Atkinson , Achille. 2015.
[2] H. Megasari, S. Amar, and I. Idris, "Analisis Perekonomian Dan Kemiskinan Di Indonesia," J. Kaji. Ekon., vol. 3, no. 6, pp. 1–18, 2015.
[3] A. Wanto and J. T. Hardinata, "Estimasi Penduduk Miskin di Indonesia sebagai Upaya Pengentasan Kemiskinan dalam Menghadapi Revolusi Industri 4.0," CESS (Journal Comput. Eng. Syst. Sci., vol. 4, no. 2, pp. 198–207, 2019.
[4] A. Zielinski, S. E. Middleton, L. N. Tokarchuk, and X. Wang, "Social media text mining and network analysis for decision support in natural crisis management," ISCRAM 2013 Conf. Proc. - 10th Int. Conf. Inf. Syst. Cris. Response Manag., no. May 2013, pp. 840–845, 2013.
[5] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and Challenges of Big Data Research," Big Data Res., vol. 2, no. 2, pp. 59–64, 2015, doi: 10.1016/j.bdr.2015.01.006.
[6] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics – Challenges in topic discovery, data collection, and data preparation," Int. J. Inf. Manage., vol. 39, pp. 156–168, Apr. 2018, doi: 10.1016/j.ijinfomgt.2017.12.002.
[7] B. Batrinca and P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms," AI Soc., vol. 30, no. 1, pp. 89–116, 2014, doi: 10.1007/s00146-014-0549-4.
[8] A. Amado, P. Cortez, P. Rita, and S. Moro, "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis," Eur. Res. Manag. Bus. Econ., vol. 24, no. 1, 2018, doi: 10.1016/j.iedeen.2017.06.002.
[9] H. Schoen, D. Gayo-avello, and P. Gloor, "The power of prediction with social media," 2015, doi: 10.1108/IntR-06-2013-0115.
[10] I. Hemalatha, D. G. P. S. Varma, and D. a. Govardhan, "Sentiment Analysis Tool using Machine Learning Algorithms," Int. J. Emerg. Trends Technol. Comput. Sci., 2013.
[11] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.
[12] M. Heikal, "ScienceDirect Sentiment Analysis of Arabic Tweets using Deep Learning Sentiment Analysis of ∗ Arabic Tweets using Deep Learning," Procedia Comput. Sci., vol. 142, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.10.466.
[13] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Bus. Horiz., vol. 53, no. 1, pp. 59–68, 2010, doi: 10.1016/j.bushor.2009.09.003.
[14] X. Wang, W. Jiang, and Z. Luo, "Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts," pp. 2428–2437, 2016.
[15] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for Aspect-level Sentiment Classification," 2016.