# **DAFTAR PUSTAKA**

Alex, G., Rahman, M. A., & Geoffrey, H. (2013). Speech Recognition with Deep Recurrent Neural Networks. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, (pp. 6645-6649).

*An Introduction to Convolutional Neural Networks*. (2013, June 9). Retrieved June 5, 2017, from Scarlet Stanford Web site: http://scarlet.stanford.edu/teach/index.php/An\_Introduction\_to\_Convolutional\_Neural\_Networks

*Apache*. (n.d.). Retrieved June 13, 2017, from https://httpd.apache.org/

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. *Proceedings of the 4th Annual ACM Web Science Conference*, (pp. 24-32). Evanston, Illinois.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 711-720.

Ben-Hur, A., Elisseeff, A., Chapelle, O., & Weston, J. A. (2013). *Patent No. B2.* United States of America.

Bird, S. (2006). NLTK: The Natural Language Toolkit. *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions* (pp. 69-72). Sydney: Association for Computational Linguistics.

Booth, D. (2004, February 11). *Web Services Architecture Working Group*. Retrieved April 23, 2017, from http://www.w3.org/TR/ws-arch/

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Müller, A. C., Grisel, O., et al. (2013). API design for machine learning software: Experiences from the Scikit-learn Project.

Carley, K. M., Malik, M., Kowalchuk, M., Pfeffer, J., & Landwehr, P. (2015). *Twitter Usage in Indonesia.*

Cervone, D., & Pervin, L. A. (2015). *Personality Theory and Research (Third Edition).* Wiley.

Chopra, A., Prashar, A., & Sain, C. (2013). Natural Language Processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 131-134.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (almost) from Scratch.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. In C. Cortes, & V. Vapnik, *Machine Learning* (pp. 273-297). Boston: Kluwer Academic Publishers.

Deng, L., & Yu, D. (2014). *Deep Learning: Methods and Applications.* now Publishers.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification.* Wiley.

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of Discriminant Methods for the Classification of Tumors Using Gene Expression Data. *Journal of the American Statistical Association, 97*(457), 77-87.

Dvorski, D. D. (2007). Installing, Configuring, and Developing with XAMPP.

Egrioglu, E., Aladag, C. H., & Gunay, S. (2008). A New Model Selection Strategy in Artificial Neural Networks. *Applied Mathematics and Computation*, 591-597.

Facebook. (2012). *Facebook Company Info*. Retrieved December 2012, from Face Sheet Web site: http://newsroom.fb.com

Farnadi, G., Zoghbi, S., Moens, M.-F., & Cock, M. D. (2013). How Well Do Your Facebook Status Updates Express Your Personality?

Frank, E., Hall, M., Holmes, G., Mayo, M., Pfahringer, B., Smith, T., et al. (n.d.). *Weka Machine Learning Project*. Retrieved 2017, from www.cs.waikato.ac.nz/ml/weka

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*(5), 1189-1232.

Fukushima, K. (1980). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 193-202.

Gers, F., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation, 12*(10), 2451-2471.

Gers, F., Schraudolph, N., & Schmidhuber, J. (2003). Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research, 3*, 115-143.

Grefenstette, E., Blunsom, P., de Freitas, N., & Hermann, K. M. (2014). A Deep Architecture for Semantic Parsing.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning.

Ho, D. (2016). *Notepad++*. Retrieved June 13, 2017, from https://notepad-plus-plus.org/

Hosmer, D. W., Jovanovic, B., & Lemeshow, S. (1989). Best Subsets Logistic Regression. *Biometrics, 45*(4), 1265-1270.

Japkowicz, N., & Stephen, S. (2002). The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis, 6*(5), 429-450.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modeling Sentences.

Kerami, D., & Murfi, H. (2004). Kajian Kemampuan Generalisasi Support Vector Machine Dalam Pengenalan Jenis Splice Sites Pada Barisan DNA. *Makara, Sains, 8*(3), 89-95.

*Keras Documentation*. (n.d.). Retrieved June 6, 2017, from Keras Web site: https://keras.io/

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist, Vol 70(6)*, (pp. 543-556).

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (n.d.). Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. American Psychologist.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521*, 436-444.

Lemaitre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, 1-5.

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks.

M. Liwicki, A. G., Fernandez, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009). A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(5).

Mahoney, M. (2017, April 28). *Large Text Compression Benchmark*. Retrieved June 13, 2017, from Matt Mahoney Web site: http://www.mattmahoney.net/dc/text.html#1218

McAdams, D., & Olson, B. (2010). Personality Development: Continuity and Change Over the Life Course. *Annual Review of Psychology, 61*, 517-542.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in Its Natural Habitat: Manifestations and Implicit Folk Theories of Personality in Daily Life. *Journal of Personality and Social Psychology, 90*(5), 862-877.

Moffit, K., & Giboney, J. S. (2012). *Splice*. Retrieved April 23, 2017, from http://splice.cmi.arizona.edu/

Nachrowi, N. D., & Usman, H. (2002). *Penggunaan Teknik Ekonometri.* Jakarta: RajaGrafindo Persada.

Naradipha, A., & Purwarianti, A. (2011). Sentiment Classification for Indonesian Message in Social Media. *International Conference on Electrical Engineering and Informatics*, (pp. 1-4). Bandung.

Nilsson, N. J. (2009). *Artificial Intelligence: A New Synthesis.* California: Morgan Kaufmann.

*NumPy developers*. (n.d.). Retrieved June 13, 2017, from http://www.numpy.org/

Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques.* Heidelberg: Springer.

Ong, V., Rahmanto, A. D., Williem, Suhartono, D., Nugroho, A. E., Andangsari, E. W., et al. (2017). Personality Prediction Based on Twitter Information in Bahasa. Bina Nusantara University.

*Pandas*. (n.d.). Retrieved June 13, 2017, from http://pandas.pydata.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The Development and Psychometric Properties of LIWC2007.* Austin: LIWC.net.

Pryke, S. D. (2004). Analysis Construction Project Coalitions: Exploring the Application of Social Network Analysis. *Construction Management and Economics, 22*(8), 787-797.

Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning.* O'Reilly Media.

Ramchoun, H., Idrissi, M. A., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *Inernational Journal of Interactive Multimedia and Artificial Intelligence, 4*(1), 26-30.

Rosenblatt, F. (1958). The perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review, 65*(6), 386-408.

Ross, C., Orr, E., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. (2009). Personality and Motivations Associated with Facebook Use.

Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach.* New Jersey: Prentice Hall.

Sak, H., Senior, A., & Beaufays, F. (2014). Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. *Interspeech*, 338-342.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE, 8*(9).

Serrat, O. (2009). Social Network Analysis. Washington DC: Asian Development Bank.

Shafey, L. E., McCool, C., Wallace, R., & Marcel, S. (2013). A Scalable Formulation of Probabilistic Linear. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(7).

Shen, Y., He, X., Gao, J., Deng, L., & Mesnil, G. (2014). A Latent Semantic Model with Convolutional-Pooling Structure for Information Retrieval.

Siddiqi, M. H., Ali, R., Khan, A. M., Park, Y.-T., & Lee, S. (2015). Human Facial Expression Recognition Using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields. *IEEE Transactions on Image Processing, 24*(4), 1386-1398.

Skiena, S. S. (2008). *The Algorithm Design Manual.* London: Springer.

Swets, D. L., & Weng, J. (1996). Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8), 831-836.

Taieb, S. B., & Hyndman, R. J. (2013). A Gradient Boosting Approach to the Kaggle Load Forecasting Competition. *International Journal of Forecasting*.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 24-54.

*Techopedia Inc.* (n.d.). Retrieved June 13, 2017, from http://searchmicroservices.techtarget.com/definition/Apache

*Theano*. (n.d.). Retrieved June 6, 2017, from Theano Web site: http://deeplearning.net/software/theano/

Turban, E., McLean, E., & Wetherbe, J. (1999). *Information Technology for Management.* New York: John Wiley & Son, Inc.

Vogel, L., & Arthorne, J. (2015). *Contributing to the Eclipse IDE Project: Principles, Plug-ins and Gerrit Code Review.* Vogella.

Wijaya, A., Febrianto, N., Prasetia, I., & Suhartono, D. (2016). *Sistem Prediksi Kepribadian "The Big Five Traits" Dari Data Twitter.* Bina Nusantara University.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition).* Morgan Khaufman.

Xhemali, D., Hinde, C. J., & Stone, R. G. (2009). Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *International Journal of Computer Science Issues, 4*(1), 16-23.

Ye, J., Janardan, R., & Li, Q. (2005). Two-Dimensional Linear Discriminant Analysis.

Zhang, Y., Huo, M., Zhou, J., & Xie, S. (2010). PKSolver: An add-in Program for Pharmacokinetic and Pharmacodynamic Data Analysis in Microsoft Excel. *Computer Methods and Programs in Biomedicine*, 306-314.

Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., & Andrzej, C. (2013). Aggregation of Sparse Linear Discriminant Analyses for Event-Related Potential Classification in Brain-Computer Interface. *International Journal of Neural Systems*.

**Lampiran 1: Pertanyaan dan hasil *Interview***

**Lampiran 2: Slang Words**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Abnormal Term** | **Normal Term** |  | **Abnormal Term** | **Normal Term** |
| 7an | tujuan |  | anter | antar |
| @ | di |  | ap | apa |
| ababil | abg\_labil |  | ap2 | apa-apa |
| abis | habis |  | apaan | apa |
| acc | accord |  | apasih | apa\_sih |
| ad | ada |  | ape | apa |
| adlah | adalah |  | apes | sial |
| adlh | adalah |  | aplot | unggah |
| adoh | aduh |  | aps | apa |
| afaik | as\_far\_as\_i\_know |  | apva | apa |
| aha | tertawa |  | aq | saya |
| ahaha | haha |  | aqu | aku |
| aing | saya |  | aquwh | aku |
| aj | saja |  | asap | sesegera\_mungkin |
| aja | saja |  | asbun | asal\_bunyi |
| ajep-ajep | dunia\_gemerlap |  | aseek | asyik |
| ajj | saja |  | aseekk | asyik |
| ak | saya |  | asek | asyik |
| aka | dikenal\_juga\_sebagai |  | asekk | asyik |
| akika | aku |  | aseknya | asyiknya |
| akko | aku |  | asem | asam |
| akkoh | aku |  | asoy | asyik |
| akku | aku |  | aspal | asli\_tetapi\_palsu |
| akuwh | aku |  | astrojim | astagfirullahaladzim |
| akyu | aku |  | astul | asal\_tulis |
| alay | norak |  | ath | kalau\_begitu |
| aljasa | asal\_jadi\_saja |  | ato | atau |
| alow | halo |  | atuh | kalau\_begitu |
| ama | sama |  | au\_ah | tidak\_mau\_tahu |
| ambilin | ambilkan |  | ava | avatar |
| ambl | ambil |  | awak | saya |
| ancur | hancur |  | aws | awas |
| anjir | anjing |  | ay | sayang |
| anjrit | anjing |  | ayang | sayang |
| ank | anak |  | ayank | sayang |
| ayok | ayo |  | b4 | sebelum |
| bacot | banyak\_bicara |  | bhubu | tidur |
| bakalan | akan |  | bijimane | bagaimana |
| bales | balas |  | bimbuluh | bimbingan\_dan\_penyuluhan |
| bandes | bantuan\_desa |  | bintal | bimbingan\_mental |
| bangdes | pembangunan\_desa |  | bisi | kalau-kalau |
| bangedh | banget |  | bkl | akan |
| bangkotan | tua |  | bkn | bukan |
| banpol | bantuan\_polisi |  | bknnya | bukannya |
| banpres | bantuan\_presiden |  | bl | beli |
| banpur | bantuan\_tempur |  | blegug | bodoh |
| bansarkas | bantuan\_sarana\_kesehatan |  | blg | bilang |
| basbang | basi |  | blh | boleh |
| bazis | badan\_amal,\_zakat, \_infak,\_dan\_sedekah |  | blm | belum |
| bcanda | bercanda |  | bln | bulan |
| bcoz | karena |  | bls | balas |
| bdg | bandung |  | blum | belum |
| beb | sayang |  | bnchi | benci |
| begajulan | nakal |  | bnci | benci |
| bejibun | banyak |  | bngung | bingung |
| beliin | belikan |  | bnran | yang\_benar |
| belom | belum |  | bnyk | banyak |
| bencong | banci |  | bodor | lucu |
| bener | benar |  | bohay | badan\_aduhai |
| bentar | sebentar |  | bokap | ayah |
| ber2 | berdua |  | bokep | video\_porno |
| ber3 | bertiga |  | boker | buang\_air\_besar |
| berdikari | berdiri\_di\_atas\_kaki \_sendiri |  | bokin | pacar |
| beresin | membereskan |  | bokis | bohong |
| bet | banget |  | bole | boleh |
| bete | bosan |  | boljug | boleh\_juga |
| beti | beda\_tipis |  | bolot | bodoh |
| beud | banget |  | bgt | banget |
| beut | banget |  | brngkt | berangkat |
| bg | abang |  | bro | saudara\_laki-laki |
| bgd | banget |  | brp | berapa |
| bgmn | bagaimana |  | bru | baru |
| bgs | bagus |  | brur | saudara\_laki-laki |
| bonek | bocah\_nekat |  | brg | bareng |
| bonyok | ayah\_ibu |  | cuex | cuek |
| boyeh | boleh |  | cuman | Cuma |
| bpk | bapak |  | cumi | Cuma\_miscall |
| br | baru |  | cungkring | kurus\_tinggi |
| brb | segera\_kembali |  | cups | culun |
| ceu | kakak\_perempuan |  | cupu | culun |
| cew | cewek |  | curanmor | pencurian\_kendaraan\_bermotor |
| cewe | cewek |  | curas | pencurian\_dengan\_kekerasan |
| chuyunk | sayang |  | curcol | curahan\_hati\_colongan |
| cibuk | sibuk |  | curhat | curahan\_hati |
| cimeng | ganja |  | cwek | cewek |
| cin | cinta |  | cwok | cowok |
| cipika\_cipiki | cium\_pipi\_kanan \_cium\_pipi\_kiri |  | cyin | cinta |
| ciye | cie |  | cz | karena |
| ciyh | sih |  | d | di |
| ckck | ck |  | dagh | sudah |
| ckepp | cakep |  | dah | sudah |
| ckp | cakep |  | damang | sehat |
| ckp | cukup |  | dapet | dapat |
| clbk | cinta\_lama\_bersemi \_kembali |  | dateng | datang |
| cmiiw | correct\_me\_if\_i'm \_wrong |  | de | adik |
| cmpr | campur |  | dech | deh |
| cmpur | campur |  | dek | adik |
| cnenk | senang |  | deket | dekat |
| cong | banci |  | demen | suka |
| congor | mulut |  | dengerin | dengarkan |
| conlok | cinta\_lokasi |  | deyh | deh |
| cow | cowok |  | dg | dengan |
| cowwyy | maaf |  | dgn | dengan |
| coz | karena |  | dh | deh |
| cp | siapa |  | diancurin | dihancurkan |
| cpa | siapa |  | digosipin | digosipkan |
| cpe | capek |  | dimaafin | dimaafkan |
| cppa | siapa |  | dimana2 | dimana-mana |
| cppe | capek |  | dimintak | diminta |
| cr | cara |  | dimna | dimana |
| cucok | cocok |  | disono | di\_sana |
| cuek | tidak\_acuh |  | diuber | dikejar |
| dket | dekat |  | elu | kamu |
| dkit | sedikit |  | emang | memang |
| dkk | dan\_kawan-kawan |  | emangnya | memangnya |
| dl | dulu |  | emg | memang |
| dll | dan\_lain-lain |  | emng | memang |
| dlm | dalam |  | emte | makan\_teman |
| dlu | dulu |  | endak | tidak |
| dmn | dimana |  | engga | tidak |
| dngn | dengan |  | enggak | tidak |
| doang | saja |  | entar | nanti |
| dodol | bodoh |  | envy | iri |
| doi | dia |  | etc | et\_cetera |
| doku | uang |  | ex | mantan |
| don\_juan | playboy |  | ext | extension |
| dongs | dong |  | fax | facsimile |
| donlot | unduh |  | fb | facebook |
| dpt | dapat |  | fifo | first\_in\_first\_out |
| dr | dari |  | finek | finansial\_ekonomi |
| dri | dari |  | folbek | follow\_back |
| drmana | darimana |  | fwd | diterukan |
| drmn | darimana |  | fyi | sebagai\_informasi |
| drpd | daripada |  | ga | tidak |
| drtd | dari\_tadi |  | gaada | tidak\_ada\_uang |
| dsb | dan\_sebagainya |  | gabisa | tidak\_bisa |
| dst | dan\_seterusnya |  | gag | tidak |
| dstu | disitu |  | gahar | sangar |
| dtg | datang |  | gaje | tidak\_jelas |
| dugem | dunia\_gemerlap |  | gak | tidak |
| duh | aduh |  | gak\_papa | tidak\_apa-apa |
| duit | uang |  | gamau | tidak\_mau |
| duren | durian |  | gan | juragan |
| dy | dia |  | gapapa | tidak\_apa-apa |
| ed | edisi |  | gaptek | gagap\_teknologi |
| edan | gila |  | gatau | tidak\_tahu |
| egp | emang\_gue\_pikirin |  | gatek | gagap\_teknologi |
| eike | aku |  | gatot | gagal\_total |
| eke | aku |  | gawe | kerja |
| elo | kamu |  | gax | tidak |
| gbs | tidak\_bisa |  | gua | saya |
| gda | tidak\_ada\_uang |  | gue | saya |
| gebetan | orang\_yang\_disuka |  | guoblok | goblok |
| geer | gede\_rasa |  | guys | teman-teman |
| geje | tidak\_jelas |  | gw | saya |
| gelo | gila |  | gws | cepat\_sembuh |
| gepeng | gelandangan\_dan \_pengemis |  | ha | tertawa |
| ghiey | lagi |  | haghaghag | tertawa |
| ghiy | lagi |  | haha | tertawa |
| gi | lagi |  | hakhak | tertawa |
| gile | gila |  | hallow | halo |
| giman | gigi\_mancung |  | handak | bahan\_peledak |
| gimana | bagaimana |  | hankam | pertahanan\_dan\_keamanan |
| gini | begini |  | hansip | pertahanan\_sipil |
| gino | gigi\_nongol |  | hehe | he |
| gitchu | gitu |  | hellow | halo |
| githu | gitu |  | helo | halo |
| gitu | begitu |  | helow | halo |
| gj | tidak\_jelas |  | hey | hai |
| gk | tidak |  | hi | hai |
| gmana | bagaimana |  | hlm | halaman |
| gmn | bagaimana |  | hlng | hilang |
| gn | begini |  | hny | hanya |
| goblog | goblok |  | hnya | hanya |
| goblok | bodoh |  | hoax | isu\_bohong |
| gokil | gila |  | houm | rumah |
| golput | golongan\_putih |  | hr | hari |
| gombal | suka\_merayu |  | hrs | harus |
| gowes | mengayuh\_sepeda |  | hrus | harus |
| gpl | tidak\_pakai\_lama |  | hubad | hubungan\_angkatan\_darat |
| gpny | tidak\_punya |  | hubdar | perhubungan\_darat |
| gpp | tidak\_apa-apa |  | hubla | perhubungan\_laut |
| gr | gede\_rasa |  | huff | mengeluh |
| gretong | gratis |  | huft | mengeluh |
| gretongan | gratisan |  | hum | rumah |
| gt | begitu |  | humas | hubungan\_masyarakat |
| gtau | tidak\_tahu |  | humz | rumah |
| gtw | tidak\_tahu |  | idk | saya\_tidak\_tahu |
| ilang | hilang |  | jurdil | jujur\_adil |
| ilfeel | tidak\_suka |  | jutek | galak |
| ilfil | tidak\_suka |  | jwb | jawab |
| imba | jago\_sekali |  | k | ke |
| imho | in\_my\_humble\_opinion |  | ka | kakak |
| imoet | imut |  | kak | kakak |
| imoetz | imut |  | kab | kabupaten |
| info | informasi |  | kabag | kepala\_bagian |
| item | hitam |  | kabor | kabur |
| itung | hitung |  | kacian | kasihan |
| itungan | hitungan |  | kacrut | kacau |
| iy | iya |  | kadit | kepala\_direktorat |
| iye | iya |  | kadiv | kepala\_divisi |
| iyo | iya |  | kaga | tidak |
| ja | saja |  | kagak | tidak |
| jablay | jarang\_dibelai |  | kaka | kakak |
| jadiin | jadikan |  | kalo | kalau |
| jadul | jaman\_dulu |  | kalo | kalau |
| jaim | jaga\_image |  | kamtib | keamanan\_dan\_ketertiban |
| jancuk | anjing |  | kamtibmas | keamanan\_dan\_ketertiban \_masyarakat |
| jayus | tidak\_lucu |  | kmrn | kemarin |
| jd | jadi |  | klu | kalau |
| jdi | jadi |  | kamuh | kamu |
| jdikan | jadikan |  | kamuwh | kamu |
| jem | jam |  | kamyu | kamu |
| jg | juga |  | kanwil | kantor\_wilayah |
| jga | juga |  | kapt | kapten |
| jgn | jangan |  | karna | karena |
| jgnkan | jangankan |  | kasat | kepala\_satuan |
| jijay | jijik |  | kasubbag | kepala\_subbagian |
| jir | anjing |  | kasubbid | kepala\_subbidang |
| jkt | jakarta |  | katrok | kampungan |
| jl | jalan |  | kau | kamu |
| jln | jalan |  | kayanya | kayaknya |
| jnj | janji |  | kbar | kabar |
| jomblo | tidak\_punya\_pacar |  | kbr | kabar |
| jth | jatuh |  | kcian | kasihan |
| jubir | juru\_bicara |  | kdu | harus |
| keburu | terlanjur |  | kmari | kemari |
| kec | kecamatan |  | kmps | kampus |
| kedubes | kedutaan\_besar |  | kmpus | kampus |
| kejurnas | kejuaraan\_nasional |  | kmrn | kemarin |
| kek | seperti |  | kn | kan |
| kekeuh | keras\_kepala |  | knal | kenal |
| keknya | kayaknya |  | knl | kenal |
| kel | kelurahan |  | knp | kenapa |
| keliatan | kelihatan |  | knpa | kenapa |
| kemaren | kemarin |  | kodya | kota\_madya |
| keneh | masih |  | kog | kok |
| kepengen | mau |  | komdis | komisi\_disiplin |
| kepikiran | terpikirkan |  | kompi | komputer |
| kepingin | mau |  | komsov | komunis\_sovyet |
| kepo | mau\_tahu\_urusan\_orang |  | komtiong | komunis\_Tiongkok |
| kepsek | kepala\_sekolah |  | kongkow | kumpul\_bareng\_teman-teman |
| kere | tidak\_punya\_uang |  | konjen | konsulat\_jenderal |
| kesbang | kesatuan\_bangsa |  | kopdar | kopi\_darat |
| kesian | kasihan |  | koq | kok |
| kesra | kesejahteraan\_rakyat |  | korup | korupsi |
| ketauan | ketahuan |  | kpd | kepada |
| ketrima | diterima |  | kpn | kapan |
| keukeuh | keras\_kepala |  | kptsan | keputusan |
| kgiatan | kegiatan |  | krenz | keren |
| khan | kan |  | krik | garing |
| kibul | bohong |  | krm | kirim |
| kibus | kaki\_busuk |  | krn | karena |
| kimpoi | kawin |  | kt | kita |
| kk | kakak |  | ktauan | ketahuan |
| kl | kalau |  | ktmu | ketemu |
| klian | kalian |  | ktny | katanya |
| klianz | kalian |  | ktr | kantor |
| klo | kalau |  | kudu | harus |
| kloter | kelompok\_terbang |  | kuper | kurang\_pergaulan |
| kluarga | keluarga |  | kuq | kok |
| klw | kalau |  | kw | imitasi |
| klwrga | keluarga |  | ky | seperti |
| km | kamu |  | kyk | seperti |
| kykny | kayanya |  | lowh | kamu |
| la | lah |  | lp | lupa |
| laka | kecelakaan |  | lu | kamu |
| lam | salam |  | luber | langsung,\_umum,\_bebas, \_dan\_rahasia |
| lambreta | lambat |  | luchu | lucu |
| lamp | lampiran |  | luchuw | lucu |
| lansia | lanjut\_usia |  | luff | cinta |
| lanud | landasan\_udara |  | lum | belum |
| lapas | lembaga\_pemasyarakatan |  | luph | cinta |
| latgab | latihan\_gabungan |  | luthu | lucu |
| lbur | libur |  | lw | kamu |
| lebay | berlebihan |  | lwn | lawan |
| leh | boleh |  | lwt | lewat |
| leh | boleh |  | maacih | terima\_kasih |
| lekong | laki-laki |  | maaciw | terima\_kasih |
| lelet | lambat |  | mabal | bolos |
| lemot | lambat\_berfikir |  | mabes | markas\_besar |
| lemot | lemah\_otak |  | macem | macam |
| lg | lagi |  | macem-macem | macam-macam |
| lgi | lagi |  | macih | masih |
| lgkp | lengkap |  | madesu | masa\_depan\_suram |
| lgsg | langsung |  | maem | makan |
| lht | lihat |  | maen | main |
| liat | lihat |  | magabut | makan\_gaji\_buta |
| linmas | perlindungan\_masyarakat |  | mahatma | maju\_sehat\_bersama |
| litbang | penelitian\_dan\_ pengembangan |  | maho | homo |
| lmyan | lumayan |  | mak | ibu |
| lmyn | lumayan |  | mak\_jang | kaget |
| lngkp | lengkap |  | makasih | terima\_kasih |
| lo | kamu |  | maksain | memaksa |
| loch | loh |  | malah | bahkan |
| loe | kamu |  | malem | malam |
| lol | tertawa |  | malu2in | memalukan |
| lola | lambat\_berfikir |  | mam | makan |
| lom | belum |  | mamz | makan |
| louph | cinta |  | maneh | kamu |
| loupz | cinta |  | manies | manis |
| low | kalau |  | maniez | manis |
| mantep | mantap |  | mumet | pusing |
| mao | mau |  | mumpung | selagi |
| markus | makelar\_kasus |  | muna | munafik |
| masukin | masukkan |  | munas | musyawarah\_nasional |
| mba | mbak |  | munaslub | musyawarah\_nasional\_luar \_biasa |
| melu | ikut |  | muntaber | muntah\_dan\_berak |
| mending | lebih\_baik |  | musda | musyawarah\_daerah |
| mepet | dekat\_sekali |  | musti | mesti |
| mgkn | mungkin |  | muup | maaf |
| mgu | minggu |  | muupz | maaf |
| mhn | mohon |  | muuv | maaf |
| migas | minyak\_dan\_gas\_bumi |  | mw | mau |
| miker | minuman\_keras |  | mw | now\_watching |
| mikol | minuman\_beralkohol |  | n | dan |
| milis | mailing\_list |  | nal | kenal |
| miras | minuman\_keras |  | nanam | menanam |
| mksd | maksud |  | nangis | menangis |
| mlah | malah |  | nanya | bertanya |
| mls | malas |  | naon | apa |
| mngkn | mungkin |  | napa | kenapa |
| mnt | minta |  | napa | kenapa |
| mo | mau |  | napi | narapidana |
| moge | motor\_gede |  | napol | narapidana\_politik |
| mokad | mati |  | napza | narkotika,\_alkohol, \_psikotropika,\_dan\_zat\_adiktif |
| mokat | mati |  | naq | anak |
| moso | masa |  | narkoba | narkotika,\_psikotropika, \_dan\_obat\_terlarang |
| mosok | masa |  | narsis | bangga\_pada\_diri\_sendiri |
| mpe | sampai |  | nasgor | nasi\_goreng |
| msh | masih |  | nax | anak |
| msk | masuk |  | nda | tidak |
| mskpn | meskipun |  | ndak | tidak |
| mslh | masalah |  | ndiri | sendiri |
| msng2 | masing-masing |  | ndut | gendut |
| mt | makan\_teman |  | ne | ini |
| muahal | mahal |  | nekolim | neokolonialisme |
| mubes | musyawarah\_besar |  | nekolin | neokolonialisme |
| muker | musyawarah\_kerja |  | nelfon | menelepon |
| mulu | melulu |  | nembak | menyatakan\_cinta |
| ngabis2in | menghabiskan |  | ngmg | bicara |
| ngabuburit | menunggu\_berbuka\_ puasa |  | ngmng | bicara |
| ngakak | tertawa |  | ngocol | lucu |
| ngaku | mengaku |  | ngomong | bicara |
| ngambek | marah |  | ngomongin | membicarakan |
| ngambil | mengambil |  | ngubek2 | mencari-cari |
| ngampus | pergi\_ke\_kampus |  | ngumpul | berkumpul |
| nganggur | tidak\_punya\_pekerjaan |  | ngurus | mengurus |
| ngantri | mengantri |  | ni | ini |
| ngapah | kenapa |  | nie | ini |
| ngapain | sedang\_apa |  | niech | nih |
| ngaret | terlambat |  | nih | ini |
| ngaruh | berpengaruh |  | nimbrung | ikut\_serta |
| ngasih | memberikan |  | niyh | nih |
| ngawur | berbicara\_sembarangan |  | nkah | nikah |
| ngebandel | berbuat\_bandel |  | nmr | nomor |
| ngeceng | kumpul\_bareng-bareng |  | nnti | nanti |
| ngegosip | bergosip |  | nntn | nonton |
| ngeh | sadar |  | no | nomor |
| ngeklaim | mengklaim |  | nobar | nonton\_bareng |
| ngekos | tinggal\_di\_kos |  | nongkrong | kumpul-kumpul |
| ngeksis | menjadi\_eksis |  | np | now\_playing |
| ngelamar | melamar |  | nrp | narapidana |
| ngeles | berkilah |  | ntar | nanti |
| ngeliat | melihat |  | ntar | sebentar |
| ngelidur | menggigau |  | ntn | nonton |
| ngemeng | bicara\_terus-terusan |  | nuhun | terima\_kasih |
| ngerampok | merampok |  | numpuk | bertumpuk |
| ngerti | mengerti |  | nunggu | menunggu |
| ngga | tidak |  | nutupin | menutupi |
| nggak | tidak |  | nyampah | membuat\_sampah |
| ngibul | berbohong |  | nyari | mencari |
| ngikut | ikut |  | nyariin | mencari |
| ngiler | mau |  | nyekar | menyekar |
| nginep | menginap |  | nyiapin | mempersiapkan |
| ngiri | iri |  | nyicil | mencicil |
| ngisi | mengisi |  | nyiram | menyiram |
| ngisiin | mengisikan |  | nyoblos | mencoblos |
| nyok | ayo |  | pw | posisi\_nyaman |
| nyokap | ibu |  | qmo | kamu |
| o/ | oleh |  | qmu | kamu |
| ogah | tidak\_mau |  | qt | kita |
| ok | ok |  | rakor | rapat\_koordinasi |
| ol | online |  | rame | ramai |
| omdo | omong\_doang |  | ranmor | kendaraan\_bermotor |
| ongkir | ongkos\_kirim |  | raskin | rakyat\_miskin |
| oon | bodoh |  | re | reply |
| oot | out\_of\_topic |  | red | redaksi |
| org | orang |  | ref | referensi |
| org2 | orang-orang |  | reg | register |
| orkes | olahraga\_dan\_kesehatan |  | rehab | rehabilitasi |
| ortu | orang\_tua |  | rejeki | rezeki |
| ospek | orientasi\_studi\_dan \_pengenalan\_kampus |  | rempong | sulit |
| otda | otonomi\_daerah |  | renstra | rencana\_strategis |
| otre | ok |  | repp | balas |
| otw | on\_the\_way,\_ sedang\_di\_jalan |  | reskrim | reserse\_kriminal |
| pa | apa |  | restik | reserse\_narkotika |
| pacal | pacar |  | rhs | Rahasia |
| pahe | paket\_hemat |  | rhs | rahasia |
| pake | pakai |  | rmbt | rambut |
| pakek | pakai |  | rmh | rumah |
| pala | kepala |  | rofl | rolling\_on\_the\_floor\_laughing |
| panja | panitia\_kerja |  | ru | baru |
| pansus | panitia\_khusus |  | rudal | peluru\_kendali |
| pantura | pantai\_utara |  | ruko | rumah\_toko |
| parpol | partai\_politik |  | rusun | rumah\_susun |
| pas2an | pas-pasan |  | rusunawa | rumah\_susun\_sewa |
| pasutri | pasangan\_suami\_istri |  | rutan | rumah\_tahanan |
| patas | cepat\_dan\_terbatas |  | ruz | terus |
| pd | pada |  | s4 | sempat |
| pdhl | padahal |  | saia | saya |
| pede | percaya\_diri |  | saiank | sayang |
| pelatda | pemusatan\_latihan\_daerah |  | salting | salah\_tingkah |
| pelatnas | pemusatan\_  latihan\_nasional |  | saltum | salah\_kostum |
| pswt | pesawat |  | sampe | sampai |
| puanas | panas |  | sampeyan | anda |
| samsek | sama\_sekali |  | siramin | siramkan |
| sapa | siapa |  | sis | kakak\_perempuan |
| sapose | siapa |  | sj | saja |
| satkorlak | satuan\_koordinator\_ pelaksana |  | skali | sekali |
| satpam | satuan\_pengamanan |  | skalian | sekalian |
| say | sayang |  | skl | sekali |
| sbb | sebagai\_berikut |  | sklh | sekolah |
| sbg | sebagai |  | skrg | sekarang |
| sbh | sebuah |  | skt | sakit |
| sblm | sebelum |  | skul | sekolah |
| sbnrny | sebenarnya |  | slesai | selesai |
| scara | secara |  | slh | salah |
| scr | secara |  | sll | selalu |
| sdg | sedang |  | slm | salam |
| sdgkn | sedangkan |  | slma | selama |
| sdh | sudah |  | slmt | selamat |
| sdkt | sedikit |  | slsai | selesai |
| sdngkn | sedangkan |  | slt | sulit |
| se7 | setuju |  | smpt | sempat |
| seabrek | banyak |  | smua | semua |
| sebelas\_dua\_belas | mirip |  | smw | semua |
| sekjen | sekretaris\_jenderal |  | smwny | semuanya |
| sembako | sembilan\_bahan\_pokok |  | sndiri | sendiri |
| sempak | celana\_dalam |  | sng | senang |
| sempet | sempat |  | soljum | sholat\_jumat |
| sempet2nya | sempat-sempatnya |  | somse | sombong\_sekali |
| sendratari | seni\_drama\_tari |  | songong | sombong |
| sertijab | serah\_terima\_jabatan |  | sorry | maaf |
| sgt | sangat |  | sory | maaf |
| sgtunya | segitunya |  | sosbud | sosial-budaya |
| shg | sehingga |  | sosek | sosial-ekonomi |
| sidak | inspeksi\_mendadak |  | sospol | sosial-politik |
| siech | sih |  | sotoy | sok\_tahu |
| sieyh | sih |  | sowry | maaf |
| sikon | situasi\_dan\_kondisi |  | spa | siapa |
| simpedes | simpanan\_pedesaan |  | spd | sepeda |
| sinetron | sinema\_elektronik |  | sppa | siapa |
| sip | ok |  | sprti | seperti |
| spt | seperti |  | tggu | tunggu |
| spy | supaya |  | tgl | tanggal |
| srtfkt | sertifikat |  | tgu | tunggu |
| stelah | setelah |  | thanks | terima\_kasih |
| stiap | setiap |  | thankz | terima\_kasih |
| stiap | setiap |  | thd | terhadap |
| stlh | setelah |  | thn | tahun |
| subbag | subbagian |  | thx | terima\_kasih |
| suk | masuk |  | tilang | bukti\_pelanggaran |
| sumbangin | sumbangkan |  | tipi | TV |
| sumpek | sempit |  | tipiwan | TvOne |
| sy | saya |  | tkg | tukang |
| syg | sayang |  | tks | terima\_kasih |
| syp | siapa |  | tll | terlalu |
| t4 | tempat |  | tlp | telepon |
| tabanas | tabungan\_pembangunan\_ nasional |  | tlpn | telepon |
| tajir | kaya |  | tls | tulis |
| tar | nanti |  | tman | teman |
| tau | tahu |  | tmbah | tambah |
| taun | tahun |  | tmbh | tambah |
| taw | tahu |  | tmen2 | teman-teman |
| tawh | tahu |  | tmn2 | teman-teman |
| td | tadi |  | tmpah | tumpah |
| tdi | tadi |  | tmph | tumpah |
| tdk | tidak |  | tmpt | tempat |
| te2p | tetap |  | tnda | tanda |
| teh | kakak\_perempuan |  | tngu | tunggu |
| tekor | rugi |  | tnh | tanah |
| telat | terlambat |  | tnyta | ternyata |
| telkom | telekomunikasi |  | togel | toto\_gelap |
| telmi | telat\_berpikir |  | tokai | tai |
| telp | telepon |  | tokai | tai |
| temen | teman |  | toserba | toko\_serba\_ada |
| temen2 | teman-teman |  | tp | tapi |
| tengil | menyebalkan |  | tpi | tapi |
| tengok | menjenguk |  | tq | terima\_kasih |
| tepar | terkapar |  | trdhulu | terdahulu |
| terbitin | terbitkan |  | trgntg | tergantung |
| trima | terima\_kasih |  | utk | untuk |
| trims | terima\_kasih |  | utul | ujian\_tertulis |
| trm | terima |  | valas | valuta\_asing |
| trnyt | ternyata |  | vol | volume |
| trs | terus |  | w/ | dengan |
| trus | terus |  | wad | buat |
| trutama | terutama |  | wadir | wakil\_direktur |
| trz | terus |  | wagub | wakil\_gubernur |
| ts | penulis |  | wamil | wajib\_militer |
| tsb | tersebut |  | waria | wanita\_pria |
| tst | tahu\_sama\_tahu |  | warkop | warung\_kopi |
| ttd | tertanda |  | warnet | warung\_internet |
| ttg | tentang |  | warteg | warung\_tegal |
| ttp | tetap |  | wartel | warung\_telekomunikasi |
| tuch | tuh |  | wat | buat |
| tuh | itu |  | wk | tertawa |
| tuir | tua |  | wkt | waktu |
| tuwh | tuh |  | wkwk | tertawa |
| tw | tahu |  | wtf | what\_the\_fuck |
| tyap | tiap |  | wth | what\_the\_hell |
| u | kamu |  | xixixi | tertawa |
| u/ | untuk |  | y | kenapa |
| ud | sudah |  | ya | iya |
| uda | sudah |  | yad | yang\_akan\_datang |
| udah | sudah |  | yap | iya |
| udh | sudah |  | yapz | iya |
| ujg | ujung |  | yaudah | ya\_sudah |
| ujung2nya | ujung-ujungnya |  | yauwda | ya\_sudah |
| ul | ulangan |  | yawdah | ya\_sudah |
| kampret | sialan |  |
| ybs | yang\_bersangkutan |  | amp | sampai |
| yg | yang |  | ind | indonesia |
| yi | yaitu |  | ina | indonesia |
| yl | yang\_lain |  | vs | lawan |
| yng | yang |  | dprd | daripada |
| yo | iya |  | amp | sampai |
| yoha | iya |  | ind | indonesia |
| yowes | ya\_sudah |  | ina | indonesia |
| yuk | ayo |  |  |  |
| ultah | ulang\_tahun |  |  |  |
| unyu | lucu |  |  |  |
| unyu | menggemaskan |  |  |  |
| uplot | unggah |  |  |  |
| uppu | apa |  |  |  |
| urang | saya |  |  |  |
| urg | saya |  |  |  |
| usah | perlu |  |  |  |
| utang | hutang |  |  |  |
| yup | iya |  |  |  |
| ywdh | ya\_sudah |  |  |  |
| apkh | apakah |  |  |  |
| apakh | apakah |  |  |  |
| lpor | lapor |  |  |  |

**Lampiran 3: Stopwords**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| a |  | her |  | she |
| about |  | here |  | should |
| above |  | hers |  | shouldn |
| after |  | herself |  | so |
| again |  | him |  | some |
| against |  | himself |  | such |
| ain |  | his |  | t |
| all |  | how |  | than |
| am |  | i |  | that |
| an |  | if |  | the |
| and |  | in |  | their |
| any |  | into |  | theirs |
| are |  | is |  | them |
| aren |  | isn |  | themselves |
| as |  | it |  | then |
| at |  | its |  | there |
| be |  | itself |  | these |
| because |  | just |  | they |
| been |  | ll |  | this |
| before |  | m |  | those |
| being |  | ma |  | through |
| below |  | me |  | to |
| between |  | mightn |  | too |
| both |  | more |  | under |
| but |  | most |  | until |
| by |  | mustn |  | up |
| can |  | my |  | ve |
| couldn |  | myself |  | very |
| d |  | needn |  | was |
| did |  | no |  | wasn |
| didn |  | nor |  | we |
| do |  | not |  | were |
| does |  | now |  | weren |
| doesn |  | o |  | what |
| doing |  | of |  | when |
| don |  | off |  | where |
| down |  | on |  | which |
| during |  | once |  | while |
| each |  | only |  | who |
| few |  | or |  | whom |
| for |  | other |  | why |
| from |  | our |  | will |
| further |  | ours |  | with |
| had |  | ourselves |  | won |
| hadn |  | out |  | wouldn |
| has |  | over |  | y |
| hasn |  | own |  | you |
| have |  | re |  | your |
| haven |  | s |  | yours |
| having |  | same |  | yourself |
| he |  | shan |  | yourselves |

**Lampiran 4: Feature Selections**

Daftar *features* dari SNA setelah dilakukan *Feature Selection*

|  |
| --- |
| **Features** |
| Network size |
| Betweeness |
| nBetweeness |
| Density |
| Brokerage |

Daftar *features* dari LIWC setelah dilakukan *Feature Selection*

|  |
| --- |
| **Features Abbrev** |
| ppron |
| percept |
| quant |
| exclam |
| see |
| allpunc |
| male |

Daftar *features* dari SPLICE setelah dilakukan *Feature Selection*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features Abbrev** |  | **Features Abbrev** |  | **Features Abbrev** |
| numChars |  | Imagery |  | numSpeculate |
| numCharsMinusSpacesAndPunctuation |  | pleasantness |  | expressivity |
| numWords |  | activation |  | pausality |
| numSentences |  | avgwordLength |  | questionCount |
| numPunctuation |  | avgSentenceLength |  | pastTense |
| numNouns |  | numSyllables |  | presentTense |
| numVerbs |  | avgSyllablesPerWord |  | swnPositivity |
| numAdjectives |  | numWordsWith3OrMoreSyllables |  | swnNegativity |
| numAdverbs |  | rateWordsWith3OrMoreSyllables |  | swnObjectivity |
| numPassiveVerbs |  | numWordsWith6OrMoreChars |  | ARI |
| firstPersonSingular |  | rateWordsWith6OrMoreChars |  | FRE |
| iCanDoIt |  | numWordsWith7OrMoreChars |  | FKGL |
| doKnow |  | rateWordsWith7OrMoreChars |  | CLI |
| posSelfImage |  | lexicalDiversity |  | LWRF |
| iCantDoIt |  | complexityComposite |  | FOG |
| numImperatives |  | hedgeVerb |  | SMOG |
| suggestionPhrases |  | hedgeModal |  | DALE |
| totalDominance |  | hedgeAll |  | LIX |
| numAgreement |  | numDisfluencies |  | RIX |
| totalSubmissiveness |  | numInterjections |  | FRY |

**Lampiran 5: Source Codes**

***Machine Learning Source Codes***

import arff

import uuid

import numpy

from pandas import DataFrame

from sklearn import preprocessing

from sklearn.pipeline import Pipeline

from sklearn.cross\_validation import KFold

from sklearn.metrics import accuracy\_score

from sklearn.naive\_bayes import GaussianNB

from sklearn.svm import LinearSVC

from sklearn.linear\_model import LogisticRegression

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

from sklearn.feature\_selection import SelectKBest, chi2

from imblearn.over\_sampling import SMOTE

from imblearn.under\_sampling import ClusterCentroids

TOTAL\_DATASET = 400

total\_features = 60

index\_class = 60

# FB\_Features=7; LIWC=85; Splice=74

# fs\_FB\_Features=5; fs\_LIWC=7; fs\_Splice=60

class **GetFeatures**(TransformerMixin):

def **transform**(*self*, features):

featuresResult = []

for i in range(0, features.shape[0]):

rows = []

for j in range(0, total\_features):

rows.append(features[i][j])

featuresResult.append(rows)

return numpy.asarray(featuresResult).reshape(features.shape[0], total\_features)

def **fit**(*self*, X, y=None):

return *self*

def **build\_data\_frame**(features, classification):

rows = []

index = []

rows.append({*'features'*: features, *'class'*: classification})

index.append(str(uuid.uuid1()))

data\_frame = DataFrame(rows, index=index)

return data\_frame

def **train**(dsetFilename, classifier):

numpy.random.seed(8)

dset = arff.load(open(dsetFilename, *'rb'*))

features = []

classification = []

labelEncoder = preprocessing.LabelEncoder()

for index\_dataset in (range(0, TOTAL\_DATASET)):

ftrs = []

for index\_features in (range(0, total\_features)):

ftrs.append(dset[*'data'*][index\_dataset][index\_features])

features.append(ftrs)

classification.append(dset[*'data'*][index\_dataset][index\_class])

encodedClassification = labelEncoder.fit([*'yes'*, *'no'*])

transformedEncodeClassification = labelEncoder.transform(classification)

data = DataFrame({*'features'*: [], *'class'*: []})

for i in range(0, TOTAL\_DATASET):

data = data.append(build\_data\_frame(features[i], transformedEncodeClassification[i]))

data = data.reindex(numpy.random.permutation(data.index))

#TRAINING USING PIPELINE

pipeline = Pipeline([

(*'features'*, GetFeatures()),

(*'classifier'*, classifier)

])

# Start of FeatureSelection #######################################################################################################

element\_data\_features = []

for i in range(0, data[*'features'*].values.shape[0]):

element\_data\_features.append(data[*'features'*].values[i])

temp\_data\_features = numpy.array(element\_data\_features)

temp\_data\_features = numpy.reshape(temp\_data\_features, (len(temp\_data\_features), total\_features))

# Univariate Selection

kBest = SelectKBest(score\_func=chi2, k=5) # FB=5 , LIWC=7, Splice=60

kBestFit = kBest.fit(temp\_data\_features, data[*'class'*].values)

numpy.set\_printoptions(precision=3)

kBestScore = kBestFit.scores\_ # Score for each class

print(kBestScore)

kBestScore.sort()

print(kBestScore)

featureSelectionResult = kBestFit.transform(temp\_data\_features)

# End of FeatureSelection #########################################################################################################

#CROSS-VALIDATING

print (*"Start training "* + dsetFilename + *" with\n"* + str(classifier))

k\_fold = KFold(n=len(data), n\_folds=10)

accuracy\_scores = []

foldIndex = 1;

for train\_indices, test\_indices in k\_fold:

print (*"kFold-"* + str(foldIndex))

foldIndex = foldIndex + 1

x\_train = data.iloc[train\_indices][*'features'*].values

y\_train = data.iloc[train\_indices][*'class'*].values

x\_test = data.iloc[test\_indices][*'features'*].values

y\_test = data.iloc[test\_indices][*'class'*].values

# Start of Oversampling #######################################################################################################

# element\_x\_train = []

# for i in range(0, len(x\_train)):

# element\_x\_train.append(x\_train[i])

# temp\_x\_train = numpy.array(element\_x\_train)

# temp\_x\_train = numpy.reshape(temp\_x\_train, (len(temp\_x\_train), total\_features))

#

# kind = ['borderline1']

# sm = [SMOTE(kind=k) for k in kind]

# for method in sm:

# x\_train\_resampled, y\_train\_resampled = method.fit\_sample(temp\_x\_train, y\_train)

#

# x\_train = x\_train\_resampled

# y\_train = y\_train\_resampled

# End of Oversampling #########################################################################################################

# Start of UnderSampling #######################################################################################################

# element\_x\_train = []

# for i in range(0, len(x\_train)):

# element\_x\_train.append(x\_train[i])

# temp\_x\_train = numpy.array(element\_x\_train)

# temp\_x\_train = numpy.reshape(temp\_x\_train, (len(temp\_x\_train), total\_features))

#

# cc = ClusterCentroids()

# x\_train\_resampled, y\_train\_resampled = cc.fit\_sample(temp\_x\_train, y\_train)

#

# x\_train = x\_train\_resampled

# y\_train = y\_train\_resampled

# End of Undersampling #########################################################################################################

pipeline.fit(x\_train, y\_train)

predictions = pipeline.predict(x\_test)

accuracyScore = accuracy\_score(y\_test, predictions)

accuracy\_scores.append(accuracyScore)

trainResult = *''*

trainResult += str(sum(accuracy\_scores)/len(accuracy\_scores)) + *','*

print (*"Trained successfully\n"*)

return trainResult

traits = [*'openness'*, *'conscientiousness'*, *'extraversion'*, *'agreeableness'*, *'neuroticism'*]

classifiers = [GaussianNB(), LinearSVC(), LogisticRegression(), GradientBoostingClassifier(), LinearDiscriminantAnalysis()]

for trait in traits:

dsetFilename = *"Dataset\Personality Prediction\Dataset\mix\Machine Learning\Arff\Splice\mix\_fs\_"*+trait+*"\_dataset.arff"*

trainResult = *""*

for classifier in classifiers:

trainResult += train(dsetFilename, classifier)

resultFileName = open(*"Dataset/Personality Prediction/Dataset/mix/Machine Learning/Training Result/tr\_ml\_splice\_fs\_nores\_"*+trait[0]+*".txt"*, *"w"*)

resultFileName.write(trainResult)

resultFileName.close

print(*"All training result has been saved."*)

***Deep Learning Source Codes***

import os

import numpy as np

from \_\_future\_\_ import print\_function

from keras.preprocessing.text import Tokenizer

from keras.preprocessing.sequence import pad\_sequences

from keras.utils import to\_categorical

from keras.models import Sequential

from keras.layers import Embedding

from keras.layers import Dense, Flatten

from keras.layers import LSTM, GRU, Conv1D, MaxPooling1D

from sklearn.metrics import accuracy\_score

from imblearn.over\_sampling import SMOTE

from imblearn.under\_sampling import ClusterCentroids

BASE\_DIR = *''*

GLOVE\_DIR = BASE\_DIR + *'Dataset/Personality Prediction/Keras/GloVe/6B/'*

MAX\_SEQUENCE\_LENGTH = 1000

MAX\_NB\_WORDS = 2800

EMBEDDING\_DIM = 100

VALIDATION\_SPLIT = 0.2

np.random.seed(8)

print(*'Indexing word vectors.'*)

embeddings\_index = {}

f = open(os.path.join(GLOVE\_DIR, *'glove.6B.100d.txt'*))

for line in f:

values = line.split()

word = values[0]

coefs = np.asarray(values[1:], dtype=*'float32'*)

embeddings\_index[word] = coefs

f.close()

def **train**(datasetFilename, classFilename):

print(*'Processing text dataset'*)

texts = []

labels = []

dataset = open(datasetFilename).read()

texts = dataset.split(*'#SEPARATOR#'*)

labelClass = open(classFilename).read()

labels = labelClass.split(*'\n'*)

tokenizer = Tokenizer(num\_words=MAX\_NB\_WORDS)

tokenizer.fit\_on\_texts(texts)

sequences = tokenizer.texts\_to\_sequences(texts)

word\_index = tokenizer.word\_index

data = pad\_sequences(sequences, maxlen=MAX\_SEQUENCE\_LENGTH)

labels = to\_categorical(np.asarray(labels))

indices = np.arange(data.shape[0])

np.random.shuffle(indices)

data = data[indices]

labels = labels[indices]

num\_validation\_samples = int(VALIDATION\_SPLIT \* data.shape[0])

x\_train = data[:-num\_validation\_samples]

y\_train = labels[:-num\_validation\_samples]

x\_test = data[-num\_validation\_samples:]

y\_test = labels[-num\_validation\_samples:]

# Start of Oversampling #######################################################################################################

# y\_train = np.reshape(y\_train, (len(y\_train)))

# y\_test = np.reshape(y\_test, (len(y\_test)))

#

# kind = ['borderline1']

# sm = [SMOTE(kind=k) for k in kind]

# for method in sm:

# print(str(x\_train.shape) + " " + str(x\_test.shape))

# x\_train\_resampled, y\_train\_resampled = method.fit\_sample(x\_train, y\_train)

# x\_test\_resampled, y\_test\_resampled = method.fit\_sample(x\_test, y\_test)

#

# x\_train = x\_train\_resampled

# x\_test = x\_test\_resampled

# y\_train = to\_categorical(np.asarray(y\_train\_resampled.tolist()))

# y\_test = to\_categorical(np.asarray(y\_test\_resampled.tolist()))

# End of Oversampling #########################################################################################################

# Start of Undersampling #####################################################################################################

# labels = np.reshape(labels, (len(labels)))

#

# cc = ClusterCentroids()

# data\_resampled, labels\_resampled = cc.fit\_sample(data, labels)

#

# data = data\_resampled

# labels = to\_categorical(np.asarray(labels\_resampled.tolist()))

# End of Undersampling #######################################################################################################

print(*'Preparing embedding matrix.'*)

num\_words = min(MAX\_NB\_WORDS, len(word\_index))

embedding\_matrix = np.zeros((num\_words, EMBEDDING\_DIM))

for word, i in word\_index.items():

if i >= MAX\_NB\_WORDS:

continue

embedding\_vector = embeddings\_index.get(word)

if embedding\_vector is not None:

embedding\_matrix[i] = embedding\_vector

embedding\_layer = Embedding(num\_words,

EMBEDDING\_DIM,

weights=[embedding\_matrix],

input\_length=MAX\_SEQUENCE\_LENGTH,

trainable=False)

print(*'Training model.'*)

# MLP

model = Sequential()

model.add(embedding\_layer)

model.add(Flatten())

model.add(Dense(50, init=*'glorot\_uniform'*, activation=*'tanh'*))

model.add(Dense(2, activation=*'softmax'*))

model.compile(loss=*'categorical\_crossentropy'*, optimizer=*'adam'*, metrics=[*'accuracy'*])

# LSTM

# model = Sequential()

# model.add(embedding\_layer)

# model.add(LSTM(32, dropout=0.2, recurrent\_dropout=0.2))

# model.add(Dense(2, activation='softmax'))

# model.compile(loss='binary\_crossentropy', optimizer='adam', metrics=['accuracy', 'mse', 'mae'])

# GRU

# model = Sequential()

# model.add(embedding\_layer)

# model.add(GRU(32, dropout=0.2, recurrent\_dropout=0.2))

# model.add(Dense(2, activation='softmax'))

# model.compile(loss='categorical\_crossentropy', optimizer='rmsprop', metrics=['accuracy', 'mse', 'mae'])

# CNN1D GLOBALMAXPOOLING

# sequence\_input = Input(shape=(MAX\_SEQUENCE\_LENGTH,), dtype='int32')

# embedded\_sequences = embedding\_layer(sequence\_input)

# x = Conv1D(32, 5, activation='tanh')(embedded\_sequences)

# x = MaxPooling1D(5)(x)

# x = Conv1D(32, 5, activation='tanh')(x)

# x = MaxPooling1D(5)(x)

# x = Conv1D(32, 5, activation='tanh')(x)

# x = MaxPooling1D(35)(x)

# x = Flatten()(x)

# x = Dense(32, activation='tanh')(x)

# preds = Dense(2, activation='softmax')(x)

# model = Model(sequence\_input, preds)

# model.compile(loss='categorical\_crossentropy', optimizer='adam', metrics=['accuracy', 'mse', 'mae'])

# CNN LSTM

# model = Sequential()

# model.add(embedding\_layer)

# model.add(Dropout(0.2))

# model.add(Conv1D(50, 3, padding='valid', activation='relu', strides=1))

# model.add(MaxPooling1D(pool\_size=4))

# model.add(LSTM(70))

# model.add(Dense(2, activation='softmax'))

# model.compile(loss='binary\_crossentropy', optimizer='adam', metrics=['accuracy', 'mse', 'mae'])

model.fit(x\_train, y\_train, batch\_size=32, epochs=10, validation\_data=(x\_test, y\_test), verbose=2)

predictions = model.predict(x\_test)

predictedRound = [round(x[1]) for x in predictions]

predicted = [x[1] for x in predictions]

tested = [round(x[1]) for x in y\_test]

accuracyScore = accuracy\_score(tested, predictedRound)

trainResult = *''*

trainResult += str(accuracyScore) + *','*

print (*"Trained successfully\n"*)

return trainResult

datasetName = *'mix'*

traits = [*'openness'*, *'conscientiousness'*, *'extraversion'*, *'agreeableness'*, *'neuroticism'*]

datasetFilename = *"Dataset\Personality Prediction\Dataset\\"* + datasetName + *"\\"* + datasetName + *"\_preprop\_status\_dataset.txt"*

for trait in traits:

classFilename = *r'Dataset\Personality Prediction\Dataset\\'* + datasetName + *'\Deep Learning\Class\\'* + datasetName + *'\_'* + trait + *'\_class.txt'*

trainResult = *""*

trainResult += train(datasetFilename, classFilename)

resultFileName = open(*"Dataset/Personality Prediction/Dataset/"* + datasetName + *"/Deep Learning/Training Result/tr\_dl\_text\_preprop\_nores\_"*+trait[0]+*".txt"*, *"a"*)

resultFileName.write(trainResult)

resultFileName.close

print(*"All training result has been saved."*)