Terbit online pada laman web jurnal: **https://journal.ilmudata.co.id/index.php/RIGGS**

RIGGS
Journal of Artificial Intelligence and Digital Business
Vol. 1 No. 1 (2022) 1 - 6 e-ISSN: 2963-914X P-ISSN: 2963-9298

fitur n-gram untuk perbandingan metode machine learning pada sentimen judul berita keuangan

Arif Mudi Priyatno[1], Fahmi Iqbal Firmanand[2]
[1,2]Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Pahlawan Tuanku Tambusai
[1]arifmudi11@gmail.com, [2]fahmi.iqbal@gmail.com

Abstrak

Sentiment analisis saat ini banyak digunakan pada aplikasi natural language processing ataupun information retrival. Analisa sentiment analisis bisa memberikan informasi terkait headline berita keuangan yang beredar dan memberikan masukan terhadap perusahaan. Sentimen positif akan memberikan dampak yang baik pula terhadap perkembangan perusahaan, akan tetapi sentiment negative akan membuat reputasi perusahaan berkurang, hal ini akan berpengaruh terhadap perkembangan perusahaan. Penelitian ini melakukan perbandingan metode machine learning pada headline berita keuangan dengan ektraksi fitur n-gram. Tujuan penelitian ini untuk mendapatkan metode terbaik dalam mengklasifikasikan sentiment headline berita keuangan perusahaan. Metode machine learning yang dibandingkan yaitu Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, multi-layer perceptron (MLP), Stochastic Gradient Descent, dan Decision Trees. Hasil menunjukan metode terbaik yaitu logistic regression dengan persentase f1-measure, precision, dan recal sebesar 73,94; 73,94; dan 74,63. Hal ini menunjukan bahwa fitur n-gram dan machine learning berhasil melakukan sentiment analisis.

Kata kunci: : n-gram, Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, multi-layer perceptron (MLP), Stochastic Gradient Descent, Decision Trees, sentiment analyst

1. Pendahuluan

Teknologi Informasi dan komunikasi pada saat ini berkembang dengan sangat pesat. Berbagai hal telah bisa dilakukan dengan teknologi terutama dengan adanya kecerdasan buatan. Salah satu contoh yang bisa dilakukan yaitu melakukan analisis sentiment terhadap judul berita. Analisa sentiment pada text dapat dilakukan menaggunakan text mining.

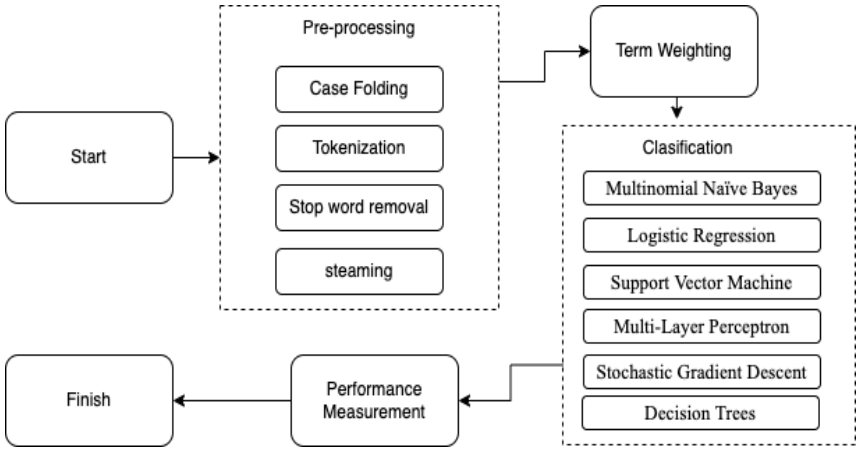
Sentiment analisis adalah proses pengkategorian pendapat / topik pada text, proses ini untuk menentukan apakah pendapat / topik bernilai positif, negative, atau netral [1]. Penelitian sentimen analisis telah banyak dilakukan diberbagai sektor, seperti keuangan [2], politik [3], perusahaan [4], dan lain sebagainya. Analisa sentiment terhadap berita keuangan dan perusahaan akan memberikan efek terhadap kredibilitas suatu perusahaan. Sentimen berpengaruh besar pada perusahaan yang bersifat terbuka atau terdaftar pada bursa saham.

Penelitian ini melakukan perbandingan metode machine learning untuk sentiment analisi dengan fitur n-gram pada kasus berita keuangan dan headline perusahaan. Data digunakan dari penelitian [5] dengan jumlah 4846 judul. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604. Data dilakukan preprocessing dengan menghapus hal yang tidak dibutuhkan. Hasil preprocesing diberikan pembobotan dengan metode n-gram. Fitur hasil n-gram dilakukan proses sentimen analyst dengan menggunakan beberapa metode machine learning, diantaranya Multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, multi-layer perceptron (MLP), Stochastic Gradient Descent, dan Decision Trees. Performa sentimen analisis di hitung menggunakan presisi, recall, dan f1-measure.

Penelitian terkait yaitu [6] melakukan analisis sentiment pada artikel berita berdasarkan lexicon. Data yang digunakan yaitu BBC News tahun 2004 hingga 2005 (http://mlg.ucd.ie/datasets/bbc.html). Penelitian tersebut mendapatkan kesimpulan bahwa artikel dengan topik bisnis dan olahraga memiliki sentiment positif, sedangkan topik terkait hiburan dan olahraga memiliki sentiment negative.

Penelitian [7] melakukan investigasi terhadap metrik pengukuran sentiment analisis pada artikel berita. Metode pembobotan yag digunakan yaitu *term frequency–inverse document frequency* (TF-IDF). Metode klasifikasi menggunakan Gaussian Naive Bayes classifiers dan *linear support vector machine*

Received: 03-03-2022 | Accepted: 23-05-2022 | Published: 31-07-2022



Gambar 1. Tahapan Penelitian

(linear SVM). Hasil terbaik didapatkan menggunakan metode linear SVM sebesar 62%, sedangkan Gaussian naïve Bayes mendapatkan 61%.

Penelitian [8] melakukan analisis sentiment pada berita keuangan VADER (Valence Aware Dictionary and Sentiment Reasoner). Penelitian bertujuan menunjukkan pengaruh sentiment berita keuangan terhadap perubahan harga pasar saham.

Penelitian [9] melakukan perbandingan metode analisis pada headline berita di belanda. Perbandingan tersebut yaitu anotasi manual, crowd coding, numerous dictionaries, machine learning, dan deep learning. Kesimpulan didapatkan bahwa hasil terbaik yaitu dilakukan oleh manusia atau crowd coding, numerous dictionaries menunjukan bahwa tingkat validitas bisa diterima, dan machine learning serta deep learning secara subtansi lebih baik dari pada numerous dictionaries tetapi masih jauh dari kinerja manusia.

2. Metode Penelitian

Proses penelitian ini diantaranya yaitu pengambilan data, pre-processing, pembobotan menggunakan n-gram, klasifikasi sentiment, dan pengukuran performa. Gambar 1 menunjukan tentang tahapan penelitian yang dilakukan.

2.1. Data

Data merupakan teks berita keuangan dan siaran pers perusahaan. Penandaan sentiment dilakukan oleh sekelompok 16 annotator dengan latar belakang pendidikan bisnis yang memadai. Data menggunakan penelitian [5] dengan jumlah total 4846 judul berita. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604.

2.2. Pre-processing

Tahapan ini dilakukan untuk memepersiapkan data dapat digunakan pada tahapan-tahapan selanjutnya. Pre-processing membersihkan data text dari kata (token) yang umum dan tidak memiliki makna yang diperlukan, hal ini dilakukan untuk mengurangi noise pada data text. Gambar 1 menunjukan proses yang terdapat pada pre-processing yaitu case folding, tokenization, stopword removal dan steaming. Case Folding merupakan tahap awal yang dilakukan untuk menyeragamkan huruf dengan cara mengubah semua huruf besar menjadi huruf kecil. Contonya “According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.” menjadi “according to gran, the company has no plans to move all production to russia, although that is where the company is growing”. Tokenization merupakan tahapan memotong kalimat menjadi token (term) yang lebih kecil. Contohnya “according to gran, the company has no plans to move all production to russia, although that is where the company is growing.” menjadi [“according“, ”to”, “gran”, “the”, “company“, “has”, “no”, “plans”, “to”, “move”, “all”, “production”, “to”, “russia", “although”, “that”, “is”, “where”, “company”, “is”, “growing”]. Stop word removal merupakan tahapan menghapus term yang bersifat umum dan tidak memiliki pengaruh. Contoh term yang dihapus yaitu 'its', 'again', 'for', 'myself', 'his', dan lain sebagainya [10]. Steaming adalah proses pengembalian term menjadi kata dasarnya [11]. Contohnya yaitu “plans” menjadi “plan”, “growing” menjadi “grow”.

Tabel 1. Contoh n-gram

N-gram		Result
Unigram	:	‘a’, ‘c’, ‘c’, ‘o’, ‘r’, ‘d’, ‘g’, ‘r’, ‘a’, ‘n’
Bigram	:	‘_a’, ‘ac’, ‘cc’, ‘co’, ‘or’, ‘rd’, ‘dg’, ‘gr’, ‘ra’, ‘an’, ‘n_’
Trigram	:	‘_ac’, ‘acc’, ‘cco’, ‘cor’, ‘ord’, ‘rdg’, ‘dgr’, ’gra’, ‘ran’, ‘an_’, ‘n__’
Quadgram	:	‘_acc’, ‘acco’, ‘ccor’, ‘cord’, ‘ordg’, ‘rdgr’, ‘dgra’, ‘gran’, ‘ran_’, ‘an__’, ‘n___’

DOI: https://doi.org/10.31004/riggs.v1i1.4

Tabel 2. Contoh n-gram pada penelitian

N-gram		Result
Unigram	:	'according', 'gran', 'company', 'plan', 'move', 'production', 'russia', 'although', 'company', 'grow'
Bigram	:	'according gran', 'gran company', 'company plan', 'plan move', 'move production', 'production russia', 'russia although', 'although company', 'company grow'
Trigram	:	'according gran company', 'gran company plan', 'company plan move', 'plan move production', 'move production russia', 'production russia although', 'russia although company', 'although company grow'
Quadgram	:	'according gran company plan', 'gran company plan move', 'company plan move production', 'plan move production russia', 'move production Russia although', 'production russia although company', 'russia although company grow'

2.3. Term Weighting

N-gram adalah pemotongan kalimat atau string menjadi leboh kecil sesuai dengan N karakter yang ditentukan. Gram didefinisikan sebagai sub urutan dari N karakter yang dikerjakan [12]. Metode n-gram digunakan dengan tujuan memberikan makna dari urutan kata ataupun karakter pada suatu kalimat . Metode n-gram digunakan untuk mengambil potongan karakter huruf sebanyak n dari kalimat yang secara kontinuitas. kontinuitas dimaksud adalah dari awal hingga akhir dokumen.

N-Gram dikelasifikasikan berdasarkan n karakter. Secara umum, n-gram dilakukan dengan memberikan tambahan blank di awal dan di akhir [13]. Contoh, kalimat "accord gran" dilakukan proses n-gram, blank dimaksud disimbolkan dengan "_" menghasilkan n-gram pada Table 1.

Table 3 Data Latih dan Data Uji

No	Persentase	Data Latih	Data Uji	Total
1	70 %	3392	1454	4868
2	80 %	3876	970	4868
3	90 %	4361	485	4868

N-Gram memiliki keuntungan yaitu matching setiap kata dapat digunakan walaupaun ada interpetasi kode tertentu seperti kode pos. Pada aplikasi natural language processing, angka dan kode tertentu dianggap sebagai noise. Hal ini dianggap keuntungan karena dekomposisi bagian-bagian kecil yang tetap memiliki makna. Hal ini secara tekstual tidak memberikan pengaruh.

Pada penelitian ini n-gram digunakan tidak dipecah menjadi karakter-karakter kecil, tetapi n-gram memecah kalimat menjadi kata perkata. Contoh kalimat "according gran company plan move production russia although company grow" dilakukan proses n-gram dan hasilnya dapat dilihat pada Table 2. Term weighting pada penelitian ini menggunakan banyaknya gram yang muncul atau lebih dikenal dengan term frequency (TF).

2.4. Klasifikasi Sentimen

Penelitian ini membandingkan beberapa metode klasifikasi. Metode yang digunakan yaitu *multinomial naïve bayes* [14][15], *logistic regression* [16][17], *support vector machine* [18][19][20], *multi layer perceptron* [21][22][23][24], *stochastic gradient descent* [25][26], dan *decition tree* [27][28].

Table 4 Result Pre-processing

No	Input	Result Pre-processing
1	According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.	according gran company plan move production russia although company grow
2	Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications, the statement said.	technopolis plan develop stage area less 100,000 square meter order host company work computer technology telecommunication statement say
3	The international electronic industry company Elcoteq has laid off tens of employees from its Tallinn facility; contrary to earlier layoffs the company contracted the ranks of its office workers, the daily Postimees reported.	international electronic industry company elcoteq laid ten employee tallinn facility contrary earlier layoff company contract rank office worker daily postimees report
4	With the new production plant the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials and therefore increase the production profitability.	new production plant company would increase capacity meet expect increase demand would improve use raw material therefore increase production profitability

DOI: https://doi.org/10.31004/riggs.v1i1.4

2.5. Pengukuran Performa

Evaluasi performa pada penelitian ini menggunakan confusion matrix. Evaluasi performa berdasarkan confusion matrix diantaranya yaitu presisi, recall, dan f1-measure[11][29].

$$recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$F1 - measure = \frac{2\ x\ precision\ x\ recall}{precision+recall} \quad (3)$$

3. Hasil dan Pembahasan

Data digunankan pada penelitian ini bejumlah 4846 judul berita. Jumlah sentimen netral 2879, jumlah sentimen positif 1363, dan jumlah sentiment negative 604. Data dilakukan pembagian data latih dan data uji dengan persentase 70:30, 80:20, dan 90:10. Table 3 menunjukan hasil dari pembagian data latih dan data uji.

Data dilakukan pre-proceesing untuk membersihkan data text. Proses pre-processing diantaranya case folding, tokenization, stop word removal, dan steaming. Table 4 menunjukan hasil dari pre-processing.

Hasil pre-processing dilakukan klasifikasi sentiment dengan beberapa metode machine learning. Table 5, 6, dan 7 menunjukan hasil klasifikasi.

Metode multinomial mendapatkan hasil terbaik Ketika pembagian persentase 70:30 yaitu f1-measure 69,79 persen. Metode logistic regression mendapatkan hasil terbaik pada pembagian 80:20 yaitu f1-measure 73,94 persen. Metode support vector machine mendapatkan hasil terbaik pada pembagian 80:20 yaitu 70,07 persen. Metode multi-layer perceptron mendapatkan hasil terbaik pada pembagian 90:10 yaitu 72,76 persen. Hal ini dikarenakan metode multi-layer perceptron membutuhkan data latih yang banyak agar dapat memaksimalkan hidden layer yang dimiliki untuk mengetahui klasifikasinya. Jika jumlah data latih sedikit maka hidden layer multi-layer perceptron tidak berjalan dengan optimal. Hal ini kedepannya jika ingin menggunakan multi-layer perceptron harap memperbanyak data latih, baik dengan menembah manual ataupun dengan metode augmentasi data. Metode stochastic gradient descent mendapatkan hasil terbaik pada pembagian 90:10 yaitu 71,16 percen. Metode decision trees mendapatkan hasil terbaik pada pembagian 70:30 yaitu 67,99 persen. Secara keseluruhan hasil terbaik didaptkan oleh metode logistic regression. Perbedaan hasil yang tidak jauh menunjukan bahwa proses metode telah sesuai dan tidak adanya underfitting ataupun overfitting. Hasil yang baik ini dapat dioptimalkan lagi dengan menggunkan metode ektraksi fitur yang lebih baik lagi seperti menggunkan pembobotan fasttext, glove ataupun yang lainnya. Hal ini tentu dengan menambahkan transfer learning yang telah dilakukan.

Table 5 Result Percentage 70:30

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	69,79	69,61	70,56
Logistic Regression	**73,93**	**73,27**	**73,93**
Support Vector Machine	66,65	70,02	70,56
Multi-Layer perceptron	70,27	70,04	70,63
Stochastic Gradient Descent	68,89	68,65	69,25
Decision Trees	67,99	68,22	67,81

Table 6 Result Percentage 80:20

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	69,73	69,85	71,03
Logistic Regression	**73,94**	**73,94**	**74,63**
Support Vector Machine	70,07	75,29	73,60
Multi-Layer perceptron	69,32	69,28	69,38
Stochastic Gradient Descent	70,27	70,17	70,41
Decision Trees	67,27	67,03	67,62

Table 7 Result Percentage 90:10

Method	F-1 Measure	Precision	Recall
Multinomial NaiveBayes	68,65	69,27	69,69
Logistic Regression	**73,70**	**73,81**	**74,43**
Support Vector Machine	68,91	71,71	71,75
Multi-Layer perceptron	72,76	72,79	72,99
Stochastic Gradient Descent	71,16	71,08	71,75
Decision Trees	66,12	65,84	66,60

DOI: https://doi.org/10.31004/riggs.v1i1.4

4. Kesimpulan

Penelitian ini melakukan perbandingan metode machine learning untuk klasifikasi sentiment dengan ektraksi fitur n-gram. Data digunakn dengan jumlah total 4868 headline berita. Pembobotan n-gram dan klasifikasi dengan machine learning menunjukan hasil yang baik. Hasil tertinggi didapatkan yaitu f1-measure, presisi, recall sebesar 73,94; 73,94; dan 74,63. Berdasarkan hasil pengujian yang telah dilakukan, n-gram mampu melakukan ektrasi fitir dari data yang ada.

Daftar Rujukan

[1] S. Kurniawan, W. Gata, D. A. Puspitawati, N. -, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 176–183, 2019, doi: 10.29207/resti.v3i2.935.

[2] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, vol. 8, pp. 131662–131682, 2020, doi: 10.1109/ACCESS.2020.3009626.

[3] E. Kušen and M. Strembeck, "Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections," *Online Soc. Networks Media*, vol. 5, pp. 37–50, 2018, doi: 10.1016/j.osnem.2017.12.002.

[4] S. M. Shuhidan, S. R. Hamidi, S. Kazemian, S. M. Shuhidan, and M. A. Ismail, "Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm," in *Advances in Intelligent Systems and Computing*, vol. 739, A. M. Lokman, T. Yamanaka, P. Lévy, K. Chen, and S. Koyama, Eds. Singapore: Springer Singapore, 2018, pp. 64–72. doi: 10.1007/978-981-10-8612-0_8.

[5] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 782–796, 2014, doi: 10.1002/asi.23062.

[6] S. Taj, B. B. Shaikh, and A. Fatemah Meghji, "Sentiment Analysis of News Articles: A Lexicon based Approach," in *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Jan. 2019, pp. 1–5. doi: 10.1109/ICOMET.2019.8673428.

[7] P. R. Nagarajan, M. T, S. R. A. M. Hari, K. K, and M. G, "Certain Investigation On Cause Analysis Of Accuracy Metrics In Sentimental Analysis On News Articles," in *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Dec. 2021, pp. 1033–1037. doi: 10.1109/ICECA52323.2021.9675846.

[8] A. Agarwal, "Sentiment Analysis of Financial News," in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2020, pp. 312–315. doi: 10.1109/CICN49253.2020.9242579.

[9] W. van Atteveldt, M. A. C. G. van der Velden, and M. Boukes, "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms," *Commun. Methods Meas.*, vol. 15, no. 2, pp. 121–140, Apr. 2021, doi: 10.1080/19312458.2020.1869198.

[10] J. Nothman, H. Qin, and R. Yurchak, "Stop Word Lists in Free Open-source Software Packages," in *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, 2018, pp. 7–12. doi: 10.18653/v1/W18-2502.

[11] A. M. Priyatno, M. M. Muttaqi, F. Syuhada, and A. Z. Arifin, "Deteksi bot spammer twitter berbasis time interval entropy dan global vectors for word representations tweet's hashtag," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 37–46, Jan. 2019, doi: 10.26594/register.v5i1.1382.

[12] R. Aulianita, L. Utami, N. Musyaffa, G. Wijaya, A. Mukhayaroh, and A. Yoraeni, "Sentiment Analysis Review Of Smartphones With Artificial Intelligent Camera Technology Using Naive Bayes and n-gram Character Selection," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, p. 012076, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012076.

[13] M. A. P. Subali and C. Fatichah, "Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, p. 219, 2019, doi: 10.25126/jtiik.2019621105.

[14] C. D. Manning, P. Raghavan, and H. Schütze, "Text classification and Naive Bayes," in *Introduction to Information Retrieval*, no. c, Cambridge University Press, 2008, pp. 234–265. doi: 10.1017/CBO9780511809071.014.

[15] "sklearn.naive_bayes.MultinomialNB — scikit-learn 1.1.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html (accessed Jul. 30, 2022).

[16] J. L. Morales and J. Nocedal, "Remark on 'algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization,'" *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 1–4, Nov. 2011, doi: 10.1145/2049662.2049669.

[17] "sklearn.linear_model.LogisticRegression — scikit-learn 1.1.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (accessed Jul. 30, 2022).

[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.

[19] J. Platt and others, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. large margin Classif.*, vol. 10, no. 3, pp. 61–74, 1999.

[20] "sklearn.svm.SVC — scikit-learn 1.1.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html (accessed Jul. 30, 2022).

[21] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, vol. 2015 Inter, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.

[23] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd Int. Conf. Learn. Represent. ICLR*

DOI: https://doi.org/10.31004/riggs.v1i1.4

2015 - Conf. Track Proc., pp. 1–15, Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[24] "sklearn.neural_network.MLPClassifier — scikit-learn 1.1.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html (accessed Jul. 30, 2022).

[25] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, 2002, p. 694. doi: 10.1145/775047.775151.

[26] "sklearn.linear_model.SGDClassifier — scikit-learn 1.1.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html (accessed Jul. 30, 2022).

[27] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," in *Ensemble Machine Learning: Methods and Applications*, Boston, MA: Springer, Boston, MA, 2012, pp. 157–175. doi: 10.1007/9781441993267_5.

[28] "Random forests - classification description." https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed Jul. 30, 2022).

[29] A. M. Priyatno, F. M. Putra, P. Cholidhazia, and L. Ningsih, "Combination of extraction features based on texture and colour feature for beef and pork classification," *J. Phys. Conf. Ser.*, vol. 1563, no. 1, p. 012007, Jun. 2020, doi: 10.1088/1742-6596/1563/1/012007.

DOI: https://doi.org/10.31004/riggs.v1i1.4