

## PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE DAN MULTI-LAYER PERCEPTRON PADA KLASIFIKASI TOPIK BERITA

Sudianto<sup>1</sup>, Asti Dwi Sripamuji<sup>2</sup>, Imada Ramadhanti<sup>3</sup>, Risa Riski Amalia<sup>4</sup>,  
Julian Saputra<sup>5</sup>, Bagas Prihatnowo<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Teknik Informatika, Institut Teknologi Telkom Purwokerto, Indonesia

email: sudianto@ittelkom-pwt.ac.id<sup>1</sup>

### Abstrak

Salah satu aktivitas masyarakat Indonesia di internet adalah mengakses berita *online* dengan persentase 15,5%. Dalam mencerna suatu informasi berita *online*, masyarakat Indonesia masih memiliki kebiasaan yang buruk ketika membaca lebih dari satu sumber media *online* pada kasus yang sama. Sehingga perlu adanya klasifikasi berita secara otomatis. Penelitian sebelumnya menggunakan *Mutual Information and Bayesian Network* untuk mengkategorikan suatu berita, dalam format data tekstual dengan akurasi 75.34%. Penelitian ini memberikan suatu inovasi baru berupa pengimplementasian algoritma *Support Vector Machine* dan *Multi-Layer Perceptron* untuk klasifikasi menggunakan pembelajaran supervised atau yang disebut *backpropagation*. Tujuan dari pengklasifikasian topik berita menggunakan Algoritma *Support Vector Machine* memudahkan pembaca untuk menemukan informasi yang pembaca butuhkan, dengan secara otomatis mengelompokkan berita ke dalam kategori tertentu sesuai informasi yang terkandung. Metode yang digunakan yaitu metode *Term Frequency – Inverse Document Frequency* untuk *feature selection* data pelatihan dan algoritma *Support Vector Machines* dan *Multi-Layer Perceptron* untuk klasifikasi. Hasil yang diperoleh menunjukkan skor akurasi sebesar 74% pada SVM dan 78% pada MLP. Sementara itu, nilai *precision* dan *recall* yaitu 76% dan 74% pada SVM serta 79% dan 78% pada MLP.

**Kata kunci:** Berita, Klasifikasi, Support Vector Machine, Multi-Layer Perceptron, TF-IDF.

### Abstract

One of the activities of Indonesian people on the internet is to access *online* news with a percentage of 15.5%. In digesting online news information, Indonesians still have a bad habit when reading more than one online media source in the same case. So there needs to be an automatic classification of news. Previous researchers used the *Mutual Information and Bayesian Network* to categorize news stories, in textual data format with 75.34% accuracy. This research provides a new innovation in the form of implementing *support vector machine* and *multi-layer perceptron* algorithms for classification using supervised learning or called *backpropagation*. The purpose of classifying news topics using *Machine's Vector Support Algorithm* makes it easy for readers to find the information readers need, by automatically grouping news into specific categories according to the information contained. The method used is the *Term Frequency – Inverse Document Frequency* method for *feature selection* of training data and algorithms *support vector machines* and *multi-layer perceptron* for classification. The results showed an accuracy score of 74% on SVM and 78% on MLP. Meanwhile, *precision* and *recall* values are 76% and 74% on SVM and 79% and 78% in MLP.

**Keywords:** News, Classification, Support Vector Machine, Multi-Layer Perceptron, TF-IDF.

## PENDAHULUAN

Kebutuhan akan ketersediaan jaringan komunikasi dan internet saat ini sangat tinggi dengan meningkatnya ketergantungan manusia akan peranan Teknologi Informasi [1]. Orang-orang menghabiskan lebih banyak waktu berinteraksi di media sosial, karena kemudahan akses dari mana dan kapan saja.

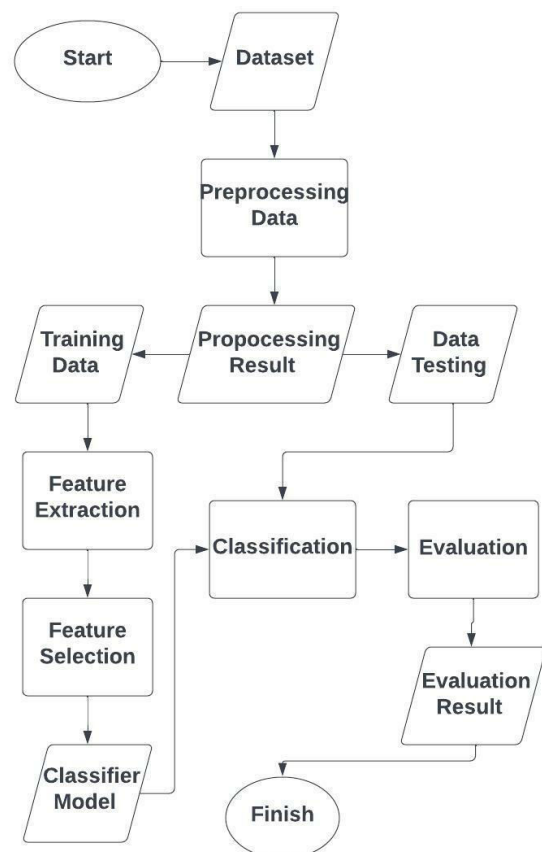
Media sosial sudah dianggap menjadi kebutuhan hidup dan menjadi hal yang sangat digemari oleh masyarakat. Berdasarkan data Kominfo, aktivitas masyarakat Indonesia dalam berinternet salah satunya untuk kegiatan mengakses berita portal *online* dengan persentase 15,5%. Namun, tingginya akses berinternet tidak berbanding lurus dengan skor informasi dan indeks literasi digital yang belum “baik” (4.00) dari tiap provinsi [2] [3]. Permasalahan yang ada saat ini, dalam mencerna suatu informasi berita *online* masyarakat memiliki kebiasaan buruk dalam membaca lebih dari satu sumber media *online* untuk isu yang sama, pembaca acap kali mengabaikan mengetahui latar belakang portal media *online* dari *website* yang dibaca [2] [4]. Berdasarkan kebiasaan negatif tersebut, maka diperlukan klasifikasi berita otomatis untuk mengatur berita dan menemukan informasi, serta berita yang pembaca butuhkan. Salah satu cara untuk mengkategorikan berita ke dalam kategori tertentu berdasarkan informasi yang terkandung melalui klasifikasi teks [5][6][7]. Terdapat beberapa metode pembelajaran mesin yang dapat digunakan untuk mengklasifikasi teks, diantaranya algoritme *K-Nearest Neighbors* (KNN), *Gaussian Naïve Bayes*, *Random Forest* (RF), *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP), *Ridge*, dan algoritme *Stochastic Gradient Descent* (SGD). Pada Penelitian ini, klasifikasi topik berita menggunakan algoritme *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP).

Penelitian sebelumnya, oleh Fahmi Salman Nurfikri, *et al* [8] melakukan klasifikasi berita menggunakan klasifikasi teks dengan metode Mutual Information (MI) dan Bayesian Network yang menunjukkan hasil 75,34%. *Dataset* yang digunakan masih sedikit dengan jumlah 360 *dataset* berita dan dilakukan pengklasifikasian secara *multi-label* [9]. Sedangkan pada penelitian ini, menggunakan metode algoritme *Support Vector Machine* (SVM) dan *Multilayer Perceptron* (MLP). Proses tahapan *preprocessing* menggunakan *case folding* dan *cleaning* yang bertujuan untuk

mendapatkan kinerja sistem yang baik, selanjutnya dilakukan *feature extraction* untuk mengurangi jumlah data berulang dari kumpulan pada *dataset*. *Feature selection* menggunakan TF-IDF yang digunakan untuk mengukur kemunculan sebuah kata dalam suatu dokumen. Batasan topik pembahasan pada penelitian ini mengklasifikasikan informasi berita (*News*) ke dalam beberapa kategori berita yaitu Otomotif, Bisnis Ekonomi, Nasional, Olahraga, *Lifestyle*, *Travel*, Teknologi dan Bola. Masing-masing kategori berita memiliki 1250 data tekstual, dengan total jumlah *dataset* yang digunakan dalam penelitian ini sebanyak 10.000 *dataset*. *Dataset* ini didapatkan dari Indonesian News Corpus [10]. Oleh karena itu, penelitian ini bertujuan untuk klasifikasi topik berita *online* sesuai dengan topik berita. Sehingga pembaca berita lebih mudah untuk memperoleh topik berita yang dibaca.

## METODE

Metode untuk klasifikasi topik berita secara otomatis, ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

1. *Dataset*

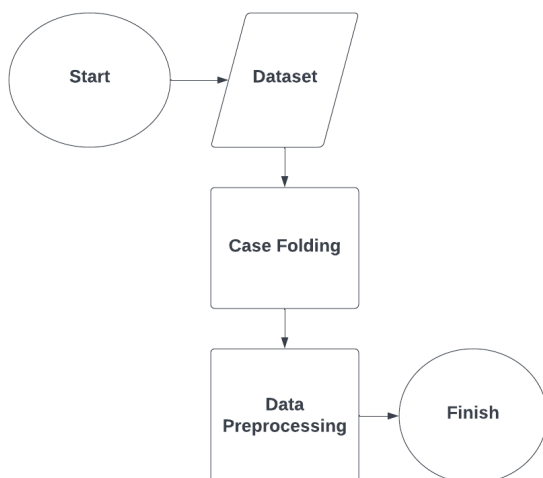
Langkah pertama adalah mengumpulkan *dataset* berupa data berita berbasis teks artikel artikel berbahasa Indonesia yang didapat dari berbagai sumber situs berita Indonesia [11][12][13][14][15][16] yang diambil dari beberapa situs *web* selama 6 bulan, yaitu dari Juli 2015 hingga Desember 2015 [10]. Pada penelitian ini, mendefinisikan 8 label kelas berdasarkan kategori berita, yaitu Otomotif, Bisnis Ekonomi, Nasional, Olahraga, *Lifestyle*, *Travel*, Teknologi dan Bola. Masing-masing label kelas memiliki 1250 data tekstual. Sehingga, pada penelitian ini jumlah data yang digunakan adalah 10.000 *dataset*. Berikut adalah distribusi data pada *dataset*, yang ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Data Setiap Label

No	Nama Label	Jumlah Dokumen
1	Otomotif	1250
2	Bisnis Ekonomi	1250
3	Nasional	1250
4	Olahraga	1250
5	Lifestyle	1250
6	Travel	1250
7	Teknologi	1250
8	Bola	1250

2. *Preprocessing Data*

Pada tahap ini data dipersiapkan untuk diproses lebih lanjut, bertujuan untuk mendapatkan kinerja sistem yang baik. Tahapan melakukan *preprocessing* data terdiri dari *case folding* dan *cleaning* seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Alur *preprocessing*

a. *Case Folding*

Pada tahap ini, semua huruf yang ada di dalam dokumen akan diubah menjadi huruf kecil. Tujuannya adalah untuk

memperkecil variasi kata dalam artikel berita [17].

b. *Cleaning*

Pada tahap ini, karakter lain selain huruf, seperti tanda baca, simbol dan nomor akan dihapus dari teks berita asli [18].

3. *Feature Extraction*

*Feature extraction* adalah proses pemeriksaan ada atau tidaknya setiap kata pada daftar kosakata yang diperoleh dari setiap dokumen teks berita. Hal ini dilakukan untuk membantu mengurangi jumlah data berulang dari kumpulan *dataset* [19].

4. *Feature Selection*

Pada tahap *feature selection*, penelitian ini menerapkan TF-IDF sebagai fitur untuk mengidentifikasi kata-kata penting atau kata-kata yang lebih jarang muncul dalam sebuah data teks. Kata-kata terpilih dianggap lebih informatif dan efektif untuk dieksekusi pada model *classifier* [20] [21].

Teknik TF-IDF diterapkan terhadap setiap kalimat sebelum dilakukan klasifikasi data. TF-IDF merupakan multiplikasi dari TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF digunakan untuk mengukur kemunculan sebuah kata dalam suatu dokumen [22] [23].

$$TF = \frac{f_k}{f_t} \tag{1}$$

Pada persamaan (1),  $f_k$  yaitu jumlah kemunculan sebuah kata dalam suatu kalimat dan  $f_t$  yaitu total kata dalam suatu dokumen. IDF memberikan nilai pada setiap kata dalam suatu korpus.

$$IDF = \log\left(\frac{N}{D_f}\right) \tag{2}$$

Dimana  $N$  merupakan total dokumen dalam suatu korpus dan  $D_f$  merupakan banyaknya dokumen yang mengandung kata tertentu dalam suatu korpus [24] [25] [26]

5. *Classifier Model*

Pada tahap ini, akan dilakukan perancangan model untuk mempersiapkan proses *training* data. Setiap kalimat dalam teks berita akan diberikan label melalui beberapa algoritme klasifikasi seperti *K-Nearest Neighbors* (KNN), *Gaussian Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP), *Ridge*, dan algoritme *Stochastic Gradient Descent* (SGD). Data akan ditraining menggunakan 7 algoritme

klasifikasi tersebut dan akan diketahui nilai akurasi.

### 6. Classification

Setelah melalui proses *training* data, tahap selanjutnya adalah memilih algoritme dengan nilai akurasi tertinggi untuk digunakan pada proses klasifikasi topik berita. Dari hasil akurasi *training* data, penelitian ini menerapkan model algoritme *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP) untuk mengklasifikasikan topik berita.

*Support Vector Machine* (SVM) adalah salah satu algoritme yang sering digunakan untuk analisis klasifikasi data. Pada dasarnya, konsep klasifikasi data pada SVM yaitu untuk mendapatkan pemisah *hyperplane* yang optimal antara positif dan negatif [27]. Untuk klasifikasi data, *hyperplane* pemisah optimal,  $h$  ditentukan oleh nilai  $x$  dimana fungsi keputusan  $D(x)$  sama dengan nol. Diberikan *dataset* pelatihan  $(x_i; y_i)$  dari ruang *input*  $x$ , fungsi keputusan didefinisikan sebagai berikut:

$$D(x) = \sum_{i=1}^n \alpha_i^* y_i H(x_i, x) \quad (3)$$

$\alpha_i^*$  adalah solusi maksimal untuk masalah optimasi kuadrat berikut:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j H(x_i, x_j) \quad (4)$$

Subjek akan dibatasi:

$$\sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq \frac{C}{n}, i = 1, \dots, n \quad (5)$$

Dimana  $H$  adalah produk dalam kernel dan  $C$  adalah parameter regularisasi. Untuk polinomial dari derajat  $q$  produk dalam kernel sama dengan:

$$H(x, x^i) = [(x \cdot x^i) + 1]^q \quad (6)$$

Untuk fungsi radial basis dari bentuk:

$$f(x) = \sin\left(\sum_{i=1}^n \alpha_i \exp\left\{-\frac{|x-x_i|^2}{\sigma^2}\right\}\right) \quad (7)$$

Dimana  $\sigma$  sama dengan lebar, maka produk kernel adalah sama dengan:

$$H(x, x^i) = \exp\left\{-\frac{|x-x^i|^2}{\sigma^2}\right\} \quad (8)$$

Pemisahan kompleksitas model dan dimensi, SVM diharapkan akan berkinerja relatif baik dibandingkan dengan metode pembelajaran lain dalam *domain* yang memiliki dimensi tinggi secara *inheren* [28][29][30][31].

*Multi-layer Perceptron* (MLP) merupakan jaringan saraf tiruan terstruktur, yang menggunakan metode pembelajaran *supervised* dan dikenal sebagai *backpropagation* untuk klasifikasi [32][33]. Pada *input layer* nya, terdapat saraf neuron yang mewakili fitur *input*. Selanjutnya, pada *hidden layer*, dibutuhkan *input* dari *layer* pertama dan menerapkan penjumlahan linier berbobot dan aktivasi non-linier pada fungsi mereka. Fungsi aktivasi yang umum digunakan adalah *sigmoid*, yang ditunjukkan oleh fungsi dibawah ini:

$$x(a_i) = \tanh(a_i) \quad (9)$$

Persamaan (9) menunjukkan tangen hiperbolik dan persamaan (10) adalah fungsi logistik.

$$x(a_i) = (1 + e^{-a_i})^{-1} \quad (10)$$

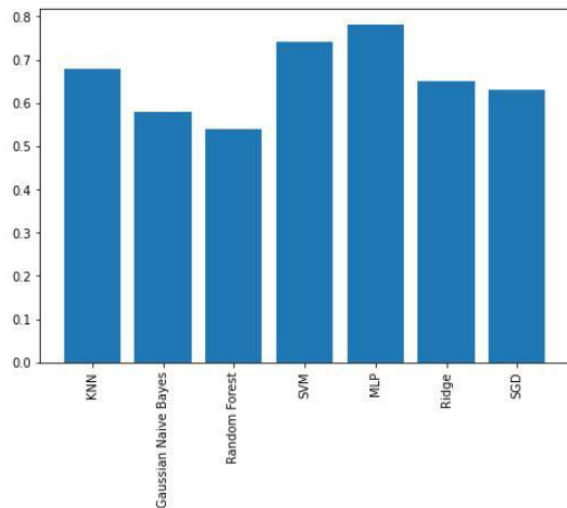
Fungsi aktivasi diambil sebagai *input* oleh *layer* berikutnya, yaitu *output layer*. Karena struktur MLP terhubung sepenuhnya, maka *node* dalam *multilayer* akan terhubung satu sama lain melalui bobot. Ketika setiap data diproses, koneksi bobot akan diubah berdasarkan pada perbedaan jumlah kesalahan antara hasil yang diharapkan dan keluarannya. Teknik pembelajaran ini disebut dengan *backpropagation* [34][35].

### 7. Evaluasi

Setelah proses klasifikasi data selesai, maka tahap terakhir adalah mengevaluasi kinerja sistem menggunakan algoritma *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM) yang ditunjukkan dengan nilai akurasi, *precision* dan *recall* pada setiap kategori topik berita.

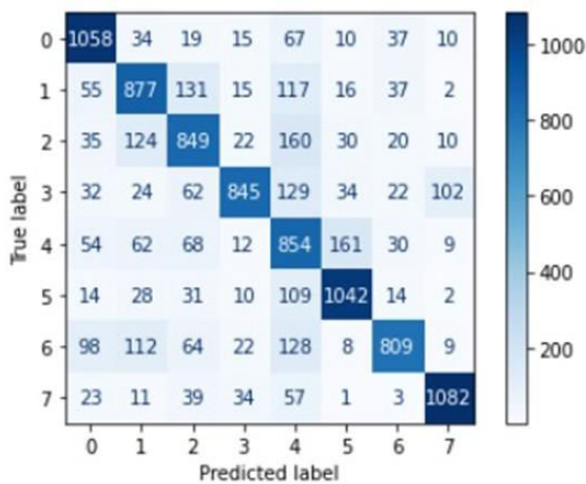
### HASIL DAN PEMBAHASAN

Proses training menggunakan 7 algoritma klasifikasi diantaranya algoritma *K-Nearest Neighbors* (KNN), *Gaussian Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP), *Ridge*, dan algoritma *Stochastic Gradient Descent* (SGD). Berdasarkan proses training terhadap 7 algoritma klasifikasi, penelitian ini mendapatkan nilai akurasi sebagai berikut.



Gambar 3. Hasil akurasi trining 7 algoritme klasifikasi

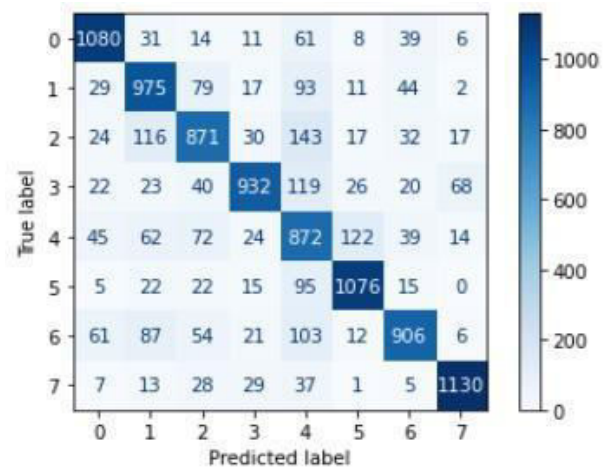
Hasil Akurasi training 7 algoritma klasifikasi diatas posisi tertinggi pertama diraih oleh *Multilayer Perceptron* (MLP) , dan algoritma tertinggi kedua diraih oleh *Support Vector Machine* (SVM) serta algoritma klasifikasi lainnya berdasarkan grafik mendapatkan nilai akurasi dibawah 70% sehingga penelitian ini akan mengevaluasi hasil prediksi klasifikasi topik berita menggunakan algoritma *Multilayer Perceptron* (MLP) dan *Support Vector Machine* (SVM).



Gambar 4. Hasil confusion matrix SVM

Hasil *confusion matrix* model SVM pada Gambar 4. menggunakan pembagian *dataset* sebanyak 1250 artikel berita terhadap setiap kategori topik berita. Topik berita yang diprediksi terdiri dari otomotif, Bisnis ekonomi, nasional, olahraga, *lifestyle*, *travel*, teknologi, dan bola yang diberi label dari 0 sampai 7. Label 0 sebagai kategori topik otomotif menghasilkan nilai prediksi sebanyak 1058, label 1 sebagai

kategori bisnis otomotif mendapatkan nilai prediksi sebanyak 877, kategori topik nasional pada label 3 mendapatkan nilai prediksi sebanyak 845, kategori topik olahraga mendapatkan nilai prediksi 854, kategori topik *lifestyle* mendapatkan nilai prediksi sebanyak 809, dan kategori bola dengan nilai prediksi 1082.



Gambar 5. Hasil confusion matrix Multilayer Perceptron

Hasil *confusion matrix* dengan model *Multilayer Perceptron* mendapatkan nilai prediksi pada pada kategori topik otomotif 1080 dari 1250 *dataset* yang diprediksi, kategori topik bisnis ekonomi dengan label 1 mendapatkan nilai prediksi sebanyak 975, kategori topik nasional mendapatkan nilai prediksi sebanyak 871, kategori olahraga mendapatkan nilai prediksi 932, label 4 dengan topik *lifestyle* mendapatkan nilai prediksi label sebanyak 872, label 5 atau kategori *travel* mendapatkan nilai prediksi label sebanyak 1076, label 6 dengan topik teknologi mendapatkan nilai prediksi label sebanyak 906 dan topik bola label 7 mendapatkan nilai prediksi sebanyak 1130.

	precision	recall	f1-score	support
0	0.77	0.85	0.81	1250
1	0.69	0.70	0.70	1250
2	0.67	0.68	0.68	1250
3	0.87	0.68	0.76	1250
4	0.53	0.68	0.59	1250
5	0.80	0.83	0.82	1250
6	0.83	0.65	0.73	1250
7	0.88	0.87	0.87	1250
accuracy			0.74	10000
macro avg	0.76	0.74	0.74	10000
weighted avg	0.76	0.74	0.74	10000

Gambar 6. Nilai *precision*, *recall*, dan *F1-Score* pada setiap kategori topik berita menggunakan algoritme Support Vector Machine (SVM)

Nilai rata-rata bobot dari 7 kategori topik klasifikasi yang terdiri dari *precision*, *recall* dan *f1-score* berturut-turut sebesar 0,76, 0,74 dan 0,74.

	precision	recall	f1-score	support
0	0.83	0.88	0.85	1250
1	0.74	0.75	0.75	1250
2	0.73	0.70	0.71	1250
3	0.84	0.77	0.80	1250
4	0.56	0.72	0.63	1250
5	0.87	0.84	0.85	1250
6	0.82	0.72	0.77	1250
7	0.94	0.88	0.91	1250
accuracy			0.78	10000
macro avg	0.79	0.78	0.79	10000
weighted avg	0.79	0.78	0.79	10000

Gambar 7. Nilai *precision*, *recall*, dan *F1-Score* pada setiap kategori topik berita menggunakan algoritma Multilayer Perceptron (MLP).

Nilai rata-rata bobot dari 7 kategori topik klasifikasi yang terdiri dari *precision*, *recall* dan *F1-score* berturut-turut sebesar 0,79, 0,78 dan 0,79.



Gambar 8. *WordCloud* dari algoritma Multilayer Perceptron beberapa kategori topik berita

Pada Gambar 8. *WordCloud term* otomotif dan penjualan paling dominan jika dibandingkan dengan *term* lainnya. Topik Bisnis mendapatkan *term* pertumbuhan yang paling besar. Kategori topik nasional mendapatkan *term* yang paling baik adalah DPR. Kategori dengan topik olahraga mendapatkan *term* Valentino Rossi yang paling banyak. Selanjutnya topik kategori *lifestyle* mendapatkan *term* yang paling besar adalah kata yang. *WordCloud* dengan topik *travel* mendapatkan *term* yang paling besar adalah wisatawan. Kemudian pada topik teknologi mendapatkan *term* yang paling besar dan berada ditengah adalah *term* pengguna. Serta *WordCloud* dengan topik bola, *term* yang paling mendominasi adalah *term* Laga Group.

### KESIMPULAN

Klasifikasi topik berita dapat dilakukan menggunakan algoritma algoritma SVM dan MLP dengan melalui *feature selection* menggunakan TF-IDF pada *preprocessing* data. Tahap *feature selection* menyiapkan data pelatihan berdasarkan tingkat kepentingan korpus pada dataset. Algoritma SVM dan MLP dikatakan dapat menjadi algoritma yang sesuai untuk tugas klasifikasi topik berita karena kedua algoritma tersebut memiliki skor akurasi yang lebih unggul dari 5 algoritma yang lainnya. Algoritma SVM dapat melakukan klasifikasi dengan skor akurasi sebesar 74%. Sementara itu, akurasi yang diperoleh menggunakan algoritma MLP yaitu sebesar 78%. Evaluasi lanjut berupa rata-rata bobot untuk *precision* dan *recall* dengan skor berturut-turut sebesar 76% dan 74% pada SVM serta 79% dan 78% pada MLP. Vektorisasi dalam tahapan *feature selection* dapat dilakukan dengan metode lainnya seperti *FastText* dan *Word2Vec* sebagai penelitian lanjutan. Selain itu, klasifikasi juga dapat melalui deep learning sehingga evaluasi lebih lanjut terkait perbandingan performa algoritma dapat diketahui sebagai pengembangan penelitian ini pada masa mendatang.

## UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih kepada Teknik Informatika, Institut Teknologi Telkom Purwokerto atas dukungannya dalam menyelesaikan penelitian ini.

## REFERENSI

- [1] I. Etin, "Pengelolaan Sistem Informasi Akademik Perguruan Tinggi Berbasis Teknologi Informasi dan Komunikasi," *J. Penelit. Pendidik.*, vol. 12, no. 1, pp. 51–67, 2006, [Online]. Available: <https://files.indihomestudy.com/pdf/ac6c9913-953a-4c0a-9426-800cfd8e9e.pdf>.
- [2] Kemkominfo, "Survei Literasi Digital Indonesia 2020," *Katadata Insight Cent.*, no. November, p. 32, 2020.
- [3] K. B. Dinata, "ANALISIS KEMAMPUAN LITERASI DIGITAL MAHASISWA COVID-19 proses Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan . Dampak yang mandiri . Salah satu kemampuan yang berperan cukup penting dalam memfasilitasi," *Edukasi J. Pendidik.*, vol. 19, no. 1, pp. 105–119, 2021, doi: 10.31571/edukasi.v19i1.
- [4] I. G. Hartari, N. M. Gelgel, and N. L. Purnawan, "Analisis Isi Berita Kekerasan Seksual Tribunnews.Com (Periode Berita Desember 2018)," *E-Jurnal Mediu.*, vol. 1, no. 2, pp. 1–12, 2019.
- [5] P. Widodo, J. A. Putra, S. Afiadi, A. Z. Arifin, and D. Herumurti, "Klasifikasi Kategori Dokumen Berita Berbahasa Indonesia Dengan Metode Kategorisasi Multi-Label Berbasis Domain Specific Ontology," *J. Ilm. Teknol. Infomasi Terap.*, vol. 2, no. 2, pp. 101–112, 2016, doi: 10.33197/jitter.vol2.iss2.2016.100.
- [6] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, p. 269, 2018, doi: 10.25126/jtiik.201853810.
- [7] F. Handayani and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2015, doi: 10.15294/jte.v7i1.8585.
- [8] F. S. Nurfikri, M. S. Mubarak, and Adiwijaya, "News topic classification using mutual information and Bayesian network," *2018 6th Int. Conf. Inf. Commun. Technol. ICoICT 2018*, vol. 0, no. c, pp. 162–166, 2018, doi: 10.1109/ICoICT.2018.8528806.
- [9] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 284, 2019, doi: 10.30865/mib.v3i4.1410.
- [10] A. RAHUTOMO, FAISAL; MIQDAD MUADZ MUZAD, "Indonesian News Corpus." Mendeley Data V1, 2018, doi: 10.17632/2zpbjs22k3.1.
- [11] W. A. S. J. H. M. Nugroho, "Kompas.com," [Online]. Available: <https://inside.kompas.com/>.
- [12] W. D. B. S. Y. T. Aryanto, "Tempo.co," [Online]. Available: <https://www.tempo.co/>.
- [13] "Merdeka.com," [Online]. Available: <https://www.merdeka.com>.
- [14] M. R. Djarot, "Republika.co.id," [Online]. Available: <https://www.republika.co.id/>.
- [15] I. W. T. Pribadi, "Viva.co.id," [Online]. Available: <https://www.viva.co.id/>.
- [16] D. D. S. F. M. Putra, "Tribunnews.com," [Online]. Available: <http://www.tribunnews.com/>.
- [17] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Inf.*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.
- [18] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, no. July, 2019, doi: 10.1109/ICIC47613.2019.8985884.
- [19] R. S. Rajput and S. R. Patra, "Survey Paper on Sentiment Analysis using NLP," vol. 1, no. 02, 2021.
- [20] A. Kumar, A. Jaiswal, S. Garg, S. Verma, and S. Kumar, "Sentiment Analysis Using Cuckoo Search for Optimized Feature Selection on Kaggle Tweets," *Int. J. Inf. Retr. Res.*, vol. 9, no. 1, pp. 1–15, 2018, doi: 10.4018/ijirr.2019010101.
- [21] Y. Zhai, W. Song, X. Liu, L. Liu, and X. Zhao, "A Chi-Square Statistics Based Feature Selection Method in Text Classification," *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, vol. 2018-November, pp. 160–163, 2019, doi: 10.1109/ICSESS.2018.8663882.
- [22] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.
- [23] S. Ghosh and M. S. Desarkar, "Class Specific TF-IDF Boosting for Short-text Classification," pp. 1629–1637, 2018, doi: 10.1145/3184558.3191621.
- [24] Z. Zhu, J. Liang, D. Li, H. Yu, and G. Liu, "Hot Topic Detection Based on a Refined TF-IDF Algorithm," *IEEE Access*, vol. 7, pp. 26996–27007, 2019, doi: 10.1109/ACCESS.2019.2893980.
- [25] S. W. Kim and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0192-7.
- [26] C. Z. Liu, Y. X. Sheng, Z. Q. Wei, and Y. Q. Yang, "Research of Text Classification Based on Improved TF-IDF Algorithm," *2018 IEEE Int. Conf. Intell. Robot. Control Eng. IRCE 2018*, no. 2, pp. 69–73, 2018, doi: 10.1109/IRCE.2018.8492945.
- [27] X. Yan and M. Jia, "A novel optimized SVM

- classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing,” *Neurocomputing*, vol. 313, pp. 47–64, 2018, doi: 10.1016/j.neucom.2018.05.002.
- [28] S. Puri and S. P. Singh, *An Efficient Hindi Text Classification Model Using SVM*, no. March 2021. Springer Singapore, 2019.
- [29] F. Wang, Z. Zhen, B. Wang, and Z. Mi, “applied sciences Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting,” doi: 10.3390/app8010028.
- [30] T. B. Shahi, “Nepali News Classification using Naïve Bayes , Support Vector Machines and Neural Networks,” pp. 18–22, 2018.
- [31] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, “A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 7, pp. 1–23, 2018, doi: 10.1142/S0218001418590127.
- [32] B. T. Pham, M. D. Nguyen, K. T. T. Bui, I. Prakash, K. Chapi, and D. T. Bui, “A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil,” *Catena*, vol. 173, no. July 2018, pp. 302–311, 2019, doi: 10.1016/j.catena.2018.10.004.
- [33] J. Jeyanathan, ... P. J.-I. J. of, and undefined 2018, “Transform based Classification of Breast Thermograms using Multilayer Perceptron Back Propagation Neural Network,” *Acadpubl.Eu*, vol. 118, no. 20, pp. 1955–1961, 2018, [Online]. Available: <http://www.acadpubl.eu/hub/2018-118-21/articles/21c/7.pdf>.
- [34] N. B. Gaikwad, V. Tiwari, A. Keskar, and N. C. Shivaprakash, “Efficient FPGA Implementation of Multilayer Perceptron for Real-Time Human Activity Classification,” *IEEE Access*, vol. 7, pp. 26696–26706, 2019, doi: 10.1109/ACCESS.2019.2900084.
- [35] C. Kar, A. Kumar, and S. Banerjee, “Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron,” *SN Appl. Sci.*, vol. 1, no. 9, 2019, doi: 10.1007/s42452-019-1134-8.