

## Klasifikasi Multi-Label pada Topik Berita Berbahasa Indonesia menggunakan Artificial Neural Network

Muhammad Fauzan<sup>1</sup>, Adiwijaya<sup>2</sup>, Mohamad Syahrul Mubarak<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung  
<sup>1</sup>zandut@student.telkomuniversity.ac.id, <sup>2</sup>adiwijaya@telkomuniversity.ac.id,  
<sup>3</sup>msyahrulmubarak@telkomuniversity.ac.id

---

### Abstrak

Ketidaksesuaian antara judul dan topik yang ada pada suatu berita menjadi masalah tersendiri dalam mencari berita. Hal ini penting dilakukan untuk membantu pembuat berita dalam menentukan topik yang tepat pada berita yang dibuatnya. Pada penelitian ini topik berita berbahasa Indonesia akan diklasifikasikan ke dalam suatu multi-label menggunakan Artificial Neural Network (ANN) sehingga didapatkan klasifikasi label yang tepat berdasarkan topik beritanya. Data berupa teks akan dijadikan masukan dan dilakukan ekstraksi ciri menggunakan *weighting* TF-IDF untuk mendapatkan data berbentuk vektor. Berdasarkan hasil penelitian didapatkan kombinasi parameter maksimum *epoch*, *learning rate*, jumlah *neuron*, *alpha* yang paling optimal yaitu 400, 0.02, 20, 0.3 dengan *loss* sebesar 0.0021.

**Kata kunci :** Topik berita, klasifikasi, multi-label, ANN.

---

### Abstract

The mismatch between the headline and the topic on a news becomes a separate issue in searching for news. This is important to help newsmakers in determining the right topic on the news that has been made. In this research the topic of Indonesian language news will be classified into a multi-label using Artificial Neural Network (ANN) so as to get the right label classification based on the news topic. Text data will be input and feature extraction using *weighting* TF-IDF to get vector-shaped data. Based on the result of the research, the maximum combination of epoch parameters, learning rate, number of neurons, alpha at the most optimum are 400, 0.02, 20, 0.3 with loss 0.0021.

**Keywords:** news topic, classification, multi-label, ANN.

---

### 1. Pendahuluan

Informasi merupakan suatu hal yang penting di era revolusi industri keempat. Pada era revolusi tersebut, terjadi revolusi digital dengan akses informasi dan pengetahuan yang tak terbatas [1] [2]. Seseorang dapat memiliki wawasan dan pengetahuan terkait informasi terbaru melalui koran, berita televisi, dan internet. Kemudahan pengaksesannya sendiri menjadikan informasi dapat mudah diakses dan dibagikan, terutama informasi melalui media digital, baik dalam bentuk berita maupun artikel. Menurut studi Media Insight Project, generasi Millenials membaca berita paling tidak sehari sekali [3]. Sering ditemui dalam pelbagai media sosial, seseorang sering membagikan berita pada laman sosial medianya. Hal ini didukung oleh kemudahan fitur *sharing* pada situs berita daring dan media sosial yang mengakibatkan seseorang dengan mudah membagikan berita. Suatu berita dapat dengan mudah dibagikan dan dipopulerkan. Namun berita yang berisi informasi yang berbobot dan benar-benar ingin dibaca menjadi suatu masalah tersendiri bagi seseorang yang ingin membaca berita. Pada masalah seperti ini muncul gagasan bagaimana mengklasifikasikan topik berita yang sesuai dengan konten beritanya.

Agar konten berita dapat dengan mudah dipahami dan dicari, terdapat suatu metode yang cukup efektif untuk menangani masalah ini yaitu dengan cara pelabelan [4]. Pelabelan dilakukan dengan memberikan tanda atau label terhadap suatu teks. Pelabelan dipilih karena label merupakan bagian dari struktur hirarkis dari teks (seperti taksonomi)[5], sehingga sangat mewakili konten dari suatu teks. Penelitian terkait pelabelan telah banyak dilakukan diantaranya metode pelabelan tunggal (single-label) dan multi-label. Metode multi-label memiliki keunggulan yaitu data yang diproses dapat dikaitkan dengan lebih dari satu label [6]–[8]. Jika berita yang akan dianalisis banyak, maka multi-label dapat dijadikan solusi untuk melabelkan berita kedalam beberapa kelas sekaligus.

Setelah berita telah terlabeli, untuk mengetahui suatu konten berita masuk kedalam label yang mana, perlu dilakukan proses klasifikasi. Penelitian terkait klasifikasi teks dengan multi-label telah banyak dilakukan, diantaranya dari Chandran dan Panicker menggunakan algoritma ML-ensembles untuk memprediksi label yang dikembangkan dengan membangun sebuah pohon (*tree*) [9]. Penelitian ini memiliki keterbatasan dalam jumlah data yang ada, karena kompleksitas secara linier berhubungan dengan jumlah pohon yang dibangun. Penelitian dari Rastin, dkk mengusulkan metode *Supervised Clustering* dengan mempertimbangkan keterkaitan hubungan (*corelations*) dari label [10][22]. Keterbatasan pada penelitian ini yaitu melakukan pengukuran dimensi dari fitur dan *label space* sebelum dilakukan supervised clustering. Pada penelitian yang dilakukan oleh Baker, dkk

digunakan *artificial neural network* (ANN) untuk menginisialisasi hirarki klasifikasi teks multi-label sehingga hubungan antar label dapat dilihat keterkaitannya [5]. Berdasarkan analisis pada berbagai penelitian sebelumnya, penelitian ini akan menggunakan metode ANN untuk klasifikasi multi-label pada topik berita berbahasa Indonesia. Metode ini dipilih karena dapat melakukan generalisasi dan ekstraksi dari suatu pola data tertentu, dapat menciptakan pola pengetahuan berdasarkan kemampuan belajar jaringan, dan penerapannya memberikan hasil klasifikasi yang lebih baik dibanding dengan menggunakan single-label [11]. Diharapkan dengan adanya penelitian ini, penulis berita dapat dengan mudah memberikan label terhadap berita yang ditulisnya.

Berkembangnya wawasan dan pengetahuan terjadi akibat kemudahan akses informasi, berita, yang dapat dinikmati dimana saja dan kapan saja. Kemudahan ini salah satu penyebabnya yaitu berkembangnya media digital, seperti *jawapos.com*, *detik.com*, *kompas.com*, *line today* dan lain sebagainya. Namun, banyaknya berita yang ada saat ini telah menyalahi kaidah bahasa Indonesia untuk lebih menarik perhatian pembaca [1]. Pada penelitian ini akan dilakukan klasifikasi topik berita berdasarkan banyak label (multi-label) dengan menggunakan ANN. Hal ini dilakukan untuk mengetahui agar konten berita dapat dengan mudah dipahami dan dicari sehingga memudahkan pembaca dalam membaca berita [21]. ANN dipilih sebab metode ini dapat melakukan generalisasi dan ekstraksi dari suatu pola data tertentu, dapat menciptakan pola pengetahuan berdasarkan kemampuan belajar jaringan.

Permasalahan pada klasifikasi teks telah banyak diteliti untuk meningkatkan akurasi untuk mencari fitur terbaik dan sekumpulan data baru [6]. Klasifikasi teks awalnya berkembang untuk mengetahui pengarang dari suatu novel. Seiring berjalannya waktu metode ini berkembang untuk menyelesaikan berbagai masalah, seperti spam filtering, kategorisasi artikel berita, analisis sentimen, dsb [5], [7]. Permasalahan yang sering ditemui yaitu penulis berita sering kali kurang tepat (terdapat *noisy information*) dalam mengklasifikasikan label (kategorisasi) dalam berita [8]. Pada penelitian ini, akan dilakukan proses pengklasifikasian multi-label dengan menggunakan ANN untuk mengetahui kelas sesungguhnya dari suatu berita. Dalam prosesnya, isi (*content*) dari berita akan diubah menjadi data vektor untuk dijadikan masukan pada ANN classifier. Selanjutnya, model ANN akan dikembangkan untuk proses klasifikasi multi-label. Setelah proses-proses tersebut dilakukan, peneliti akan menentukan pengaturan parameter-parameter yang paling optimal untuk mengurangi *loss*.

Penelitian ini dibatasi pengerjaannya antara lain kelas pelabelan dibatasi ke dalam 13 kelas yaitu politik, hukum, ekonomi, sosial, budaya, teknologi, gaya hidup, olahraga, entertainment, pendidikan, HanKam, kesehatan, lainnya. Klasifikasi multi-label yang dilakukan dalam tugas akhir ini mencakup maksimal tiga label dari 13 kelas. Sumber berita yang digunakan dalam pengerjaan tugas akhir ini diperoleh dari media digital JawaPos dan Pikiran Rakyat.

Tujuan yang ingin dicapai pada penelitian ini diantaranya yaitu untuk mengetahui cara mengklasifikasi multi-label dengan menggunakan ANN. Sebelumnya, data yang berupa teks akan diubah menjadi vektor sehingga dapat dijadikan masukan pada ANN classifier. Selanjutnya, model ANN untuk klasifikasi multi-label akan dilakukan dan dilakukan pengaturan parameter-parameter yang paling optimal untuk mengurangi *loss*.

Lanjutan pembahasan pada penelitian ini disajikan dalam sub bagian berikut ini. Bagian 1 membahas teori dan metode yang digunakan dalam penelitian ini. Pada bagian 2 membahas rancangan sistem dan sistem yang akan dibangun. Selanjutnya, bagian 3 membahas hasil pengujian dan analisis hasil pengujian. Terakhir, bagian 4 berisi kesimpulan dan saran terhadap penelitian yang telah dilakukan.

## 2. Tinjauan Pustaka

### 2.1 Multi-label Text Classification

*Multi-label text classification* merupakan bagian dari *supervised classification* dimana masing-masing data dapat dikaitkan dengan lebih dari satu label [19][20]. Klasifikasi *multi-label* bertujuan untuk melakukan proses *learning* dari sekumpulan contoh dimana setiap data berada pada satu kelas atau lebih [9]. Berbeda dengan *single-label classification*, permasalahan pada *multi-label* dipengaruhi oleh hubungan antar label, dalam arti bahwa keanggotaan sebuah kelas dapat membantu memprediksi labelnya.

2.2 Preprocessing

Preprocessing adalah suatu proses perubahan bentuk data yang belum terstruktur menjadi data yang sesuai dengan kebutuhan. Proses ini digunakan pada pengolahan teks atau dokumen seperti *sentiment analysis*, *summarization*, *clustering*, dan lain-lain [23][24]. Beberapa proses yang digunakan adalah : *Case folding* merupakan proses yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Sedangkan, karakter selain huruf akan dihilangkan, misal seperti ( , “ , ! , % , ? , dll). *Tokenization* merupakan proses untuk memisahkan teks dapat berupa kalimat, paragraph atau dokumen, menjadi token-token. Token-token dipisahkan dengan tanda spasi antara kata. *Stopword Removal* merupakan proses untuk menghilangkan kata yang sering muncul, misalnya kata penghubung seperti “dan”, “atau”, “tapi” dan lainnya. *Lemmatization* adalah proses untuk menemukan bentuk dasar dari sebuah kata (Ingason dkk, 2008). Proses ini bertujuan untuk normalisasi bentuk kata berdasarkan kata dasar. Proses normalisasi pada *lemmatization* mengidentifikasi dan menghapus prefiks serta suffiks dari sebuah kata. Pada tugas akhir ini digunakan *lemmatization* untuk berita bahasa Indonesia dari jsastrawi [10].

2.3 Feature Extraction

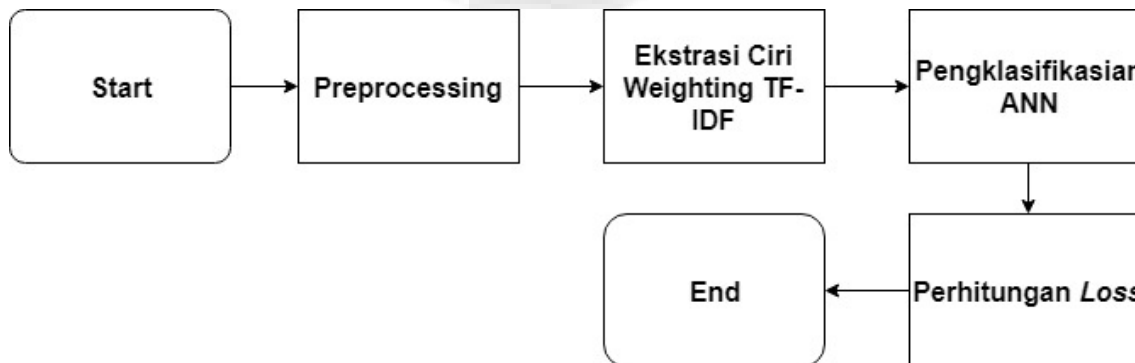
Feature Extraction merupakan tahapan untuk mengubah teks atau dokumen yang sebelumnya masih berbentuk kata kedalam bentuk vektor [11][15][16]. Tahapan ini bertujuan agar teks atau dokumen dapat diklasifikasikan ke dalam kelas-kelas yang sudah ada. Pada tugas akhir ini, metode *feature extraction* yang digunakan adalah *weighting TF-IDF*

2.4 Artificial Neural Network

Artificial Neural Network (ANN) merupakan metode komputasi yang meniru jaringan syaraf biologis manusia. Jaringan syaraf tiruan digunakan untuk memecahkan masalah-masalah seperti pada pengenalan pola atau proses klasifikasi karena memerlukan proses pembelajaran. ANN dapat digunakan untuk memodelkan hubungan yang kompleks antara *input* dan *output* [12]. Umumnya lapisan ANN dibagi menjadi tiga bagian (Riedmiller, 1994), *Input layer* merupakan lapisan masukan yang terdiri dari *neuron* yang berfungsi untuk menerima data masukan. *Hidden layer* merupakan lapisan yang terdiri dari *neuron* yang menerima data dari *input layer*. *Output layer* merupakan lapisan yang terdiri *neuron* yang menerima data dari *hidden layer* atau langsung dari *input layer*.

3. Sistem yang Dibangun

Rancangan sistem klasifikasi yang dibangun menjadi 3 tahap sesuai yang ditunjukkan oleh Gambar 1.



Gambar 1: Rancangan Sistem Klasifikasi menggunakan ANN

3.1 Preprocessing

Data yang didapat dari media digital dimasukan dalam file excel yang dapat dilihat pada lampiran 1. Data memiliki beberapa label dengan perbandingan jumlah antar label dapat dilihat pada lampiran 2. Selanjutnya data teks berita dilakukan *preprocessing*. *Preprocessing* merupakan tahapan awal yang dilakukan untuk mengolah data yang belum sesuai dengan yang diharapkan. Tahapan tersebut seperti : *case folding*, *tokenization*, *stopword removal*, dan *lemmatization*. Tujuan *preprocessing* adalah untuk mendapatkan data yang sesuai sehingga mudah dalam pemrosesan data.

3.2 Ekstraksi Ciri

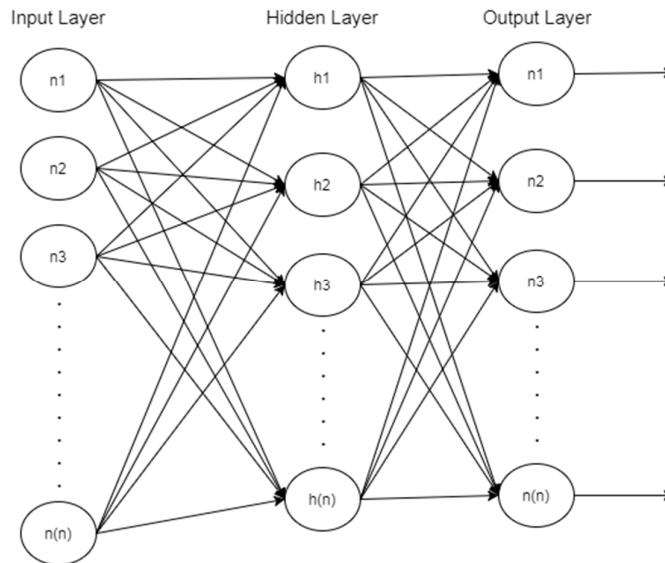
Sebelum dilakukan pengklasifikasian, data berita yang berupa teks diubah menjadi vektor dengan menggunakan ekstraksi ciri. Pada tahap ini bobot dari setiap kata dihitung berdasarkan frekuensi kemunculan *term* atau kata dalam satu dokumen (t) dan frekuensi kemunculan *term* atau kata dalam keseluruhan dokumen (IDF). Rumus *weighting* TF-IDF dapat dilihat pada persamaan 1.

$$W = \log(t + 1) * \log\left(\frac{N}{df}\right) \tag{1}$$

Dengan keterangan W adalah hasil ekstraksi ciri TF-IDF, t adalah jumlah kemunculan *term* pada suatu dokumen, N adalah jumlah keseluruhan dokumen, df adalah jumlah dokumen-dokumen yang memiliki *term* yang bersangkutan.

3.3 Klasifikasi ANN

Hasil ekstraksi ciri adalah masukan data untuk diklasifikasikan dengan ANN. Model ANN dapat dilihat pada gambar 2.



Gambar 2: Model ANN

Pada *input layer* terdapat neuron sesuai dengan dimensi hasil ekstraksi ciri. Neuron-neuron pada *input layer* (lapisan masukan) terhubung dengan neuron-neuron pada *hidden layer* (lapisan tersembunyi). Jumlah neuron pada *output layer* (lapisan luaran) sesuai dengan kelas yang akan diklasifikasikan berjumlah 13 kelas. Fungsi aktivasi yang digunakan pada klasifikasi ANN adalah Sigmoid dengan alpha -0.1.

Algoritma pelatihan ANN (Kusumadewi, 2004) :

0. Inisialisasi bobot (diambil bobot awal dengan nilai random), maksimum Epoch, target *error*, dan *learning rate*.
1. Kerjakan langkah-langkah berikut selama (Epoch < maksimum Epoch) dan (MSE < target *error*) :
2. Untuk tiap-tiap pasangan elemen yang akan dilakukan pembelajaran, kerjakan langkah berikutnya.
3. Untuk langkah 3 hingga 5 merupakan proses Umpan Maju (*Feed Forward*). Tiap-tiap unit input ( $X_i, i = 1, 2, \dots, n$ ) menerima sinyal  $X_i$  dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada di atasnya (lapisan tersembunyi).
4. Tiap-tiap unit pada suatu lapisan sembunyi ( $Z_j, j = 1, 2, 3, \dots, p$ ) menjumlahkan sinyal-sinyal input terbobot :

$$Z_{in_j} = b1_j + \sum_{i=1}^n X_i V_{ij} \tag{2}$$

gunakan fungsi aktivasi untuk menghitung sinyal output :

$$Z_j = f(Z_{in_j}) \quad (3)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan atasnya (unit - unit output).

5. Tiap-tiap unit output  $Y_k$  ( $k = 1, 2, \dots, m$ ) menjumlahkan sinyal-sinyal input terbobot :

$$Y_{in_k} = b2_k + \sum_{j=1}^p Z_j W_{jk}$$

gunakan fungsi aktivasi untuk menghitung sinyal output :

$$Y_k = f(Y_{in_k}) \quad (4)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan atasnya (unit - unit output).

6. Untuk langkah 6 hingga 9 merupakan proses Umpan Mundur (*Backpropagation*). Tiap-tiap unit output  $Y_k$  ( $k = 1, 2, \dots, m$ ) menerima pola target berhubungan dengan pola input pelatihan, hitung informasi errornya :

$$\delta = (t_k - Y_k) f'(Y_{in_k}) \quad (5)$$

$$\varphi2_{jk} = \delta_k Z_j \quad (6)$$

$$\beta2_k = \delta_k \quad (7)$$

kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai) :

$$\Delta W_{jk} = \alpha \varphi2_{jk} \quad (8)$$

Hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai) :

$$\Delta b2_k = \alpha \beta2_k \quad (9)$$

7. Tiap-tiap unit tersembunyi  $Z_j$  ( $j = 1, 2, \dots, p$ ) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan di atasnya) :

$$\delta_{in_j} = \sum_{k=1}^m \delta2_k W_{jk} \quad (10)$$

Kalikan nilai ini dengan turunan dari fungsi aktivasinya untuk menghitung informasi *error* :

$$\delta1_j = \delta_{in_j} f'(Z_{in_j}) \quad (11)$$

$$\varphi1_{ij} = \delta1_j X_j \quad (12)$$

$$\beta1_j = \delta1_j \quad (13)$$

Kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai) :

$$\Delta V_{ij} = \alpha \varphi1_{ij} \quad (14)$$

Hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai) :

$$\Delta b1_j = \alpha \beta1_j \quad (15)$$

8. Tiap-tiap unit output  $Y_k$  ( $k = 1, 2, \dots, m$ ) memperbaiki bias dan bobotnya ( $j = 0, 1, 2, \dots, p$ ) :

$$V_{ij}(\text{baru}) = V_{ij}(\text{lama}) + \Delta V_{ij} \quad (16)$$

$$b2_k(\text{baru}) = b2_k(\text{lama}) + \Delta b2_k \quad (17)$$

Tiap-tiap unit tersembunyi  $Z_j$  ( $j = 1, 2, \dots, p$ ) memperbaiki bias dan bobotnya ( $i = 1, 2, \dots, n$ ) :

$$V_{ij}(\text{baru}) = V_{ij}(\text{lama}) + \Delta V_{ij}$$

(18)

$$b1_j(\text{baru}) = b1_j(\text{lama}) + \Delta b1_j$$

(19)

9. Hitung MSE

10. Hitung *hamming loss* [14][17][18] :

$$HL = \frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^L [y_{ij} \neq t_{ij}]$$

(20)

Dengan keterangan HL merupakan *hamming loss*, N merupakan indek baris, L merupakan indek kolom, y merupakan label luaran, t merupakan label target. Penelitian ini menggunakan perbandingan data latih dan uji sebesar 80:20 persen. Data latih dan uji yang digunakan bersifat acak.

**4. Evaluasi**

4.1 Hasil Pengujian

Sebelum dilakukan pengujian metode ANN untuk klasifikasi teks berita, dilakukan pengujian terhadap ekstrasi ciri menggunakan *weighting* TF-IDF. Perhitungan *weighting* TF-IDF dapat dilihat pada tabel 2 berdasarkan contoh perhitungan kemunculan kata pada table 1.

Kata	Jumlah Kemunculan					
	Politik_1	Politik_2	Ekonomi_1	Ekonomi_2	Tekno_1	Tekno_2
program	0	1	2	0	2	3
Produk	0	0	1	3	5	2
Uang	4	1	5	3	2	0
Pasar	0	0	4	2	3	1
potensi	1	0	2	1	0	2

Tabel 1. Contoh Jumlah Kemunculan Kata

Kata	<i>Weighting</i> TF-IDF					
	Politik_1	Politik_2	Ekonomi_1	Ekonomi_2	Tekno_1	Tekno_2
program	0	0.0530	0.0840	0	0.0840	0.1060
Produk	0	0	0.0530	0.1060	0.1370	0.0840
Uang	0.0553	0.0238	0.0616	0.0477	0.0378	0
Pasar	0	0	0.1231	0.0840	0.1060	0.0530
potensi	0.0530	0	0.0840	0.0530	0	0.0840

Tabel 2. Hasil Ekstrasi Ciri sesuai dengan Tabel 1

Setelah proses mengubah teks menjadi bentuk vektor dengan mencari nilai *weighting* TF-IDF, proses selanjutnya adalah menggunakan metode *Artificial Neural Network* (ANN). Pada penelitian ini menggunakan 16 skema pengujian yaitu nilai maksimum epoch sebanyak 200 dan 400, *learning rate* sebesar 0.008 dan 0.02, nilai *alpha* pada fungsi aktifasi *sigmoid* sebesar 0.1 dan 0.3, jumlah neuron pada lapisan tersembunyi sebanyak 10 dan 20. Berdasarkan pembagian data latih dengan data uji sebesar 80:20 persen untuk data latih dan data uji, didapatkan

nilai *loss* sebesar 0.0021 dengan maksimum epoch 400, *learning rate* 0.02, jumlah neuron 20, dan *alpha* 0.3. Tabel perbandingan hasil aktual dengan hasil prediksi dapat dilihat dalam lampiran 3.

#### 4.2 Analisis Hasil Pengujian

Hasil pengujian dari delapan skema pengujian dapat dilihat pada tabel 3.

Skema	Parameter				Hamming loss
	Maksimum Epoch	Learning rate	Jumlah neuron	Alpha	
1	200	0.008	10	0.1	0.1175
2	200	0.008	10	0.3	0.1047
3	200	0.008	20	0.1	0.1132
4	200	0.008	20	0.3	0.0598
5	200	0.02	10	0.1	0.0897
6	200	0.02	10	0.3	0.0534
7	200	0.02	20	0.1	0.0385
8	200	0.02	20	0.3	0.0128
9	400	0.008	10	0.1	0.1132
10	400	0.008	10	0.3	0.0577
11	400	0.008	20	0.1	0.0513
12	400	0.008	20	0.3	0.0256
13	400	0.02	10	0.1	0.0406
14	400	0.02	10	0.3	0.0214
15	400	0.02	20	0.1	0.0064
16	400	0.02	20	0.3	0.0021

Tabel 3. Hasil Pengujian Metode ANN

Berdasarkan tabel 3 dan grafik pada lampiran 4, dapat ditunjukkan bahwa:

1. Semakin banyak jumlah maksimum epoch, semakin kecil nilai *loss*. Namun jika maksimum epoch terlalu besar, maka algoritma ANN tidak mampu mencapai minimum global. Hal ini terlihat pada skema 1 dan 9, 2 dan 10, 3 dan 11, 4 dan 12, 5 dan 13, 6 dan 14, 7 dan 15, 8 dan 16.
2. Nilai *learning rate* yang dapat menunjukkan penurunan nilai *loss* yang cukup signifikan sebesar 0.02. Jika nilai *learning rate* terlalu besar, maka algoritma ANN tidak mampu mencapai minimum global. Hal ini terlihat pada skema 1 dan 5, 2 dan 6, 3 dan 7, 4 dan 8.
3. Semakin banyak jumlah neuron, semakin kecil nilai *loss*. Hal ini terlihat pada skema 1 dan 2, 3 dan 4, 5 dan 6, 7 dan 8.
4. Nilai *alpha* untuk fungsi aktivasi ANN yang paling optimal sebesar 0.3. Nilai *alpha* yang terlalu besar, dapat menyebabkan *overfitting*. Hal ini terlihat pada skema 1 dan 2, 3 dan 4, 5 dan 6, 7 dan 8.

#### 5. Kesimpulan dan Saran

Berdasarkan hasil pengujian dan analisis yang telah dilakukan dapat disimpulkan bahwa nilai *hamming loss* yang paling optimal sebesar 0.0021 dengan rentang dari 0 sampai 1. Parameter yang menentukan performansi dari metode ANN adalah maksimum epoch, *learning rate*, jumlah neuron, *alpha*. Parameter Maksimum epoch, *learning rate*, jumlah neuron, dan *alpha* yang paling optimal adalah 400, 0.02, 20, 0.3.

Dalam proses pengembangan sistem selanjutnya dalam klasifikasi teks berita berbahasa Indonesia agar dapat ditambahkan lagi jumlah dataset untuk sampel, karena dengan dataset yang lebih banyak diharapkan dapat menambah keragaman sampel sehingga dapat meningkatkan performansi dari sistem.

## Daftar Pustaka

- [1] S. P. Sari, "Masuki Era Digital, Nih Persiapan yang Perlu Dilakukan Pemuda Indonesia," *Okezone News*, 2017. [Online]. Available: <https://news.okezone.com/read/2017/08/11/65/1754124/masuki-era-digital-nih-persiapan-yang-perlu-dilakukan-pemuda-indonesia>. [Accessed: 02-Feb-2018].
- [2] A. P. Prasetyono, "Revolusi Industri Ke-4 dan Integrasinya dalam Tata Kelola Negara," *Ristek Dikti*, 2017. [Online]. Available: <https://ristekdikti.go.id/revolusi-industri-ke-4-dan-integrasinya-dalam-tata-kelola-negara/>. [Accessed: 02-Feb-2018].
- [3] S. Tanaka and A. Jatowt, "Supporting News Article Understanding by Detecting Subject-Background Event Relations," *2016 IEEE/WIC/ACM Int. Conf. Web Intell.*, pp. 256–263, 2016.
- [4] T. Saito and O. Uchida, "Automatic Labeling for News Article Classification Based on Paragraph Vector," *2017 9th Int. Conf. Inf. Technol. Electr. Eng.*, 2017.
- [5] S. Kanj, F. Abdallah, and T. Denœux, "Editing training data for multi-label classification with the k-nearest neighbors rule," *Pattern Anal. Appl.*, vol. 19, no. January 2015, 2016.
- [6] N. Zechner, "The Past, Present and Future of Text Classification," in *2013 European Intelligence and Security Informatics Conference*, 2013, p. 230.
- [7] B. S. Kumar and V. Ravi, "A Survey of the Applications of Text Mining in Financial Domain," *Know.-Based Syst.*, vol. 114, no. C, pp. 128–147, Dec. 2016.
- [8] V. Korde, "Text Classification and Classifiers:A Survey," *Int. J. Artif. Intell. Appl.*, vol. 3, pp. 85–99, 2012.
- [9] M. S. Sorower, "A Literature Survey on Algorithms for Multi-label Learning," *Semantic Scholar*, 2010.
- [10] Librian, Andy, "Natural Language Processing (NLP) Tools for Bahasa Indonesia," *Github*, 2016. [Online]. Available: <https://github.com/jsastrawi/jsastrawi>. [Accessed: 30-April-2018].
- [11] C. S. Utami, M. A. Mukid, Sugito, *Klasifikasi Kinerja Perusahaan di Indonesia dengan Menggunakan Metode Weighted K Nearest Neighbor (Studi Kasus: 436 Perusahaan Yang Terdaftar Di Bursa Efek Indonesia Tahun 2015)*, 2017
- [12] H. Kurt, S. Maxwell dan W. Halbert, "Multilayer Feedforward Network are Universal Approximators," *Neural Networks*, vol. 2, no. 5, pp. 359-366, 1989.
- [13] Kusumadewi, Sri, "Membangun Jaringan Syaraf Tiruan Menggunakan MATLAB & EXCEL LINK", Graha Ilmu, Yogyakarta, 2004
- [14] Read, Jesse, "Multi-label classification", Department of Information and Computer Science, Finland, 2015
- [15] Adiwijaya. 2014. *Aplikasi Matriks dan Ruang Vektor*. Yogyakarta, Graha Ilmu 1, 210
- [16] Adiwijaya. 2016. *Matematika Diskrit dan Aplikasinya*, Bandung, Alfabeta
- [17] Reynaldi Ananda Pane, Mohamad Syahrul Mubarak, Nanang Saiful Huda, Adiwijaya, 2018. A Multi-label Classification on Topics of Quranic Verses in English Translation using multinomial Naïve Bayes. In 2018 6<sup>th</sup> International Conference on Information and Communication Technology (ICoICT). IEEE
- [18] Al Mira Khonsa Izzaty, Mohamad Syahrul Mubarak. 2018. A Multi-label Classification on Topics of Quranic Verses in English Translation Using Tree Augmented Naïve Bayes. In 2018 6<sup>th</sup> International Conference on Information and Communication Technology (ICoICT). IEEE
- [19] Asriyanti Indah Pratiwi, Adiwijaya. 2018. On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. *Applied Computational Intelligence and Soft Computing*, 2018.
- [20] Mubarak. M.S., Adiwijaya and Aldhi, M.D., 2017. Aspect-based sentiment analysis to review products using Naive Bayes. In AIP Conference Proceedings (Vol. 1867, No. 1, p. 020060). AIP Publishing.
- [21] Alamsyah A., Syawiluna M., 2018. Mapping Organization Knowledge Network and Social Media Based Reputation Management, *Journal of Data Science and Its Applications (JDSA)*, 1(1), pp.39-48.
- [22] Firmansyah A. F. B., Pramana, S., 2018. Ensemble Based Gustafson Kessel Fuzzy Clustering, *Journal of Data Science and Its Applications (JDSA)*, 1(1), pp.1-9.
- [23] Adiwijaya, M. N. Aulia, M. S. Mubarak, W. U. Novia, and F. Nhita. A comparative study of mfcc-knn and lpc-knn for hijaiyyah letters pronunciation classification system. In *Information and Communication Technology (ICoICT)*, 2017 5th International Conference on, pages 1–5. IEEE, 2017.
- [24] H. Ayadenta and Adiwijaya. A clustering approach for feature selection in microarray data classification using random forest. *Journal of Information Processing System*, 14(5), 2018.

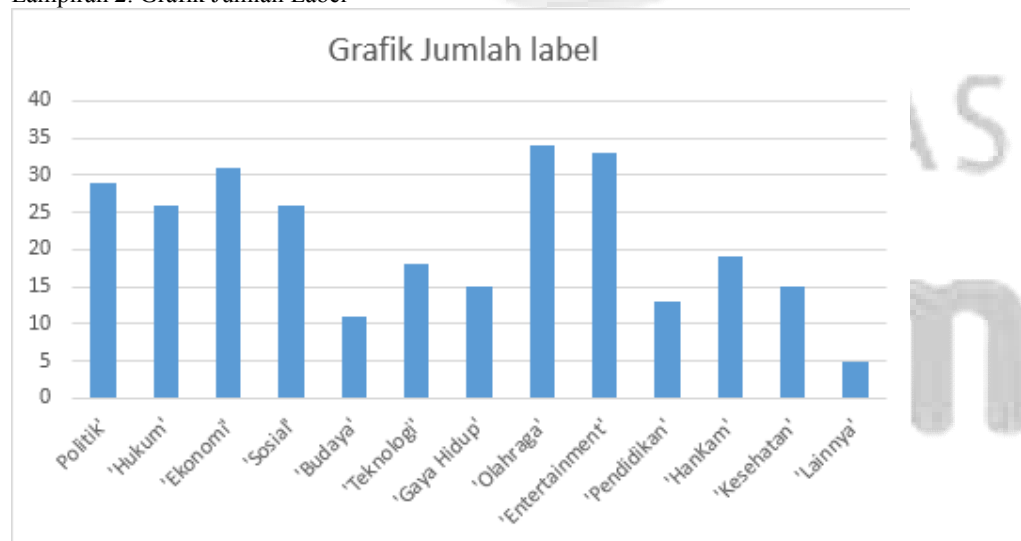


Lampiran

Lampiran 1. Screenshot Dataset

A	B	C	D	E	F
20	Tiang listrik mendadak jadi topik perbin	6,9	<a href="https://www.jawapos.com/read/2017/11/19/169423">https://www.jawapos.com/read/2017/11/19/169423</a>		
21	Memasuki usia ke-35 tahun, Epson Si	3,6	<a href="https://www.jawapos.com/read/2017/11/07/166989">https://www.jawapos.com/read/2017/11/07/166989</a>		
22	Dinas Pariwisata Kabupaten Gianyar n	3	<a href="https://www.jawapos.com/radarbali/read/2017/11/3">https://www.jawapos.com/radarbali/read/2017/11/3</a>		
23	Kunjungan ke Daerah Tujuan Wisata (f	3	<a href="https://www.jawapos.com/baliexpress/read/2017/11">https://www.jawapos.com/baliexpress/read/2017/11</a>		
24	Buruknya perekonomian keluarga men	4	<a href="https://www.jawapos.com/radarmadura/read/2017/">https://www.jawapos.com/radarmadura/read/2017/</a>		
25	Dampak Siklon Tropis Cempaka yang	4	<a href="https://www.jawapos.com/read/2017/11/28/171238">https://www.jawapos.com/read/2017/11/28/171238</a>		
26	Badan Keamanan Laut (Bakamla) aka	11,6	<a href="https://www.jawapos.com/read/2017/11/21/169710">https://www.jawapos.com/read/2017/11/21/169710</a>		
27	Pasangan ganda putri Indonesia, Grey	8	<a href="https://www.jawapos.com/read/2017/11/28/171099">https://www.jawapos.com/read/2017/11/28/171099</a>		
28	Pemenuhan hak pencipta lagu dan kar	9	<a href="https://www.jawapos.com/read/2018/01/10/180525">https://www.jawapos.com/read/2018/01/10/180525</a>		
29	Wakil Presiden Republik Indonesia, Ju	1,3,8	<a href="https://www.jawapos.com/read/2018/01/08/179937">https://www.jawapos.com/read/2018/01/08/179937</a>		
30	Akhirnya kabut misteri saldo nasabah	2,6	<a href="https://www.jawapos.com/read/2018/03/16/196442">https://www.jawapos.com/read/2018/03/16/196442</a>		
31	Aktor kawakan, Reza Rahadian memp	4,10	<a href="https://www.jawapos.com/read/2018/01/10/180508">https://www.jawapos.com/read/2018/01/10/180508</a>		
32	Dilansir dari Yonhap News, Pada bular	8,9	<a href="https://www.jawapos.com/read/2018/01/10/180444">https://www.jawapos.com/read/2018/01/10/180444</a>		
33	Utusan Khusus PBB untuk HAM di My	2,4,6	<a href="https://www.jawapos.com/read/2018/03/16/196371">https://www.jawapos.com/read/2018/03/16/196371</a>		
34	Pemerintah Indonesia melalui Kement	4,11	<a href="https://www.jawapos.com/read/2018/03/16/196358">https://www.jawapos.com/read/2018/03/16/196358</a>		
35	Empat kontainer berisi daging sapi be	2,11,3	<a href="https://www.jawapos.com/read/2018/03/16/196433">https://www.jawapos.com/read/2018/03/16/196433</a>		
36	Sebagai salah satu warisan budaya ba	5,7	<a href="https://www.jawapos.com/read/2018/03/16/196418">https://www.jawapos.com/read/2018/03/16/196418</a>		
37	Kemiskinan di Ibu Kota DKI Jakarta se	1,4	<a href="https://www.jawapos.com/read/2018/03/16/196353">https://www.jawapos.com/read/2018/03/16/196353</a>		
38	Megahnya kehidupan di Jakarta bagai	4	<a href="https://www.jawapos.com/read/2018/03/16/196337">https://www.jawapos.com/read/2018/03/16/196337</a>		
39	Pemerintah Provinsi DKI Jakarta masi	4	<a href="https://www.jawapos.com/read/2018/03/15/196343">https://www.jawapos.com/read/2018/03/15/196343</a>		
40	Dinas Sosial (Dinsos) Kota Surabaya t	4,3	<a href="https://www.jawapos.com/read/2018/03/14/195997">https://www.jawapos.com/read/2018/03/14/195997</a>		
41	Badan Narkotika Nasional Kota Suraba	4,2,9	<a href="https://www.jawapos.com/read/2018/03/10/194986">https://www.jawapos.com/read/2018/03/10/194986</a>		
42	Kota Pahlawan akan segera punya tra	4,3,6	<a href="https://www.jawapos.com/read/2018/03/08/194250">https://www.jawapos.com/read/2018/03/08/194250</a>		
43	Pelarangan cadar di beberapa lembag	10,4	<a href="https://www.jawapos.com/read/2018/03/07/194064">https://www.jawapos.com/read/2018/03/07/194064</a>		
44	Gubernur DKI Jakarta Anies Baswedar	5,8,1	<a href="https://www.jawapos.com/read/2018/03/04/193155">https://www.jawapos.com/read/2018/03/04/193155</a>		
45	Warga Surabaya penggila kesenian Kc	5,9	<a href="https://www.jawapos.com/read/2018/03/02/192888">https://www.jawapos.com/read/2018/03/02/192888</a>		
46	Pemerintah Provinsi DKI Jakarta tanga	10,1	<a href="https://www.jawapos.com/read/2018/03/01/192309">https://www.jawapos.com/read/2018/03/01/192309</a>		
47	Prostitusi sesama jenis kian menjamu	4,2	<a href="https://www.jawapos.com/read/2018/02/28/192282">https://www.jawapos.com/read/2018/02/28/192282</a>		

Lampiran 2. Grafik Jumlah Label



Lampiran 3. Tabel Hasil Aktual dan Hasil Prediksi

No Berita	Isi berita	Target	Output
35	Empat kontainer berisi daging sapi	2,11,3	2,11,3
30	Akhirnya kabut misteri saldo nasabah	2,6	2,6
19	Usulan anggota Komisi X DPR An	6,1	6,1
34	Pemerintah Indonesia melalui Ker	4,11	4,11
18	Sebuah momen menarik terlihat p	8	8
13	Tren busana muslim (modest wea	7	7
9	Dinamika bongkar pasangan calo	1	1
22	Dinas Pariwisata Kabupaten Gian	3	3
29	Wakil Presiden Republik Indonesi	1,3,8	1,3,8
36	Sebagai salah satu warisan buday	5,7	7

Lampiran 4. Grafik Hasil Pengujian

