

KLASIFIKASI KALIMAT PADA BERITA OLAHRAGA SECARA OTOMATIS MENGUNAKAN METODE ARTIFICIAL NEURAL NETWORK

Asep Saepul Ridwan¹, Yulison H. Chrisnanto² dan Ridwan Ilyas³

^{1,2,3} Program Teknik Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani

¹Email: asepsaepulridwan99@gmail.com

²Email: y.chrisnanto@gmail.com

³Email: ilyas@lecture.unjani.ac.id

ABSTRAK

Berita olahraga sangat diminati oleh masyarakat sekarang ini. Ini dikarenakan olahraga telah berkembang menjadi sebuah hiburan. Ada banyak berita olahraga yang ada sekarang ini yang meliputi berbagai macam cabang olahraga, dari mulai cabang olahraga yang menggunakan bola sebagai objek permainannya seperti sepakbola, sampai olahraga di arena balap otomotif seperti Formula 1. Selain itu substansi dari berita olahraga itu sendiri sangat beragam seperti berita tentang manajerial dari sebuah klub olahraga, hasil pertandingan, cedera pemain, dan lain-lain. tentunya hal yang demikian akan menyulitkan pembaca berita yang menginginkan kumpulan satu bidang pembahasan dalam berita tersebut. Tumpang tindih data terjadi dalam dokumen berita olahraga karena mencampuradukan data kalimat satu bahasan dengan data kalimat bahasan yang lainnya. Beberapa bahasan yang ada dalam berita olahraga adalah tentang manajerial, pemain, jadwal, *preview*, *review*, klasemen, statistik, juara, dan lain-lain. Hal ini menjadi masalah ketika pembaca menginginkan bahasan pada berita yang memfokuskan pada satu pembahasan tertentu. Pada penelitian ini telah dibangun sistem klasifikasi kalimat menggunakan *Artificial Neural Network* (ANN) dengan metode pembelajaran *Backpropagation*. Fitur yang digunakan yaitu frekuensi dari kemunculan suatu *term* dalam kalimat yang terkait dan hasil perhitungan dari sebuah *term* yang didistribusikan. Pengujian dari metode yang kami ajukan menunjukkan akurasi mencapai 99% untuk hasil yang terbaik pada data latih dan 57% pada data uji.

Kata kunci: klasifikasi, *artificial neural network*, *Term Frequency-Inverse Document Frequency* (TF-IDF), *backpropagation*

ABSTRACT

Sports news is of great interest to today's society. This is because sports have grown into entertainment. There is a lot of sports news today that covers a wide range of sports, from branches that use the ball as objects for games like football, to sports in automotive race like formula 1. Beyond that, the substance of the sports news itself is as diverse as the news of managerial from a sports club, results matches, *player injuries*, et cetera. Surely such a thing would be difficult. The network wants one in the field of discussion in the news. Overlap data occurs in the sports news document because it mixes one sentence data with the other. Some of the content in the sports news is about managerial, *players*, schedules, previews, reviews, standings, statistics, champions, etc. Becomes a problem when the reader wants a topic on the news that focuses on one particular discussion. This study has built a book on sentence classification meaning *Artificial Neural Network* (ANN) with a method of learning *Backpropagation*. The feature used is the frequency of the occurrence of a term in the corresponding sentence and the calculating result of a distributed term. The testing of our proposed methods shows an accuracy of 99% to best results on training data and 57% on test data.

Keywords: classification, artificial neural networks, term frequency-inverse document frequency (TF-IDF), *backpropagation*

1. PENDAHULUAN

Berita olahraga sangat diminati oleh masyarakat sekarang ini. Ini dikarenakan olahraga telah berkembang menjadi sebuah hiburan. Ada banyak berita tentang olahraga yang ada sekarang ini meliputi berbagai macam cabang olahraga, dari mulai cabang olahraga yang menggunakan bola sebagai obyek permainannya seperti sepakbola, sampai olahraga di arena balap otomotif seperti Formula 1. Selain itu substansi dari berita olahraga itu sendiri sangat beragam seperti berita tentang manajerial dari sebuah klub olahraga, hasil pertandingan, cedera pemain, dan lain-lain. tentunya hal yang demikian akan menyulitkan

pembaca berita yang menginginkan kumpulan satu bidang pembahasan dalam berita tersebut. Dengan pengklasifikasian kalimat berdasarkan bahasan pada berita olahraga secara otomatis, pembaca yang mengkhususkan mencari bahasan tertentu akan lebih mudah membaca berita yang dicari. Dengan demikian manfaat yang diharapkan dari penelitian ini yaitu sistem klasifikasi kalimat berita olahraga yang dapat memilah kalimat penting yang ada pada berita olahraga tersebut ke dalam kelas berdasarkan isi pembahasan yang ada pada setiap kalimat berita olahraga.

Pada tahun 2018, salah satu situs berita olahraga *online* yakni Bolasport.com menduduki posisi ke 30 *top site* di Indonesia menurut situs Alexa.com. Selain itu 2 situs berita olahraga lainnya masuk dalam 50 *top site*. Itu semua belum termasuk berita olahraga yang dimuat oleh situs portal berita *online* yang tidak menghususkan memuat konten berita olahraga seperti Tribunnews.com yang mana pada situs tersebut berita olahraga sangat digemari dan dibutuhkan oleh para pembaca berita. Permasalahan muncul ketika pembaca berita menginginkan topik bahasan tertentu. Menjadi masalah karena dalam satu artikel berita terdapat tumpang tindih data topik berita. Seperti dalam satu berita memuat banyak topik yakni tentang manajerial, pemain, klub, hasil pertandingan, dan lain-lain.

Event atau topik pada berita olahraga sangat banyak, seperti skor pertandingan, cedera pemain, statistik pertandingan, klasemen, manajerial sebuah klub, transfer pemain, dan masih banyak lagi *event* lain yang ada pada berita olahraga. Ini semua memungkinkan untuk dilakukannya pengklasifikasian baik itu pengklasifikasian berita maupun pengklasifikasian kalimat yang ada pada berita olahraga berdasarkan *event* yang ada pada kalimat di berita tersebut. Penelitian terhadap klasifikasi telah banyak dilakukan seperti klasifikasi berita *online*.

Dalam klasifikasi, terdapat banyak pilihan metode yang bisa digunakan salah satunya yaitu metode *Artificial Neural Network* (ANN). Kemudian untuk ekstraksi fiturnya bisa menggunakan *Term Frequency-Inverse Document Frequency* TF-IDF. Penelitian terdahulu telah banyak menggunakan metode ANN dengan ekstraksi fitur TF-IDF, seperti digunakan untuk klasifikasi topik berita berbahasa Indonesia [1]. Selain itu ANN juga bisa digunakan untuk kasus klasifikasi lainnya seperti digunakan untuk klasifikasi kepribadian berdasarkan status Facebook [2] yang mana metode pembelajaran yang digunakan adalah *Backpropagation*. Penelitian lain yang menggunakan *Backpropagation* adalah untuk analisis sentimen *tweet* [3], klasifikasi sentimen *review* produk otomotif [4], perancangan pengenalan karakter alfabet menggunakan metode *hybrid* [5], dan prediksi jumlah pengguna internet terbanyak [6]. Untuk input data selain teks, *Backpropagation* juga bisa digunakan seperti untuk prediksi laju inflasi [7], prediksi pasang surut air laut [8], prediksi awal musim hujan berdasarkan suhu permukaan laut [9], klasifikasi hewan vertebrata [10], prediksi cuaca [11], prediksi cuaca dengan menggunakan *moving average* [12], dan prediksi stok obat di apotek (studi kasus : apotek ABC) [13].

Hasil beberapa penelitian terdahulu menunjukkan akurasi atau presisi dari metode ANN menggunakan metode pembelajaran *Backpropagation* sangat baik. Keakuratan dari ANN menggunakan metode pembelajaran *Backpropagation* bisa lebih dari 80% [2]. Ini membuktikan bahwa metode ANN sangat cocok untuk kasus klasifikasi.

Penelitian terdahulu banyak menggunakan metode pembelajaran *Backpropagation* untuk memperbarui bobot. Arsitektur yang digunakan adalah jaringan dengan banyak lapis (*multilayer*). Untuk pemrosesan berupa teks, salah satu praproses yang bisa dilakukan untuk data tersebut adalah proses segmentasi dan kemudian data bisa ditransformasikan ke bentuk vektor menggunakan metode TF-IDF. *Backpropagation* bisa diaplikasikan pada data berupa teks. Algoritma *Backpropagation* merupakan algoritma pembelajaran untuk memperoleh tingkat *error* yang paling kecil dengan cara menyesuaikan bobotnya berdasarkan perbedaan *output* dan target yang diinginkan. Selain itu *Backpropagation* juga merupakan metode sistematis untuk pelatihan *multilayer* ANN.

Penelitian terdahulu untuk klasifikasi berita belum spesifik membahas tentang klasifikasi yang berdasarkan pada substansi bahasan per-kalimat. Ini tentu mengakibatkan terjadinya tumpang tindih data kalimat dengan substansi bahasan yang berbeda beda pada satu berita. Ini tentu menyulitkan sebagian pembaca berita yang menginginkan pembahasan tentang topik tertentu.

Berbeda dari penelitian yang terdahulu, penelitian ini lebih mengedepankan pada klasifikasi per-kalimat. Klasifikasinya sendiri berdasarkan pada apa yang dibahas pada kalimat tersebut. Apa yang dibahas pada sebuah kalimat bisa diketahui dengan cara dianalisis terlebih dahulu.

Penelitian ini bertujuan untuk membuat sebuah sistem yang bisa mengklasifikasi kalimat pada berita olahraga berdasarkan pembahasan dari kalimat pada berita tersebut, dengan harapan bisa membantu pembaca berita yang menginginkan bacaan tentang satu bahasan dengan ringkas. Penelitian ini diharapkan dapat berkontribusi terhadap ilmu pengetahuan yakni pengembangan metode yang dipakai dapat digunakan untuk berbagai jenis klasifikasi.

Penelitian ini akan membuat sistem untuk klasifikasi kalimat pada berita olahraga secara otomatis menggunakan ANN. Metode yang dipakai adalah metode *Backpropagation* karena sangat baik dalam hal proses klasifikasi mengingat *Backpropagation* memiliki kemampuan dalam mengadaptasikan kondisi jaringan dengan data yang diberikan melalui proses pembelajaran.

Praproses menggunakan TF-IDF untuk merepresentasikan setiap kata di dalam konteks sebagai vektor dengan N dimensi. Data berita olahraga yang berupa paragraf terlebih dahulu dipecah menjadi beberapa kalimat. Pemisahan setiap kalimatnya sendiri berdasarkan tanda baca titik (.). pemisahan setiap kalimat dikenal dengan istilah segmentasi. Barulah setelah proses segmentasi selesai, data kalimat berita diproses dengan menggunakan TF-IDF. TF-IDF digunakan karena dapat memberikan nilai pada kalimat yang sebagai inti pada sebuah berita.

Tujuan yang ingin dicapai dari penelitian ini yakni membangun sistem yang bisa mengklasifikasi kalimat pada berita olahraga berdasarkan isi pembahasan dalam konten kalimat di berita tersebut dengan terlebih dahulu mencari isi dari berita. Sistem akan diimplementasikan dalam bentuk perangkat lunak.

Manfaat yang terdapat pada penelitian ini yaitu sistem klasifikasi kalimat berita olahraga yang dapat memilah kalimat yang ada pada berita olahraga tersebut ke dalam kelas berdasarkan isi pembahasan yang ada pada setiap kalimat berita olahraga.

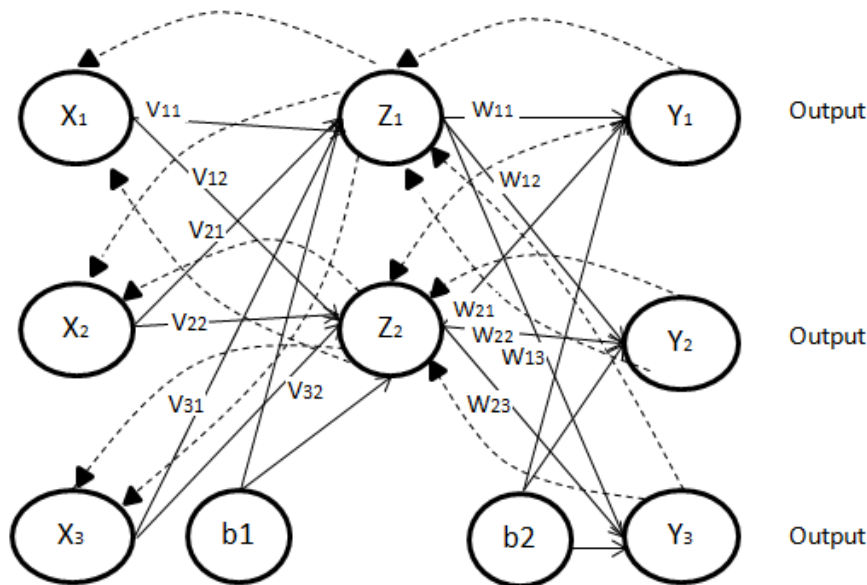
2. MATERI DAN METODE

Artificial Neural Network (ANN)

ANN adalah metode untuk mengelompokkan dan memisahkan data dengan cara kerjanya sama dengan jaringan syaraf pada manusia. ANN dirancang untuk memecahkan masalah tertentu seperti mengenali pola tertentu atau mengklasifikasi suatu data melalui proses pembelajaran. Salah satu kelebihan dari ANN adalah kemampuannya dalam perhitungan secara paralel sehingga proses menjadi lebih singkat.

Backpropagation

Karakteristik utama dari *neural network* yaitu kemampuan dalam mempelajari hubungan input-output nonlinear yang sangat kompleks, dengan menggunakan proses pelatihan sekuensial sehingga dapat beradaptasi dengan data yang ada. Salah satu metode pelatihan ANN yaitu *Backpropagation*. Arsitektur dari *Backpropagation* dapat dilihat seperti pada gambar 1.



Gambar 1. Arsitektur *Backpropagation*

Pada metode pembelajaran *Backpropagation*, terdapat fungsi aktivasi. Ada banyak pilihan fungsi aktivasi yang dapat digunakan dalam metode pembelajaran *Backpropagation*, seperti fungsi sigmoid bipolar dan sigmoid biner. Fungsi aktivasi dari sigmoid biner memiliki tingkat akurasi yang lebih baik ketimbang fungsi aktivasi sigmoid bipolar [14].

TF-IDF

TF-IDF adalah suatu teknik pembobotan kata yang basisnya pada statistik kemunculan kata dan juga tingkat kepentingan dokumen yang mengandungnya. Pembobotan yang didapat dari pendekatan temu kembali informasi ini merupakan hasil dari perkalian *term frequency* dan *inverse document frequency* yang mana tiap nilainya didapatkan dari persamaan 1.

$$TF * IDF = TF (w_i, d) * IDF (w_i).....(1)$$

di mana:

- w_i = kata ke-i
- D = dokumen
- TF(w_i, d) = jumlah kemunculan kata w_i pada dokumen D
- IDF (w_i) = *Invers Document Frequency* dari kata w_i .

Pembobotan dihitung tiap kata. *Term Frequency* (TF) menunjukkan bahwa semakin tinggi kemunculan kata pada sebuah dokumen maka semakin penting kata tersebut mewakili sebuah dokumen. Nilai *Inverse Document Frequency* (IDF) merupakan indikasi bahwa jika kemunculan sebuah kata memiliki frekuensi yang lebih tinggi pada dokumen tertentu saja maka kata tersebut menjadi ciri penting dan mewakili dokumen tersebut. Sedangkan untuk kata yang muncul pada seluruh dokumen, maka kata tersebut bersifat umum dan tidak mewakili dokumen manapun sehingga memiliki nilai IDF 0. Nilai dari IDF didapatkan dengan persamaan 2.

$$IDF (w_i) = \log \frac{|D|}{DF(w_i)}.....(2)$$

di mana:

- w_i = kata ke i
- |D| = jumlah seluruh dokumen
- IDF (w_i) = jumlah dokumen yang mengandung kata w_i .

Pada klasifikasi kalimat berita olahraga, untuk dokumen pada persamaan diwakili oleh kalimat yang digunakan. Pembobotan TF-IDF sangat membantu mengambil kata-kata dari kalimat yang paling mewakili kelasnya.

Analisis dan Perancangan Sistem

Pada klasifikasi menggunakan ANN terdapat dua tahapan yaitu pembelajaran menggunakan data latih dan pengujian menggunakan data uji. Data latih dan data uji diambil dari *dataset*. *Dataset* terdiri dari kumpulan kalimat yang berasal dari 200 berita olahraga *online* yang diambil dari situs Bolasport.com dalam kurun waktu 4 bulan dari tanggal 10 Agustus 2019 sampai dengan 29 Desember 2019. Untuk keperluan keseimbangan *dataset*, maka tidak semua berita di situs Bolasport.com dalam kurun waktu tersebut diambil. Juga tidak setiap kalimat yang ada pada berita yang diambil, digunakan. Proses pelabelan *dataset* dilakukan oleh anotator secara manual. Untuk proses pembelajaran dan pengujian dapat dilihat pada gambar 2.

Data latih berasal dari *dataset* yang sudah melewati tahapan praproses untuk mendapatkan nilai fitur yakni nilai similaritas dan label yang merupakan kelas. Dari *dataset* yang berjumlah 1000 *dataset*, x1 sampai x2912 merupakan nilai fitur sedangkan label merupakan kelas yang terdiri dari kelas manajemen, kelas cedera, kelas jadwal, kelas *preview*, kelas *review*, kelas klasemen, kelas statistik, kelas juara, kelas pemain/atlet, dan kelas lainnya. Kelas lainnya adalah satu kelas untuk kalimat dengan pembahasan selain manajemen, cedera, jadwal, *preview*, *review*, klasemen, statistik, juara, atau pemain/atlet. Data latih digunakan untuk proses pelatihan untuk menentukan bobot pelatihan yang memisahkan antara sepuluh kelas yaitu manajemen, cedera, jadwal, *preview*, *review*, klasemen, statistik, juara, pemain/Atlet, dan lainnya. Pada tahap pengujian data tes yang digunakan berasal dari 20% *dataset* yang diambil secara acak. Tahap pertama dalam pembelajaran ANN adalah dengan melakukan *feed forward* menggunakan aktivasi sigmoid biner seperti persamaan 3.

$$f(x) = \frac{1}{1+e^{-x}}.....(3)$$

Setelah dilakukan *feed forward* selanjutnya melakukan *back forward* untuk memperbaiki *error* sesuai persamaan 4.

$$E = y - o.....(4)$$

di mana:

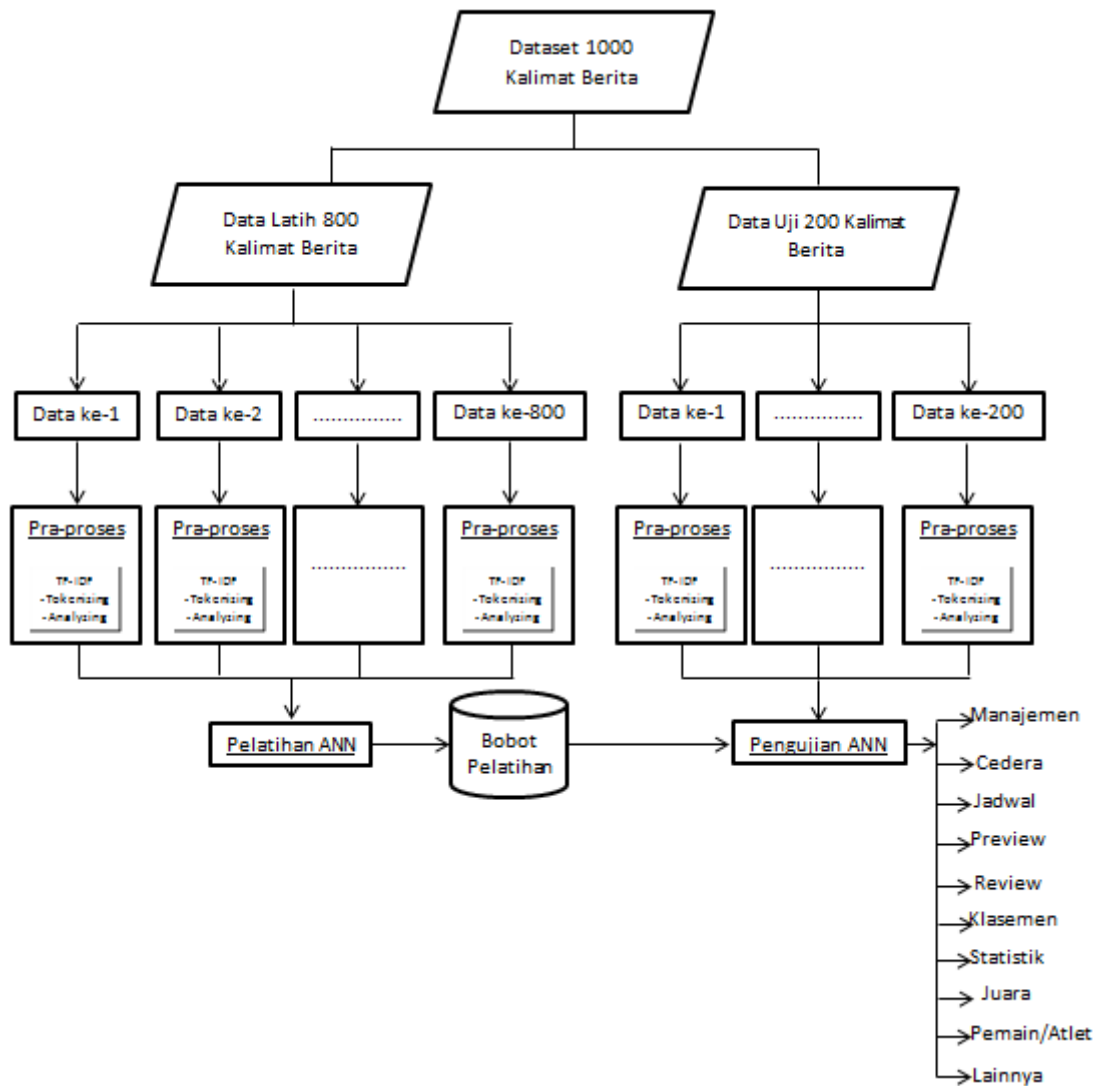
- E = error
- y = target
- o = output

Bahasa pemrograman yang digunakan pada penelitian ini adalah Python versi 3. selain itu untuk mempermudah penelitian digunakan juga *library* Keras. Sedangkan *Integrated Development Environment* (IDE) yang digunakan adalah Jupyter Notebook.

3. HASIL DAN PEMBAHASAN

Hasil

Dataset yang digunakan pada penelitian ini terdiri dari 1000 kalimat berita olahraga. Dari *dataset* 1000 kalimat, diambil 800 kalimat yang digunakan untuk data latih dan 200 kalimat digunakan untuk data uji. Pembelajaran terhadap data latih 800 kalimat berita olahraga dilakukan dengan menggunakan 3 *layer*. 3 *layer* tersebut terdiri dari 1 *layer* untuk input, 1 *layer* untuk *hidden*, dan 1 *layer* untuk *output*. Neuron input berjumlah 2912 buah neuron. Jumlah neuron input tersebut diambil dari jumlah fitur yang ada pada *dataset*. neuron *hidden* pada percobaan pertama berjumlah 2 buah neuron, pada percobaan kedua berjumlah 4 neuron, pada percobaan ketiga berjumlah 6 neuron, dan pada percobaan keempat berjumlah 8 neuron. neuron output berjumlah 10 neuron. Dari *dataset* 1000 kalimat, diambil 800 kalimat digunakan untuk data latih dan 200 kalimat digunakan untuk data uji. Hasil pembelajaran dari data latih dapat dilihat pada tabel 1.



Gambar 2. Perancangan sistem pembelajaran dan pengujian ANN

Tabel 1. Hasil pembelajaran data latih

Neuron <i>Hidden Layer</i>	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
2 Neuron	Manajemen	1.00	0.71	0.83	82
	Cedera	1.00	0.95	0.97	76
	Jadwal	0.98	0.95	0.96	83
	<i>Preview</i>	0.00	0.00	0.00	72
	<i>Review</i>	0.00	0.00	0.00	86
	Klasemen	0.20	1.00	0.34	88
	Statistik	0.97	0.87	0.92	79
	Juara	0.95	0.96	0.96	80
	Pemain/Athlet	0.00	0.00	0.00	77

Neuron Hidden Layer	Kelas	Precision	Recall	F1-Score	Support
4 Neuron	Lainnya	0.00	0.00	0.00	77
	Accuracy			0.55	800
	Macro avg	0.51	0.54	0.50	800
	Weighted avg	0.51	0.55	0.50	800
	Manajemen	0.97	0.90	0.94	82
	Cedera	0.99	1.00	0.99	76
	Jadwal	0.39	1.00	0.56	83
	Preview	1.00	0.82	0.90	72
	Review	1.00	0.91	0.95	86
	Klasemen	0.98	0.97	0.97	88
	Statistik	1.00	0.87	0.93	79
	Juara	0.00	0.00	0.00	80
	Pemain/Atlet	0.91	0.90	0.90	77
	Lainnya	0.98	0.79	0.88	77
6 Neuron	Accuracy			0.82	800
	Macro avg	0.82	0.82	0.80	800
	Weighted avg	0.82	0.82	0.80	800
	Manajemen	0.96	1.00	0.98	82
	Cedera	1.00	1.00	1.00	76
	Jadwal	1.00	1.00	1.00	83
	Preview	1.00	1.00	1.00	72
	Review	1.00	1.00	1.00	86
	Klasemen	1.00	1.00	1.00	88
	Statistik	1.00	1.00	1.00	79
	Juara	1.00	0.97	0.99	80
	Pemain/Atlet	1.00	1.00	1.00	77
	Lainnya	1.00	0.99	0.99	77
	8 Neuron	Accuracy			1.00
Macro avg		1.00	1.00	1.00	800
Weighted avg		1.00	1.00	1.00	800
Manajemen		1.00	0.98	0.99	82
Cedera		1.00	0.96	0.98	76
Jadwal		0.00	0.00	0.00	83
Preview		0.00	0.00	0.00	72
Review		0.32	1.00	0.48	86
Klasemen		1.00	0.98	0.99	88
Statistik		1.00	0.94	0.97	79
Juara		1.00	0.95	0.97	80
Pemain/Atlet		1.00	0.83	0.91	77
Lainnya		0.99	0.95	0.97	77
Accuracy				0.77	800
Macro avg	0.73	0.76	0.73	800	
Weighted avg	0.73	0.77	0.73	800	

Setelah melakukan pembelajaran terhadap 800 kalimat data latih, tahap selanjutnya adalah pengujian. Tahap pengujian menggunakan data uji yang terdiri dari 200 kalimat. Hasil dari pengujian dapat dilihat pada tabel 2.

Tabel 2. Hasil pengujian

Neuron Hidden Layer	Kelas	Precision	Recall	F1-Score	Support
2 Neuron	Manajemen	0.64	0.50	0.56	18
	Cedera	0.92	0.50	0.65	24
	Jadwal	0.61	0.65	0.63	17
	Preview	0.00	0.00	0.00	28
	Review	0.00	0.00	0.00	14
	Klasemen	0.12	0.92	0.21	12
	Statistik	0.32	0.57	0.41	21

Neuron Hidden Layer	Kelas	Precision	Recall	F1-Score	Support	
	Juara	0.89	0.80	0.84	20	
	Pemain/Atlet	0.00	0.00	0.00	23	
	Lainnya	0.00	0.00	0.00	23	
	<i>Accuracy</i>			0.36	200	
	<i>Macro avg</i>	0.35	0.39	0.33	200	
	<i>Weighted avg</i>	0.35	0.35	0.32	200	
	4 Neuron	Manajemen	0.45	0.50	0.47	18
		Cedera	0.76	0.67	0.71	24
		Jadwal	0.40	0.82	0.54	17
		<i>Preview</i>	0.39	0.39	0.39	28
<i>Review</i>		0.60	0.64	0.62	14	
Klasemen		0.60	0.75	0.67	12	
Statistik		0.47	0.33	0.39	21	
Juara		0.00	0.00	0.00	20	
Pemain/Atlet		0.48	0.43	0.45	23	
Lainnya		0.43	0.43	0.43	23	
<i>Accuracy</i>			0.48	200		
<i>Macro avg</i>	0.46	0.50	0.47	200		
<i>Weighted avg</i>	0.45	0.47	0.46	200		
6 Neuron	Manajemen	0.76	0.89	0.82	18	
	Cedera	1.00	0.67	0.80	24	
	Jadwal	1.00	0.59	0.74	17	
	<i>Preview</i>	0.80	0.29	0.42	28	
	<i>Review</i>	0.58	0.50	0.54	14	
	Klasemen	0.86	0.50	0.63	12	
	Statistik	0.28	0.71	0.40	21	
	Juara	0.94	0.75	0.83	20	
	Pemain/Atlet	0.53	0.43	0.48	23	
	Lainnya	0.34	0.52	0.41	23	
<i>Accuracy</i>			0.57	200		
<i>Macro avg</i>	0.71	0.59	0.61	200		
<i>Weighted avg</i>	0.70	0.57	0.60	200		
8 Neuron	Manajemen	0.59	0.56	0.57	18	
	Cedera	0.93	0.54	0.68	24	
	Jadwal	0.00	0.00	0.00	17	
	<i>Preview</i>	0.00	0.00	0.00	28	
	<i>Review</i>	0.16	0.57	0.25	14	
	Klasemen	0.86	1.00	0.92	12	
	Statistik	0.44	0.76	0.56	21	
	Juara	0.68	0.65	0.67	20	
	Pemain/Atlet	0.61	0.48	0.54	23	
	Lainnya	0.41	0.48	0.44	23	
<i>Accuracy</i>			0.47	200		
<i>Macro avg</i>	0.47	0.50	0.46	200		
<i>Weighted avg</i>	0.46	0.47	0.44	200		

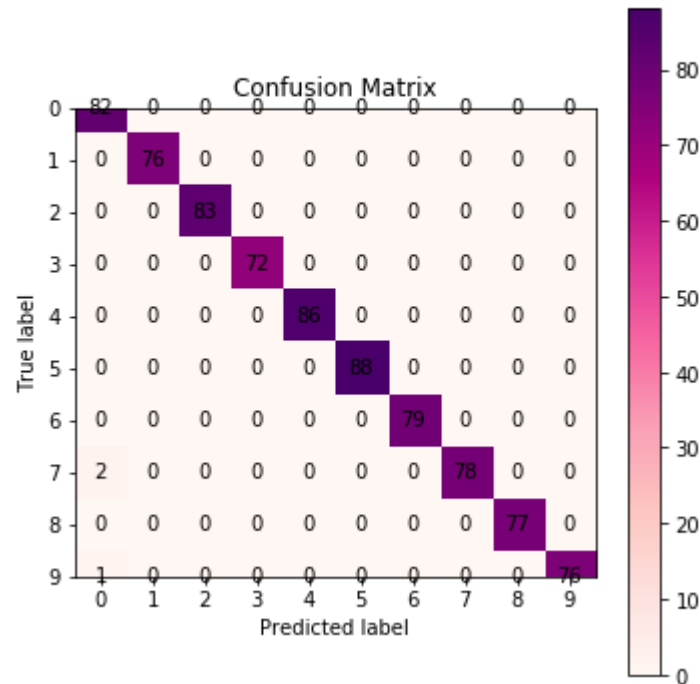
Pembahasan

Pada percobaan pertama yang menggunakan *hidden layer* berjumlah 2 buah neuron menghasilkan akurasi 55% untuk data latih. Sedangkan pada data uji menghasilkan akurasi 35%. Pada percobaan kedua yang menggunakan *hidden layer* berjumlah 4 buah neuron menghasilkan akurasi 55% untuk data latih. Sedangkan pada data uji menghasilkan akurasi 47%. Pada percobaan ketiga yang menggunakan *hidden layer* berjumlah 6 buah neuron menghasilkan akurasi 99% untuk data latih. Sedangkan pada data uji menghasilkan akurasi 57%. Pada percobaan keempat yang menggunakan *hidden layer* berjumlah 8 buah neuron menghasilkan akurasi 76% untuk data latih. Sedangkan pada data uji menghasilkan akurasi 46%. Berdasarkan dari percobaan tersebut didapatkan kesimpulan bahwa percobaan ketiga dengan menggunakan *hidden layer* berjumlah 6 buah neuron menghasilkan akurasi yang paling optimal. Akurasi pada data latih

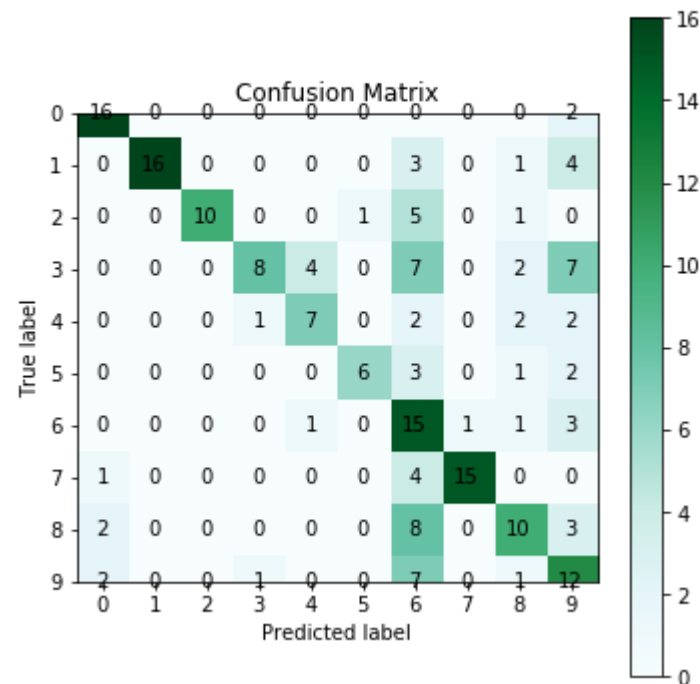
lebih besar ketimbang akurasi pada data uji. Hasil klasifikasi yang didapatkan dari data latih pada percobaan ketiga ditunjukkan melalui *confusion matrix* yang dapat dilihat pada gambar 3.

Hasil klasifikasi yang didapatkan dari data uji percobaan ketiga ditunjukkan melalui *confusion matrix* yang dapat dilihat pada gambar 4.

Number of training data : 800
 Data structure : (800, 2912)



Gambar 3. *Confusion matrix* data latih percobaan ketiga
 Number of testing data : 200
 Data structure : (200, 2912)



Gambar 4. *Confusion matrix* data uji percobaan ketiga

4. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini telah menghasilkan sistem klasifikasi kalimat berita olahraga yang terdiri dari sepuluh kelas yakni manajemen, cedera, jadwal, *preview*, *review*, klasemen, statistik, juara, pemain/atlet, dan kelas lainnya. Sistem dibangun menggunakan ANN dengan metode pembelajaran *Backpropagation*. Berdasarkan hasil pengujian, parameter yang paling optimal yaitu *hidden layer* berjumlah 6 buah neuron dengan menghasilkan akurasi 99% pada data latih dan 57% pada data uji. Hasil pengujian penambahan neuron pada *hidden layer* pada percobaan keempat tidak mampu meningkatkan akurasi karena semakin banyak jumlah neuron yang ditambahkan pada *hidden layer*, akurasi yang dihasilkan malah mengalami penurunan. *Dataset* yang *balance* sangat berpengaruh pada akurasi. *Dataset* yang *balance* lebih berpengaruh pada akurasi data uji ketimbang data latih. Sistem ini bisa membantu *user* memilah kalimat berita olahraga yang diperlukan.

Saran

Saran untuk penelitian selanjutnya supaya menghasilkan klasifikasi yang lebih baik adalah dengan menambahkan input data latih yang lebih banyak sehingga hasil analisa menjadi semakin optimal dan akurat. Pada proses *training* dapat menggunakan variabel yang lebih bervariasi, sehingga menghasilkan hasil prediksi yang lebih akurat. Diharapkan hasil analisis yang dilakukan ini dapat dikembangkan dan diimplementasikan dengan membangun sebuah sistem yang menerapkan metode ANN untuk melakukan prediksi kalimat pada berita olahraga atau berita lainnya.

DAFTAR PUSTAKA

- [1] G. Tika and Adiwijaya, "Klasifikasi Topik Berita Berbahasa Indonesia menggunakan Multilayer Perceptron," *e-Proceeding Eng.*, vol. 6, no. 1, pp. 2137–2143, 2019.
- [2] K. M. Lhaksana, F. Nhita, and D. Anggraini, "Klasifikasi Kepribadian Berdasarkan Status Facebook Menggunakan Metode *Backpropagation*," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 5174–5183, 2017.
- [3] M. A. Assuja and Saniati "Analisis Sentimen Tweet Menggunakan *Backpropagation* Neural Network," *Jurnal TEKNOINFO*, vol. 10, no. 2, pp. 23–28, 2016.
- [4] F. A. Shiddiq, S. Al Faraby, and Adiwijaya "Klasifikasi Sentimen Review Produk Otomotif Menggunakan *Backpropagation* Neural Network," *e-Proceeding of Engineering*, vol. 5, no. 3, pp. 7790–7794, 2018.
- [5] R. R. M. Salim and A. S. Jauhari, "Perancangan Pengenalan Karakter Alfabet menggunakan Metode Hybrid Jaringan Syaraf Tiruan," *J. SIFO Mikroskil*, vol. 17, no. 1, pp. 109–118, 2016.
- [6] S. Setti and A. Wanto, "Analysis of *Backpropagation* Algorithm in Predicting the Most Number of Internet Users in the World," *J. JOIN*, vol. 3, no. 2, pp. 110–115, 2018.
- [7] J. S. D. Raharjo, "Model Artificial Neural Network Berbasis Particle Swarm Optimization Untuk Prediksi Laju Inflasi," *J. SISKOM*, vol. 3, no. 1, pp. 10–21, 2013.
- [8] N. Nikentari, H. Kurniawan, N. Ritha, and D. Kurniawan, "Optimasi Jaringan Syaraf Tiruan *Backpropagation* Dengan Particle Swarm Optimization Untuk Prediksi Pasang Surut Air Laut," *J. JTIK*, vol. 5, no. 5, pp. 605-612, 2018.
- [9] L. S. Lubis and A. Buono, "Pemodelan Jaringan Syaraf Tiruan Untuk Memprediksi Awal Musim Hujan Berdasarkan Suhu Permukaan Laut Artificial Neural Network Modeling To Predict The Beginning of Rainy Season Based On Sea Surface Temperature," *J. Ilmu Komputer*, vol. 1, no. 2, pp. 52-61, 2012.
- [10] Tursina, "Pendekatan Artificial Neural Network (ANN) Untuk Mengklasifikasikan Hewan Vetebrata Menggunakan Kohonen Self Organizing Map (SOM)," *J. Informatika*, vol. 13, no. 1, pp. 63–70, 2013.
- [11] Yunita, "Prediksi Cuaca Menggunakan Metode Neural Network," *J. Paradigma*, vol. XVII, no. 2, pp. 47–53, 2015.
- [12] F. Rozi and F. Sukmana, "Penggunaan Moving Average Dengan Metode Hybrid Artificial Neural Network Dan Fuzzy Inference System Untuk Prediksi Cuaca," *J. JIPI*, vol. 1, no. 02, pp. 38–42.
- [13] N. Yanti, "Penerapan Metode Neural Network Dengan Struktur *Backpropagation* Untuk Prediksi Stok Obat Di Apotek (Studi Kasus : Apotek ABC)," *J. SNATI*, pp. 17–18, 2011.
- [14] Julpan, E. B. Nababan, and M. Zarlis, "Analisis Fungsi Aktivasi Sigmoid Biner Dan Sigmoid Bipolar Dalam Algoritma *Backpropagation* Pada Prediksi Kemampuan Siswa," *J. TEKNOVASI*, vol. 02, no. 1, pp. 103-116, 2015.

