

## ANALISIS SENTIMEN PADA DATA SARAN MAHASISWA TERHADAP KINERJA DEPARTEMEN DI PERGURUAN TINGGI MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK

Yuliska<sup>\*1</sup>, Dini Hidayatul Qudsi<sup>2</sup>, Juanda Hakim Lubis<sup>3</sup>, Khairul Umam Syaliman<sup>4</sup>, Nina Fadilah Najwa<sup>5</sup>

<sup>1,2,4,5</sup> Politeknik Caltex Riau, <sup>3</sup> Universitas Medan Area Medan

Email: <sup>1</sup>yuliska@pcr.ac.id, <sup>2</sup>dinihq@pcr.ac.id, <sup>3</sup>juandahakim@gmail.com, <sup>4</sup>khairul@pcr.ac.id, <sup>5</sup>nina@pcr.ac.id

\*Penulis Korepondensi

(Naskah masuk: 15 Maret 2021, diterima untuk diterbitkan: 19 Oktober 2021)

### Abstrak

*Review* atau saran dari *customer* dapat menjadi sangat penting bagi penyedia layanan, begitu pula saran dari mahasiswa mengenai layanan sebuah unit kerja di perguruan tinggi. *Review* menjadi penting karena dapat menjadi indikator kinerja penyedia layanan. Pengolahan *review* juga sangat penting karena dapat menjadi referensi untuk pengambilan keputusan dan peningkatan layanan yang lebih baik ke depannya. Penelitian ini menerapkan analisis sentimen pada data saran atau *review* mahasiswa terhadap kinerja unit kerja atau departemen di perguruan tinggi, yaitu Politeknik Caltex Riau. Analisis sentimen dilakukan dengan menggunakan *Convolutional Neural Network (CNN)* dan *word embedding Word2vec* sebagai representasi kata. *CNN* merupakan metode yang memiliki performa yang baik dalam mengklasifikasi teks, yaitu dengan teknik *convolutional* yang menggabungkan beberapa *window* kata pada kalimat dan mengambil *window* yang paling *representative*. *Word2Vec* digunakan sebagai representasi data saran dan inputan awal pada *CNN*, dimana *Word2Vec* merupakan *dense vectors* yang dapat merepresentasikan hubungan antar kata pada data saran dengan baik. Saran mahasiswa dapat mengandung kalimat yang sangat panjang, karena itu perpaduan *Word2Vec* sebagai representasi kata dan *CNN* dengan teknik *convolutional*, dapat menghasilkan representasi yang *representative* dari kalimat panjang tersebut. Penelitian ini menggunakan dua arsitektur *CNN*, yaitu *Simple CNN* dan *DoubleMax CNN* untuk mengidentifikasi pengaruh kompleksitas arsitektur terhadap hasil klasifikasi sentimen. Berdasarkan hasil pengujian, *DoubleMax CNN* dapat mengklasifikasi sentimen pada saran mahasiswa dengan sangat baik, yaitu mencapai Akurasi tertinggi sebesar 98%, *Recall* 97%, *Precision* 98% dan *F1-Score* 98%.

**Kata kunci:** Analisis Sentimen, Convolutional Neural Network, Kinerja Department, Saran Mahasiswa

## SENTIMENT ANALYSIS IN THE STUDENT'S REVIEWS OF COLLEGE DEPARTMENT PERFORMANCE USING CONVOLUTIONAL NEURAL NETWORK

### Abstract

*Student's reviews about department performance can be essential for a college for it can be used to evaluate the department performance and to take an immediate action to improve its performance. This research applies sentiment analysis in the student's reviews of college department in Politeknik Caltex Riau. Convolutional Neural Network and Word2Vec are employed to analyze the sentiment. CNN is known for its good performance in text classification by applying a convolutional technique to the input sentences. Word2Vec is used as word representation and as an input to the CNN. Word2Vec are dense vectors which can represent the relationship between words excellently. Student's reviews can be a long sentence; hence the combination of Word2Vec as word representation and CNN with convolutional technique can produce a representative figure from that long sentence. This research utilizes two CNN architectures, which are Simple CNN dan DoubleMax CNN to identify the effect of the complexity of CNN architecture to final result. Our experiments show that DoubleMax CNN has a great performance in classifying sentiment in the student's reviews with the best Accuracy value of 98%, Recall 97%, Precision 98% and F1-Score value of 98%.*

**Keywords:** Convolutional Neural Network, Department Performance, Sentiment Analysis, Student's Reviews

### 1. PENDAHULUAN

Semakin berkembangnya teknologi, semakin banyak pula pengguna yang menggunakan internet

sebagai sumber informasi. Peranan internet sebagai sumber informasi membuat banyak pengguna menggunakan internet untuk berbagi pendapat pribadinya yang dituangkan dalam bentuk saran atau

review. Saran atau review ini dapat berupa pendapat pribadi pengguna mengenai kualitas sebuah film, kualitas produk, pelayanan toko online, pelayanan sebuah lembaga, pelayanan hotel, dan lain sebagainya (Pang & Lee, 2008). Banyak situs atau *website* yang menyediakan fitur kuesioner agar pengguna dapat menyampaikan pendapat mereka mengenai hotel, restoran, atau tempat-tempat wisata (Tsytsarau & Palpanas, 2012). Hasil review kemudian akan digunakan untuk pengambilan keputusan yang lebih baik (Raut & Londhe, 2014). Fitur kuesioner atau review ini juga diterapkan pada perguruan tinggi, salah satunya untuk memberikan review terhadap kinerja dosen (Salam, Supianto & Perdanakusuma, 2019). Namun, kinerja dosen bukan satu-satunya indikator untuk mengukur kualitas sebuah perguruan tinggi. Pelayanan yang baik dari unit-unit atau departemen pada sebuah perguruan tinggi juga menjadi salah satu indikator penting, karena itu pengolahan data saran secara efisien sangat diperlukan. Salah satu teknik yang dapat diterapkan agar pengolahan data saran menjadi lebih efisien yaitu teknik analisis sentimen. Analisis sentimen adalah suatu proses klasifikasi menjadi dua *polarity* atau kecenderungan (*binary classification*), yaitu positif dan negatif (Gupta, Tiwari & Robert, 2016). Dengan penerapan analisis sentimen terhadap data saran, akan diketahui kecenderungan saran mahasiswa atau kepuasan mahasiswa terhadap kinerja sebuah unit kerja, apakah positif atau negatif. Sehingga pengambilan keputusan dan peningkatan layanan untuk kinerja yang lebih baik ke depannya dapat segera diterapkan.

Hamzah (2014) mengukur tingkat kepuasan mahasiswa dalam proses pembelajaran di AKPRIND dengan memanfaatkan analisis sentimen. Penelitian ini menggunakan metode *Naïve Bayes* dalam mengklasifikasi sentimen, dengan representasi kalimat menggunakan teknik *TF-IDF*. Dengan menggunakan *Naïve Bayes* dan *TF-IDF* sebagai representasi kata, penelitian ini berhasil mendapatkan rata-rata akurasi 85.95%.

Salam, Supianto & Perdanakusuma (2019) juga melakukan analisis sentimen pada opini mahasiswa terhadap kinerja dosen di Fakultas Ilmu Komputer Universitas Brawijaya (FILKOM UB). Penelitian ini dilakukan dengan tujuan agar data saran mahasiswa dapat diolah secara mendalam dan dijadikan bahan evaluasi untuk peningkatan kinerja dosen. Kombinasi *TF-IDF* dan *K-Nearest Neighbor (KNN)* digunakan dalam mengklasifikasi sentimen pada opini mahasiswa yang menghasilkan rata-rata akurasi sebesar 81%.

Penelitian selanjutnya menggunakan *Support Vector Machine (SVM)* untuk menganalisis sentimen data kritik dan saran dari sebuah pelatihan bertema teknologi informasi (APLINET) (Al Kabir, Basuki & Wicaksono, 2019). Analisis sentimen dilakukan untuk mengevaluasi tanggapan peserta dari pelatihan yang telah dilakukan. Proses analisis sentimen diawali dengan melakukan beberapa langkah *text preprocessing*, kemudian dilakukan pembobotan *TF-*

*IDF*. Proses akhir dari penelitian ini yaitu klasifikasi sentimen data saran menggunakan *Support Vector Machine* yang mendapatkan akurasi sebesar 82.08%.

Kemudian, analisis sentimen terhadap teks berbahasa Indonesia dengan menggunakan salah satu metode *deep learning* yaitu *Convolutional Neural Network (CNN)* telah dilakukan oleh (Juwiantho et al., 2020), yaitu terhadap 999 data *Twitter* berbahasa Indonesia. Berbeda dengan penelitian-penelitian sebelumnya, penelitian ini menggunakan *Word2Vec* dengan berbagai dimensi (50 s/d 300) sebagai representasi kata dengan arsitektur *CNN* yang digunakan adalah *CNN* dengan sebuah *convolutional layer* yang diikuti dengan sebuah operasi *Max-Pooling*. Hasil percobaan dari penelitian ini adalah akurasi terbaik sebesar 76.40% dengan pilihan dimensi *Word2Vec* terbaik yaitu 100. Walau akurasi yang didapat sudah cukup baik, akurasi masih dapat ditingkatkan dengan penggunaan arsitektur *CNN* yang lebih kompleks, seperti penambahan layer konvolusi dan pemilihan *filter* yang tepat.

Penggunaan *CNN* dengan berbagai arsitektur untuk menganalisis sentimen dilakukan oleh (Kim & Jeong, 2019), namun untuk korpus dengan teks berbahasa Inggris, yaitu *user review* terhadap kualitas film dan produk. Penelitian ini menggunakan arsitektur *CNN* mulai dari yang sangat sederhana yang hanya terdiri dari sebuah layer konvolusi, hingga arsitektur yang terdiri dari 3 buah layer konvolusi. Hasil dari penelitian ini menunjukkan bahwa *CNN* yang memiliki lebih dari sebuah layer konvolusi bekerja lebih baik dibandingkan dengan *CNN* dengan sebuah layer konvolusi. Namun, *CNN* dengan banyak *convolutional layer* tidak selalu memiliki performa yang tinggi. Kompleksitas atau kedalaman *network* harus ditentukan berdasarkan karakteristik data, karena *network* yang terlalu kompleks terkadang dapat mengakibatkan *overfitting*.

Pada penelitian ini, data yang digunakan adalah data review atau kuisisioner mahasiswa terhadap kinerja 7 departemen kerja di Politeknik Caltex Riau (PCR). Politeknik Caltex Riau adalah sebuah perguruan tinggi swasta di provinsi Riau yang berdiri pada tahun 2001. Data saran mahasiswa terhadap kinerja suatu unit dapat mencapai 1500 saran per departemen, dengan jumlah kata dapat mencapai 150 kata per data saran. Kontribusi dari penelitian ini adalah penerapan metode *Convolutional Neural Network (CNN)* untuk mengklasifikasi sentimen data saran mahasiswa terhadap kinerja atau pelayanan sebuah unit kerja. Metode *CNN* yang diusulkan memiliki 2 buah arsitektur. Arsitektur pertama terdiri dari 2 *layer* dengan operasi *Max-Pooling* di masing-masing *layer* (*DoubleMax CNN*). Arsitektur kedua terdiri dari 2 *convolutional layer* dengan operasi *Max-Pooling* di layer terakhir (*Simple CNN*). Dua buah arsitektur *CNN* ini digunakan untuk mengidentifikasi pengaruh kompleksitas arsitektur *CNN* terhadap hasil klasifikasi sentimen terhadap data saran, terutama klasifikasi sentimen terhadap kalimat yang panjang karena saran mahasiswa terhadap sebuah unit dapat berupa kalimat

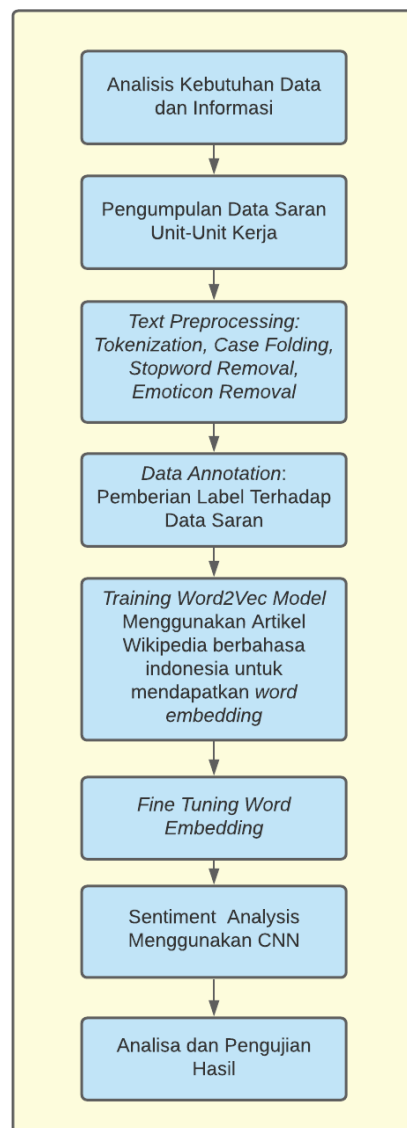
yang sangat panjang (150 kata). Oleh sebab itu, penelitian ini tidak menggunakan *TF-IDF* sebagai representasi kata, karena *TF-IDF* tidak dapat merepresentasikan keterkaitan antar kata dengan baik. Penelitian ini menggunakan *Word2Vec* berbahasa Indonesia sebagai representasi kata. *Word2Vec* terbukti dapat menjadi representasi yang lebih baik dibanding *TF-IDF* untuk teks berbahasa Indonesia (Juwiantho et al., 2020). *Word2Vec* akan menjadi input awal *CNN*, kemudian *CNN* akan menentukan pola data saran yang dilabeli positif, negatif dan netral. Hasil penelitian ini diharapkan dapat menghasilkan teknik atau metode yang dapat mengklasifikasi sentimen dengan 2 dan 3 kecenderungan yang memiliki performa tinggi.

## 2. METODE PENELITIAN

Tahapan-tahapan yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 1. Penelitian diawali dengan melakukan analisis terhadap masalah yang terjadi terkait pengolahan data saran mahasiswa terhadap kinerja unit kerja atau departemen di PCR. Dari tahapan analisis masalah ini, teridentifikasi bahwa pengolahan data saran mahasiswa terhadap kinerja unit kerja PCR masih belum menerapkan analisis sentimen. Sehingga data saran belum terfilter menjadi saran positif dan negatif yang dapat digunakan untuk pengambilan keputusan dan perbaikan kualitas pelayanan. Kemudian dilanjutkan dengan pengumpulan data saran dimana data saran yang dikumpulkan adalah dalam bentuk file.csv yang didapatkan dari Badan Perencanaan, Pengembangan dan Penjaminan Mutu (BP3M) PCR. Proses selanjutnya yaitu proses *preprocessing* terhadap data yang telah terkumpul.

Setelah itu, karena metode yang akan digunakan untuk melakukan analisis sentimen atau *sentiment analysis* adalah metode *supervised*, maka perlu dilakukan proses *labelling* terhadap data saran agar proses *training* dapat dilakukan. *Labelling* yang dilakukan pada penelitian ini adalah *automatic labelling*, yaitu memberikan label sentimen positif, negatif dan netral kepada setiap data saran, yang dilakukan secara otomatis dengan menggunakan beberapa *rules*. Setelah tahapan *labelling*, proses selanjutnya yaitu melakukan *training word2vec model* untuk mendapatkan *word embedding* atau representasi kata yang nantinya akan digunakan dalam mengklasifikasi sentimen. *Training word2vec model* pada penelitian ini dilakukan dengan menggunakan artikel wikipedia berbahasa indonesia, yang terdiri dari 415.521 kalimat dan 443.979 kata unik. Proses *fine tuning* juga dilakukan terhadap *word embedding* yang sudah didapat, agar *word embedding* menjadi lebih sesuai dengan kata-kata yang digunakan pada penelitian ini. Kemudian, proses klasifikasi sentimen dengan menggunakan *CNN* dilakukan. Proses terakhir dari penelitian ini adalah melakukan pengujian terhadap analisis sentimen yang dilakukan dengan menggunakan *CNN* dan menganalisa hasil pengujian. Hasil pengujian dengan *CNN* juga akan dibandingkan dengan beberapa metode *machine learning* (*Support Vector Machine*,

*logistic regression* dan *Naive Bayes*) untuk memastikan efektifitas performa *CNN*.



Gambar 1. Tahapan Penelitian

### 2.1. Analisis Sentimen

Analisis sentimen atau *Sentiment analysis* adalah sebuah bidang yang sangat umum dalam *text classification* atau klasifikasi teks. Analisis sentimen adalah sebuah proses yang menganalisa dan mendeteksi sentimen sebuah inputan teks memiliki sentimen positif, negatif atau netral (Gupta, Tiwari & Robert, 2016). Namun, hingga sekarang, sentimen yang dapat dideteksi menjadi lebih beragam dan mendetil dan tidak terbatas hanya kepada positif dan negatif, yaitu dapat mendeteksi *happiness* (bahagia), *sadness* (sedih), *anger* (marah), *fear* (takut), *disgust* ( jijik) dan *surprise* (terkejut) (Seyeditabari, Tabari & Zadrozny, 2018).

Analisis sentimen dapat digunakan salah satunya untuk melakukan monitoring terhadap kualitas atau performa sebuah produk dan pelayanan sebuah lembaga. Dengan menerapkan analisis sentimen pengembang produk dan pemilik layanan dapat dengan

mudah mengetahui apakah sebuah produk dan layanan diterima secara positif oleh pelanggan atau malah sebaliknya. Kegunaan analisis sentimen selanjutnya yaitu untuk menganalisa sebab dibalik diterima atau tidak diterimanya sebuah produk atau layanan (Liu, 2010). Oleh karena itu, analisis sentimen dapat digunakan oleh hampir di semua bidang.

Penelitian di bidang analisis sentimen untuk teks berbahasa Indonesia telah banyak dilakukan, seperti yang sudah dijelaskan di pendahuluan. Data yang digunakan mulai dari *review* mahasiswa terhadap kinerja dosen, *review* produk (Attabi, Muflikhah & Fauzi, 2018), hingga analisis sentimen terhadap data *Twitter*. Metode yang digunakanpun beragam, mulai seperti *SVM*, *Naïve Bayes*, *KNN*, hingga metode berbasis *Deep Learning*, seperti *CNN*.

## 2.2. Convolutional Neural Network (CNN)

*Convolutional neural network* pada awalnya digunakan pada bidang *image processing* atau *computer vision*. Kemudian pada tahun 2015, *CNN* digunakan pada bidang *natural language processing (NLP)*, yaitu untuk melakukan klasifikasi teks (Kim, 2014). Inti dari *CNN* dalam mengklasifikasi teks yaitu dengan menerapkan teknik *convolution* terhadap kalimat, paragraf, atau keseluruhan dokumen teks. *Convolution* pada dasarnya adalah *sliding window* yang disebut juga dengan *filter*, yaitu membagi representasi teks yang berupa matriks menjadi beberapa *window*, kemudian menjumlahkannya. Hasil penjumlahan inilah yang menjadi representasi baru dari teks yang disebut dengan *feature maps*. Pada *CNN*, dikenal pula teknik *max-pooling*, yaitu pemilihan nomor terbesar dari setiap *feature map*.

Berikut adalah keterangan poin-poin arsitektur *CNN* pada Gambar 2:

- 1) Representasi Kalimat: Setiap kata pada kalimat input merupakan sebuah vector yang berdimensi  $k$ . Jadi, sebuah kalimat merupakan *concatenation* atau operasi gabungan dari  $n$  representasi kata, yang menghasilkan sebuah matriks  $\mathbf{S} \in \mathbb{R}^{k \times n}$ , matriks ini yang digunakan sebagai input ke *convolutional layer*. Layer pertama merupakan *embedding layer* yang mengandung sebuah *lookup table* di mana *word embedding* direpresentasikan sebagai matriks  $\mathbf{X} \in \mathbb{R}^{k \times |V|}$ ,  $V$  adalah *vocabulary* atau kamus kata. Jadi, kolom ke- $i$  pada  $\mathbf{X}$  merepresentasikan kata ke- $i$  pada kamus kata  $V$ .
- 2) *Convolutional Layer*: Pada layer ini, operasi *Convolution* dilakukan dengan melibatkan sebuah

*filter*  $\mathbf{w} \in \mathbb{R}^{hk}$ , yang diterapkan pada sebuah *window* dari  $h$  kata untuk menghasilkan sebuah fitur baru. Sebagai contoh, sebuah fitur  $c_i$  dihasilkan dari sebuah *window* kata  $\mathbf{s}_{i:i+h-1}$  dengan menggunakan persamaan sebagai berikut:

$$c_i = f(\mathbf{w} \cdot \mathbf{s}_{i:i+h-1} + b) \quad (1)$$

Dari persamaan (1) di atas,  $b \in \mathbb{R}^m$  adalah sebuah bias,  $f$  adalah sebuah fungsi non-linier. *Filter* ini digunakan pada setiap *window* kata yang terdapat pada kalimat  $\{\mathbf{s}_{1:h}, \mathbf{s}_{2:h+1}, \dots, \mathbf{s}_{n-h+1:n}\}$  untuk menghasilkan sebuah *feature map* sebagai berikut:

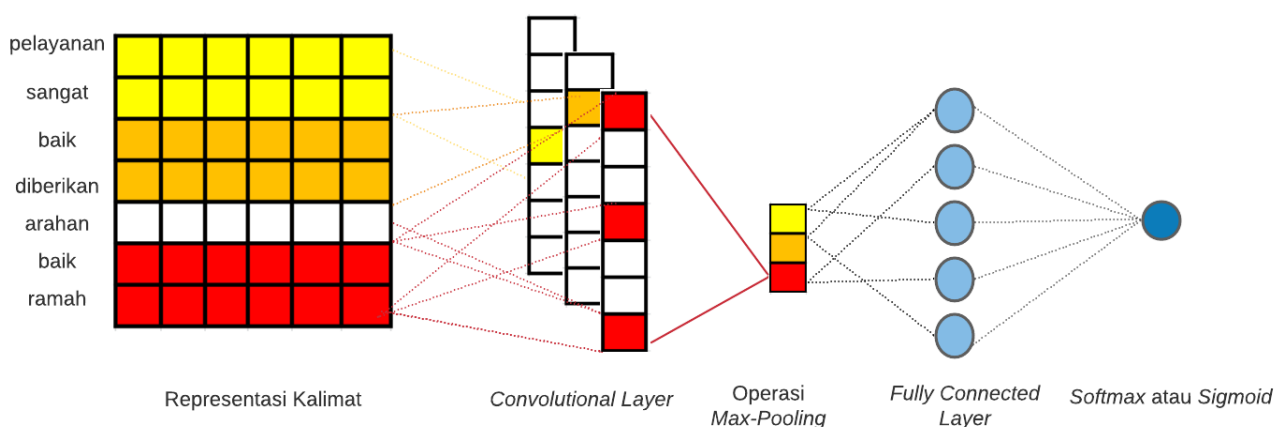
$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}] \quad (2)$$

Di mana  $\mathbf{c} \in \mathbb{R}^{n-h+1}$ .

- 3) *Max-Pooling*: Operasi *Max-Pooling* adalah mengambil nilai maksimum dari sebuah *feature map*.

$$\hat{\mathbf{c}} = \max\{\mathbf{c}\} \quad (3)$$

- 4) *Fully Connected Layer*: Sebuah *Fully Connected Layer* menghitung transformasi  $\alpha(\mathbf{W} * \mathbf{x} + b)$ , dimana  $\mathbf{W} \in \mathbb{R}^{m \times m}$  adalah *weight matrix*,  $\alpha$  adalah sebuah *activation function* dan  $\mathbf{x} \in \mathbb{R}^m$  adalah representasi baru dari kalimat (*sentence embedding*), yang merupakan output dari layer ini.
- 5) *Dropout*: Adalah sebuah teknik alternatif untuk mengurangi terjadinya *overfitting*, yaitu keadaan di mana dapat diklasifikasi dengan baik pada saat *training*, namun bekerja sebaliknya pada saat *evaluation* dan *testing*. Pada setiap fase *training*, setiap node atau unit *network* dimatikan sebanyak probabilitas  $p$ .
- 6) *Optimization*: Adalah sebuah teknik yang digunakan untuk mengubah atribut dari *neural network*, seperti *weight* dan *learning rate* untuk mengurangi *loss* dari model *neural network* selama *training*. Ada banyak *optimizer* yang dapat digunakan, yaitu *Adadelta* (Zeiler, 2014), *Adam* (Kingma & Ba, 2015), *Nadam* (Dozat, 2016) dan *SGD* (Ruder, 2016). Penelitian ini menggunakan *Adam* sebagai *optimizer*.
- 7) *Sigmoid*: Output dari *Fully Connected Layer*  $\mathbf{x} \in \mathbb{R}^m$  terhubung ke sebuah *sigmoid layer*, yang memberikan output berupa 2 buah *class*, dimana  $\hat{y} \in [0, 1]$ .

Gambar 2. Arsitektur *Convolutional Neural Network* (CNN)

$$\hat{y} := \frac{1}{1+e^{-(x^T W+b)}} \quad (4)$$

Output akan bernilai 1 jika  $\hat{y}$  bernilai lebih besar dari *threshold* (pada penelitian ini, *threshold*=0.5), dan akan bernilai 0 jika sebaliknya.

- 8) *Softmax*: Pada *multiclass classification*, Output dari *Fully Connected Layer*  $\mathbf{x} \in \mathbb{R}^m$  terhubung ke sebuah *softmax layer*, yang memberikan output berupa  $K$  buah *class*, dimana  $\hat{y} \in [0, K]$  dengan probabilitas terbesar.

$$\hat{y} := \operatorname{argmax}_j \frac{e^{(x^T w_j + b_j)}}{\sum_{k=1}^K e^{(x^T w_k + b_k)}} \quad (5)$$

### 2.3. Word2Vec

*Word2Vec* (Mikolov et al., 2013) adalah salah satu cara merepresentasikan teks dalam bentuk *dense vector* yang disebut juga dengan *word embedding*. *Word2Vec* dapat memiliki 50 sampai dengan 300 dimensi. *Word2Vec* mulai ramai digunakan dalam bidang *natural language processing* di tahun 2013, karena *Word2Vec* merupakan *dense vectors* yang dapat merepresentasikan hubungan antar kata dengan lebih baik (dibandingkan dengan *TF-IDF*), secara semantik maupun sintaksis. Google telah menyediakan *pretrained Word2Vec* yang dapat langsung digunakan tanpa perlu melatih atau *training model Word2Vec* untuk mendapatkan *word embedding*. Namun, *pretrained Word2Vec* berbahasa Indonesia belum tersedia, sehingga pada penelitian ini dilakukan *training model Word2Vec* untuk mendapatkan *word embedding* berbahasa Indonesia.

### 2.4. Parameter Pengujian (*Evaluation Metrics*)

Parameter pengujian yang digunakan pada penelitian ini yaitu beberapa metrik *confusion matrix*, yaitu akurasi, *recall*, *precision*, dan *f-measure*.

- 1) *Recall*: Menunjukkan jumlah data saran dengan label positif yang terklasifikasi secara benar diantara data yang diprediksi sebagai positif.

$$\operatorname{Recall} = \frac{TP}{TP+FN} \quad (6)$$

- 2) *Precision*: Menunjukkan jumlah data saran dengan label positif yang terklasifikasi secara benar diantara semua data dengan label positif.

$$\operatorname{Precision} = \frac{TP}{TP+FP} \quad (7)$$

- 3) Akurasi: Menunjukkan jumlah data saran yang terklasifikasi secara benar (baik yang berlabel positif maupun negatif).

$$\operatorname{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

- 4) *F1-Score*: Menunjukkan *harmonic mean* dari *precision* dan *recall*.

$$\operatorname{F1\ Score} = \frac{2TP}{2TP+FP+FN} \quad (9)$$

Keterangan:

- *True Positif (TP)*: *Review* dengan label positif yang terklasifikasi sebagai *review* positif
- *True Negatif (TN)*: *Review* dengan label negatif yang terklasifikasi sebagai *review* negatif
- *False Positif (FP)*: *Review* dengan label positif yang terklasifikasi sebagai *review* negatif
- *False Negatif (FN)*: *Review* dengan label negatif yang terklasifikasi sebagai *review* positif

## 3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dibahas data yang digunakan pada penelitian ini, *preprocessing* yang dilakukan terhadap data yang digunakan, *automatic labeling*, *training Word2Vec model* untuk mendapatkan *word embedding*, *training model CNN*, *testing model* dan menganalisa hasil.

### 3.1. Data yang Digunakan

Data yang digunakan pada penelitian ini adalah data hasil *Quality Assurance (QA)* penilaian performa layanan unit kerja atau departemen di Politeknik Caltex Riau di semester genap Tahun Ajaran 2018/2019 yang didapatkan dari BP3M PCR. Data QA setiap unit terdiri dari nilai angka dengan range 1-4 untuk beberapa poin pelayanan di masing-masing unit dan data saran atau *review* dari mahasiswa mengenai pelayanan di masing-masing unit. Yang digunakan pada penelitian ini yaitu data saran atau *review* dari

mahasiswa. Jumlah data saran di masing-masing unit adalah mulai dari 1470 hingga 1500, namun hanya 1200 data saran saja yang akan digunakan, sedangkan 300 data saran lainnya dieliminasi berdasarkan banyak kata yang terkandung di dalamnya dan juga demi keseimbangan data saran di masing-masing kelas (kelas data saran dengan label positif, negatif dan netral). Jumlah kata pada data saran mulai dari 2 kata hingga 150 kata. Tabel 1 berikut menunjukkan contoh data saran di salah satu unit.

Tabel 1. Contoh Data Saran di Salah Satu Unit

Nomor	Saran
1	Pelayanan Sangat baik dan diberikan arahan dengan baik dan ramah 😊😊
2	Terapkan pembayaran angsuran agar mahasiswa yang tidak mampu dapat kemudahan
3	Jangan sampai antrian panjang menanti
4	Pelayanan yang d berikan sudah baik, penjaga nya juga ramah, dalam menindaklanjuti pembludakan pembayaran serentak yang dilakukan mahasiswa harus d perbaiki.
5	Saran dari saya supaya tempat pembayaran SPP di tambah karna menurut saya kalau 1 bank masi kurang cukup, sebab admin bank akan kerepotan mengurus mahasiswa yang mengurus SPP tersebut
6	Bagian keuangan sudah memberikan pelayanan dan tanggapan yang memuaskan
7	Sebaiknya untuk antrian di kasi nomor, agar tidak ada yang memotong saat ada urusan keuangan

### 3.2. Text Preprocessing

Penelitian ini menggunakan *NLTK (Natural Language ToolKit)* sebagai *library* untuk pengolahan kata. Tahapan-tahapan *preprocessing* yang diterapkan pada penelitian ini yaitu sebagai berikut:

- 1) *Case Folding*: Proses pengubahan huruf besar menjadi huruf kecil.

Tabel 2. Contoh *Case Folding*

Saran	Case Folding
Pelayanan Sangat baik dan diberikan arahan dengan baik dan ramah 😊😊	pelayanan sangat baik dan diberikan arahan dengan baik dan ramah 😊😊

- 2) *Stopword Removal*: Proses mengeliminasi kata-kata yang dianggap tidak penting.

Tabel 3. Contoh *Stopword Removal*

Saran	Stopword Removal
pelayanan sangat baik dan diberikan arahan dengan baik dan ramah 😊😊	pelayanan sangat baik diberikan arahan baik ramah 😊😊

- 3) *Emoticon Removal*: Proses mengeliminasi *emoticon-emoticon* yang terdapat dalam kalimat.

Tabel 4. Contoh *Emoticon Removal*

Saran	Emoticon Removal
pelayanan sangat baik dan diberikan arahan dengan baik dan ramah 😊😊	pelayanan sangat baik diberikan arahan baik ramah

- 4) *Tokenization*: Pemecahan kalimat menjadi kumpulan kata.

Tabel 5. Contoh *Tokenization*

Saran	Tokenization
pelayanan sangat baik diberikan arahan baik ramah	["pelayanan", "sangat", "baik", "diberikan", "arahan", "baik", "ramah"]

### 3.3. Automatic Labelling

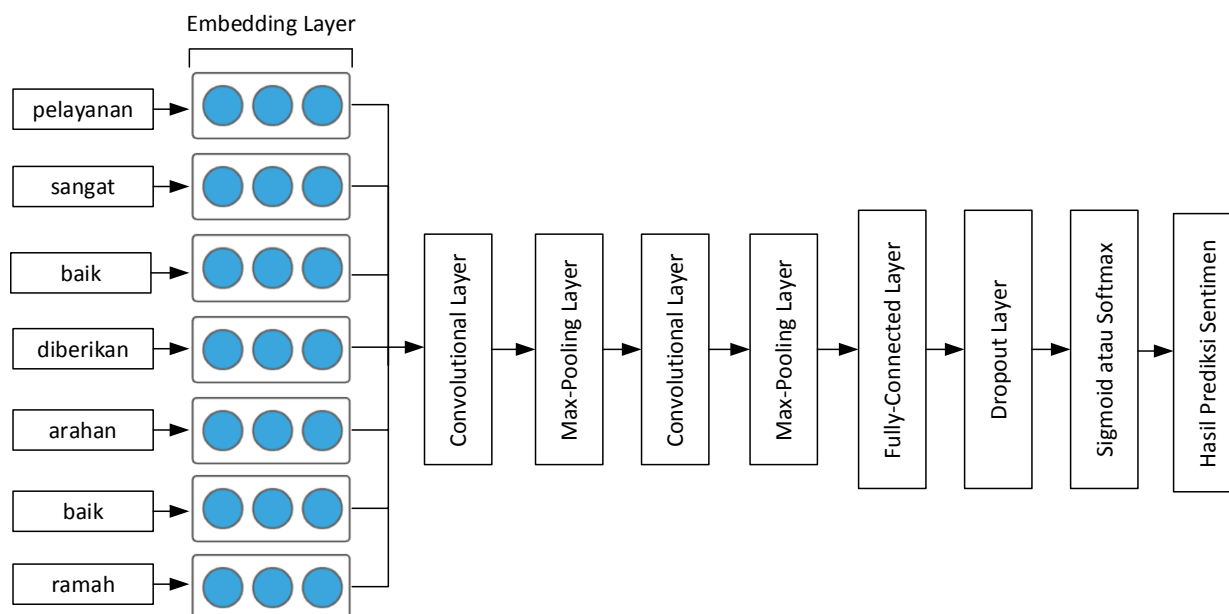
*Data labeling* pada penelitian ini dilakukan secara otomatis, karena untuk melakukan *manual labeling* dibutuhkan waktu yang cukup lama. *Automatic labeling* telah banyak dilakukan, salah satunya melabeli kalimat sebagai positif, jika kalimat tersebut mengandung lebih banyak kata positif dan sebaliknya. Kata-kata negasi, seperti tidak dan bukan, juga akan mempengaruhi proses pelabelan (Wicaksono et al., 2014). Pada penelitian ini, pelabelan data dilakukan dengan cara yang hampir sama, yaitu memberikan skor sentimen berdasarkan kata-kata yang dikandung dalam data tersebut dengan menggunakan *SentiStrength* (Wahid & SN, 2016). Berdasarkan skor yang didapat, data saran kemudian diberikan label positif yang ditandai dengan angka 1 (jika skor total kata dengan sentimen positif lebih besar dari skor total kata dengan sentiment negative dan netral), label negatif yang ditandai dengan angka 0 (jika skor total kata dengan sentiment positif lebih kecil dari skor total kata dengan sentiment negative dan netral) dan label netral yang ditandai dengan angka 2 (jika skor total kata dengan sentiment netral lebih besar dari skor total kata dengan sentiment negative dan positif).

Tabel 6. Data Saran dengan Label

Departemen	Jumlah Saran	Label		
		Positif	Negatif	Netral
BAAK	1200	600	300	300
BSTI	1200	600	300	300
Kemahasiswaan	1200	600	300	300
Kerja Sama	1200	600	300	300
Keuangan	1200	600	300	300
Perpustakaan	1200	600	300	300
Sarana	1200	600	300	300
<b>TOTAL</b>	<b>8400</b>	<b>4200</b>	<b>2100</b>	<b>2100</b>

### 3.4. Training Word2Vec Model

*Training Word2vec* model dilakukan untuk mendapatkan *word embedding* yang akan digunakan dalam proses klasifikasi sentiment nantinya. *Training Word2vec model* pada penelitian ini dilakukan dengan menggunakan kumpulan artikel Wikipedia berbahasa Indonesia (terdiri dari 415.521 kalimat dan 443.979 kata unik) dan *tool* yang digunakan yaitu *gensim*. Kemudian, *word embedding* yang sudah didapat di *fine-tune* dengan kata-kata yang ada pada data saran, agar *word embedding* yang terbentuk lebih sesuai konteks. Dimensi *word embedding* yang digunakan pada penelitian ini yaitu 50 dimensi.

Gambar 3. Arsitektur *Double-Max CNN*

### 3.5. Skema Training

Untuk melakukan *training* model *CNN*, data pada Tabel 5 digabungkan dan diacak dan kemudian dibagi menjadi data *training*, data *evaluation*, dan data *testing* dengan *ratio* 8:1:1. Data dibagi menjadi data *training*, data *evaluation*, dan data *testing* untuk memastikan performa model pada data *evaluation* terlebih dahulu sebelum menerapkannya pada data *testing*. Jika performa model pada saat *training* dan *evaluation* tidak jauh berbeda, maka dapat dikatakan model telah bekerja dengan baik. *Network* pertama yang digunakan yaitu *Convolutional neural network* dengan 2 buah *Convolutional Layer* dan penerapan *Maxpooling* setelah *Convolutional Layer* kedua (*Simple CNN*). Untuk dapat memahami arsitektur *Simple CNN*, dapat dilihat pada Gambar 3, dengan keadaan *Max-Pooling layer* setelah *Convolutional Layer* pertama dihapuskan. Pada *CNN* dengan arsitektur ini, *hidden size* yang digunakan berukuran 50, *activation function* yang digunakan yaitu *relu*, jumlah *filter* yaitu 2, 3, dan 5, *max-pooling* dengan ukuran 2, dengan *optimizer* yang digunakan yaitu *adam* (*learning rate*=0.001). *Simple CNN* juga menerapkan *dropout* senilai 0.5. Jumlah *epoch* yang digunakan adalah 10 dan *batch size* yaitu 100.

*Network* kedua yang digunakan yaitu *Convolutional neural network* dengan 2 buah *Convolutional Layer* dan penerapan *Maxpooling* setelah masing-masing *Convolutional Layer* (*DoubleMax CNN*), seperti yang dijelaskan pada Gambar 3. Pada *CNN* dengan arsitektur ini, *hidden size* yang digunakan berukuran 50, *activation function* yang digunakan yaitu *relu*, dan jumlah *filter* yaitu 2, 3, dan 5, *max-pooling* dengan ukuran 2, dengan *optimizer* yang digunakan yaitu *adam* (*learning rate*=0.001). *DoubleMax CNN* juga menerapkan *dropout* senilai 0.2. Jumlah *epoch* yang digunakan adalah 10 dan *batch size* yaitu 100.

Semua proses pada sub bab ini diimplementasikan dengan menggunakan library *keras*.

### 3.6. Hasil Pengujian

Pengujian dilakukan dengan menggunakan beberapa *metrics* dari *confusion matrix*, yaitu akurasi dan *f1-score*.

Tabel 7. Hasil Pengujian *CNN* dengan 2 Kelas Sentimen

<i>Network</i>	<i>Filter</i>	<i>Recall</i>	<i>Precision</i>	Akurasi	<i>F1-Score</i>
<i>Simple CNN</i>	2	98%	96%	97%	96%
	3	98%	96%	97%	97%
	5	<b>98%</b>	<b>96%</b>	<b>97%</b>	<b>97%</b>
<i>DoubleMax CNN</i>	2	89%	93%	92%	91%
	3	86%	93%	90%	88%
	5	<b>98%</b>	<b>97%</b>	<b>98%</b>	<b>98%</b>

Tabel 7 menjelaskan mengenai hasil pengujian yang didapat dari klasifikasi 2 kelas sentimen dengan menggunakan *Simple CNN* dan *DoubleMax CNN* dengan dengan *filter* yang digunakan yaitu 2, 3, dan 5. Dari hasil pengujian di tabel 7 dapat diketahui bahwa hasil terbaik didapatkan dengan menggunakan *filter* sebesar 5, baik pada *Simple CNN* maupun *DoubleMax CNN*. Pada *Simple CNN*, hasil pengujian antara menggunakan *filter* dengan ukuran 2, 3, dan 5 tidak terlalu mencolok, dimana *Recall*, *Precision*, Akurasi dan *F1-Score* tampak stabil. Namun, pada *DoubleMax CNN*, ukuran *filter* menjadi sangat berpengaruh, dimana *filter* dengan ukuran 5 mendapatkan hasil terbaik dengan perbedaan hasil yang sangat jauh dibandingkan dengan penggunaan *filter* dengan ukuran 2 dan 3, yaitu mencapai *Recall* 98%, *Precision* 97%, Akurasi 98% dan *F1-Score* 98%.

Tabel 8 menjelaskan mengenai hasil pengujian yang didapat dari klasifikasi 3 kelas sentimen dengan menggunakan *Simple CNN* dan *DoubleMax CNN* dengan dengan *filter* yang digunakan yaitu 2, 3, dan 5.

Tabel 8. Hasil Pengujian CNN dengan 3 Kelas Sentimen

Network	Filter	Recall	Precision	Akurasi	F1-Score
Simple CNN	2	90%	90%	91%	90%
	3	88%	88%	89%	88%
	5	88%	89%	90%	89%
DoubleMax CNN	2	80%	79%	78%	78%
	3	62%	77%	70%	62%
	5	91%	90%	91%	90%

Dari hasil pengujian di tabel 8 didapatkan tren yang sama dengan hasil pengujian pada Tabel 7, dimana *DoubleMax CNN* dengan *filter 5* mendapatkan hasil pengujian terbaik. Namun berbeda dengan saat klasifikasi 2 kelas sentimen, dapat diketahui bahwa hasil klasifikasi 3 kelas sentimen terbaik dengan menggunakan *Simple CNN* didapatkan dengan menggunakan *filter* ukuran 2 yang mencapai *Recall*, *Precision* dan *F1-Score* 90%, dengan akurasi 91%. Walau hasil pengujian ketika menggunakan *filter* ukuran yang lebih besar yaitu 2 dan 3 tidak terlalu signifikan. Pada *DoubleMax CNN*, sama seperti saat mengklasifikasi 2 kelas sentimen, hasil terburuk didapatkan pada saat penggunaan *filter* ukuran 3 dengan *Recall* 62%, *Precision* 77%, *F1-Score* 62% dan Akurasi 70%. Penggunaan ukuran *filter* tampak sangat berpengaruh pada *DoubleMax CNN* yang ditunjukkan dengan perbedaan hasil pengujian *Recall*, *Precision*, *F1-Score* dan akurasi yang besar antara penggunaan *filter* ukuran 2, 3 dan 5.

Tabel 9 menjelaskan perbandingan hasil klasifikasi sentimen dengan menggunakan *Support Vector Machine (SVM)*, *Logistic Regression*, *Naïve Bayes* dan *CNN*.

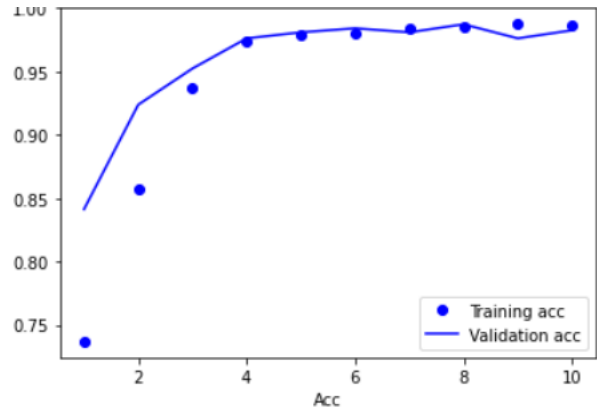
Tabel 9. Perbandingan Hasil Klasifikasi Sentimen berbagai Metode

Classifier	2 Kelas Sentimen		3 Kelas Sentimen	
	Akurasi	F1-Score	Akurasi	F1-Score
SVM	94%	93%	85%	84%
Naïve Bayes	61%	61%	46%	45%
Logistic Regression	93%	92%	85%	84%
Simple CNN	97%	97%	91%	90%
DoubleMax CNN	98%	98%	91%	90%

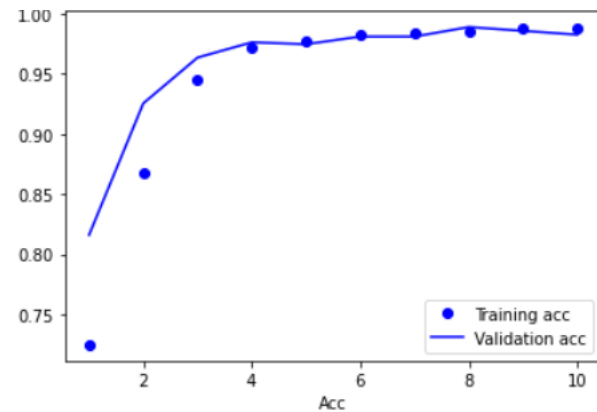
Dari hasil perbandingan di tabel 9, dapat diketahui bahwa *CNN*, baik *Simple CNN* maupun *DoubleMax CNN* memiliki performa terbaik, dari segi Akurasi, *Recall*, *Precision* dan *F1-Score*, pada saat klasifikasi 2 dan 3 kelas sentimen, jika dibandingkan dengan 3 metode lainnya, yaitu *SVM*, *Naïve Bayes*, dan *Logistic Regression* dengan *TF-IDF* sebagai representasi teks. *DoubleMax CNN* berhasil mencapai Akurasi dan *F1-Score* tertinggi sebesar 98% pada klasifikasi 2 kelas sentimen. Pada klasifikasi 3 kelas sentimen, *DoubleMax CNN* dan *Simple CNN* menunjukkan performa yang sama dengan Akurasi dan *F1-Score* terbaik mencapai 91% dan 90%. Sementara itu, di antara metode non *CNN*, *SVM* memiliki hasil terbaik, yang disusul oleh *Logistic Regression*. Hasil

terendah ditunjukkan oleh metode *Naïve Bayes* mencapai akurasi terendah mencapai 46% pada klasifikasi 3 kelas sentimen.

Sementara itu, Gambar 4 dan 5 menjelaskan perkembangan akurasi yang didapat pada setiap *epoch* pada saat *training* dan *evaluation Simple CNN* dan *DoubleMax CNN*.



Gambar 4. Akurasi pada saat *Training* dan *Evaluation* di setiap *Epoch* dengan menggunakan *Simple CNN*



Gambar 5. Akurasi pada saat *Training* dan *Evaluation* di setiap *Epoch* dengan menggunakan *DoubleMax CNN*

Berdasarkan Gambar 4 dan 5 dapat disimpulkan bahwa *CNN model* yang digunakan bekerja dengan baik, karena tidak terdapat perbedaan yang besar antara akurasi pada data *training* dan data *evaluation* di setiap *epoch*. Akurasi meningkat secara stabil mulai dari epoch 1 sampai dengan epoch 4. Namun, mulai dari epoch 5 sampai dengan 10, akurasi tidak mengalami peningkatan yang signifikan dan cenderung stagnan.

Berdasarkan beberapa pengujian yang telah dilakukan di atas, dapat disimpulkan bahwa *CNN*, baik *Simple CNN* maupun *DoubleMax CNN* dapat bekerja dengan sangat baik dalam melakukan analisis sentimen data saran mahasiswa terhadap kinerja departemen di perguruan tinggi. Walau data saran mahasiswa yang digunakan pada penelitian ini dapat berisi kalimat yang sangat panjang, yakni mencapai 150 kata, *CNN* dapat menghasilkan representasi kalimat yang sangat baik, sehingga performa analisis sentimen juga menjadi lebih baik. *Simple CNN* berhasil mencapai akurasi sebesar 97% dan 91%, masing-masing untuk 2 dan 3 kelas sentimen. Sedangkan *DoubleMax CNN* mendapatkan



akurasi yang lebih baik dibandingkan *Simple CNN*, yaitu akurasi klasifikasi sentimen yang mencapai 98% dan 91%, masing-masing untuk 2 dan 3 kelas sentimen.

#### 4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan dapat disimpulkan bahwa *DoubleMax CNN* dengan *filter* ukuran 5 memiliki hasil terbaik pada klasifikasi 2 dan 3 kelas sentimen. Pada klasifikasi 2 kelas sentimen, *DoubleMax CNN* mendapatkan Akurasi dan *F1-Score* sebesar 98%. Pada klasifikasi 3 kelas sentimen, *DoubleMax CNN* mendapatkan Akurasi dan *F1-Score* sebesar 91% dan 90%. Dibandingkan dengan metode klasik *machine learning* seperti *SVM*, *Logistic Regression* dan *Naïve Bayes*, *DoubleMax CNN* bekerja rata-rata 17% lebih baik dalam melakukan klasifikasi sentimen data saran. Hal ini menunjukkan bahwa *CNN* dengan arsitektur yang lebih kompleks memiliki performa yang lebih baik dalam menganalisis sentimen data saran mahasiswa, bahkan untuk kalimat yang panjang. Kemudian, dapat pula disimpulkan bahwa pada saat menggunakan *CNN* dengan penerapan *Maxpooling* di masing-masing layer, pemilihan *filter* yang tepat menjadi sangat penting. Penerapan filter pada teks sama seperti penerapan *N-Gram*, sehingga pemilihan *filter* yang tepat akan menyebabkan representasi teks yang baik pula. Pada penelitian ini, penggunaan *filter* ukuran 5 pada arsitektur *DoubleMax CNN* merupakan ukuran *filter* yang paling tepat dibandingkan dengan *filter* ukuran 2 dan 3.

Pada penelitian mendatang, *CNN* dengan arsitektur yang mirip akan digunakan untuk melakukan *multiclass classification* pada data saran, yaitu mengklasifikasi data saran berdasarkan emosi yang tertuang pada data saran menjadi bahagia/senang, marah, jijik, netral dan sedih. Untuk dapat melakukan penelitian mendatang, hal yang perlu dilakukan adalah mengumpulkan lebih banyak data saran dari tahun akademik terbaru dan melakukan *manual labelling* karena belum terdapat metode untuk melakukan *automatic labelling* untuk memberikan label emosi pada teks.

#### DAFTAR PUSTAKA

- AL KABIR, A.H., BASUKI, S. & WICAKSONO, G.W., 2019. Analisis sentimen kritik dan saran pelatihan aplikasi teknologi informasi (PATI) menggunakan algoritma support vector machine (SVM). *Jurnal Repositor*, 1(1), hal.39.
- ATTABI, A.W., MUFLIKHAH, L. & FAUZI, M.A., 2018. Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), hal.4548–4554.
- DOZAT, T., 2016. Incorporating Nesterov Momentum into Adam. *ICLR Workshop*, (1), hal.2013–2016.
- GUPTA, P., TIWARI, R. & ROBERT, N., 2016. Sentiment analysis and text summarization of online reviews: A survey. In: *International Conference on Communication and Signal Processing, ICCSP 2016*, hal.241–245.
- HAMZAH, A., 2014. Sentiment Analysis Untuk Memanfaatkan Saran Kuesioner Dalam Evaluasi Pembelajaran Dengan Menggunakan Naive Bayes Classifier (NBC). In: *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, hal.159.
- JUWIANTHO, H., SETIAWAN, E.I., SANTOSO, J. & PURNOMO, M.H., 2020. Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2Vec Menggunakan Deep Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(1), hal.181–188.
- KIM, H. & JEONG, Y.S., 2019. Sentiment classification using Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 9(11), hal.1–14.
- KIM, Y., 2014. Convolutional neural networks for sentence classification. In: *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. hal.1746–1751.
- KINGMA, D.P. & BA, J.L., 2015. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, hal.1–15.
- LIU, B., 2010. Sentiment Analysis and Subjectivity. In: N. Indurkha & F. J. Damerou (ed.), *Handbook of Natural Language Processing*, Chapman & Hall/CRC, hal.272–277.
- MIKOLOV, T., CHEN, K., CORRADO, G. & DEAN, J., 2013. Efficient estimation of word representations in vector space. In: *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, hal.1–12.
- PANG, B. & LEE, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), hal.1-135.
- RAUT, V.B. & LONDHE, D.D., 2014. Opinion mining and summarization of hotel reviews. In: *6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*, hal.556–559.
- RUDER, S., 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. [online] Tersedia di: <<https://arxiv.org/abs/1609.04747>>
- SALAM, N.S.N., SUPIANTO, A.A. & PERDANAKUSUMA, A.R., 2019. Analisis

Sentimen Opini Mahasiswa Terhadap Saran Kuesioner Penilaian Kinerja Dosen dengan Menggunakan TF-IDF dan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, [online] 3(6), hal.6148–6156. Tersedia di: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5649>>.

- SEYEDITABARI, A., TABARI, N. & ZADROZNY, W., 2018. Emotion Detection in Text: a Review. *arXiv preprint arXiv: 1806.00674*. [online] Tersedia di: <<https://arxiv.org/abs/1806.00674>>
- TSYTSARAU, M. & PALPANAS, T., 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), hal.478–514.
- WAHID, D.H. & SN, A., 2016. Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), hal.207.
- WICAKSONO, A.F., VANIA, C., DISTIAWAN, B.T. & ADRIANI, M., 2014. Automatically building a corpus for sentiment analysis on Indonesian tweets. In: *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation, PACLIC 2014*. hal.185–194.
- ZEILER, M.D., 2014. Adadelata: An Adaptive Learning Rate Method. *arXiv preprint arXiv: 1212.5701*. [online] Tersedia di: <<https://arxiv.org/abs/1212.5701>>