

# Analisis Sentimen terhadap Ulasan Aplikasi Bank Digital Syariah di Indonesia dengan Menggunakan Metode *Support Vector Machine* dan *Convolutional Neural Network*

Fatimatuz Zahro, R. Mohamad Atok, dan Prilyandari Dina Saputri  
 Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)  
*e-mail:* moh\_atok@its.ac.id

**Abstrak**—Teknologi dan digitalisasi beberapa tahun terakhir mengalami perkembangan sangat pesat. Di Indonesia sendiri, hingga pada periode 2021-2022 pengguna teknologi internet mencapai 210,03 juta. Teknologi memudahkan sebagian besar kebutuhan manusia dalam menjalankan pekerjaan dan aktivitas, termasuk dalam hal transaksi keuangan. Salah satunya dengan hadirnya Bank Digital. Layanan perbankan berbasis digital tak hanya dilakukan oleh bank konvensional namun juga hadirnya beberapa bank digital syariah yang beroperasi di Indonesia. Pada akhir 2021, Bank Aladin Syariah muncul sebagai bank digital murni syariah pertama di Indonesia. Sebagai *pioneer* bank digital syariah, sangat penting bagi Bank Aladin untuk mendapatkan reputasi yang baik sehingga evaluasi secara berkala diperlukan untuk menjaga kualitas layanan. Ulasan pengguna terhadap Bank Aladin dapat dijadikan sebagai alat efektif dan efisien dalam menemukan informasi. Oleh karena itu, perlu dilakukan analisis sentimen terhadap komentar pada layanan aplikasi. Analisis sentimen Ulasan pengguna terhadap Bank Aladin akan dikategorikan menjadi sentimen positif dan negatif. Klasifikasi awal dalam dua kategori sentimen ini akan ditentukan dengan menggunakan pelabelan berdasarkan *rating* ulasan pengguna. Hasil klasifikasi awal akan dibandingkan dengan menggunakan metode *Support Vector Machine* dan *Convolutional Neural Network* dengan ekstraksi fitur *Word2Vec*. Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari kumpulan ulasan atau *review* pengguna layanan aplikasi Bank Aladin di *Google Play Store* pada periode 1 Oktober 2022 hingga 31 Desember 2022. Hasil penelitian menunjukkan bahwa metode *Convolutional Neural Network* lebih baik dibandingkan *Support Vector Machine*. Pada metode *Support Vector Machine* nilai akurasi paling baik didapatkan pada parameter  $C = 1000$  dan  $\gamma = 10$  yaitu sebesar 81,9%. Sedangkan pada metode *Convolutional Neural Network* model terbaik yang didapatkan menggunakan kombinasi layer yaitu *Embedding Layer* dengan bobot *Word2Vec*, *Max Pooling 1D*, *Layer 1* dengan *node 8*, dan *Output layer* dengan fungsi aktivasi *sigmoid* dengan kebaikan model data *testing* didapatkan *accuracy* yang baik sebesar 89,8%.

**Kata Kunci**—Analisis Sentimen, Bank Aladin, *Convolutional Neural Network*, *Support Vector Machine*, *Word2Vec*.

## I. PENDAHULUAN

BEBERAPA tahun terakhir, teknologi telah berkembang pesat di Indonesia, terutama dalam hal penggunaan internet. Jumlah pengguna internet terus meningkat dari tahun ke tahun, mencapai 210,03 juta pengguna pada periode 2021-2022, dengan tingkat penetrasi internet sebesar 77,02%. Perkembangan ini juga didukung oleh digitalisasi yang semakin maju, termasuk dalam sektor perbankan.

Bank digital telah mengalami perkembangan yang signifikan di Indonesia. Pada tahun 2022, nilai transaksi bank

Tabel 1.  
*Confusion matrix*

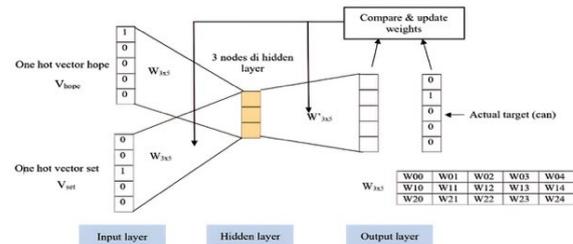
| Kelas Aktual | Kelas Prediksi               |                              |
|--------------|------------------------------|------------------------------|
|              | Positif                      | Negatif                      |
| Positif      | TP ( <i>True Positive</i> )  | FN ( <i>False Negative</i> ) |
| Negatif      | FP ( <i>False Positive</i> ) | TN ( <i>True Negative</i> )  |

Tabel 2.  
Contoh proses *cleaning*

| Sebelum   | Sesudah <i>Cleaning</i>   |
|---|---|
| Keren deh aplikasi Aladin, prosesnya lancar dan cepat, aplikasinya sangat mudah di pahami | Keren deh aplikasi Aladin prosesnya lancar dan cepat aplikasinya sangat mudah di pahami |

Tabel 3.  
Contoh proses *folding*

| Sebelum   | Sesudah <i>Cleaning</i>   |
|---|---|
| Keren deh aplikasi Aladin prosesnya lancar dan cepat aplikasinya sangat mudah di pahami | keren deh aplikasi aladin prosesnya lancar dan cepat aplikasinya sangat mudah di pahami |



Gambar 1. Ilustrasi CBOW pada *word2vec*.

digital meningkat 28,72% menjadi Rp52.545,8 triliun, dan proyeksi nilai transaksi di tahun 2023 mencapai Rp64.175,1 triliun. Bank digital sendiri merupakan bank yang beroperasi melalui saluran elektronik tanpa kantor fisik, dan semakin diminati oleh masyarakat Indonesia. Selain bank konvensional, juga ada bank digital syariah yang beroperasi di Indonesia, seperti Aladin, Jago Syariah, dan Hijrah. Bank Aladin Syariah merupakan bank digital syariah pertama di Indonesia dan berhasil meraih 1,1 juta nasabah pada tahun 2022.

Meskipun literasi keuangan syariah meningkat, pangsa pasar keuangan syariah masih rendah dibandingkan dengan keuangan konvensional. Hal ini menimbulkan pertanyaan mengingat mayoritas penduduk Indonesia adalah Muslim. Oleh karena itu, perlu inovasi produk dan layanan keuangan syariah untuk menjangkau lebih banyak masyarakat. Sebagai bank digital syariah *pioneer*, Bank Aladin perlu

Tabel 4.  
Contoh proses *stemming*

| Sebelum   | Sesudah <i>Cleaning</i>   |
|---|---|
| keren deh aplikasi aladin prosesnya lancar dan cepat aplikasinya sangat mudah di pahami | keren deh aplikasi aladin proses lancar dan cepat aplikasi sangat mudah di pahami |

Tabel 5.  
Contoh proses *stopword removal*

| Sebelum  | Sesudah <i>Cleaning</i>               |
|--|---------------------------------------|
| keren deh aplikasi aladin proses lancar dan cepat aplikasinya sangat mudah di pahami | keren proses lancar cepat mudah paham |

Tabel 6.  
Contoh proses *tokenization*

| Sebelum                               | Sesudah <i>Cleaning</i>                                  |
|---------------------------------------|--|
| keren proses lancar cepat mudah paham | ['keren', 'proses', 'lancar', 'cepat', 'mudah', 'paham'] |

Tabel 7.  
Jumlah data sebelum dan sesudah *random oversampling*

| <i>Random Oversampling</i> | Data     | Kelas Sentimen |              | Total        |
|----------------------------|----------|----------------|--------------|--------------|
|                            |          | Positif        | Negatif      |              |
| Sebelum                    | Training | 3.336          | 802          | <b>4.138</b> |
|                            | Testing  | 834            | 201          | <b>1.035</b> |
| <b>Total</b>               |          | <b>4.170</b>   | <b>1.003</b> | <b>5.173</b> |
| Sesudah                    | Training | 3.336          | 3.336        | <b>6.672</b> |
|                            | Testing  | 834            | 201          | <b>1.035</b> |
| <b>Total</b>               |          | <b>4.170</b>   | <b>3.537</b> | <b>7.707</b> |

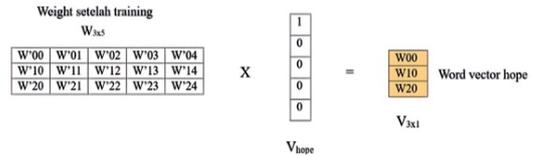
Tabel 8.  
Hasil uji f pada model PCR

| No    | Ulasan (x)  | Klasifikasi Sentimen (y) |
|-------|---|--------------------------|
| 1     | 'mudah', 'nyaman', 'guna', 'cocok', 'buat', 'males', 'buat', 'keluar', 'rumah', 'sukany', 'rebahn', 'jdi', 'apaapa', 'rum', 'terimakasih' | Positif                  |
| ⋮     | ⋮   | ⋮                        |
| 6.672 | 'pinjam', 'online'  | Negatif                  |

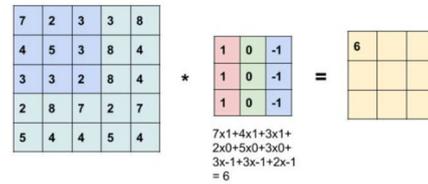
mempertahankan reputasinya dan meningkatkan kualitas layanannya. Analisis sentimen digunakan untuk memahami pendapat pengguna terhadap aplikasi Bank Aladin. Analisis sentimen melibatkan teknik menganalisis sentimen dan pendapat seseorang terhadap suatu topik, produk, atau layanan.

Sejumlah penelitian sebelumnya telah menggunakan metode seperti *Support Vector Machine* dan *Convolutional Neural Network* dalam analisis sentimen. Penelitian tersebut melakukan analisis sentimen terkait pengaruh panjang kalimat pada klasifikasi suatu komentar terhadap kedua metode. Diperoleh hasil bahwa Algoritma *Convolutional Neural Network* dikombinasikan dengan ekstraksi fitur *Word2Vec* memberikan akurasi terbaik dalam penelitian ini dibandingkan *Support Vector Machine* yang dikombinasikan dengan ekstraksi fitur TF-IDF dan *Support Vector Machine* yang dikombinasikan dengan ekstraksi fitur *Word2Vec*. Hasil penelitian menunjukkan bahwa *Support Vector Machine* memiliki akurasi yang tinggi, sedangkan *Convolutional Neural Network* memiliki akurasi lebih tinggi dalam analisis sentimen dibandingkan metode *deep neural network* lainnya [1].

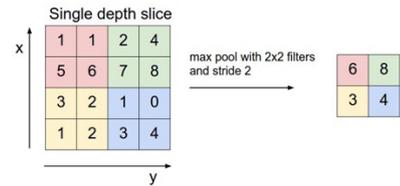
Penulis akan melakukan penelitian analisis sentimen



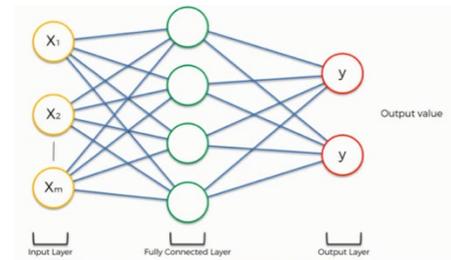
Gambar 2. Pembuatan *word vector* setiap kata.



Gambar 3. Ilustrasi *convolutional layer*.



Gambar 4. Ilustrasi *pooling layer*.



Gambar 5. Ilustrasi *fully connected layer*.

menggunakan data ulasan pengguna di *Google Play Store* tentang Bank Aladin. Tujuan penelitian ini adalah untuk membandingkan metode *Support Vector Machine* dan *Convolutional Neural Network* dalam analisis sentimen, dengan menggunakan ekstraksi fitur *Word2Vec*. Penelitian sebelumnya menunjukkan bahwa metode ini memberikan hasil akurasi yang baik.

## II. TINJAUAN PUSTAKA

### A. Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi dan mengklasifikasikan pendapat dan perasaan seseorang terhadap suatu topik. Ini dilakukan melalui pendekatan seperti algoritma kamus (*dictionary based* dan *corpus based*) atau *machine learning* (*supervised* dan *unsupervised learning*) [2].

### B. Text Mining

*Text mining* adalah bidang dalam *data mining* yang fokus pada analisis teks, di mana data yang ditambang adalah teks dari dokumen. *Text mining* dapat mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta mengidentifikasi kesamaan antara dokumen untuk mengetahui hubungannya dengan variabel lain.

### C. Text Preprocessing

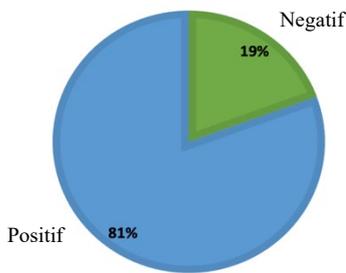
*Text Preprocessing* adalah tahapan awal dalam *text mining* untuk mengubah teks asli dengan menghilangkan unsur

Tabel 9.  
Struktur data setelah *sequence dan padding*

| Ulasan ke- <i>n</i> | Kata Ke-1 | Kata Ke-2 | ... | Kata Ke-53 | Klasifikasi Sentimen ( <i>y</i> ) |
|---------------------|-----------|-----------|-----|------------|-----------------------------------|
| 1                   | 2         | 69        | ... | 0          | 0                                 |
| ⋮                   | ⋮         | ⋮         | ⋮   | ⋮          | ⋮                                 |
| 6.672               | 141       | 158       | ... | 0          | 1                                 |

Tabel 10.  
Struktur data setelah *word2vec*

| No       | Vektor Kata ke-1   | ... | Vektor Kata ke-53  | Klasifikasi Sentimen ( <i>y</i> ) |
|----------|--|-----|--|-----------------------------------|
| 1        | [ <i>x</i> <sub>1,1,1</sub> ... <i>x</i> <sub>1,1,100</sub> ]                | ... | [ <i>x</i> <sub>53,1,1</sub> ... <i>x</i> <sub>53,1,100</sub> ]                | <i>y</i> <sub>1</sub>             |
| ⋮        | ⋮  | ⋮   | ⋮  | ⋮                                 |
| <i>n</i> | [ <i>x</i> <sub>1,<i>n</i></sub> ,1 ... <i>x</i> <sub>1,<i>n</i></sub> ,100] | ... | [ <i>x</i> <sub>53,<i>n</i></sub> ,1 ... <i>x</i> <sub>53,<i>n</i></sub> ,100] | <i>y</i> <sub><i>n</i></sub>      |



Gambar 8. *Pie chart* sentimen positif dan negatif.

tekstual yang tidak berguna untuk meningkatkan akurasi dalam klasifikasi teks. Proses ini melibatkan *Cleaning* (pembersihan tanda baca dan karakter tidak diperlukan), *Case Folding* (menyeragamkan karakter menjadi huruf kecil), *Stemming* (mencari kata dasar dari kata yang sesuai dengan struktur Bahasa Indonesia), *Stopword Removal* (menghilangkan kata umum yang tidak penting), dan *Tokenization* (memotong teks menjadi kata-kata).

D. *Word Cloud*

*Word Cloud* adalah metode visualisasi yang menggambarkan frekuensi kata dalam sebuah dokumen dengan ukuran kata yang menunjukkan tingkat keseringan kata tersebut [3]. Semakin sering sebuah kata muncul, semakin besar pula ukuran kata tersebut dalam *Word Cloud*. Warna latar belakang, ukuran *font* minimum dan maksimum, serta ukuran lebar dan tinggi output dapat disesuaikan dalam *Word Cloud*.

E. *Stratified Holdout Validation*

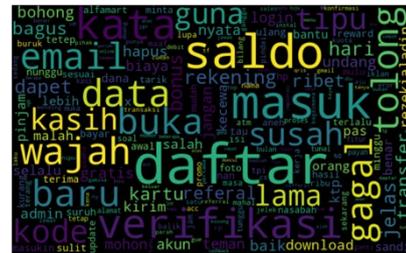
*Stratified random sampling* merupakan metode mengambil sampel dengan memisahkan elemen populasi ke dalam beberapa kelompok yang tidak saling tumpang tindih (*strata*), dari setiap kelompok dilakukan pengambilan sampel secara acak [4]. *Stratified holdout validation* adalah metode membagi *dataset* menjadi dua bagian yaitu data *train* digunakan untuk membangun suatu model *machine learning* dan data *test* digunakan untuk melihat seberapa baik model yang telah dibangun dengan proporsi yang sama dari setiap kelompok yang ada, pembagian proporsi yang paling sering digunakan adalah 80% untuk data *train* dan 20% untuk data *test*.

F. *Random Oversampling*

Tidak seimbangnya *dataset* dapat menyebabkan klasifikasi yang dihasilkan dari proses pengklasifikasian pada data minor menjadi kurang tepat karena tertutupi oleh data mayor,



Gambar 6. Sentimen positif.



Gambar 7. Sentimen negatif.

perlu dilakukan pengendalian *dataset* tidak seimbang agar rasio ketimpangan data dapat dikurangi [5]. *Random oversampling* merupakan metode untuk mengatasi *dataset* yang tidak seimbang dengan cara menambahkan data secara acak dari kelas minoritas ke dalam data *training*. Proses penambahan data minor akan diulang hingga jumlah data pada kelas minor menyamai jumlah data pada kelas mayor.

G. *Word2Vec*

*Word2Vec* adalah salah satu metode *embedding word* yang berguna untuk merepresentasikan kata menjadi sebuah vektor. Vektor tersebut tidak hanya merepresentasikan kata secara sintatik tapi juga secara semantik atau secara makna. *Cosine similarity* digunakan untuk mengukur kesamaan antara dua vektor.

$$Similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \tag{1}$$

di mana *A* adalah vektor *word embedding* dari kata ke-1 dan *B* adalah vektor *word embedding* dari kata ke-2.

Terdapat dua jenis *Word2Vec* yaitu skip-gram dan CBOW (*Continuous Bag of Words*). Gambar 1 merupakan ilustrasi dari metode CBOW dalam *Word2Vec*. Nilai *weight* matriks ideal yang dimiliki akan dikalikan dengan vektor *one-hot vector* setiap kata yang dimiliki, maka akan menghasilkan *word vector* dari setiap kata seperti pada Gambar 2.

H. *Support Vector Machine*

*Support Vector Machine* (SVM) adalah teknik prediksi yang dapat digunakan dalam klasifikasi maupun regresi [6]. SVM memiliki prinsip dasar sebagai *linier classifier*, yang berarti dapat memisahkan kelas-kelas data secara linier. Namun, SVM juga dapat bekerja pada problem non-linier dengan menggunakan konsep kernel untuk mengubah ruang kerja menjadi berdimensi tinggi. Dalam ruang berdimensi tinggi, SVM mencari *hyperplane* yang dapat memaksimalkan jarak antara kelas-kelas data, yang disebut *margin*. Pada SVM, kedua kelas data dipisahkan oleh bidang pembatas yang sejajar atau linear. Data yang berada pada bidang pembatas tersebut disebut *support vector*. Persamaan *hyperplane* klasifikasi linier SVM dituliskan sebagai berikut:

$$f(x) = w^T x + b \tag{2}$$

Tabel 11.  
Hasil akurasi peramalan dari tiap saham

| No   | Vektor Kata Ke-1 |     |              | ... | Vektor Kata Ke-53 |     |     | Klasifikasi Sentimen (y) |
|------|------------------|-----|--------------|-----|-------------------|-----|-----|--------------------------|
|      | 1                | ... | 100          |     | 1                 | ... | 100 |                          |
| 1    | 0.00467194       | ... | -0.000718026 | ... | 0                 | ... | 0   | 0                        |
| ⋮    | ⋮                | ⋮   | ⋮            | ⋮   | ⋮                 | ⋮   | ⋮   | ⋮                        |
| 6672 | -0.001085056     | ... | 0.0077961013 | ... | 0                 | ... | 0   | 1                        |

dimana  $w$  adalah vektor bobot,  $x$  adalah data *training*, dan  $b$  adalah bias atau konstanta.

Pada SVM (*Support Vector Machine*) yang digunakan untuk memisahkan data non-linear (yang tidak dapat dipisahkan secara linear), SVM menggunakan fungsi kernel yang dikenal sebagai "*Kernel Trick*". Beberapa fungsi kernel yang umum digunakan dalam SVM adalah Kernel Linear, Kernel Polynomial, Kernel Sigmoid (*Tangent Hyperbolic*), dan Kernel RBF. Adapun fungsi kernel yang digunakan dalam penelitian ini adalah fungsi kernel RBF. Bentuk kernel dapat ditunjukkan pada persamaan (3) berikut [7].

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (3)$$

Kernel tersebut menggunakan parameter  $\gamma$  dan  $C$ , di mana  $C$  merupakan parameter untuk menentukan besar penalti akibat kesalahan data dan nilainya ditentukan oleh pengguna, sehingga parameter  $C$  memiliki peran untuk meminimalkan kesalahan pelatihan dan mengurangi kompleksitas dari model. Sementara  $\gamma$  (gamma) merupakan percepatan fungsi kernel untuk memperoleh nilai klasifikasi model optimal.

### I. Convolutional Neural Network

*Convolutional Neural Network* (CNN) merupakan algoritma *deep learning* yang umumnya menggunakan gambar sebagai *input*, menetapkan kepentingan (bobot dan bias yang dapat dipelajari) untuk berbagai aspek dan objek dalam gambar agar dapat membedakan satu dengan yang lain. Pada proses pengolahan terdapat tiga tahapan layer yaitu *convolutional layer*, *pooling layer*, dan *fully connected layer* [8].

#### 1) Embedding Layer

*Embedding layer* memungkinkan kita mengubah setiap kata menjadi vektor dengan panjang tertentu sebagai representasi dari setiap kata. Persamaan pada *embedding layer* adalah sebagai berikut:

$$x_i = o \times W \quad (4)$$

dimana  $o$  merupakan vektor hasil *one-hot encoding* dari setiap kata pada ulasan dan  $W$  adalah matriks bobot *Word2Vec*.

#### 2) Convolutional Layer

Data yang masuk pada *convolutional layer* akan mengalami proses konvolusi, lapisan akan mengkonversi setiap filter ke seluruh data dan akan menghasilkan *activation map* atau *feature map 2D*. Ilustrasi *convolutional layer* ditunjukkan pada Gambar 3. Jika input memiliki panjang *feature map*  $N_h$  dan lebar  $N_w$  serta kernel dengan panjang  $C$  dan lebar  $D$  maka dimensi output yang dihasilkan dapat dihitung dengan persamaan  $(N_h - C + 1) \times (N_w - D + 1)$ . Pada Gambar 3 ukuran *input feature map* adalah  $N_h = N_w = 5$ , serta ukuran kernel bobot adalah  $C = D = 3$ , sehingga dengan

menggunakan persamaan tersebut didapatkan ukuran output sebesar  $3 \times 3$ .

#### 3) Pooling Layer

Tahapan setelah *convolutional layer* adalah *pooling layer*, *pooling layer* terdiri dari sebuah filter dengan ukuran dan *stride* tertentu. Setiap pergeseran akan ditentukan oleh jumlah *stride* yang akan digeser pada seluruh area *feature map* atau *activation map*. Dalam penerapannya terdapat dua jenis *pooling* yang biasa digunakan, yaitu *Max Pooling* dengan mengambil nilai maksimal pada area tertentu, dan *Average Pooling* dengan mengambil nilai rata-rata pada area tertentu. Ilustrasi *pooling layer* ditunjukkan pada Gambar 4.

#### 4) Fully Connected Layer

Ilustrasi *fully connected layer* ditunjukkan pada Gambar 5. *Fully connected layer* memiliki beberapa fungsi aktivasi, untuk klasifikasi dua kelas, fungsi aktivasi yang digunakan pada *output layer* adalah *sigmoid* karena memiliki *range* nilai 0 hingga 1. Fungsi terakhir adalah *loss function* untuk menghitung *loss* (nilai *error*) dengan menggunakan *binary cross-entropy*. Kemudian terdapat optimasi parameter pada layer CNN dilakukan dengan cara memperbaharui parameter seperti bobot dan bias dengan proses *backpropagation* yang bertujuan untuk meminimumkan nilai *loss function* menggunakan metode *gradient descent* ADAM.

### J. Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi dilakukan untuk mengevaluasi performa klasifikasi. *Confusion matrix* adalah suatu tabel yang digunakan untuk visualisasi hasil klasifikasi pada mesin pembelajaran yang melibatkan dua kategori atau lebih [9]. Tabel *confusion matrix* disajikan pada Tabel 1.

Pengukuran performa klasifikasi pada model data *balanced* yang sering digunakan untuk menghitung ketepatan klasifikasi adalah *accuracy*, *precision*, dan *recall* (*sensitivity*).

$$\begin{aligned} \text{Accuracy} &: \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Recall} &: \frac{TP}{TP+FN} \\ \text{Precision} &: \frac{TP}{TP+FP} \end{aligned}$$

di mana  $TP$  adalah *True Positive* (jumlah prediksi benar pada kelas positif)  $FP$  adalah *False Positive* (jumlah prediksi salah pada kelas positif)  $FN$  adalah *False Negative* (jumlah prediksi salah pada kelas negatif), dan  $TN$  adalah *True Negative* (jumlah prediksi benar pada kelas negatif). kedalam nilai positif.

### K. Bank Aladin

PT Bank Aladin Syariah Tbk merupakan *pioneer* bank murni digital syariah di Indonesia yang disahkan 7 April 2021. Bank Aladin merupakan merek yang ramah di telinga publik dan berarti "dengan atau di atas" dan "*way of life* atau *faith*" dalam suku kata. Bank Aladin memiliki fokus pada

Tabel 16.  
Akurasi data training dan testing untuk penentuan parameter terbaik SVM

|               | C     | $\gamma$ |       |       |       |       |       |       |
|---------------|-------|----------|-------|-------|-------|-------|-------|-------|
|               |       | 0,001    | 0,01  | 0,1   | 1     | 10    | 100   | 1000  |
| Data Training | 0,001 | 0,704    | 0,704 | 0,705 | 0,710 | 0,661 | 0,539 | 0,526 |
|               | 0,01  | 0,704    | 0,704 | 0,705 | 0,710 | 0,661 | 0,539 | 0,526 |
|               | 0,1   | 0,704    | 0,704 | 0,705 | 0,703 | 0,818 | 0,570 | 0,552 |
|               | 1     | 0,704    | 0,704 | 0,704 | 0,793 | 0,942 | 0,983 | 0,984 |
|               | 10    | 0,704    | 0,704 | 0,787 | 0,884 | 0,977 | 0,984 | 0,984 |
|               | 100   | 0,704    | 0,785 | 0,846 | 0,951 | 0,984 | 0,984 | 0,984 |
| Data Testing  | 1000  | 0,786    | 0,839 | 0,902 | 0,977 | 0,984 | 0,984 | 0,984 |
|               | 0,001 | 0,604    | 0,604 | 0,604 | 0,610 | 0,484 | 0,288 | 0,243 |
|               | 0,01  | 0,604    | 0,604 | 0,604 | 0,610 | 0,484 | 0,288 | 0,243 |
|               | 0,1   | 0,604    | 0,604 | 0,604 | 0,648 | 0,707 | 0,332 | 0,300 |
|               | 1     | 0,604    | 0,604 | 0,629 | 0,729 | 0,814 | 0,807 | 0,806 |
|               | 10    | 0,604    | 0,627 | 0,726 | 0,778 | 0,818 | 0,804 | 0,806 |
|               | 100   | 0,627    | 0,727 | 0,743 | 0,799 | 0,816 | 0,804 | 0,806 |
|               | 1000  | 0,727    | 0,741 | 0,771 | 0,808 | 0,819 | 0,804 | 0,806 |

Tabel 17.  
Confusion matrix data training SVM

|               | Kelas Aktual | Kelas Prediksi |         |
|---------------|--------------|----------------|---------|
|               |              | Positif        | Negatif |
| Data Training | Positif      | 3315           | 21      |
| Data Training | Negatif      | 83             | 3253    |
| Data Testing  | Positif      | 776            | 58      |
| Data Testing  | Negatif      | 129            | 72      |

Tabel 18.  
Ketepatan klasifikasi metode SVM

| Data     | Accuracy | Precision | Recall |
|----------|----------|-----------|--------|
| Training | 98,4%    | 97,6%     | 99,4%  |
| Testing  | 81,9%    | 85,7%     | 93%    |

Tabel 19.  
Contoh hasil prediksi metode SVM

| Ulasan  | Sentimen | Prediksi |
|---|----------|----------|
| Bank syariah digital yang sangat bagus dan modern memberikan kebutuhan yang sangat praktis di lengkapi dengan fitur fitur menarik                                 | Positif  | Positif  |
| Katanya login di aplikasi ini dapat saldo 15k ternyata Cuma tipuan semata agar kita tertarik krna saya sudah login tapi tidak ada saldo yang masuk. Cuma penipuan | Negatif  | Positif* |
| Susah pas daftar foto wajah ga bisa   | Negatif  | Negatif  |

digitalisasi dan inklusi keuangan, serta menyediakan fitur-fitur seperti tarik setor tunai melalui gerai Alfamart, penyaluran ZISWAF, dan sub-rekening untuk memudahkan nasabah.

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari kumpulan ulasan pengguna layanan aplikasi Bank Aladin di *Google Play Store*. Data tersebut diambil pada periode 1 Oktober 2022 hingga 31 Desember 2022 sebanyak 5173 data.

#### B. Langkah Analitis

Adapun langkah-langkah pada penelitian ini adalah sebagai berikut:

- Melakukan pengumpulan data ulasan layanan aplikasi Bank Aladin pada periode 1 Oktober 2022 hingga 31 Desember 2022 dengan menggunakan *web scraping* pada *software Python* untuk disimpan dalam *database CSV*.

Tabel 12.  
Confusion matrix dan ketepatan klasifikasi arsitektur model pertama

|               | Kelas Aktual | Kelas Prediksi |         | Precision | Recall |
|---------------|--------------|----------------|---------|-----------|--------|
|               |              | Positif        | Negatif |           |        |
| Data Training | Positif      | 3313           | 23      | 0,977     | 0,993  |
| Data Training | Negatif      | 78             | 3258    |           |        |
| Data Testing  | Positif      | 782            | 52      | 0,935     | 0,938  |
| Data Testing  | Negatif      | 54             | 147     |           |        |

Tabel 13.  
Confusion matrix dan ketepatan klasifikasi arsitektur model kedua

|               | Kelas Aktual | Kelas Prediksi |         | Precision | Recall |
|---------------|--------------|----------------|---------|-----------|--------|
|               |              | Positif        | Negatif |           |        |
| Data Training | Positif      | 3308           | 28      | 0,977     | 0,992  |
| Data Training | Negatif      | 78             | 3258    |           |        |
| Data Testing  | Positif      | 762            | 72      | 0,940     | 0,914  |
| Data Testing  | Negatif      | 48             | 153     |           |        |

Tabel 14.  
Confusion matrix dan ketepatan klasifikasi arsitektur model ketiga

|               | Kelas Aktual | Kelas Prediksi |         | Precision | Recall |
|---------------|--------------|----------------|---------|-----------|--------|
|               |              | Positif        | Negatif |           |        |
| Data Training | Positif      | 3297           | 39      | 0,978     | 0,988  |
| Data Training | Negatif      | 75             | 3261    |           |        |
| Data Testing  | Positif      | 762            | 72      | 0,950     | 0,914  |
| Data Testing  | Negatif      | 40             | 161     |           |        |

Tabel 15.  
Confusion matrix dan ketepatan klasifikasi arsitektur model keempat

|               | Kelas Aktual | Kelas Prediksi |         | Precision | Recall |
|---------------|--------------|----------------|---------|-----------|--------|
|               |              | Positif        | Negatif |           |        |
| Data Training | Positif      | 3283           | 53      | 0,980     | 0,984  |
| Data Training | Negatif      | 65             | 3271    |           |        |
| Data Testing  | Positif      | 761            | 73      | 0,944     | 0,912  |
| Data Testing  | Negatif      | 45             | 156     |           |        |

- Pemberian label sentimen ke dalam dua kategori yaitu kategori sentimen positif dan negatif berdasarkan *rating*. Di mana untuk *rating* 1, 2, dan 3 diklasifikasi negatif (nilai 1), *rating* 4 dan 5 diklasifikasi positif (nilai 0).
- Preprocessing* data, ada beberapa tahapan yang harus dilakukan pada proses ini, yaitu *Cleaning*, *Case folding*, *Stemming*, *Stopwords removal*, dan *Tokenization*.
- Melakukan visualisasi data dengan *word cloud*.
- Membagi data ulasan ke dalam data *training* dan *testing*. Data akan dibagi menjadi 80% data *training* dan 20% data *testing* menggunakan *stratified holdout validation*.
- Melakukan *random oversampling* pada data *training* apabila terdapat kelas sentimen *imbalance* yang bertujuan untuk menghindari bias pada ketepatan klasifikasi.
- Melakukan pembobotan kata dengan menggunakan algoritma *Word2Vec*.
- Melakukan analisis sentimen menggunakan SVM dengan mencoba beberapa parameter.
  - Membuat dan melatih model SVM menggunakan nilai parameter *C* dan  $\gamma$  pada kernel RBF dengan rentang nilai parameter  $C = (0,001; 0,01; 0,1; 1; 10; 100; 1000)$  dan nilai  $\gamma = (0,001; 0,01; 0,1; 1; 10; 100; 1000)$ .
  - Melakukan klasifikasi menggunakan metode SVM yang telah dilatih.

Tabel 20.

*Confusion matrix* dan ketepatan klasifikasi arsitektur model kelima

|          | Kelas Aktual | Kelas Prediksi |         | <i>Precision</i> | <i>Recall</i> |
|----------|--------------|----------------|---------|------------------|---------------|
|          |              | Positif        | Negatif |                  |               |
| Data     | Positif      | 3278           | 58      | 0,980            | 0,983         |
| Training | Negatif      | 65             | 3271    |                  |               |
| Data     | Positif      | 759            | 75      | 0,945            | 0,910         |
| Testing  | Negatif      | 44             | 157     |                  |               |

Tabel 21.

Contoh hasil prediksi metode CNN

| Ulasan   | Sentimen | Prediksi |
|--|----------|----------|
| Bank syariah digital yang sangat bagus dan modern memberikan kebutuhan yang sangat praktis di lengkapi dengan fitur fitur menarik  | Positif  | Positif  |
| Katanya login di aplikasi ini dapat saldo 15k ternyata Cuma tipuan semata agar kita tertarik krna saya sudah login tapi tidak ada saldo yang masuk. Cuma penipuan $\hat{O} \leq -\infty$ , | Negatif  | Negatif  |
| Susah pas daftar foto wajah ga bisa  | Negatif  | Positif* |

Tabel 22.

Perbandingan metode SVM dan CNN

| Metode | <i>Accuracy</i> | <i>Precision</i> | <i>Recall</i> |
|--------|-----------------|------------------|---------------|
| SVM    | 81,9%           | 85,7%            | 93%           |
| CNN    | 89,8%           | 93,5%            | 93,8%         |

- c. Menghitung ketepatan klasifikasi dari model SVM pada data *training*.
- d. Melakukan klasifikasi pada data *testing*.
- e. Menghitung ketepatan klasifikasi dari model SVM pada data *testing*.
9. Melakukan analisis sentimen menggunakan CNN dengan mencoba beberapa arsitektur.
  - a. Membangun model arsitektur CNN
  - b. Melakukan klasifikasi menggunakan metode CNN yang telah dibangun menggunakan aktivasi *sigmoid*.
  - c. Menghitung ketepatan klasifikasi dari model CNN pada data *training*.
  - d. Melakukan klasifikasi pada data *testing*.
  - e. Menghitung ketepatan klasifikasi dari model CNN pada data *testing*.
10. Membandingkan hasil ketepatan klasifikasi antara metode SVM dan CNN berdasarkan nilai ketepatan klasifikasi.
11. Melakukan interpretasi hasil analisis ketepatan klasifikasi.
12. Menarik kesimpulan berdasarkan analisis dan memberikan saran untuk pengembangan penelitian lebih lanjut.

#### IV. HASIL DAN PEMBAHASAN

##### A. Karakteristik Data

Dari 5173 data ulasan, diperoleh data dengan klasifikasi sentimen positif sebanyak 4170 data dan klasifikasi sentimen negatif sebanyak 1003 data. Gambar 6 merupakan perbandingan jumlah antar klasifikasi sentimen yang disajikan dalam bentuk diagram lingkaran.

##### B. Preprocessing Text

Ada beberapa tahapan atau langkah-langkah yang harus dilakukan pada *preprocessing text*.

##### 1) Cleaning

Pada langkah ini, *cleaning* data dilakukan dengan menghilangkan angka, tanda baca, simbol, dan emotikon. Penerapan dari proses *cleaning* ditunjukkan pada Tabel 2.

##### 2) Case Folding

Langkah ini dilakukan untuk mengubah semua huruf kapital menjadi *lowercase* yang ditunjukkan pada Tabel 3.

##### 3) Stemming

Langkah selanjutnya adalah stemming, di mana ulasan akan diubah menjadi kata dasar dengan menghilangkan awalan, akhiran, sisipan, dan *confixes* (awalan dan akhiran). Penerapan dari proses stemming ditunjukkan pada Tabel 4.

##### 4) Stopword Removal

*Stopword Removal* dilakukan untuk menghilangkan kata-kata umum dan sering muncul tetapi tidak memiliki pengaruh signifikan terhadap makna kalimat. Proses ini dilakukan dengan menggunakan kamus Sastrawi di *Python*. Beberapa kata di Sastrawi dihapus karena penulis menganggap kata tersebut penting untuk suatu kalimat yaitu “baik”, “lebih”, “tidak”, “bisa”, “guna”, “tolong”, “belum”, “dapat”, “ok”, “ada”, “tanpa”. Penulis juga menambahkan beberapa kata pada daftar *stopword* seperti kata “aplikasi”, “aladin”, “bank”, “digital” dan seterusnya karena dianggap tidak berpengaruh pada sentiment. Penerapan dari proses *Stopword removal* ditunjukkan pada Tabel 5.

##### 5) Tokenization

Pada langkah ini kalimat akan dipecah menjadi kata per kata untuk memudahkan proses analisis. Penerapan dari proses *tokenization* ditunjukkan pada Tabel 6.

##### C. Visualisasi Word Cloud

Ulasan divisualisasikan berdasarkan klasifikasi sentimen positif maupun negatif untuk melihat kata yang sering muncul dalam data ulasan setelah dilakukan *preprocessing text*. Gambar 7 menunjukkan *word cloud* dari sentimen positif.

Kata yang paling sering muncul pada kumpulan ulasan yang bersentimen positif di antaranya “mudah”. Kata “mudah” berhubungan dengan aplikasi Bank Aladin yang dinilai mudah digunakan dan dapat memudahkan transaksi online para pengguna, sehingga kata “guna” juga memiliki frekuensi kemunculan yang sering. Selanjutnya kata “aman” berhubungan dengan Bank Aladin yang sudah terdaftar di OJK sedangkan kata yang paling sering muncul pada kumpulan ulasan yang bersentimen negatif di antaranya “daftar”, “saldo”, “masuk.” Kata “daftar” menunjukkan adanya kendala yang dihadapi oleh pengguna aplikasi Bank Aladin saat melakukan pendaftaran dan juga banyaknya protes karena tidak mendapatkan bonus saldo saat daftar sesuai yang diberitakan. Untuk itu, kata “saldo” juga menempati frekuensi kata yang paling banyak muncul. Selanjutnya kata “masuk” juga berhubungan dengan pengguna yang sulit masuk akun karena proses verifikasi, serta berhubungan dengan protes karena bonus saldo tidak masuk. Gambar 8 menunjukkan *word cloud* dari sentimen negative.

##### D. Pembagian Data Training dan Testing

Setelah *preprocessing text*, data dibagi menjadi dua bagian yaitu *training* dan *testing* dengan proporsi 80:20. Pada

penelitian ini diterapkan metode *Random Oversampling* agar jumlah antar kelas menjadi seimbang. Data *training* dengan kelas negatif bertambah menyesuaikan dengan jumlah data *training* dengan kelas positif. Jumlah data sebelum dan sesudah *Random Oversampling* ditunjukkan oleh Tabel 7.

Adapun sebagian data ulasan layanan aplikasi Bank Aladin pada data *training* ditunjukkan oleh Tabel 8.

E. *Word Embedding*

Sebelum dilakukan *Word Embedding*, agar dapat dijadikan *input*, kita lakukan *sequence* dan *padding* dari struktur data awal yang sudah terbentuk pada Tabel 8. *Sequence* merupakan proses pemberian indeks kata. Indeks kata didasarkan pada jumlah frekuensi data *training* terbanyak, adapun untuk Indeks 1 dikhususkan terhadap kata-kata *out of vocabulary (OOV)* yaitu kata-kata yang tidak terdapat di kamus. Kemudian, indeks 2 akan ditempati oleh kata yang memiliki frekuensi paling banyak, indeks 3 akan ditempati oleh kata yang memiliki frekuensi paling banyak kedua, dan seterusnya.

Sementara itu *padding* merupakan proses membuat setiap *vector* kalimat memiliki panjang yang sama. Pada data *training* ulasan, ulasan dengan kalimat terpanjang memiliki jumlah 53. Misalkan suatu kalimat hanya memiliki panjang 20 kata, maka 33 kata di belakangnya akan diberikan nilai 0. Struktur data setelah dilakukan *sequence* dan *padding* ditunjukkan pada Tabel 9.

Selanjutnya dilakukan proses pembobotan kata hingga terbentuknya *vector* pada sebuah kata yang nantinya akan menjadi prediktor pada analisis sentimen. Data *training* digunakan untuk membentuk model, dilakukan pembobotan representasi kedekatan antar kata menjadi numerik dengan menggunakan *Word2Vec*. Jumlah *vocabulary* pada data secara keseluruhan adalah 3354 kata dengan dimensi 100.

Kemudian, indeks kata akan diubah menjadi bentuk vektor  $1 \times 3354$  dengan nilai 1 pada kolom yang mewakili kata tersebut dan 0 pada kolom lainnya. Vektor tersebut akan dilakukan dengan matriks  $W$  berukuran  $3354 \times 100$  dan akan diperoleh vektor baru dari setiap kata dengan ukuran  $1 \times 100$ . Matriks  $W$  yang didapatkan pada data penelitian ini yaitu sebagai berikut:

$$W = \begin{bmatrix} -0.0070443717 & -0.006845473 & \dots & 0.0030869301 \\ 0.00467194 & 0.006726348 & \dots & -0.0007180266 \\ \vdots & \vdots & \ddots & \vdots \\ -0.00053381635 & 0.00024672798 & \dots & 0.0063960026 \end{bmatrix}$$

Kemudian setiap kata pada data ulasan dikalikan dengan matriks  $W$  dengan *one-hot encoding*, misal kata pada indeks 2 di *vocabulary* akan dirubah menjadi vektor dengan ukuran  $1 \times 3354$  dan dikalikan dengan matriks  $W$  sebagai berikut:

$$[0 \ 1 \ \dots \ 0] \times \begin{bmatrix} -0.0070443717 & -0.006845473 & \dots & 0.0030869301 \\ 0.00467194 & 0.006726348 & \dots & -0.0007180266 \\ \vdots & \vdots & \ddots & \vdots \\ -0.00053381635 & 0.00024672798 & \dots & 0.0063960026 \end{bmatrix}$$

Maka dari itu didapatkan vektor kata pertama pada ulasan pertama yaitu kata “mudah,” sebagai berikut:

$$[x_{1,1,1} \ x_{1,1,2} \ \dots \ x_{1,1,100}] = [0.00467194 \ 0.006726348 \ \dots \ -0.000718026]$$

Jika setiap ulasan memiliki 53 vektor kata, maka struktur data setelah dilakukan *Word2Vec* ditunjukkan oleh Tabel 10.

Pembangunan model *Word2Vec* pada data *training* akan ditransformasi pada data *testing* untuk dilakukan analisis klasifikasi dengan menggunakan SVM dan CNN .

F. *Analisis Klasifikasi Menggunakan Support Vector Machine*

Pada metode *Support Vector Machine*, digunakan 5300 input yang terdiri dari 53 vektor kata di mana setiap vektor kata berisi 100 dimensi seperti yang ditunjukkan pada Tabel 11.

Analisis klasifikasi *Support Vector Machine* dengan pembobotan *Word2Vec* digunakan kernel RBF dengan parameter  $C$  dan  $\gamma$ . Optimasi hyperparameter dilakukan dengan menggunakan *gridsearch*. Parameter  $C$  dan  $\gamma$  yang digunakan memiliki nilai  $10^{-3}$  sampai  $10^3$  seperti yang ditunjukkan pada Tabel 12. Setelah dilakukan perhitungan, didapatkan bahwa parameter terbaik terletak pada parameter  $C = 1000$  dan  $\gamma = 10$ .

Setelah didapatkan parameter terbaik, selanjutnya dilakukan analisis ketepatan klasifikasi pada metode SVM berdasarkan *confusion matrix* yang disajikan di Tabel 13.

Lalu dilakukan pengukuran performa ketepatan klasifikasiseperti pada Tabel 14.

Dari model *Support Vector Machine* yang telah dibangun, berikut ini disajikan contoh ulasan, sentimen aktual dan sentimen hasil prediksi dengan metode SVM pada Tabel 15.

G. *Analisa Klasifikasi Menggunakan Convolutional Neural Network*

Pada klasifikasi CNN dilakukan proses konvolusi pada *convolutional layer*, proses *pooling* dengan tujuan menyeleksi atau memilih informasi penting pada *feature map*, selanjutnya *feature map* diproses pada *output layer* dengan aktivasi *sigmoid*. Pembelajaran model CNN pada penelitian ini menggunakan *epoch* sebanyak 30 berdasarkan konvergen.

Model arsitektur pertama adalah *embedding layer* dengan bobot *Word2Vec*, *max pooling* 1D, *layer 1* dengan besar *node* 8 menggunakan aktivasi ReLU, dan *output layer* dengan aktivasi *Sigmoid*. Didapatkan nilai akurasi pada data *training* sebesar 0,985 dengan *loss* 0,0342 dan akurasi data *testing* sebesar 0,898 dengan nilai *loss* sebesar 0,505. Nilai akurasi pada data *testing* sebesar 89,8% memperlihatkan jika kemampuan model dalam melakukan klasifikasi dikategorikan baik. *Confusion matrix* dan ketepatan klasifikasi ditunjukkan pada Tabel 16.

Model arsitektur kedua adalah *embedding layer* dengan bobot *Word2Vec*, *max pooling* 1D, *layer 1* dengan besar *node* 16, *layer 2* dengan besar *node* 8 menggunakan aktivasi ReLU, dan *output layer* dengan aktivasi *Sigmoid*. Didapatkan nilai akurasi pada data *training* sebesar 0,984 dengan *loss* sebesar 0,0341 dan akurasi data *testing* sebesar 0,884 dengan nilai *loss* sebesar 0,561. Nilai akurasi pada data *testing* sebesar 88,4% memperlihatkan jika kemampuan model dalam melakukan klasifikasi dikategorikan baik. *Confusion matrix* dan ketepatan klasifikasi ditunjukkan pada Tabel 17.

Model arsitektur ketiga adalah *embedding layer* dengan bobot *Word2Vec*, *max pooling* 1D, *layer 1* dengan besar *node* 32, *layer 2* dengan besar *node* 16, *layer 3* dengan besar *node* 8 menggunakan aktivasi ReLU, dan *output layer* dengan aktivasi *Sigmoid*. Didapatkan nilai akurasi pada data *training* sebesar 0,983 dengan *loss* sebesar 0,0332 dan akurasi data *testing* sebesar 0,892 dengan nilai *loss* sebesar 0,624. Nilai akurasi pada data *testing* sebesar 89,2% memperlihatkan jika kemampuan model dalam melakukan klasifikasi

dikategorikan baik (*good classification*). *Confusion matrix* dan ketepatan klasifikasi ditunjukkan pada Tabel 18.

Model arsitektur keempat yang digunakan dalam penelitian ini adalah *embedding layer* dengan bobot *Word2Vec, max pooling* 1D, *layer* 1 dengan besar *node* 64, *layer* 2 dengan besar *node* 32, *layer* 3 dengan besar *node* 16, *layer* 4 dengan besar *node* 8 menggunakan aktivasi ReLU, dan *output layer* dengan aktivasi *Sigmoid*. Didapatkan nilai akurasi pada data *training* sebesar 0,982 dengan *loss* sebesar 0,0339 dan akurasi data *testing* sebesar 0,886 dengan *loss* sebesar 0,632. Nilai akurasi pada data *testing* sebesar 88,6% memperlihatkan jika kemampuan model dalam melakukan klasifikasi dikategorikan baik. *Confusion matrix* dan ketepatan klasifikasi ditunjukkan pada Tabel 19.

Model arsitektur kelima adalah *embedding layer* dengan bobot *Word2Vec, max pooling* 1D, *layer* 1 dengan besar *node* 128, *layer* 2 dengan besar *node* 64, *layer* 3 dengan besar *node* 32, *layer* 4 dengan besar *node* 16, *layer* 5 dengan besar *node* 8 menggunakan aktivasi ReLU, dan *output layer* dengan aktivasi *Sigmoid*. Pada model arsitektur kelima didapatkan nilai akurasi pada data *training* sebesar 0,982 dengan *loss* sebesar 0,0421 dan akurasi data *testing* sebesar 0,886 dengan *loss* sebesar 0,576. Nilai akurasi pada data *testing* sebesar 88,6% memperlihatkan jika kemampuan model dalam melakukan klasifikasi dikategorikan baik. *Confusion matrix* dan ketepatan klasifikasi ditunjukkan pada Tabel 20.

Didapatkan bahwa model arsitektur pertama merupakan model yang paling baik dengan nilai akurasi yang baik pada data *training* dan *testing* hasil pembelajaran pada *epoch* terakhir berturut-turut sebesar 98,5% dan 89,8%.

Dari model *Support Vector Machine* yang telah dibangun, berikut ini disajikan contoh ulasan, sentimen aktual dan sentimen hasil prediksi dengan metode CNN pada Tabel 21.

#### H. Perbandingan Ketepatan Klasifikasi

Perbandingan ketepatan klasifikasi dari metode SVM dan CNN seperti yang disajikan pada Tabel 22.

Hasil menunjukkan bahwa metode *Convolutional Neural Network* lebih baik dibandingkan *Support Vector Machine* dalam pengimplementasian analisis sentimen pada layanan aplikasi Bank Aladin Oktober 2022 – Desember 2022.

### V. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan terhadap data ulasan layanan Bank Aladin, maka didapatkan kesimpulan sebagai berikut: (1) Ulasan aplikasi Bank Aladin pada 1 Oktober 2022 – 31 Desember 2020 mayoritas mengandung ulasan bersentimen positif. Pada visualisasi *word cloud* Kata yang paling sering muncul pada kumpulan ulasan yang bersentimen positif adalah “mudah”. Kata “mudah” berhubungan dengan aplikasi Bank Aladin yang dinilai mudah digunakan dan dapat memudahkan transaksi online para pengguna. Sedangkan kata yang paling

sering muncul pada kumpulan ulasan yang bersentimen negatif adalah “daftar”. Kata “daftar” menunjukkan adanya kendala yang dihadapi oleh pengguna aplikasi Bank Aladin saat melakukan pendaftaran dan juga banyaknya protes karena tidak mendapatkan bonus saldo saat daftar sesuai yang diberitakan. (2) Pada metode SVM parameter yang digunakan yaitu  $C = 1000$  dan  $\gamma = 10$ , pada data *testing* didapatkan *accuracy* sebesar 81,9%, *precision* sebesar 85,7% dan *recall* sebesar 93%. Sementara itu pada metode *Convolutional Neural Network*, didapatkan bahwa model arsitektur pertama di mana *layer* 1 dengan *node* 8 merupakan model yang paling baik di data *testing* didapatkan *accuracy* sebesar 89,8%, *precision* sebesar 93,5% dan *recall* sebesar 93,8%. (3) Berdasarkan perbandingan metode SVM dan CNN untuk ulasan pada layanan aplikasi Bank Aladin didapatkan bahwa metode CNN lebih baik ketepatan klasifikasinya dibandingkan dengan metode SVM. Hal ini ditunjukkan dengan nilai *accuracy* pada model CNN sebesar 89,8% sedangkan pada metode SVM sebesar 81,9%.

Saran yang dapat diberikan kepada peneliti selanjutnya adalah mencari kamus lain untuk *stopword removal* seperti *colloquial Indonesian lexicon*, mengembangkan variasi data atau tingkat kompleksitas data. Bagi Bank Aladin, berdasarkan keluhan di ulasan, dapat meningkatkan kemudahan dan sosialisasi pada proses pendaftaran bagi pengguna baru.

### DAFTAR PUSTAKA

- [1] A. Pambudi and S. Suprpto, “Effect of sentence length in sentiment analysis using support vector machine and convolutional neural network method,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, p. 21, Jan. 2021, doi: 10.22146/ijccs.61627.
- [2] S. Pandya and P. Mehta, “A review on sentiment analysis methodologies, practices and applications,” *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 601–609, Feb. 2020.
- [3] I. N. Dewi, R. Nurcahyo, and Farizal, “Word cloud result of mobile payment user review in indonesia,” in *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, 2020, pp. 989–992. doi: 10.1109/ICIEA49774.2020.9102048.
- [4] I. R. L. O. Richard L. Scheaffer; William Mendenhall, *Elementary Survey Sampling 7th Edition*, 7th ed. Boston: Cengage Learning, 2011, ISBN: 9780840053619.
- [5] R. D. Fitriani, H. Yasin, and Tarno, “Penanganan klasifikasi kelas data tidak seimbang dengan random oversampling pada naive bayes (studi kasus: status peserta kb iud di kabupaten kendal),” *Jurnal Gaussian*, vol. 10, no. 1, pp. 11–20, Feb. 2021, doi: 10.14710/j.gauss.10.1.11-20.
- [6] B. Santosa, *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis, Teori & Aplikasi*, 1st ed. Yogyakarta: Graha Ilmu, 2007, ISBN: 978-979-756-224-3.
- [7] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, “Kernlab - an S4 package for kernel methods in R,” *J Stat Softw*, vol. 11, no. 9, pp. 1–20, 2004, doi: 10.18637/jss.v011.i09.
- [8] A. Santoso and G. Ariyanto, “Implementasi deep learning berbasis keras untuk pengenalan wajah,” *Jurnal Teknik Elektro*, vol. 18, no. 01, 2016.
- [9] J. Han, M. Kamber, and J. Pei, *Data Mining - Concepts And Techniques*, 3rd ed. Burlington: Morgan Kaufman, 2012, ISBN: 978-0-12-381479-1.