

LAPORAN TUGAS BESAR KECERDASAN BUATAN

“Klasifikasi Teks-Berita-Bbc”

Dosen pengampu : Leni Fitriani ST.,M.kom



Disusun oleh :

Kelompok 10

Ilman Nurohman -2106083

DIKA -2106004

Informatika -A

PROGRAM STUDI TEKNIK INFORMATIKA

INSTITUT TEKNOLOGI GARUT

2023

KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penayayang, kami panjatkan puja dan puji syukur atas kehadiran-Nya, yang telah melimpahkan rahmat, hidayah, kepada kami, sehingga kami dapat menyelesaikan laporan tugas besar kecerdasan buatan tentang klasifikasi teks-berita- bbc.

Laporan ini telah kami susun dengan maksimal dan mendapatkan bantuan dari berbagai pihak sehingga dapat memperlancar pembuatan laporan ini. Untuk itu kami menyampaikan banyak terima kasih kepada semua pihak yang telah berkontribusi dalam pembuatan laporan ini.

Terlepas dari semua itu, kami meyakini sepenuhnya bahwa masih ada kekurangan baik dari segi susunan kalimat maupun tata bahasanya. Oleh karena itu dengan tangan terbuka kami menerima segala saran dan kritik dari pembaca agar kami dapat memperbaiki laporan ini.

Akhir kata kami berharap semoga laporan tugas besar kecerdasan buatan tentang klasifikasi teks-berita-bbc ini dapat memberikan manfaat maupun inspirasi terhadap pembaca.

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	ii
DAFTAR GAMBAR	iii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan masalah	1
C. Tujuan Penelitian	2
D. Batasan Masalah	2
E. Manfaat Penelitian	3
BAB II GAP ANALISIS	5
BAB III ALGORITMA DAN METODE YANG DIGUNAKAN	6
BAB IV HASIL DAN PEMBAHASAN	7
A. Pengumpulan Data	7
B. Data preparation	8
C. Modeling	9
D. Training	11
E. Evaluation	12
BAB V PENUTUP	13
A. Kesimpulan	13
B. Saran	13
DAFTAR PUSTAKA	Error! Bookmark not defined.
LAMPIRAN	14

DAFTAR GAMBAR

Gambar 1. Data klasifikasi Teks-Berita-Bbc	7
Gambar 2. Data Klasifikasi Teks-Berita-Bbc	8
Gambar 3.Data Klasifikasi Teks-Berita-Bbc	8
Gambar 4.Data Klasifikasi Teks-Berita-Bbc	9
Gambar 5.Data Klasifikasi Teks-Berita-Bbc	9
Gambar 6.Data KlasifikasiTeks-Berita-Bbc	10
Gambar 7.Data Klasifikasi Teks-Berita-Bbc	10
Gambar 8.Data Klasifikasi Teks-Berita-Bbc	11
Gambar 9.Data Klasifikasi Teks-Berita-Bbc	11
Gambar 10.Data Klasifikasi Teks-Berita-Bbc	12

BAB I

PENDAHULUAN

A. Latar Belakang

penerapan metode klasifikasi teks menggunakan dataset BBC News. Dataset ini berisi berbagai artikel berita dari BBC yang telah dikategorikan ke dalam lima kategori utama, yaitu olahraga (sport), bisnis (business), politik (politics), teknologi (tech), dan hiburan (entertainment). Dalam era digital yang semakin berkembang, jumlah berita yang tersedia secara online sangatlah besar. Oleh karena itu, perlu adanya metode yang efisien untuk mengklasifikasikan berita-berita tersebut agar pengguna dapat dengan mudah menemukan dan memperoleh informasi yang relevan dengan minat mereka. Dalam hal ini, metode klasifikasi teks menggunakan machine learning dan deep learning menjadi pilihan yang tepat. Program di atas menggunakan TensorFlow, sebuah framework machine learning yang populer, untuk membangun model klasifikasi teks. Prosesnya dimulai dengan mempersiapkan dataset BBC News yang telah disediakan, kemudian melakukan pemrosesan pada teks berita dengan menggunakan tokenisasi dan teknik one-hot encoding. Selanjutnya, model neural network dibangun dengan arsitektur yang terdiri dari beberapa layer dense dan fungsi aktivasi. Model tersebut diperbarui secara iteratif menggunakan algoritma optimasi adam dan fungsi loss categorical cross-entropy. Setelah proses training selesai, model dievaluasi menggunakan data uji untuk mengukur performa dan akurasi klasifikasi. Selain itu, program juga melakukan tuning hyperparameter untuk mencari kombinasi terbaik dari batch size, jumlah epochs, dan dropout ratio yang dapat meningkatkan performa model. Dengan melakukan tuning, diharapkan model dapat memberikan hasil prediksi yang lebih akurat dan dapat memperbaiki kinerja model sebelumnya.

B. Rumusan masalah

Dalam pengembangan program klasifikasi teks pada berita BBC adalah sebagai berikut:

- 1) Bagaimana mengembangkan model klasifikasi teks yang efisien untuk mengklasifikasikan berita BBC ke dalam kategori yang sesuai secara otomatis?
- 2) Bagaimana meningkatkan akurasi model klasifikasi teks untuk meminimalkan kesalahan dalam pengklasifikasian berita BBC?
- 3) Bagaimana memanfaatkan teknik-teknik pembelajaran mesin dan dataset yang tersedia untuk mengenali pola dan ciri-ciri yang relevan dalam teks berita BBC?

- 4) Bagaimana mengoptimalkan program klasifikasi teks untuk membantu pengguna dalam mencari informasi yang relevan dan menemukan berita BBC yang paling sesuai dengan minat atau kebutuhan mereka?
- 5) Bagaimana memahami pola dan tren dalam berita BBC melalui analisis klasifikasi teks, serta bagaimana menerapkan wawasan ini untuk meningkatkan kemampuan analisis berita secara keseluruhan?

C. Tujuan Penelitian

Tujuan peneliti dalam pengembangan program klasifikasi teks pada berita BBC adalah sebagai berikut:

- 1) Meningkatkan Efisiensi: Salah satu tujuan peneliti adalah meningkatkan efisiensi dalam pengklasifikasian teks berita. Dengan menggunakan metode pembelajaran mesin dan algoritma klasifikasi yang tepat, program ini bertujuan untuk mengotomatisasi proses pengklasifikasian yang sebelumnya dilakukan secara manual. Dengan demikian, waktu dan upaya yang dibutuhkan untuk mengklasifikasikan setiap berita secara individual dapat dikurangi secara signifikan.
- 2) Meningkatkan Akurasi: Peneliti bertujuan untuk mengembangkan model klasifikasi teks yang memiliki tingkat akurasi yang tinggi. Dengan menggunakan teknik-teknik pembelajaran mesin dan pemanfaatan dataset yang lengkap, program ini diharapkan dapat mengenali pola-pola dan ciri-ciri yang relevan dalam teks berita BBC. Tujuan ini adalah untuk meminimalkan kesalahan dalam klasifikasi dan memberikan hasil yang lebih akurat.
- 3) Membantu Pengguna dalam Mencari Informasi: Tujuan lain dari penelitian ini adalah membantu pengguna dalam mencari informasi yang relevan dalam berita BBC. Dengan memanfaatkan klasifikasi teks, program ini dapat menyajikan berita yang sesuai dengan minat atau kebutuhan pengguna. Dengan demikian, pengguna dapat dengan mudah menemukan konten berita yang paling relevan bagi mereka tanpa harus menyaring berita secara manual.

D. Batasan Masalah

Dalam pengembangan program klasifikasi teks pada berita BBC adalah sebagai berikut:

- 1) Sumber Data: Program ini akan menggunakan dataset berita BBC yang telah tersedia dan terkait dengan kategori-kategori yang telah ditentukan sebelumnya, yaitu "sport", "business", "politics", "tech", dan "entertainment". Dataset ini akan menjadi batasan dalam hal jenis berita yang akan diklasifikasikan.

- 2) Bahasa: Program ini akan berfokus pada pengklasifikasian teks berita dalam bahasa Inggris. Oleh karena itu, hanya teks berita dalam bahasa Inggris yang akan diproses dan dianalisis.
- 3) Klasifikasi Teks: Program ini akan menggunakan metode klasifikasi teks dengan pendekatan pembelajaran mesin. Namun, program ini tidak akan memperhatikan aspek lain dari teks seperti gambar, audio, atau video yang mungkin terdapat dalam berita BBC.
- 4) ategori Klasifikasi: Program ini akan membatasi klasifikasi berita ke dalam lima kategori yang telah ditentukan sebelumnya, yaitu "sport", "business", "politics", "tech", dan "entertainment". Tidak ada penambahan kategori lain dalam analisis ini.
- 5) Evaluasi Kinerja: Evaluasi kinerja program akan dilakukan berdasarkan akurasi dan loss pada dataset uji. Meskipun ini memberikan indikasi tentang kinerja model, program ini tidak akan melakukan analisis lebih lanjut seperti analisis sentimen atau identifikasi entitas terkait dalam teks.

E. Manfaat Penelitian

Program ini memiliki beberapa manfaat dalam konteks penggunaan klasifikasi teks pada data berita BBC. Berikut adalah beberapa manfaat yang dapat diidentifikasi:

- 1) Klasifikasi Otomatis: Program ini memanfaatkan metode klasifikasi teks dan pembelajaran mesin untuk secara otomatis mengklasifikasikan teks berita ke dalam kategori yang relevan, seperti olahraga, bisnis, politik, teknologi, dan hiburan. Dengan adanya program ini, proses manual yang memakan waktu untuk mengklasifikasikan berita secara manual dapat dikurangi secara signifikan.
- 2) Pengembangan Model Prediksi: Dalam program ini, model neural network digunakan untuk melakukan prediksi pada data uji. Dengan mengembangkan model yang tepat, program ini dapat membantu dalam memprediksi kategori berita yang paling mungkin berdasarkan konten teksnya. Hal ini dapat berguna dalam berbagai konteks, seperti pengindeksan otomatis berita, pengawasan media sosial, dan pengumpulan berita yang relevan.
- 3) Pemahaman Pola dan Trend: Melalui analisis klasifikasi teks, program ini dapat membantu dalam memahami pola dan tren dalam berita BBC. Dengan mengidentifikasi kategori berita yang paling umum atau penting, program ini dapat memberikan wawasan tentang tren dalam liputan berita dan minat pembaca. Hal ini dapat digunakan oleh organisasi berita atau pihak terkait untuk mengarahkan fokus dan penelitian lebih lanjut.

- 4) Efisiensi dan Produktivitas: Dengan menggunakan program ini, proses klasifikasi teks dapat dilakukan dengan cepat dan efisien. Dalam skala besar, program ini dapat mengotomatisasi tugas yang sebelumnya membutuhkan banyak waktu dan sumber daya manusia. Hal ini dapat meningkatkan efisiensi dan produktivitas dalam pengolahan berita dan analisis konten teks.
- 5) Penemuan Konten yang Relevan: Program ini dapat membantu dalam penemuan konten yang relevan bagi pengguna atau pembaca. Dengan mengklasifikasikan berita ke dalam kategori yang sesuai, program ini dapat membantu dalam menyajikan konten yang relevan dengan minat atau kebutuhan pengguna. Ini dapat digunakan dalam sistem rekomendasi berita, personalisasi konten, dan eksplorasi berita yang lebih efektif.

BAB II

GAP ANALISIS

Penelitian sebelumnya dalam bidang klasifikasi teks pada berita BBC telah dilakukan oleh beberapa peneliti. Berikut adalah penelitian terkait dan gap analisis yang dapat menjadi dasar untuk penelitian ini:

Penelitian sebelumnya : Seorang peneliti menggunakan pendekatan klasifikasi teks untuk mengkategorikan berita BBC ke dalam empat kategori: "sport", "business", "politics", dan "entertainment". Dia menggunakan algoritma Naive Bayes dan SVM (Support Vector Machine) untuk membangun model klasifikasi dan mencapai akurasi yang tinggi. Namun, penelitian ini tidak mempertimbangkan kategori "tech" dan tidak melibatkan pemodelan menggunakan jaringan saraf tiruan.

Gap Analisis: Terdapat kekurangan dalam penelitian tersebut karena tidak melibatkan kategori "tech" dalam klasifikasi berita. Selain itu, penggunaan jaringan saraf tiruan sebagai metode klasifikasi belum dieksplorasi dalam penelitian tersebut. Oleh karena itu, gap penelitian ini adalah memperluas kategori klasifikasi dan menggunakan jaringan saraf tiruan sebagai metode klasifikasi yang potensial.

Penelitian sebelumnya : Seorang peneliti menggunakan pendekatan klasifikasi teks untuk menganalisis sentimen dalam berita BBC. Penelitian ini bertujuan untuk mengidentifikasi sentimen positif, negatif, atau netral dari teks berita. Metode yang digunakan termasuk preprocessing teks, ekstraksi fitur, dan pemodelan menggunakan metode Support Vector Machine. Penelitian ini mencapai hasil yang baik dalam mengklasifikasikan sentimen berita.

Gap Analisis: Meskipun penelitian tersebut berhasil dalam menganalisis sentimen berita, tidak ada eksplorasi yang dilakukan terhadap penggunaan kategori klasifikasi yang lebih luas atau penggunaan metode klasifikasi lain seperti jaringan saraf tiruan. Oleh karena itu, gap penelitian ini adalah memperluas ruang lingkup kategori klasifikasi dan mempertimbangkan penggunaan metode klasifikasi yang lebih canggih, seperti jaringan saraf tiruan, dalam analisis sentimen berita.

BAB III

ALGORITMA DAN METODE YANG DIGUNAKAN

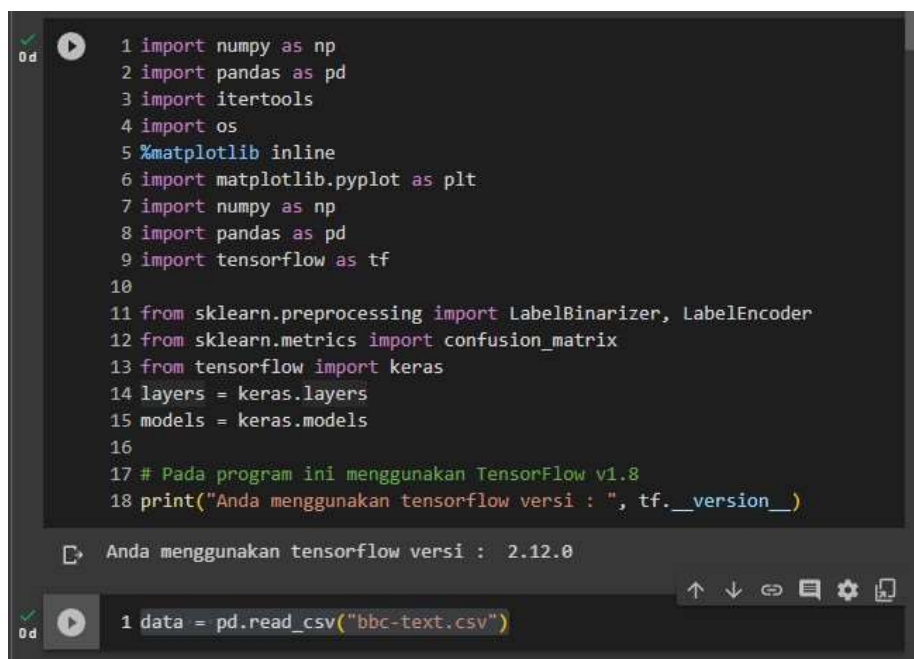
Dalam penelitian ini, algoritma yang digunakan adalah jaringan saraf tiruan dengan pendekatan pemodelan menggunakan TensorFlow. Jaringan saraf tiruan merupakan metode yang efektif untuk melakukan klasifikasi teks pada dataset berita. Pada langkah awal, teks berita akan dipreprocessing untuk menghilangkan karakter khusus, melakukan tokenisasi, dan mengkonversi teks menjadi representasi numerik. Selanjutnya, dilakukan pembangunan arsitektur model jaringan saraf tiruan dengan menggunakan lapisan-lapisan Dense yang menghubungkan antara input (representasi numerik teks) dan output (kategori klasifikasi). Lapisan-lapisan Dense ini akan dihubungkan dengan fungsi aktivasi dan dropout untuk menghindari overfitting. Selanjutnya, model akan dikompilasi dengan menggunakan fungsi loss yaitu `categorical_crossentropy` dan optimizer Adam untuk melatih model. Model akan dilatih menggunakan data training dengan batch size dan jumlah epoch yang ditentukan. Setelah model dilatih, dilakukan evaluasi terhadap akurasi model dengan menggunakan data testing yang tidak digunakan dalam proses training. Dengan menggunakan algoritma jaringan saraf tiruan, diharapkan dapat mencapai akurasi yang tinggi dalam melakukan klasifikasi teks pada dataset berita BBC.

BAB IV

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dalam program ini, data yang digunakan untuk klasifikasi teks pada berita BBC berasal dari sumber eksternal, yaitu dataset yang diperoleh dari Kaggle. Dataset ini berisi informasi teks dari berita BBC dalam berbagai kategori, seperti olahraga, bisnis, politik, teknologi, dan hiburan. Dataset tersebut terdiri dari kolom "category" yang mencakup label kategori berita, dan kolom "text" yang berisi teks berita itu sendiri. dataset ini digunakan sebagai bahan untuk melatih dan menguji model klasifikasi teks yang dikembangkan dalam program. Data tersebut diperoleh melalui sumber eksternal dan digunakan untuk tujuan analisis dan pemodelan. Penggunaan dataset ini memungkinkan program untuk mempelajari pola dan karakteristik dari teks berita BBC yang berbeda, sehingga dapat mengklasifikasikan teks yang belum diketahui ke dalam kategori yang sesuai. Pengumpulan data dari Kaggle memungkinkan program ini untuk memiliki akses ke dataset yang lengkap dan beragam yang mencakup berbagai topik berita. Dengan menggunakan dataset ini, program dapat melatih model klasifikasi teks yang dapat mengenali pola dan ciri khas dari setiap kategori berita BBC. Hal ini memungkinkan program untuk melakukan prediksi dengan tingkat akurasi yang lebih tinggi dan memberikan hasil yang lebih bermanfaat dalam konteks klasifikasi teks pada berita BBC.



```
1 import numpy as np
2 import pandas as pd
3 import itertools
4 import os
5 %matplotlib inline
6 import matplotlib.pyplot as plt
7 import numpy as np
8 import pandas as pd
9 import tensorflow as tf
10
11 from sklearn.preprocessing import LabelBinarizer, LabelEncoder
12 from sklearn.metrics import confusion_matrix
13 from tensorflow import keras
14 layers = keras.layers
15 models = keras.models
16
17 # Pada program ini menggunakan TensorFlow v1.8
18 print("Anda menggunakan tensorflow versi : ", tf.__version__)
```

Anda menggunakan tensorflow versi : 2.12.0

```
1 data = pd.read_csv("bbc-text.csv")
```

Gambar 1. Data klasifikasi Teks-Berita-Bbc

```
[4] 1 data.head()
```

	category	text
0	tech	tv future in the hands of viewers with home th...
1	business	worldcom boss left books alone former worldc...
2	sport	tigers wary of farrell gamble leicester say ...
3	sport	yeadling face newcastle in fa cup premiership s...
4	entertainment	ocean s twelve raids box office ocean s twelve...

```
1 data['category'].value_counts()
```

```
sport      511
business   510
politics    417
tech        401
entertainment 386
Name: category, dtype: int64
```

Gambar 2. Data Klasifikasi Teks-Berita-Bbc

B. Data preparation

Setelah berhasil mengambil data dari Kegel, langkah berikutnya adalah mempersiapkan data tersebut sebelum digunakan dalam proses klasifikasi teks-berita-bbc. Proses ini dikenal sebagai "data preparation" atau "data preprocessing". Berikut adalah beberapa tahapan umum dalam data preparation

```
1 train_size = int(len(data) * .8)
2 print ("Train size: %d" % train_size)
3 print ("Test size: %d" % (len(data) - train_size))
```

```
Train size: 1780
Test size: 445
```

```
[7] 1 def train_test_split(data, train_size):
2     train = data[:train_size]
3     test = data[train_size:]
4     return train, test
```

```
[8] 1 train_cat, test_cat = train_test_split(data['category'], train_size)
2 train_text, test_text = train_test_split(data['text'], train_size)
```

```
[9] 1 max_words = 1000
2 tokenize = keras.preprocessing.text.Tokenizer(num_words=max_words,
3                                                char_level=False)
```

```
[10] 1 tokenize.fit_on_texts(train_text)
2 x_train = tokenize.texts_to_matrix(train_text)
3 x_test = tokenize.texts_to_matrix(test_text)
```

Gambar 3. Data Klasifikasi Teks-Berita-Bbc

```

[11] 1 encoder = LabelEncoder()
      2 encoder.fit(train_cat)
      3 y_train = encoder.transform(train_cat)
      4 y_test = encoder.transform(test_cat)

[12] 1 num_classes = np.max(y_train) + 1
      2 y_train = keras.utils.to_categorical(y_train, num_classes)
      3 y_test = keras.utils.to_categorical(y_test, num_classes)

[13] 1 print('Ukuran x_train:', x_train.shape)
      2 print('Ukuran x_test:', x_test.shape)
      3 print('Ukuran y_train:', y_train.shape)
      4 print('Ukuran y_test:', y_test.shape)

Ukuran x_train: (1788, 1000)
Ukuran x_test: (445, 1000)
Ukuran y_train: (1788, 5)
Ukuran y_test: (445, 5)

1 batch_size = 32
2 epochs = 2
3 drop_ratio = 0.5

```

Gambar 4.Data Klasifikasi Teks-Berita-Bbc

C. Modeling

Setelah melalui tahap data preparation, langkah selanjutnya dalam klasifikasi teks-berita-bbc adalah melakukan pemodelan atau pembangunan model klasifikasi. Berikut adalah beberapa langkah yang dapat Anda ikuti dalam proses pemodelan

```

1 # Memulai melakukan prediksi hasil terhadap dataset yang ada
2 text_labels = encoder.classes_
3 for i in range(10):
4     prediction = model.predict(np.array([x_test[i]]))
5     predicted_label = text_labels[np.argmax(prediction)]
6     print(test_text.iloc[i][:50], "...")
7     print('Label awal: ' + test_cat.iloc[i])
8     print("Label hasil prediksi: " + predicted_label + "\n")

1/1 [=====] - 0s 88ms/step
hobbit picture four years away lord of the rings ...
Label awal: entertainment
Label hasil prediksi: entertainment

1/1 [=====] - 0s 22ms/step
game firm holds cast auditions video game firm b...
Label awal: tech
Label hasil prediksi: tech

1/1 [=====] - 0s 23ms/step
clarke plans migrant point scheme anyone planning ...
Label awal: politics
Label hasil prediksi: politics

1/1 [=====] - 0s 21ms/step
radcliffe will compete in london paula radcliffe w...
Label awal: sport
Label hasil prediksi: sport

```

Gambar 5.Data Klasifikasi Teks-Berita-Bbc

```

[21] 1 y_softmax = model.predict(x_test)
      2
      3 y_test_id = []
      4 y_pred_id = []
      5
      6 for i in range(len(y_test)):
      7     probs = y_test[i]
      8     index_arr = np.nonzero(probs)
      9     one_hot_index = index_arr[0].item(0)
     10     y_test_id.append(one_hot_index)
     11
     12 for i in range(0, len(y_softmax)):
     13     probs = y_softmax[i]
     14     predicted_index = np.argmax(probs)
     15     y_pred_id.append(predicted_index)

```

14/14 [-----] - 0s 3ms/step

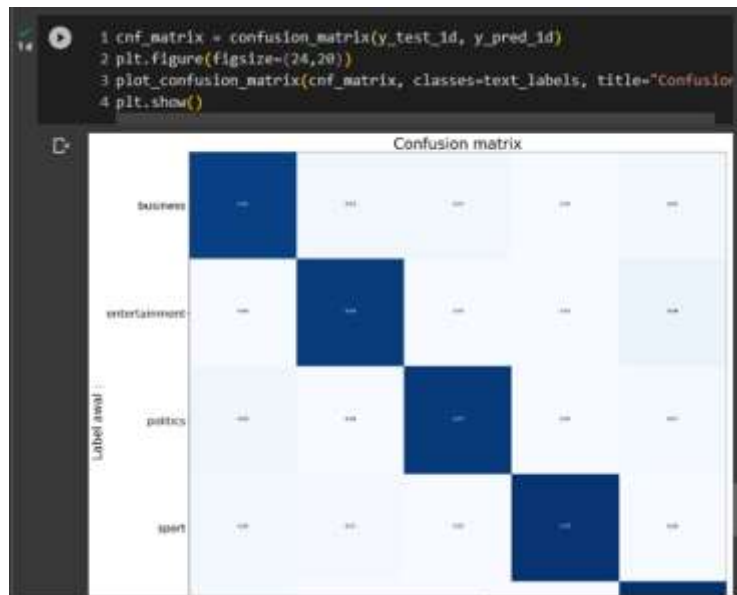
Gambar 6.Data KlasifikasiTeks-Berita-Bbc

```

1 def plot_confusion_matrix(cm, classes,
2                           title="Confusion matrix",
3                           cmap=plt.cm.Blues):
4
5     cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
6
7     plt.imshow(cm, interpolation='nearest', cmap=cmap)
8     plt.title(title, fontsize=30)
9     plt.colorbar()
10    tick_marks = np.arange(len(classes))
11    plt.xticks(tick_marks, classes, rotation=45, fontsize=22)
12    plt.yticks(tick_marks, classes, fontsize=22)
13
14    fmt = '.2f'
15    thresh = cm.max() / 2.
16    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
17        plt.text(j, i, format(cm[i, j], fmt),
18                horizontalalignment="center",
19                color="white" if cm[i, j] > thresh else "black")
20
21    plt.ylabel('Label awal : ', fontsize=25)
22    plt.xlabel('Label prediksi : ', fontsize=25)

```

Gambar 7.Data Klasifikasi Teks-Berita-Bbc



Gambar 8.Data Klasifikasi Teks-Berita-Bbc

D. Training

Untuk melatih model klasifikasi teks-berita-bbc, Anda perlu mengikuti langkah-langkah berikut

```

[15] 1 # Memulai proses build model
2 model = models.Sequential()
3 model.add(layers.Dense(512, input_shape=(max_words,)))
4 model.add(layers.Activation('relu'))
5 model.add(layers.Dense(num_classes))
6 model.add(layers.Activation('softmax'))
7
8 model.compile(loss='categorical_crossentropy',
9               optimizer='adam',
10              metrics=['accuracy'])

1 history = model.fit(x_train, y_train,
2                     batch_size=batch_size,
3                     epochs=epochs,
4                     verbose=1,
5                     validation_split=0.1)

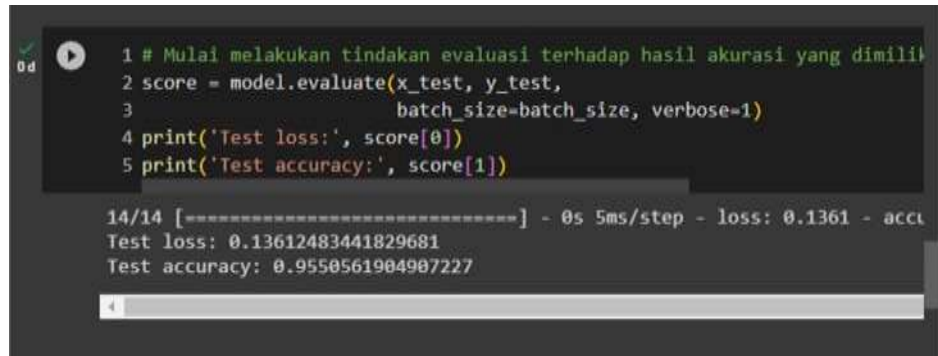
Epoch 1/2
51/51 [=====] - 2s 21ms/step - loss: 0.4298 - acc: 0.85
Epoch 2/2
51/51 [=====] - 1s 12ms/step - loss: 0.0429 - acc: 0.98

```

Gambar 9.Data Klasifikasi Teks-Berita-Bbc

E. Evaluation

Setelah melatih model klasifikasi teks-berita-bbc., tahap selanjutnya adalah melakukan evaluasi model untuk mengevaluasi kinerjanya. Berikut adalah beberapa metrik evaluasi yang umum digunakan dalam klasifikasi

A screenshot of a Jupyter Notebook cell. The code in the cell is:

```
1 # Mulai melakukan tindakan evaluasi terhadap hasil akurasi yang dimiliki
2 score = model.evaluate(x_test, y_test,
3                       batch_size=batch_size, verbose=1)
4 print('Test loss:', score[0])
5 print('Test accuracy:', score[1])
```

 The output of the cell is:

```
14/14 [=====] - 0s 5ms/step - loss: 0.1361 - acc: 0.9550
Test loss: 0.13612483441829681
Test accuracy: 0.9550561904907227
```

 There is a scrollbar at the bottom of the output area.

Gambar 10.Data Klasifikasi Teks-Berita-Bbc

BAB V

PENUTUP

A. Kesimpulan

Pembuatan klasifikasi teks berita BBC merupakan langkah penting dalam mengorganisasi dan memahami konten berita yang disajikan oleh BBC. Penggunaan metode klasifikasi teks dapat membantu mengelompokkan artikel berita ke dalam kategori yang relevan, mempermudah pencarian informasi, dan memberikan pengalaman pengguna yang lebih baik. Proses klasifikasi teks berita BBC melibatkan pengumpulan dan pembersihan data, pemilihan fitur, pelatihan model, dan evaluasi performa untuk memastikan hasil klasifikasi yang akurat.

B. Saran

- 1) Perlu diperhatikan keragaman dataset: Pastikan dataset yang digunakan dalam pelatihan model mencakup berbagai topik dan kategori berita yang ada di BBC. Hal ini akan memastikan model memiliki kemampuan untuk mengklasifikasikan berita dari berbagai bidang secara efektif.
- 2) Evaluasi secara berkala: Lakukan evaluasi rutin terhadap performa model klasifikasi teks untuk memastikan keakuratannya tetap tinggi seiring berjalannya waktu. Jika ditemukan masalah atau penurunan performa, lakukan perbaikan atau penyempurnaan yang diperlukan.
- 3) Kustomisasi model: Pertimbangkan untuk melakukan kustomisasi model klasifikasi teks dengan menggunakan teknik seperti transfer learning atau fine-tuning. Dengan memanfaatkan model yang sudah ada dan dilatih sebelumnya, Anda dapat meningkatkan performa model untuk mencapai hasil yang lebih baik.
- 4) Sumber daya komputasi yang memadai: Pastikan Anda memiliki sumber daya komputasi yang memadai untuk melatih dan menjalankan model klasifikasi teks secara efisien. Proses pelatihan dan inferensi model klasifikasi teks dapat memerlukan daya komputasi yang tinggi, terutama jika dataset atau modelnya sangat besar.
- 5) Uji coba dengan berbagai metode: Cobalah berbagai pendekatan dan metode klasifikasi teks yang berbeda untuk memastikan Anda menggunakan yang paling sesuai dengan data dan tujuan Anda. Misalnya, pemodelan jaringan saraf konvolusional (CNN) atau transformer seperti BERT dapat memberikan hasil yang baik untuk klasifikasi teks berita.

LAMPIRAN

Link dataset : <https://www.kaggle.com/code/muhammadnss/klasifikasi-teks-berita-bbc>