

**PERBANDINGAN KINERJA ALGORITME *CONVOLUTIONAL NEURAL NETWORK* DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN PERFORMA *MANCHESTER UNITED* DI TWITTER**

**(Skripsi)**

**Oleh**

**MUHAMMAD NAUFAL HUMAM  
1617051115**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS LAMPUNG**

**BANDAR LAMPUNG**

**2023**

## Abstrak

# PERBANDINGAN KINERJA ALGORITME *CONVOLUTIONAL NEURAL NETWORK* DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN PERFORMA *MANCHESTER UNITED* DI TWITTER

Oleh

**Muhammad Naufal Humam**

Analisis sentimen digunakan untuk melihat kecenderungan suatu sentimen atau pendapat, apakah pendapat tersebut cenderung beropini positif atau negatif. Dalam penelitian terbaru, data biasanya diperoleh dari media sosial, termasuk Twitter, di mana pengguna sering memberikan pendapat pribadi mereka tentang subjek tertentu. Manchester United menjadi klub yang sering dibicarakan di twitter mengenai performanya dalam sepakbola, baik dalam performa bagus ataupun performa buruk. Penelitian ini membandingkan dua algoritme, yaitu *Convolutional Neural Network* dan *Naive Bayes* dalam analisis sentimen menggunakan data sentimen Manchester United dari Twitter. Data yang digunakan pada penelitian ini berbahasa Indonesia dan berbahasa Inggris. Berdasarkan hasil penelitian, pada dataset berbahasa Inggris *Convolutioanl Neural Network* menunjukkan kinerja terbaik dengan nilai akurasi 94% dibandingkan dengan nilai *Naive Bayes* 79%. Begitupun pada dataset berbahasa Indonesia, *Convolutioanl Neural Network* menunjukkan kinerja terbaik dengan nilai akurasi 91% dibandingkan dengan nilai *Naive Bayes* 75%. Untuk sentimen soal Manchester United, sentimen positif cenderung muncul ketika hasil pertandingan dimenangkan oleh Manchester United, dan pada sentimen berbahasa Inggris maupun Indonesia sentimen positif banyak diisi oleh kata-kata pujian. Sementara sentimen negatif cenderung muncul ketika Manchester United kalah dalam pertandingan atau ketika pemain bermain buruk dan berperilaku indisipliner.

**Kata Kunci :** Analisis Sentimen, *Convolutional Neural Network*, *Naive Bayes*, Manchester United

*Abstract*

**COMPARISON OF CONVOLUTIONAL NEURAL NETWORK AND NAÏVE BAYES ALGORITHM PERFORMANCE ON MANCHESTER UNITED PERFORMANCE SENTIMENT ANALYSIS ON TWITTER**

*By*

**Muhammad Naufal Humam**

*Sentiment analysis is used to see the tendency of a sentiment or opinion, whether the opinion tends to be positive or negative. In recent research, data is usually obtained from social media, including Twitter, where users often give their personal opinions on certain subjects. Manchester United is a club that is often discussed on Twitter regarding its performance in football, both in good performance and in bad performance. This study compares two algorithms, namely Convolutional Neural Network and Naive Bayes in sentiment analysis using Manchester United sentiment data from Twitter. The data used in this study are in Indonesian and English. Based on the results of the study, the Convolutional Neural Network in English dataset shows the best performance with an accuracy value of 94% compared to a Naive Bayes value of 79%. In the Indonesian language dataset, the Convolutional Neural Network shows the best performance with an accuracy value of 91% compared to the Naive Bayes value of 75%. For sentiment about Manchester United, positive sentiment tends to arise when the result of the match is won by Manchester United, and in English and Indonesian, positive sentiment is filled with words of praise. Meanwhile, negative sentiment tends to arise when Manchester United loses a match or when a player plays badly and behaves indisciplinately.*

**Keywords :** *Sentiment Analysis, Convolutional Neural Network, Naïve Bayes, Manchester United*

**PERBANDINGAN KINERJA ALGORITME *CONVOLUTIONAL NEURAL NETWORK* DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN PERFORMA *MANCHESTER UNITED* DI TWITTER**

**Oleh**

**MUHAMMAD NAUFAL HUMAM**

**Skripsi**

**Sebagai Salah Satu Syarat untuk Mencapai Gelar  
SARJANA ILMU KOMPUTER**

**Pada**

**Jurusan Ilmu Komputer  
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG**

**2023**

Judul Skripsi : **PERBANDINGAN KINERJA ALGORITME  
CONVOLUTIONAL NEURAL NETWORK DAN  
NAÏVE BAYES PADA ANALISIS SENTIMEN  
PERFORMA MANCHESTER UNITED DI  
TWITTER**

Nama : Muhammad Naufal Humam

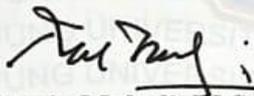
Nomor Pokok Mahasiswa : 1617051115

Program Studi : Ilmu Komputer

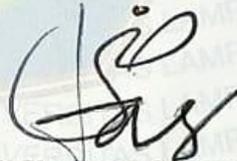
Fakultas : Matematika dan Ilmu Pengetahuan Alam

**MENYETUJUI**

1. Komisi Pembimbing

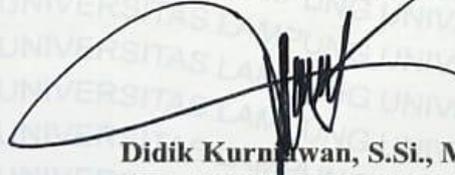


**Dr. Ir. Kurnia Muludi, M.S.Sc**  
NIP. 196406161989021001



**Dewi Asiah Shofiana, S.Komp., M.Kom**  
NIP. 199509292020122030

2. Ketua Jurusan Ilmu Komputer,

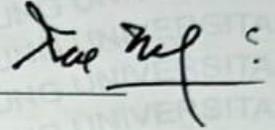


**Didik Kurniawan, S.Si., MT**  
NIP. 198004192005011004

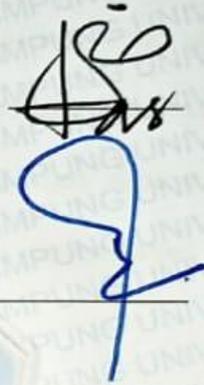
**MENGESAHKAN**

1. Tim Penguji

Ketua : **Dr. Ir. Kurnia Muludi, M.S.Sc.**



Sekretaris : **Dewi Asiah Shofiana, S.Komp., M.Kom.**



Penguji  
Bukan Pembimbing : **Prof. Admi Syarif, P.hD.**

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



**Dr. Eng. Heri Satria, S.Si., M.Si.,**  
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : 08 Juni 2023

## PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul **“Perbandingan Kinerja Algoritme Convolutional Neural Network dan Naïve Bayes pada Analisis Sentimen Performa Mancheser United di Twitter”** ini merupakan hasil karya sendiri dan bukan hasil karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar akademik yang telah saya terima.

Bandar Lampung, 08 Juni 2023



Muhammad Naufal Humam  
1617051115

## RIWAYAT HIDUP



Penulis dilahirkan pada 1 April 1998 di Kabupaten Lampung Timur, merupakan putra pertama dari pasangan Bapak Imam Kambali dan Ibu Khoirul Liummah. Penulis menyelesaikan pendidikan formal pertamanya di Taman Kanak-Kanak (TK) Daarul Huda Teluk Dalem pada tahun 2003, melanjutkan sekolah dasar di SDIT Baitul Muslim Way Jepara dan selesai pada tahun 2009. Kemudian, penulis melanjutkan pendidikan sekolah menengah pertama di SMPIT Baitul Muslim Way Jepara yang diselesaikan pada tahun 2012, serta menyelesaikan pendidikan sekolah menengah atas pada tahun 2015 di SMAIT Baitul Muslim Way Jepara dengan Jurusan Ilmu Pengetahuan Alam.

Pada tahun 2016, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dengan jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Selama menjadi mahasiswa, penulis mengikuti beberapa kegiatan antara lain:

1. Anggota KMB BIROHMAH Universitas Lampung periode 2016/2017.
2. Ketua Bidang Kajian ROIS FMIPA Universitas Lampung periode 2017.
3. Pada bulan Januari 2017 penulis melaksanakan karya wisata ilmiah di Desa Margosari, Kecamatan Pagelaran Utara, Kabupaten Pringsewu
4. Penulis pernah menjabat sebagai Ketua Badan Bimbingan Baca Qur'an ROIS Fakultas Matematika dan Ilmu Pengetahuan Alam pada tahun 2018.
5. Penulis juga pernah menjadi anggota departemen MTQ dan Seni Islam di BIROHMAH Universitas Lampung 2019.

6. Pada bulan Januari 2019 penulis melaksanakan kerja praktik di PT. Andaglos Global Teknologi.
7. Pada bulan Juni 2019 penulis melaksanakan KKN di Desa Paku, Kecamatan Kelumbayan, Kabupaten Tanggamus.
8. Pada bulan Oktober 2020 penulis mengikuti ujian sertifikasi yang diselenggarakan oleh BNSP dengan kualifikasi atau kompetensi sebagai *Junior Web Developer*.

## MOTTO

“Ilmu pengetahuan lebih baik dari harta. Karena pengetahuan akan menjagamu, sementara harta harus engkau jaga.”

( Ali Bin Abi Thalib)

*“Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning.”*

(Albert Einstein)

## SANWACANA

Puji syukur kehadirat Allah Subhanahu wa ta'ala yang telah melimpahkan rahmat, hidayah serta inayah-Nya sehingga penulis dapat menyelesaikan skripsi ini berjudul “Perbandingan Kinerja Algoritme *Convolutional Neural Network* dan *Naive Bayes* pada Analisis Sentimen Performa Manchester United di Twitter”. Tidak lupa shalawat serta salam kepada Nabi Muhammad Shallahu ‘alaihi wasallam, yang kita nanti- nantikan syafaatnya di yaumul akhir kelak.

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dan memiliki peran besar dalam penyusunan skripsi ini, yaitu:

1. Kedua orang tuaku tercinta, Bapakku Imam Kambali dan ibuku Khoirul Liummah yang selalu memberikan semua dukungan, kasih sayang dan segala doa yang tiada hentinya.
2. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc., selaku dosen pembimbing utama atas kesediaannya dan kesabarannya untuk memberikan dukungan, bimbingan, kritik, dan saran dalam proses penyelesaian skripsi.
3. Ibu Dewi Asiah Shofiana, S.Komp., M.Kom., selaku dosen pemimbing kedua yang telah memberikan saran dan masukan guna penyempurnaan penulisan skripsi.
4. Bapak Prof. Admi Syarif, Ph.D, selaku dosen pembahas atas kesediaannya dalam memberikan nasihat, kritik yang membangun juga masukan, dan saran selama proses pengerjaan skripsi.
5. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
6. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc., selaku Sekretaris Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
7. Bapak dan Ibu Dosen Jurusan Ilmu Komputer Universitas Lampung yang telah memberikan ilmu dan pengetahuan hidup selama penulis menjadi mahasiswa.

8. Bapak Dekan Dr. Eng. Heri Satria, S.Si., M.Si., M.T., selaku Dekan FMIPA Universitas Lampung.
9. Ibu Nora Ilmu Komputer yang selalu membantu pemberkasan dan informasi selama perkuliahan.
10. BIROHMAH UNILA dan ROIS FMIPA UNILA yang telah memberikan pengalaman ilmu dunia maupun akhirat dalam proses skripsi ini.
11. Teman-teman grup “Mbah Yang Sama”, Ahmad Taqiyudin, Hikmawan Saputra, M Razif Rizqullah, dan Ariston Rais Zidane yang telah memberikan motivasi serta semangat dalam menyusun skripsi ini, meskipun kadang yang diberi semangat dan motivasi tidak semangat dan termotivasi.
12. Robin, Irfan, Hanggara, Augusto, Akbar, Pahlevi, Hikmawan, Raul, Ghani, Pujo, Fikri, dan teman-teman dari Queens Harbor yang telah memberikan dukungan dan tempat untuk berproyek serta doa terbaik dalam menyusun skripsi ini.
13. Teman-teman Ilmu Komputer 2016 yang selalu kebersamai selama bimbingan skripsi ini berjalan.
14. Semua pihak yang secara langsung maupun tidak langsung yang telah membantu menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, akan tetapi semoga skripsi ini dapat membawa manfaat dan keberkahan bagi perkembangan ilmu pengetahuan terutama bagi semua civitas Ilmu Komputer Universitas Lampung.

Bandar Lampung, 08 Juni 2023

Muhammad Naufal Humam  
1617051115



## DAFTAR ISI

<b>DAFTAR GAMBAR</b> .....	viii
<b>DAFTAR TABEL</b> .....	ix
<b>I. PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan .....	5
1.5 Manfaat .....	5
<b>II. TINJAUAN PUSTAKA</b> .....	6
2.1 <i>Convolutional Neural Network (CNN)</i> .....	6
2.2 Naïve Bayes .....	9
2.3 <i>Text Mining</i> .....	10
2.4 Analisis Sentimen .....	11
2.5 Praproses Teks .....	12
2.6 Ketepatan Klasifikasi .....	13
2.7 Twitter .....	15
2.8 Manchester United .....	15
2.9 Penelitian Terkait .....	16
<b>III. METODOLOGI PENELITIAN</b> .....	19
3.1 Waktu dan Tempat Penelitian .....	19

3.2	Alat dan Bahan Penelitian.....	19
3.3	Tahapan Penelitian.....	20
3.3.1	Analisis Permasalahan .....	21
3.3.3	Pengumpulan Data .....	21
3.3.4	Pengolahan <i>Dataset</i> .....	21
3.3.5	Klasifikasi Menggunakan Algoritme CNN dan <i>Naïve Bayes</i> .....	22
3.3.6	Evaluasi.....	23
<b>IV. HASIL DAN PEMBAHASAN.....</b>		<b>25</b>
3.4	Pengumpulan Data .....	25
3.5	Pembagian Data .....	27
3.6	Praproses Data.....	28
3.7	<i>Training</i> Klasifikasi .....	30
3.8	Evaluasi Hasil Klasifikasi .....	37
<b>V. KESIMPULAN DAN SARAN .....</b>		<b>53</b>
5.1	Kesimpulan .....	53
5.2	Saran .....	54
<b>DAFTAR PUSTAKA.....</b>		<b>56</b>

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Sliding Window.....	7
Gambar 2. ID Convolution.....	8
Gambar 3. Max pooling dan average pooling.....	9
Gambar 4. Diagram alir penelitian.....	20
Gambar 5. Flowchart implementasi algoritme CNN .....	23
Gambar 6. Hasil Normalisasi Teks Berbahasa Inggris .....	29
Gambar 7. Hasil Normalisasi Teks Berbahasa Indonesia .....	29
Gambar 8. Contoh Hasil Stopword Berbahasa Indonesia.....	30
Gambar 9. Contoh Hasil Stopword Berbahasa Inggris .....	30
Gambar 10. Model Sequential Summary .....	31
Gambar 11. Grafik Training CNN <i>Dataset</i> Berbahasa Inggris <i>Epoch</i> 8 .....	34
Gambar 12. Grafik Training CNN <i>Dataset</i> Berbahasa Indonesia <i>Epoch</i> 8 .....	34
Gambar 13. Grafik Training CNN <i>Dataset</i> Berbahasa Inggris <i>Epoch</i> 10 .....	35
Gambar 14. Grafik Training CNN <i>Dataset</i> Berbahasa Indonesia <i>Epoch</i> 10 .....	35
Gambar 15. Grafik Training CNN <i>Dataset</i> Berbahasa Inggris <i>Epoch</i> 12 .....	36
Gambar 16 Grafik Training CNN <i>Dataset</i> Berbahasa Indonesia <i>Epoch</i> 12 .....	36

## DAFTAR TABEL

Tabel	Halaman
Tabel 1 <i>Confusion Matrix</i> .....	14
Tabel 2. Penelitian Terdahulu .....	17
Tabel 3. Alat Penelitian.....	19
Tabel 4. Contoh Dataset Berbahasa Inggris.....	26
Tabel 5. Contoh Dataset Berbahasa Indonesia.....	27
Tabel 6 Pembagian data berbahasa inggris .....	28
Tabel 7 Pembagian data berbahasa indonesia.....	28
Tabel 8. <i>Confusion Matrix CNN Epoch 8</i> dataset berbahasa inggris. ....	37
Tabel 9. <i>Confusion Matrix CNN Epoch 10</i> dataset berbahasa inggris. ....	38
Tabel 10. <i>Confusion Matrix CNN Epoch 12</i> dataset berbahasa inggris. ....	38
Tabel 11. <i>Confusion Matrix Naïve Bayes</i> dataset berbahasa inggris.....	38
Tabel 12. <i>Confusion Matrix CNN Epoch 8</i> dataset berbahasa Indonesia. ....	39
Tabel 13. <i>Confusion Matrix CNN Epoch 10</i> dataset berbahasa Indonesia. ....	39
Tabel 14. <i>Confusion Matrix CNN Epoch 12</i> dataset berbahasa Indonesia. ....	39
Tabel 15. <i>Confusion Matrix Naïve Bayes</i> dataset berbahasa Indonesia. ....	40
Tabel 16. <i>Training Data</i> Berbahasa Inggris <i>CNN Epoch 8</i> .....	41
Tabel 17. <i>Training Data</i> Berbahasa Indonesia <i>CNN Epoch 8</i> .....	42
Tabel 18. <i>Training Data</i> Berbahasa Inggris <i>CNN Epoch 10</i> .....	43
Tabel 19. <i>Training Data</i> Berbahasa Indonesia <i>CNN Epoch 10</i> .....	44
Tabel 20. <i>Training Data</i> Berbahasa Inggris <i>CNN Epoch 12</i> .....	45
Tabel 21. <i>Training Data</i> Berbahasa Indonesia <i>CNN Epoch 12</i> .....	46
Tabel 22. <i>Training Data Naïve Bayes</i> Berbahasa Inggris. ....	48

Tabel 23 <i>Training Data</i> Naïve Bayes Berbahasa Indonesia. ....	49
Tabel 24. Hasil Perbandingan dan Perhitungan Klasifikasi CNN <i>Dataset</i> Berbahasa Inggris dan Indonesia .....	51
Tabel 25. Hasil Perbandingan dan Perhitungan Klasifikasi CNN <i>Dataset</i> Berbahasa Inggris dan Indonesia. ....	51

## I. PENDAHULUAN

### 1.1 Latar Belakang

Teknologi informasi memegang peranan di berbagai sendi kehidupan manusia karena dapat menghubungkan dan menyajikan berbagai informasi melalui *web*. *Web* atau *website* merupakan halaman informasi yang disediakan melalui jalur internet, sehingga dapat diakses di seluruh dunia selama terkoneksi dengan jaringan internet. *Web* memuat dua tipe tekstual yaitu fakta dan opini (Buntoro et al., 2014). Fakta adalah suatu keadaan yang benar terjadi di dunia, sesuatu tersebut dapat dinyatakan fakta jika dapat dibuktikan kebenarannya dan memiliki sumber yang jelas, sedangkan opini adalah sikap, pandangan, atau reaksi seseorang terhadap suatu fakta dan memiliki nilai kebenaran yang relatif, karena secara subjektif dipengaruhi unsur-unsur pribadi. Opini disebut juga gagasan atau argumentasi. *Web* telah membagikan berbagai fakta dan opini tentang banyak hal melalui blog pribadi, jejaring sosial, dan lain-lain sehingga ketika seseorang atau sekelompok organisasi ataupun perusahaan yang ingin mengetahui opini publik tentang sesuatu kejadian, hal, dan sebagainya, maka penggunaan data yang terdapat dalam *web* dapat menjadi alternatif yang efisien selain menggunakan survei konvensional.

Twitter adalah layanan mikro blog populer yang memungkinkan pengguna membuat pesan status yang disebut juga sebagai *tweet*. *Tweet* berisi banyak ekspresi manusia, suka atau tidak suka dan setuju atau tidak setuju terhadap

topik yang sedang dibicarakan (Shelar dan Huang, 2018). Twitter adalah *platform* penting dalam komunikasi digital saat ini, setiap orang dapat bergabung berpartisipasi untuk berbagi segalanya. Sosial media memberikan kemudahan dalam berkomunikasi dan memiliki kemampuan untuk menyajikan komunikasi dua arah (Fuchs et al., 2010). Hal tersebut memudahkan suatu perusahaan untuk berinteraksi dengan pelanggan. Melalui sosial media perusahaan dapat membagikan sesuatu dan memperoleh kritik, masukan ataupun saran, yang positif maupun negatif dari pelanggan, dengan cara yang sangat mudah dan cepat.

Analisis sentimen merupakan serangkaian cara, teknik, dan alat yang bertujuan untuk melakukan deteksi dan mengekstrak informasi yang subjektif seperti opini dari suatu tulisan (Mäntylä et al., 2018). Analisis sentimen digunakan untuk melihat kecenderungan suatu sentimen atau pendapat, apakah pendapat tersebut cenderung beropini positif atau negatif.

Sepakbola merupakan olahraga yang digemari banyak orang. Pendukung atau fansnya datang dari berbagai kalangan. Kecintaan para pendukung sepakbola tidak mengenal umur, ras, suku, agama, golongan, dan jenis kelamin. Tidak hanya mendukung klub lokal saja, banyak dari fans ini mendukung klub-klub dari liga papan atas Eropa seperti Manchester United, Juventus, dan Barcelona. Fans adalah mereka yang memberikan dukungan dari jauh dan biasanya menggunakan media dalam mendukung dan menyaksikan klub kesayangannya bertanding karena keterbatasan geografis maupun ekonomi (Giulianotti, 2018). Dalam bidang olahraga sepakbola,

analisis sentimen dapat digunakan untuk melakukan analisis argumen pendukung suatu klub sepak bola terhadap performa klub tersebut.

Manchester United *Football Club* adalah sebuah klub sepak bola profesional Inggris yang berbasis di Old Trafford, Manchester Raya, yang bermain di Liga Inggris. Perusahaan akuntansi dan industri olahraga konsultan Deloitte memperkirakan bahwa Manchester United memiliki 75 juta penggemar di seluruh dunia, sementara perkiraan lain menempatkan angka ini mendekati 333 juta (Sigalingging, 2018). Dengan jumlah penggemar yang sangat banyak, Manchester United menjadi klub yang sering dibicarakan mengenai performanya dalam permainan sepakbola, baik ketika dalam performa yang bagus ataupun performa yang buruk.

Terdapat banyak metode klasifikasi dalam ilmu statistika yang digunakan untuk analisis sentimen namun metode yang sering digunakan dalam klasifikasi teks adalah metode *Naïve Bayes Classifier* (NBC). Selain menggunakan *naïve bayes*, algoritme *deep learning* juga digunakan untuk mengklasifikasikan teks, salah satunya adalah algoritme *convolutional neural network* (CNN). Dalam penelitian analisis sentimen terhadap film melalui media twitter, akurasi tertinggi diperoleh algoritme CNN sebesar 80,99% (Ratnawati & Winarko, 2018). Falahah dan Nur (2015) melakukan penelitian mengenai pengembangan aplikasi *sentiment analysis* menggunakan metode *Naïve Bayes*. Penelitian tersebut menggunakan praproses teks *case folding*, *parsing*, dan transformasi sehingga mendapatkan hasil akurasi klasifikasi 73%. Selain itu, penelitian lain yang berjudul *Indonesian news classification using convolutional neural network*

yang juga menggunakan metode CNN menyimpulkan bahwa algoritme CNN tidak hanya digunakan untuk memproses data citra tetapi juga dapat digunakan untuk data teks. CNN dapat melakukan klasifikasi berita Indonesia dengan rata-rata hasil pengujian sekitar 90,74% (Ramdhani et al., 2020).

Berdasarkan hasil dari sejumlah penelitian terdahulu maka dalam penelitian ini bertujuan untuk membandingkan dua algoritme, yaitu CNN, dan juga algoritme yang sudah sangat sering digunakan dalam klasifikasi, *naive bayes* untuk mengetahui kinerja algoritme dengan tingkat efektivitas, efisiensi, dan akurasi paling baik untuk mengklasifikasikan sentimen para pendukung Manchester United.

## **1.2 Rumusan Masalah**

Rumusan masalah yang akan diselesaikan pada penelitian ini adalah bagaimana mengklasifikasikan sentimen menggunakan algoritme *Convolutional Neural Network* dan *Naïve Bayes* serta mengetahui sejauh mana perbedaan akurasi antara algoritme dalam mengklasifikasikan data sentimen performa Manchester United.

## **1.3 Batasan Masalah**

Batasan masalah penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan studi kasus performa Manchester United pada musim 2022/2023.
2. Klasifikasi sentimen data awal ditentukan secara subjektif.

3. Penelitian tidak memperhatikan latar belakang atau demografi dari pemilik akun Twitter.
4. *Dataset* pada penelitian ini menggunakan dua bahasa, yaitu bahasa Inggris dan bahasa Indonesia.

#### **1.4 Tujuan**

Adapun tujuan dari penelitian ini adalah mendapatkan hasil perbandingan dari kedua algoritme berupa tingkat ketepatan atau akurasi klasifikasi dan menentukan algoritme mana yang lebih baik tingkat akurasinya.

#### **1.5 Manfaat**

Manfaat dari penelitian ini adalah sebagai berikut :

1. Mengetahui algoritme mana yang memiliki akurasi lebih tinggi dalam analisis sentimen pada kasus ini dan menguji akurasi, presisi, *recall*, dan *f1-score* algoritme *Convolutional Neural Network* dan *Naive Bayes* pada analisis sentimen dan melakukan perbandingan antara keduanya.
2. Memberikan tambahan informasi kepada pembaca kata apakah yang biasanya digunakan untuk sentimen positif maupun sentimen negatif dan pada situasi apa sentimen tersebut cenderung kepada sentiment positif atau negatif.

## II. TINJAUAN PUSTAKA

### 2.1 *Convolutional Neural Network (CNN)*

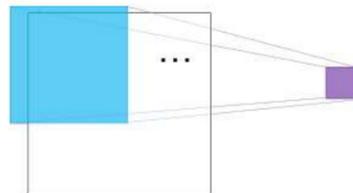
Menurut Eka Putra *Convolutional Neural Network (CNN)* adalah pengembangan dari *Multilayer Perceptron (MLP)* yang didesain untuk mengolah data dua dimensi. CNN termasuk dalam jenis Deep Neural Network karena kedalaman jaringan yang tinggi (Putra, 2016). CNN pada umumnya digunakan pada klasifikasi dua dimensi, video atau gambar namun CNN juga dapat digunakan untuk mengklasifikasikan teks (Kim, 2014). Kemampuan utama dari CNN adalah arsitektur yang mampu mengenali informasi prediktif dari suatu objek seperti teks, potongan suara, dan gambar.

CNN merupakan salah satu metode penerapan *deep learning* yang berbeda dengan metode *machine learning*. Dilihat dari waktu eksekusinya, model-model *deep learning* membutuhkan waktu yang lebih lama daripada *machine learning* karena penghitungan bobot yang besar dan adanya parameter tambahan. Selain itu dalam hal perangkat keras, *deep learning* membutuhkan spesifikasi yang lebih untuk menghitung operasi matriks dalam jumlah besar. *Machine learning* cocok untuk data yang kecil atau data yang jumlahnya tidak terlalu banyak, sedangkan *deep learning* bagus diterapkan untuk data yang lebih besar atau banyak.

CNN dibagi menjadi 2 bagian besar yaitu, *feature extraction layer* dan *fully connected layer*. *Feature extraction layer* terdiri dari dua bagian yaitu *convolutional layer* dan *pooling layer*.

### 2.1.1 Convolutional Layer

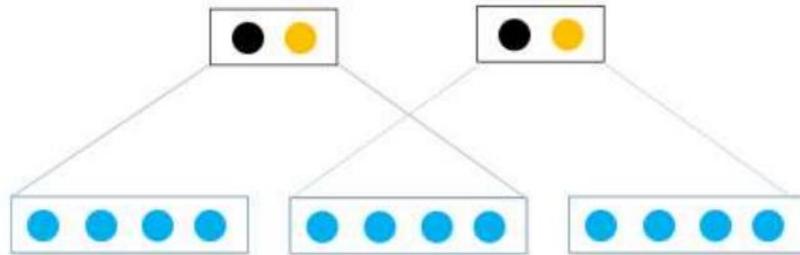
*Convolutional layer* menggunakan prinsip *sliding window* dan *weight sharing* yang bertujuan untuk mengurangi kompleksitas perhitungan. Operasi pada *window* digunakan untuk mengetahui aspek lokal yang paling informatif atau dikenal dengan *filter* yang dapat dikenali oleh *window*. Seperti yang disajikan pada Gambar 1, warna biru merepresentasikan *window* dan warna ungu merepresentasikan *filter*. *Window* digeser sebanyak  $T$  yang dapat menghasilkan vektor dengan ukuran tertentu. *Window* ditransformasikan menjadi suatu nilai numerik.



Gambar 1. *Sliding Window*

Dalam CNN 1 dimensi atau CNN 1-D Lingkaran warna biru merepresentasikan *feature vector* dari suatu input, disajikan seperti Gambar 2. Setiap dua input ditransformasikan menjadi dua dimensi atau *channel* yang dapat menghasilkan vektor dengan 4 dimensi ( $2 \text{ window} \times 2$ ). Suatu input  $x$  dapat menggunakan stride sebesar  $s$

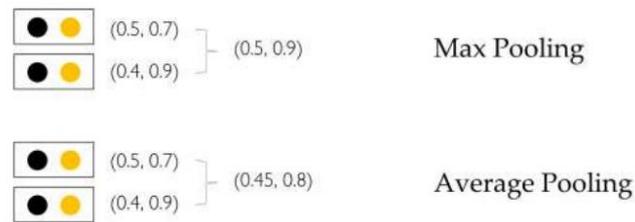
untuk menentukan seberapa banyak data dapat digeser untuk *window* baru.



Gambar 2. 1D Convolution

### 2.1.2 Pooling Layer

*Pooling layer* berada setelah *convolutional layer* yang digunakan untuk merangkum informasi yang dihasilkan dari *convolutional layer*. Vektor-vektor yang dihasilkan kemudian dikombinasikan (*pooled*) menjadi vektor baru. *Pooling* yang biasa digunakan adalah *max pooling* dan *average pooling* yang diilustrasikan seperti Gambar 3. *Max pooling* mengambil nilai terbesar dari lingkaran hitam dan mengambil nilai terbesar dari lingkaran jingga. *Average pooling* mengambil nilai rata-rata dari lingkaran hitam dan rata-rata lingkaran jingga untuk menghasilkan nilai baru. Tujuan dari penggunaan *pooling layer* adalah untuk mengurangi dimensi dari *feature map* (*downsampling*), sehingga mempercepat komputasi karena parameter yang harus diupdate semakin sedikit dan mengatasi *overfitting*.



Gambar 3. *Max pooling* dan *average pooling*

Setelah melalui *convolutional layer* dan *pooling layer*, *feature map* yang dihasilkan masih berbentuk multidimensional *array*, sehingga perlu dilakukan *flatten* atau *reshape feature map* menjadi sebuah vektor agar bisa kita gunakan sebagai input dari *fully connected layer*. *Fully connected layer* sama dengan *multi layer perceptron* yang memiliki *hidden layer*, *activation function*, *output layer*, dan *loss function*.

## 2.2 Naïve Bayes

Algoritme *Naïve Bayes* merupakan algoritme yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger., 2007). Metode klasifikasi *naive bayes* merupakan metode yang dapat digunakan untuk mengklasifikasikan teks. Kelebihan Algoritme *Naïve Bayes* adalah algoritmenya sederhana tetapi memiliki akurasi yang tinggi. Terdapat dua tahap dalam klasifikasi tweet. Tahap pertama adalah pelatihan terhadap tweet yang telah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi tweet yang belum diketahui kategorinya (Falahah & Nur, 2015).

*Naive Bayes* memiliki persamaan seperti berikut ini.

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$  = Probalitas hipotesis H berdasarkan kondisi x

$P(H)$  = Probalitas hipotesis H

$P(X|H)$  = Probalitas X berdasarkan kondisi tersebut

$P(X)$  = Probalitas dari X

*Naive Bayes* memiliki tingkat kesalahan paling sedikit dibandingkan dengan metode klasifikasi lainnya. Namun, dalam prakteknya hal ini tidak selalu terjadi, karena ketidakakuratan asumsi yang dibuat untuk penggunaannya, seperti kondisi kelas independen, dan kurangnya data probabilitas yang tersedia. Pengklasifikasi bayesian juga berguna dalam memberikan pembenaran teoritis untuk pengklasifikasi lain yang tidak secara eksplisit menggunakan teorema *bayes*.

### **2.3 Text Mining**

*Text Mining* adalah penggalian data untuk menyelesaikan masalah kebutuhan informasi dengan menerapkan teknik *data mining*, *machine learning*, *natural language processing*, pencarian informasi, dan manajemen pengetahuan. *Text mining* melibatkan praproses dokumen seperti kategorisasi teks, ekstraksi informasi, dan ekstraksi kata. Metode ini digunakan untuk mengekstraksi informasi dari sumber data melalui

identifikasi dan eksplorasi pola yang menarik (Feldman & Sanger, 2007). *Text Mining* merupakan teknik yang digunakan untuk menangani permasalahan klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Berry & Kogan, 2010). Pada dasarnya proses kerja dari *Text Mining* banyak mengadopsi dari penelitian *data mining* namun yang menjadi perbedaan adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alami yang tidak terstruktur sedangkan dalam *data mining* pola yang diambil dari *database* yang terstruktur. Tahap-tahap *text mining* secara umum adalah praproses teks dan *feature selection*.

#### **2.4 Analisis Sentimen**

Analisis sentimen adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Analisis sentimen sama dengan analisis media sosial. Faktanya, analisis sentimen sekarang berpusat pada media sosial. Oleh karena itu, penelitian dalam analisis sentimen tidak hanya berdampak pada penggunaan NLP saja, tetapi mungkin juga berdampak besar pada ilmu manajemen, politik, sains, ekonomi, dan ilmu sosial karena semuanya dipengaruhi oleh pendapat orang (Liu, 2012). Tugas dasar analisis sentimen adalah mengklasifikasikan teks dalam suatu kalimat atau dokumen, dan kemudian menentukan apakah pendapat yang diungkapkan dalam kalimat atau dokumen itu positif atau negatif. Analisis sentimen juga dapat menyatakan perasaan emosional sedih, gembira, atau marah. Kita dapat mencari pendapat tentang produk-produk, merek atau

orang-orang dan menentukan apakah mereka dilihat positif atau negatif di web.

Ekspresi atau sentiment mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Oleh karena itu pada beberapa penelitian, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses analisis sentimennya.

## **2.5 Praproses Teks**

Praproses teks merupakan tahap awal dalam pengolahan teks yang digunakan untuk mengubah bentuk dokumen menjadi data yang terstruktur sesuai dengan kebutuhannya, sehingga dapat diproses lebih lanjut dalam proses *text mining*. Tahapan praproses teks dalam klasifikasi bertujuan untuk meningkatkan akurasi klasifikasi data. Tahapan dalam praproses teks adalah sebagai berikut :

### *2.5.1 Case Folding*

*Case Folding* merupakan proses untuk mengubah semua karakter teks menjadi huruf kecil serta menghilangkan tanda baca dan angka. Cara kerja *case folding* adalah memproses huruf alfabet dari “a” hingga “z” saja sehingga karakter selain huruf tersebut akan dihapus (Weiss dkk., 2010).

### *2.5.2 Tokenizing*

*Tokenizing* adalah proses memecah yang semula kalimat menjadi kata-kata atau memutus urutan string menjadi potongan-potongan, seperti kata-kata berdasarkan tiap kata yang menyusunnya.

### 2.5.3 *Stopwords Removal*

*Stopwords* merupakan kosakata yang bukan termasuk kata unik atau ciri pada suatu dokumen atau tidak menyampaikan pesan apapun secara signifikan pada teks atau kalimat (Dragut et al., 2009). Kosakata yang dimaksud yaitu seperti kata penghubung dan kata keterangan yang bukan merupakan kata unik, misalnya “from”, “the”, “or”, dan sebagainya.

### 2.5.4 *Stemming*

*Stemming* adalah proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, dan kombinasi antara awalan dan akhiran.

## 2.6 Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi dilakukan untuk melihat performa klasifikasi yang telah dilakukan. Dalam mengukur ketepatan klasifikasi, perlu diketahui jumlah pada setiap kelas prediksi dan kelas aktual yang terdiri dari TP (*True positif*) yaitu jumlah ulasan bersentimen positif yang tepat terprediksi dalam kelas positif, TN (*True negatif*) yaitu ulasan bersentimen negatif yang tepat terprediksi dalam kelas negatif, FP (*False positif*) yaitu ulasan bersentimen negatif yang terprediksi dalam kelas positif, dan FN (*False Negatif*) yaitu ulasan bersentimen positif yang terprediksi dalam kelas negatif. Berikut merupakan *Confusion Matrix* yang memuat keempat nilai tersebut.

Tabel 1 *Confussion Matrix*

<b>Kelas Aktual</b>	<b>Kelas Prediksi</b>	
	Positif	Negatif
<b>Positif</b>	TP	FN
<b>Negatif</b>	FP	TN

Pengukuran yang sering digunakan untuk menghitung ketepatan klasifikasi adalah akurasi, *specificity*, dan *sensitivity* (Hotho et al., 2005). Akurasi merupakan persentase dokumen yang teridentifikasi secara tepat dari total dokumen dalam proses klasifikasi. Akurasi digunakan untuk menghitung ketepatan klasifikasi sebuah dokumen yang mempunyai data yang seimbang pada tiap kategorinya. Berikut merupakan rumus dalam menghitung akurasi, *specificity* dan *sensitivity*.

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TN + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

## 2.7 Twitter

Twitter adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh Twitter inc, yang menawarkan jaringan sosial berupa microblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan *tweet* (Bilton, 2013). Mikroblog merupakan jenis alat komunikasi online dimana pengguna dapat membagikan aktivitas yang sedang dilakukan ataupun pendapat tentang suatu objek dan fenomena, dengan cara menulis status yang disebut *tweet*. *Tweet* adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. Namun sejak tahun 2020 jumlah batas maksimal karakter twitter sudah ditambah menjadi 280 karakter. Pengguna twitter dapat membatasi *tweet* nya untuk bisa dilihat secara publik ataupun secara privat ke daftar orang orang yang telah ditentukan.

*Trending topic* pada twitter merupakan suatu isu atau fenomena yang sedang banyak diulas oleh para pengguna twitter. Pada twitter juga terdapat *Hashtag* yang digunakan sebagai penanda topik tertentu dan mengindeks kata kunci atau topik tersebut. Fungsi ini dibuat twitter supaya memungkinkan pengguna untuk mengikuti topik yang diminati dengan mudah.

## 2.8 Manchester United

Manchester United *Football Club* adalah sebuah klub sepak bola profesional Inggris yang berbasis di Old Trafford, Manchester Raya, yang bermain di Liga Inggris. Perusahaan akuntansi dan industri olahraga konsultan Deloitte memperkirakan bahwa Manchester United memiliki 75 juta penggemar di seluruh

dunia, sementara perkiraan lain menempatkan angka ini mendekati 333 juta (Sigalingging, 2018). Di antara klub liga Inggris lain Manchester United berada di urutan teratas untuk penggemar paling banyak dengan total lebih dari 91 juta fans di Facebook dan Twitter. Manchester United jauh meninggalkan Liverpool di peringkat dua yang memiliki sekitar 43 juta fans di dua media sosial tersebut.

## **2.9 Penelitian Terkait**

Penelitian terkait dengan *text mining* pada umumnya untuk masalah *sentiment analysis* telah banyak dilakukan sebelumnya dengan metode yang berbeda-beda. Tabel 2 menunjukkan penelitian terdahulu.

Tabel 2. Penelitian Terdahulu

<b>Perbandingan teori</b>	<b>Peneliti (tahun penelitian)</b>	<b>Tujuan</b>	<b>Langkah-langkah</b>	<b>Metode</b>	<b>Output</b>
Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode <i>Naive Bayes</i>	(Falahah dan Nur 2015)	Membangun aplikasi klasifikasi opini yang menerapkan pendekatan <i>Naive Bayes</i> untuk mengklasifikasikan kata-kata dan difokuskan pada tweet dalam bahasa Indonesia	<ol style="list-style-type: none"> <li>Pengumpulan awal data</li> <li>Pengolahan <i>dataset</i> dan data <i>Training</i></li> <li>Klasifikasi data</li> <li>Evaluasi</li> </ol>	Algoritme klasifikasi <i>Naive Bayes</i>	<ul style="list-style-type: none"> <li><i>Naive Bayes</i>: 73%</li> </ul>
Implementasi <i>Text Mining</i> Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan <i>Naive Bayes Classifier</i> Dan <i>Support Vector Machine</i> Terhadap Media Mainstream Menggunakan <i>Naive Machine</i>	(Kurniawan 2017)	Mendapatkan hasil ketepatan klasifikasi sentiment pengguna twitter terhadap media mainstream menggunakan metode <i>Naive Bayes classifier</i> dan SVM, kemudian membandingkan hasil keduanya	<ol style="list-style-type: none"> <li>Pengumpulan data <i>tweet</i></li> <li>Penentuan sentimen secara manual</li> <li>Pembagian data</li> <li><i>Indexing</i> data latih dan data uji</li> <li>Seleksi fitur</li> <li>Fungsi klasifikasi <i>multinomial Naive Bayes</i></li> <li>Evaluasi</li> </ol>	Algoritme klasifikasi <i>Naive Bayes</i> dan SVM	<ul style="list-style-type: none"> <li><i>Naive Bayes</i>: 95.8% dan 97.8%</li> <li>SVM: 97,9% dan 99.3%</li> </ul>

Tabel 2. Penelitian Terdahulu (lanjutan)

<b>Perbandingan teori</b>	<b>Peneliti (tahun penelitian)</b>	<b>Tujuan</b>	<b>Langkah-langkah</b>	<b>Metode</b>	<b>Output</b>
<i>Indonesian news classification using convolutional neural network</i>	(Ramdhani 2020)	Melakukan klasifikasi data teks berita berbahasa indonesia menggunakan algoritme CNN sebagai salah satu algoritme <i>Deep Learning</i>	<ol style="list-style-type: none"> <li>a. Pengumpulan data dan pelabelan</li> <li>b. Eksplorasi analisis data</li> <li>c. Praproses data</li> <li>d. Klasifikasi data</li> <li>e. Evaluasi</li> </ol>	Menggunakan metode CNN	Nilai akurasi CNN: 90.74%

### III. METODOLOGI PENELITIAN

#### 3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada waktu semester genap tahun ajaran 2021/2022. Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang berada di Jalan Soemantri Brodjonegoro No.1 Gedung Meneng, Bandar Lampung.

#### 3.2 Alat dan Bahan Penelitian

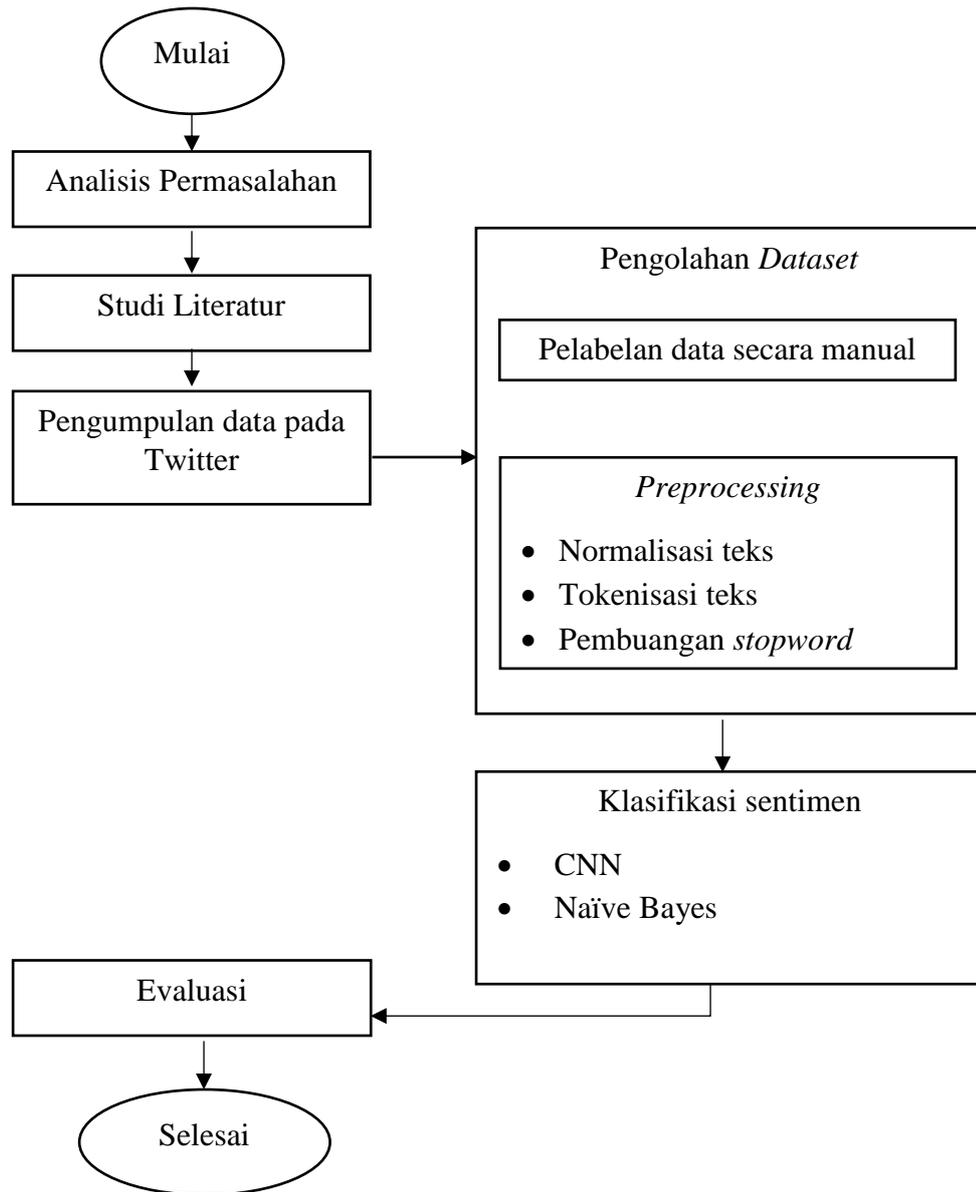
Penelitian perbandingan kinerja algoritme CNN dan *Naive Bayes* ini menggunakan alat yaitu *hardware* dan *software*. Spesifikasi dari alat penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. Alat Penelitian

No.	Alat Penelitian	Spesifikasi
1.	<i>Hardware</i>	
	<i>Processor</i>	AMD A4-5000 APU with Radeon(TM) HD Graphics (4CPU), ~1,5GHz.
	<i>RAM</i>	DDR3 4GB
2.	<i>Software</i>	
	<i>Sistem Operasi</i>	Windows 10 home 64-bit
	<i>Tools</i>	Python 3.9 64-bit <i>using</i> Jupyter Notebook

### 3.3 Tahapan Penelitian

Tahapan penelitian pada penelitian ini dilakukan dengan beberapa langkah yang telah tersaji pada diagram alir penelitian dan dapat dilihat pada Gambar 4



Gambar 4. Diagram alir penelitian

### 3.3.1 Analisis Permasalahan

Pada tahap awal penelitian ini dilakukan analisis permasalahan yaitu membandingkan dua algoritma klasifikasi pada kasus sentiment performa Manchester United yang ada pada media sosial twitter. Algoritme yang dibandingkan adalah *Convolutional Neural Network* (CNN) dan *Naïve Bayes*, untuk nanti diketahui algoritme yang paling baik dalam mengklasifikasikan data tersebut.

### 3.3.2 Studi Literatur

Pada tahap studi literatur dilakukan pengumpulan informasi dari penelitian terdahulu yang berkaitan dengan topik permasalahan pada penelitian ini. Kemudian mempelajari berbagai metode penelitian yang telah digunakan dari literatur yang telah dikumpulkan sebagai pedoman penelitian.

### 3.3.3 Pengumpulan Data

Tahap pertama adalah pengumpulan data. Data yang digunakan dalam penelitian ini adalah data teks dari kumpulan *tweet* mengenai performa Manchester United di seluruh kompetisi sepakbola baik domestik maupun Eropa seperti *Champions League* dan *Europa League*. Kumpulan *tweet* diambil dari *tweet* yg dikirim pada tanggal 1 November 2021 hingga 31 Desember 2021.

### 3.3.4 Pengolahan *Dataset*

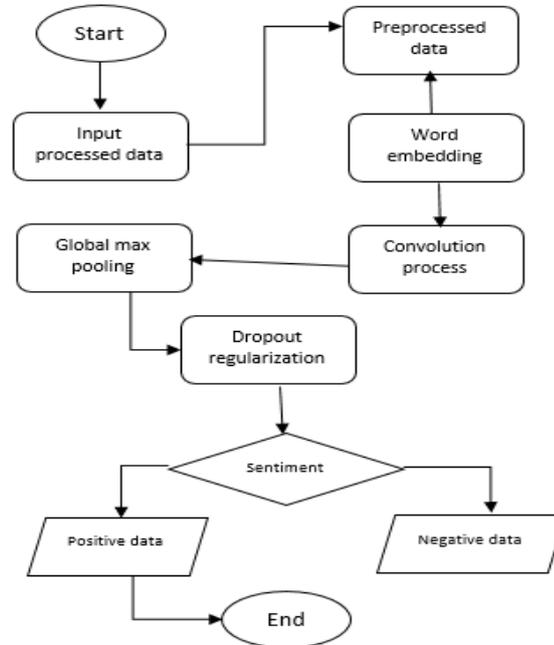
Setelah mengumpulkan data, terlebih dahulu data harus melalui praproses data. Tujuannya untuk mengubah data yang telah didapatkan menjadi bentuk data yang diinginkan sebelum masuk ke tahap implementasi. Beberapa proses pada tahapan ini adalah sebagai berikut :

- a. Menyiapkan data sentimen tentang Manchester United dari Twitter, daftar *Stopwords*, dan kata dasar. Data yang sudah diambil dan disimpan kemudian disiapkan ke praproses teks.
- b. Praproses teks adalah proses mengekstraksi teks untuk kemudian diolah ke proses klasifikasi. Pada proses ini data ulasan yang telah disiapkan kemudian diproses dengan cara menghilangkan tag html. Lalu melakukan *case folding*, mengubah semua teks dengan huruf kecil, serta menghilangkan tanda baca. Setelah itu kata-kata pada ulasan yang terdaftar dalam daftar *Stopword* dihapus. Selanjutnya, dilakukan tokenisasi atau pembobotan pada setiap kata yang ada. Kemudian melakukan *stemming* untuk mendapatkan kata dasar dengan menghilangkan imbuhan.

### 3.3.5 Klasifikasi Menggunakan Algoritme CNN dan *Naïve Bayes*

Pada tahapan ini dilakukan proses klasifikasi, terdapat dua tahapan utama dalam proses ini yaitu tahap implementasi CNN dan *Naïve Bayes*. Data yang sebelumnya telah melalui praproses akan diklasifikasikan ke dalam sentimen positif atau negatif menggunakan algoritme CNN dan *Naïve Bayes*. Selanjutnya data akan diproses menggunakan algoritme CNN yang diantaranya terdapat proses konvolusi dan *max-pooling*. Pada tipe data dua dimensi atau gambar, operasi ini akan diulang sebanyak jumlah *pixel* input gambar. Karena tipe data yang digunakan pada penelitian ini adalah data satu dimensi atau teks, teks harus diubah terlebih dahulu menjadi vektor melalui proses *word embedding*. Setelah proses tersebut data diklasifikasikan menjadi beberapa kelas antara lain kelas positif, kelas

negatif, maupun kelas netral. *Flowchart* untuk implementasi algoritme CNN disajikan pada Gambar 4.



Gambar 5. *Flowchart* implementasi algoritme CNN

Kemudian untuk algoritme *Naïve Bayes*, setelah data melalui praproses, berikutnya data akan dilatih untuk membentuk pengetahuan berupa nilai probabilitas kata pada suatu sentimen baik negatif maupun positif. Kemudian akan dilakukan proses klasifikasi menggunakan Algoritme *Naïve Bayes* terhadap data uji dengan data latih tadi sebagai acuan pengetahuannya. Lalu data juga akan dibagi ke beberapa kelas antara lain kelas positif, dan kelas negatif.

### 3.3.6 Evaluasi

Hasil dari implementasi selanjutnya dilakukan proses pengujian. Pada tahap ini dilihat lima aspek yaitu *accuracy*, *precision*, *recall*, *specificity*, dan *F-*

*measure*. lima aspek tersebut yang akan digunakan untuk melihat keakuratan atau performa kinerja dari metode yang digunakan.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Kesimpulan pada penelitian ini adalah :

1. Hasil klasifikasi dari metode *Convolutional Neural Network* menunjukkan nilai klasifikasi yang cukup baik. Pada *dataset* berbahasa inggris hasil klasifikasi dengan nilai paling tinggi ada pada akurasi sebesar 94%, presisi 94%, *recall* 93%, dan *f1-score* 94%. Pada *dataset* berbahasa Indonesia hasil klasifikasi dengan nilai paling tinggi ada pada akurasi sebesar 91%, presisi 91%, *recall* 91%, dan *f1-score* 91%. Ini menunjukkan pada *dataset* berbahasa inggris kinerja dari metode CNN sedikit lebih baik daripada *dataset* berbahasa Indonesia.
2. Untuk perbandingan hasil klasifikasi antara metode CNN dengan *Naïve Bayes* memiliki nilai klasifikasi yang berbeda. Pada *dataset* berbahasa inggris CNN lebih unggul dibandingkan *Naïve Bayes* dengan akurasi 94%, presisi 94%, *recall* 93%, dan *f1-score* 94%, dan *Naïve Bayes* dengan akurasi 79%, presisi 87%, *recall* 71%, dan *f1-score* 72%. Pada *dataset* berbahasa Indonesia CNN juga lebih unggul dibandingkan *Naïve Bayes* dengan akurasi 91%, presisi 91%, *recall* 91%, dan *f1-score*

91%, dan *Naïve Bayes* dengan akurasi 75%, presisi 85%, *recall* 68%, dan *f1-score* 68%.

3. Untuk sentimen soal Manchester United, sentimen positif cenderung muncul ketika hasil pertandingan dimenangkan oleh Manchester United, dan pada sentimen berbahasa Inggris maupun Indonesia sentimen positif banyak diisi oleh kata-kata pujian seperti *great*, *amazing*, menang, main bagus, *good game*, dan *well done*. Sementara sentimen negatif cenderung muncul ketika Manchester United kalah dalam pertandingan atau ketika pemain bermain buruk dan berperilaku indisipliner. Kata-kata yang sering muncul pada sentiment negatif baik pada sentimen berbahasa Inggris maupun Indonesia adalah ekspresi kekesalan seperti *worst defense*, *undeserved*, *problem*, sampah banget, dan diobok-obok.

## 5.2 Saran

Terdapat beberapa hal yang perlu ditambahkan atau diperbaiki untuk penelitian selanjutnya, yaitu:

1. Data *tweet* yang digunakan dapat diperbanyak jumlahnya, agar data latih yang digunakan dapat lebih banyak dan menyetarakan kelas data positif dan negatif sehingga sistem klasifikasi dapat memiliki nilai akurasi yang lebih baik.
2. Pada data berbahasa Indonesia dapat ditambahkan dengan perbaikan kata tidak baku, karena data *tweet* dari pengguna Twitter di Indonesia

masih banyak *tweet* yang bahasanya bercampur dengan bahasa inggris atau bahasa daerah.

3. Dapat mengembangkan analisis sentimen ini dengan mengkombinasikan metode *Deep Learning* yang lain selain CNN seperti *Deep Neural Networks*, dan *Artificial Neural Network* untuk dilakukan pengujian klasifikasi pada data teks, dan mendapatkan nilai akurasi yang lebih baik.

## DAFTAR PUSTAKA

- Afaq, S., & Rao, S. (2020). Significance Of *Epochs* On Training A Neural Network. *International Journal of Scientific & Technology Research*, 9(06), 485–488. [www.ijstr.org](http://www.ijstr.org)
- Berry, M. W., & Kogan, J. (2010). Text Mining: Applications and Theory. In *Text Mining: Applications and Theory*. <https://doi.org/10.1002/9780470689646>
- Bilton, N. (2013). *Hatching Twitter: A True Story of Money, Power, Friendship and Betrayal*. Hodder & Stoughton.  
<https://books.google.co.id/books?id=jXgPAAAAQBAJ>
- Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. *Citee*.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*.
- Dragut, E., Fang, F., Sistla, P., Yu, C., & Meng, W. (2009). Stop word and related problems in web interface integration. *Proceedings of the VLDB Endowment*, 2(1), 349–360. <https://doi.org/10.14778/1687627.1687667>
- Eka Putra, W. S. (2016). Klasifikasi Citra Menggunakan Convolutional Neural

Network (CNN) pada Caltech 101. *Jurnal Teknik ITS*, 5(1).

<https://doi.org/10.12962/j23373539.v5i1.15696>

Eliane Birba, D. (2020). A Comparative study of data splitting algorithms for machine learning model selection. *Degree Project in Computer Science and Engineering*, 1–23. [https://www.diva-](https://www.diva-portal.org/smash/get/diva2:1506870/FULLTEXT01.pdf)

[portal.org/smash/get/diva2:1506870/FULLTEXT01.pdf](https://www.diva-portal.org/smash/get/diva2:1506870/FULLTEXT01.pdf)

Falahah, & Nur, D. D. A. (2015). Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes. *Seminar Nasional Sistem Informasi Indonesia, November*, 335–340.

Feldman, R., Sanger, J., & Press, C. U. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. [https://books.google.co.id/books?id=U3EA\\_zX3ZwEC](https://books.google.co.id/books?id=U3EA_zX3ZwEC)

Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical Foundations of the Web: Cognition, Communication, and Co-Operation. Towards an Understanding of Web 1.0, 2.0, 3.0. *Future Internet*, 2(1), 41–59. <https://doi.org/10.3390/fi2010041>

Giulianotti, R. (2018). Supporters, followers, fans, and flâneurs: A taxonomy of spectator identities in football. In *Sport, Power, and Society: Institutions and Practices: A Reader*. <https://doi.org/10.4324/9780429497216>

Hotho, A., Nürnberger, A., & Paass, G. (2005). A Brief Survey of Text Mining. *LDV Forum*, 20, 19–62.

Kim, Y. (2014). Convolutional neural networks for sentence classification.

*EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1746–1751.

<https://doi.org/10.3115/v1/d14-1181>

Kurniawan, T. (2017). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine. *IT Journal*, 23, 1.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*.

<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. In *Computer Science Review*. <https://doi.org/10.1016/j.cosrev.2017.10.002>

Ramdhani, M. A., Ramdhani, M. A., Maylawati, D. S. adillah, & Mantoro, T. (2020). Indonesian news classification using convolutional neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 1000–1009. <https://doi.org/10.11591/ijeecs.v19.i2.pp1000-1009>

Ratnawati, F., & Winarko, E. (2018). Sentiment Analysis of Movie Opinion in Twitter Using Dynamic Convolutional Neural Network Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 12(1), 1. <https://doi.org/10.22146/ijccs.19237>

Shelar, A., & Huang, C. Y. (2018). Sentiment analysis of twitter data. *Proceedings - 2018 International Conference on Computational Science and*

*Computational Intelligence, CSCI 2018*, 1301–1302.

<https://doi.org/10.1109/CSCI46756.2018.00252>

Sigalingging, B. P. H. (2018). *Red Devils the Legend: Sejarah Manchester United dari Masa ke Masa*. Media Pressindo.

<https://books.google.co.id/books?id=flGGDwAAQBAJ>

Weiss, S. M., Indurkha, N., Zhang, T., & Damerau, F. (2010). *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer New

York. <https://books.google.co.id/books?id=NZteXd4qf9sC>

Zhang, Y., & Wallace, B. (2015). *A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification*.

<http://arxiv.org/abs/1510.03820>