

ANALISIS SENTIMEN MASYARAKAT TENTANG
PERKEMBANGAN INDUSTRI E-SPORT INDONESIA
PADA TWITTER MENGGUNAKAN MULTINOMIAL
NAÏVE BAYES

Tugas Akhir

Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat Sarjana
Informatika



Dibuat Oleh:

Joshua Restu Valentino

160709021

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2020

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN MASYARAKAT TENTANG PERKEMBANGAN INDUSTRI E-SPORT
INDONESIA PADA TWITTER MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

yang disusun oleh

JOSHUA RESTU VALENTINO

160709021

dinyatakan telah memenuhi syarat pada tanggal 05 Oktober 2020

		Keterangan
Dosen Pembimbing 1	: Prof. Ir. Suyoto, MSc., PhD	Telah menyetujui
Dosen Pembimbing 2	: Dr. Pranowo, S.T., M.T.	Telah menyetujui
Tim Penguji		
Penguji 1	: Prof. Ir. Suyoto, MSc., PhD	Telah menyetujui
Penguji 2	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah menyetujui
Penguji 3	: Paulus Mudjihartono, ST., MT., PhD	Telah menyetujui

Yogyakarta, 05 Oktober 2020

Universitas Atma Jaya Yogyakarta

Fakultas Teknologi Industri

Dekan

ttd

Dr. A. Teguh Siswanto, M.Sc



PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Joshua Restu Valentino
NPM : 160709021
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : ANALISIS SENTIMEN MASYARAKAT
TENTANG PERKEMBANGAN INDUSTRI E-
SPORT INDONESIA PADA TWITTER
MENGUNAKAN MULTINOMIAL NAÏVE
BAYES.

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 5 Oktober 2020

Yang menyatakan,

Josuha Restu Valentino

160709021

HALAMAN PERSEMBAHAN

Tugas akhir ini saya persembahkan kepada

- ❖ Bapak, Ibu, dan Adik serta keluarga besar saya.
- ❖ Teman-teman senasib dan seperjuangan.



KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir “ANALISIS SENTIMEN MASYARAKAT TENTANG PERKEMBANGAN INDUSTRI E-SPORT INDONESIA PADA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER” ini dengan baik.

Penyusunan tugas akhir ini bertujuan untuk memenuhi salah satu ketentuan buat menggapai derajat sarjana Informatika dari Program Studi Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis menyadari jika dalam pembuatan tugas akhir ini penulis sudah memperoleh dukungan, bimbingan, serta dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang menemani saya dalam tiap pengerjaan tugas akhir ini melalui berkat, rahmat, serta karunia-Nya.
2. Bapak Dr. A. Teguh Siswanto, M.Sc., selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D, selaku dosen pembimbing I yang turut serta selalu membimbing dan mendukung penulis untuk tetap memberikan motivasi dan semangat dalam menyelesaikan tugas akhir ini.
4. Bapak Dr. Pranowo, S.T., M.T., selaku dosen pembimbing II yang turut serta selalu membimbing dan mendukung penulis untuk tetap memberikan motivasi dan semangat dalam menyelesaikan tugas akhir ini.
5. Bapak Paulus Mudjihartono, ST., MT., Ph.D., selaku dosen pembimbing akademik yang telah banyak memberi saran selama masa perkuliahan.
6. Seluruh Dosen dan Staff Pengajar Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta yang telah membimbing dan membantu penulis dalam masa perkuliahan yang penulis telah tempuh.

7. Orang tua (Britelhori dan Tariana), adik (Nanda Jessica Shabrina) dan keluarga besar yang mendukung serta memberikan doa dan semangat kepada penulis sehingga tugas akhir ini dapat selesai dengan tepat waktu.
8. Sahabat Perkuliahan yaitu Hakeem Nanda Saputra, Frans Josua Damanik, Theo Walasary, Vladimir Juino, Aga, Yudho, Yafet, Vian, Yodhi dan semua teman yang sudah memberikan banyak motivasi dan bantuan bagi penulis selama masa perkuliahan dan masa penulisan Tugas Akhir ini.
9. Fransiscus yang telah banyak membantu saya sejak SMA hingga perkuliahan.
10. Reynaldo Suarez Wawan yang telah menjadi teman seperjuangan beda universitas untuk saling bertukar pikiran.
11. Shani Indira Natio yang telah memberikan penulis semangat selama menempuh perkuliahan.
12. Seluruh teman-teman yang mendukung penulis dan tidak dapat disebutkan satu per satu yang tentunya telah memberikan semangat dan kasih sayang kepada penulis.

Pembuatan projek magang dan juga penulisan tugas akhir tentunya dibuat dengan sepenuh hati. Namun tidak dipungkiri bahwa dalam proses penulisan tugas akhir ini terdapat banyak kekurangan. Maka dari itu, apabila pembaca menemukan kekurangan dan kekeliruan dalam hal penulisan, penggunaan bahasa, pengartian makna kalimat satu dan lainnya mohon dimaafkan. Semoga laporan ini dapat bermanfaat bagi pembaca.

Yogyakarta, 5 Oktober 2020

Joshua Restu Valentino

16 07 09021

DAFTAR ISI

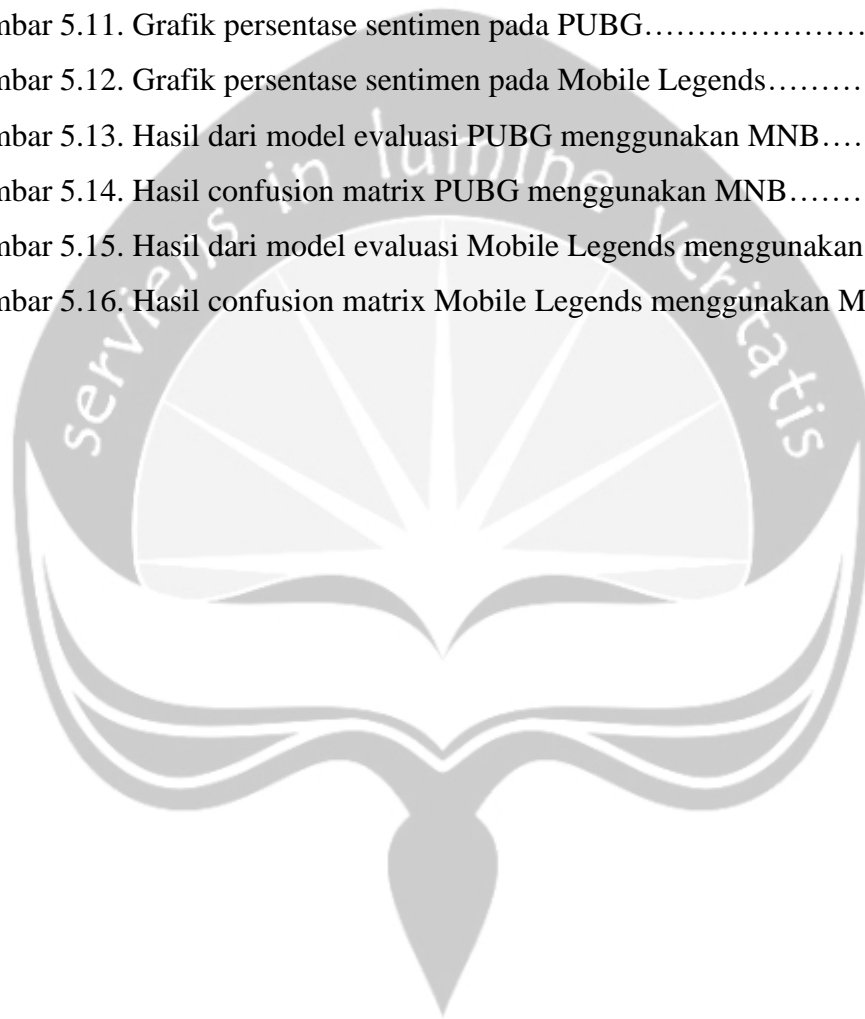
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH.....	ii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
INTISARI.....	xii
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5. Sistematika Penulisan	4
BAB II. TINJAUAN PUSTAKA.....	5
BAB III. LANDASAN TEORI.....	10
3.1 Analisis Sentimen	10
3.2 Jejaring Sosial Twitter	11
3.3 Data Mining	11
3.4 Naïve Bayes Classifier (NBC).....	13
3.5. Multinomial Naïve Bayes	14
3.6. Confusion Matrix	15
3.7. Python 3	16
BAB IV. METODOLOGI PENELITIAN	20
4.1 Proses pengumpulan Data.....	21
4.2 Pre-processing Data	23
4.2.1 Repeat Character Normalize.....	23
4.2.2 Spell Normalize.....	24

4.2.3 Tokenize Text.....	24
4.2.4 Expand Contractions.....	24
4.2.5 Stemming.....	24
4.2.6 Remove Special Characters.....	24
4.2.7 Remove Stopwords.....	25
4.2.8 Normalizr Corpus.....	25
4.3 Pemberian Bobot dan Label	25
4.4 Model Evaluasi.....	35
4.4.1 Klasifikasi.....	35
4.4.1 Validasi.....	36
BAB V. HASIL PEMBAHASAN	38
5.1 Pengumpulan Data.....	38
5.2 Pre-processing Data	45
5.2.1 Repeat Character Normalize.....	46
5.2.2 Spell Normalize.....	46
5.2.3 Tokenize Text.....	46
5.2.4 Expand Contractions.....	46
5.2.5 Stemming.....	46
5.2.6 Remove Special Characters.....	47
5.2.7 Remove Stopwords.....	47
5.2.8 Normalizr Corpus.....	47
5.3 Pemberian Bobot dan Label	49
5.4 Model Evaluasi.....	55
BAB VI. KESIMPULAN DAN SARAN	59
6.1 Kesimpulan	59
6.2 Saran	60
DAFTAR PUSTAKA	61

DAFTAR GAMBAR

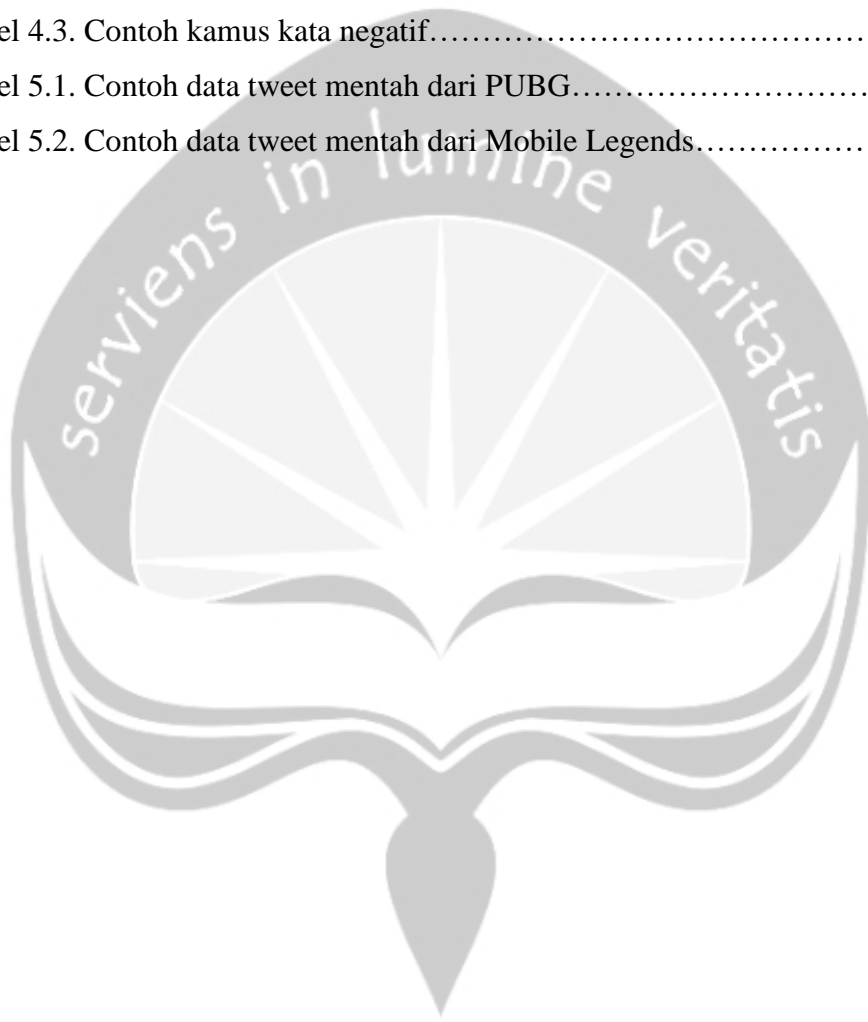
Gambar 3.1. Contoh import library python.....	17
Gambar 3.2. Contoh import data CSV.....	17
Gambar 3.3. Contoh melakukan definisi pada Python.....	18
Gambar 3.4. Contoh cara menampilkan data pada Python.....	18
Gambar 3.5. Contoh hasil dari menampilkan data pada Python.....	18
Gambar 3.6. Contoh cara menampilkan matrix pada Python.....	19
Gambar 3.7. Contoh hasil dari menampilkan matrix pada Python.....	19
Gambar 4.1. Alur dari proses analisis sentimen.....	20
Gambar 4.2. Alur dari pengambilan data tweets.....	21
Gambar 4.3. Contoh dari fitur Search Twitter.....	22
Gambar 4.4. Contoh dari fungsi select attributes.....	22
Gambar 4.5. Contoh dari fungsi write excel.....	23
Gambar 4.6. Contoh dari Normalize corpus.....	25
Gambar 4.7. Contoh hasil dari kolom sentiment positif.....	29
Gambar 4.8. Contoh hasil dari kolom sentiment negatif.....	30
Gambar 4.9. Contoh hasil dari perhitungan skor negatif dan positif.....	31
Gambar 4.10. Contoh hasil dari pemberian label.....	32
Gambar 4.11. Contoh hasil dari jumlah data positif.....	33
Gambar 4.12. Contoh hasil dari jumlah data negatif.....	33
Gambar 4.13. Contoh hasil dari jumlah data netral.....	34
Gambar 4.14. Contoh hasil dari jumlah total data.....	34
Gambar 4.15. Contoh hasil dari diagram sentimen.....	35
Gambar 4.16. Contoh hasil dari model evaluasi menggunakan MNB.....	33
Gambar 4.17. Contoh hasil dari model evaluasi menggunakan MNB.....	36
Gambar 5.1. Data mentah PUBG dari Rapidminer.....	39
Gambar 5.2. Data mentah PUBG dari GetOldTweet.....	40
Gambar 5.3. Data mentah Mobile Legends dari Rapidminer.....	41

Gambar 5.4. Data mentah Mobile Legends dari GetOldTweet.....	42
Gambar 5.5. Data PUBG yang telah melalui normalize corpus.....	48
Gambar 5.6. Data Mobile Legends yang telah melalui normalize corpus.....	48
Gambar 5.7. Hasil pemberian bobot dan labeling dari PUBG.....	49
Gambar 5.8. Hasil pemberian bobot dan labeling dari Mobile Legends.....	50
Gambar 5.9. Grafik kelas sentimen pada PUBG.....	51
Gambar 5.10. Grafik kelas sentimen pada Mobile Legends.....	52
Gambar 5.11. Grafik persentase sentimen pada PUBG.....	53
Gambar 5.12. Grafik persentase sentimen pada Mobile Legends.....	54
Gambar 5.13. Hasil dari model evaluasi PUBG menggunakan MNB.....	55
Gambar 5.14. Hasil confusion matrix PUBG menggunakan MNB.....	56
Gambar 5.15. Hasil dari model evaluasi Mobile Legends menggunakan MNB.....	57
Gambar 5.16. Hasil confusion matrix Mobile Legends menggunakan MNB.....	58



DAFTAR TABEL

Tabel 2.1. Perbandingan dengan penelitian yang telah dilakukan sebelumnya.....	8
Tabel 4.1. Pemberian score pada sentiment.....	26
Tabel 4.2. Contoh kamus kata positif.....	27
Tabel 4.3. Contoh kamus kata negatif.....	27
Tabel 5.1. Contoh data tweet mentah dari PUBG.....	43
Tabel 5.2. Contoh data tweet mentah dari Mobile Legends.....	44



INTISARI

ANALISIS SENTIMEN MASYARAKAT TENTANG PERKEMBANGAN INDUSTRI E-SPORT INDONESIA PADA TWITTER MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

Joshua Restu Valentino

160709021

Perkembangan teknologi yang ada saat ini telah banyak berkembang, tak terkecuali pada media sosial. Media sosial merupakan satu dari sekian media yang mempunyai data yang berkapasitas besar. Salah satu dari manfaat media sosial yaitu menjadi sarana untuk mengetahui opini dari suatu pembahasan dimana terdapat topik tertentu.

Topik yang menjadi fokus utama dari penelitian ini adalah perkembangan industri e-sport Indonesia. Pada media sosial masyarakat bisa mengungkapkan opini tentang perkembangan industri e-sport Indonesia. Opini-opini itu disampaikan melalui tweets yang ada pada Twitter. Analisis sentimen merupakan satu dari sekian banyak cabang yang meneliti tentang Text Mining. Penelitian ini digunakan untuk melakukan klasifikasi pada dokumen yang berupa teks dimana teks ini berisikan opini-opini yang beralaskan sentiment. Opini tentang perkembangan industri e-sport Indonesia dikumpulkan, kemudian dijadikan suatu data yang akan dilakukan analisa serta pengamatan. Naïve Bayes Classifier merupakan salah satu metode yang digunakan untuk melakukan analisis sentimen. Multinomial Naïve Bayes merupakan turunan dari algoritma yang telah dilakukan perkembangan yang berasal dari Naïve Bayes Classifier.

Hasil dari metode Multinomial Naïve Bayes untuk data tweets PUBG memperoleh hasil *f-1 score* sebesar 68%, sedangkan untuk data tweets *Mobile Legends* memperoleh hasil *f-1 score* sebesar 66%. Kemudian untuk sentimen, baik PUBG ataupun *Mobile legends* memperoleh lebih dari 50% sentimen netral. Yang artinya opini masyarakat pada industri e-sport masih bisa menjadi lebih baik untuk kedepannya.

Kata Kunci: *Twitter, Data Mining, Analisis Sentimen, Multinomial Naïve Bayes, Perkembangan Industri E-sport Indonesia, PUBG, Mobile Legends, pabji, mobil lejen.*

Dosen Pembimbing I : Prof. Ir. Suyoto, M.Sc., Ph.D

Dosen Pembimbing II : Dr. Pranowo, S.T., M.T.

Jadwal Sidang Tugas Akhir : 5 Oktober 2020

BAB I. PENDAHULUAN

1.1 Latar Belakang

Teknologi pada saat ini sangat berkembang pesat yang membuat suatu kejadian yang terjadi dapat tersebar dengan cepat. Teknologi yang sangat populer pada masa ini salah satunya adalah internet karena dengan tersedianya internet maka informasi-informasi yang tersebar dan data-data penting yang ada dapat ditemukan dan disimpan dengan mudah dan cepat. Informasi-informasi tersebut bisa didapatkan dari berbagai situs-situs portal informasi yang ada. Informasi-informasi tersebut sebagian besarnya berbentuk teks. Masyarakat menjadi sulit memahami informasi yang ada karena informasi-informasi tersebut masih banyak yang kurang terstruktur, sehingga masyarakat menjadi sulit menentukan informasi yang benar dan informasi yang salah dari suatu informasi.

Data mining adalah sebuah proses dimana informasi diekstrak dari data-data yang terkumpul pada bidang statistic. Data mining merupakan suatu kegiatan dimana data yang berukuran besar dikumpulkan, kemudian dicari pola-pola yang memiliki keteraturan. Data mining merupakan sebuah proses analisa dari pandangan yang berbeda untuk mendapatkan kesimpulan agar menjadi informasi-informasi berharga yang mampu meningkatkan keuntungan dengan biaya yang minim [1].

Indonesia merupakan pengguna internet terbesar ketiga yang ada di Asia, dimana terdapat 143,26 juta jiwa dan diperkirakan bisa mencapai 269,54 juta jiwa yang artinya adalah 53% dari total populasi yang ada (internetworldstats.com, Jun 2019). *Twitter* merupakan salah satu situs jejaring sosial yang dapat membantu manusia untuk saling berhubungan dan berbagi informasi satu sama lain. *Twitter* telah berdiri sejak bulan Maret pada tahun 2006 dan sudah sekarang sudah memiliki lebih dari 352 juta pengguna yang aktif di berbagai pelosok dunia [2].

Twitter juga dikenal sebagai layanan *micro-blogging*, dimana pengguna-penggunanya diberikan fasilitas untuk menulis, membaca, serta mengirimkan informasi kepada forum umum. *Twitter* mampu menciptakan fasilitas agar perusahaan dan pengguna dapat saling berinteraksi walau tidak secara langsung. Perusahaan bisa menyebarkan informasi terbaru kepada pengguna, dan memperoleh *feedback* dari pengguna dengan waktu yang singkat, dengan demikian perusahaan mampu mengevaluasi layanan-layanan yang telah diberikan. Oleh karena itu perusahaan pengembang industri *e-sport* saling bersaing agar bisa meningkatkan layanan dan kualitas dari produk yang mereka miliki [3].

Perkembangan pasar game dunia pada tahun 2019 akan mencapai nilai US\$152 miliar atau sekitar Rp.2,15 kuadriliun, telah mengalami peningkatan sebesar 9,6% dibanding dengan tahun sebelumnya. Pada wilayah Asia Pasifik sendiri, Indonesia merupakan salah satu pasar *video game* terbesar yang mencapai nilai US\$941 juta atau sekitar Rp.13 triliun. Dari angka yang terlihat, maka bukanlah hal yang mengherankan jika industri di dunia *e-sport* mengalami perkembangan yang pesat. *E-sport* merupakan permainan kompetitif yang berbasis *video game* [4].

Perkembangan industri *e-sport* yang ada di Indonesia telah mendapatkan respon yang positif dari pemerintah. Indonesia e-Sport Association atau biasa disebut juga dengan IeSPA merupakan wadah resmi yang didirikan pemerintah sebagai asosiasi yang menaungi komunitas ini di Indonesia. IeSPA juga membantu para pemain agar bisa mengembangkan bakat dan kemampuan dalam dunia game kompetitif. Selain itu juga, pemerintah menyelenggarakan kejuaraan seperti IEC University Series 2019, Youth National eSport Championship, dan Piala Presiden [5].

1.2 Rumusan Masalah

Berdasarkan dari latar belakang, permasalahan yang muncul diantaranya yaitu:

1. Bagaimana menganalisis serta mengevaluasi pendapat masyarakat tentang perkembangan industri e-sport di Indonesia melalui twitter?
2. Bagaimana cara melakukan klasifikasi pada setiap sentiment pada tweet yang akan dianalisis oleh penulis?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini melingkupi hal berikut yaitu:

1. Twitter merupakan sumber data yang digunakan pada penelitian ini.
2. Data yang digunakan oleh peneliti hanya menggunakan data yang berasal dari twitter yaitu *tweets*.
3. PUBG dan *Mobile Legends* adalah industri *e-sport* yang akan menjadi target analisis opini oleh penulis.
4. Penelitian ini mengumpulkan data yang hanya menggunakan Bahasa Indonesia.
5. Hasil Output berupa sentiment positif, negative, dan netral dari analisis sentiment menggunakan metode Multinomial Naïve Bayes.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Melakukan analisis serta melakukan evaluasi terhadap sentimen masyarakat tentang Perkembangan Industri E-sport Indonesia terkhusus PUBG dan *Mobile Legends* pada Twitter menggunakan metode Multinomial Naïve Bayes.
2. Melakukan klasifikasi terhadap sentiment pada tweets.

1.5. Sistematika Penulisan

Sistematika penulisan untuk penelitian ini adalah:

BAB I: Pendahuluan

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, metodologi penelitian serta sistematika pada penulisan laporan tugas akhir.

BAB II: Tinjauan Pustaka

Bab ini menjelaskan tentang topik yang pernah diteliti sebelumnya. Kemudian menjelaskan tentang perbandingan yang ada pada penelitian yang telah dilakukan dengan penelitian yang akan dilakukan.

BAB III: Landasan Teori

Bab ini menjelaskan tentang teori-teori yang memiliki hubungan atau keterkaitan dengan permasalahan-permasalahan yang akan dibahas.

BAB IV: Metodologi Penelitian

Bab ini menjelaskan bagaimana proses serta langkah-langkah yang dilakukan untuk mendapatkan hasil dari analisis sentimen.

BAB V: Hasil dan Pembahasan

Bab ini menjelaskan tentang hasil yang diperoleh dari penelitian yang telah dilakukan serta melakukan pembahasan terhadap hasil yang didapatkan.

BAB VI: Kesimpulan dan Saran

Bab ini berisikan kesimpulan dari keseluruhan pembahasan dan juga saran yang kiranya akan bermanfaat pada penelitian-penelitian setelahnya

BAB II. TINJAUAN PUSTAKA

Penelitian yang dilakukan ini tidak terlepas dari penelitian-penelitian yang telah dilakukan pada masa sebelumnya sebagai kajian serta perbandingan. Penelitian-penelitian yang telah ada dijadikan perbandingan berdasarkan hasil-hasil dari penelitian. Penelitian yang dijadikan referensi tidak lepas dari topik pembahasan yaitu perihal analisis sentiment pada twitter. Menurut Rofiqoh, Perdana, dan Fauzi³ analisis sentiment pada twitter dapat dilakukan dengan metode Lexicon Based Features dan Support Vector Machine [6].

Pada proses analisis sentiment dilakukan implementasi dengan metode Support Vector Machine terhadap tingkat kepuasan pengguna dengan hasil klasifikasi yang dibagi menjadi dua yaitu sentimen positif dan sentimen negatif. Penelitian ini mengambil data yang berupa teks dari *twitter* melalui tweets yang ada. Data yang ada berjumlah 300 kemudian dibagi menjadi dua bagian yaitu 70%:30% yang merupakan data training dan juga data testing. Data training adalah 210 data kemudian data testing sebanyak 90 data[6] .

Data training dan data testing dilakukan pengujian melalui analisis sentimen dengan langkah pre-processing, memberi bobot pada kata tf-idf, kemudian memberikan bobot pada Lexicon based Feature yang merupakan sebuah perhitungan dimana frekuensi kata yang muncul mengandung sentimen negatif ataupun positif. Kata-kata yang ada kemudian akan melalui proses normalisasi menggunakan metode Min-max, dan juga pengklasifikasian Support Vector Machine agar bisa memperoleh hasil analisis yang berbentuk sentimen negatif maupun positif. Pada saat proses Support Vector Machine dan Lexicon Based Features, diperoleh angka-angka dari parameter Support Vector Machine dimana nilai yang didapatkan adalah *value* yang optimal. Nilai optimal ini adalah *degree* yang merupakan sebuah perhitungan *kernel polynomial* yang memiliki nilai 2 serta nilai dari *konstanta learning rate*. Jumlah iterasi maksimum pada perhitungan tersebut

adalah 0,0001 dan 50 yang merupakan perhitungan *sequential training SVM*. Jika nilai dari *degree* dibesarkan maka akurasi dari hasil yang muncul akan mengalami sebuah kemunduran. Sedangkan jika nilai konstanta *learning rate* dan jumlah dari iterasi maksimum dibesarkan maka support vector akan berubah. Support vector yang telah berubah akan menjadi titik data yang jaraknya paling kecil dengan garis fungsi yang memisahkan kelas data, hal ini akan menghasilkan tingkat dari akurasi menurun [6].

Lexicon Based features yang ditambahkan akan mempengaruhi analisis sentimen pada tingkat atau level kepuasan yang berasal dari pengguna penyedia layanan telekomunikasi seluler pada twitter, tetapi tingkat pada akurasi yang diperoleh lebih rendah jika dilakukan perbandingan dengan penelitian analisis yang tidak menggunakan metode Lexicon Based Features. Hal ini disebabkan karena terdapat kata-kata sindrian yang mempunyai sentimen positif pada data testing yang seharusnya memiliki sentiment negative atau sebaliknya [6].

Mujib Ridwan, Hadi Suyono, dan M. Sarosa melakukan penelitian yang dilakukan dengan beberapa tahap yaitu membuat proposal, mengumpulkan data, mengolah data, melakukan implementasi NBC, pengujian, dan yang terakhir adalah hasil analisis. Pada tahapan pengelolaan terdapat beberapa fase yang terdapat pada data mining, yaitu membersihkan data, mengintegrasikan data, menyeleksi data, melakukan transformasi pada data, dan membentuk dataset yang terdapat pada penelitian yang digunakan [7].

Pada tahun 2014 Ling melakukan penelitian dimana dia membangun aplikasi yang mampu menangani analisis sentimen menggunakan metode *Naïve Bayes* dan juga fitur *Chi Square*. Pre-processing pada saat aplikasi ini dibangun memiliki beberapa tahapan yaitu *tokenization*, *stemming*, dan yang terakhir adalah *stopwords*. *Tokenization* memiliki tugas untuk melakukan pemotongan terhadap urutan-urutan dari karakter dan merupakan sebuah set dimana potongan-potongan kata yang dibutuhkan atau sesuai dengan sistem menjadi dokumen. Potongan-potongan itulah yang disebut dengan token. *Stemming* merupakan tahapan dimana token yang memiliki imbuhan

dijadikan kata dasar. Kemudian, kata-kata yang dianggap tidak mempunyai *stopwords* atau bisa disebut juga tendensi sentiment akan dibuang. Seleksi fitur *chi square* berfungsi sebagai fitur yang akan melakukan reduksi terhadap fitur-fitur yang tidak relevan pada tahapan proses klasifikasi Naïve Bayes Classifier. Fitur ini menggunakan teori yang ada pada statistika sebagai acuan untuk melakukan pengujian independensi pada term berdasarkan kategorinya. Sehingga penelitian ini memperoleh hasil dimana tingkat akurasi yaitu 83% dengan rata-rata harmonik 90.713% [8].

Pada tahun 2016 Wicaksono, Suyoto, & Pranowo melakukan penelitian terhadap pemilihan Presiden Amerika Serikat menggunakan analisis sentimen pada sosial media. Pada penelitian ini terdapat empat buah metode yang digunakan untuk melakukan perbandingan analisis sentiment yaitu Multinomial Naïve Bayes, Binarized Multinomial Naïve Bayes, SentWordNet, serta AFINN-111. Binarized Multinomial Naïve Bayes dan Multinomial Naïve Bayes adalah sebuah bentuk analisis yang menggunakan pembelajaran (*learning-based*). Bentuk analisis ini memerlukan adanya data latih yang berupa pengetahuan agar bisa dilakukan pengklasifikasian. Sementara SentiWordNet dan AFINN-11 adalah metode klasifikasi dimana digunakannya leksikon yang berisikan kata-kata atau disebut juga *synset* yang telah diberikan nilai sentiment [9].

Tabel 2.1. merupakan tabel perbandingan dari tinjauan pustaka menurut penelitian yang telah dilakukan sebelumnya mengenai implementasi data mining.

Tabel 2.1. Perbandingan dengan penelitian yang telah dilakukan sebelumnya :

No	Pembanding/ Peneliti	Algoritma	Pokok Penelitian	Melakukan Klasifikasi Pada	Hasil
1	Umi Rofiqoh, Rizal Setya Perdana, & M. Ali [6].	Metode Support Vector Machine dan Lexicon Based Features	Sentimen positif dan sentiment negatif	Data Twitter	Tingkat akurasi dipengaruhi oleh kata-kata yang sentimennya tidak sesuai
2	Mujib Ridwan, Hadi Suyono, dan M. Sarosa [7].	Naïve Bayes Classifier	Nilai akurasi dari algoritma	Data Akademik Mahasiswa	Naïve Bayes Classifier menghasilkan nilai precision 83%, recall 50%, dan akurasi 70%
3	Ling, Kencana, & Oka [8].	Naïve Bayes Classifier dan fitur Chi Square	Nilai Akurasi dari algoritma	Data dari 200 buah opini	Naïve Bayes Classifier memperoleh akurasi sebesar 83% dengan harmonik sebesar 90,713%.

No	Pembanding/ Peneliti	Algoritma	Pokok Penelitian	Melakukan Klasifikasi Pada	Hasil
4	Wicaksono, Suyoto, & Pranowo [10].	Multinomial Naïve Bayes, Binarized Multinomial Naïve Bayes, SentWordNet, serta AFINN-111	Nilai Akurasi dari algoritma	Data media sosial	Multinomial Naïve Bayes memperoleh tingkat akurasi yang paling baik.

BAB VI. KESIMPULAN DAN SARAN

Setelah peneliti melakukan penelitian serta analisis sentimen terhadap industri *e-sport* yang ada di Indonesia yaitu PUBG dan *Mobile Legends* dengan menggunakan metode *Multinomial Naïve Bayes*. Pada bagian yang terakhir, peneliti akan melakukan penjelasan tentang kesimpulan yang terkait dengan analisis sentimen yang telah dikerjakan. Setelah memberikan kesimpulan dari penelitian ini, peneliti juga akan memberikan saran, sehingga di kemudian harinya dapat berguna untuk penelitian-penelitian selanjutnya.

6.1 Kesimpulan

Berdasarkan dari hasil yang diperoleh berdasarkan penelitian serta analisis akan dijelaskan pada kesimpulan sebagai berikut :

1. Hasil ketepatan klasifikasi dengan menggunakan metode *Multinomial Naïve Bayes* pada industri e-sport PUBG memperoleh hasil *f-1 score* sebesar 68%, sedangkan untuk hasil klasifikasi yang diperoleh dari industri e-sport *mobile legends* dengan menggunakan *Multinomial Naïve Bayes* mendapatkan hasil sebesar 66%. Dapat diambil kesimpulan bahwa dengan menggunakan metode *Multinomial Naïve Bayes* mampu memberikan hasil yang cukup baik untuk melakukan klasifikasi pada data.
2. Persentase dari analisis sentimen pada PUBG mendapatkan respon positif lebih banyak jika dibandingkan *Mobile legends*. PUBG memperoleh sentimen positif sebanyak 25%, sedangkan untuk *mobile legends* memperoleh sebanyak 21%.
3. Masyarakat di Indonesia bisa terbilang masih cukup netral dalam menanggapi industri e-sport yang ada di Indonesia khususnya PUBG dan *Mobile Legends*. Karena dapat dilihat dari hasil penelitian bahwa baik PUBG ataupun *Mobile legends* memperoleh lebih dari 50% sentiment netral. Jadi dapat diartikan bahwa opini masyarakat pada industri e-sport masih bisa menjadi lebih baik untuk kedepannya.

6.2 Saran

Saran yang bisa peneliti sampaikan berdasarkan hasil penelitian yang telah diperoleh ini adalah sebagai berikut :

1. Pada penelitian yang telah dilakukan ini, terdapat beberapa kekurangan yang telah peneliti alami yaitu masih kurang lengkapnya kamus yang digunakan. Sekiranya pada penelitian yang akan ada dimasa yang akan datang, agar bisa menggunakan kamus yang lebih lengkap lagi agar hasil yang dapat diperoleh menjadi maksimal.
2. Pada saat *pre-processing data* , data-data tweet banyak yang mengandung karakter-karakter yang tidak dikenali oleh program yang bisa mengakibatkan error pada code yang telah dibuat, sehingga diharapkan untuk kedepannya proses pembersihan data bisa menambahkan karakter-karakter special tersebut agar program mampu membersihkannya. Ini semua agar data yang akan melalui proses berjalan dengan baik.

DAFTAR PUSTAKA

- [1] L. R. Mabruur angga, “PENERAPAN DATA MINING UNTUK MEMPREDIKSI Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA),” *J. Komput. dan Inform.*, vol. 1, pp. 53–57, 2012.
- [2] C. Bridge, “Unstructured data and the 80 percent rule,” 2019. [Online]. Available: <https://www.internetlivestats.com/twitter-statistics/>.
- [3] R. A. Setiawan and D. B. Setyohadi, “Analisis Komunikasi Sosial Media Twitter sebagai Saluran Layanan Pelanggan Provider Internet dan Seluler di Indonesia,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 1, p. 16, 2017.
- [4] Newzoo, “Global Mobile Market Report,” *Newzoo*, 2019. [Online]. Available: <https://newzoo.com/insights/trend-reports/newzoo-global-games-market-report-2019-light-version/>.
- [5] Mostar, “Perkembangan Industri E-sport Indonesia di Tahun 2019,” *Mostar*, 2019. [Online]. Available: <https://mostar.co.id/perkembangan-industri-e-sport-indonesia-di-tahun-2019/>.
- [6] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017.
- [7] M. Ridwan, H. Suyono, and M. Sarosa, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” vol. 7, no. 1, pp. 59–64, 2013.
- [8] J. LING, I. P. E. N. KENCANA, and T. B. OKA, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square,” *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014.
- [9] N. Monarizqa, L. E. Nugroho, and B. S. Hantono, “Penerapan Analisis

- Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating,” *J. Penelit. Tek. Elektro dan Teknol. Inf.*, vol. 1, pp. 151–155, 2014.
- [10] A. J. WICAKSONO, “Prediksi pemilihan presiden amerika serikat 2016 dengan menganalisis sentimen pada sosial media,” *J. Buana Inform.*, pp. 1–56, 2016.
- [11] I. Rozi, S. Pramono, and E. Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi,” *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [12] A. R. Naradhipa and A. Purwarianti, “Sentiment classification for Indonesian message in social media,” *Proc. - Int. Conf. Cloud Comput. Soc. Netw. 2012 Cloud Comput. Soc. Netw. Smart Product. Soc. ICCCSN 2012*, no. July, pp. 2–5, 2012.
- [13] M. Rivki and A. M. Bachtiar, “Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia,” *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017.
- [14] S. P. Beasiswa, A. Musthafa, H. Suyono, and M. Sarosa, “Perbandingan Kinerja Algoritma C.45 dan AHP-TOPSIS Sebagai Pendukung Keputusan Proses Seleksi Penerima Beasiswa,” *J. EECCIS Vol. 9, No. 2*, vol. 9, no. Sistem Pendukung Keputusan, pp. 109–114, 2015.
- [15] T. Novianti and A. Aziz, “Aplikasi Data Mining Menggunakan Metode decision Tree untuk menampilkan laporan Hasil Nilai Akhir Mahasiswa Fakultas Teknik UMSURABAYA,” *Netw. Eng. Res. Oper. [NERO]*, vol. 1, no. 3, pp. 198–203, 2015.
- [16] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [17] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *J. Teknol. Informasi, Komputer, dan Apl. (JTIKA)*, vol. 2, no. 1, pp. 89–100, 2020.

- [18] M. T. H. Bunga *et al.*, “Multinomial Naive Bayes Untuk Klasifikasi Status Kredit Mitra Binaan Di Pt. Angkasa Pura I Program Kemitraan,” *J-Icon*, vol. 6, no. 2, pp. 30–34, 2018.
- [19] M. C. Sinaga, “Kriptografi dan Python,” 2017.
- [20] S. Suryono, E. Utami, and E. T. Luthfi, “Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naïve Bayes Classifier,” *Semin. Nas. Geotik 2018*, pp. 9–15, 2018.
- [21] Injosoft AB, “ASCII Code - The extended ASCII table,” *ASCII Code - The extended ASCII table*, 2020. .

