

ACC Semprom
29 Okt'2021



Oktalia Juwita



Priza Pandunata



**ANALISIS SENTIMEN TERHADAP KARIR *E-SPORTS* DI INDONESIA
PADA MEDIA *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES***

PROPOSAL SKRIPSI

Oleh:

Fendy Marzuki

NIM 172410101106

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS JEMBER

2021

DAFTAR ISI

DAFTAR ISI.....	2
BAB 1. PENDAHULUAN	4
1.1. Latar Belakang	4
1.2. Rumusan Masalah.....	5
1.3. Tujuan Penelitian	6
1.4. Manfaat Penelitian	6
1.5. Batasan Masalah	6
BAB 2. TINJAUAN PUSTAKA.....	7
2.1. E-Sport	7
2.2. Media Sosial.....	8
2.3. Twitter.....	10
2.4. Sentimen Analisis	10
2.5. Data Mining	11
2.6. Pengelompokan Data Mining.....	14
2.7. Klasifikasi	15
2.8. Metode Naive Bayes	17
2.9. Confusion Matrix	20
2.10. Penelitian Terdahulu	21
BAB 3. METODOLOGI PENELITIAN.....	22
3.1. Jenis Penelitian.....	22
3.2. Tahapan Penelitian.....	22
3.3. Metode Pengumpulan Data	23
3.4. Pembuatan Kamus Data	25
3.5. Preprocessing	26

3.6.	Penyusunan Model.....	28
3.7.	Pengujian Model	29
3.7.1.	Confusion Matrix	29
3.7.2.	Pengukuran Akurasi	30
DAFTAR PUSTAKA.....		31

BAB 1. PENDAHULUAN

1.1. Latar Belakang

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 (newzoo.com, Juni 2019). 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*.

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. Perkembangan industri *e-sport* di Indonesia sangat pesat, ekosistem ini dapat menciptakan banyak sekali lapangan pekerjaan potensial. Tidak harus ahli dalam bermain game, *e-sport* juga membutuhkan peran-peran seperti pelatih, tim yang memiliki struktur organisasi yang cukup lengkap, *streamer*, *caster*, *event organizer* agar keberlangsungan *e-sport* dapat berjalan dengan baik. *E-sport* ini juga dilirik sebagai industri bisnis yang menjanjikan, oleh karena itu cukup banyak sponsor maupun investor yang turut meramaikan industri ini.

Pelaku *e-sport* menggunakan media sosial sebagai tempat mereka bertukar informasi dan melakukan komunikasi yang berkaitan tentang *game*. Salah satu media sosial yang digunakan adalah media sosial *Twitter*. *Twitter* mengklaim dirinya sebagai rumahnya para *gamers* baik untuk mendapatkan informasi, menonton potongan video pertandingan yang sedang berlangsung, atau

berinteraksi dengan atlet *e-sports* favorit mereka. Sebanyak 91 Juta pengguna aktif *twitter* diseluruh dunia telah melakukan *tweet* terkait *game* per detik dan telah menghasilkan 156 Juta *tweet* tentang *game* sepanjang Juni 2020 hingga Juni 2021 (marketing.twitter.com).

Twitter seringkali digunakan untuk mengungkapkan emosi mengenai sesuatu hal, baik memuji ataupun mencela dalam bentuk emosi. Tidak hanya itu, *Twitter* sering digunakan oleh penggunanya sebagai tempat untuk mendapatkan informasi berisi teks maupun link berita yang *up to date*. Sebagian pengguna percaya bahwa konten informasi di media sosial Indonesia bermula dari *Twitter*, kemudian konten tersebut disebar ulang ke media sosial *Facebook*, *Instagram*, dan *Youtube*.

Terkait dengan opini tentang karir *e-sport* di media sosial *Twitter*, polaritas opini positif atau negatif menggunakan teknik klasifikasi dokumen. *Naive Bayes Classifier* adalah salah satu metode populer yang digunakan untuk pengklasifikasian dokumen. Metode ini mempunyai akurasi dan kecepatan analisis yang tinggi ketika digunakan untuk pengolahan data pada basis data yang besar dan data yang beragam. Hasil klasifikasi ini nantinya dapat menjadi komponen pendukung bagaimana tanggapan masyarakat di media sosial *Twitter* terhadap karir *e-sport* di Indonesia.

Penelitian ini sebelumnya telah dilakukan oleh Simorangkir (2018) dengan judul “Analisis Sentimen pada *Twitter* untuk *Games Online Mobile Legends* dan *Arena of Valor* dengan Metode *Naive Bayes Classifier*”. Hasil penelitian beliau memiliki akurasi cukup baik. Berdasarkan apa yang dipaparkan pada latar belakang di atas serta mengacu pada penelitian-penelitian sebelumnya, maka peneliti tertarik untuk melakukan penelitian yang berjudul “**Analisis Sentimen Terhadap Karir E-Sports Di Indonesia Pada Media Twitter Menggunakan Metode Naive Bayes**”.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat dibuat rumusan masalah sebagai berikut :

1. Bagaimana hasil implementasi metode *Naive Bayes* untuk mengklasifikasikan sentimen data *Twitter* terhadap karir *E-Sports* di Indonesia?
2. Berapa tingkat akurasi dari hasil implementasi metode *Naive Bayes* dalam mengklasifikasikan sentimen data *Twitter* terhadap karir *E-Sports* di Indonesia?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang penulis buat, adapun tujuan dari penelitian ini ialah sebagai berikut:

1. Mengetahui hasil implementasi metode *Naive Bayes* dalam mengklasifikasikan sentimen data *Twitter* terhadap karir *E-Sports* di Indonesia.
2. Mengetahui tingkat akurasi dari hasil implementasi metode *Naive Bayes* dalam mengklasifikasikan sentimen data *Twitter* terhadap karir *E-Sports* di Indonesia.

1.4. Manfaat Penelitian

Adapun hasil dari penelitian ini diharapkan mampu memberikan manfaat sebagai berikut:

1. Pengembangan metode data mining terhadap social network analityc.
2. Dengan mengetahui hasil analisis sentimen karir e-sports di Indonesia pada media sosial Twitter, maka dapat dijadikan sebagai informasi serta sarana pertimbangan masyarakat untuk berkarir di bidang e-sports.

1.5. Batasan Masalah

Agar dalam penelitian ini dapat mencapai sasaran dan tujuan yang diharapkan, maka permasalahan yang ada hanya dibatasi pada data Twitter yang dianalisis tidak mencakup data retweet, gambar atau foto, video dan link.

BAB 2. TINJAUAN PUSTAKA

2.1. E-Sport

E-Sport merupakan olahraga yang dimainkan dengan menggunakan media elektronik, dimana di beberapa negara, seperti : Korea Selatan, Swedia dan Indonesia game sudah dikategorikan sebagai salah satu cabang olahraga dan memiliki lembaga resmi untuk menaungi para pemain game tersebut, dibawah kementerian olahraga di negara-negara tersebut. E-Sport adalah salah satu lini yang paling berkembang dalam beberapa tahun ke belakang. Walaupun begitu, usia E-Sport bisa dibilang hampir setara sejarah dunia gaming sendiri. Saat mode multiplayer mulai diperkenalkan pada konsumen, saat itu pula para gamers seperti mendapatkan sebuah dunia baru yang disebut dengan dunia virtual, dimana para gamers bisa bermain, bersosialisasi dan saling berkompetisi dengan gamers lainnya.

Albert F, 2010) E-Sport Team atau tim E-Sport merupakan kumpulan dari gamers-gamers yang bermain bersama dalam kategori game tertentu dan memiliki visi dan misi yang sama dalam mencapai tujuan dari tim E-Sport tersebut. Para gamers tersebut memainkan game yang sama dalam satu divisi E-Sport Team, namun E-Sport Team bisa memiliki lebih dari satu divisi E-Sport Team dengan kategori game yang berbeda, misalnya : salah satu E-Sport Team di Indonesia yang telah memiliki banyak penghargaan baik penghargaan lokal maupun internasional, yaitu nextlevelgaming atau lebih dikenal dengan nama team nxl> memiliki beberapa divisi eSport Team diantaranya : divisi CS:GO, Mobile Legends dan DOTA2.

E-Sport Team terdiri dari 2 atau lebih gamers yang memiliki kesamaan terhadap suatu jenis tertentu dan untuk mencari gamers yang sesuai untuk membentuk suatu E-Sport Team tidaklah mudah. Hal ini dikarenakan oleh banyak faktor, yaitu :

1. Lokasi

Lokasi sangat menentukan E-Sport Team di dalam berlatih. Hal ini dikarenakan apabila terdapat salah satu anggota eSport Team yang berbeda kota hal ini dapat menghambat komunikasi antara satu anggota eSport Team dengan anggota E-Sport Team yang lain.

2. Jenis Game

Jenis Game merupakan salah satu faktor yang sangat mempengaruhi di dalam pembentukan suatu E-Sport Team. Karena E-Sport Team merupakan kumpulan gamers yang sama-sama bermain di dalam satu jenis game yang sama.

3. Posisi Pemain

Di dalam E-Sport Team setiap pemain memiliki fungsi dan perannya masing-masing. Pembagian peran atau tugas ini berdasarkan keahlian dari masing-masing gamer tersebut. Pembagian peran ini bertujuan agar di dalam setiap pertandingan setiap strategi yang telah ditentukan dapat berjalan sesuai dengan rencana.

4. Usia

Usia juga dapat mempengaruhi terbentuknya suatu E-Sport Team karena usia dapat mempengaruhi tingkat mentalitas, kecepatan respon mata dan reflek seseorang. Karena untuk jenis game tertentu seperti jenis game FPS (First Person Shooter) diperlukan kejelian mata di dalam melihat setiap pergerakan lawan.

2.2. Media Sosial

Media sosial merupakan salah satu media instan yang saat ini memang memiliki berbagai fungsi dalam perannya. Selain berfungsi sebagai alat untuk berkomunikasi, media massa juga menjadi sarana untuk penggunaanya dalam menggali berbagai informasi. Definisi media sosial tidak serta merta merupakan gagasan yang tidak berdasar yang dikemukakan oleh para ahli tersebut. media sosial memiliki peran dan dampak bagi kehidupan masyarakat yang harus didesain sedemikian rupa agar media sosial tetap pada fungsi dan tujuan media sosial itu sendiri dan memiliki manfaat dalam kehidupan setiap individu.

Seperti yang dikemukakan oleh Henderi (2013) bahwa pengertian media sosial adalah situs jaringan sosial berbasis web yang memungkinkan bagi setiap individu untuk membangun profil publik ataupun semi public dalam sistem terbatas, daftar pengguna lain dengan siapa mereka terhubung, dan melihat serta menjelajahi daftar koneksi mereka yang dibuat oleh orang lain dengan suatu sistem (Henderi, 2007: 3)

Sedangkan menurut Phillip Kotler dan Kevin Keller media sosial adalah sarana bagi konsumen untuk berbagi informasi teks, gambar, video dan audio dengan satu sama lain dan dengan perusahaan dan sebaliknya (Kotler, Keller 2012: 568)

Teknologi-teknologi web baru memudahkan semua orang untuk membuat dan yang terpenting menyebarkan konten mereka sendiri. Post di Blog, tweet, instagram, facebook, atau video di YouTube yang dapat direproduksi dan dilihat oleh jutaan orang secara gratis. Pemasang iklan tidak harus membayar banyak uang kepada penerbit atau distributor untuk memasang iklannya. Sekarang pemasang iklan dapat membuat konten sendiri yang menarik dan dilihat banyak orang (Zarrella, 2010).

Pada perannya saat ini, media sosial telah membangun sebuah kekuatan besar dalam membentuk pola perilaku dan berbagai bidang dalam kehidupan masyarakat. hal ini yang membuat fungsi media sosial sangat besar. Adapun fungsi media sosial diantaranya sebagai berikut :

- a) Media sosial mendukung demokratisasi pengetahuan dan informasi.
- b) Mentransformasi manusia dari pengguna isi pesan menjadi pembuat pesan itu sendiri.
- c) Media sosial adalah media yang didesain untuk memperluas interaksi sosial manusia dengan menggunakan internet dan teknologi web.
- d) Media sosial berhasil mentransformasi praktik komunikasi searah media siaran dari satu institusi media ke banyak audience ke dalam praktik komunikasi dialogis antara banyak audience.

2.3. Twitter

Menurut Wikipedia, Twitter adalah layanan jejaring sosial dan mikroblog dalam jaringan (daring) yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis text hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan (tweet). Secara standar, tweet pengguna dapat terlihat oleh umum, namun pengguna dapat membatasi pengiriman kicauan hanya bagi pengikut mereka. Pengguna bisa "berkicau" melalui situs Twitter, aplikasi eksternal yang kompatibel (seperti untuk telepon pintar), ataupun melalui layanan pesan singkat (SMS) yang tersedia di negara-negara tertentu. Layanan-layanan tersebut bersifat gratis, kecuali layanan SMS, yang dikenakan biaya oleh penyedia layanan seluler.

Pengguna bisa berlangganan kicauan pengguna lain dengan cara mengikuti (follow) pengguna yang bersangkutan, dan pengguna yang mengikuti tersebut akan menjadi pengikut (followers) bagi pengguna yang diikutinya. Pengguna dapat mengelompokkan kicauan menurut topik atau jenis dengan menggunakan tagar (hashtag), kata atau frasa yang diawali dengan tanda #. Sedangkan tanda @, yang diikuti dengan nama pengguna, digunakan untuk mengirim atau membalas kicauan pada pengguna lain.

2.4. Sentimen Analisis

Sentimen analisis adalah merupakan salah satu bidang dari ilmu komputer yang mempelajari komputasi linguistik, pengolahan bahasa alami, dan text mining yang bertujuan untuk menganalisa emosi, penilaian, sikap, pendapat, sentimen, evaluasi seseorang terhadap seorang pembicara atau penulis berkenaan dengan suatu produk, layanan, organisasi, individu, topik publik, topik, acara, ataupun kegiatan tertentu (Liu, 2012). Proses utama dalam analisis sentimen yaitu mengelompokkan teks yang terdapat dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan tersebut apakah bersifat positif, negatif, atau netral. Analisis sentimen dapat digunakan untuk mencari pendapat tentang produk, merek atau tokoh publik dan menentukan apakah mereka dilihat positif atau negatif (Saraswati, 2011). Analisis sentimen atau

opinion mining adalah deteksi sikap-sikap terhadap objek atau orang (beineke et al. 2014).

Analisis sentimen dapat digunakan untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, institusi, produk atau pada sebuah kondisi tertentu. Nilai dari analisis sentimen bisa dipecah menjadi 3 yakni, sentimen positif, sentimen negatif dan sentimen netral atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif. Tugas analisis sentimen yaitu mengelompokkan teks ke dalam kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen yang dianalisis apakah bersifat positif, negatif, atau netral (Dehhaf, 2010). Sentimen atau opini mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Alasan tersebut menyebabkan beberapa penelitian terutama pada review produk didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining (Manalu, 2014).

Analisis sentimen bertujuan untuk melakukan penilaian terhadap emosi, sikap, pendapat, evaluasi yang disampaikan oleh seseorang pembicara atau penulis terhadap sebuah produk atau terhadap tokoh masyarakat. Alasan tersebut menyebabkan beberapa penelitian terutama pada review produk didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining. Contoh kalimat yang mengandung sentimen positif yaitu “Jokowi sangat cerdas cocok menjadi Presiden”, contoh kalimat yang mengandung sentimen negatif yaitu “saya benci dengan Prabowo”, dan contoh kalimat yang mengandung sentimen netral yaitu “sipapun presidennya saya selalu mendukung untuk kemajuan Indonesia”.

2.5. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine

learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk. 2005).

Definisi umum dari data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (hidden pattern) berupa pengetahuan (knowledge) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data warehouse, atau media penyimpanan informasi yang lain. Hal penting yang terkait di dalam data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat

Data mining dilakukan dengan tool khusus, yang mengeksekusi operasi data mining yang telah didefinisikan berdasarkan model analisis. Data mining merupakan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah data besar yang disimpan ketika menjalankan bisnis perusahaan. Menurut Larose (2005), kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan internet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan

Istilah data mining dan knowledge discovery in databases (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi

tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lainnya. Salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD itu ada 5 tahapan yang dilakukan secara terurut, yaitu:

1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing / cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan

metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation / evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini

merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.6. Pengelompokan Data Mining

Menurut Larose, Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpul suara mungkin tidak menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun dengan record lengkap menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analysis).

2.7. Klasifikasi

Klasifikasi adalah sebuah proses untuk menemukan sebuah model yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui (Tan et al.,2004). Dalam klasifikasi, diberikan sejumlah record yang dinamakan data latih, yang terdiri dari beberapa atribut yang dapat berupa kontinu ataupun kategoris, salah satu atribut menunjukkan kelas untuk record. Tujuan dari klasifikasi adalah untuk:

1. Menemukan model dari data latih yang membedakan record ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan record yang kelasnya belum diketahui sebelumnya pada testing set.
2. Mengambil keputusan dengan memprediksi suatu kasus, berdasarkan hasil klasifikasi yang diperoleh.

Untuk mendapatkan model, harus dilakukan analisis terhadap data latih, Sedangkan data uji digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu objek data. Proses klasifikasi data dapat dibedakan dalam dua tahap, yaitu:

1. Pembangunan Model

Tiap-tiap record yang digunakan dalam pembangunan model dianalisis berdasarkan nilai-nilai atributnya dengan menggunakan suatu algoritma klasifikasi untuk mendapatkan model.

2. Klasifikasi

Pengujian dilakukan untuk mengetahui tingkat akurasi dari model yang dihasilkan. Jika tingkat akurasi yang diperoleh sesuai dengan nilai yang ditentukan, maka model tersebut dapat digunakan untuk mengklasifikasikan record data baru yang belum pernah dilatihkan atau diujikan sebelumnya. Pembuatan model menguraikan sebuah set dari penentuan kelas-kelas sebagai:

- a. Setiap record diasumsikan sudah mempunyai kelas yang dikenal seperti ditentukan oleh label kelas atribut.
- b. Kumpulan record yang digunakan untuk membuat model disebut data pelatihan.
- c. Model direpresentasikan sebagai pola dalam penentuan klasifikasi. Penggunaan model menguraikan pengklasifikasian data yang akan diuji ataupun objek yang belum diketahui.

Adapun parameter keberhasilan dari model yang terdiri dari:

- a. Label yang telah diketahui dari data latih dibandingkan dengan hasil klasifikasi dari model.
- b. Nilai akurasi adalah persentase dari kumpulan data uji yang diklasifikasikan secara tepat oleh model.
- c. Data uji tidak sama dengan data latih.
- d. Jika sesuai, gunakan model untuk mengklasifikasi data record yang label kelasnya belum diketahui.

2.8. Metode Naive Bayes

Naïve Bayes merupakan metode probabilistik pengklasifikasian sederhana berdasarkan Teorema Bayes di mana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien (Hadiyani, 2013). Naïve Bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain (Pang-Ping, etc, 2006). Teorema Bayes sendiri dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Di mana persamaan teori Bayes tersebut adalah:

$$P(C|X) = \frac{p(X|C)p(c)}{p(X)}$$

Keterangan:

X : Data dengan kelas yang belum diketahui

C : Hipotesis data X merupakan suatu kelas spesifik

P(C|X): Probabilitas hipotesis C berdasar kondisi X (probabilitas posterior)

P(C) : Probabilitas hipotesis C (probabilitas prior)

P(X|C) : Probabilitas X berdasarkan kondisi pada hipotesis C

P(X) : Probabilitas X

Untuk menjelaskan Teorema Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, Teorema Bayes pada persamaan (1) disesuaikan menjadi persamaan (2):

$$P(C|X_1 \dots X_n) = \frac{p(c)p(X_1, \dots, X_n | C)}{p(X_1, X_n)}$$

Di mana Variabel C merepresentasikan kelas, sementara variabel X_1 , X_n merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi atau kriteria. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, sering kali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C

(disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana pada persamaan:

$$\text{Posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|X_1, \dots, X_n)$ menggunakan aturan perkalian sebagai berikut:

$$P(C|X_1, \dots, X_n) = P(C) P(X_1, \dots, X_n|C)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (naïf), bahwa masing-masing kriteria (X_1, X_2, \dots, X_n) saling bebas (independen) satu sama lain.

Naïve Bayes Classifier adalah konsep probabilitas penentuan kelompok kelas dokumen. Algoritma klasifikasi ini dapat mengolah data dalam jumlah besar dengan hasil akurasi yang tinggi. Algoritma Naive Bayes Classifier terdiri dari dua tahap. Tahap pertama adalah pelatihan terhadap himpunan dokumen contoh (data latih) dan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya (kelas).

Algoritma ini memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Karena asumsi atribut tidak saling terkait, maka:

$$V_{map} = \frac{\argmax_{vj \in V} P(Vj) \prod_k P(Wk|Vj)}{vj \in V}$$

Setelah diperoleh perhitungan untuk masing-masing kategori, maka kategori yang dipilih adalah yang memiliki nilai V_{map} terbesar. Nilai $P(V_j)$ ditentukan pada saat pelatihan, yang nilainya berdasarkan persamaan:

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

Keterangan :

$P(V_j)$: probabilitas setiap dokumen terhadap sekumpulan dokumen.

$|docs\ j|$: banyaknya dokumen yang memiliki kategori j dalam pelatihan.

$|contoh|$: banyaknya dokumen dalam contoh yang digunakan saat pelatihan

$$\text{Prior prob} = \frac{\sum \text{sentimen}}{\sum \text{kamus}}$$

Keterangan :

$\sum \text{sentimen}$ = total sentimen pada kamus positif atau negatif

$\sum \text{kamus}$ = total kamus positif dan negatif yang terbentuk

Untuk membobot tiap kata dengan pendekatan kamus yaitu titik pembobotan dengan nilai lebih terdapat pada kata yang telah terhimpun pada data latih, berikut persamaan yang digunakan pada pembobotan tiap kata.

$$\text{bobot} = \frac{1}{\sum \text{kamus} + \sum \text{sentimen}}$$

Dengan syarat pada sentimen data uji tidak memiliki kamus atau kata yang serupa pada kamus data latih sehingga nilai akan dikembalikan pada term frequency. Untuk pembobotan kata pada data uji yang memiliki kamus kata yang sama pada data latih akan membentuk persamaan seperti berikut.

$$\text{bobot} = \frac{1}{\sum \text{kamus} + \sum \text{sentimen}} \times 2$$

Kondisi pada persamaan diatas akan membentuk bobot lebih terhadap kata pada data uji yang memiliki bentuk yang sama terhadap kamus data latih. Sehingga akan membentuk persamaan klasifikasi seperti berikut.

$$\text{class} = n1 \times n2 \times n3 \times n4 \dots \dots \dots \times \text{prior prob}$$

Keterangan :

class : output, dalam kasus ini adalah sentimen yaitu positif dan negatif.

N : kata pada tiap tweet.

Prior prob : prior probabilitas pada output tiap class yang dihasilkan data latih.

2.9. Confusion Matrix

Data mining digunakan untuk mengukur kinerja model, ada beberapa cara untuk mengukur kinerja dari model yang dihasilkan salah satunya menggunakan confusion matriks (akurasi). Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau confidence adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau sensitivity adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

Tabel 2.1 Tabel *Confusion Matrix*

Aktual	Classified as	
	+	-
+	TP (<i>True positives</i>)	FN (<i>False negative</i>)
-	FP (<i>False positives</i>)	TN (<i>True negative</i>)

Perhitungan akurasi dengan tabel *Confusion Matrix* adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN}$$

Presisi didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut. Rumus presisi adalah:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Presisi dapat diberi nilai dalam bentuk angka dengan menggunakan perhitungan persentase (1-100%) atau dengan menggunakan bilangan antara 0-1. Sistem rekomendasi akan dianggap baik jika nilai presisi tinggi.

2.10. Penelitian Terdahulu

Sebelum adanya penelitian yang akan dilakukan ini, sudah terdapat beberapa penelitian relevan mengenai topik yang akan diangkat pada penelitian ini, pertama ialah penelitian yang dilakukan oleh Mujib Ridwan, Hadi Suyono, dan M. Sarosa (2013) 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.

Penelitian kedua yakni penelitian yang dilakukan oleh Ling (2014) mengenai aplikasi yang mampu menangani analisis sentimen menggunakan metode Naïve Bayes dan juga fitur Chi Square. Hasil Penelitian ini memperoleh hasil dimana tingkat akurasinya yaitu 83% dengan rata-rata harmonik 90.713%

Penelitian ketiga yakni dilakukan oleh Falahah dan Nur (2015) dalam judulnya “Pengembangan Aplikasi Sentimen Analisis Menggunakan Metode Naïve Bayes (Studi Kasus Sentimen Analisis dari Media Twitter)”, Penelitian ini mengklasifikasikan opini publik Twitter terkait layanan pemerintah terhadap masyarakat, berdasarkan sentimen positif, negatif atau netral. Metode Naïve Bayes Classifier dapat diterapkan sebagai metode untuk melakukan klasifikasi sentimen analisis.

Penelitian keempat yakni dilakukan oleh Wicaksono, Suyoto, & Pranowo (2016) yang menganalisis “Analisis sentimen terhadap pemilihan Presiden Amerika Serikat menggunakan pada sosial media dengan metode Naive Bayes. Pada penelitian ini yaitu Multinomial Naïve Bayes”. Hasil dari penelitian tersebut memaparkan bahwa Naive Bayes Classifier memperoleh tingkat akurasi sebesar 83%.

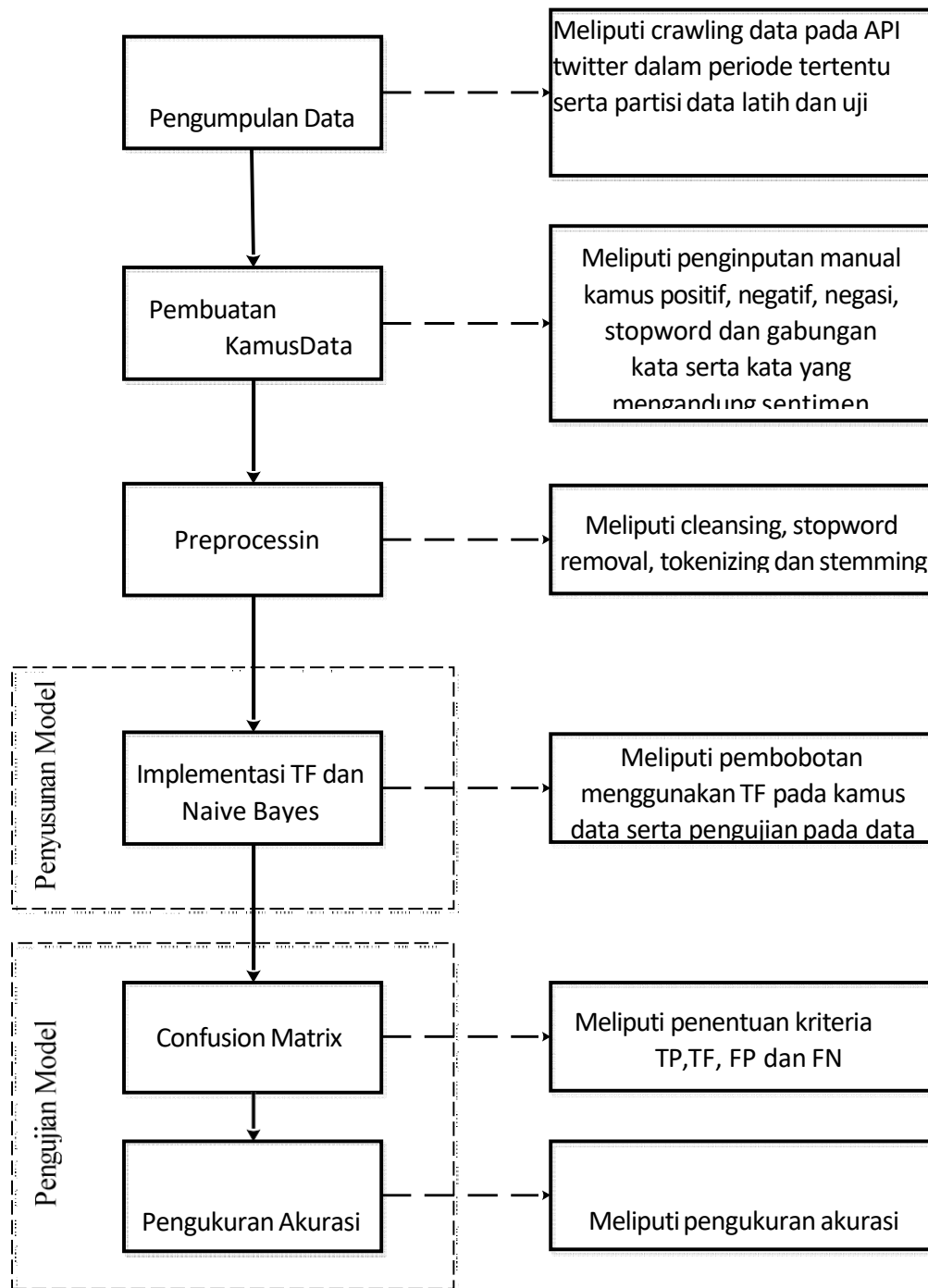
BAB 3. METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Penelitian ini adalah fokus terhadap menganalisis sentimen masyarakat terhadap sebuah topik, dimana sentimen masyarakat yang akan diteliti adalah opini masyarakat terkait karir *e-sport* di Twitter. Jenis penelitian yang dilakukan merupakan penelitian kuantitatif, mengingat kebutuhan akan banyaknya data dikarenakan penelitian ini menggunakan pendekatan pada kamus yang dipakai masyarakat dalam mengutarakan sentimennya. Penelitian kuantitatif dilakukan pada tahap penghitungan dan pemrosesan data berupa angka, hasil tweet berupa kalimat akan melalui berbagai macam proses dan pembobotan sehingga setiap kata akan mempunyai nilai. Perhitungan tersebut dilakukan menggunakan term frequency serta pengklasifikasian menggunakan metode Naive Bayes.

3.2. Tahapan Penelitian

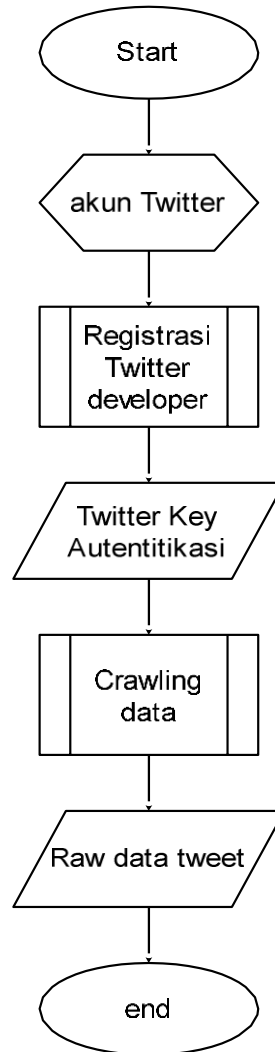
Tahapan penelitian untuk mengetahui klasifikasi sentimen masyarakat pada Twitter terhadap karir *e-sport* dilakukan dalam beberapa tahapan. Gambar proses dari tahapan penelitian dapat dilihat pada Gambar 3.1



Gambar 3.1. Tahapan Penelitian

3.3. Metode Pengumpulan Data

Pengumpulan data didapat dari hasil crawling opini masyarakat pada media sosial Twitter dari tanggal 6 November 2021 sampai 13 November 2021 dengan berbagai kata kunci dan tagar yang berhubungan dengan karir e-sports. Berikut alur flowchart pengambilan data.



Gambar 3.2. Alur Pengumpulan Data

Berdasarkan gambar 3.2 dapat dijelaskan bahwa langkah awal untuk mendapatkan data tweet adalah dengan melakukan registrasi pada laman Twitter developer yaitu <https://apps.Twitter.com>. Twitter akan melakukan verifikasi tujuan dan data kita dan memberikan key autentikasi untuk melakukan crawling data tweet. Selanjutnya data hasil akan dipecah sesuai kebutuhan penelitian, dalam kasus ini berdasar tujuan sentimen, data latih dan uji. Setelah tweet terkelompokkan menjadi beberapa bagian, selanjutnya akan disusun menjadi data latih dan data uji. Dalam penelitian ini persentase data latih adalah 80% dari total data dan data uji 20% dari total data.

3.4. Pembuatan Kamus Data

Pembuatan kamus data terdiri atas kamus positif, negatif, stopwords dan negasi.

a. Kamus positif dan negatif

Hal pertama yang dilakukan dalam pembentukan kamus positif dan negatif adalah menentukan nilai aktual dari sentimen tiap *tweet*. Dengan rujukan data *tweet* yang memiliki arah sentimen, selanjutnya peneliti dapat menganalisis dan menghimpun kata yang dianggap memiliki arah sentimen pada tiap *tweet* dan menjadikannya kamus data. Untuk contoh penerapan dapat dilihat pada Tabel 3.2.

Tabel 3.2. Teknik Pembentukan Kamus Positif dan Negatif

No.	ScreenName	Text	Aktual
1	Inggarpramudya h8	Saya setuju dengan @jessnolimits dengan perkembangan karir e- sport#kariresport	Positif
2	ananda_official	Aduhh @jessnolimits mana ada karir esport, emang ada pangkat2an #kariresport	Negatif

Berdasarkan Tabel 3.2 dapat dijelaskan bahwa nilai aktual adalah hasil dari kuesioner oleh ahli bahasa yang dijadikan nilai asli atau aktual. Teknik pembentukan kamus dengan menganalisis satu persatu tiap kata yang merujuk pada hasil sentimen. Pada contoh tabel 3.2. pada dokumen pertama memiliki arah sentimen positif, hasil analisis menunjukkan kata “setuju” merujuk pada sentimen positif, maka kata “setuju akan dijadikan kamus positif”. Sedangkan pada dokumen nomor dua, kata “Aduh” memiliki arti negasi bahwa tweet tersebut mengarah pada sentimen negatif. Selanjutnya kata “Aduhh” dijadikan kamus negatif.

b. Stopword

Stopword adalah kamus data yang berisi kata- kata yang keberadaannya dianggap tidak berpengaruh pada sentimen *tweet*. Dalam penelitian ini penulis memakai dua cara untuk penghimpunan kamus *stopword*. Pertama penulis menggunakan pihak ketiga yang telah menyediakan kamus *stopword* yaitu Sastrawi yang merujuk pada KBBI, artinya kamus dari pihak ketiga ini memiliki tatanan bahasa baku. Kedua, penulis menganalisis tiap *tweet*. Dengan mencari kata-kata yang disingkat oleh peng-*tweet*.

c. Negasi

Pembentukan kamus negasi akan digabung dengan pembentukan kamus yang diambil dari dua kata atau lebih yang memiliki satu arti.

Tabel 3.3. Contoh Kamus Negasi

No.	Kamus	Hasil konversi
1	gak jelas	gakjelas
2	ndak visioner	ndakvisioner
3	ga akan menang	gaakanmenang

Berdasarkan Tabel 3.3. dapat dijelaskan bahwa teknik pembentukan negasi adalah dengan membuat persamaan yaitu, *ingkaran + kata bermakna positif = kata bermakna negatif* dan jika *ingkaran + kata bermakna negatif = kata bermakna positif*, hal itu terjadi pada tabel no. 1 dan 2. Sedangkan tabel no. 3 adalah gabungan dua kata atau lebih yang memiliki satu arti dan keberadaannya tidak dapat dipisah, jika dipisah maka sentimen tidak akan didapat.

3.5. Preprocessing

Preprocessing bertujuan untuk mendapatkan data siap diproses pada pemodelan. Pada proses ini data *tweets* yang digunakan untuk data training dan data testing dibersihkan dari noise seperti link, “RT”, “@”, *stopword*, simbol, angka, konversi gambar dan video dan hashtag. Proses *preprocessing* terdiri dari berbagai tahapan yaitu *case folding*, *cleansing*, *stopword*, *convert emoticon*,

convert negation, tokenizer dan stemming. Berikut adalah penjelasan dari masing-masing tahapan:

a. *Case Folding*

Case Folding ialah proses merubah huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*). Hal ini dilakukan agar semua huruf menjadi seragam. Berikut ini adalah contoh *case folding*.

b. *Cleansing*

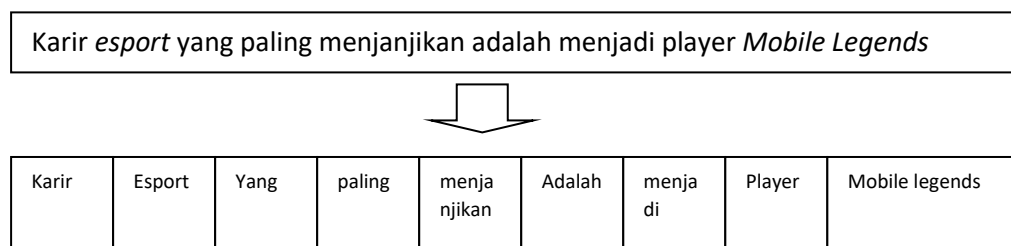
Tweet yang berhubungan dengan Pilpres memiliki berbagai komponen atau karakteristik *tweet* yaitu “#”, link dan RT, angka dan simbol. Komponen-komponen tersebut tidak memiliki pengaruh apapun terhadap sentimen, maka akan dibuang. Berikut ini adalah contoh cleansing.

c. *Convert Negation*

Convert Negation merupakan proses konversi kata-kata negasi yang terdapat pada suatu *tweet*, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada suatu *tweet*. Jika terdapat kata negasi pada suatu *tweet* maka kata tersebut akan disatukan dengan kata setelahnya. Contoh kata-kata negasi tersebut diantaranya “bukan”, “bkn”, “tidak”, “enggak”, “g”, “ga”, “jangan”, “nggak”, “tak”, “tdk”, dan “gak”.

d. *Tokenizing*

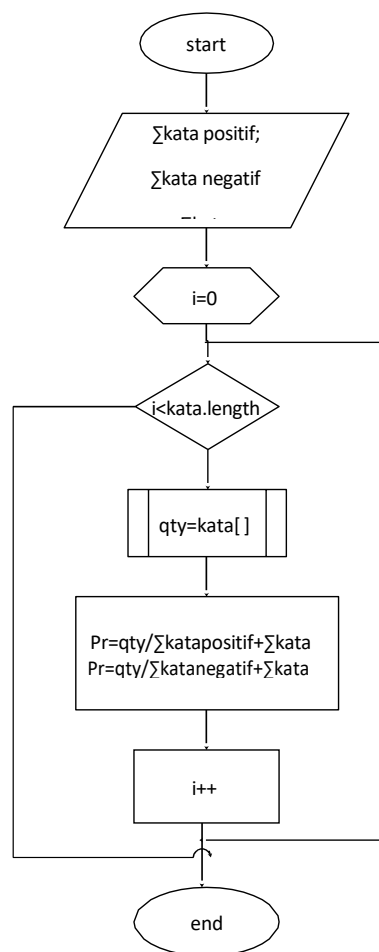
Proses *tokenizing* setiap kata pada *tweet* dipisahkan, pada proses ini tahap yang dilakukan adalah memisahkan setiap kata yang dipisahkan oleh spasi. Hal ini dilakukan agar tahap *preprocessing* selanjutnya dapat berjalan. Berikut ini adalah contoh *tokenizing*.



Gambar 3.3. Gambaran *Tokenizing*

3.6. Penyusunan Model

Tahap selanjutnya adalah tahap penyusunan model, dari hasil *preprocessing* akan dibobot menggunakan algoritma TF kemudian diklasifikasikan ke dalam dua class; positif dan negatif menggunakan metode Naïve Bayes Classifier. Berikut alur flowchart perhitungan Naive Bayes.



Gambar 3.4. Alur perhitungan *Naive Bayes*

Langkah-langkah dalam implementasi algoritma TF dan metode *Naive Bayes Classifier* adalah sebagai berikut :

a. Menghitung *prior probability*

Prior probability adalah menghitung probabilitas dari total kamus data. Berikut tergambar pada Tabel 3.5.

Tabel 3.5. Contoh hasil hitung *Prior Probability*

	Jumlah kata	Jumlah kata positif	Jumlah kata negatif	Probabilitas positif	Probabilitas negatif
Kata keseluruhan	26	22	4	0,846153846	0,181818182

Berdasarkan Tabel 3.5 dapat dijelaskan jumlah kata adalah total kamus data positif + kamus data negatif sedangkan untuk menghitung probabilitasnya adalah $\sum \text{kamus sentimen} / \sum \text{kamus data}$.

3.7. Pengujian Model

Pengujian model yang dihasilkan menggunakan *confusion matrix* untuk menentukan kriteria dan pengukuran tingkat akurasi.

3.7.1. Confusion Matrix

Hasil dari klasifikasi menggunakan *Naïve Bayes classifier* selanjutnya akan diukur tingkat akurasi dan presisi. Hasil klasifikasi ini dibandingkan dengan hasil klasifikasi manual dari kuesioner. Berdasarkan tabel *confusion matrix*, perhitungan di atas masuk dalam kriteria *True negative*.

Merujuk pada tabel *confusion matrix* pada bab 2 dapat dijabarkan untuk menentukan kriteria adalah sebagai berikut

- Jika nilai aktual *positive* dan klasifikasi *positive* maka kriteria bernilai *true positive* (TP)
- Jika nilai aktual *positive* dan klasifikasi *negative* maka kriteria bernilai *false negative* (FN).

- c. Jika nilai aktual *negative* dan klasifikasi *negative* maka kriteria bernilai *true negative* (TN).
- d. Jika nilai aktual *negative* dan klasifikasi *positive* maka kriteria bernilai *false positive* (FP).

3.7.2. Pengukuran Akurasi

Peneliti menggunakan pengukuran tingkat akurasi pada tiap profesi esport untuk mengetahui hasil kualitas model klasifikasi yang dihasilkan, dengan merujuk pada persamaan :

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN}$$

Persamaan di atas digunakan untuk mengukur tingkat akurasi semua calon.

3.8 Jadwal Kegiatan

Rencana jadwal penelitian adalah sebagai berikut :

No.	Nama Kegiatan	November				Desember				Januari				Februari			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Pengumpulan data																
2.	Pembuatan Kamus data																
3.	Preprocessing																
4.	Implementasi TF dan <i>Naïve Bayes Classifier</i>																
5.	Confusion Matrix																
6.	Pengukuran Akurasi																

DAFTAR PUSTAKA

- C. Aggarwal, C. 2015. Data Classification: Algorithms and Applications. *Berilustrasi peyunt*. New York: CRC Press.
- Larose, Daniel T. 2005. Discovering Knowledge in Data : An Introduction to Data Mining. John Willey & Sons, Inc.
- Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Manalu B. 2014. Analisis Sentimen Pada Twitter Menggunakan Text Mining. Prodi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi. Universitas Sumatera Utara, Medan
- Natalius, Samuel. 2010. Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen. *Jurnal Sekolah Teknik Elektro dan Informatika*. Institut Teknologi Bandung. Bandung.
- Simorangkir,H & Lhaksmana. 2018. Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naive Bayes Classifier. *e-Proceeding of Engineering : Vol.5*. Fakultas Informatika, Telkom University