

**ANALISIS SENTIMEN PADA *FINAL MOBILE LEGENDS*
PROFESSIONAL LEAGUE INDONESIA SEASON 5**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar Magister
Ilmu Komputer (M.Kom)

DESIANA NUR KHOLIFAH

14002219

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI

2020

**ANALISIS SENTIMEN PADA *FINAL MOBILE LEGENDS*
PROFESSIONAL LEAGUE INDONESIA SEASON 5**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar Magister
Ilmu Komputer (M.Kom)

DESIANA NUR KHOLIFAH

14002219

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI

2020

SURAT PERNYATAAN ORISINALITAS DAN BEBAS PLAGIARISME

Yang bertanda tangan di bawah ini:

Nama : Desiana Nur Kholifah
NIM : 14002219
Perguruan Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: **“Analisis Sentimen Pada *Final Mobile Legends Professional League Indonesia Season 5*”** adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 05 Agustus 2020
Yang menyatakan,

Desiana Nur Kholifah

HALAMAN PERSETUJUAN DAN PENGESAHAN TESIS

Tesis ini diajukan oleh:

Nama : Desiana Nur Kholifah
NIM : 14002219
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*
Judul Tesis : Analisis Sentimen Pada Final Mobile Legends Professional League Indonesia Season 5

Telah dipertahankan pada periode 2020-1 dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Magister Ilmu Komputer (M.Kom) pada Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 12 Agustus 2020

PEMBIMBING TESIS

Pembimbing I : Dr. Windu Gata, M.Kom



DEWAN PENGUJI

Penguji I : Dr. Yan Rianto, M.Eng



Penguji II : Dr. Rifki Sadikin, M.Kom



Penguji III /
Pembimbing I : Dr. Windu Gata, M.Kom



**LEMBAR KONSULTASI BIMBINGAN****TESIS****STMIK NUSA MANDIRI**

NIM : 14002219
Nama Lengkap : Desiana Nur Kholifah
Dosen Pembimbing : Dr. Windu Gata, M.Kom
Judul Tesis : "Analisis Sentimen Pada *Final Mobile Legends Professional League Indonesia Season 5*"

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	24 April 2020	Pengajuan Judul dan Proposal Tesis	
2	26 April 2020	Pengajuan BAB I dan BAB II	
3	16 Mei 2020	Revisi BAB I + BAB II dan Pengajuan BAB III	
4	30 Mei 2020	Acc BAB I + BAB II dan Revisi BAB III	
5	2 Juni 2020	Pengajuan Model dan Tahapan Penelitian	
6	22 Juni 2020	Acc BAB III + Model dan Pengajuan BAB IV + Deployment + BAB V	
7	25 Juni 2020	Revisi BAB IV + Deployment + BAB V	
8	04 Juli 2020	Acc Keseluruhan	

Catatan untuk dosen pembimbing

Bimbingan Tesis

- Dimulai pada tanggal : 24 April 2020
- Diakhiri pada tanggal : 04 Juli 2020
- Jumlah pertemuan bimbingan : 8 (delapan) kali bimbingan

Disetujui oleh,
Dosen Pembimbing

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis yang penulis ambil sebagai berikut “Analisis Sentimen Pada *Final Mobile Legends Professional League Indonesia Season 5*”.

Tujuan penulisan tesis ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

Tesis ini diambil berdasarkan hasil penelitian atau riset dari ulasan komentar pada *Final Mobile Legends Professional League Indonesia Season 5* di Live Facebook. Penulis juga lakukan melakukan pencarian dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dll yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinkanlah penulis kesempatan ini untuk mengucapkan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Windu Gata, M.Kom selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Bapak Supardi dan Ibu Nunuk Hartuti selaku orang tua, Riswandi Hasibuan selaku orang tercinta yang telah memberikan dukungan material dan moral kepada penulis.
3. Seluruh staf pengajar (dosen) Program Studi Ilmu Komputer STMIK Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
4. Seluruh staf dan karyawan Program Studi Ilmu Komputer STMIK Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.

5. Dll.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca pada umumnya.

Jakarta, 06 Agustus 2020

Yang menyatakan,



Desiana Nur Kholifah

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya:

Nama : Desiana Nur Kholifah
NIM : 14002219
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalti-Free Right)** atas karya ilmiah kami yang berjudul : “Analisis Sentimen Pada *Final Mobile Legends Professional League Indonesia Season 5*” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau bentuk-kan, mengelolanya dalam pangkalan data (database), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini .

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 05 Agustus 2020
Yang menyatakan,



Desiana Nur Kholifah

ABSTRAK

Nama : Desiana Nur Kholifah
NIM : 14002219
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*
Judul Tesis : Analisis Sentimen Pada *Final Mobile Legends Professional League Indonesia Season 5*

Smartphone dapat mempermudah aktivitas manusia yang antara lain dalam dunia hiburan. Salah satu hiburan yang digemari saat ini adalah bermain *game online*. *Game online* yang sedang populer saat ini adalah *Mobile Legends: Bang Bang* (MLBB). Dan saat ini game yang dimainkan pada perangkat mobile atau komputer, bertransformasi menjadi cabang olahraga yang dipertandingkan. Dan emosi pendukung masing-masing tim berubah ketika menyaksikan tim yang didukung bertanding. Dengan adopsi perangkat seluler yg luas dan akses yg mudah ke internet, emosi penggemar dapat bermanifestasi dalam tulisan mereka di media sosial. Penelitian ini akan menganalisa seberapa akurat komentar-komentar pada *final Mobile Legends Professional League Indonesia Season 5* saat live di facebook dengan menerapkan algoritma Support Vector Machine, Random Forest, dan K-Nearest Neighbor untuk mengukur seberapa tinggi akurasi yang didapatkan. Algoritma SVM dengan partitioning menghasilkan *accuracy* tertinggi sebesar 76,67% dan AUC sebesar 0,822.

Kata kunci:

Sentimen, Game Online, Text Mining, SVM, RF, KNN

ABSTRACT

Name : Desiana Nur Kholifah
NIM : 14002219
Study of Program : Computer Science
Levels : Strata Dua (S2)
Concentration : Data Mining
Title : Sentiment Analysis On Final Mobile Legends Professional League Indonesia Season 5

Smartphones can facilitate human activities, including in the world of entertainment. One of the popular pastimes is playing online games. The online game that is currently popular is Mobile Legends: Bang Bang (MLBB). And now, games that are played on a mobile device or computer are transformed into a sport that is contested. And the emotions of the supporters of each team change when watching the supported team compete. With the widespread adoption of mobile devices and easy access to the internet, fans' emotions can manifest in their posts on social media. This study will analyze how accurate the comments in the final Mobile Legends Professional League Indonesia Season 5 when live on Facebook by applying the Support Vector Machine, Random Forest, and K-Nearest Neighbor algorithms to measure how high the accuracy is obtained. The SVM algorithm with cross-validation produces the highest accuracy of 76,67% and AUC of 0,822.

Keywords:

Sentiment, Online Game, Text Mining, SVM, RF, KNN

DAFTAR ISI

	Halaman
ANALISIS SENTIMEN PADA <i>FINAL MOBILE LEGENDS PROFESSIONAL LEAGUE</i> INDONESIA SEASON 5	i
ANALISIS SENTIMEN PADA <i>FINAL MOBILE LEGENDS PROFESSIONAL LEAGUE</i> INDONESIA SEASON 5	1
KATA PENGANTAR	5
ABSTRAK	8
ABSTRACT.....	9
DAFTAR ISI.....	10
BAB I PENDAHULUAN.....	12
BAB II LANDASAN TEORI	5
BAB III METODE PENELITIAN.....	11
3.1 <i>Business Understanding</i>	13
3.2 Pemahaman Data (<i>Data Understanding</i>)	14
3.3 Persiapan Data (<i>Data Preparation</i>).....	14
3.4 Pemodelan (<i>Modelling</i>)	14
3.5 Evaluasi (<i>Evaluation</i>)	14
3.6 Penerapan (<i>Deployment</i>)	15
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	16
4.1 <i>Business Understanding</i>	16
4.2 <i>Data Understanding</i>	16
4.3 <i>Data Preparation</i>	17
4.3.1 @Anotation Removal	17
4.3.2 Tokenization: Regexp	18
4.3.3 Indonesian Stemming	18
4.3.4 Transformation: Not (Negative).....	18
4.3.5 Indonesian Stop Word Removal	19
4.3.6 N Chars Filter	19
4.3.7 Stop Word Filter.....	20
4.3.8 Bag of Words Creator	20
4.3.9 TF-IDF	21
4.5 <i>Evaluation</i>	22
4.6 <i>Deployment</i>	24

BAB V PENUTUP.....	25
5.1 Kesimpulan.....	25
5.2 Saran.....	25
DAFTAR PUSTAKA	26

BAB I

PENDAHULUAN

Perkembangan globalisasi saat ini memberi pengaruh besar terhadap kehidupan manusia terutama dalam bidang teknologi yang salah satunya adalah *smartphone* yang dapat mempermudah aktivitas manusia yang antara lain dalam bidang hiburan. Salah satu hiburan yang digemari saat ini adalah bermain *game online*. *Game online* merupakan sebuah permainan dengan jaringan, dimana interaksi antara satu orang dengan yang lainnya untuk mencapai tujuan, melaksanakan misi, dan meraih nilai tertinggi dalam dunia virtual. *Game online* yang sedang populer saat ini adalah *Mobile Legends: Bang Bang* (MLBB) yang merupakan jenis *game* bertajuk *Multiplayer Online Battle Arena* atau sering disebut MOBA [1].

Game Mobile Legends: Bang Bang ini berada dikategori populer dan sempat menjadi peringkat pertama di *Play Store*. Layaknya *game* MOBA pada umumnya, ada karakter-karakter yang tentunya menjadi favorit bagi setiap pemain. *Game* yang dirancang untuk *smartphone* ini memiliki aturan main yang berfokus pada salah satu dari kedua tim yang akan berjuang untuk mencapai dan menghancurkan basis musuh sambil mempertahankan basis mereka sendiri untuk mengendalikan jalan setapak, yaitu tiga jalur yang dikenal sebagai *top lane*, *middle lane*, dan *bottom lane* [2] [3].

E-Sport itu sendiri tidak hanya sebagai *game* yang dimainkan pada sebuah perangkat mobile atau komputer, saat ini bertransformasi menjadi cabang olahraga yang dipertandingkan dalam skala nasional maupun internasional. *E-Sport* berkembang dengan adanya turnamen dan kompetisi yang diselenggarakan oleh sponsor [4]. Seperti babak final *Mobile Legends Professional League* (MPL) Indonesia Season 5 ketika tim *e-sport* RRQ Hoshi melawan Evos Legends dan RRQ sukses keluar menjadi juara. Laga final MPL ID Season 5 berlangsung dengan format *best of 5*, pertarungan berjalan dengan sengit dan RRQ berhasil mengalahkan Evos dengan skor 3-2 [5].

Emosi pendukung masing-masing tim berubah selama pertandingan, wajah dan gerakan tubuh mereka adalah cara alami untuk mengekspresikan emosi mereka. Dengan adopsi perangkat seluler yg luas dan akses yg mudah ke internet, dapatkah emosi penggemar dapat bermanifestasi dalam tulisan mereka di media sosial dan melalui interaksi mereka dengan orang lain melalui internet dan perangkat seluler? [6]. Berdasarkan ringkasan [7] tentang motivasi para penggemar olahraga untuk mengkonsumsi permainan yang dimediasi, penelitian terbaru di media sosial dan komunikasi elektronik menunjukkan bahwa penggemar olahraga termotivasi untuk menggunakan media sosial untuk berbagai tujuan, termasuk pelepasan emosional [6] [8] [9].

Media sosial telah menjadi fenomena yang muncul karena kemajuan besar dan cepat dalam teknologi informasi. Orang-orang menggunakan media sosial setiap hari untuk mengkomunikasikan pendapat mereka satu sama lain tentang beragam subjek, produk, dan layanan, yang menjadikannya sumber daya yang kaya untuk penambangan teks dan analisis sentimen[10]. Facebook merupakan sosial online terbesar dan terpopuler, telah menjadi bagian penting dari kehidupan sehari-hari untuk hampir 1,4 miliar orang di seluruh dunia [11].

Terdapat penelitian terkait penelitian tentang klasifikasi sentimen seperti klasifikasi pertandingan sepak bola menggunakan data tweet dari *twitter* menggunakan *Support Vector Machine* (SVM), *Multinomial Naïve Bayes* (MNB), dan *Random Forest* (RF) yang mengkombinasikan dengan variasi parameter dan performa *accuracy* sebesar 79,4%, *precision* sebesar 72,68%, *recall* sebesar 51,53% dan *f1 score* sebesar 60,3% untuk model SVM, MNB menghasilkan skor *accuracy* sebesar 75,48%, *precision* sebesar 71,11%, *recall* sebesar 67,14% dan *f1 score* sebesar 68,3%, sedangkan *accuracy* sebesar 78,7%, *precision* sebesar 77,2%, *recall* sebesar 70,02% dan *f1-score* sebesar 71,94% untuk model RF [12]. Kemudian pada penelitian tentang klasifikasi sentimen terhadap polemik persepakbolaan Indonesia dengan pembobotan TF-IDF dan KNN dengan menghasilkan akurasi optimal pada nilai $k=23$ sejumlah 79,99% [13].

KNN ini merupakan metode yang sederhana untuk pengklasifikasian tanpa harus melakukan perhitungan secara kompleks, oleh sebab itu KNN ini juga biasa disebut *lazy learning* [13]. Kelebihan dari algoritma *Support Vector Machines*

adalah dari kemampuannya untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi [14]. *Random Forest* dapat menghindari *overfitting* pada sebuah dataset saat mencapai akurasi yang maksimum [15]

Dalam facebook pada halaman Mobile Legends: Bang bang tentang final MPL yang banyak berisi komentar maupun pendapat dari pengguna *facebook* yang mengekspresikan emosinya terkait final MPL ID *Season 5 Mobile Legends Bang Bang* yang digunakan sebagai sumber data untuk dilakukan penelitian terkait dengan analisis sentimen kemudian akan dilakukan klasifikasi dengan mengusulkan metode *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN) dan *Random Forest* (RF).

Opini atau pendapat pengguna facebook terhadap final *Mobile Legends Professional League Season 5* sangat beragam dalam mengekspresikan emosinya dalam tulisan yang pada komentar ketika berlangsungnya turnamen Final MPL Indonesia *Season 5 Mobile Legends Bang Bang*. Dan algoritma manakah yang lebih efektif diantara SVM, KNN, dan RF dalam melakukan analisis sentimen MPL Indonesia *Season 5*.

Adapun tujuan dari penelitian ini untuk mengidentifikasi komentar yang beragam dalam mengekspresikan emosinya terhadap Final MPL Indonesia *Season 5*. Oleh karena itu agar komentar-komentar tersebut mudah dipahami tentang makna cuitan mereka, diklasifikasikan menjadi dua kelas yaitu positif atau negatif. Dan untuk mengukur keberhasilan klasifikasi menggunakan metode SVM, KNN, dan RF.

Pengumpulan data dilakukan dengan crawling pada facebook halaman Mobile Legends: Bang Bang dan analisis sentimen komentar-komentar pengguna facebook terhadap Final turnamen MPL Indonesia *Season 5 Mobile Legends Bang Bang* dengan menerapkan metode usulan yaitu SVM, KNN, dan RF.

BAB II

LANDASAN TEORI

Penelitian yang akan dilakukan perlu adanya tinjauan pustaka yang sesuai dengan tujuan penelitian untuk mengetahui metode dalam penerapan analisis sentimen serta mengetahui dasar sumbernya.

Media sosial *facebook* merupakan sosial online terbesar dan terpopuler, telah menjadi bagian penting dari kehidupan sehari-hari untuk hampir 1,4 miliar orang di seluruh dunia [11]. Berkantor pusat di Menlo Park, California, Amerika Serikat yang diluncurkan pada bulan Februari 2004. Pengguna harus mendaftar sebelum dapat menggunakan situs ini. Setelah itu, pengguna dapat membuat profil pribadi, menambahkan pengguna lain sebagai teman, dan bertukar pesan, termasuk pemberitahuan otomatis ketika mereka memperbarui profilnya. Selain itu, pengguna dapat bergabung dengan grup pengguna dengan ketertarikan yang sama, diurutkan berdasarkan tempat kerja, sekolah atau perguruan tinggi, atau ciri khas lainnya, dan mengelompokkan teman-teman mereka ke dalam daftar seperti "Rekan Kerja" atau "Teman Dekat".

Data mining sering disebut juga dengan *Knowledge Discovery in Database* (KDD), adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [16]. Secara garis besar *data mining* dikelompokkan menjadi 2 kategori utama yaitu:

1. *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

Text mining merupakan sebuah cara untuk mendapatkan pola yang signifikan dan menarik dengan tujuan mendapatkan pengetahuan dari sumber data berupa teks. *Text mining* mengolah data berupa data yang tersimpan dalam bentuk semi terstruktur maupun tidak terstruktur. Dalam *text mining* terdapat beberapa teknik yang biasa dilakukan seperti *summarization*, *classification*, dan *clustering* yang bisa diterapkan untuk mengekstrak pengetahuan [17].

Secara umum proses yang dilakukan untuk mengekstrak pengetahuan dalam text mining dijelaskan sebagai berikut:

1. Mengumpulkan data yang tidak terstruktur dari berbagai sumber yang tersedia.
2. Melakukan tahapan *preprocessing* terhadap data yang didapat. Tahap ini berfungsi untuk membersihkan data untuk membuang yang tidak diperlukan dan hanya menjaga data yang penting.
3. Memproses data dan mencari atau mengenali pola dari data tersebut.
4. Mengekstrak pengetahuan yang relevan dari data.

Crawling data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari suatu database [18] [19] [20].

Data yang sudah dikumpulkan kemudian dilakukan preprocessing untuk menghindari data yang belum siap untuk diolah seperti terdapat gangguan (*noise*) dan kata yang tidak konsisten [13] [24]. Tahapan pada preprocessing sebagai berikut:

1. *Case Folding dan Cleansing*

Pada tahap ini dilakukan penyeragaman seluruh teks menjadi huruf kecil (*lowecase*) dan pembersihan atau penghapusan pada semua dokumen yang berisi angka, url (<http://>), *username* (@), *hashtag* (#), *delimiter* seperti koma(,), dan titik (.) dan juga tanda baca lainnya.

2. *Tokenizing*

Di tahap ini dilakukan proses pemotongan pada sebuah dokumen ataupun kalimat menjadi kata atau biasa disebut *token*.

3. Normalisasi Kata

Pada tahap ini dilakukan proses untuk menormalisasi kata tidak baku maupun *noise* menjadi kata yang baku dan siap diolah.

4. *Stopword Removal*

Stopword Removal adalah suatu proses untuk menghapus kata yang dianggap tidak seperti contoh kata ‘di’, ‘yang’, ‘dan’, ‘ke’, dan semua kata yang terdapat dalam kamus *stopword* yang sudah dibuat. Tujuan dari proses ini adalah untuk mengurangi jumlah kata yang disimpan dalam daftar *token* yang nantinya akan dilakukan proses selanjutnya.

5. Stemming

Pada tahap terakhir preprocessing ini dilakukan proses untuk mengubah semua kata-kata pada dokumen menjadi kata dasar dengan menghilangkan semua kata imbuhan. Kata imbuhan yang dihilangkan terdiri dari awalan (*prefix*), akhiran (*suffix*), sisipan (*infix*), dan gabungan awalan-akhiran (*confix*).

SVM dikembangkan pada tahun 1992 oleh Vladimir Vapnik dan rekannya, Bernhard Boser dan Isabelle Guyon yang dikembangkan dari teori *structural risk minimization*. SVM adalah sebuah algoritma yang bekerja menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi dalam dimensi yang baru, kemudian akan mencari linear optimal pemisah *hyperplane* (yaitu, “*decision boundary*” yang memisahkan tupel dari satu kelas dengan kelas lainnya) [22]. Banyak peneliti telah melaporkan bahwa *Support Vector Machine* metode yang paling akurat untuk teks klasifikasi [14].

Langkah awal suatu algoritma SVM adalah pendefinisian persamaan suatu *hyperplane* pemisah yang dituliskan dengan persamaan 2.1 [22].

$$w \cdot X + b = 0 \dots\dots\dots (2.1)$$

Dimana w merupakan bobot vektor, yaitu $w = \{w_1, w_2, \dots, w_n\}$; n adalah jumlah atribut dan b merupakan suatu skalar yang disebut dengan bias. Jika berdasarkan atribut A_1 dan A_2 , dan jika b dianggap sebagai suatu pembobot tambahan w_0 , maka persamaan suatu *hyperplane* pemisah dapat ditulis ulang seperti persamaan 2.2.

$$w_0 + w_1x_1 + w_2x_2 = 0 \dots\dots\dots (2.2)$$

Setelah persamaan dapat didefinisikan nilai x_1 dan x_2 dapat dimasukkan ke dalam persamaan untuk mencari bobot w_1 , w_2 , dan w_0 atau b . SVM menemukan *hyperlane* pemisah maksimum, yaitu *hyperlane* yang mempunyai jarak maksimum antara tupel pelatihan terdekat. *Support vector* juga ditunjukkan

dengan batasan tebal pada titik tupel. Dengan demikian, setiap titik yang terletak diatas *hyperplane* pemisah memenuhi persamaan 2.3.

$$w_0 + w_1x_1 + w_2x_2 > 0 \dots\dots\dots (2.3)$$

Sedangkan, titik yang terletak dibawah *hyperlane* pemisah memenuhi rumus seperti pada persamaan 2.4.

$$w_0 + w_1x_1 + w_2x_2 < 0 \dots\dots\dots (2.4)$$

Melihat dua kondisi di atas, maka didapatkan dua persamaan *hyperplane*, seperti pada persamaan 2.5 dan 2.6

$$H_1: w_0 + w_1x_1 + w_2x_2 \geq 0 \dots\dots\dots (2.5)$$

$$H_2: w_0 + w_1x_1 + w_2x_2 \leq 0 \dots\dots\dots (2.6)$$

Artinya, setiap tupel yang berada di atas H_1 memiliki kelas +1, dan setiap tupel yang berada di bawah H_2 memiliki kelas -1. Perumusan model SVM menggunakan trik matematika yaitu *lagrangian formulation* [22]. Berdasarkan *Lagrangian Formulation, Maksimum Margin Hyperplane* (MMH) dapat ditulis ulang sebagai suatu batas keputusan (*decision boundary*) ditulis dengan persamaan 2.7.

$$d(X^T) = \sum_{i=1}^l y_i a_i X_i X^T + b_0 \dots\dots\dots (2.7)$$

y_i adalah label kelas dari support vector X_i , X^T merupakan suatu tuple *test*. a_i dan b_0 adalah parameter numerik yang ditentukan secara otomatis oleh optimalisasi algoritma SVM dan l adalah jumlah *vector support*.

KNN merupakan suatu pendekatan klasifikasi yang mencari semua data latih yang relatif mirip dengan data uji. Teknik klasifikasi ini disebut *lazy learning* karena teknik ini tidak membangun model model klasifikasi terlebih dahulu, seperti: pohon keputusan (*decision tree*), klasifikasi berbasis aturan (*rule-based*) [23].

Cara kerja KNN ini adalah melakukan klasifikasi berdasarkan data latih yang jaraknya paling dekat dengan objek data uji. Setiap data pada data latih memiliki atribut kelas, selanjutnya adalah dengan menguji model dengan data uji yang tidak memiliki kelas [13]. Adapun tahapan proses pada KNN adalah sebagai berikut:

1. Tahapan awal adalah menentukan nilai K, misal k=23 artinya 23 dokumen terdekat dengan dokumen uji yang akan diambil.

2. Menghitung jarak antara data baru disetiap label data Untuk menghitung tingkat kesamaan dalam dokumen ini menggunakan *Euclidean distance*. Cara perhitungan *euclidean distance* adalah dengan rumus berikut.

$$D(X, Y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \dots\dots\dots (2.8)$$

Keterangan (2.8):

D = Jarak antara dua titik x dan y

X = Data uji

Y = Sampel data

n = Dimensi data

3. Kemudian urutkan hasil euclidean distance berdasarkan jarak K yang ditentukan jika K=3 artinya akan dipilih 3 jarak terkecil dari hasil *euclidean distance*.
4. Selanjutnya gunakan mayoritas atribut kelas pada 3 tetangga terdekat yang sudah dipilih untuk menentukan prediksi kelas pada data baru tersebut. Misalkan pada 3 tetangga terdekat memiliki 2 atribut kelas positif dan 1 atribut kelas negatif maka kelas pada data baru tersebut adalah positif.

Random Forest adalah pengembangan dari metode CART, yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection*. Metode ini merupakan metode pohon gabungan. Dalam *random forest*, banyak pohon ditumbuhkan sehingga terbentuk suatu hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut [15].

Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari simpul akar, simpul internal, dan simpul daun dengan mengambil atribut data acak dan ketentuan terkait yang belaku. Simpul akar adalah simpul yang terletak dibagian atas, atau biasa disebut akar dari pohon keputusan. Node internal adalah node cabang, dimana ia memiliki output setidaknya dua dan hanya satu input. Sedangkan node daun atau terminal node adalah node terakhir yang hanya memiliki satu input dan tidak memiliki output [24]. Untuk menghitung nilai *entropy* dan nilai *information gain* dapat menggunakan persamaan sebagai berikut:

$$Entropy(Y) = - \sum_i p(c|Y) \log^2 p(c|Y) \dots\dots\dots (2.9)$$

Dimana Y adalah himpunan kasus dan $p(c|Y)$ nilai Y) adalah proporsi dari kelas c. *Information Gain*:

$$(Y, a) = Entropy(Y) \sum_{v \in Value(a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \dots \dots \dots (2.10)$$

Dimana nilai (a) mewakili semua nilai yang mungkin dalam satu set kasus. YV adalah sub kelas dari kelas Y yang berkaitan dengan kelas a. Pemilihan atribut sebagai node, baik *root* (akar) atau internal node berdasarkan *information gain* adalah yang tertinggi dari atribut yang ada.

$$Split\ Information(S, A) = \sum_{i=1}^c \left(\frac{|S_i|}{|S|} \right) \log^2 \left(\frac{|S_i|}{|S|} \right) \dots \dots \dots (2.11)$$

Dimana informasi *split* (S,A) adalah nilai estimasi input variabel *entropy* dari S yang memiliki kelas c dan $|S_i|/|S|$ adalah probabilitas atribut kelas i.

$$Gain\ Ratio(S, A) = \frac{Information\ Gain(S, A)}{Split\ information(S, A)} \dots \dots \dots (2.12)$$

Penggunaan metode *random forest* memiliki keuntungan yang kuat dalam hal kompleksitas komputasi, mudah dan cepat dalam penerapannya, khususnya dalam menghasilkan nilai prediksi akurasi, dapat menangani *missing value*, tidak terlalu sensitif pada *outlier* dan dapat menangani sejumlah besar inputan variabel tanpa harus *overtfitting* [25].

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah suatu proses untuk melakukan transformasi data dari data tekstual ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur. TF- IDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen. TF adalah frekuensi kemunculan kata pada di tiap dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam tiap dokumen tersebut. DF adalah frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. IDF adalah inverse dari nilai DF. Hasil dari pembobotan kata menggunakan TF-IDF ini adalah hasil perkalian dari TF dikalikan dengan IDF. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen [13]. Adapun rumus dari pembobotan kata TD-IDF adalah:

$$W_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \log \frac{N}{df_t} \dots \dots \dots (2.13)$$

Keterangan :

$W_{t,d}$ = Bobot TF-IDF

$tf_{t,d}$ = Jumlah frekuensi kata

idf_t = Jumlah inverse frekuensi dokumen tiap kata

df_t = Jumlah frekuensi dokumen tiap kata

N = Jumlah total dokumen

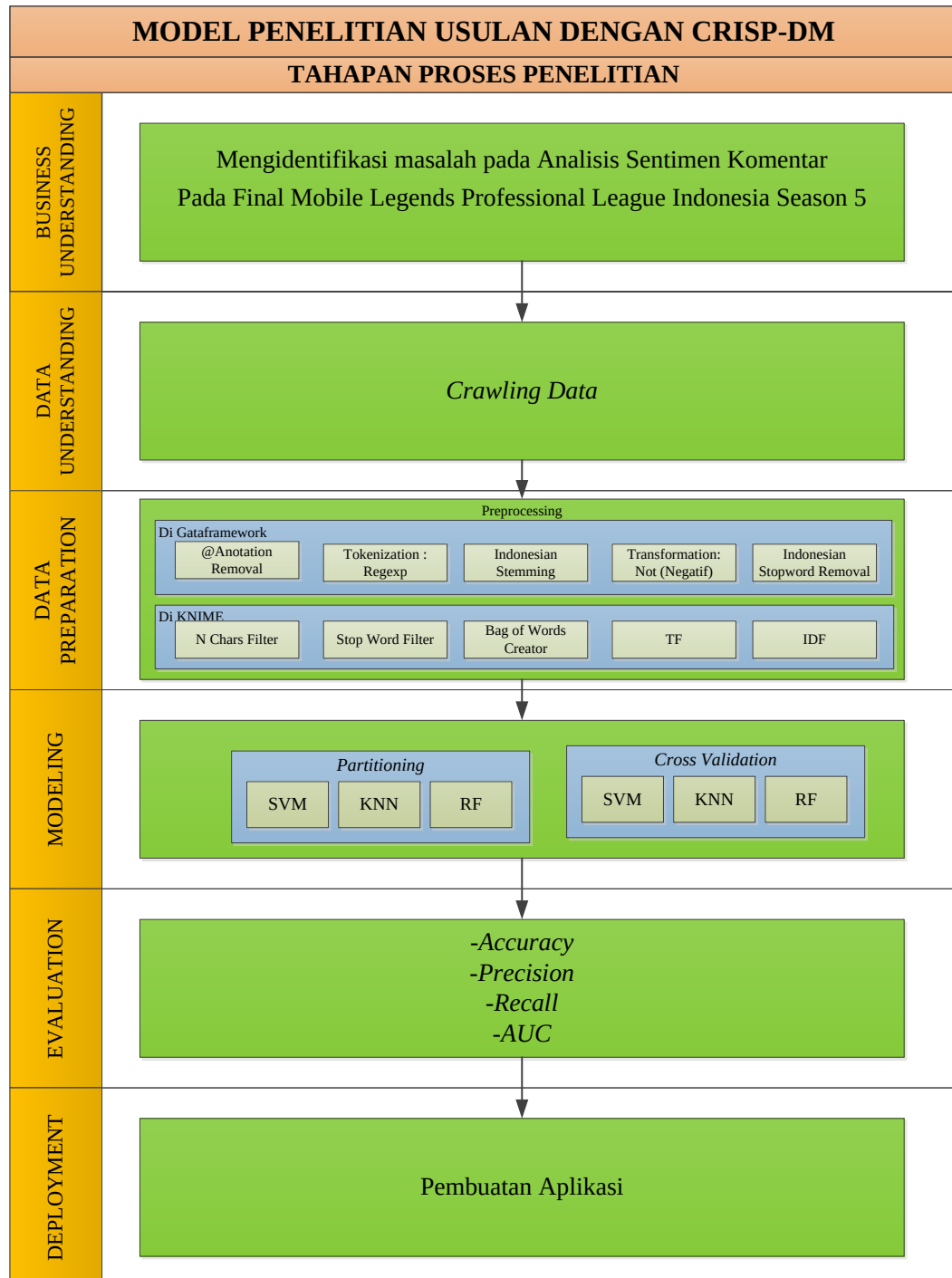
BAB III

METODE PENELITIAN

Penelitian merupakan suatu proses kegiatan mencari kebenaran terhadap suatu fenomena ataupun fakta yang terjadi dengan cara yang terstruktur dan sistematis. Metode penelitian yang digunakan dalam penelitian ini yaitu dengan menggunakan metode eksperimen. Tujuan dari metode eksperimen adalah untuk menguji efektifitas dan efisiensi dari suatu pendekatan, metode, teknik atau media pengajaran dan pembelajaran sehingga hasilnya bisa diterapkan jika memang baik atau tidak digunakan jika memang tidak baik. Penelitian ini bertujuan untuk mengukur keberhasilan metode yang diusulkan dalam mengklasifikasi komentar pada halaman *facebook Mobile Legends: Bang Bang* yang menampilkan final MPL Indonesia Season 5 menggunakan metode SVM, KNN, dan RF.

Metode penelitian yang digunakan dalam penelitian ini menggunakan metode penelitian standar yang digunakan dalam *data mining* yaitu CRISP-DM. Berikut 6 tahap dalam CRISP-DM yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation* dan *Deployment*.

Berikut ini adalah model yang diusulkan dalam penelitian dengan menggunakan metode CRISP-DM:



Gambar 3.1 Model Penelitian Usulan

3.1 *Business Understanding*

Pada tahap ini merupakan tahap pemahaman penelitian untuk memahami bidang masalah dan menentukan tujuan dari penelitian. Opini atau pendapat pengguna facebook pada Final *Mobile Legends Professional League Season 5*

sangat beragam dalam mengekspresikan emosinya dalam tulisan pada kolom komentar ketika *Live Final* turnamen MPL Indonesia *Season 5 Mobile Legends Bang Bang*. Oleh karena itu agar komentar-komentar tersebut mudah dipahami tentang makna cuitan mereka, diklasifikasikan menjadi dua kelas yaitu positif dan negatif.

3.2 Pemahaman Data (*Data Understanding*)

Tahap kedua dari CRISP-DM adalah *Data Understanding*. Ditahap ini peneliti bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki. Data yang digunakan dalam penelitian ini adalah data dari komentar penonton di *Live facebook Final MPL Indonesia Season 5*.

3.3 Persiapan Data (*Data Preparation*)

Tahap ketiga dari CRISP-DM adalah *Data Preparation*. Ditahap ini data awal yang diperoleh dari *crawling* data yaitu dari komentar penonton di *Live facebook Final MPL Indonesia Season 5*. Setelah itu dilakukan *cleansing* data untuk mengurangi seperti data yang redundan, kemudian pada tahap *preprocessing* awal dilakukan pemrosesan data menggunakan *gataframework*. Proses ini mencakup proses penyeragaman dokumen menjadi lowercase, memfilter semua dokumen yang mempunyai notasi @, hashtag (#), tanda baca titik (.), koma (,) dan tanda baca lainnya, Indonesian stemming, *transformation not (negatif)*, dan *Indonesian stop word removal*. Kemudian proses *preprocessing* pada KNIME meliputi proses memfilter kata yang kurang dari N karakter, memfilter kata yang terdapat dalam daftar *stop word*, *bag of words*, dan pembobotan menggunakan TF-IDF.

3.4 Pemodelan (*Modelling*)

Dalam tahap modelling akan dilakukan teknik pengklasifikasian data. Untuk mengkomparasi, pada penelitian ini akan digunakan algoritma SVM, RF dan KNN menggunakan validasi *partitioning* dengan algoritma SVM, RF, dan KNN menggunakan validasi *cross validation* dengan tools KNIME.

3.5 Evaluasi (*Evaluation*)

Model yang terbentuk akan ditentukan pengujian untuk nilai akurasi. Tahap pengujiannya digambarkan dalam hasil akurasi mulai dari prediksi positif yang benar dan yang salah, dan prediksi negatif yang benar dan yang salah. Semakin

tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Pengujian juga diukur dengan menggunakan *ROC Curve*, yang akan menggambarkan kelas positif dalam bentuk kurva. Pengujian akan dilakukan dengan menghitung nilai AUC, semakin tinggi nilai AUC dan *ROC Curve*, maka semakin baik model yang terbentuk.

3.6 Penerapan (*Deployment*)

Tahap ini adalah tahap terakhir pada CRISP-DM, yaitu hasil dari seluruh tahapan yang sebelumnya diterapkan secara nyata. Penerapan dalam penelitian ini akan dikembangkan dengan menggunakan PHP dan MySQL.

Adapun rincian waktu kegiatan penelitian (*Timeline*) yang dilakukan dibuat dalam bentuk table yang ditampilkan seperti pada table 3.1 di bawah ini:

Tabel 3.1 Jadwal Kegiatan (*Timeline*)

No	Keterangan	April				Mei				Juni				Juli			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Analisis kebutuhan																
2	Studi literature																
3	Penentuan judul dan latar belakang																
4	Merancang metode penelitian																
5	Pengolahan data																
6	Penerapan model																
7	<i>Deployment</i>																
8	Dokumentasi																

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 *Business Understanding*

Pada tahap *Business Understanding*, dilakukan pemahaman terhadap objek penelitian yang dilakukan dengan menggali informasi melalui kolom komentar pada *live Final MPL Indonesia Season 5* di Facebook dari tanggal 12 April 2020 dari pukul 18.00 WIB sampai 23.01 WIB.

Pada tahap ini juga dilakukan pemahaman untuk mencari metode dengan pendekatan model analisis sentimen terbaik agar dapat membantu untuk menentukan model yang tepat untuk digunakan berdasarkan perbandingan hasil algoritma. Adapun algoritma yang digunakan yaitu algoritma SVM, RF, dan KNN menggunakan *partitioning* dan algoritma SVM, RF, dan KNN menggunakan *cross validation*.

4.2 *Data Understanding*

Ditahap *data understanding*, dimana proses memahami data yang akan digunakan sebagai bahan yang akan diteliti. Tahap ini dilakukan proses pengambilan data mentah sesuai dengan atribut yang dibutuhkan. Data diperoleh dari kolom komentar *Live Final MPL Indonesia Season 5* pada url https://m.facebook.com/story/graphql_permalink/?graphql_id=UzpfSTgyMDc1MTU1ODA3NjY2MDpWSzoyMDk4NjUxNDY5NzQ4OTQ%3D pada 12 April 2020

4.3 Data Preparation

Tahap *data preparation* merupakan tahap proses penyiapan data yang bertujuan mendapatkan data menjadi bersih dan siap untuk digunakan dalam penelitian. Data yang diambil pada awalnya adalah 4.403 data. Proses selanjutnya akan dilakukan proses *cleansing* seperti menghilangkan *duplicate data* dan menghilangkan data yang tidak berhubungan dengan topik penelitian, dan diperoleh 1.037 data. Dengan menggunakan sumber data yang diperoleh, dibuat dataset dengan atribut Komentar yang berisi komentar yang diambil yang sudah di *cleansing* lalu atribut sentimen yang merupakan label yang berisi positif dan negatif. Proses pelabelan dilakukan *annotator* dari pihak komunitas *Mobile legends*. Adapun hasil dari tahap ini sebagai berikut:

Tabel 4.1 Tabel Hasil Proses Pelabelan

Label	Jumlah
Positif	528
Negatif	509

Pada tahap ini dilakukan *text preprocessing*, pada tahap ini peneliti menggunakan tools Knime. Berikut tahapan dalam *preprocessing*:

4.3.1 @Anotation Removal

Pada *@anotation removal* ini adalah proses untuk menghilangkan teks yang memiliki anotasi @ dan menyeragamkan tulisan huruf besar (*uppercase*) diubah ke huruf kecil (*lowercase*).

Tabel 4.2 Perbandingan Teks Sebelum dan Sesudah Dilakukan @Anotation Removal

@Anotation Removal	
Data Sebelum	Data Sesudah
#viva_RRQ #RRQHoshi semangat @XIN @LEMON @R7 @XYI @LJ	#viva_rrq #rrqhoshi semangat

4.3.2 *Tokenization: Regexp*

Proses *tokenization: regexp* dilakukan untuk menghilangkan tanda baca,, simbol, serta karakter khusus ataupun bukan huruf dalam suatu teks.

Tabel 4.3 Perbandingan Teks Sebelum dan Sesudah Dilakukan *Tokenization: Regexp*

<i>Tokenization: Regexp</i>	
Data Sebelum	Data Sesudah
Sayangnya Final taun ini serasa hambar . Karna kita sebagai suporter dri kedua finalis tidak bisa menonton langsung ke area turnamen ... #VivaRRQ. Hormat untuk #evosRoar ..	sayangnya final taun ini serasa hambar karna kita sebagai suporter dri kedua finalis tidak bisa menonton langsung ke area turnamen vivarrq hormat untuk evosroar

4.3.3 *Indonesian Stemming*

Pada *Indonesian stemming* ini untuk mencari kata dasar (*root*) dari tiap kata hasil dari proses *filtering* dokumen. Pencarian kata dasar dapat memperkecil hasil indeks tanpa harus menghilangkan makna.

Tabel 4.4 Perbandingan Teks Sebelum dan Sesudah Dilakukan *Indonesian Stemming*

<i>Indonesian Stemming</i>	
Data Sebelum	Data Sesudah
Yeah RRQ tim favorit saya memenangkan MPL musim 5	yeah rrq tim favorit saya menang mpl musim

4.3.4 *Transformation: Not (Negative)*

Pada tahap ini dilakukan untuk menghilangkan kata-kata yang memiliki terbalik (ambigu) seperti kata yang tidak konsisten yang mengandung arti positif.

Tabel 4.5 Perbandingan Teks Sebelum dan Sesudah Dilakukan *Transformation: Not (Negative)*

<i>Transformation: Not (Negative)</i>

Data Sebelum	Data Sesudah
Sayangnya Final taun ini serasa hambar . Karna kita sebagai suporter dri kedua finalis tidak bisa menonton langsung ke area turnamen ... #VivaRRQ. Hormat untuk #evosRoar ..	sayang final taun ini rasa hambar karna kita bagai suporter dri dua finalis tidak_bisa tonton langsung ke area turnamen vivarrq hormat untuk evosroar

4.3.5 Indonesian Stop Word Removal

Pada *Indonesian stop word removal* pada *gata framework* digunakan untuk menghapus kata-kata yang tidak memiliki arti.

Tabel 4.6 Perbandingan Teks Sebelum dan Sesudah *Indonesian Stop Word Removal*

<i>Indonesian Stop Word Removal</i>	
Data Sebelum	Data Sesudah
RRQ, EVOS dan ONIC tetap 3 Team Terbaik Indonesia...Siapapun yang Juara cuma faktor beruntung saja dan hasil dari kerja keras selama ini ?	rrq evos onic team indonesia juara faktor untung hasil kerja keras

4.3.6 N Chars Filter

Yang dimana untuk memfilter semua istilah yang terkandung dalam dokumen input dengan jumlah karakter N kurang dari yang ditentukan.

Tabel 4.7 Perbandingan Teks Sebelum dan Sesudah Dilakukan *N Chars Filter*

<i>N Chars Filter</i>	
Data Sebelum	Data Sesudah
aq dukung rrq semangat tim tanding mreka masuk mpl tim	dukung rrq semangat tim tanding mreka masuk mpl tim

4.3.7 Stop Word Filter

Proses ditahap ini merupakan tambahan *preprocessing* menggunakan Knime untuk menyempurnakan pembuangan *stop word* yang sebelumnya sudah menggunakan gata *framework*. *Stop word filter* yaitu memfilter semua ketentuan dokumen input, yang terkandung dalam daftar *stop word* yang ditentukan dan / atau dalam tabel input kedua dengan menginput *stop word* bahasa indonesia. Apa bila sebuah kata masuk ke dalam daftar *stop word* maka kata tersebut tidak akan diproses.

Tabel 4.8 Perbandingan Teks Sebelum dan Sesudah Dilakukan *Stop Word Filter*

<i>Stop Word Filter</i>	
Data Sebelum	Data Sesudah
pokok malam rrq is the best	pokok malam best

4.3.8 Bag of Words Creator

Pada proses tokenisasi pada node *bag of words creator* di Knime yaitu semua kata yang ada didalam tiap dokumen dikumpulkan dan dihilangkan tanda baca, serta simbol atau karakter khusus lainnya yang bukan termasuk huruf, akan kalimat tersebut akan dipecahkan menjadi perkata.

Tabel 4.9 Perbandingan Teks Sebelum dan Sesudah Dilakukan *Bag of Word* Pada KNIME

<i>Bag Of Words Creator</i>	
Data Sebelum	Data Sesudah
sayang final hambar suporter finalis tidak_bisa tonton langsung area turnamen vivarrq hormat evosroar	sayang[] final[] hambar[] suporter[]

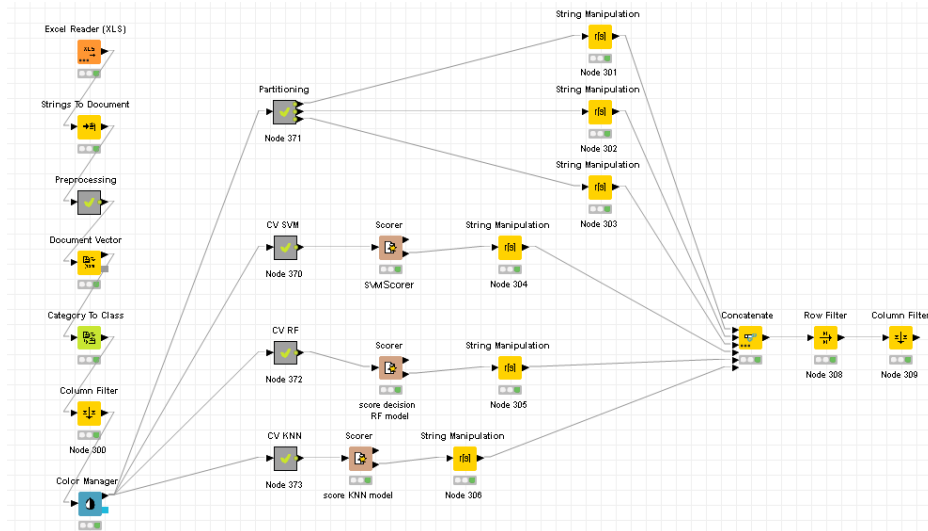
	finalis[]
	tidak_bisa[]
	tonton[]
	langsung[]
	area[]
	turnamen[]
	vivarrq[]
	hormat[]
	evosroar[]

Sumber : Hasil Penelitian (2020)

4.3.9 TF-IDF

Pada proses ini yaitu menghitung frekuensi istilah relatif (TF) dari setiap istilah sesuai dengan setiap dokumen dan menambahkan kolom yang berisi nilai TF. Nilai tersebut dihitung dengan membagi frekuensi absolut dari suatu istilah menurut suatu dokumen dengan jumlah semua syarat dari dokumen itu.

Pada tahap IDF ini menghitung tiga varian frekuensi dokumen terbalik (IDF) untuk setiap istilah sesuai dengan kumpulan dokumen yang diberikan dan menambahkan kolom yang berisi nilai IDF. IDF lancar, normal, dan probabilistik.



Sumber : Hasil Penelitian (2020)

Gambar 4.1 Desain Model Menggunakan KNIME

Berdasarkan gambar 4.1 dapat dilihat desain model penelitian diawali dengan *Excel Reader (XLS)*, *Strings To Document*, *Preprocessing*, *Document Vector*, *Category To Class*, *Column Filter*, dan *Color Manager*.

4.5 Evaluation

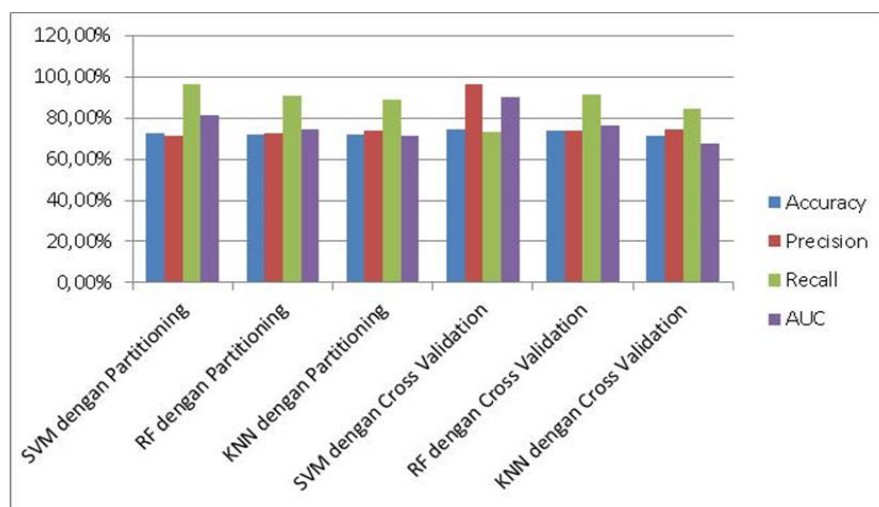
Adapun perbandingan *accuracy*, *precision*, *recall* dan AUC antara algoritma SVM, RF, dan KNN dengan *Partitioning* dan algoritma SVM, RF, dan KNN dengan *Cross Validation* yang telah dilakukan sebagai berikut:

Tabel 4.10 Perbandingan Accuracy, Precision, Recall dan AUC

Validasi	Algoritma	Sentimen	Accuracy	Precision	Recall	AUC
Partitioning	SVM	Positif	76,67%	74,12%	82,89%	0,822
		Negatif		80,0%	70,27%	
	RF	Positif	71,33%	66,84%	86,18%	0,795
		Negatif		79,81%	56,08%	
	KNN	Positif	73,33%	79,03%	64,47%	0,775
		Negatif		69,32%	82,43%	
	SVM	Positif	74,4%	73,68%	77,17%	0,868

Cross Validation		Negatif		75,21%	71,54%	
	RF	Positif	71,8%	67,17%	87,01%	0,844
		Negatif		80,7%	56,1%	
	KNN	Positif	71,2%	77,78%	60,63%	0,81
		Negatif		66,89%	82,11%	

Sumber: Hasil Penelitian (2020)



Sumber : Hasil Penelitian (2020)

Gambar 4.2 Grafik Perbandingan Accuracy, Precision, Recall, Dan AUC

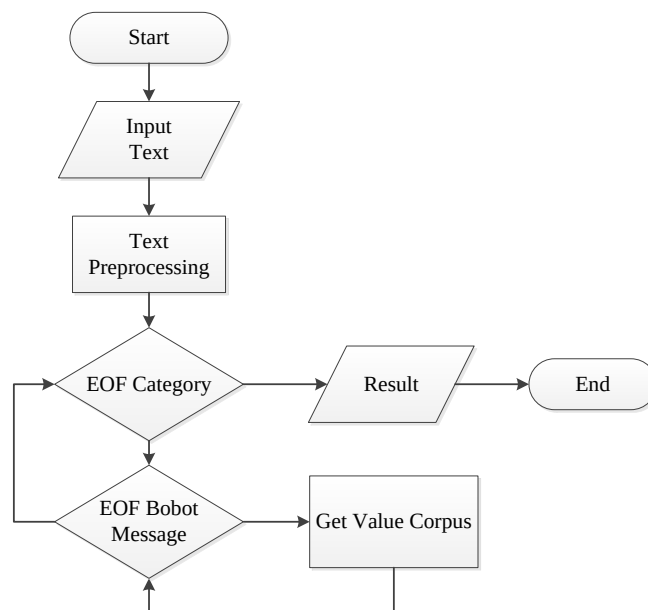
Dapat disimpulkan bahwa metode SVM dengan *cross validation* menghasilkan nilai *accuracy* dan AUC terbaik. Dengan hasil *accuracy* sebesar 76,67% dan AUC sebesar 0,822 sehingga termasuk *good classification*.

Model klasifikasi *text* dari komentar yang digunakan dapat memudahkan untuk mengetahui kategori sentimen positif atau negatif. Berdasarkan data komentar yang diolah menggunakan tools *Knime Analytics Platform*, data komentar akan dipecah secara perkata yang memiliki bobot. Kata-kata tersebut akan digunakan untuk melihat kata-kata yang berhubungan dengan kategori yang sering muncul dan memiliki bobot tertinggi akan dapat digunakan untuk mengetahui komentar tersebut positif atau negatif pada euforia yang terjadi.

4.6 Deployment

Setelah didapat korpus hasil dari data yang sudah dievaluasi menggunakan SVM dengan cross validation maka korpus data tersebut dimasukkan ke dalam database php untuk proses deployment aplikasi analisis sentimen pada *final Mobile Legends Professional League Indonesia Season 5*.

Berikut ini merupakan flowchart dari aplikasi analisis sentimen pada *final Mobile Legends Professional League Indonesia Season 5* yang dibuat:



Gambar 4.3 Flowchart Aplikasi Analisis Sentimen pada final Mobile Legends Professional League Indonesia Season 5

Berdasarkan flowchart pada gambar 4.3 dapat dilihat bahwa dalam aplikasi analisis sentimen pada *final Mobile Legends Professional League Indonesia Season 5*, cara dalam melakukan input analisis text yaitu dengan diinput secara manual yang kemudian akan dilakukan proses *text preprocessing*.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa penerapan *data mining* dapat membantu dalam menganalisis sentimen komentar pengguna facebook pada *Live* pertandingan MPL Indonesia *Season 5* di *facebook*. Algoritma metode SVM merupakan metode terbaik dalam penelitian ini dengan *cross validation* menghasilkan nilai *accuracy* dan AUC terbaik. Dengan hasil *accuracy* sebesar 76,67% dan AUC sebesar 0,822 sehingga termasuk *good classification*.

5.2 Saran

Adapun saran yang diberikan berdasarkan penelitian ini sebagai berikut:

1. Menggunakan algoritma pengklasifikasian lain seperti menggunakan metode *deep learning*.
2. Menggunakan metode pemilihan fitur seperti PSO, N-Gram, SMOTE untuk mengetahui seberapa berpengaruhnya.
3. Penelitian selanjutnya perlu dikembangkan dengan menggunakan data yang lebih banyak

DAFTAR PUSTAKA

- [1] E. Valentina and W. P. Sari, “Studi Komunikasi Verbal dan Non Verbal Game Mobile Legends : Bang Bang,” pp. 300–306, 2018.
- [2] B. Ulum, “Game ‘ mobile legends bang bang ’ di kalangan mahasiswa uin sunan ampel surabaya dalam tinjauan ‘ one dimensional man ’ herbert marcuse,” UNIVERSITAS ISLAM NEGERI SUNAN AMPEL, 2018.
- [3] M. Z. Mustamiin, “Konseling individu dengan sikap kecanduan,” 2019.
- [4] Y. Azzery and I. Krisnadi, “Strategi Pengembangan e-Sport di Indonesia Dengan Menggunakan Metode SWOT,” pp. 2–6.
- [5] G. Nurikhsani, “El Grande Clasico di MPL ID Season 5, RRQ Hoshi Kalahkan EVOS Legends!,” Jakarta, Apr. 2020.
- [6] Y. Yu and X. Wang, “Computers in Human Behavior World Cup 2014 in the Twitter World : A big data analysis of sentiments in U . S . sports fans ’ tweets,” *Comput. Human Behav.*, vol. 48, pp. 392–400, 2015, doi: 10.1016/j.chb.2015.01.075.
- [7] J. B. Arthur A. Raney, *Handbook of Sports and Media*. New York: Lawrence Erlbaum Associates, 2006.
- [8] X. Wang, “Computers in Human Behavior Applying the integrative model of behavioral prediction and attitude functions in the context of social media use while viewing mediated sports,” *Comput. Human Behav.*, vol. 29, no. 4, pp. 1538–1545, 2013, doi: 10.1016/j.chb.2013.01.031.
- [9] X. Wang, “Using attitude functions , self- efficacy , and norms to predict attitudes and intentions to use mobile devices to access social media during sporting event attendance,” 2015, doi: 10.1177/2050157914548932.
- [10] B. R. Naiknaware, S. Kawathekar, and S. N. Deshmukh, “Sentiment Analysis of Indian Government Schemes Using Twitter Datasets,” pp. 70–78, 2017.
- [11] M. Kosinski, S. C. Matz, and S. D. Gosling, “Facebook as a Research Tool for the Social Sciences,” pp. 543–556, 2015.
- [12] M. R. Pahlevi, “ANALISIS SENTIMEN PERTANDINGAN SEPAK BOLA PADA TWEET TWITTER MENGGUNAKAN TEXT MINING,” Universitas Gadjah Mada, 2019.

- [13] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," no. September, 2019.
- [14] W. G. A. B. P. Imam Santoso, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," vol. 1, no. 10, pp. 5–11, 2019.
- [15] B. Bawono and R. Wasono, "PERBANDINGAN METODE RANDOM FOREST DAN NAÏVE BAYES UNTUK KLASIFIKASI DEBITUR BERDASARKAN KUALITAS KREDIT," pp. 343–348, 2019.
- [16] P. Oktaviani, I. Asror, and M. A. Bijaksana, "Analisis Implementasi Sistem OLAP dan Klasifikasi Ketepatan Waktu Lulus dan Undur Diri Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Random forest," vol. 5, no. 2, pp. 3564–3574, 2018.
- [17] S. R. M. Pakpahan, Indriati, and Marji, "Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku," vol. 3, no. 7, pp. 7259–7267, 2019.
- [18] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter , a Social Network or a News Media ?," pp. 591–600, 2010.
- [19] G. Valkanas, A. Saravanou, and D. Gunopulos, "A Faceted Crawler for the Twitter Service," pp. 178–188, 2014.
- [20] J. E. Sembodo, E. B. Setiawan, and A. Baizal, "Data Crawling Otomatis pada Twitter," no. October 2018, pp. 10–16, 2016, doi: 10.21108/INDOSC.2016.111.
- [21] I. Hemalatha, "Preprocessing the Informal Text for efficient," no. July 2012, 2019.
- [22] E. Listiana and M. A. Muslim, "PENERAPAN ADABOOST UNTUK KLASIFIKASI SUPPORT VECTOR MACHINE GUNA MENINGKATKAN AKURASI PADA DIAGNOSA CHRONIC KIDNEY DISEASE," no. 2015, pp. 875–881, 2017.
- [23] W. Gata and Purnomo, "Akurasi Text Mining Menggunakan Algoritma K-

- Nearest Neighbour pada Data Content Berita SMS,” vol. 6, pp. 1–14, 2016.
- [24] W. Gata and H. Basri, “Algorithm Implementations Naïve Bayes , Random Forest . C4 . 5 on Online Gaming for Learning Achievement Predictions,” vol. 258, no. Icream 2018, 2019.
- [25] B. H. Mahendra *et al.*, “Kategorisasi Berita Multi-Label Berbahasa Indonesia Menggunakan Algoritma Random,” vol. 6, no. 2, pp. 9030–9041, 2019.