

Analisis Ulasan *Indie Video Game* Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis *Latent Dirichlet Allocation*

Mochamad Yudha Febrianta¹, Sri Widiyanesti², Syahrul Robbiansyah Ramadhan³

^{1,2,3}Manajemen Bisnis Telekomunikasi dan Informatika, Fakultas Ekonomi dan Bisnis, Universitas Telkom

¹yudhafebrianta@telkomuniversity.ac.id, ²widiyanesti@telkomuniversity.ac.id,

³robbystr@student.telkomuniversity.ac.id

Abstrak

Video game merupakan produk Ekonomi Kreatif yang berkembang pesat dan mempengaruhi perekonomian dunia. Di antara para pelakunya, pengembang *indie video game* memiliki keterbatasan sumber daya dan bergantung pada jasa distribusi digital untuk menjual produknya. Steam merupakan platform distribusi digital *video game* dengan fitur ulasan yang dapat dijadikan acuan pengembangan *video game*. Ulasan produk *online* cenderung berjumlah banyak dan beragam sehingga menimbulkan tantangan bagi para pengembang *indie video game* untuk menganalisisnya. Penelitian ini bertujuan untuk melakukan analisis berbasis *machine learning* terhadap ulasan produk *indie video game* di Steam secara otomatis. Metode yang digunakan adalah analisis sentimen dengan algoritma klasifikasi Naïve Bayes dan pemodelan topik berbasis *LDA* menggunakan perangkat lunak Rapidminer dan RStudio. Hasil penelitian menunjukkan sentimen positif dominan sebesar 69.8% dengan akurasi algoritma 75.45%. Penelitian ini juga menunjukkan bahwa “*story*”, “*character*”, “*music*”, dan “*art*” termasuk istilah yang sering muncul di antara topik-topik dominan.

Kata kunci: *video game*; analisis sentimen; model topik; *machine learning*; ulasan produk

Analysis of Local Indie Video Game Reviews on Steam Using Sentiment Analysis and Latent Dirichlet Allocation-based Topic Modeling

Abstract

Video game is a product of Creative Economy in rapid development, affecting the global economy. Within the industry, indie video game developers have limited resources and are dependent on digital video game distribution services to sell their products. Steam is a such platform that features online reviews, which can be used as references to develop better video games. Online product reviews commonly come in big number and variety, imposing challenges for indie developers to analyze them. This research intends to analyze indie video game reviews on Steam automatically using machine learning. The method used is sentiment analysis with Naïve Bayes Classifier algorithm and LDA-based topic modeling using Rapidminer and RStudio softwares. The results show that positive sentiment is dominant by 69.8% with 75.45% of the algorithm’s accuracy, and identified dominant topics includes most used words such as “*story*”, “*character*”, “*music*”, and “*art*”.

Keywords: *video game*; sentiment analysis; topic model; machine learning; product review

Pendahuluan

Pada tahun 2009 pemerintah Indonesia menerbitkan Instruksi Presiden Nomor 6 Tahun 2009 perihal pengembangan ekonomi kreatif yang menjadi salah satu pionir landasan hukum ekonomi kreatif di medan perekonomian Indonesia. Landasan hukum tersebut mendorong terbitnya Peraturan Presiden Republik Indonesia Nomor 6 Tahun 2015 sebagai landasan pembentukan Badan Ekonomi Kreatif Indonesia (Badan Ekonomi Kreatif, 2018). Lalu pada tahun 2019 institusi tersebut dilebur dengan Kementerian Pariwisata Republik Indonesia menjadi Kementerian Pariwisata dan Ekonomi Kreatif Republik Indonesia. Dengan itu, Indonesia telah menjadi negara pertama dan satu-satunya hingga saat ini yang memiliki institusi pemerintahan dengan nomenklatur “Ekonomi Kreatif” (Antariksa, 2012; United Nations, 2018s).

John Howkins merupakan tokoh pertama yang mengemukakan istilah “ekonomi kreatif” dan menjelaskannya sebagai hasil dari kegiatan yang menggunakan imajinasi serta mengeksploitasi nilai ekonomisnya. Produk kreatif adalah hal yang ditawarkan dari ekonomi kreatif. Produk tersebut berupa kekayaan intelektual yang berasal dari kreativitas manusia dalam bentuk barang atau jasa yang dapat diakui nilai ekonomisnya. Secara singkat, ekonomi kreatif adalah cara orang menghasilkan uang dari ide (Howkins, 2007).

Perkembangan industri kreatif pertama kali dilaporkan oleh firma Ernst and Young pada tahun 2015 dan telah menunjukkan bahwa pada tahun tersebut *Cultural and Creative Industries* telah menghasilkan US\$2.250 miliar yang merupakan 3% dari produk domestik bruto (PDB) total seluruh negara di dunia, serta menyediakan 29.5 juta lapangan kerja baru (Ernst and Young, 2015). Sementara Badan Ekonomi Kreatif Indonesia (2018) melaporkan bahwa pada tahun 2018 di Indonesia sektor ekonomi kreatif telah menyumbang Rp1.105 triliun yang merupakan 7.4% PDB nasional pada tahun itu (Badan Pusat Statistik Republik Indonesia, 2019).

Ekonomi kreatif di Indonesia pada tahun 2018 terbagi dalam 16 subsektor, di antaranya terdapat subsektor “Aplikasi dan *Game Developer*” yang telah menunjukkan potensi bisnis besar terhadap perekonomian dunia. Pada tahun 2016 subsektor tersebut telah menghasilkan nilai PDB nasional sebesar Rp17.42,8 miliar dan memperingkatkan posisi kelima di antara subsektor ekonomi kreatif lainnya berdasarkan laju pertumbuhan

PDB-nya. Lalu pada tahun 2018 subsektor tersebut telah memperoleh valuasi pasar *video game* global sebesar US\$137,9 juta.

Perkembangan subsektor Aplikasi dan *Game Developer* di Indonesia dimungkinkan oleh kemunculannya perusahaan-perusahaan *startup* di bidang *video game* seperti Agate Studio, Altermyth Studio, Toge Production, Tinker Game, Touch Ten Game, dan lain-lain. Tidak hanya itu, peningkatan penggunaan perangkat *mobile* yang telah meningkatkan penggunaan internet secara nasional, serta adanya pendanaan dari investor dalam jumlah besar terhadap perusahaan-perusahaan *startup* bidang *video game* di Indonesia juga menjadi alasan pesatnya perkembangan subsektor Aplikasi dan *Game Developer* di Indonesia (Badan Ekonomi Kreatif, 2018).

Sebelum secara resmi diakui sebagai produk ekonomi kreatif, industri *video game* sudah menunjukkan potensinya dalam bisnis sebagai alternatif dari permainan arkade seperti mesin slot, *pinball*, bola sodok, *air hockey*, dan lainnya. Alternatif yang disebut *video game* ini diawali dengan penciptaan Computer Space yang merupakan mesin permainan arkade pertama, serta penciptaan Magnavox Odyssey sebagai konsol *video game* pertama pada akhir 1960-an yang telah menjadi asal mula *video game* modern (Smith, 2019).

Industri *video game* pada tahun 2020 diperkirakan memiliki 2.7 miliar konsumen di dunia dan menghasilkan pendapatan global senilai US\$159,3 miliar. Melalui pendapatan tersebut, industri *video game* memiliki tingkat pertumbuhan tahun ke tahun sebesar +9.3% dengan *compound annual growth rate (CAGR)* sebesar +7.7.% (Newzoo, 2020). Sementara itu, khususnya di Indonesia pendapatan industri *video game* mencapai nilai US\$1.084 juta dan menempatkan Indonesia pada urutan ke-17 berdasarkan peringkat pendapatan pasar *video game* secara global pada tahun 2019. Peringkat pendapatan pasar *video game* global tersebut diungguli oleh Cina dengan pendapatan senilai US\$34.400 juta dan Amerika Serikat dengan nilai US\$31.535 juta (Knoema, 2019; Newzoo, 2019).

Industri *video game* dapat dibagi dalam 3 segmen utama berdasarkan saluran produknya, yaitu *mobile*, *Personal Computer (PC)*, dan *console* (Newzoo, 2020). Berdasarkan tipe produksinya, *video game* dapat dikategorikan sebagai produksi “AAA” (dibaca sebagai “*triple-A*”) atau “*indie*” (merupakan singkatan dari kata *independent*) (Mathews & Wearn, 2016). Wong dalam Mathews (2016) menjelaskan bahwa AAA

adalah produksi *video game* yang memiliki anggaran bernilai jutaan Dollar Amerika Serikat dan dihasilkan oleh *video game developer* (pengembang *video game*) ternama yang didukung oleh *publisher* (penerbit) perihal pemasaran dan promosinya yang berskala besar. Sementara *indie video games* menurut Saltman dalam Mathews (2016) adalah proyek *video game* yang pada umumnya bersifat eksploratif dan eksperimental dengan anggaran yang kecil oleh sekelompok kecil individu.

Pengembangan *indie video game* sering mengutamakan desain sebagai estetika, yang mana lebih baik dikelola dalam skala kecil oleh para pengembang *indie* (Zackariasson & Wilson, 2012). Berbeda dengan para pengembang *indie*, para pengembang AAA memiliki budaya dan kepentingan untuk terus mengembangkan teknologi canggih terhadap produksi *video game* demi memperoleh “realisme” dalam permainannya. AAA *video game* juga didominasi oleh genre yang lebih umum, menantang, realistis, rumit, dan menarik, selain memaparkan inovasi teknologi. Namun, untuk mencapai kepentingan tersebut memerlukan tim pengembang yang besar dengan pengorbanan yang juga besar dalam segi tenaga dan waktu (Oakley & O’Connor, 2015).

Dengan semakin berkembangnya industri *video game*, pengembangan dan produksi AAA *video game* pun mengalami peningkatan dalam segi risiko dan biayanya, sehingga mendesak para pengembang *video game* untuk mengubah produksi *video game* secara fisik (melalui CD-ROM atau disket) menjadi digital dan *downloadable* (dapat diunduh secara *online*). Dalam hal ini segmen *PC video game* menjadi pionir dalam mengadopsi jenis distribusi baru yang disebut dengan distribusi digital karena kemampuan komputer pribadi atau *PC* yang mampu terhubung dengan internet dengan mudah, sehingga memungkinkan permainan bersama (yang disebut *multiplayer*) dan akses terhadap komunitas *video game online* (Toivonen & Sotamaa, 2010).

Dalam hal pendistribusian produk *video game* terhadap konsumennya, penerbit memiliki peran yang sangat penting karena para pengembang *video game* sering kali memiliki keterbatasan dalam kemampuan untuk mempromosikan *video game*-nya. *Video game* berstatus AAA memiliki anggaran sekitar US\$15 juta sampai US\$20 juta untuk pengembangan, promosi, hingga penjualannya (Zackariasson & Wilson, 2012). Sementara para pengembang *indie* dengan semakin umumnya saluran distribusi digital yang tidak bergantung pada toko ritel, memanfaatkan internet secara penuh untuk menjual dan mendistribusikan *video game* yang berukuran lebih kecil dengan biaya yang

lebih sedikit kepada para konsumennya (Oakley & O'Connor, 2015).

Peran jasa distribusi digital sebagai alternatif saluran distribusi dan promosi menjadi sesuatu yang tidak dapat terpisahkan dari para pengembang *indie video game*. Karena mereka menerbitkan dan mempublikasikan produknya secara mandiri, pemanfaatan distribusi digital menjadi salah satu opsi yang memungkinkan model bisnis dengan keuntungan lebih besar bagi para pengembang *video game*. Melalui penerapan distribusi digital, para pengembang *video game* juga dapat mengendalikan faktor-faktor distribusi secara langsung (Garda & Grabarczyk, 2016), bahkan khususnya bagi para pengembang *indie*, platform distribusi digital memberikan akses terhadap pasar *video game* yang lebih luas meskipun keterbatasan finansial yang mereka miliki (Martin & Deuze, 2009).

Pada segmen *PC video game* terdapat beberapa platform distribusi *video game* digital yang populer, seperti Steam, GOG.com, itch.io, GameFlay, Green Man Gaming, Humble Bundle, Epic Games Store, Origin, Uplay, dan Xbox For PC (Minor, 2020). Namun di antara sejumlah platform tersebut, Steam yang dikelola oleh Valve Corporation merupakan platform distribusi digital yang paling unggul berdasarkan jumlah pengguna dan katalognya (Wilson, 2020).

Hanya dengan biaya penerbitan sebesar US\$100 di Steam, para pengembang *indie video game* mampu menjual produknya pada pasar *online PC video game* terbesar yang memiliki 20 juta pengunjung aktif setiap harinya, meskipun terdapat itch.io yang hadir sebagai platform khusus *indie video game*. Namun, keunggulan mendistribusikan *video game* di Steam oleh para pengembang *indie* harus dikorbankan dengan berhadapan langsung dengan perusahaan-perusahaan besar beserta ribuan produk *indie video game* lain yang dapat mempengaruhi visibilitas produk yang dipromosikan (Henges, 2020).

Persaingan industri *video game* yang sangat ketat mendesak para pengembang *indie video game* untuk menerapkan strategi pemasaran dan penjualan yang tepat demi memperoleh keuntungan dan kepuasan konsumennya (Denby, 2019), terutama karena konsumen *video game*, disebut sebagai *gamer*, merupakan jenis konsumen yang sangat sulit untuk merasa puas dan menjadikan kualitas *video game* sangat penting untuk diperhatikan (Chambers et al., 2005). Oleh karena itu, Lin et al. (2019) menjelaskan salah satu cara untuk mencapai kepuasan konsumen *video game* adalah dengan menganalisis hasil ulasan produk *video game* yang bersangkutan karena mengandung informasi

penting dan bermanfaat terhadap perbaikan dan pengembangan *video game* yang diulas maupun yang masih dalam pengembangan. Dalam hal ini, Steam menyediakan fitur *review* (ulasan) terhadap seluruh *video game* yang dijual di platformnya, yang dapat ditulis oleh siapapun yang telah membeli dan memainkan produk terkait dan ditampilkan secara terbuka untuk dibaca oleh kalangan umum.

Ulasan daring juga dapat dimanfaatkan sebagai sumber informasi oleh calon pelanggan dalam mempertimbangkan keputusan pembelian produk yang telah diulas tanpa dipengaruhi oleh informasi sepihak dari produsennya (Kang et al., 2017). Terkait hal ini, Ramadhan (2019) telah melakukan penelitian yang menunjukkan adanya pengaruh positif secara langsung dan signifikan antara ulasan konsumen di Steam dengan keputusan pembelian konsumen.

Lin et al. (2019) menjelaskan bahwa satu *video game* di Steam secara rata-rata menerima 2 ulasan setiap hari. Meskipun jumlah tersebut dapat dianggap relatif rendah, adanya kemungkinan bahwa para pengembang *video game*, khususnya pengembang *indie*, tidak mampu menganalisis ulasan yang mereka terima setiap hari secara menyeluruh, karena mereka memiliki waktu dan tenaga yang terbatas untuk mengelola *video game* yang mereka telah produksi (bahkan yang masih dalam pengembangan), sehingga jumlah ulasan yang harus diperiksa dapat menumpuk dengan cepat.

Zuo (2018) menjelaskan bahwa *video game* sebagai produk digital sangat berbeda dengan produk tradisional fisik. Karena hanya dapat dibeli secara daring, maka para konsumen tidak dapat mencoba produk sebelum membelinya. Ulasan yang tampil secara daring memiliki peran penting dalam proses pengambilan keputusan konsumen, sehingga prediksi akurat terkait sentimen yang terkandung dalam ulasan tersebut dapat meningkatkan potensi perolehan keuntungan.

Ulasan *video game* di Steam memiliki karakteristik jumlah yang banyak, terbentuk secara cepat dan memiliki keberagaman yang luas. Karakteristik-karakteristik tersebut sesuai dengan tiga dimensi *big data*, yaitu *volume* (jumlah data yang besar), *velocity* (kecepatan pembentukan yang tinggi), dan *variety* (keragaman yang luas (Kang et al., 2017). Oleh karena itu, pengetahuan atau wawasan yang penting dan bermanfaat sebagai dasar dari pengambilan keputusan dapat diperoleh dari ulasan *video game* di Steam melalui teknik analisa *big data* seperti *text analytics* berupa analisis sentimen (Vashisht & Gupta, 2016) yang merupakan metode untuk memahami sentimen terhadap

suatu entitas (Fang & Zhan, 2015), dilengkapi dengan pemodelan topik untuk memperoleh pengetahuan makroskopis melalui identifikasi topik (Ignatow & Mihalcea, 2018).

Analisis sentimen merupakan bagian tugas dari *opinion mining* (penggalian opini). *Opinion mining* didefinisikan sebagai tugas untuk mengidentifikasi opini yang berupa ekspresi keadaan pribadi seperti emosi, sentimen, evaluasi, kepercayaan, dan spekulasi dalam bahasa natural. Opini-opini tersebut memiliki atribut, yang mencakup siapa yang mengekspresikan opini, jenis sikap yang diekspresikan, mengenai siapa atau apa opini yang diekspresikan, serta sentimen terhadap opini tersebut, dan lainnya. Berdasarkan hal tersebut, analisis sentimen dapat dijelaskan sebagai bagian tugas dari *opinion mining* untuk mengklasifikasikan opini dalam bentuk positif, negatif, atau netral. Penentuan sentimen antara positif dan negatif juga merupakan bentuk dari *text classification* (klasifikasi teks) yang merupakan kegiatan untuk menggolongkan teks terhadap satu atau lebih kategori yang telah ditentukan, berbeda dengan *text clustering* (pengelompokkan teks) yang merupakan proses mengelompokkan teks berdasarkan kelompok yang didasari oleh kemiripannya (Ignatow & Mihalcea, 2018).

Menurut Joshi dan Itkat (2014) ada 3 kategori algoritma klasifikasi yang dapat diterapkan, yaitu:

- i. *Supervised Learning Algorithms*, yaitu algoritma yang membutuhkan pengawasan melalui persiapan data pembelajaran. Termasuk dalam kategori ini adalah algoritma klasifikasi *Naïve Bayes*, *Maximum Entropy*, *Decision Tree*, dan *Support Vector Machine*.
- ii. *Unsupervised Learning Algorithms*, yaitu algoritma yang tidak membutuhkan pengawasan. Disebut juga sebagai *lexicon based techniques*, algoritma ini melibatkan pembelajaran terhadap pola pada *input* ketika tidak ada nilai *output* spesifik yang disediakan, yang artinya algoritma pembelajaran hanya menerima *examples set* yang tidak dilabeli. Termasuk dalam kategori ini adalah *K-Means Clustering*, dan *Cluster Analysis*.
- iii. *Semi-Supervised Learning Algorithms*, yaitu algoritma hibrida yang menggabungkan pendekatan *supervised learning* dan *lexicon based*.

Sementara, untuk melengkapi sentimen analisis dalam upaya menganalisis teks, dapat dilakukan pemodelan topik, yaitu melibatkan prosedur otomatis untuk

mengkodifikasi kumpulan teks dalam kategori yang bermakna yang merepresentasikan topik utama yang dibahas dalam teks. *Topic model* mengasumsi bahwa makna saling memiliki hubungan dan makna-makna yang berkaitan dengan suatu topik dalam pembicaraan dapat dipahami sebagai satu kelompok kata-kata. *Topic model* memperlakukan teks sebagai *bag of words* (kantong kata-kata) yang menangkap kemunculan bersama antar kata tanpa mementingkan sintaks, narasi, ataupun lokasinya dalam teks. Sebuah topik dapat dipahami sebagai kelompok kata-kata yang cenderung muncul dalam diskusi, sehingga dapat saling bermunculan lebih sering dari seharusnya ketika topik sedang dibahas. Pemodelan topik juga merupakan bentuk dari *probabilistic modeling* (pemodelan probabilistik), contohnya *Latent Dirichlet Allocation (LDA)* yang didasari oleh pemahaman bahwa setiap teks dalam kumpulan teks menyerupai suatu *bag of words* yang dihasilkan terhadap campuran topik yang diharapkan oleh penulisnya untuk didiskusikan. Selain *LDA*, terdapat *probabilistic model* berupa *Latent Semantic Analysis (LSA)* yang didasari oleh kemiripan makna dari kata-kata yang muncul dalam teks dan menunjukkan kata-kata dan teks menggunakan *vector space modeling* yang mengumpulkan data tekstual menjadi *term-by-document matrix* (matriks istilah per dokumen), dengan menunjukkan frekuensi berbobot dari setiap istilah untuk merepresentasikan dokumen-dokumen dalam kumpulan istilah (Ignatow & Mihalcea, 2018).

Ignatow dan Mihalcea (2018) menjelaskan bahwa pada *Latent Dirichlet Allocation (LDA)* setiap topik merupakan hasil distribusi terhadap seluruh kata-kata yang terobservasi dalam teks sehingga kata-kata yang berkaitan kuat dengan topik-topik dominan dalam teks memiliki peluang lebih tinggi untuk dicantumkan. Di antara sejumlah hasil prosedur *LDA* terdapat distribusi set topik per kata yang mengasosiasikan suatu probabilitas dengan setiap pasangan topik dengan kata dan suatu distribusi set teks per topik yang serupa dan menjelaskan probabilitas pemilihan topik khusus untuk setiap teks yang spesifik. *LDA* memiliki keunggulan akurasi yang lebih tinggi dalam penggunaannya sebagai model probabilistik untuk pemodelan topik.

Di antara *indie video game* asal Indonesia, Asosiasi Game Indonesia (2020) mencatat setidaknya 24 dari 40 judul *video game* PC yang telah dirilis melalui platform Steam pada tahun 2020. Berdasarkan latar belakang mengenai perkembangan industri *video game* secara global dan nasional yang telah dipaparkan, peneliti melihat adanya

peluang yang dapat diraih oleh para pengembang *indie video game* lokal untuk bersaing hingga mengungguli pasar *indie video game* global. Namun, para pengembang *indie video game* lokal, layaknya para pengembang *indie* di seluruh dunia pada umumnya memiliki keterbatasan finansial maupun sumber daya dalam menciptakan produk *video game* hingga menjual dan memasarkannya. Oleh karena itu, penelitian ini bertujuan untuk menyarankan metode analisis sentimen otomatis dengan algoritma klasifikasi *Naïve Bayes* dan pemodelan topik berbasis *Latent Dirichlet Allocation* terhadap ulasan *indie video game* di Steam guna mengidentifikasi sentimen dan topik dominan terkait ulasan *indie video game* lokal yang dapat dimanfaatkan dalam pengembangan produk *video game*. Hasil penemuan dari penelitian ini diharapkan dapat membantu meningkatkan keuntungan bisnis bagi para pengembang *indie video game*.

Pembahasan

Pengambilan dan Persiapan Data

Penelitian ini menggunakan data berupa ulasan dari 24 judul *indie video game* lokal di Steam. Data tersebut diperoleh menggunakan teknik *web scraping* menggunakan perangkat lunak Web Scraper. Pengambilan data dilakukan pada tanggal 2 Februari 2021 dan menghasilkan 3621 data ulasan. Jumlah perolehan data ulasan dari setiap judul *indie video game* terpilih adalah sebagai berikut:

Tabel 1 Jumlah Perolehan Data Ulasan.

No.	Judul Video Game	Jumlah Data Ulasan
1.	A Day Without Me	33
2.	Aquaculture Land: Fish Farming Simulation	49
3.	Azure Saga: Pathfinder	56
4.	Ciel Fledge	37
5.	Coffee Talk	1315
6.	DreadOut 2	519
7.	Eurgava – Tomb of Senza	1
8.	Ghost Parade	21
9.	Hell's Pharma	8
10.	Insurgence – Chains of Renegade	32
11.	Insurgence – Second Assault	14
12.	Legrand Legacy: Tale of the Fatebounds	248

No.	Judul <i>Video Game</i>	Jumlah Data Ulasan
13.	Lobster Empire	26
14.	My Lovely Daughter	265
15.	Necronator: Dead Wrong	212
16.	Nusakana	50
17.	Orbiz	3
18.	Pulang: Insanity	61
19.	Rage in Peace	52
20.	Retrograde Arena	62
21.	Rising Hell	63
22.	Sheh and the Light Bearer	79
23.	Ultra Space Battle Brawl	34
24.	When The Past Was Around	381
	TOTAL	3621

Penelitian ini hanya menggunakan atribut *review* dan *recommendation* dari data ulasan yang telah diperoleh. Seluruh data ulasan dikumpulkan dalam satu *spreadsheet* menggunakan Google Sheet dan dipersiapkan melalui tahapan-tahapan berikut sebelum dilanjutkan pada tahapan *pre-processing*:

i. *Data Labeling* (Pelabelan Data)

Pelabelan data dilakukan secara otomatis berdasarkan rekomendasi pada atribut *recommendation* yang tertera pada data ulasan. *Recommendation* dengan label “Recommended” dikonversi menjadi label “positive”, sementara label “Not Recommended” dikonversi menjadi label “negative”. Proses ini menghasilkan 3187 data dengan label “positive” dan 434 data dengan label “negative”.

ii. *Data Selection* (Pemilihan Data)

Data yang akan diolah pada penelitian ini hanya ulasan berbahasa Inggris dengan alasan bahwa ulasan berbahasa Inggris memiliki frekuensi kemunculan yang jauh lebih tinggi daripada bahasa lain, serta teks berbahasa Inggris lebih kompatibel dengan algoritma *pre-processing* dan *machine learning* yang digunakan dan tersedia pada waktu penelitian ini berlangsung. Dengan ini, bahasa setiap ulasan diidentifikasi menggunakan formula “DETECTLANGUAGE” pada Google Sheets secara otomatis. Melalui proses tersebut, ulasan yang teridentifikasi tidak menggunakan bahasa Inggris dihapus sekaligus, menyisakan sejumlah 3497 data

ulasan berbahasa Inggris.

iii. *Data Balancing* (Penyeimbangan Data)

Penyeimbangan data perlu dilakukan karena data yang tidak seimbang mampu mempengaruhi akurasi klasifikasi *machine learning* karena bias, *overfitting*, atau *underfitting* (Khomsah, 2020). Pada penelitian ini penyeimbangan data dilakukan dengan cara *undersampling*, yaitu mengurangi jumlah sampel data dengan atribut “sentiment” yang berjumlah lebih banyak.

Terbukti bahwa pada sejumlah data yang akan digunakan terdapat 3072 data dengan label “positive” pada atribut “sentiment”, sementara data dengan label “negative” hanya berjumlah 425 data. Karena data dengan label “positive” berjumlah 7.2 kali lipat lebih banyak daripada data dengan label “negative” maka *undersampling* diterapkan untuk menyeimbangkan kedua label tersebut, dengan cara memangkas sekaligus secara acak sejumlah data dengan label “negative” agar seimbang dengan jumlah data dengan label “positive”. Proses ini menyisakan sejumlah 425 data untuk label “positive” maupun label “negative”. Kemudian 70% data dari setiap label atribut “sentiment” diambil secara acak untuk menentukan data *training set* yang akan digunakan sebagai data latih atau pembelajaran *machine learning* yang akan diterapkan terhadap data *testing set*. Dari keseluruhan data, pada proses ini telah diperoleh 298 data dengan label “positive” dan 298 data dengan label “negative” sebagai data *training set*, serta 2774 data dengan label “positive” dan 127 data dengan label “negative” sebagai data *testing set*.

***Data Pre-Processing* (Prapengolahan Data)**

Dataset *training* dan *testing* harus dibersihkan dan dirapikan terlebih dahulu demi mengoptimalkan efektivitas data dan akurasi proses pengolahan data melalui tahapan prapengolahan data. Google Sheet dan RapidMiner digunakan sebagai alat untuk melakukan prapengolahan data secara otomatis melalui sejumlah tahapan berupa:

i. *Trim Whitespace*

Menghapus bagian teks yang tampil seperti kosong berupa *tabs*, *line feed/breaks*, *carriage return*, *form feed*, dan *vertical tab* menggunakan fitur “*trim whitespace*”.

- ii. *Remove URL*
Menghapus teks berupa *uniform resource locator (URL)*, yaitu teks berupa tautan menuju halaman *website*.
- iii. *Correct Misspelled Words*
Memperbaiki kata-kata yang mengandung kesalahan ketik.
- iv. *Tokenize Regular Expression*
Memangkas teks berdasarkan *regular expressions* berupa “https?://[-a-zA-Z0-9+&@#/%?~_!:,;]*[-a-zA-Z0-9+&@#/%?~_].”
- v. *Replace Tokens*
Menggantikan suatu kata dengan kata lain yang memiliki makna sama. Peneliti melakukan proses ini secara manual dengan menggantikan sejumlah kata, simbol, atau ekspresi dengan makna yang sama atau memiliki makna khusus terhadap konteks ulasan *video game* seperti “*indie*” menjadi “*independent*”, “*uwu*” menjadi “*cute*”, “*devs*” menjadi “*developer*” dan seterusnya
- vi. *Case Folding / Lowercasing*
Mengubah semua huruf menjadi lowercase (huruf kecil).
- vii. *Tokenize*
Memecah teks menjadi kata-kata individu, serta mengidentifikasi kata-kata dengan memisahkan tanda baca.
- viii. *Filter Tokens by Length*
Menerapkan filter terhadap teks berdasarkan jumlah huruf pada setiap kata. Penelitian ini menerapkan filter untuk hanya menerima kata-kata dengan kriteria jumlah huruf lebih besar atau sama dengan 2 dan lebih kecil atau sama dengan 20.
- ix. *Stopwords Removal*
Penghapusan *stopwords* yang merupakan daftar kata-kata yang mencakup pronom, preposisi, penentu, dan lainnya. Dalam kasus ini peneliti menggunakan daftar *stopwords* yang disediakan oleh NLTK ditambah dengan sejumlah kata-kata yang peneliti anggap tidak relevan berdasarkan frekuensi kemunculan dan jumlah karakter melalui hasil *word cloud* terhadap dataset yang belum diproses untuk mengidentifikasi seluruh kata yang terkandung pada keseluruhan dataset.

Proses Analisis Sentimen

Metode ini diterapkan terhadap data yang telah terkumpul dan melewati tahap *pre-processing*. Penelitian ini menggunakan klasifikasi *Naïve Bayes* yang merupakan algoritma klasifikasi yang didasari oleh teorema Bayes. Teorema tersebut menyatakan kondisi probabilitas suatu kejadian terhadap suatu kejadian lain. Model ini melibatkan penyederhanaan asumsi kondisional independen, di mana berdasarkan kelas positif atau negatif, kata-kata yang digunakan secara kondisional tidak bergantung antar sesamanya. Asumsi ini tidak mempengaruhi akurasi dalam klasifikasi teks secara signifikan, tetapi mampu menerapkan algoritma klasifikasi yang cepat terhadap permasalahan (Joshi & Itkat, 2014:5422).

Klasifikasi *Naïve Bayes* digunakan sebagai algoritma klasifikasi *machine learning* pada penelitian ini dengan alasan bahwa algoritma tersebut merupakan model klasifikasi probabilistik dengan kemampuan latih paling sederhana, sangat efektif digunakan pada beragam situasi (Potts, 2021) serta lebih akurat dan cepat ketika diterapkan pada *big data* (Glen, 2019).

Berdasarkan hasil persiapan data, penelitian ini menggunakan sejumlah 596 data *training* dan 2901 data *testing* yang diproses melalui RapidMiner dengan susunan operator sebagai berikut:

i. *Input Training dan Input Testing Data (Read CSV)*

Operator yang digunakan untuk membaca dataset *training* dan *testing* dalam format file “.csv” yang akan digunakan dalam proses klasifikasi. Karena sejumlah teks ulasan yang digunakan pada penelitian ini mengandung tanda baca koma (,), maka peneliti menggunakan format file “.tsv” (yang memisahkan kolom data menggunakan tab) sebagai alternatif demi mencegah terjadinya error.

ii. *Set Role*

Operator yang digunakan untuk mengubah peran dari satu atribut atau lebih pada data *training*. Dalam kasus ini atribut “sentiment” pada set data ditargetkan dengan peran “label”.

iii. *Nominal to Text*

Operator yang digunakan untuk mengubah tipe atribut nominal terpilih menjadi teks dan memetakan seluruh nilai dari atribut-atribut tersebut terhadap *string value* yang sesuai.

iv. *Process Documents from Data*

Operator yang digunakan untuk melakukan sejumlah tahapan prapengolahan data yang telah dijelaskan sebelumnya.

v. *Cross Validation*

Operator yang digunakan untuk melakukan *cross validation* demi menghasilkan estimasi performa statistik model *machine learning* yang meliputi operator *Naïve Bayes* untuk membentuk model klasifikasi *Naïve Bayes*, operator *apply model* untuk menerapkan model pada *ExampleSet*, dan operator *performance (classification)* untuk mengevaluasi performa statistik kerja klasifikasi.

Klasifikasi algoritma *Naïve Bayes* yang digunakan merupakan penerapan dari teorema Bayes sebagai pernyataan dari teori probabilitas yang memungkinkan perhitungan probabilitas kondisional tertentu. Probabilitas kondisional merupakan probabilitas yang menggambarkan pengaruh salah satu kejadian terhadap probabilitas kejadian lain. Dalam teorema Bayes terdapat *prior probability* dari sebuah hipotesis atau kejadian sebagai probabilitas asli yang diperoleh sebelum informasi tambahan diperoleh. Adapun *posterior probability* yang merupakan hasil revisi probabilitas dari hipotesis menggunakan informasi atau bukti tambahan yang telah diperoleh (Sharma et al. 2015:705). Teorema Bayes dirumuskan sebagai berikut:

$$P(A|B) = P(B|A)P(A)/P(B)$$

Dengan keterangan $P(A)$ sebagai *prior probability* dari A, $P(B)$ sebagai *prior probability* dari B, $P(A|B)$ sebagai *posterior probability* dari A dengan diketahui B, dan $P(B|A)$ sebagai *posterior probability* dari B dengan diketahui A.

Performa Klasifikasi Teks

Hasil performa klasifikasi teks menggunakan *machine learning* ditunjukkan melalui tingkat *accuracy* (akurasi), *class precision* (presisi kelas), *class recall* (penarikan kelas), *F1 Measure*, dan *Cohen's Kappa* terhadap data yang dapat diukur dalam nilai *True Positive* (Positif Asli), *True Negative* (Negatif Asli), *False/Predictive Positive* (Positif Palsu), dan *False/Predictive Negative* (Negatif Palsu). Angka performa klasifikasi teks yang dihasilkan melalui RapidMiner ditampilkan dalam bentuk *confusion matrix* sebagai berikut:

Tabel 2 *Confusion Matrix* Performa Klasifikasi *Naïve Bayes*

[accuracy: 75.45% +/- 6.84% (micro average: 75.45%)] [kappa: 0.454 +/- 0.156 (micro average: 0.455)]	True Positive	True Negative	Class Precision
Predictive Positive	118	31	79.19%
Predictive Negative	23	48	67.61%
Class Recall	83.69%	60.76%	

Akurasi performa yang diperoleh oleh klasifikasi *machine learning* memperoleh nilai 75.45%, artinya algoritma klasifikasi *Naïve Bayes* terhadap dataset yang digunakan memiliki kemampuan ketepatan klasifikasi yang baik, karena berada di atas peluang 50%. Sementara Kappa memiliki nilai 0.454, yang dapat diartikan berdasarkan tabel interpretasi nilai Kappa oleh Landis & Koch (1977:165) bahwa kekuatan kesepakatan algoritma *Naïve Bayes* dalam mengklasifikasikan label sentimen positif dan label sentimen negatif terhadap dataset yang digunakan adalah *moderate* (sedang), karena memiliki nilai di antara 0.41 dan 0.60. Berdasarkan interpretasi tersebut, algoritma yang digunakan dalam penelitian ini memiliki akurasi 45% dalam peluang acak.

Hasil performa hasil performa berdasarkan *class recall true positive* bernilai 83.69%, *class recall true negative* bernilai bernilai 60.76%, *precision class predictive positive* bernilai 79.19%, dan *precision class predictive negative* bernilai 67.61% didasari oleh nilai jumlah *True Negative/Positive* dan *Predictive Negative/Positive* berikut:

- i. *True Positive* merupakan jumlah data dengan nilai sentimen positif yang telah diklasifikasikan secara tepat oleh *machine learning*. Melalui proses yang telah dilakukan, sejumlah 118 data dinyatakan terklasifikasi positif asli.
- ii. *True Negative* merupakan jumlah data dengan nilai sentimen negatif yang telah diklasifikasikan secara tepat oleh *machine learning*. Melalui proses yang telah dilakukan, sejumlah 48 data dinyatakan terklasifikasi negatif asli.
- iii. *Predictive Positive* merupakan jumlah data dengan nilai sentimen negatif yang telah diklasifikasikan sebagai positif oleh *machine learning*. Melalui proses yang telah dilakukan, sejumlah 31 data positif dinyatakan terklasifikasi negatif.

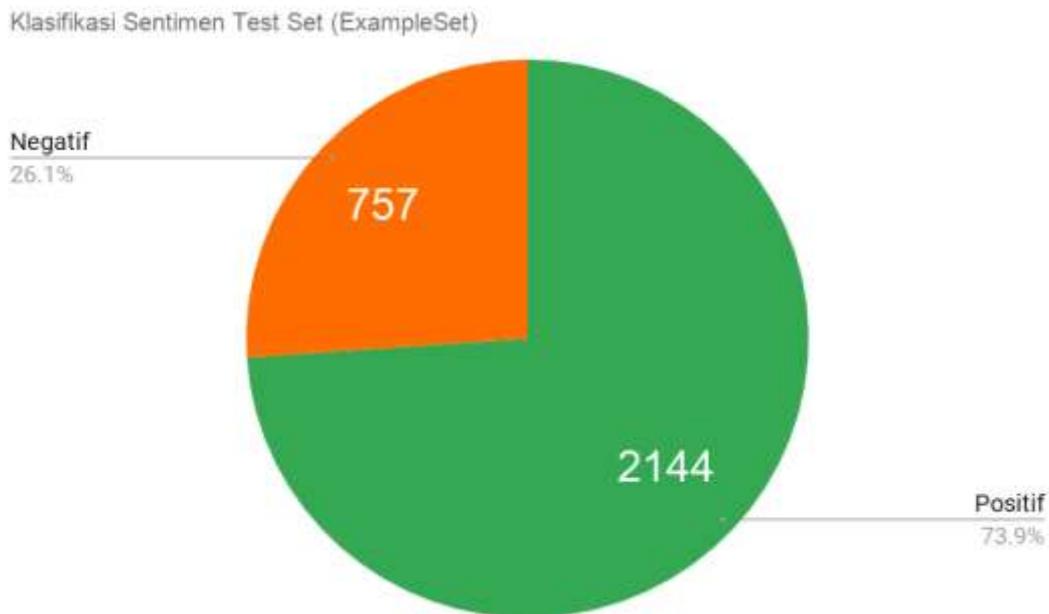
- iv. *Predictive Negative* merupakan jumlah data dengan nilai sentimen positif yang telah diklasifikasikan sebagai negatif oleh *machine learning*. Melalui proses yang telah dilakukan, sejumlah 23 data negatif dinyatakan terklasifikasi positif.

Hasil performa dilengkapi dengan perhitungan nilai F1 yang merupakan perbandingan nilai rata-rata *precision* dan *recall* label “positive” beserta label “negative” dengan nilai secara urut sebesar 81.37% dan 64%.

Melalui hasil perhitungan performa klasifikasi berdasarkan *confusion matrix* tersebut, dapat disimpulkan bahwa kemampuan algoritma *Naïve Bayes* untuk mengklasifikasikan setiap data dalam *dataset* yang digunakan pada penelitian ini adalah baik karena memiliki nilai performa di atas peluang 50% dengan nilai kesepakatan *moderate*.

Hasil Analisis Sentimen

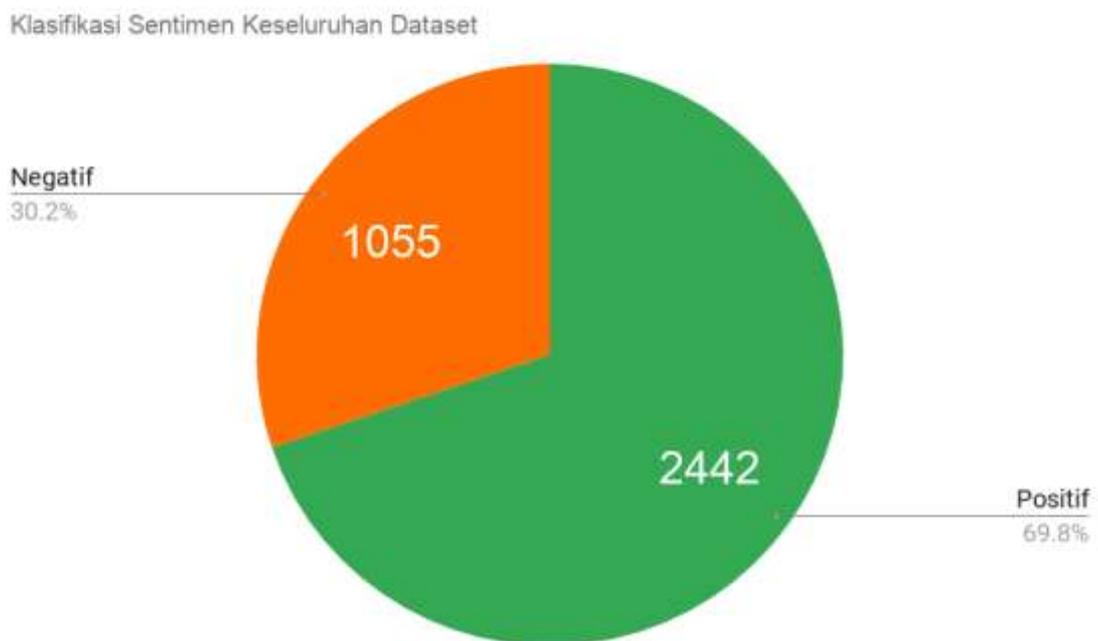
Proses klasifikasi menggunakan algoritma *Naïve Bayes* menunjukkan hasil sebagai berikut:



Gambar 1 Hasil Klasifikasi Sentimen pada Data *Testing Set*

Jumlah prediksi klasifikasi sentimen terhadap data *testing set* didominasi oleh sentimen positif. Proses klasifikasi tersebut telah memprediksikan 2144 (73.9%) ulasan

memiliki sentimen positif, sementara 757 (26.1%) ulasan memiliki sentimen negatif dari total 2901 data *testing*.



Gambar 2 Hasil Klasifikasi Sentimen pada Keseluruhan Dataset

Data *training set* yang telah dilabeli sebelumnya pada tahapan persiapan data digabungkan dengan data *testing set* yang telah terklasifikasi menggunakan *machine learning* untuk mengidentifikasi jumlah seluruh sentimen terkait produk *indie video game* lokal di Steam. Hasil penggabungan tersebut menunjukkan ulasan dengan sentimen positif berjumlah 2442 data (69.8%), yang mendominasi ulasan dengan sentimen negatif berjumlah 1055 data (30.2%), dengan selisih 1387 data.

Proses Pemodelan Topik

Pemodelan topik pada penelitian ini dilakukan menggunakan perangkat lunak RStudio. Untuk melaksanakan algoritma pemodelan topik dengan RStudio sejumlah alat berupa kode dalam *package* digunakan. Sejumlah *package* yang digunakan adalah “dplyr”, “tidyr”, “lubridate”, “ggplot2”, “textclean”, “tm”, “snowball”, “stringr”, “tidytext”, “topicmodels”.

Pemodelan topik yang digunakan adalah dengan algoritma *Latent Dirichlet*

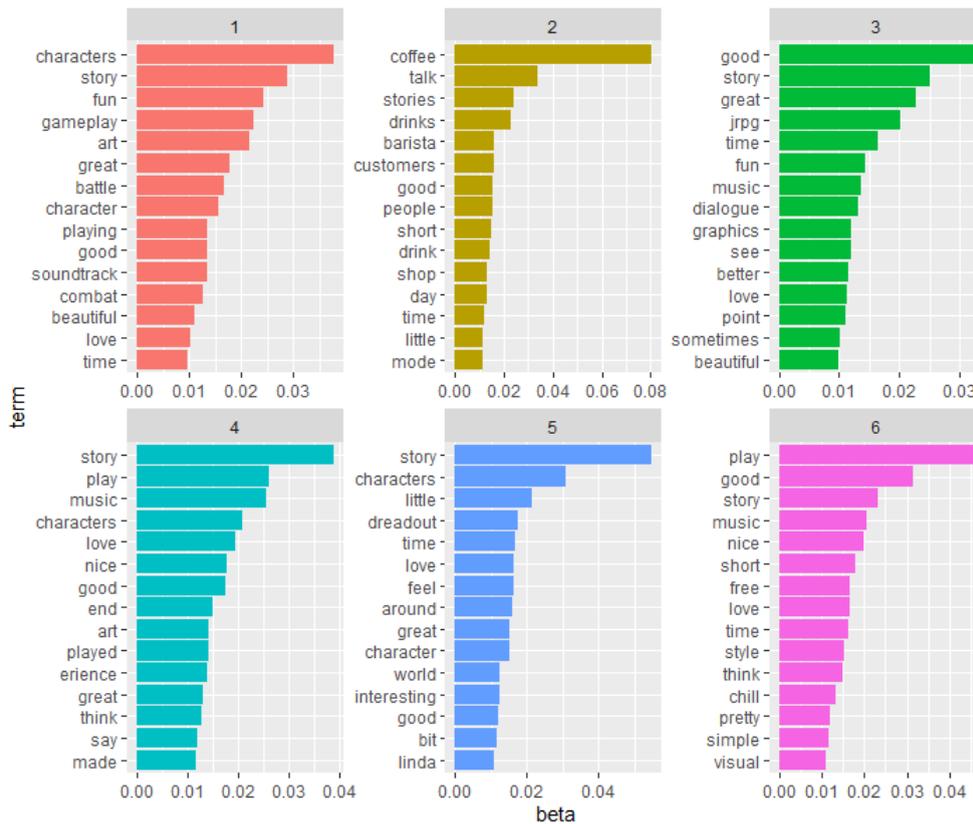
Allocation (LDA) yang merupakan model matematis untuk menemukan campuran kata-kata terhadap setiap topik, serta menentukan campuran topik-topik yang mendeskripsikan setiap dokumen teks. *LDA* menjawab prinsip-prinsip pemodelan topik di mana setiap dokumen merupakan campuran topik-topik dan setiap topik adalah campuran dari kata-kata. *LDA* sebagai model statistik generatif memungkinkan kumpulan hasil observasi untuk dijelaskan oleh kelompok-kelompok tidak terobservasi yang menjelaskan kemiripan sejumlah bagian dari data. Contohnya, jika hasil observasi adalah kumpulan kata-kata yang dikumpulkan dalam dokumen-dokumen, maka diasumsikan bahwa setiap dokumen adalah campuran dari sejumlah topik-topik dan setiap kata yang muncul dapat dikaitkan terhadap salah satu topik dalam dokumen tersebut (Cristian, 2020). Proses *LDA* mengasumsi bahwa setiap dokumen-dokumen baru terbentuk melalui tahapan berikut:

- i. Menentukan jumlah kata dalam dokumen
- ii. Memilih campuran topik terhadap dokumen dari kumpulan topik-topik yang tetap
- iii. Membentuk kata-kata dalam dokumen dengan cara:
 - Memilih topik berdasarkan distribusi multinomial dokumen
 - Memilih kata berdasarkan distribusi multinomial topik
- iv. Mengulang proses tersebut untuk sejumlah n iterasi hingga distribusi kata-kata dalam topik-topik memenuhi kriteria campuran topik pada tahapan ii.

Dalam penelitian ini, melalui RStudio, *package* “tidytext” akan membentuk pemodelan topik dan menyediakan metode untuk mengekstrak probabilitas per-topik-per-kata yang disebut sebagai β (beta) dari model yang digunakan. Kata-kata yang diambil secara otomatis melalui algoritma adalah kata-kata yang memiliki frekuensi kemunculan tertinggi yang setidaknya muncul pada 50 teks ulasan untuk membentuk sejumlah 6 topik dengan masing-masing maksimal 15 kata baik pada dataset ulasan dengan sentimen “positif” maupun dataset ulasan dengan sentimen “negatif”. Hasil proses algoritma *LDA* lalu ditampilkan dalam bentuk plot untuk diinterpretasikan.

Hasil Pemodelan Topik

Berdasarkan pengolahan data menggunakan pemodelan topik berbasis *LDA* melalui RStudio terhadap teks ulasan dengan sentimen positif dan sentimen negatif telah dihasilkan plot diagram sebagai berikut:

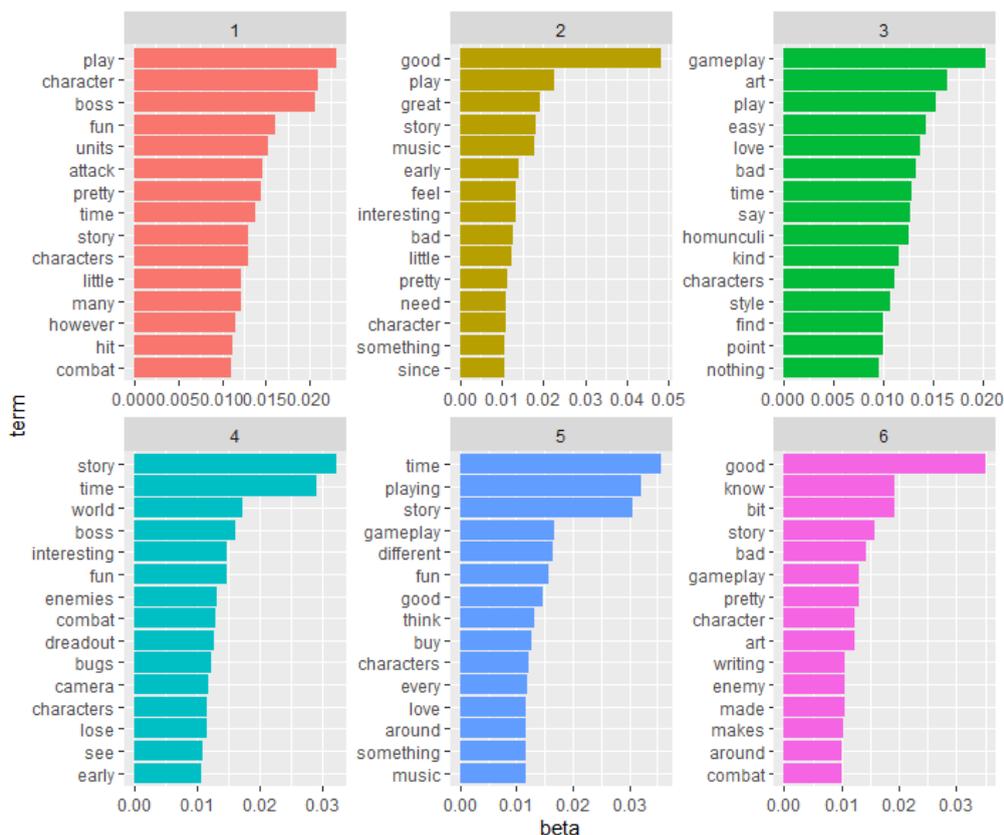


Gambar 3 Hasil Pemodelan Topik Dengan Sentimen Positif

Hasil plot terhadap ulasan dengan sentimen positif menunjukkan 6 topik yang paling dominan di antara seluruh ulasan *indie video game* lokal. Topik-topik yang teridentifikasi tersebut dapat diasumsikan sebagai berikut:

- i. Topik 1: Membahas aspek-aspek *video game* yang bagus terkait karakter, alur cerita, *gameplay*, seni, pertarungan, *soundtrack*, dan bagaimana para pemain menikmati waktu bermain dan mencintai *video game* yang dimainkan.
- ii. Topik 2: Membahas *video game* yang memiliki tema kopi dan barista beserta orang-orang dan pelanggannya, di mana *video game* tersebut memiliki cerita dan mode permainan yang singkat dan disukai oleh para pemain.
- iii. Topik 3: Membahas tentang *video game* dengan genre “JRPG” yang menyenangkan melalui alur cerita, musik, dan grafis yang indah sehingga

- membuat para pemain menikmati waktu mereka bermain dan mencintai *video game* tersebut, bahkan berharap adanya perkembangan/kelanjutan terkait *video game* tersebut di masa depan.
- iv. Topik 4: Membahas aspek-aspek *video game* terkait alur cerita, musik, karakter, seni dan pengalaman bermain yang baik sehingga membuat para pemain berpikir dan menyatakan bahwa game tersebut dibuat dengan baik dan membuat mereka mencintai *video game* tersebut.
 - v. Topik 5: Membahas aspek-aspek *video game* berjudul “DreadOut” dengan kandungan alur cerita, karakter, tema dunia yang sangat menarik dengan karakter utama bernama “Linda”. Para pemain menikmati waktu mereka bermain dan merasakan bahwa *video game* tersebut sangat bagus.
 - vi. Topik 6: Membahas aspek-aspek *video game* berupa *gameplay*, alur cerita, musik, *style*, dan visual yang sangat bagus sehingga membuat para pemain.



Gambar 4 Hasil Pemodelan Topik Dengan Sentimen Negatif

Hasil plot terhadap ulasan dengan sentimen negatif menunjukkan 6 topik yang paling dominan di antara seluruh ulasan *indie video game* lokal. Topik-topik yang

teridentifikasi tersebut dapat diasumsikan sebagai berikut:

- i. Topik 1: Membahas aspek-aspek *video game* yang bermasalah namun tetap dianggap bagus dan menyenangkan seperti *gameplay*, karakter, dan segala aspek pertarungan yang menyita waktu bermain.
- ii. Topik 2: Membahas aspek-aspek *video game* berupa *gameplay*, alur cerita, musik, dan karakter yang dianggap cukup buruk namun tetap menarik dan membutuhkan pengembangan karena masih dianggap baru.
- iii. Topik 3: Membahas tentang aspek-aspek *video game* berupa *gameplay*, seni, karakter, dan *style* pada *video game* yang melibatkan tema “homunculi” yang membuat para pemain menyatakan bahwa *video game* tersebut mudah dimainkan dan cukup buruk.
- iv. Topik 4: Membahas aspek-aspek *video game* terkait alur cerita, dunia, pertarungan, dan karakter yang, khususnya *video game* berjudul “DreadOut” yang dianggap memiliki banyak *bugs (error)* pada beberapa fiturnya seperti kamera, dan membuat para pemain kalah meskipun mereka tetap merasa senang menghabiskan waktu mereka bermain *video game* tersebut.
- v. Topik 5: Membahas aspek-aspek *video game* terkait alur cerita, *gameplay*, karakter, dan musik yang berbeda sepanjang permainan sehingga membuat para pemain senang membeli dan menghabiskan waktu bermain *video game* tersebut bahkan mencintainya.
- vi. Topik 6: Membahas aspek-aspek *video game* berupa alur cerita, *gameplay*, seni, penulisan cerita, dan pertarungan yang membuat para pemain menganggap *video game* tersebut cukup buruk.

Word Cloud (Awan Kata)

Word cloud memberikan tampilan visual terhadap jumlah kata dari satu teks atau lebih. Semakin sering suatu kata muncul, maka semakin besar tampilan visual kata tersebut dalam *word cloud* (Viégas & Wattenberg, 2008 dalam Ignatow & Mihalcea, 2018:298). Penelitian ini menggunakan perangkat lunak Orange untuk memproses data dan menghasilkan *word cloud*. Hasil *word cloud* dipisahkan antara teks ulasan dengan sentimen positif dan teks ulasan dengan sentimen negatif. Dengan itu, *word cloud* yang diperoleh adalah sebagai berikut:

menunjukkan bahwa sejumlah kata-kata yang memiliki frekuensi kemunculan tertinggi adalah “*time*”, “*story*”, “*good*”, “*character*”, dan “*play*”, masing-masing dengan nilai frekuensi kemunculan sebanyak 576, 565, 499, 482, dan 344 secara urut.

Kesimpulan

Berdasarkan hasil penelitian yang telah dijelaskan, peneliti dalam penelitian ini telah berhasil melakukan klasifikasi sentimen, identifikasi topik melalui pemodelan topik, hingga memvisualisasikan daftar kata menggunakan *word cloud* terhadap dataset ulasan *indie video game* lokal yang telah diperoleh dari platform Steam menggunakan teknik *text mining* melalui dukungan perangkat lunak *Web Scraper* pada tanggal 2 Februari 2021

Klasifikasi sentimen menggunakan *machine learning* menghasilkan 2144 ulasan (73.9% dari keseluruhan data *testing*) dengan sentimen positif dan 757 ulasan (26.1% dari keseluruhan data *testing*) dengan sentimen negatif. Akurasi performa yang diperoleh oleh klasifikasi *machine learning* memperoleh nilai 75.45% dengan nilai Kappa sebesar 0.454. Sementara hasil performa *class recall true positive* bernilai 83.69%, *class recall true negative* bernilai 60.76%, *precision class predictive positive* bernilai 79.19%, dan *precision class predictive negative* bernilai 67.61% dengan nilai F1 *precision* dan *recall* label “positive” dan label “negative” secara urut sebesar 81.37% dan 64%. Hal ini mengartikan bahwa performa klasifikasi yang dilakukan dianggap baik karena berada di atas peluang 50% dengan tingkat kesepakatan *moderate* (sedang) pada peluang acak 45%.

Penggabungan antara data *training set* dengan data *testing set* yang telah terklasifikasi menunjukkan adanya ulasan dengan sentimen positif berjumlah 2442 data (69.8%) dan ulasan dengan sentimen negatif berjumlah 1055 data (30.2%). Melalui hasil *sentiment analysis* tersebut dapat disimpulkan bahwa ulasan produk *indie video game* lokal di Steam secara garis besar memiliki sentimen positif, yang dipengaruhi oleh sejumlah topik dan kata/istilah yang diidentifikasi melalui *topic modeling* dan *word cloud*.

Ulasan produk *video game indie* lokal yang memiliki sentimen positif didominasi oleh topik dengan istilah “*good*”, “*story*”, “*characters*”, “*love*”, “*time*”, “*great*”, “*play*”, “*music*”, “*art*”, “*beautiful*”. Melalui pengidentifikasian kata-kata tersebut, dapat diasumsikan bahwa aspek-aspek yang menimbulkan sentimen positif kepada para pemain *video game indie* lokal adalah alur cerita, karakter, musik, dan seni yang bagus, indah, bahkan mereka cintai sehingga mereka menikmati waktu mereka memainkan *video game* tersebut.

Sementara ulasan produk *video game indie* lokal yang memiliki sentimen negatif didominasi oleh topik dengan istilah “*character*”, “*story*”, “*play*”, “*time*”, “*bad*”, “*combat*”, “*fun*”, “*gameplay*”, “*good*”, “*pretty*”. Melalui pengidentifikasian kata-kata tersebut, dapat diasumsikan bahwa aspek-aspek yang menimbulkan sentimen negatif kepada para pemain *indie video game* lokal adalah karakter, alur cerita, cara permainan, dan pertarungan yang cukup buruk sehingga mereka tidak menikmati waktu mereka bermain. Namun, meskipun para pemain memberikan opini negatif terhadap *video game* yang dimainkannya, mereka masih berusaha untuk memaparkan sisi positif dari produk tersebut yang dapat dianggap cukup baik bahkan menyenangkan. Asumsi tersebut didasari oleh fakta bahwa di antara data ulasan produk *video game indie* lokal yang digunakan dalam penelitian ini terdapat banyak konsumen yang memberikan ulasan secara detail dengan memaparkan baik keunggulan maupun kekurangan produk *video game* yang dimainkannya, melainkan hanya terfokus pada satu aspek yang dominan.

Penelitian ini telah membuktikan bahwa produk *indie video game* lokal untuk saat ini didominasi oleh sentimen positif. Oleh karena itu para pengembang *video game indie* lokal dapat mempertahankan kualitas produknya, bahkan meningkatkannya berdasarkan aspek-aspek *video game* yang terbukti mempengaruhi sentimen para *gamers*, yaitu “*story*” (alur cerita), “*characters*” (karakter), “*gameplay*” (cara permainan), “*music*” (musik), “*art*” (seni), dan “*combat*” (pertarungan) untuk bersaing di pasar *video game* global.

Disarankan untuk para peneliti selanjutnya yang bertujuan untuk menganalisis ulasan produk daring menggunakan teknik analisis sentimen dan pemodelan topik untuk meningkatkan efektivitas persiapan dan prapengolahan data demi menghasilkan hasil klasifikasi yang lebih akurat. Tidak hanya itu, para peneliti selanjutnya juga disarankan untuk membangun algoritma *machine learning* yang mampu melakukan klasifikasi terhadap data ulasan yang bersifat sangat kompleks, dalam kasus ini khususnya data ulasan *video game* yang cenderung mengandung istilah-istilah khusus dengan konteks tertentu baik berupa singkatan, kesalahan penulisan kata yang dilakukan secara sengaja, teks yang terdiri dari banyak kalimat dan frasa, hingga mengandung sarkasme yang mampu mempengaruhi performa *machine learning* secara negatif.

Sementara bagi para praktisi yang berkaitan dengan bisnis *video game* disarankan untuk memanfaatkan teknik *data mining*, khususnya dengan metode analisis sentimen dan

pemodelan topik demi memperoleh informasi penting terhadap pengembangan produk *video game* yang lebih baik. Bagi para pengembang *indie video game* disarankan untuk memfokuskan kualitas sejumlah aspek yang telah teridentifikasi dalam penelitian ini terhadap pengembangan *video game*, berupa aspek karakter, musik, alur cerita, seni, dan *gameplay*. Contohnya, berdasarkan hasil data ulasan yang terklasifikasi, para pemain *indie video game* lokal menunjukkan sentimen negatif terhadap karakter dan cerita yang tidak memiliki latar belakang luas dan menarik, begitupun *gameplay* yang terlalu kompleks dan dianggap tidak masuk akal, maupun *gameplay* yang terlalu sederhana. Sebaliknya, para pemain menunjukkan sentimen positif terhadap musik yang *catchy* (menarik), gambar/seni dunia dan karakter yang indah dan menarik, beserta tentu saja latar belakang karakter dan alur cerita yang menawan. Berdasarkan penelitian ini, peneliti menyimpulkan bahwa sentimen para pemain *indie video game* sangat dipengaruhi oleh aspek-aspek tersebut dibandingkan dengan sejumlah aspek *video game* lainnya.

Dengan memperbaiki *video game* yang telah dirilis ataupun meningkatkan kualitas *video game* yang akan datang terkait aspek-aspek yang telah diidentifikasi dalam penelitian ini, diharapkan para pemainnya akan merasa puas dan menuliskan ulasan yang menunjukkan sentimen positif serta merekomendasi *video game* tersebut. Ulasan *video game* mempengaruhi secara langsung keputusan pembelian, artinya ulasan yang memaparkan kualitas-kualitas terbaik *video game* tersebut dan merekomendasikannya dapat memberikan peluang yang tinggi dalam menarik calon-calon konsumen untuk membelinya, sehingga menghasilkan profit bagi pengembangnya. Tidak hanya itu, *video game* dengan ulasan yang dominan positif dan direkomendasi juga diharapkan dapat memberikan citra merek yang baik terhadap pengembangnya, sehingga selain menarik calon konsumen baru juga dapat menciptakan *customer base* yang berpeluang besar akan membeli *video game* terbaru selanjutnya oleh pengembang tersebut.

Referensi

- Antariksa, B. (2012). *Konsep Ekonomi Kreatif: Peluang dan Tantangan Dalam Pembangunan di Indonesia*.
- Asosiasi Game Indonesia. (2020). *Katalog Game Lokal Indonesia: Hari Game Indonesia 2020 08.08.2020*.
<https://static1.squarespace.com/static/5ddb7b78259a2f4dfb3423be/t/5f3678506eab486af6656959/1597405280903/Katalog+Game+Lokal+HARGAI+2020.pdf>
- Badan Ekonomi Kreatif. (2018). Opus: Ekonomi Kreatif Outlook 2019. In *Badan Ekonomi Kreatif*.
- Badan Pusat Statistik Republik Indonesia. (2019). *Berita Resmi Statistik: Pertumbuhan Ekonomi*.
- Chambers, C., Feng, W. C., Sahu, S., & Saha, D. (2005). Measurement-based characterization of a collection of on-line games. *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*, 1–14.
<https://doi.org/10.1145/1330107.1330109>
- Cristian, J. (2020). *Topic Modeling LDA using textmineR and tidytext*. R Pubs.
<https://rpubs.com/jojoecp/643113>
- Denby, L. (2019). *Indie Game Marketing 101 – Part 1 – A beginner’s guide to games marketing*. Game If You Are. <https://gameifyouare.com/2019/05/15/indie-game-marketing-101-part-1-a-beginners-guide-to-games-marketing/>
- Ernst and Young. (2015). *Cultural times - The first global map of cultural and creative industries* (Vol. 1, Issue 1). <https://doi.org/10.1016/j.physletb.2006.08.021>
- Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(1). <https://doi.org/10.1186/s40537-015-0015-2>
- Garda, M. B., & Grabarczyk, P. (2016). Is Every Indie Game Independent? Towards the Concept of Independent Game. *The International Journal of Computer Game Research*, 16(1). <http://gamestudies.org/1601/articles/gardagrabczyk>
- Henges, L. (2020). *Examining the indie ins and outs of today’s game distribution platforms*. Gamasutra.
https://www.gamasutra.com/view/news/363919/Examining_the_indie_ins_and_outs_of_todays_game_distribution_platforms.php
- Howkins, J. (2007). *The Creative Economy: How People Make Money From Ideas* (2nd ed., Vol. 1, Issue 1). Penguin Group.
- Ignatow, G., & Mihalcea, R. (2018). *An Introduction to Text Mining: Research Design, Data Collection, and Analysis*. SAGE Publications.
- Joshi, N., & Itkat, S. (2014). A Survey on Feature Level Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, 5(4), 5422–5425. <http://www.ijcsit.com/docs/Volume5/vol5issue04/ijcsit20140504135.pdf>
- Kang, H.-N., Yong, H.-R., & Hwang, H.-S. (2017). A Study of Analyzing on Online Game Reviews Using a Data Mining Approach: STEAM Community Data. *International Journal of Innovation, Management and Technology*, 8(2), 90–94.
<https://doi.org/10.18178/ijimt.2017.8.2.709>

- Khomsah, S. (2020). Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews. *Jurnal Penelitian Pos Dan Informatika*, 10(2), 157–168. <https://doi.org/10.17933/jppi.2020.100206>
- Knoema. (2019). *Top 100 Countries by Game Revenues*. Knoema. <https://knoema.com/infographics/tqldbq/top-100-countries-by-game-revenues>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.
- Lin, D., Bezemer, C. P., Zou, Y., & Hassan, A. E. (2019). An empirical study of game reviews on the Steam platform. In *Empirical Software Engineering* (Vol. 24, Issue 1). Empirical Software Engineering. <https://doi.org/10.1007/s10664-018-9627-4>
- Martin, C. B., & Deuze, M. (2009). The independent production of culture: A digital games case study. *Games and Culture*, 4(3), 276–295. <https://doi.org/10.1177/1555412009339732>
- Mathews, C. C., & Wearn, N. (2016). How Are Modern Video Games Marketed? *The Computer Games Journal*, 5(1–2), 23–37. <https://doi.org/10.1007/s40869-016-0023-2>
- Minor, J. (2020). *The Best Places to Buy and Rent PC Games Online in 2020*. PC Mag. <https://sea.pcmag.com/console-games/38669/the-best-places-to-buy-and-rent-pc-games-online>
- Newzoo. (2019). *2019 Global Games Market Report*. <https://newzoo.com/insights/trend-reports/newzoo-global-games-market-report-2019-light-version/>
- Newzoo. (2020). *2020 Global Games Market Report*. <https://platform.newzoo.com/reports>
- Oakley, K., & O'Connor, J. (2015). The Routledge companion to the cultural industries. In *Cultural Trends* (1st ed., Vol. 25, Issue 2). Routledge.
- Ramadhan, A. F. (2019). *Pengaruh Online consumer review, Potongan Harga, dan Citra Merek terhadap Keputusan Pembelian Game pada Aplikasi STEAM (Studi Pada Mahasiswa Universitas Brawijaya)*. <https://www.semanticscholar.org/paper/Pengaruh-Online-consumer-review%2C-Potongan-Harga%2Cdan-Ramadhan/a0436648fb02569c6710f5df372a1f1273442558?p2df>
- Sharma, P., Singh, D., dan Singh, A. (2015). Classification algorithms on a large continuous random dataset using rapid miner tool. 2015 2nd *International Conference on Electronics and Communication Systems*, 704-709.
- Smith, A. (2019). They Create Worlds. In *They Create Worlds* (1st Edition). CRC Press. <https://doi.org/10.1201/9780429423642>
- Toivonen, S., & Sotamaa, O. (2010). Digital distribution of games: The players' perspective. *Future Play 2010: Research, Play, Share - International Academic Conference on the Future of Game Design and Technology*, 199–206. <https://doi.org/10.1145/1920778.1920806>
- United Nations. (2018). *UNCTAD Creative Economy Outlook and Country Profile report (2018)* (p. 445). United Nations Conference on Trade and Development. https://unctad.org/en/PublicationsLibrary/ditcted2018d3_en.pdf
- Vashisht, P., & Gupta, V. (2016). Big data analytics techniques: A survey. *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 264–269. <https://doi.org/10.1109/ICGCIoT.2015.7380470>
- Wilson, J. L. (2020). *Steam*. PC Mag. <https://sea.pcmag.com/pc-games/5787/steam>

Zackariasson, P., & Wilson, T. L. (2012). *The video game industry: Formation, present state, and future* (1st ed.). Routledge.

Zuo, Z. (2018). *Sentiment Analysis of Steam Review Datasets using Naive Bayes and Decision Tree Classifier*.