

Rancang Bangun Ekstraksi Ekspresi Kata Kerja pada Ulasan Pelanggan Dengan *Text Chunking* untuk Memaparkan Pengalaman Penggunaan Produk

Almira Fiana Dhara dan Rully Hendrawan

Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: ruhendrawan@gmail.com

Abstrak—Saat ini, hampir seluruh situs e-commerce memiliki fitur ulasan pelanggan mengenai produk atau jasa yang telah dibeli dan dapat dibaca oleh seluruh pengguna. Seiring berjalannya waktu, ulasan pelanggan tersebut terus bertambah semakin banyak, terutama pada produk dan jasa yang populer. Sehingga, e-commerce menambahkan beberapa fitur seperti rating, ulasan positif dan netral populer, dan beberapa fitur lainnya untuk memudahkan pengguna. Tetapi, hal yang luput diperhatikan oleh e-commerce adalah e-commerce tidak menonjolkan pengalaman seperti apa yang didapat pengguna saat menggunakan produk atau jasa. Padahal, pengalaman tersebut merupakan informasi penting bagi pengguna agar mengetahui kegunaan dan manfaat dari produk dan jasa tersebut. Oleh karena itu, diperlukan fitur yang dapat menampilkan pengalaman penggunaan produk secara ringkas sehingga pelanggan dapat mengetahui secara umum kegunaan dan manfaat dari produk tersebut. Pada penelitian ini, pengalaman penggunaan produk terbatas pada aktivitas yang dilakukan pengguna ketika memakai produk tersebut. Metode yang akan digunakan adalah membuat prototipe dengan *Text Chunking* untuk mengekstrak informasi mengenai pengalaman penggunaan produk dari ulasan pelanggan. Luaran yang diharapkan dari penelitian ini adalah perangkat lunak yang menampilkan daftar kata kerja yang sering muncul pada ulasan sebagai representasi aktivitas yang dilakukan pengguna ketika memakai produk. Pada penelitian ini ditemukan bahwa ekstraksi ekspresi kata kerja menggunakan *text chunking* masih kurang efektif dikarenakan kalimat yang dihasilkan masih kurang sesuai dengan harapan setelah melakukan percobaan dengan menggunakan 3, 4 dan 5 chunks pada ekspresi frase kata kerja yang dihasilkan. Selain itu, tata bahasa dari kalimat yang dihasilkan juga masih kurang tepat dibandingkan dengan ulasan pelanggan yang sesungguhnya. Permasalahan tersebut dikarenakan dataset yang digunakan memiliki ejaan kata dan tata bahasa yang kurang tepat sehingga diharapkan penelitian selanjutnya dapat membuat model yang dapat melakukan *chunking* yang memiliki akurasi lebih tinggi dan memilih dataset yang tepat.

Kata Kunci: *Online Reviews, CRM, Text Chunking.*

I. PENDAHULUAN

E-COMMERCE yang beredar pada saat ini jumlahnya cukup banyak, dan masing-masing memiliki proses bisnis dan fitur yang berbeda yang dapat digunakan untuk memudahkan calon pengguna untuk membuat keputusan dalam pembelian produk selain hanya dengan memberikan deskripsi mengenai

spesifikasi produk. Salah satunya adalah adanya fitur ulasan pelanggan. Ulasan dari pelanggan tersebut telah menjadi salah satu sumber terpenting bagi pelanggan lainnya dalam membuat keputusan untuk membeli suatu produk maupun jasa, dimana sebelumnya sumber utama bagi pelanggan dalam pembelian suatu produk hanya pada iklan komersial maupun informasi produk yang disediakan oleh vendor [1].

Ulasan pelanggan biasanya memuat banyak informasi mengenai produk tersebut, mulai dari deskripsi produk dan kesesuaian produk dengan spesifikasi yang ditawarkan, pendapat pelanggan mengenai produk tersebut, hingga pengalaman pengguna saat menggunakan produk tersebut. Banyak calon pengguna mencari pendapat pelanggan mengenai produk tersebut yang tercermin dalam ulasan pelanggan. Pihak e-commerce juga telah menyadari hal tersebut, maka dapat dilihat perkembangan fitur ulasan pengguna sekarang, selain menampilkan ulasan pelanggan, dapat dilihat juga sentimen pengguna yang tercermin dalam *rating*. Selain itu, banyak e-commerce yang menyediakan fitur ulasan pelanggan juga dapat dilihat dari perspektif *helpfulness*, positif maupun netral.

Namun, selain sentimen pelanggan, salah satu hal yang paling sering dicari calon pengguna adalah pengalaman pelanggan lain tentang bagaimana memanfaatkan produk tersebut dalam aktivitasnya. Aktivitas tersebut tercermin dalam pengalaman pelanggan tersebut pada saat menggunakan produk tersebut. Saat ini, untuk memahami hal tersebut membutuhkan usaha yang cukup besar karena calon pengguna harus menyimpulkan sendiri pengalaman penggunaan produk tersebut dengan membaca ulasan pelanggan yang jumlahnya cukup banyak. Sehingga, waktu yang dibutuhkan calon pengguna untuk memutuskan pembelian suatu produk lebih lama sehingga calon pengguna menunda merasakan kebermanfaatannya dari produk tersebut. Hal tersebut juga merupakan hal yang baik bagi pemilik e-commerce karena penjualan menjadi tertunda. Sehingga, butuh cara yang lebih efisien untuk konsumen agar dapat mendapatkan informasi tersebut dengan lebih mudah.

Oleh karena itu, penelitian ini akan berfokus pada pengembangan sebuah *prototype* untuk melakukan ekstraksi pengalaman pengguna dalam menggunakan suatu produk berdasarkan ulasan yang telah ditulis pengguna. Dasar teori yang digunakan untuk melakukan ekstraksi ini adalah dengan *Natural Language Processing* yaitu cabang penelitian untuk menganalisis sekumpulan teks yang didasari oleh sekumpulan

teori dan teknologi. Model yang dapat dilakukan untuk mengolah informasi tersebut bermacam-macam, dan pada penelitian ini metode yang akan dikembangkan adalah model statistik *Chunking* berdasarkan algoritma *Maximum Entropy* dengan *library* yang terdapat pada OpenNLP. Ulasan akan diekstrak menjadi kata kerja yang mencerminkan pengalaman pengguna saat menggunakan produk. Selanjutnya, informasi yang berhasil diekstrak tersebut akan diurutkan mulai dari yang terbanyak muncul hingga yang paling sedikit muncul.

Penelitian ini diharapkan dapat membantu pengguna untuk mencari produk yang sesuai berdasarkan dengan pengalaman penggunaan produk tersebut. Selain itu, penelitian ini juga dapat membantu pengguna untuk mendapatkan informasi lebih cepat karena tidak perlu membaca banyak ulasan pelanggan untuk memahami fitur produk secara keseluruhan. Sehingga, pelanggan dapat dengan mudah untuk mengidentifikasi produk yang dibutuhkan berdasarkan pengalaman pengguna lain pada saat menggunakan produk tersebut.

II. DASAR TEORI

A. Apache OpenNLP

Apache OpenNLP adalah sebuah *library* yang berbasis *machine learning* yang dapat digunakan untuk melakukan pemrosesan teks dengan basis *natural language processing*. Apache OpenNLP mendukung beberapa teknik untuk melakukan NLP seperti tokenisasi, segmentasi kalimat, *Part-of-Speech Tagging*, *Named Entity Extraction*, *Chunking*, *Parsing* dan *Coreference Resolution*. Teknik-teknik ini biasanya dibutuhkan untuk membangun sistem pemrosesan teks yang lebih mumpuni. OpenNLP menggunakan *Maximum Entropy* dan *Perceptron* sebagai basis model dari *machine learning* yang digunakan pada *library* tersebut.

B. Google Play

Google Play adalah sebuah layanan distribusi digital meliputi pasar media digital yang dioperasikan dan dikembangkan oleh Google. Google Play dibuat sebagai *application store* resmi untuk sistem operasi berbasis Android, yang membuat pengguna dapat menjelajahi dan mengunduh aplikasi yang dikembangkan dengan Android SDK (*Software Development Kit*) dan diluncurkan melalui Google [2].

C. Natural Language Processing

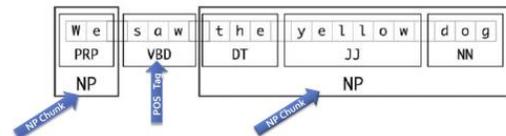
Natural Language Processing (NLP) adalah pendekatan terkomputerisasi yang digunakan untuk menganalisis sekumpulan teks yang didasari oleh sekumpulan teori dan teknologi. Saat ini, NLP telah menjadi salah satu area penelitian yang bertujuan untuk mengumpulkan ilmu pengetahuan mengenai bagaimana manusia dapat memahami dan menggunakan bahasa. Sehingga, teknologi sesuai dapat dikembangkan untuk memahami dan memanipulasi bahasa tersebut untuk menyelesaikan berbagai macam permasalahan yang diinginkan [3].

Tujuan utama dari NLP adalah untuk mencapai pemrosesan bahasa seperti manusia pada umumnya dengan menggunakan mesin [4]. Saat ini, NLP terus dikembangkan untuk mencapai tujuan utamanya tersebut dengan menggunakan berbagai macam pendekatan, model dan algoritma. Pada penelitian ini,

pendekatan yang digunakan adalah pendekatan dengan menggunakan algoritma *statistical machine learning*. Algoritma ini akan mengambil masukan berupa *features* yang didapatkan dari data tertentu.

D. Text Chunking

Text Chunking, atau biasa disebut juga dengan *Shallow Parsing*, adalah metode yang digunakan untuk membagi setiap kalimat menjadi beberapa *segment* yang tidak saling tumpang tindih [5]. *Chunking* merupakan salah satu alternatif metode yang berguna selain melakukan *full parsing*. Pada mulanya, *text chunking* digunakan sebagai tahap yang berguna sebelum melakukan *parsing*. Penelitian banyak menggunakan metode ini untuk mengekstrak kata benda yang berada pada level rendah dengan menggunakan berbagai jenis model statistika tertentu. Namun, metode ini terus dikembangkan sehingga *chunking* dapat digunakan untuk mengekstrak informasi dengan level yang lebih tinggi. Secara umum, *Chunking* berguna untuk menemukan frase benda, kelompok kata kerja, frase lainnya [6]. Secara umum, konsep *Chunking* dapat dilihat pada gambar berikut :



Gambar 1. Konsep Text Chunking

Chunking melakukan ekstraksi informasi dengan memecah setiap kalimat menjadi beberapa bagian yang biasa disebut *phrasal chunks*. *Phrasal chunks* tersebut lalu dapat disimpan dan diambil bagian yang diinginkan, pada penelitian ini dapat digunakan untuk mengekstrak kata kerja. *Chunking* mengekstrak informasi yang lebih akurat dari *POS Tagging* karena *POS Tagging* hanya memecah kalimat menjadi kata per kata yang diberi label dan mengabaikan keterkaitan kata tersebut dengan kata lainnya. Dengan melakukan *chunking*, kata yang saling terkait masih berada dalam satu *phrasal chunk* yang sama sehingga dapat memberikan informasi yang lebih jelas dibandingkan dengan *POS Tagging* [5].

E. Maximum Entropy Model

Maximum Entropy Model merupakan suatu model probabilistik statistika yang seringkali digunakan untuk memprediksi *Part-of-Speech* (POS) *Tagging*, yang merupakan metode dasar untuk melakukan *text mining* yang merupakan bagian dari NLP pada suatu *corpus*. Tingkat akurasi dari *Maximum Entropy Model* ini mencapai 96.6% dan telah menjadi salah satu model yang paling populer digunakan dalam penelitian berbasis NLP [7]. *Maximum Entropy Model* ini telah digunakan untuk *language modeling*, *machine translation*, *prepositional phrase attachment* dan *word morphology*.

Model probabilitas tersebut didefinisikan sebagai $H \times T$, dimana H merupakan sekumpulan kata yang mungkin dan konteks *tag*, sedangkan T adalah sekumpulan *tag* yang diperbolehkan. *Maximum Entropy Model* pada penelitian ini digunakan oleh OpenNLP sebagai dasar pembuatan model

untuk melakukan klasifikasi kata yang terdapat pada ulasan pelanggan. Model probabilitas dari *Maximum Entropy* dapat dilihat sebagai berikut [7] :

$$p(h, t) = \pi \mu \prod_{j=1}^k \alpha_j^{f_j(h,t)} \quad (1)$$

Dengan keterangan :

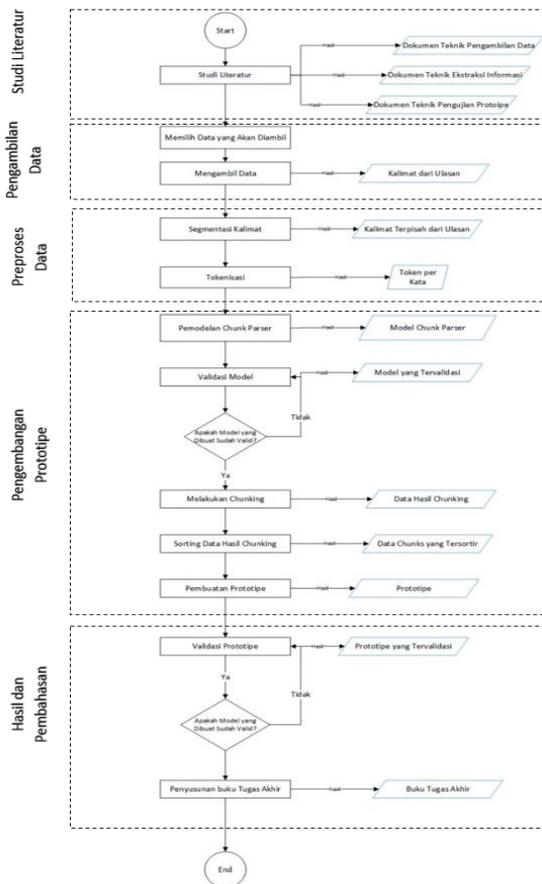
Π = konstan yang telah dinormalisasi

$\{\mu, \alpha_1, \dots, \alpha_k\}$ = parameter positif dari model

$\{f_1, \dots, f_k\}$ = *features* dari model

III. METODOLOGI

Pengerjaan penelitian ini tersusun atas beberapa langkah yang sistematis. Langkah-langkah pengerjaan penelitian dari awal hingga akhir ditunjukkan pada Gambar 2.



Gambar 2. Metodologi Pengerjaan Studi

IV. RANCANGAN PROTOTYPE

A. Ekstraksi Data Ulasan Pelanggan

Data yang digunakan pada penelitian ini merupakan dataset berupa teks dari salah satu situs E-Commerce, yaitu Google Play. Data yang diambil dari Google Play dan digunakan pada penelitian ini adalah data mengenai ulasan pengguna pada produk-produk yang dijual pada situs tersebut. Berikut ini merupakan hasil dari ekstraksi data tersebut :

Tabel 1.
Hasil Ekstraksi Data Ulasan Pelanggan

My plant vs zombie 2 app has stopped, how to fix it without clearing the data.
Game is not allowing me to log onto after the update. Please fix think the other premium plants can bought using real money or gems, and the others can only be bought with gems please read this!
It nice
Best game for our minds
It was a very nice game. More challenges.

Data ulasan pelanggan yang dimuat pada program didapat pada satu sumber, yaitu data teks yang dimuat sebelumnya yang telah dimasukkan ke dalam satu basis data. Dari 4830 ulasan pelanggan, data yang akan digunakan untuk diteliti lebih lanjut berjumlah 4000 ulasan pelanggan

B. Segmentasi Kalimat dan Tokenisasi

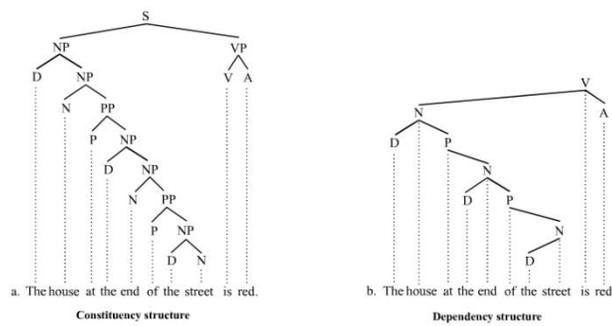
Segmentasi kalimat dan Tokenisasi merupakan proses yang dilakukan untuk merubah kalimat menjadi kata per kata yang disebut dengan token agar dapat dilakukan proses *labeling*. Sebagai contoh, Hasil dari segmentasi kalimat dan tokenisasi pada satu kalimat ulasan dapat dilihat pada tabel berikut :

Tabel 2.
Contoh Hasil Segmentasi Kalimat dan Tokenisasi

Teks Asli	Hasil Tokenisasi
Very pleased that Sun Bean is back like the old days. Homing Thistle still attacking air a lot. But with recent level up, it's attack is upgraded and it's not bad. Sadly Surfer Zombie and Balloon Zombie also upgraded.	[Very, pleased, that, Sun, Bean, is, back, like, the, old, days, ., Homing, Thistle, still, attacking, air, a, lot, ., But, with, recent, level, up, ., it, 's, attack, is, upgraded, and, it, 's, not, bad, ., Sadly, Surfer, Zombie, and, Balloon, Zombie, also, upgraded]

C. Labeling dengan POS Tagging

Labeling merupakan proses pemberian label terhadap kata sesuai dengan jenis kata tersebut. Pada penelitian ini digunakan Bahasa Inggris, maka label yang digunakan merupakan jenis-jenis kata yang terdapat pada bahasa tersebut. Dalam konsep linguistik, label ini diambil dari konsep *treebank*. *Treebank* adalah sebuah bagian dari korpus teks yang menganotasikan struktur kalimat yang sintetis atau semantik [8]. Konstruksi dari bagian korpus ini dikembangkan sejak akhir tahun 1990an sebagai bagian dari revolusi di linguistik komputasional yang merupakan keuntungan dari besarnya data empiris. Lebih jelasnya, konsep *treebank* tersebut dapat dilihat pada gambar berikut :



Gambar 3. Konsep Treebank Pada *English Linguistics* [3]

Sebagai contoh, berikut adalah hasil dari labeling dari kalimat yang telah dilakukan tokenisasi pada tahap sebelumnya :

Tabel 3. Contoh Hasil *Labeling* dengan POS Tagging

No	Kata	Label yang Dihasilkan	Kecocokan kata dengan label
1	Very	RB (Adverb)	Telah sesuai
2	Pleased	VBD (Past Tense)	Telah sesuai
3	That	IN (Preposition)	Telah sesuai
4	Sun	NNP (Proper Noun, Singular)	Telah sesuai
5	Bean	NNP (Proper Noun, Singular)	Telah sesuai
6	Is	VBZ (3rd person singular present)	Telah sesuai

D. VALIDASI MODEL

Validasi model dilakukan untuk memastikan bahwa model yang dibuat untuk melakukan ekstraksi ekspresi kata kerja ini telah mampu melakukan proses *chunking* tersebut dengan baik. Untuk melakukan validasi terhadap model, maka digunakan Apache OpenNLP, dengan *tools* validasi yang telah disediakan oleh *library* tersebut dengan menggunakan konsep Precision, Recall dan F-Measure sesuai dengan rancangan validasi pada penelitian ini.

Setelah menentukan target tersebut, maka dilakukan validasi terhadap model yang digunakan tersebut. Hasil validasi yang dilakukan terhadap model dapat dilihat pada tabel berikut :

Tabel 4. Hasil Validasi Model

Model yang Digunakan	Hasil Precision	Hasil Recall	Hasil F-Measure
en-chunker.bin	92,50%	92,22%	92,40%

E. PEMODELAN CHUNK PARSER

Setelah proses POS Tagging dilakukan, maka langkah selanjutnya adalah melakukan *chunking* dengan model Chunk Parser yang telah dibuat. Proses *chunking* yang dilakukan oleh model tersebut meliputi pemberian label *chunks* yang sesuai. Perbedaan label *chunks* dengan label yang diberikan pada POS Tagging pada dasarnya terletak pada kata yang dilabeli. *Chunks* akan membentuk frase-frase yang terdiri dari beberapa kata, sedangkan POS Tagging hanya mengandung satu kata yang dilabeli dengan jenis tertentu. Hal ini juga yang menyebabkan label yang terdapat pada konsep *chunking* jauh lebih sedikit dari label yang akan diberikan pada model POS Tagging. Karena itulah, meskipun *chunking* juga merupakan proses *labeling* terhadap token dari kata-kata yang terbentuk dari segmentasi

kalimat, namun *label* yang diberikan oleh model *chunking* merupakan *label* hasil dari penggabungan beberapa label dari POS Tagging. Secara sederhana, *chunking* menggabungkan label-label yang telah dihasilkan oleh POS Tagging menjadi frase-frase yang memuat informasi sederhana yang dapat lebih mudah dipahami oleh orang awam.

Hasil kalimat pemodelan chunk parser dengan batas maksimal *chunks* sejumlah 3 dan kalimat-kalimat tersebut telah dilakukan *sorting* dengan frekuensi paling sering muncul dapat dilihat pada tabel berikut :

Tabel 5. Hasil *Chunking* dan *Sorting* dengan 3 *Chunks*

Hasil Chunks	Frekuensi Muncul
i love this game	20
i love it	14
i love the game	13
i don't like	7
i think it	6
i play it	6

Hasil kalimat pemodelan chunk parser dengan batas maksimal *chunks* sejumlah 4 dan kalimat-kalimat tersebut telah dilakukan *sorting* dengan frekuensi paling sering muncul dapat dilihat pada tabel berikut :

Tabel 6. Hasil *Chunking* dan *Sorting* dengan 4 *chunks*

Hasil Chunks	Frekuensi Muncul
i love this game	4
i love it so	4
i lost all of	4
i opened it again	3
i installed it again	3
i love this game	4

Hasil kalimat pemodelan chunk parser dengan batas maksimal *chunks* sejumlah 5 dan kalimat-kalimat tersebut telah dilakukan *sorting* dengan frekuensi paling sering muncul dapat dilihat pada tabel berikut :

Tabel 7. Hasil *Chunking* dan *Sorting* dengan 5 *chunks*

Hasil Chunks	Frekuensi Muncul
i lost all of my progress	6
this gametry to add a new plant, the app got updated	5
there is playing time especially since the app crashes	5
i clear the data and i came back	5
my friends have where i can get free gems costume and coins	3
that need real money with a lot gems	3

V. VALIDASI PROTOTIPE

Validasi prototipe dilakukan untuk memastikan bahwa prototipe yang dihasilkan telah sesuai dengan kebutuhan dan tujuan dari penelitian ini. Metode yang digunakan untuk melakukan validasi prototipe ini adalah *Usability Testing*. *Usability Testing* merupakan salah satu metode yang digunakan untuk melakukan validasi terhadap prototipe suatu sistem. Dalam hal ini, *usability testing* digunakan untuk menguji

informasi yang dihasilkan dari *chunks* yang telah dilakukan oleh model ke pengguna.

Usability testing dilakukan dengan menyebarkan kuisioner yang berisi 10 pertanyaan mengenai prototipe yang telah dibuat, dan difokuskan pada informasi yang dihasilkan oleh model yang telah dibuat. Kuisioner tersebut lalu diberikan kepada sampel berupa 10 pengguna aplikasi *Plant vs Zombies* versi Android, yang otomatis mengunduh aplikasi tersebut melalui *Google Play* dan telah membaca ulasan pelanggan tersebut. Hasil *System Usability Scale* (SUS) yang diberikan oleh pengguna terhadap prototipe pada penelitian ini memiliki kegunaan sebesar 74,5% berdasarkan rata-rata nilai SUS yang diberikan oleh setiap pengguna. Hasil tersebut menunjukkan bahwa setiap rata-rata pengguna telah setuju bahwa prototipe ini berguna bagi mereka dalam mempermudah pemahaman informasi mengenai ulasan pelanggan terhadap aplikasi *Plant vs Zombies 2* yang terdapat pada *platform* *Google Play*.

Hasil dari SUS tersebut dapat dilihat pada gambar berikut :

SUS Calculation													
Responden	Jenis Kelamin	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Total	
Ashma Hanifah		2	3	4	3	3	2	3	1	3	3	4	77,5
Nadya Chandra		2	2	3	1	4	3	3	3	2	3	72,5	
Faisal S Putra		1	3	3	2	4	1	4	1	4	1	5	72,5
M Alvin Darari		1	3	4	2	3	2	3	3	3	1	4	72,5
Umar Al Agsho		1	2	3	2	4	2	3	1	4	2	3	67,5
Pri Rezki		1	3	3	1	3	3	4	2	3	3	4	75
Ni Ketut Ratih L.		2	3	4	3	5	2	3	1	3	2	3	77,5
Robbigh F		1	2	3	3	5	2	3	3	3	2	4	77,5
Hanum Fitriani Ayu K		2	3	5	1	3	3	3	2	3	3	3	77,5
Bambang Setiawan		1	3	3	1	4	2	5	2	4	2	3	75
Rata-rata SUS												74,5	

Gambar 4. Hasil Perhitungan Usability Testing

VI. KESIMPULAN

Dari hasil pengembangan penelitian yang telah dilakukan, pengembangan rancang bangun prototipe untuk ekstraksi kata kerja dengan metode *text chunking* dan menggunakan studi kasus ulasan pelanggan dari aplikasi *Plant vs Zombies 2* pada *platform* *Google Play*. Tujuan dari penelitian ini adalah untuk memaparkan pengalaman penggunaan produk yang berfungsi untuk memudahkan pelanggan dalam memahami ulasan pelanggan sehingga dapat Kesimpulan yang dapat diambil dari penelitian ini antara lain :

1. Label setelah melakukan praproses data, yaitu menggunakan model POS Tagging dari *library* yang sama, yaitu *OpenNLP* telah menghasilkan label-label yang akurat pada kata-kata yang berhasil dilabeli setelah melakukan pengecekan secara manual pada salah satu sampel ulasan pelanggan yang diambil. Pada sampel tersebut tidak terdapat kesalahan ejaan kata dan secara keseluruhan tata bahasa yang terdapat pada kalimat tersebut telah sesuai dengan susunan tata Bahasa Inggris sehingga model dapat dengan mudah memberikan label pada kalimat tersebut.
2. *Labeling* yang dilakukan pada model *text chunking* dengan menggunakan *library* *OpenNLP* pada dasarnya telah memiliki tingkat akurasi yang cukup tinggi dan dapat melabeli kata sesuai dengan konsep *text chunking* dengan benar, namun ekstraksi ekspresi kata kerja yang dihasilkan kurang sesuai dengan tujuan penelitian ini.

3. Penyebab utama dari kesalahan tata bahasa yang dihasilkan oleh prototipe terletak pada data yang digunakan. Ulasan-ulasan yang ditulis pelanggan terdapat banyak kesalahan ejaan pada kata dan tata bahasa yang tidak sesuai sehingga membuat model yang digunakan pada prototipe ini sulit untuk membentuk kalimat dengan tata bahasa yang sesuai.
4. Metode *sorting* yang digunakan pada prototipe ini menggunakan kombinasi *HashMap* dan *Arraylist* yang diharapkan dapat meningkatkan akurasi prototipe dalam mendeteksi kalimat yang sama. Namun, pada kenyataannya prototipe masih mengalami kesulitan dalam menemukan kalimat yang sama, terutama ketika menggunakan kalimat dengan 5 *chunks*, karena semakin banyak kata yang terkandung dalam suatu kalimat, maka semakin sulit bagi algoritma *sorting* untuk mendeteksi kesamaan yang dimiliki antar kalimat tersebut.

SARAN

Saran yang dapat diberikan berdasarkan hasil yang diperoleh dari penelitian ini adalah sebagai berikut :

- Data yang digunakan untuk melakukan penelitian ini sebaiknya dilakukan *cleansing* terlebih dahulu, yang bertujuan untuk mengeliminasi ulasan-ulasan pelanggan dengan kesalahan ejaan maupun tata bahasa yang dapat memperbesar peluang kesalahan ekstraksi pada ekspresi kata kerja yang telah dilakukan oleh model.
- Perlu dilakukan eksplorasi kembali terhadap model *chunker* yang digunakan, apakah telah sesuai dan dapat digunakan untuk melakukan ekstraksi ekspresi kata kerja yang berfungsi untuk memaparkan pengalaman penggunaan produk.
- Untuk pengembangan yang lebih baik, perlu dilakukan eksplorasi pelabelan lain agar meningkatkan efektivitas dan akurasi dari kalimat yang dihasilkan agar dapat memaparkan pengalaman penggunaan produk secara lebih akurat.

DAFTAR PUSTAKA

- [1] Y. D. and J. A. C. D. Mayzlin, *An Empirical Investigation of Online Review Manipulation*. Cambridge: National Bureau of Economic Research, 2012.
- [2] J. Callaham, "The new Google Store is your one-stop web shop for Chromebooks, Nexus devices and more," *AndroidCentral.com*, 2015. [Online]. Available: <http://www.androidcentral.com/new-google-store-your-one-stop-web-shop-chromebooks-nexus-devices-and-more>.
- [3] G. G. Chowdhury, "Natural Language Processing," 2003.
- [4] E. D. Liddy, *Natural Language Processing*, Encycloped. New York: Marcel Decker Inc, 2001.
- [5] L. A. R. and M. P. Marcus, "Text Chunking using Transformation-Based Learning," *J. Comput. Sci.*, pp. 82–95.
- [6] D. Zhu, *Text Chunking Using NLTK*. china: Microsoft Corporation, 2014.
- [7] A. Ratnaparkhi, "A Maximum Entropy for Part-of-Speech Tagging," *J. Artif. Intell.*, vol. 2, no. 1, p. 10, 2009.
- [8] C. F. a. S. L. A. Clark, *The handbook of computational linguistics and natural language processing*. Wiley, 2010.