

Analisis Pengaruh Normalisasi Teks pada Klasifikasi Sentimen Ulasan Produk Kecantikan

1st Zhafirah Rizqy Nur Shabrina Prasetija
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
zhafirahr@students.telkomuniversity.ac.id

2nd Ade Romadhony
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
aderomadhony@telkomuniversity.ac.id

3rd Erwin Budi Setiawan
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

Abstrak

Teks ulasan produk adalah suatu teks yang berisi ulasan, penilaian atau *review* terhadap sebuah karya. Terdapat banyak jenis teks ulasan produk, salah satunya berupa teks ulasan produk kecantikan yang berfungsi sebagai masukan untuk meningkatkan kualitas suatu produk kecantikan. Tetapi, banyak teks ulasan produk kecantikan yang menggunakan bahasa informal sehingga diperlukan normalisasi teks. Normalisasi teks merupakan istilah yang digunakan untuk mengkonversi format teks dengan tujuan yang sudah ditentukan. Penggunaan normalisasi teks pada tahap *pre-processing* biasanya mampu menaikkan kualitas teks ulasan produk kecantikan yang akan diklasifikasi. Pada Tugas Akhir ini, diimplementasikan normalisasi teks untuk ulasan produk kecantikan, dengan tujuan untuk menaikkan kinerja klasifikasi sentimen. Normalisasi teks dilakukan dengan metode berbasis kamus, yang didefinisikan berdasarkan daftar *internet slang* dan *spelling correction*. Hasil eksperimen menunjukkan bahwa dengan normalisasi teks berbasis kamus, terdapat selisih akurasi sebesar 0,54% antara proses klasifikasi tanpa normalisasi teks dan proses klasifikasi dengan normalisasi teks.

Kata kunci: Normalisasi teks, Teks ulasan produk kecantikan, Internet slang, Spelling correction

Abstract

Product review text is a text that contains a review, rating or review of a work. There are many types of product review texts, one of which is a beauty product review text that serves as input to improve the quality of a beauty product. However, many beauty product review texts use informal language, so text normalization is necessary. Text normalization is a term used to convert text formats with a predetermined purpose. The use of text normalization at the pre-processing stage is usually able to improve the quality of the text of the beauty product reviews to be classified. In this final project, text normalization is implemented for beauty product reviews, with the aim of improving the performance of sentiment classification. Text normalization is carried out using a dictionary-based method, which is defined based on a list of internet slang and spelling correction. The experimental results show that with dictionary-based text normalization, there is an accuracy difference of 0.54% between the classification process without text normalization and the classification process with text normalization.

Keywords : Text normalization, Beauty product review text, Internet slang, Spelling correction

Di era globalisasi, informasi merupakan sebuah hal yang tidak dapat dipisahkan dari berbagai aspek masyarakat. Setiap orang dapat mengakses informasi dengan mudah. Terdapat banyak informasi yang beredar, mulai dari berita, artikel, hingga teks ulasan. Teks ulasan adalah suatu teks yang berisi ulasan, penilaian atau *review* terhadap sebuah karya [1]. Selain karya, teks ulasan dapat juga berisi ulasan, penilaian atau *review* dari sebuah produk. Salah satu contoh teks ulasan produk adalah teks ulasan produk kecantikan. Saat ini jumlah teks ulasan produk kecantikan yang dibuat dan diunggah oleh masyarakat semakin bertambah. Karena setiap pihak, baik konsumen maupun produsen membutuhkan teks ulasan untuk meningkatkan kualitas produk dan mendapatkan kepercayaan dari konsumen. Namun, penggunaan bahasa pada teks ulasan produk umumnya berupa teks bebas yang bersifat informal, dan terdapat penggunaan istilah bahasa asing di dalamnya.

Berdasarkan kondisi di atas, dibutuhkan sebuah proses untuk mengolah teks ulasan produk berupa normalisasi teks. Normalisasi teks merupakan salah satu bagian dari pengembangan sistem pada Pemrosesan Bahasa Alami (*Natural Language Processing*) yang ditujukan untuk *preprocessing* data berbasis teks. *Preprocessing* pada data dapat berupa pembentukan model bahasa hingga pemotongan teks ke kalimat maupun kata. Teks dengan penggunaan bahasa yang bersifat informal perlu dilakukan normalisasi, agar komputer dapat memahami bahasa yang digunakan oleh manusia. Salah satu metode normalisasi teks yaitu pengecekan terhadap *internet slang* pada teks ulasan produk. Menurut Magali et al. (2014) *internet slang* merupakan kelas yang menggabungkan :

- 1) kata yang dituliskan dengan cara yang berbeda dan menggunakan singkatan dari ekspresi berulang;
- 2) penggunaan huruf dan tanda baca yang berulang (contohnya : “cocok sama skin tone orang indo juga. lovveeeee!”);
- 3) urutan huruf yang terkait dengan pengekspresian emosi, seperti *emoticon* (contohnya : “berhubung bibirku bukan tipe bibir kering jadi aku tidak pernah bermasalah dengan lipstick matte :)”, “ngasih efek brighten up your face :)”), tertawa (contohnya : “namanya juga cewek hihhi” *fyi, i'm foodzilla kkk~”) yang digunakan pada tujuan tertentu [2].

Normalisasi teks yang difokuskan pada Tugas Akhir ini adalah normalisasi teks pada teks ulasan produk menggunakan prosedur normalisasi teks dengan pendekatan penggunaan kamus Bahasa Indonesia. kemudian teks yang

I. PENDAHULUAN

A. Latar Belakang

sudah dinormalisasi akan dianalisis pengaruh hasil normalisasi terhadap keakuratan hasil klasifikasi sentimen positif dan negatif dari teks ulasan yang diuji.

B. Rumusan Masalah

Adapun rumusan masalah yang didapat berdasarkan latar belakang yaitu bagaimana pengaruh hasil normalisasi teks ulasan produk kecantikan dengan menggunakan Bahasa Indonesia terhadap keakuratan hasil klasifikasi sentimen positif dan negatif.

C. Tujuan

Tujuan dari Tugas Akhir ini adalah untuk menganalisis pengaruh hasil normalisasi teks ulasan produk kecantikan dengan menggunakan Bahasa Indonesia yang mempertimbangkan penggunaan *internet slang* dan *spelling correction* terhadap keakuratan hasil klasifikasi sentimen positif dan negatif.

D. Organisasi Tulisan

Laporan tugas akhir ini disusun atas lima bagian. Bagian pertama menjelaskan pendahuluan yang berisi latar belakang, rumusan masalah dan tujuan dari penelitian ini; bagian kedua menjelaskan studi terkait dan penelitian yang berhubungan dengan normalisasi teks; bagian ketiga menjelaskan alur proses normalisasi pada sistem yang dibangun; bagian keempat menjelaskan tentang analisis hasil normalisasi teks pada sistem; dan bagian kelima merupakan penjelasan kesimpulan dari hasil penelitian ini.

II. KAJIAN TEORI

A. Normalisasi Teks

Normalisasi Teks merupakan istilah yang digunakan untuk menyampaikan gagasan mengkonversi format teks dengan tujuan yang sudah ditentukan [2]. Dalam bidang Pemrosesan Bahasa Alami, terdapat banyak proses normalisasi teks yang bergantung pada penggunaannya yaitu : i) jenis teks yang dimasukkan; ii) format keluaran/output yang diinginkan; iii) tujuan dari normalisasi teks yang dilakukan; iv) metode yang digunakan dalam proses normalisasi teks. Menurut Jurafsky [3], normalisasi teks dilakukan sebelum memulai pemrosesan bahasa alami pada teks. Setidaknya ada tiga tahapan yang dilakukan dalam proses normalisasi teks, yaitu :

- Tokenisasi (pemecahan kalimat menjadi kata)
- Normalisasi format kata
- Segmentasi pada kalimat (pemecahan kalimat untuk diambil maknanya)

Saat ini sudah banyak penelitian mengenai normalisasi teks, salah satunya pada konten yang dibuat oleh user (*user generated content*) berupa teks ulasan produk. Pada penelitian Magali S. Duran [2] yang berjudul "*Some Issues on the Normalization of a Corpus of Product Review*" disebutkan bahwa beberapa masalah yang diangkat pada proses normalisasi teks, yaitu *true casing* (mengubah huruf kapital menjadi huruf kecil), tanda baca, kesalahan pengejaan (*spelling correction*) dan *internet slang*.

True casing pada penelitian sebelumnya dilakukan dengan tujuan untuk mengecek apakah teks ulasan tersebut memiliki penggunaan huruf kapital yang buruk atau tidak, dan melindungi entitas bernama (*named entity*) saat

pemeriksaan ejaan agar mengurangi resiko kesalahan ketika pemeriksaan ejaan. Prosedur *true casing* yang dilakukan pada penelitian diatas yaitu :

- 1) Menyerahkan sampel untuk diberi *named entity* Bahasa Portugis,
- 2) Mengubah entitas yang diketahui ke dalam huruf kapital
- 3) Mengembalikan teks yang sudah diubah ke huruf kapital (kecuali *named entity*) ke huruf kecil.

Pada penggunaan tanda baca, beberapa aturan tata bahasa dapat digunakan untuk memperbaiki penggunaan koma, tetapi masalah yang dihadapi lebih kompleks. Untuk mengatasi kasus ini dibangun alat segmentasi kalimat yang tidak tergantung pada adanya tanda baca atau pelipatan huruf, karena ini noise yang utama dalam korpus.

Dalam kasus kesalahan pengejaan, analisis secara manual dilakukan untuk menentukan apakah kata tersebut telah dikoreksi secara akurat atau tidak. Seperti yang disebutkan, peninjau produk memiliki beberapa tingkat literasi dan mereka yang tingkatannya lebih rendah sering menukar huruf konsonan yang menyampaikan nilai fonetis yang sama. Aturan-aturan ini harus dimasukkan dalam pemeriksa ejaan. Dengan cara yang sama, kata-kata asing yang digunakan dalam bahasa Portugis Brasil harus dimasukkan dalam pemeriksa ejaan untuk meningkatkan saran perbaikan.

Selain tiga kasus di atas, terdapat satu kasus berupa *internet slang*. Menurut Magali et al. (2014), *internet slang* merupakan kelas yang menggabungkan : 1) kata yang dituliskan dengan cara yang berbeda dan menggunakan singkatan dari ekspresi berulang; 2) penggunaan huruf dan tanda baca yang berulang (contohnya "!!!!!!!", dan "ameiiiiiiiiiiiiiiiiiiii", di mana kata "amei" = "cinta" ditekan.); 3) urutan huruf yang terkait dengan pengekspresian emosi, seperti emoticon (contohnya : ":"), " : = (") tertawa (contohnya : "srsrsrs, hehehe, kkkkkkkk") yang digunakan pada tujuan tertentu. Prosedur yang berkaitan dengan *internet slang* akan diterapkan dengan hati-hati untuk memungkinkan pengguna untuk mengaktifkan masing-masing dari tiga prosedur secara terpisah, tergantung pada kebiasaan dalam menggunakan ekspresi emosi atau tidak.

B. Teks Ulasan Produk

Teks ulasan adalah suatu teks yang berisi ulasan, penilaian atau review terhadap sebuah karya [1]. Selain karya, teks ulasan dapat juga berisi ulasan, penilaian atau review dari sebuah produk. Teks ulasan produk adalah suatu teks yang berisi ulasan, penilaian atau review terhadap sebuah produk. Tujuan dari teks ulasan produk adalah untuk meningkatkan kualitas produk dan mendapatkan kepercayaan dari konsumen. Ciri-ciri dari teks ulasan produk yaitu:

- Strukturnya terdiri atas: Orientasi, Tafsiran, Evaluasi, dan Rangkuman.
- Memuat informasi berdasarkan pandangan atau opini penulis terhadap suatu karya atau produk.
- Opinions berdasarkan fakta yang diinterpretasikan.
- Dikenal dengan istilah lain yaitu resensi.

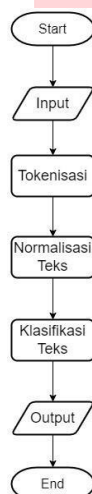
C. Pengaruh Normalisasi Teks pada Klasifikasi Teks

Pada proses klasifikasi teks, terdapat proses awal berupa *pre-processing* untuk membersihkan data maupun melakukan normalisasi pada teks. Proses normalisasi teks memiliki pengaruh yang cukup baik pada proses klasifikasi. Pada penelitian yang dilakukan oleh Magali et al [2], proses normalisasi yang dilakukan pada penelitiannya terhadap tool untuk normalisasi teks pada kasus klasifikasi opini terhadap teks yang dibuat oleh user (*user-generated content*) menunjukkan bahwa dengan adanya normalisasi teks sebelum proses klasifikasi opini, mampu meningkatkan performansi sebanyak 2% dibandingkan dengan tanpa adanya normalisasi teks.

III. METODE

A. Deskripsi Umum Sistem

Deskripsi umum sistem berisikan rancangan keseluruhan sistem yang akan digunakan dalam penelitian yang dilakukan. Berikut merupakan perancangan sistem yang diperlukan untuk kasus pada tugas akhir:



Gambar 1. Diagram Alir Sistem

Tahapan dari sistem di atas yaitu :

- **Tokenisasi**
Pada tahap ini, kumpulan teks ulasan produk akan ditokenisasi atau dipecah menjadi kata per kata untuk melanjutkan ke proses selanjutnya
- **Normalisasi Teks**
Setelah dipecah menjadi kata per kata, kemudian dicek apakah terdapat *internet slang* pada teks ulasan produk yang ada dan kata yang sudah dicek pada proses sebelumnya akan diperbaiki ejaannya.
- **Klasifikasi Teks**
Pada tahapan ini, setelah teks ulasan produk dinormalisasi dengan tiga tahapan sebelumnya kemudian akan diklasifikasikan sesuai dengan kelas yang ditentukan (berupa *rating* dengan nilai 1 sampai 5).

B. Perancangan Sistem

1. Tokenisasi

Proses tokenisasi merupakan salah satu contoh tahapan pada normalisasi teks berupa pemecahan kalimat menjadi kata. Proses tokenisasi pada tugas akhir ini dilakukan ketika dataset yang berupa kumpulan teks ulasan produk sudah dibagi ke dalam data latih (*data train*) dan data uji (*data testing*), kemudian kumpulan teks ulasan produk yang sudah dibagi akan ditokenisasi untuk memecah kalimat menjadi kata demi kata (termasuk memisahkan tanda baca). Hal ini dilakukan agar mempermudah proses selanjutnya.

2. Normalisasi Teks

Setelah kumpulan teks ulasan produk ditokenisasi, pada proses ini dilakukan pengecekan terhadap *internet slang* yang terdapat pada kumpulan teks ulasan produk untuk menghilangkan kata yang memiliki ciri-ciri dari *internet slang* (penggunaan huruf dan tanda baca yang berlebihan maupun urutan huruf yang terkait dengan pengekspresian emoticon, contohnya : “cocok sama skin tone orang indo juga. love”) untuk mempermudah proses selanjutnya.

Pada penelitian Tugas Akhir ini terdapat 3 jenis kamus yang digunakan. *Internet slang* keseluruhan (domain kosmetik dan umum), *Internet slang* dengan domain kosmetik dan *Internet slang* yang bersifat umum. *Domain* kosmetik pada kasus ini adalah istilah yang digunakan pada ranah pembahasan ulasan produk kosmetik. Proses pengecekan *internet slang* pada kasus ini akan dilakukan secara otomatis. Jumlah keseluruhan kata yang digunakan pada kamus adalah 678 kata, yang dibagi menjadi 2 kamus untuk pengecekan berbasis *domain*. Kata yang diperoleh untuk kamus didapatkan dari *data set* yang digunakan pada penelitian ini.

Kemudian, kumpulan teks ulasan produk yang sudah dilakukan pengecekan terhadap *internet slang* akan dicek kembali apakah terdapat kata yang salah penulisannya maupun kata yang disingkat agar dapat dikembalikan ke dalam bentuk kata yang sesuai dengan Ejaan Bahasa Indonesia yang baik dan benar (contohnya : “ warna merah nya pas tidak terlalu gelap dan tidak terlalu terang”).

Kamus yang digunakan pada *spelling correction* merupakan kamus yang sama dengan kamus pengecekan *internet slang*. Terdapat 110 kata yang terdapat pada kamus. Cara penyusunannya yaitu jika ada kata yang memiliki huruf lebih dari 1 huruf yang sama atau ada huruf yang kurang, maka akan diperbaiki menjadi kata aslinya.

3. Klasifikasi Teks

Proses Klasifikasi Teks merupakan proses penggolongan teks kedalam suatu kategori tertentu berdasarkan isi teks tersebut. Pada tugas akhir ini, teks ulasan produk yang sudah dinormalisasi dengan tiga tahap sebelumnya akan dikategorikan kedalam lima kelas, yaitu kelas dengan sentimen sangat baik (bernilai 5), kelas dengan sentimen baik (bernilai 4), kelas dengan sentimen cukup (bernilai 3), kelas dengan sentimen kurang (bernilai 2) dan kelas dengan sentimen sangat kurang (bernilai 1) berdasarkan ulasan produk yang terdapat di *dataset*

untuk dilihat pengaruhnya terhadap proses normalisasi yang dilakukan di tiga tahap sebelumnya.

Metode klasifikasi teks yang akan digunakan pada kasus ini adalah *Multinomial Naive Bayes*. Pada penelitian sebelumnya yang dilakukan oleh Lohar et al.(2017) [4], *Multinomial Naive Bayes* merupakan contoh spesifik dari metode klasifikasi *Naive Bayes* yang menggunakan distribusi multinomial untuk setiap fitur alih-alih mengacu pada independensi bersyarat dari masing-masing fitur dalam model. Sehingga sesuai untuk melakukan proses klasifikasi teks dengan banyak kelas. Pada penelitian ini, total data uji yang digunakan berjumlah 3100 ulasan dengan persebaran sebagai berikut:

Tabel 1. Persebaran jumlah data uji berdasarkan kelas sentiment

Kelas Sentimen	Jumlah
Sentimen sangat baik (bernilai 5)	869 ulasan
Sentimen baik (bernilai 4)	734 ulasan
Sentimen cukup (bernilai 3)	623 ulasan
Sentimen kurang (bernilai 2)	613 ulasan
Sentimen sangat kurang (bernilai 1)	261 ulasan

IV. HASIL DAN PEMBAHASAN

Dalam eksperimen ini, dilakukan pengujian sistem yang digunakan untuk mengetahui pengaruh normalisasi teks pada sistem klasifikasi sentimen ulasan produk yang di bangun. Terdapat tiga skenario yang dilakukan untuk mendapatkan akurasi yang menunjukkan perbedaan pada akurasi pada sistem yang dibangun, diantaranya yaitu menggunakan kamus berisi *internet slang* keseluruhan, *internet slang* dengan *domain* kosmetik dan *internet slang* Bahasa Indonesia yang bersifat umum. Tabel dibawah ini merupakan daftar kamus yang digunakan dalam skenario pengujian pada eksperimen ini.

Tabel 2. Daftar kamus yang digunakan pada pengujian

Skenario	Kamus yang digunakan
Skenario 1	<i>Internet slang</i> keseluruhan (<i>domain</i> kosmetik dan umum)
Skenario 2	<i>Internet slang</i> dengan <i>domain</i> kosmetik
Skenario 3	<i>Internet slang</i> yang bersifat umum

A. Analisis Data Uji

Analisis data uji yang dilakukan pada eksperimen ini dilakukan dengan mengecek kata yang paling sering muncul pada data latih dan data uji yang sudah dilakukan proses normalisasi teks dan digabungkan menjadi satu *file* tersendiri. Dari keseluruhan jumlah ulasan produk

yang terdapat pada data tersebut, diambil 20 kata yang paling sering muncul di setiap *rating*. Berikut adalah tabel dari kata yang paling sering muncul pada setiap *rating*

Tabel 3. Daftar kata-kata yang paling sering muncul

Rating	Kata yang Paling Sering Muncul (Sebelum Normalisasi)	Kata yang Paling Sering Muncul (Sesudah Normalisasi)
5	aku,banget, jadi, kulit, pake, ga, tapi, juga, bikin, pakai, sama, lebih, produk, suka, saya, buat, gak, muka, bisa, setelah	aku, pakai, banget, buat, kulit, jadi, tidak, produk, kalau, sama, muka, suka, lebih, gak, wajah, beli, kering, cocok, udah, warna
4	aku, jadi, tapi, kulit, ga, pake, banget, bikin, juga, lebih, pakai, suka, buat, saya, produk, bisa, sama, kalo, gak, setelah	pakai, aku, kulit, buat, tidak, jadi, banget, kalau, lebih, produk, suka, warna, sama, sih, kering, muka, wajah, gak, bibir, dipakai
3	aku, banget, jadi, kulit, pake, ga, tapi, juga, bikin, pakai, sama, lebih, produk, suka, saya, buat, gak, muka, bisa, setelah	aku, pakai, banget, buat, kulit, jadi, tidak, produk, kalau, sama, muka, suka, lebih, gak, wajah, beli, kering, cocok, udah, warna
2	aku, ga, jadi, tapi, kulit, pake, bikin, produk, gak, juga, pakai, saya, setelah, sama, banget, beli, cocok, muka, lebih, suka	aku, tidak, pakai, jadi, kulit, buat, banget, produk, muka, cocok, gak, kalau, sama, beli, kering, sih, lebih, suka, aja, bibir
1	aku, ga, jadi, gak, kulit, pake, tapi, produk, sama, bikin,banget, muka,cocok, juga, pakai,beli, buat, saya, pas, setelah	aku, tidak, pakai, buat, kulit, banget, jadi, gak, sama, produk, cocok, muka, kering, beli, jerawat, sih, dipakai, wajah, bibir, kalau

B. Pengujian Pada Skenario 1

Tabel 4 Hasil pengujian pada Skenario 1

Kamus yang digunakan	Kondisi	Akurasi
<i>Internet slang</i> keseluruhan (<i>domain</i> kosmetik dan umum)	Sebelum dilakukan Normalisasi Teks	47.69%
	Sesudah dilakukan Normalisasi Teks	47.15%

Dari hasil pengujian di atas, akurasi yang diperoleh pada kondisi sebelum normalisasi teks sebesar 47. 69% sedangkan akurasi pada kondisi setelah normalisasi teks sebesar 47.15%. Dari kedua kondisi tersebut terjadi perbedaan sebesar 0.54%. Pada pengujian skenario 1, kamus

yang digunakan tidak memperhatikan keberadaan *domain* sehingga hasil normalisasi teks lebih sesuai. Berikut adalah contoh hasil normalisasi teks pada skenario 1.

C. Pengujian Pada Skenario 2

Tabel 5 Hasil pengujian pada Skenario 2

Kamus yang digunakan	Kondisi	Akurasi
Internet slang dengan domain kosmetik	Sebelum dilakukan Normalisasi Teks	47.69%
	Sesudah dilakukan Normalisasi Teks	46.55%

Dari hasil pengujian di atas, akurasi yang diperoleh pada kondisi sebelum normalisasi teks sebesar 47.69% sedangkan akurasi pada kondisi setelah normalisasi teks sebesar 46.55%. Dari kedua kondisi tersebut terjadi perbedaan sebesar 1.14%. Pada pengujian skenario 2, kamus yang digunakan memperhatikan keberadaan *domain* sehingga hasil normalisasi teks lebih menyesuaikan pada kata yang terikat pada *domain*. Berikut adalah contoh hasil normalisasi teks pada skenario 2.

Tabel 6 Contoh hasil normalisasi pada Skenario 2

Input	Output
bener2 hg bgt! aku beli ini kebetulan pas s*ci*Ila disc 100rb, langsung lah cobain sekalian yg 250ml karna kebetulan makeup remover m*yb*Ilin*e aku habis pas aku coba, bener2 ga lengket dan licin2 aneh spt merek lain. pas setelah pakai malah wajah berasa jadi moist gituu terus a little goes a long way! duhh pokoknya suka banget deh, gabakal beli merek yg kmrn aku pakai lagi lol sayang, harganya termasuk mahal sih untuk makeup remover, but at least sebanding dengan performance nyaa	bener produk yang selalu digunakan di rutinitas kecantikan banget aku beli kebetulan pas scilla disc rb langsung lah cobain sekalian yang ml karna kebetulan riasan remover myblline aku habis pas aku coba bener tidak lengket licin aneh spt merek pas pakai malah wajah berasa jadi lembab gituu terus a little goes a long way duhh pokoknya suka banget deh gabakal beli merek yang kmrn aku pakai lol sayang harganya termasuk mahal sih riasan remover but at least sebanding performance nyaa
seneng banget akhir nya bedak dingin bisa gampang dibeli,, aku lagi rajinin pakai bedak dingin ini,, seminggu 2x semoga muka makin kinclong ya,, ovale ini di aku juga lumayan efek buat keringin bruntusan,, jadi kadang aku totolin ke bruntusan,,	seneng banget akhir nya bedak dingin gampang dibeli aku rajinin pakai bedak dingin seminggu x semoga muka makin bersinar ovale aku lumayan efek buat keringin bintik halus pada kulit jadi kadang aku totolin bintik halus pada kulit
katanya si ini dupe nya glamglow. langsung aja aku beli di olshop langganan ku. teksturnya kental bgt, bau nya mint kaya pepsodent. pas di pake agak susah di ratain sih, pas udah di cuci bau nya masih aja perasaan dan emang ngecilin pori, ngempesein jerawat juga	katanya si duplikat nya glamglow langsung aja aku beli olshop langganan ku teksturnya kental banget bau nya mint kaya pepsodent pas pakai susah ratain sih pas udah cuci bau nya aja perasaan emang mengecilkan pori ngempesein jerawat

D. Pengujian Pada Skenario 3

Tabel 7 Hasil pengujian pada Skenario 3

Kamus yang digunakan	Kondisi	Akurasi
Internet slang yang bersifat umum	Sebelum dilakukan Normalisasi Teks	47.69%
	Sesudah dilakukan Normalisasi Teks	47.03%

Dari hasil pengujian di atas, akurasi yang diperoleh pada kondisi sebelum normalisasi teks sebesar 47.69% sedangkan akurasi pada kondisi setelah normalisasi teks sebesar 47.03%. Dari kedua kondisi tersebut terjadi perbedaan sebesar 0.66%. Pada pengujian skenario 3, kamus yang digunakan hanya memperhatikan *domain* yang bersifat umum sehingga hasil normalisasi teks menyesuaikan dengan *domain* yang digunakan. Berikut adalah contoh hasil normalisasi teks pada skenario 3.

Tabel 8 Contoh hasil normalisasi pada Skenario 3

Input	Output
bener2 hg bgt! aku beli ini kebetulan pas s*ci*Ila disc 100rb, langsung lah cobain sekalian yg 250ml karna kebetulan makeup remover m*yb*Ilin*e aku habis pas aku coba, bener2 ga lengket dan licin2 aneh spt merek lain. pas setelah pakai malah wajah berasa jadi moist gituu terus a little goes a long way! duhh pokoknya suka banget deh, gabakal beli merek yg kmrn aku pakai lagi lol sayang, harganya termasuk mahal sih untuk makeup remover, but at least sebanding dengan performance nyaa	bener hg banget aku beli kebetulan pas scilla disc rb langsung lah cobain sekalian yang ml karna kebetulan makeup remover myblline aku habis pas aku coba bener tidak lengket licin aneh spt merek pas pakai malah wajah berasa jadi moist gituu terus a little goes a long way duhh pokoknya suka banget deh gabakal beli merek yang kmrn aku pakai lol sayang harganya termasuk mahal sih makeup remover but at least sebanding performance nyaa
my second tube and still loving it so.. menurut aku cc cream ini teksturnya ringan, tapi coverage nya juga lumayan.. cocok banget buat dipakai sebagai daily makeup. cocok banget sama wajahku yang oily, jadinya matte dan tahan sampai sore pas pulang kantor.. suka suka sukaa!!!	my second tube dan still loving it so aku cc cream teksturnya ringan coverage nya lumayan cocok banget buat dipakai daily makeup cocok banget sama wajahku oily jadinya matte tahan sore pas pulang kantor suka suka sukaa
katanya si ini dupe nya glamglow. langsung aja aku beli di olshop langganan ku. teksturnya kental bgt, bau nya mint kaya pepsodent. pas di pake agak susah di ratain sih, pas udah di cuci bau nya masih aja perasaan dan emang ngecilin pori, ngempesein jerawat juga	katanya si duplikat nya glamglow langsung aja aku beli olshop langganan ku teksturnya kental bgt bau nya mint kaya pepsodent pas pake susah ratain sih pas udah cuci bau nya aja perasaan emang ngecilin pori ngempesein jerawat

E. Analisis Hasil Normalisasi

Berdasarkan hasil eksperimen dengan tiga skenario yang digunakan, skenario pertama memiliki hasil akurasi klasifikasi yang paling baik diantara dua skenario lainnya. Pada skenario pertama, hasil akurasi sebelum normalisasi memiliki nilai sebesar 47.69% dan hasil akurasi sesudah normalisasi memiliki nilai sebesar 47.15%. Hal ini menunjukkan terdapat penurunan akurasi sebesar 0.54% antara proses sebelum normalisasi dan sesudah normalisasi.

Proses normalisasi teks memiliki pengaruh yang cukup baik pada proses klasifikasi. Normalisasi yang dilakukan

pada penelitian terhadap tool untuk normalisasi teks pada kasus klasifikasi opini terhadap teks yang dibuat oleh *user* (*user-generated content*) menunjukkan bahwa dengan adanya normalisasi teks sebelum proses klasifikasi opini, mampu meningkatkan performansi sebanyak 2% dibandingkan dengan tanpa adanya normalisasi teks [2]. Namun hal ini tidak dihasilkan pada penelitian ini. Skenario pertama yang dilakukan pada penelitian ini menunjukkan angka penurunan akurasi. Penurunan akurasi ini disebabkan karena proses normalisasi pada data ulasan masih belum tepat. Penggunaan tanda baca yang berulang (*emoticon*) berpengaruh pada proses ini, pengulangan huruf untuk menunjukkan emosi pengguna juga mempengaruhi penurunan akurasi normalisasi teks [5].

Normalisasi kalimat tidak hanya memperhatikan tanda baca dan pengulangan huruf, akan tetapi, ada beberapa hal seperti akhiran *-x* menjadi *-nya* juga perlu diperhatikan. Sehingga keakuratan normalisasi menjadi meningkat. Penggunaan tanda baca (*emoticon*) sebaiknya dihilangkan atau diabaikan, jadi normalisasi teks berfokus pada isi atau konteks ulasan yang diberikan. Kata yang seharusnya masuk ke sentiment negatif dapat berubah menjadi positif karena penggunaan emoticon [6].

Selain itu, terdapat pertimbangan dari *domain* yang digunakan pada kamus. Pada kamus yang digunakan saat pengujian skenario pertama secara keseluruhan berisi *internet slang* dari *domain* kosmetik maupun *internet slang* yang bersifat umum. Kemudian, *internet slang* yang terdapat pada data uji diubah sesuai dengan kata yang ada di dalam kamus. Hal ini mengakibatkan perbedaan pada sentimen, sehingga berpengaruh terhadap kesalahan klasifikasi pada beberapa teks ulasan.

V. KESIMPULAN

Berdasarkan hasil eksperimen diatas, kesimpulan yang dapat diambil yaitu, Normalisasi teks yang dilakukan sebelum proses klasifikasi teks pada teks ulasan produk kecantikan berpengaruh terhadap proses klasifikasi teks. Pada Tabel 4 dijelaskan bahwa dengan menggunakan kamus *internet slang* keseluruhan (*domain* kosmetik dan umum), kondisi sebelum dilakukan normalisasi teks memiliki akurasi sebesar 47.69% dan kondisi setelah dilakukan normalisasi teks memiliki akurasi sebesar 47.15%. Penurunan akurasi yang dihasilkan cukup besar yaitu 0.54%. Proses normalisasi teks yang belum sesuai dan banyaknya *internet slang* yang dimasukkan ke dalam kamus dengan kondisi melihat kedua *domain* memiliki pengaruh yang cukup besar pada sistem. Semakin banyak jumlah *internet slang* yang masuk ke dalam kamus, semakin bagus akurasinya.

Saran untuk penelitian selanjutnya yaitu memperbanyak kamus *internet slang* yang dapat berpengaruh pada hasil normalisasi teks, sehingga dapat meningkatkan akurasi dari sistem yang dibangun. Selain itu, fungsi normalisasi teks pada program dapat dimodifikasi sehingga membantu meningkatkan akurasi pada sistem.

REFERENSI

- [1] Bilal Syahid. (2021, Juni) gurupendidikan.co.id. [Online]. <https://www.gurupendidikan.co.id/teks-ulasan/>
- [2] Lucas V. Avanco, Sandra M. Aluisio, Thiago A.S. Pardo, Maria G.V. Nunes Magali S. Duran, "Some Issues on the Normalization of A Corpus of Product Reviews in Portuguese," p. 6, April 2014.
- [3] James H. Martin Daniel Jurafsky, *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Third Edition draft*, 3rd ed., 2020.
- [4] Sudi Suryadi, "IMPLEMENTASI NORMALISASI DALAM PERANCANGAN DATABASE RELATIONAL," *U-NET : Jurnal Teknik Informatika*, vol. 03, p. 5, Agustus 2019.
- [5] Teguh Bharata Adji, Adhitya Erna Purnamasari Ghulam Asrofi Buntoro, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *CITEE 2014*, p. 5, Oktober 2014.
- [6] Lucas Avanco, M. Gracas Volpe Nunes Magali Sanches Duran, "A Normalizer for UCG in Brazilian Portuguese," p. 31, Juli 2015.
- [7] Jing Jiang Swapna Gottipati, "Extracting and Normalizing Entity-Actions from User Comments," *COLLING*, p. 9, December 2012.
- [8] Maria das Grascas Vulpe Nunes Thales Felipe Costa Bertagila, "Exploring Word Embeddings for Unsupervised Textual User-Generated Content Normalization," in *The COLING 2016 Organizing Committee*, Osaka, 2016, p. 5.
- [9] Paloma Moreda Alejandro Mosquera, "Improving Web 2.0 Opinion Mining Systems Using Text Normalization Techniques," in *Proceedings of Recent Advances in Natural Language Processing*, 2013, p. 4.
- [10] Lucas V. Avanco, Pedro P. Balage, Magali S. Duran, Maria G. V. Nunes, Thiago A. S. Pardo, Sandra M. Aluisio Nathan S. Hartmann, "A Large Corpus of Product Reviews in Portuguese: Tackling Out-Of-Vocabulary Words," in *Ninth International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, 2014, p. 7.