TEKNOLOGI WEB CRAWLER SEBAGAI ALAT PENGEMBANGAN MARKET SEGMENTASI UNTUK MENCAPAI KEUNGGULAN BERSAING PADA E-MARKETPLACE

Ade Surahman[1], A. Ferico Octaviansyah[2], Dedi Darwis[3]
[1,2,3] Program Studi Informatika, Universitas Teknokrat Indoneisa,
Jl. ZA. Pagar Alam No.9-10, Labuhan Ratu, Kec. Kedaton, Bandar Lampung, Lampung 35132
E-mail: [1]adesurahman@teknokrat.ac.id, [2]fericopasaribu@teknokrat.ac.id,
[3]darwisdedi@teknokrat.ac.id

ABSTRAK

Berdasarkan situs wearesocial.com menunjukan data perkembangan pengguna internet sebesar 4.437 miliar yang disertai 75 % pembelian produk online, 82 % pencarian produk dan kunjungan on-line retail 92% yang kemudian memicu sebuah penelitian oleh asosiasi yang mengungkapkan bahwa jika sebuah produk tetap online untuk waktu lama (lebih dari 550 hari) maka 78 % sangat mungkin tidak akan dibeli, seharusnya aktivitas peletakan produk di online shop harus tersegmentasi secara tepat, karena segmentasi produk yang baik akan meningkatkan potensi e-commerce atau e-marketplace pada tingkat negara seperti meningkatkan efisiensi pasar, efesiensi operational, memperluas akses terhadap pasar, dan adanya keterkaitan, selanjutnya dari tren tersebut maka peneliti membuat suatu teknologi script otomatis atau program untuk menulusuri situs-situs e-marketplace indonesia, teknologi ini dikenal dengan nama web crawler atau web spider, dan peneliti mengimplementasikan sebuah prototipe teknologi yang dapat menganalisis teks – teks dari produk yang diletakan di e-marketplace Indonesia, dan menghasilkan clustering segmentasi pasar sebagai solusi yang efektif untuk pelaku bisnis yang akan meletakan produknya di e-commerce maupun di e-marketplace, sehingga akan mendapatkan keunggulan bersaing guna mencapai Store Image yang baik bagi pelaku bisnis di indonesia.

Kata kunci*—Web Crawler, Market Segmentasi, E-Marketplace, Keunggulan Bersaing, Store Image*

ABSTRACT

Based on the site wearesocial.com shows data on the development of internet users amounting to 4,437 billion accompanied by 75% of purchases of online products, 82% of product searches and 92% on-line retail visits which then triggered a study by an association that revealed that a product remained online for time long time (more than 550 days), 78% is very unlikely to be purchased, the product laying activity in the online shop should be segmented appropriately, because good product segmentation will increase the potential of e-commerce or e-marketplace at the country level such as increasing market efficiency , operational efficiency, expanding access to the market, and linkages, then from that trend, the researcher makes an automated script technology or program to track Indonesian e-marketplace sites, this technology is known as web crawlers or web spiders, and researchers implement a technology prototype that can analysis of texts from products placed in Indonesian e-marketplace, and produce clustering market segmentation as an effective solution for business people who will put their products in e-commerce and in e-marketplace, so that they will gain competitive advantages to reach the Store Image good for business people in Indonesia.

Keywords— *Web Crawler, Market Segmentation, E-Marketplace, Competitive Advantage, Store Image*

1. **PENDAHULUAN**

Web Crawler adalah sebuah program atau *code* otomatis yang dapat mengunjungi dan mencari informasi yang sesuai dengan kata kunci dalam sebuah website, *web crawler* memiliki nama lain diantaranya *web spiders*, *crawler*, *web robot*, dan *web walkers*. *Web crawler* melakukan penjelajahan dan pengambilan halaman website yang berada di internet dengan melakukan proses indeks dalam pencariannya, dan pencarian berdasarkan *keyword* adalah fungsi utama *web crawler* [1] [2]. *Web Crawler* yang di implementasikan oleh peneliti digunakan untuk mengumpulkan data berupa teks – teks yang berada di *e-marketplace* untuk di ambil informasinya, informasi ini berupa judul, harga, deskripsi, gambar dari produk dan komunikasi ulasan produk yang dijual di *e-marketplace* Indonesia yaitu Bukalapak.com dan Elevenia.co.id.

Informasi berupa judul, harga, deskripsi, gambar dari produk dan komunikasi ulasan produk yang dijual di *e-marketplace* merupakan sumber data yang akan diambil, sebagaimana diketahui bahwa komunikasi adalah salah satu aktivitas manusia untuk pengalihan suatu informasi dari manusia 1 (satu) ke manusia lainnya agar dipahami, ketika manusia berkomunikasi mereka melakukan dengan banyak cara [3]. Proses pengambilan dilakukan dengan *web crawler* yaitu dengan menggunakan *library simple php dom* dengan memodifikasi *taq* yang disesuaikan dengan struktur html dari situs *e-marketplace*, dimana pembeli dan penjual melakukan transaksi penjualan dan berkomunikasi tentang harga maupun penawaran produk secara *online* [4]. Data dari *e-marketplace* berupa teks tersebut akan menghasikan pengelompokan data yang memiliki karakteristik dari produk yang dijual, dan karakteristik tersebut sesuai dengan kategori dalam segmentasi pasar, proses segmentasi dilakukan dengan cara melakukan pengelompokan berdasarkan *geographics* dari hasil *crawling* data. Segmentasi pasar sendiri memiliki pokok pembahasan yaitu pembagian jenis pelanggan, potensi pelanggan, kelompok pelanggan yang berbeda-beda seperti produk dan jasa, demografi, *geographics*, dan *psychographics* [5].

Segmentasi menghasilkan suatu upaya menciptakan pengetahuan berupa gambaran umum tentang perusahaan baik tempat maupun produk yang dimiliki sehingga persepsi masyarakat umum menjadi positif terhadap perusahaan tersebut yang kemudian disebut *store image*. Secara operasional, *store image* umumnya dinilai dengan meminta konsumen memberikan *feedback* baik atau buruk terhadap kemampuan *store* dalam hal penjualan produk kepada pelanggannya. Kategori *Store image* seperti barang yang didagangkan, layanan, fasilitas yang ada seperti fisik fasilitas, promosi dari produk, dan keadaan toko yang terasa nyaman [6]. *Store image* ini yang akan menjadi keunggulan bersaing dimana kondisi toko atau perusahaan menjadi pilihan konsumen yang ada dipasar dikarenakan perusahaan memberikan nilai yang lebih menguntungkan untuk konsumennya [7], keunggulan bersaing merupakan cakupan aktivitas yang berusaha dicapai suatu badan usaha guna menghasilkan tiga strategi bersaing yang berguna menciptakan posisi aman bagi perusahaan diantaranya pertama *cost leadership* yang mengutamakan menjadi produsen berbiaya cukup rendah dalam proses bisnisnya, kedua *differentiation* yang dalam bisnisnya menjadi perusahan yang unik atau berbeda degan pesaingnya, dan ketiga *focus* yang mengutamakan pelayanan hanya konsentrasi pada kelompok tentetu dengan baik, sehingga pebisnis di *e-marketplace* lebih dapat memenuhi keinginan konsumen yang difokuskan.

Berdasarkan kajian teoritis mengenai *web crawler*, dan tujuan penelitian untuk menghasilan *cluster* segmentasi pasar yang digunakan oleh pelaku bisnis di *e-marketplace*, diharpkan nantinya pelaku bisnis dapat meletakan produknya sesuai karakteristik segmentasi pasar, karena berdasarkan suatu penelitian oleh asosiasi yang mengungkapkan bahwa jika sebuah produk tetap online untuk waktu lama (lebih dari 550 hari) maka 78 % sangat mungkin

tidak akan dibeli [8], seharusnya aktivitas peletakan produknya di *online shop* harus tersegmentasi secara tepat, dengan adanya *cluster* segmentasi pasara dan dapat diterapkan pelaku bisnis di *e-marketplace* maka diharapkan pelaku bisnis akan mendapatkan *store image* yang baik dan menjadi keunggulan dalam bersaing.

2. METODE PENELITIAN

Penelitian tentang penerapan Teknologi *Web Crawler* sebagai alat pengembangan segmentasi pasar untuk mencapai keunggulan bersaing pada *e-marketplace* memiliki gambaran penelitian berupa:

1. Bukalapak.com dan Elevenia.co.id adalah situs *e-marketplace* sebagai uji coba objek penelitian yang akan di analisis teks-teksnya.
2. Proses extraksi dan analisis teks – teks dari situs *e-marketplace*, hanya berkaitan dengan pengambilan data produk, data toko penjual dan data-data pendukung untuk penyajian data segmentasi pasar.
3. Penyajian data yang dihasilkan dari analisis teks – teks yang dilakukan, disajikan dalam bentuk grafik berdasarkan segmentasi pasar yang diinginkan.
4. Pada *Prototipe* ini tidak terdapat media atau perantara yang digunakan untuk pengujian tingkat ketahanan dari mesin *web crawler*.
5. Pembuatan prototipe ini di buat dengan cara memodifikasi *library php dom* untuk proses *crawling* dengan cara melengkapi kriteria *crawling*, hasil *crawling* yang disimpan dalam array 1 dimensi yang kemudian hasil *crawling* di sajikan menjadi segmentasi pasar.
6. Pemanfaatan *Prototipe* ini hanya untuk digunakan pada klasifikasi segmentasi pasar yang untuk bukalapak.com dan elevenia.co.id.
7. Metode Pengembangan sistem yang digunakan pada penelitian ini yaitu model *prototipe evolusioner* dan untuk pengujian hasil *prototipe* menggunakan model kesuksesan sistem informasi yaitu Model DeLone dan Mclean.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Proses Bisnis *E-Maketplace*

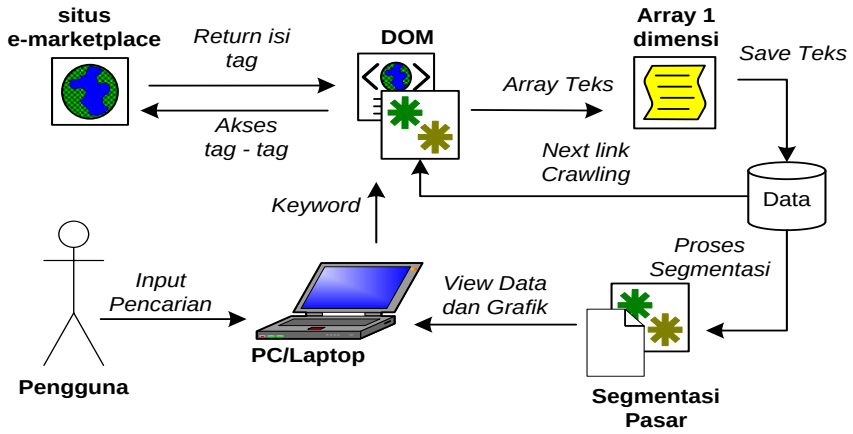
Proses Bisnis E-marketplace seperti situs Bukalapak.com dan Elevenia.com yaitu Model C2C (*Customer to Customer*), dimana transaksi berupa rekening bersama dijadikan fasilitas antara pemilik toko di *e-marketplace* dengan pelanggan yang membeli produk. Produk -produk yang dijual di *e-marketplace* seperti barang, jasa maupun informasi dilakukan secara elektronik atau *online shop*, pebisnis di e-marketplace dapat mendaftar dan membuka toko online sesuai dengan ketentuan penyedia layanan e-marketplace seperti bukalapak.com. Jaminan transaksi elektronik diatur oleh organisasi penyedia layanan e-marketplace dan dijamin aman [9]. Dengan adanya transaksi bisnis secara elektronik di *e-marketplace* yang terhubung dengan pemilik toko dan pelanggan yang membeli diperlukan jaringan internet untuk komunikasi, seperti web browser atau aplikasi *mobile* [10].

3.2 Arsitektur *Prototipe*

Aristektur dari *prototipe* yang dibuat adalah menggunakan *library* php *DOM (document object model)* dimana *document object model* adalah antarmuka *platform* dan bahasa netral yang akan memungkinkan *program* dan *script* untuk mengakses dan memperbarui konten, struktur dan gaya dokumen dinamis. dokumen dapat diproses lebih lanjut dan hasil pengolahan yang

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

dapat dimasukkan kembali ke dalam halaman yang disajikan. Ini adalah gambaran dari *libarary DOM* untuk proses *crawling* isi dari tag – tag html berdasarkan *keyword* dari pencarian yang diinputkan, proses *crawling* dilakukan dari 2 (dua) situs *e*-marketplace yaitu bukalapak.com dan elevenia.com, kemudian isi *tag* - *tag* di simpan sementara kedalam sebuah *array* teks yang kemudian nantinya teks –teks tersebut di kategorikan sesuai hubungan segmentasi pasar. Pada Gambar 6 memperlihatkan arsitektur *prototipe* yang dibuat dimana proses penginputan dilakukan oleh pengguna menggunakan pc/laptop, kemudian *keyword* atau kata kunci yang diinputkan di proses ke *library* tag –tag dengan kriteria yang peneliti inginkan yaitu kriteria produk – produk yang ada hubungannya dengan *keyword* dari teks yang diinputkan oleh pengguna, setelah *keyword* di proses oleh *library* DOM maka proses selanjutnya adalah mengakses tag – tag yang ada di situs *e-marketplace* yang hasilnya berupa isi dari tag – tag tersebut, kemudian menghasilkan teks – teks yang disimpan kedalam data sementara berupa *array* yang kemudian di proses untuk dilakukan pemilihan kriteria yang disimpan kedalam *datastore*, jika terdapat *next link* dari situs *e-marketplace* maka akan dilakukan proses sampai analisis teks – teks dari tag di situs e-*marketplace* selesai, setelah proses *save datastore* selesai kemudian akan dilakukan segmentasi berdasarkan kriteria yang dipilih oleh user.



Gambar 1 Arsitektur Prototipe

3.3 Analisis Kebutuhan Sistem

3.3.1 Fungsional

Fungsional dari *prototipe* yang dibuat sebagai berikut :

a. *User Friendly*, *Prototipe* harus dapat memudahkan *user* dalam penginputan produk yang ingin di cari, baik dari segi *user interface* maupun kemudahan dalam menggunakan *prototipe*.
b. *Prototipe* harus dapat menampilkan hasil pencarian yang diinput *user* sebelumnya pada *user interface* pencarian.
c. *Prototipe* harus mampu menyajikan grafik dari *clustering* segmentasi pasar yang dipilih oleh *user*.
d. *Prototipe* harus dapat melakukan pencarian ulang ketika *user* ingin mencari produk yang berbeda.

3.3.2 *Non* Fungsional

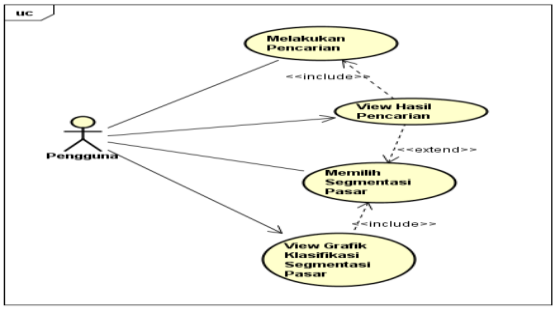
Kebutuhan non fungsional dalam membuat dan menjalankan prototipe sebagai berikut :

a. Kebutuhan *Software*
   a) Bahasa Pemrogram PHP
   b) Sistem Operasi Windows 10
   c) Database Mysql

b. Kebutuhan *Hardware*
   a) PC dengan sepesifikasi Procecor minimal Core i3
   b) Ram Minimal 2 GB
   c) Harddisk SATA 512 GB
   d) Akses Internet dengan Jaringan minimal 3G/4G

3.4 Rancangan Fungsional Sistem

Secara fungsional *Prototipe* diilustrasikan berupa *use case diagram* pada gambar 2.

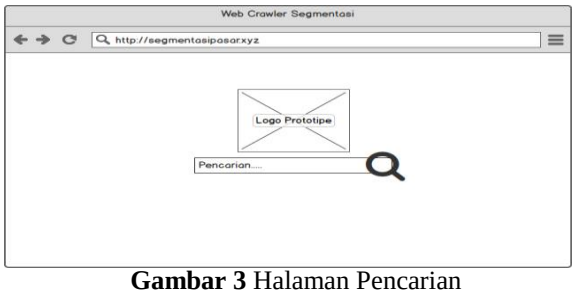


Gambar 2 Arsitektur Prototipe

3.5 *User Interface*

3.5.1 Halaman Pencarian

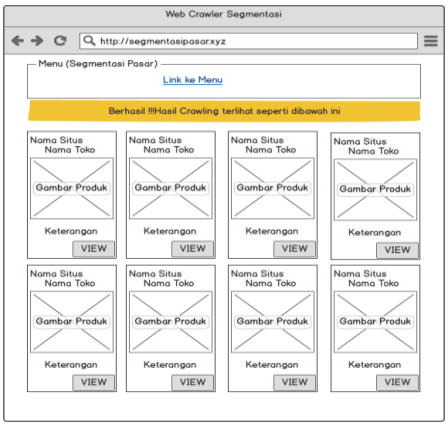
Halaman pencarian digunakan untuk menginputkan *keyword* produk yang diinginkan. Adapun tampilannya pada gambar 3.



Gambar 3 Halaman Pencarian

3.5.2 Halaman Hasil Pencarian

Halaman hasil pencarian yang telah dilakukan dan telah di *crawl* berdasarkan *keyword* produk. Adapun halaman hasil pencarian terlihat pada gambar 4.

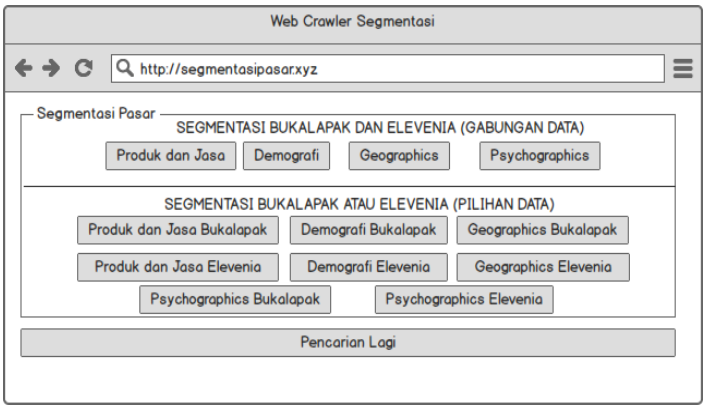


ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Gambar 4 Halaman Hasil Pencarian

3.5.3 Halaman Menu

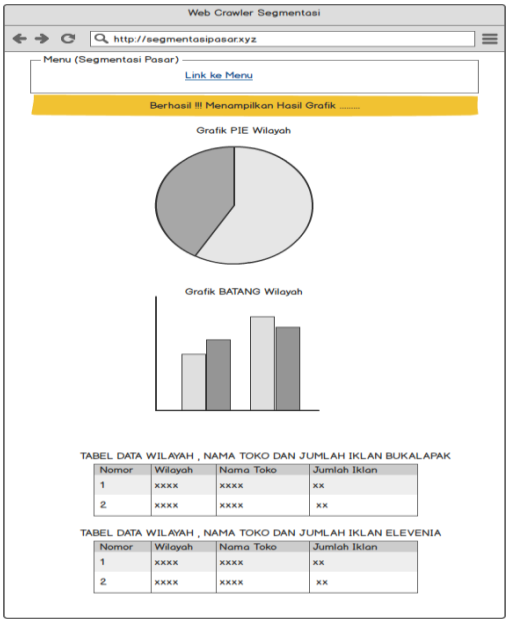
Halaman Menu adalah halaman yang digunakan mengakses halaman – halaman segmentasi pasar, halaman menu ini terlihat setelah *user* menginputkan *keyword* di kolom pencarian dan menampilkan hasilnya. Adapun tampilan halaman Menu terlihat pada gambar 5.



Gambar 5 Halaman Menu

3.5.4 Halaman *Cluster* Segmentasi Pasar

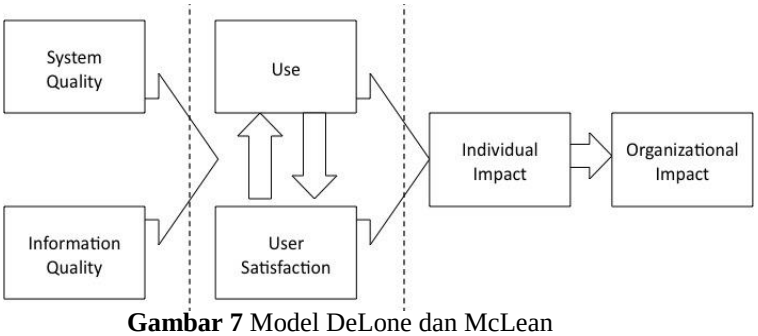
Halaman *clustering* segmentasi pasar digunakan untuk menyajikan data berupa grafik sesuai pilihan dari pengguna, penyajian grafik memiliki 2 kategori yaitu grafik *pie* dan grafik batang. Adapun tampilan halaman pencarian terlihat pada gambar 6.



Gambar 6 Halaman *Clustering* Segmentasi Pasar

3.6 Pengujian

Pengukuran Kesuksesan sistem informasi telah banyak di teliti dan model DeLone dan McLean yang menyebutkan bahwa *information quality*, *system quality dan service quality* akan berpengaruh positif pada *use* dan *user satisfaction* dan selanjutnya akan berpengaruh positif pada *net benefit* atau hasil akhir [11] [12].



Gambar 7 Model DeLone dan McLean

Model ini dapat dilakukan untuk mengukur kesuksesan sistem informasi dari sebuah sistem informasi yang dihasilkan seperti yang dilakukan oleh penelitian terdahulu [13] [14] [15], model ini akan dipadukan dengan *kuesioner* dan hasil dari kuesioner akan dibandingkan dengan kriteria tanggapan responden. Dimana pengujian prototipe dilakukan dengan pengujian dengan analisis deskriptif, analisis ini digunakan untuk menggambarkan fakta-fakta yang ada untuk selanjutnya diolah menjadi data. Data yang dianalisis diolah untuk mendapatkan sebuah kesimpulan [16]. Langkah pengujian yang dilakukan sebagai berikut.

1. Setiap indikator yang dinilai oleh responden, diklasifikasikan dalam lima alternatif jawaban dengan menggunakan skala ordinal yang menggambarkan peringkat jawaban.
2. Dihitung total skor setiap valiabel / subvariabel = jumlah skor dari seluruh indikator variabel untuk semua responden.
3. Dihitung skor setiap variabel / subvariabel = rata – rata dari total skor.
4. Untuk mendeskripsikan jawaban responden, juga digunakan statistik deskriptif seperti distribusi frekuensi dan tampilan dalam bentuk tabel atau grafik
5. Untuk menjawab deskripsi tentang variabel penelitian ini, digunakan rentang kriteria penelitian sebagai berikut.

$$Skor\ Total = \frac{skor\ Aktual}{skor\ Ideal}\ x\ 100\ \%$$

Skor aktual adalah jawaban seluruh responden atas kuesioner yang telah diajukan. Skor ideal adalah skor atau bobot tertinggi atau semua responden diasumsikan memilih jawaban dengan skor tertinggi. Penjelasan bobot nilai skor aktual dapat dilihat pada tabel 1.

Tabel 1 Kriteria Tanggapan Responden

% Jumlah Skor	Kriteria
20,00 % - 36,00 %	Tidak Baik
36,01 % - 52,00 %	Kurang Baik
52,01 % - 68,00 %	Cukup
68,01 % - 84,00 %	Baik
84,01 % - 100 %	Sangat Baik

Model DeLone dan McLean digunakan sebagai model pengujian kesuksesan informasi yang dihasilkan dari prototipe degan hasil berdasarkan 5 kriteria.

Tabel 2 Hasil Pengujian *Prototipe*

No	Pengujian dengan Model DeLone dan McLean			
	Kriteria	***Skor Aktual***	***Skor Ideal***	***% Skor Aktual***
1.	Kualitas Informasi	4988	6250	80%
2.	Kualitas Sistem	4961	6250	79%
3.	Kualitas Layanan	2906	3750	77%
4.	Penggunaan	1984	2500	79%
5.	Kepuasan Pengguna	1946	2500	78%
Total		16785	21250	79%

Hasil Pengujian dengan jumlah responden sebanyak 250 responden, dan menggunakan 5 kriteria dari Model DeLone dan McLean maka didapat hasil dari kriteria Kualitas Informasi (*Information Quality*) sebesar **80 %,** Kualitas Informasi (*Information Quality*) sebesar **79 %,** Kualitas Layanan (*Service Quality*) sebesar **77 %,** Penggunaan (*Use)* **sebesar 79 %**, Kepuasan Pengguna (*User Satisfaction*) sebesar **78 %**. Dari keseluruhan kriteria Model DeLone dan McLean untuk kesuksesan informasi segmentasi pasar sebesar **79 %**.

Dari data diatas membuktikan bahwa penelitian ini memberikan solusi dalam upaya meningkatkan segmentasi pasar bagi pelaku bisnis dan bagi pemulai bisnis dalam berbisnis di *e-marketplace* hasilnya dikategorikan **Baik** sebesar **79 %**.

4. KESIMPULAN

Berdasarkan hasil kajian teori dan pengujian yang dilakukan pada prototipe, maka kesimpulannya sebagai berikut :

1. Penggunaan prototipe ini menjadi solusi yang baik untuk memulai bisnis dan pelaku bisnis untuk mengetahui segmentasi pasar di situs bukalapak.com dan elevenia.co.id untuk meletakan produk atau membuka gudang secara konvensional sesuai dengan hasil segmentasi pasar yang disajikan oleh *prototipe*.
2. Dari hasil pengujian menggunakan Model DeLone dan McLean yang dilakukan oleh peneliti, didapat bahwa kesuksesan informasi yang disajikan dalam bentuk segmentasi pasar memiliki kategori **Baik**.
3. Berdasarkan kuesioner yang disebarkan kepada 250 responden dengan 40,4 % wanita dan 59,6 % pria yang dibedakan berdasarkan pekerjaan yaitu wiraswasta 9 responden, pesagwai swasta 19 responden, mahasiswa 210 dan lainnya 12 responden yang mengisi kuesioner berdasarkan 5 kriteria kesuksesan sistem informasi dari Model Delone dan McLean dengan masing – masing nilai kriteria % skor aktual yaitu: Kualitas Informasi (*Information Quality*) sebesar 80 %, Kualitas Sistem (*System Quality*) sebesar 79 %, Kualitas Layanan (*Service Quality*) sebesar 77 %, Penggunaan (*Use*) sebesar 79 %, Kepuasan Pengguna (*User Satisfaction*) sebesar 78 % dengan kesimpulan pengujian kualitas informasi segmentasi pasar sebesar **79 %** dan nilai tersebut memiliki *presentase* tanggapan responden dalam kriteria **Baik**.

DAFTAR PUSTAKA

[1] M. Schrenk, 2012, *Webbots Spiders and Screen Scrapers a Guide to Developing Internet Agents with PHP/CURL*, No Starch Pres, San Francisco.
[2] A. V. Singh dan A. Mishra, 2014, A Review of Web Crawler Algorithms, *IJCSIT*, Vol. 5(5) 6689–6691.
[3] J. Fiske, 1990, *Introduction to Communication Studies, Second edition*. West 35th Street, New York: Taylor & Francis Group.
[4] M. H. Pattinson, dan D. R. Low, 2011, *E-Novation for Competitive Advantage in Collaborative Globalization: Technologies for Emerging E-Business Strategies*. United States of America: IGI Global.
[5] M. McDonald, dan I. Dunbar, 2012, *Market Segmentation: How to do it and how to profit from it*. Southern Gate, United Kingdom: John Wiley & Sons Ltd.
[6] J. P. Peter, dan J. C. Olson, 2010, *Consumer Behavior & Marketing*. Dana. New York: Paul Ducham.
[7] M. E. Porter, 1998, *Competitive Advantage Creating and Sustaining Superior Performance*. new york: Simon & Schuster Inc.
[8] E. Russ, 2001, E-marketplaces: New Challenges for Enterprise Policy, Competition and Standardization. eEurope go Digital. *Workshop Report*, Brussels, 23-24 April.
[9] AkashKalmegh, dkk, 2016, A Survey on Web Mining in E-commerce: Pattern Discovery, Issues and Applications, *IJIRCCE*, Vol. 4, Issue 1.
[10] E. Turban, D. King dan J. K. Lee, 2015, *Electronic Commerce A Managerial and Social Networks Perspective*. new york: Springer International Publishing Switzerland.
[11] W. H. DELONE and E. R. MCLEAN, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003.
[12] T. McGill, V. Hobbs, and J. Klobas, "User-developed applications and information systems success: A test of DeLone and McLean's model," *Inf. Resour. Manag. J.*, vol. 16, no. 1, pp. 24–45, 2003.
[13] P. H. Saputro, D. Budiyanto, and J. Santoso, "Model Delone and Mclean Untuk Mengukur Kesuksesan E-Government Kota Pekalongan," *Sci. J. Informatics*, vol. 2, no. 1, pp. 1–8, 2016.
[14] J. W. Lian, "Establishing a cloud computing success model for hospitals in Taiwan," *Inq. (United States)*, vol. 54, 2017.
[15] I. Nofikasari and A. Sunyoto, "Evaluasi Penerapan Sistem Informasi Akademik Dengan Pendekatan Model Delone & McLean, HOT FIT dan UTAUT (Studi Kasus: STMIK Duta Bangsa Surakarta)," *J. Tek. Inform.*, vol. 12, no. April, pp. 81–90, 2017.
[16] U. Narimawati, *Riset Manajemen Sumber Daya Manusia*. Jakarta: Agung Media, 2007.