



PENERAPAN ALGORITMA *K-MEDOIDS* DALAM PENENTUAN FAKTOR TERBESAR PEMILIHAN JURUSAN DI FAKULTAS TEKNIK UNIVERSITAS HALU OLEO

Saskia Randawula Silondae^{*1}, Sutardi², Statisfaty³

^{1,2,3} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo, Kendari

e-mail: ^{*1}saskia.randawula@gmail.com, ²statisfaty@uho.ac.id, ³sutardi_hapal@yahoo.com

Abstrak

Menentukan perguruan tinggi dan jurusan yang tepat bukanlah persoalan yang mudah. Kurangnya informasi tentang fasilitas fakultas atau jurusan, lingkungan fakultas, jaminan kerja dari fakultas, biaya kuliah, dan promosi dapat menyebabkan calon mahasiswa salah dalam memilih jurusan sehingga mahasiswa tidak dapat mengikuti perkuliahan dengan baik karena tidak tertarik di bidang pilihannya, serta tidak dapat menyelesaikan perkuliahan dengan baik bahkan berujung pada *Drop Out* (DO). Permasalahan tersebut dapat diantisipasi oleh fakultas atau program studi dengan mengetahui faktor-faktor yang menentukan mahasiswa baru memilih jurusannya. Penelitian ini mengimplementasikan metode *data mining* dengan algoritma *K-Medoids Clustering* untuk menentukan faktor terbesar sumber informasi pemilihan jurusan di Fakultas Teknik Universitas Halu Oleo. Dari hasil pengujian menunjukkan jumlah klaster yang terbaik adalah klaster 4 dengan nilai *silhouette coefficient* sebesar 0.23045321036704783 dan klaster yang paling buruk adalah klaster 3 dengan dengan nilai *silhouette coefficient* adalah 0.11000736495160417.

Kata kunci; *K-Medoids*, Sumber Informasi, Perguruan Tinggi, Klasterisasi

Abstract

Selecting the right college and department is not an easy matter. Lack of information about faculty or department facilities, faculty environment, job guarantees from the faculty, tuition fees, and promotions can cause prospective students to choose the wrong department so that students cannot attend lectures properly because they are not interested in their chosen field, and cannot complete lectures properly. even lead to Dropping Out (DO). The above problems can be anticipated by the faculty or study program by knowing the factors that determine new students choosing their major. The researcher implemented the data mining method with the K-Medoids Clustering algorithm to determine the biggest factor in the source of information for the selection of majors at the Faculty of Engineering, University of Halu Oleo. The test results show that the best number of clusters is cluster 4 with a silhouette coefficient value of 0.23045321036704783 and the worst cluster is cluster 3 with a silhouette coefficient value of 0.11000736495160417.

Keywords; *K-Medoids*, Information Sources, University, Clustering



1. PENDAHULUAN

Dalam perkembangannya, lembaga perguruan tinggi atau universitas merupakan salah satu tujuan masyarakat dalam meningkatkan pendidikan yang fokus pada satu bidang konsentrasi yang dibutuhkan untuk memasuki dunia kerja. Meningkatnya keinginan masyarakat untuk memiliki karier yang baik dalam ketatnya dunia persaingan, mendorong meningkatnya jumlah universitas yang tersebar di seluruh Indonesia. Saat ini banyak universitas baik negeri maupun swasta yang menawarkan berbagai jurusan atau program studi. Oleh karena itu, suatu universitas harus mampu bersaing dengan universitas lainnya, selain untuk mendapatkan keuntungan yang maksimal, juga agar nama dari universitas tersebut tidak tertinggal[1].

Menentukan perguruan tinggi dan jurusan yang tepat bukanlah persoalan yang mudah. Banyak faktor yang berpengaruh dalam menentukan pemilihan jurusan. Kurangnya informasi tentang fasilitas fakultas atau jurusan, lingkungan fakultas, jaminan kerja dari fakultas, biaya kuliah, dan promosi dapat menyebabkan calon mahasiswa salah dalam memilih jurusan sehingga mahasiswa tidak dapat mengikuti perkuliahan dengan baik karena tidak tertarik di bidang pilihannya, serta tidak dapat menyelesaikan perkuliahan dengan baik bahkan berujung pada *Drop Out* (DO).

Permasalahan tersebut dapat diantisipasi oleh fakultas atau program studi dengan mengetahui faktor-faktor yang menentukan mahasiswa baru memilih jurusannya. Dengan mengetahui faktor-faktor tersebut, ke depannya fakultas dapat mengambil langkah-langkah antisipasi agar calon mahasiswa dapat mengambil jurusan yang sesuai.

Berdasarkan uraian permasalahan tersebut, peneliti akan membuat sistem untuk menentukan faktor terbesar sumber informasi di Fakultas Teknik Universitas Halu Oleo, dengan mengelolah data tersebut menggunakan metode *data mining* dengan algoritma *K-Medoid clustering*.

2. METODE PENELITIAN

2.1 Minat

Widyastuti mendefinisikan minat adalah keinginan yang didorong oleh suatu keinginan setelah melihat, mengamati dan membandingkan serta mempertimbangkan dengan kebutuhan yang diinginkannya[2].

Berdasarkan pendapat tersebut, maka minat memilih program studi atau jurusan pada mahasiswa, dapat diartikan sebagai kecenderungan yang mengarahkan mahasiswa untuk memilih program studi yang ditandai dengan adanya perasaan senang terhadap program studi tersebut, perasaan tertarik dan perasaan bahwa program studi tersebut bersangkutan paut dengan kebutuhannya. Minat yang besar terhadap program studi yang telah mahasiswa pilih akan mempermudah terciptanya konsentrasi dalam studi. Jadi tanpa minat maka konsentrasi belajar juga akan sulit berkembang bahkan dipertahankan.

2.2. Skala likert

Skala likert adalah skala pengukuran yang dikembangkan oleh Likert tahun 1932. Menurut Preston dan Colman, skala likert mempunyai empat atau lebih butir-butir pertanyaan yang dikombinasikan sehingga membentuk sebuah skor atau nilai yang merepresentasikan sifat individu, misalkan pengetahuan, sikap, dan perilaku. Dalam proses analisis data, komposit skor, biasanya jumlah atau rataan, dari semua butir pertanyaan dapat digunakan. Jumlah titik 2, 3 dan 4 menghasilkan indeks reliabilitas, validitas, dan kekuatan diskriminasi yang jelek. Hasil ini berbeda nyata dibandingkan jumlah titik 5, 6, dan 7 yang mempunyai indeks reliabilitas, validitas, dan kekuatan diskriminasi lebih baik. Ukuran stabilitas (*test-retest* validitas) menunjukkan jumlah titik 1 dan 2 mempunyai stabilitas yang paling buruk dan semakin bagus saat jumlah titik meningkat menjadi 3, 3 atau 4 [3].

2.3 Clustering

Clustering atau klusterisasi adalah metode pengelompokan data. Menurut Tan, *clustering* adalah sebuah proses untuk mengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang

minimum[4]. *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang di dalam *cluster* memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan grup atau kelompok yang tidak dikenal dalam data. *Clustering* banyak digunakan dalam berbagai aplikasi seperti misalnya pada *business intelligence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). Di dalam *business intelligence*, *clustering* bisa mengatur banyak *customer* ke dalam banyaknya kelompok. Contohnya mengelompokkan pelanggan ke dalam beberapa *cluster* dengan kesamaan karakteristik yang kuat. *Clustering* juga dikenal sebagai data segmentasi karena *clustering* mempartisi banyak data set ke dalam banyak group berdasarkan kesamaannya. Selain itu *clustering* juga bisa sebagai *outlier detection*.

2.4 Partitioning Around Medoids (PAM)

Algoritma *Partitioning Around Medoids* (PAM) atau dikenal juga dengan *K-Medoids* adalah algoritma pengelompokan yang berkaitan dengan algoritma *K-Means* dan algoritma *Medoidshift*. Algoritma *K-Medoids* ini diusulkan pada tahun 1987. Algoritma *K-Medoid* dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma ini sangat mirip dengan algoritma *K-Means*, terutama karena kedua algoritma ini *partitional*. Dengan kata lain, kedua algoritma ini memecah *dataset* menjadi kelompok-kelompok dan kedua algoritma ini berusaha untuk meminimalkan kesalahan. Tetapi algoritma *K-Medoids* bekerja dengan menggunakan *Medoids*, yang merupakan entitas dari dataset yang mewakili kelompok dimana ia dimasukkan. *Medoids* adalah objek representatif dari kumpulan data atau *cluster* dengan kumpulan data yang ketidaksamaan rata-rata untuk semua objek dalam *cluster* minimal.

Algoritma *K-Medoids* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan n (jumlah) objek menjadi sejumlah *cluster*. Algoritma ini menggunakan objek pada kumpulan objek

untuk mewakili sebuah *cluster*. Objek yang terpilih untuk mewakili sebuah *cluster* disebut dengan *Medoid*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara *Medoid* dengan objek *non-Medoid*.

Algoritma dari *K-Medoids* adalah sebagai berikut [5]:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance*.
3. Pilih secara acak pilih objek pada masing-masing *cluster* sebagai kandidat *Medoids* baru
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *Medoids* baru.
5. Hitung total biaya simpangan (S), dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *Medoid*.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *Medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

Dalam mengukur jarak dalam *clustering* dapat dilakukan dengan menggunakan *Euclidean Distance*. *Euclidean Distance* merupakan pengukuran jarak objek dan pusat *cluster* yang banyak digunakan secara luas dalam berbagai kasus *pattern matching*, termasuk *clustering*. *Euclidean distance* dinyatakan dengan Persamaan 1.

$$dist = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (1)$$

Dimana :

n = jumlah fitur dalam suatu data.

k = indeks data.

p_k = nilai atribut (fitur) ke- k dari p .

q_k = nilai atribut (fitur) ke- k dari q .

Nilai S dinyatakan dengan Persamaan 2.

$$S = \text{Total cost baru} - \text{Total cost lama} \quad (2)$$

Dimana :

Total *cost* baru = jumlah biaya/*cost non-Medoids*.

Total *cost* baru = jumlah biaya/*cost Medoids*.

K-Medoids sangat mirip dengan *K-Means*. Perbedaan utama di antara dua algoritma tersebut adalah jika pada *K-Means cluster* diwakili dengan pusat dari *cluster*, sedangkan pada *K-Medoids cluster* diwakili oleh objek terdekat dari pusat *cluster*

2.5 Silhouette Coefficient (SC)

Silhouette Coefficient (SC) merupakan metode evaluasi yang digunakan untuk melihat kualitas dan kekuatan *cluster* dan kualitas suatu objek dalam suatu *cluster*[6]. Untuk menghitung *Silhouette Coefficient* terdapat beberapa tahapan yaitu:

- Menghitung rata-rata jarak setiap dokumen (disimbolkan dalam a_i) dengan setiap dokumen lain dalam *cluster* yang sama. Simpan nilai rata-rata pada suatu variabel. misalkan a_i .
- Menghitung rata-rata nilai jarak setiap dokumen suatu *cluster* (disimbolkan dalam b_i) dengan setiap dokumen pada *cluster* lain dan diambil nilai terkecil. Simpan nilai tersebut pada suatu variabel misalkan b_i .
- Hitung nilai *Silhouette Coefficient* dengan menggunakan Persamaan 3.

$$SC = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (3)$$

2.6 Principal Component Analysis (PCA)

Metode PCA merupakan metode yang digunakan untuk menyederhanakan suatu variabel yang akan diteliti dengan mereduksi dimensinya. Reduksi dimensi dilakukan dengan cara menghilangkan korelasi yang dimiliki antar variabel bebas dengan transformasi variabel asal menjadi variabel baru yang tidak berkorelasi sama sekali.

Menurut Margono dalam *Principal Component Analysis* (PCA) memiliki beberapa kriteria sebagai berikut:

- Communalities* sebagai faktor bersama.

Communalities merupakan peranan faktor untuk satu kesatuan yang berasal dari masing-masing variabel atau sub variabel yang menyusun satu faktor secara bersama. Hasil yang ditunjukkan dalam *communalities* merupakan nilai nital dan *extraction*. Nilai initial merupakan besarnya peranan variabel secara individu untuk membentuk faktor baru bersama. Dan nilai *extraction* merupakan

besarnya peranan variabel yang berupa persentase masing-masing variabel untuk membentuk faktor baru.

- Nilai akar karakteristik (*eigen values*)

Memiliki persamaan karakteristik $|M - \lambda I| = 0$, dengan $\lambda > 1$. *Screen plot* merupakan diagram yang menggambarkan kecenderungan penurunan *eigen value*. Sumbu Y pada diagram menunjukkan *eigen value* dan sumbu X menunjukkan jumlah faktor. Dalam menentukan banyaknya faktor yang terbentuk dapat dilihat pada grafik dengan kemiringan yang paling tajam yang terbentuk dari faktor satu ke faktor lainnya[7].

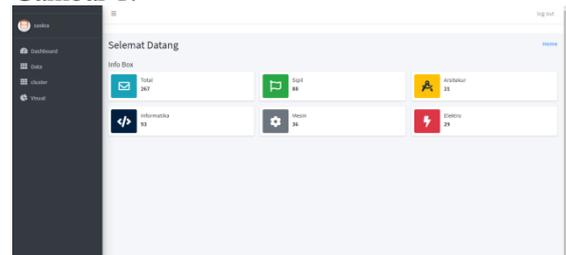
3. HASIL DAN PEMBAHASAN

Pada tahap implementasi merupakan tahap dimana hasil analisis dan perancangan diimplementasikan dalam bentuk Penentuan Faktor Terbesar Sumber Informasi Pemilihan Jurusan di Fakultas Teknik Universitas Halu Oleo menggunakan bahasa pemrograman Python yang siap untuk dioperasikan. Sistem ini menggunakan metode *Clustering K-Medoids*.

3.1 Tampilan Aplikasi

- Tampilan Dashboard

Pada halaman *dashboard* terdapat empat menu yaitu Menu Data Mahasiswa, Menu Hasil *Cluster*, dan Menu *Visual*. Tampilan halaman menu utama dapat dilihat pada Gambar 1.



Gambar 1. Tampilan Halaman Dashboard

- Tampilan data Mahasiswa

Halaman ini merupakan halaman yang menampilkan data mahasiswa yang dapat diolah oleh admin dari kuisioner diupload dalam bentuk csv.

Gambar 2. Tampilan Halaman Data Mahasiswa

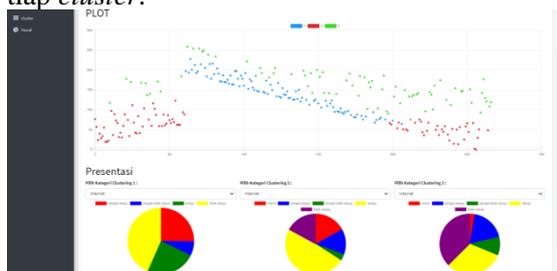
3. Menu Hasil Cluster

Halaman ini merupakan halaman yang menampilkan proses analisa pengelompokan data yang telah di *clustering* dan faktor dominan yang memprngaruhi pemilihan jurusan

Gambar 3. Tampilan Halaman Hasil Cluster

4. Tampilan Visualisasi

Halaman ini merupakan halaman yang menampilkan pengelompokan data menggunakan algoritma *K-Medoids*. pada halaman ini dapat dilihat persebaran dalam bentuk *plot* dan melihat perbandingan jumlah tiap *cluster*.



Gambar 4. Tampilan Halaman Visualisasi

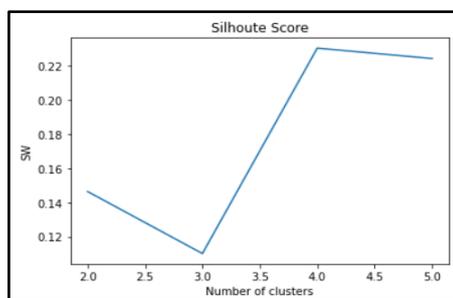
3.2 Hasil Uji Silhouette Coefficient

Pengujian dilakukan dengan mengevaluasi kualitas hasil pengelompokan data algoritma *K-Medoids* menggunakan *Silhouette Coefficient* berdasarkan kombinasi masukan nilai kluster yang berbeda. Hasil pengujian ditunjukkan pada Tabel 1.

Tabel 1. Hasil *Silhouette Coefficient*

No	Jumlah Kluster	Nilai <i>Silhouette Coefficient</i>
1	2	0.14637431590665603
2	3	0.11000736495160417
3	4	0.23045321036704783
4	5	0.22435958179742807

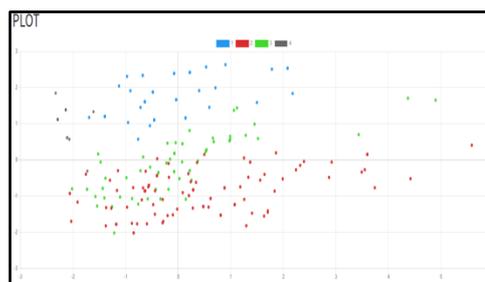
Dari hasil pengujian pada Tabel 1 menunjukkan jumlah kluster yang terbaik adalah kluster 4 dengan dengan nilai *silhouette coefficient* sebesar 0.23045321036704783 dan kluster yang paling buruk adalah kluster 3 dengan dengana nilai *silhouette coefficient* adalah 0.11000736495160417.



Gambar 5. Grafik pengaruh perbedaan jumlah kluster dengan 264 data terhadap *Silhouette Coefficient*.

1) Plot

Dari hasil uji coba pada sistem ini, didapatkan *plot* yang dapat dilihat pada Gambar 2.

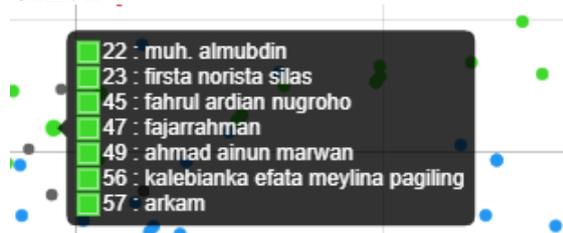


Gambar 6. Ilustrasi Plot

Plot terbentuk dari *Principal Component Analysis* (PCA). Fungsi dari *principal component analysis* (PCA) adalah dengan merangkum banyak variable dari kuisisioner yang memiliki indikasi saling berkorelasi ataupun saling mempengaruhi menjadi satu atau lebih variable baru yang memuat kombinasi dari variabel-variabel kuisisioner sebelumnya. Dalam penelitian ini *Principal Component Analysis* (PCA)

digunakan untuk membentuk dua variable yang berfungsi untuk membentuk *plot*.

Dari *plot* yang terbentuk terdapat beberapa orang yang mempunyai nilai yang sama. Hal ini terjadi karena mereka memiliki jawaban yang sama. Selain itu terdapat *cluster* yang mempunyai nilai yang berdekatan sehingga menyebabkan dua *cluster* saling bercampur. Berikut diilustrasikan pada Gambar 7.



Gambar 7. Ilustrasi Data *Plot* yang Sama

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Penerapan Algoritma K-Medoids dalam Penentuan Faktor Terbesar Sumber Informasi Pemilihan Jurusan di Fakultas Teknik Universitas Halu Oleo memiliki beberapa kesimpulan antara lain:

- 1) Penerapan Algoritma K-Medoids dalam Penentuan Faktor Terbesar Sumber Informasi Pemilihan Jurusan di Fakultas Teknik Universitas Halu Oleo telah berhasil dibangun untuk menghasilkan *cluster* terbesar sumber informasi pemilihan jurusan di Fakultas Teknik Universitas Halu Oleo.
- 2) Berdasarkan hasil pengujian dengan menerapkan metode *clustering* K-Medoids pada pengelompokan data dengan jumlah *cluster* 3 menampilkan hasil yang sama antara perhitungan manual dan perhitungan pada sistem yaitu *cluster* pertama terdiri dari data mahasiswa yang mendapatkan informasi melalui alumni UHO dan *cluster* kedua terdiri dari data mahasiswa yang mendapatkan informasi melalui media masa, *cluster* ketiga adalah data mahasiswa yang mendapatkan informasi melalui internet.
- 3) Berdasarkan hasil analisa dengan menerapkan metode K-Medoids menghasilkan kesimpulan bahwa sangat setuju ditempati oleh kriteria peluang

kerja dan sangat tidak setuju ditempati oleh kriteria media.

- 4) Dari hasil pengujian menunjukkan jumlah *cluster* yang terbaik adalah *cluster* 4 dengan dengan nilai *Silhouette Coefficient* sebesar 0.23045321036704783 dan *cluster* yang paling buruk adalah *cluster* 3 dengan dengan nilai *Silhouette Coefficient* adalah 0.1100073649516041.
- 5) terdapat beberapa orang yang mempunyai nilai yang sama. Hal ini terjadi karena mereka memiliki jawaban yang sama. Selain itu terdapat *cluster* yang mempunyai nilai yang berdekatan sehingga menyebabkan dua *cluster* saling bercampur.

5. SARAN

Untuk pengembangan sistem selanjutnya diharapkan dapat dilakukan pengembangan sistem dengan metode pengelompokan lainnya dan dapat menambahkan beberapa kriteria lainnya.

DAFTAR PUSTAKA

- [1] S. Mulyatini and T. Handayani, "Faktor-faktor yang berpengaruh terhadap keputusan memilih program studi," 2010.
- [2] E. Widyastuti *et al.*, "Hubungan Antara Minat Belajar Matematika Keaktifan Siswa Dan Fasilitas Belajar Disekolah Dengan Prestasi Belajar Matematika Siswa Kelas X Smk Se-Kecamatan Umbulharjo", vol. 0, no. 0. 2018.
- [3] W. Budiaji, "Skala Pengukuran dan Jumlah Respon Skala Likert (The Measurement Scale and The Number of Responses in Likert Scale)," *Ilmu Pertan. dan Perikan.*, vol. 2, no. 2, pp. 127–133, 2013, [Online]. Available: <http://umbidharma.org/jipp>.
- [4] M. Simanjuntak and Dkk, "Penerapan Data Mining Pengelompokan Kejahatan Elektronik Sesuai UU ITE dengan Menggunakan Metode Clustering," *J. Mahajana Inf.*, vol. 3, no. 2, p. 3, 2018.
- [5] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has.*

- Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [6] R. Handoyo, R. Rumani M, and S. Michrandi Nasution, “Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K - Means Pada Pengelompokan Dokumen,” *J. SIFO Mikroskil*, vol. 15, no. 2, pp. 73–82, Oct. 2014.
- [7] R. Puspitasari, “Analisis Faktor Penyebab Terjadinya Perubahan Pada Kontrak Lump Sum (Studi Kasus : Proyek Apartment And Soho Ciputra World) Naskah Publikasi,” 2016.
-

