



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

TUGAS AKHIR - 141501

**RANCANG BANGUN PERANGKAT LUNAK VISUALISASI AKTIVITAS REMAJA DI INSTAGRAM (STUDI KASUS: SISWA SMP DI SURABAYA)**

**AN APPLICATION OF VISUALIZATION TEENAGER'S ACTIVITY ON INSTAGRAM (CASE STUDY: STUDENT OF JUNIOR HIGH SCHOOL IN SURABAYA)**

ALVIN RAHMAN KAUTSAR  
NRP 5213100031

Dosen Pembimbing  
Nur Aini R., S.Kom., M.Sc.Eng., Ph.D  
Irmasari Hafidz, S.Kom, M.Sc

DEPARTEMEN SISTEM INFORMASI  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya, 2017

*Halaman ini sengaja dikosongkan*

TUGAS AKHIR - 141501

**RANCANG BANGUN PERANGKAT LUNAK VISUALISASI AKTIVITAS REMAJA DI INSTAGRAM (STUDI KASUS: SISWA SMP DI SURABAYA)**

**ALVIN RAHMAN KAUTSAR**

**NRP 5213100031**

Dosen Pembimbing

**Nur Aini R., S.Kom., M.Sc.Eng., Ph.D**

**Irmasari Hafidz, S.Kom, M.Sc**

**DEPARTEMEN SISTEM INFORMASI**

**Fakultas Teknologi Informasi**

**Institut Teknologi Sepuluh Nopember**

**Surabaya, 2017**

*Halaman ini sengaja dikosongkan*

UNDERGRADUATE THESIS - 141501

**AN APPLICATION OF VISUALIZATION TEENAGER'S ACTIVITY ON INSTAGRAM (CASE STUDY: STUDENT OF JUNIOR HIGH SCHOOL IN SURABAYA)**

**ALVIN RAHMAN KAUTSAR**

**NRP 5213100031**

**Supervisor**

**Nur Aini R., S.Kom., M.Sc.Eng., Ph.D**

**Irmasari Hafidz, S.Kom, M.Sc**

**DEPARTMENT OF INFORMATION SYSTEM**

**Faculty of Information Technology**

**Institut Teknologi Sepuluh Nopember**

**Surabaya, 2017**

*Halaman ini sengaja dikosongkan*

**LEMBAR PERSETUJUAN**

**RANCANG BANGUN PERANGKAT LUNAK VISUALISASI  
AKTIVITAS REMAJA DI INSTAGRAM (STUDI KASUS:  
SISWA SMP DI SURABAYA)**

**TUGAS AKHIR**

Diajukan Guna Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Bidang Studi Akuisisi Data dan Diseminasi Informasi  
Program Studi S1 Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

**ALVIN RAHMAN KAUTSAR**

**NRP: 5213100031**

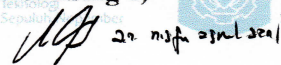
Disetujui Tim Penguji: Tanggal Ujian: 7 Juli 2017

Periode Wisuda: September 2017

**Nur Aini R., S.Kom., M.Sc.Eng., Ph.D**

  
**(Pembimbing 1)**


**Irmasari Hafidz, S.Kom, M.Sc**

  
**(Pembimbing 2)**

**Faizal Johan Atletiko, S.Kom., M.T.**

  
**(Penguji 1)**

**Radityo Prasetyanto, S.Kom., M.Kom.**

  
**(Penguji 2)**

*Halaman ini sengaja dikosongkan*



**LEMBAR PENGESAHAN**

**RANCANG BANGUN PERANGKAT LUNAK VISUALISASI  
AKTIVITAS REMAJA DI INSTAGRAM (STUDI KASUS:  
SISWA SMP DI SURABAYA)**

**TUGAS AKHIR**

Diajukan Guna Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Bidang Studi Akuisisi Data dan Diseminasi Informasi  
Program Studi S1 Departemen Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

**ALVIN RAHMAN KAUTSAR**

**NRP: 5213100031**

Surabaya, September 2017

**KEPALA  
DEPARTEMEN SISTEM INFORMASI**

**Dr. Ir. Aris Tjahyanto, M.Kom.**

**NIP. 19650310 199102 1 001**

*Halaman ini sengaja dikosongkan*

## **RANCANG BANGUN PERANGKAT LUNAK VISUALISASI AKTIVITAS REMAJA DI INSTAGRAM (STUDI KASUS: SISWA SMP DI SURABAYA)**

Nama : ALVIN RAHMAN KAUTSAR  
NRP : 5213100031  
Departemen : Sistem Informasi FTIf  
Pembimbing I : Nur Aini R., S.Kom., M.Sc.Eng., Ph.D  
Pembimbing II : Irmasari Hafidz, S.Kom, M.Sc

### **Abstrak**

*Indonesia memiliki jumlah pengguna internet (media sosial) berusia remaja yang tinggi, yakni berjumlah 24,4 juta pengguna. Di media sosial, remaja dengan leluasa dapat mengakses informasi kapanpun. Keleluasaan remaja dalam mengakses media sosial ini dikhawatirkan dapat mengganggu jam harian mereka seperti jam sekolah, jam belajar dan jam istirahat yang cukup. Istirahat (tidur) dan kesehatan mental memiliki keterkaitan yang erat. Kualitas tidur yang buruk dapat menyebabkan kesehatan mental yang buruk dan juga sebaliknya. Dari masalah tersebut, hadirnya sebuah media yang dapat memvisualisasikan pola aktivitas remaja dapat menjadi salah satu solusi untuk mencegah aktivitas online terhadap harian yang kurang sesuai. Dalam penelitian ini, penulis membuat sebuah rancang bangun perangkat lunak untuk visualisasi aktivitas remaja di Instagram. Studi kasus penelitian ini ditujukan pada siswa di 18 SMP di Surabaya. Teknik pengambilan data yang diterapkan dalam penelitian ini diantaranya adalah survey dan crawling data. Dari proses pengambilan data didapatkan 6 atribut utama yaitu tipe online, hari, jam, aktivitas harian, jenis sekolah dan jenis kelamin. Selanjutnya 6 atribut tersebut di kluster menggunakan metode K-Modes Clustering menggunakan aplikasi R-studio dan divisualisasikan dengan R-shiny.*

*Kata kunci: Remaja, instagram, crawling, K-Modes Clustering, Perangkat Lunak.*

## **AN APPLICATION OF VISUALIZATION TEENAGER'S ACTIVITY ON INSTAGRAM (CASE STUDY: STUDENT OF JUNIOR HIGHSCHOOL IN SURABAYA)**

Name : ALVIN RAHMAN KAUTSAR  
NRP : 5213100031  
Major : Information System FTIf  
Supervisor I : Nur Aini R., S.Kom., M.Sc.Eng., Ph.D  
Supervisor II : Irmasari Hafidz, S.Kom, M.Sc

### **Abstract**

*Indonesia has a high number of internet users (social media), which amounted to 24.4 million users. In Social media, adolescents can freely access information Anytime. The flexibility of adolescents in accessing this social media feared could interfere with their daily hours like clockwork School, study hours and adequate breaks. Rest (sleep) and mental health have a close relationship. Poor sleep quality can lead to poor mental health and vice versa. From the problem, the presence of a media that can visualize patterns of adolescent activity Can be one solution to prevent online activity To less suitable daily. In this study, the authors Create a software design for visualization of teen activity in Instagram. This case study study is aimed at students at 18 junior high schools in Surabaya. The data retrieval technique applied in this research are survey and crawling data. From the data retrieval process obtained 6 main attributes are online type, day, hour, daily activity, type School and sex. The next 6 attributes are in the cluster Using the K-Modes Clustering method using the application R-studio and visualized with R-shiny .*

*Keywords: Teenager, instagram, crawling, K-Modes Clustering, Software*

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Segala puji dan syukur pada Allah SWT yang telah melimpahkan rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Rancang Bangun Perangkat Lunak Visualisasi Aktivitas Remaja di Instagram (Studi Kasus: Siswa SMP di Surabaya)” dengan tepat waktu.

Harapan dari penulis semoga apa yang tertulis di dalam buku Tugas Akhir ini dapat bermanfaat bagi pengembangan ilmu pengetahuan saat ini, serta dapat memberikan kontribusi nyata bagi kampus Sistem Informasi, ITS, dan bangsa Indonesia.

Dalam pelaksanaan dan pembuatan Tugas Akhir ini tentunya sangat banyak bantuan yang penulis terima dari berbagai pihak, tanpa mengurangi rasa hormat penulis ingin menyampaikan terimakasih kepada:

1. Bapak, Ibu, Kunny dan keluarga yang selalu memberikan dukungan berupa materiil maaupun imateriil sehingga penulis terus bersemangat menjalani hari-hari perkuliahan.
2. Ibu Nur Aini Rakhmawati, S.Kom., M.Sc., Eng., Ph.D dan Ibu Irmasari Hafidz, S.Kom., M.Sc. selaku dosen pembimbing penulis yang telah memberikan ide, bimbingan, saran, kritik, ilmu, dan pengalamannya yang sangat bermanfaat sehingga penulis dapat menyelesaikan Tugas Akhir ini.
3. Bapak Prof. Ir. Arif Djunaidy, M.Sc.,Ph.D.. selaku dosen wali penulis yang selalu membimbing dan memberikan arahan ke penulis.
4. Bapak Dr. Ir. Aris Tjahyanto, M.Kom. selaku Kepala Departemen Sistem Informasi yang telah memberikan ilmu dan pengalaman kepada penulis.
5. Seluruh dosen Departemen Sistem Informasi ITS yang telah

memberikan ilmu pengetahuan dan pengalaman yang sangat berharga dan bermanfaat bagi penulis.

6. Teman-teman: Tetha, Oliv, Aziz, Oyong, Wiqa, Stezar, Uci, Marina, Master Harun, ngaqaq gan, sekitar kita, Beltranis, Partmaps dan masih banyak lagi yang belum bisa saya sebutkan satu per satu yang senantiasa menemani dan memberikan motivasi bagi penulis selama perkuliahan hingga dapat menyelesaikan tugas akhir.
7. Kakak dan adik angkatan 2011, 2012 dan 2014 yang selalu membantu dan memberikan semangat bagi penulis.
8. Serta seluruh pihak-pihak lain yang tidak dapat disebutkan satu per satu yang telah banyak membantu penulis selama perkuliahan hingga dapat menyelesaikan tugas akhir ini.

Tugas Akhir ini merupakan persembahan bagi penulis untuk kedua orang tua dan keluarga besar yang selalu memberikan motivasi terbaik bagi penulis untuk dapat menuntut ilmu setinggi-tingginya dan dapat meraih kesuksesan.

Tugas Akhir ini juga masih jauh dari kata sempurna, sehingga penulis mengharapkan saran dan kritik yang membangun dari pembaca untuk perbaikan ke depan. Semoga Tugas Akhir ini dapat bermanfaat bagi perkembangan ilmu pengetahuan dan semua pihak.



## DAFTAR ISI

<b>ABSTRAK</b>	<b>xi</b>
<b>ABSTRACT</b>	<b>xiii</b>
<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR TABEL</b>	<b>xxiii</b>
<b>DAFTAR GAMBAR</b>	<b>xxv</b>
<b>DAFTAR KODE</b>	<b>xxvii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Masalah . . . . .	3
1.4 Tujuan . . . . .	4
1.5 Manfaat . . . . .	4
1.6 Relevansi . . . . .	5

<b>2</b>	<b>TINJAUAN PUSTAKA</b>	<b>7</b>
2.1	Penelitian Sebelumnya . . . . .	7
2.1.1	Penelitian 1 . . . . .	7
2.1.2	Penelitian 2 . . . . .	8
2.1.3	Penelitian 3 . . . . .	8
2.2	Dasar Teori . . . . .	9
2.2.1	Instagram . . . . .	9
2.2.2	Crawler . . . . .	10
2.2.3	Data Mining . . . . .	11
2.2.4	Klateralisasi . . . . .	13
2.2.5	R-studio . . . . .	15
2.2.6	Rshiny . . . . .	15
2.2.7	Klasifikasi Naive Bayes . . . . .	17
<b>3</b>	<b>METODOLOGI</b>	<b>19</b>
3.1	Tahapan pengerjaan tugas akhir . . . . .	19
3.1.1	Studi literatur . . . . .	19
3.1.2	Pengumpulan Data . . . . .	21
3.1.3	Pengolahan Data . . . . .	23

3.1.4	Visualisasi . . . . .	27
3.1.5	Perancangan Web . . . . .	28
3.1.6	Arsitektur Teknologi . . . . .	31
3.1.7	Penyusunan Buku Tugas Akhir . . . . .	33
<b>4</b>	<b>PERANCANGAN</b>	<b>35</b>
4.1	Pengumpulan data . . . . .	35
4.1.1	Survey Data . . . . .	36
4.1.2	Crawl Data . . . . .	40
4.2	Pengolahan Data . . . . .	42
4.2.1	Penggabungan Data . . . . .	42
4.2.2	Data reduksi . . . . .	43
4.2.3	Proses Klasterisasi . . . . .	45
4.3	Perancangan Visualisasi Data . . . . .	46
<b>5</b>	<b>IMPLEMENTASI DAN VISUALISASI</b>	<b>49</b>
5.1	Lingkungan Implementasi . . . . .	49
5.2	Crawling . . . . .	50
5.3	Penjadwalan <i>Crawling</i> . . . . .	54
5.4	<i>Query</i> Atribut Klaster . . . . .	55

5.5	Menyambungkan database dengan R-studio . . . . .	57
5.6	Klasterisasi . . . . .	59
5.7	Uji Coba . . . . .	60
5.7.1	K=3 . . . . .	60
5.8	Klasifikasi . . . . .	60
5.9	Visualisasi . . . . .	62
<b>6</b>	<b>HASIL DAN PEMBAHASAN</b>	<b>69</b>
6.1	Hasil . . . . .	69
6.1.1	<i>Crawling</i> . . . . .	69
6.1.2	Menyimpan Hasil <i>Crawling</i> . . . . .	71
6.1.3	Penjadwalan <i>Crawling</i> . . . . .	72
6.1.4	Data terkumpul . . . . .	72
6.1.5	Hasil <i>query</i> pemilihan atribut . . . . .	73
6.1.6	Penyambungan MySQL dengan R-studio . . . . .	73
6.1.7	Klasterisasi . . . . .	78
6.1.8	Klasifikasi . . . . .	81
6.1.9	Visualisasi . . . . .	82
6.1.10	Visualisasi PHP . . . . .	84

6.2	Pembahasan Hasil . . . . .	84
6.2.1	Pengujian Fungsional . . . . .	84
6.2.2	Pengujian non-Fungsional . . . . .	85
<b>7</b>	<b>KESIMPULAN DAN SARAN</b>	<b>89</b>
7.1	Kesimpulan . . . . .	89
7.2	Saran . . . . .	90
	<b>DAFTAR PUSTAKA</b>	<b>91</b>
<b>A</b>	<b>KODE</b>	<b>93</b>
A.1	Crawling . . . . .	93
A.2	Penjadwalan <i>Crawling</i> . . . . .	95
A.3	<i>Query</i> . . . . .	96
A.4	Koneksi DB MySQL ke R-studio . . . . .	96
A.5	Pengolahan di R-studio . . . . .	97
A.6	Visualisasi . . . . .	99
A.7	Halaman Instagram . . . . .	102
A.8	Hasil Data Crawling . . . . .	102
A.8.1	Visualisasi . . . . .	105



## DAFTAR TABEL

4.1	Metode pengumpulan data . . . . .	35
4.2	Daftar sekolah yang dikunjungi . . . . .	37
4.3	Potongan akun instagram dari hasil survey ke SMP	39
4.4	Formulir survey aktivitas harian . . . . .	39
4.5	Data yang digabungkan . . . . .	42
4.6	Daftar tabel dalam database . . . . .	43
4.7	Tabel aktivitas . . . . .	43
4.8	Tabel jam aktivitas . . . . .	44
4.9	Tabel time akun . . . . .	44
4.10	Tabel time hari . . . . .	44
4.11	Tabel waktu online . . . . .	45
4.12	Atribut hasil query . . . . .	47
5.1	Spesifikasi Perangkat Keras . . . . .	49
5.2	Spesifikasi Perangkat Lunak . . . . .	50
6.1	Data mentah . . . . .	73
6.2	Gabungan data atribut . . . . .	73

6.3	Data atribut . . . . .	74
6.4	Jumlah komponen atribut . . . . .	76
6.5	Withindiff . . . . .	80
6.6	Jumlah Withindiff . . . . .	80
6.7	Size . . . . .	80
6.8	Total Size . . . . .	81
6.9	Total Withindiff . . . . .	81



## DAFTAR GAMBAR

2.1	Proses web crawler . . . . .	11
2.2	Proses data mining . . . . .	12
2.3	Tampilan R-studio . . . . .	16
2.4	Dashboard rshiny . . . . .	16
2.5	Visualisasi klaster rshiny . . . . .	17
3.1	Metodologi Penelitian . . . . .	20
3.2	Halaman <i>following</i> . . . . .	22
3.3	Halaman <i>you</i> . . . . .	23
3.4	Proses validasi akun . . . . .	24
3.5	Identifikasi atribut . . . . .	25
3.6	Pemindahan data . . . . .	26
3.7	Visualisasi plot klaster . . . . .	28
3.8	Tahapan <i>Extreme Programming</i> [4] . . . . .	29
3.9	Visualisasi plot klaster . . . . .	31
4.1	Proses validasi akun . . . . .	38

5.1	Semua data yang tersimpan didalam database MySQL	56
6.1	Hasil crawling . . . . .	70
6.2	Hasil crawling di simpan di database . . . . .	71
6.3	Penjadwalan <i>crawling</i> setiap satu jam . . . . .	72
6.4	Hasil crawling . . . . .	74
6.5	Hasil plot $k=3$ . . . . .	79
6.6	Hasil akurasi klasifikasi . . . . .	82
6.7	Hasil visualisasi Teenstagram . . . . .	83
6.8	Visualisasi <i>traffic</i> aktivitas remaja di instagram . . . . .	84
6.9	Tampilan dari komputer . . . . .	86
6.10	Tampilan dari <i>handphone</i> . . . . .	87
A.1	Halaman <i>following</i> . . . . .	102
A.2	Halaman <i>you</i> . . . . .	103
A.3	Hasil crawling . . . . .	104
A.4	Hasil visualisasi Teenstagram . . . . .	105

## DAFTAR KODE

2.1	Potongan Kode Penanda Link . . . . .	10
3.1	crawl data . . . . .	31
3.2	K-Modes di R-studio . . . . .	33
4.1	potongan kode melakukan crawling . . . . .	40
4.2	library klaR . . . . .	45
4.3	K-Modes di R . . . . .	45
5.1	potongan kode crawl (1) . . . . .	50
5.2	potongan kode crawl (2) . . . . .	51
5.3	potongan kode crawl (3) . . . . .	51
5.4	potongan kode crawl (4) . . . . .	51
5.5	potongan kode crawl (5) . . . . .	52
5.6	potongan kode crawl (6) . . . . .	52
5.7	potongan kode crawl (7) . . . . .	53
5.8	perintah cron job . . . . .	54
5.9	query pemilihan atribut . . . . .	56
5.10	koneksi database dengan R-studio . . . . .	57
5.11	koneksi database dengan R-studio . . . . .	58

5.12 koneksi database dengan R-studio . . . . .	58
5.13 koneksi database dengan R-studio . . . . .	58
5.14 Data frame . . . . .	59
5.15 menghilangkan atribut id akun . . . . .	59
5.16 K-Modes . . . . .	59
5.17 plot klaster . . . . .	59
5.18 uji coba klasterisasi dengan jumlah k . . . . .	60
5.19 metode naive bayes untuk mengetahui persentase klasifikasi data . . . . .	61
5.20 server.R . . . . .	62
5.21 ui.R . . . . .	63
5.22 web teenstagram dari HTML . . . . .	65
5.23 kode PHP untuk grafik garis aktivitas online (1) . .	65
5.24 kode PHP untuk grafik garis aktivitas online (2) . .	65
5.25 kode PHP untuk grafik garis aktivitas online (3) . .	66
A.1 potongan kode crawl . . . . .	93
A.2 perintah cron job . . . . .	95
A.3 query pemilihan atribut . . . . .	96
A.4 koneksi database dengan R-studio . . . . .	96

A.5	Data frame	97
A.6	menghilangkan atribut id akun	97
A.7	K-Modes	98
A.8	plot klaster	98
A.9	klasifikasi	98
A.10	server.R	99
A.11	ui.R	100

*Halaman ini sengaja dikosongkan*

# BAB 1

## PENDAHULUAN

Pada bab pedahuluan ini akan membahas terkait latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan relevansi terhadap pengerjaan tugas akhir.

### 1.1 Latar Belakang

Minat remaja di Indonesia saat ini dalam mengakses sebuah layanan internet memang tinggi, menurut hasil survey yang dilakukan oleh APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) yang diterbitkan pada bulan November 2016, dari 132,7 juta pengguna internet di Indonesia, 18,4 persen atau 24,4 juta adalah pengguna yang masih dalam usia remaja. Usia remaja adalah masa transisi dimana ketertarikan untuk mencoba hal-hal baru sangat tinggi, seperti halnya beraktivitas di media sosial. Pengguna media sosial di Indonesia tergolong tinggi yakni sebanyak 129,2 juta pengguna [2]. Dari beberapa jenisnya, 2 media sosial yang sering diakses adalah *Facebook* (71,6 juta pengguna) dan *Instagram* (19,9 juta pengguna) [2]. Jika dicermati, *Instagram* merupakan media sosial dengan peningkatan jumlah pengguna aktif tercepat dan terbesar dimana saat ini terdapat sekitar 500 juta pengguna di dunia dan Indonesia memiliki 19,9 juta pengguna serta terus akan bertambah di tiap waktu [2].

Dari tingginya data yang didapatkan tersebut, terdapat sebuah kekhawatiran dimana media sosial dapat menjadi sarana yang membuat anak-anak usia remaja bebas beraktivitas *online*/mengakses in-

formasi setiap saat. Keleluasaan anak remaja dalam mengakses media sosial ini dikhawatirkan dapat mengganggu jam harian mereka seperti jam sekolah, belajar dan khususnya istirahat yang cukup. Istirahat (tidur) dan kesehatan mental memiliki keterkaitan yang erat. Kesehatan mental yang buruk dapat menyebabkan tidur yang buruk dan tidur yang buruk dapat menyebabkan kesehatan mental yang buruk [24]. Tidur sangat penting bagi remaja dan anak muda karena ini adalah kunci untuk berkembang. Otak tidak berkembang sepenuhnya sampai seseorang berusia dua puluhan dan tiga puluhan [26]. Tidur sangat penting untuk memungkinkan tubuh kita berfungsi dengan baik selama berjam-jam dan remaja membutuhkan sekitar 1-2 jam lebih banyak tidur setiap harinya daripada orang dewasa. Tidur yang buruk dikaitkan dengan berbagai kondisi kesehatan fisik dan mental pada orang dewasa termasuk tekanan darah tinggi, diabetes, obesitas, serangan jantung, stroke dan depresi [rs-ph]. Menurut penelitian yang dilakukan UNICEF, sekitar 50 persen orang tua di Indonesia cukup mengawasi anak dalam hal penggunaan internet, namun persoalannya *gadget* yang dipakai tersebut adalah bersifat *mobile* sehingga sulit untuk dikontrol terus-menerus [3].

Melihat dari kejadian yang disebutkan diatas, maka, dibangunnya sebuah media yang dapat memvisualisasikan aktivitas anak remaja di media sosial khususnya instagram dapat menjadi salah satu cara untuk mengetahui pola aktivitas remaja menggunakan media sosial. Pada proses visualisasi penulis menggunakan metode klasterisasi[] dengan tujuan untuk menemukan karakteristik dari aktivitas pengguna (remaja) di instagram. Beberapa atribut data yang didapatkan untuk klasterisasi diantaranya adalah tipe online, hari, jam, aktivitas harian, jenis sekolah dan jenis kelamin. Metode klaster yang digunakan dalam pengolahan data ini adalah *K-Modes Clustering*, dikarenakan data yang diolah dalam penelitian berupa data



kategori (kategorikal)[]. Studi kasus penelitian ini ditujukan untuk menganalisa pola waktu aktivitas akun instagram anak remaja yang mengenyam pendidikan di 18 SMP di daerah Surabaya. Harapan dari hasil penelitian ini adalah untuk memberikan gambaran mengenai pola aktivitas remaja untuk orang tua agar dapat mengarahkan putra/putrinya menggunakan media sosial dan juga sebagai acuan penelitian selanjutnya dalam hal analisis kebiasaan pengguna media sosial.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, tugas akhir yang akan diajukan ini menitikbertatkan permasalahan pada beberapa hal sebagai berikut:

1. Bagaimana cara mendapatkan data waktu aktivitas *online* pengguna instagram dan aktivitas harian siswa SMP di Surabaya?
2. Bagaimana cara menganalisis aktivitas *online* dan aktivitas harian siswa SMP di Surabaya?
3. Bagaimana cara membuat rancang bangun visualisasi perilaku anak remaja di instagram?

## 1.3 Batasan Masalah

Batasan-batasan dalam pembuatan Usulan Tugas Akhir adalah sebagai berikut:

1. Studi kasus penelitian ini dikhususkan pada data akun instagram anak remaja di 18 SMP daerah Surabaya.
2. Data yang digunakan sebagai identifikasi aktivitas *online* adalah berdasarkan waktu *like*, *comment* dan *follow* di *insta-*

*gram.*

3. Jadwal aktivitas harian terdiri dari 3 hal yaitu jam sekolah, belajar dan istirahat yang didapatkan dari survey ke pelajar SMP di Surabaya.
4. Proses klasterisasi menggunakan K-Modes pada penelitian ini hanya sebatas pada kebutuhan visualisasi dan persebaran data.

## **1.4 Tujuan**

Tujuan dalam pembuatan Usulan Tugas Akhir ini adalah sebagai berikut:

1. Mendapatkan data waktu aktivitas online pengguna instagram dan aktivitas harian siswa SMP di Surabaya.
2. Mengetahui pola aktivitas remaja di Surabaya.
3. Membuat rancang bangun visualisasi perilaku anak remaja di instagram.

## **1.5 Manfaat**

Manfaat yang akan diperoleh dengan tugas akhir ini antara lain:

1. Membantu orang tua dan sekolah untuk mengetahui pola perilaku anak remaja.
2. Memberikan rekomendasi orang tua dan sekolah untuk mencegah aktivitas *online* anak remaja yang tidak wajar.
3. Sebagai bahan penelitian lebih mendalam terkait analisa perilaku pengguna media sosial.

## **1.6 Relevansi**

k Tugas akhir ini berkaitan dengan mata kuliah Pemrograman Berbasis Web, Analisa dan Desain Perangkat Lunak, Konstruksi Pengembangan Perangkat Lunak dan Pemrograman Integratif.

*Halaman ini sengaja dikosongkan*

## BAB 2

### TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan mengenai penelitian yang pernah dilakukan sebelumnya dan dasar teori sebagai landasan penyusunan laporan tugas akhir ini. Dasar teori akan memberikan gambaran secara umum mengenai landasan pengerjaan tugas akhir.

#### 2.1 Penelitian Sebelumnya

##### 2.1.1 Penelitian 1

- Judul penelitian : A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining.
- Penulis : Zhexue Huang

Penelitian ini menerapkan metode K-Modes pada proses data mining untuk mengklaster data set yang besar dan berupa kategorikal. Dalam suatu kasus tertentu, menggunakan algoritma K-Means dalam proses data mining sangat efisien untuk memproses kumpulan data yang besar, namun algoritma K-Means terbatas pada data berupa numerik. Dalam penelitian ini *K-Modes Clustering* di klaim dapat lebih efisien memproses kumpulan data kategorikal. K-Modes merupakan hasil dari modifikasi K-Means, yakni dengan cara mengganti means (rata-rata) dengan modes. Metode ini diuji dengan *Soybeans diseas data set* yang umum digunakan. Dan percobaan yang dilakukan pada data asuransi kesehatan yang terdiri dari setengah juta rekaman dan 34 atribut kategorikal menunjukkan algoritma terukur pada kondisi jumlah klaster dan jumlah

rekaman [8].

### 2.1.2 Penelitian 2

- Judul penelitian : Pengklasifikasian Karakteristik dengan Metode *K-Means Cluster Analysis*.
- Penulis : Ediyanto, Muhlasah Novitasari Mara, Neva Satyahadewi

Metode *K-Means Cluster Analysis* cukup efektif diterapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian. Algoritma *K-Means Cluster Analysis* pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran dan algoritma *K-Means* ini juga dapat digunakan untuk meringkas objek dari jumlah besar sehingga lebih memudahkan untuk mendiskripsikan sifat-sifat atau karakteristik dari masing-masing kelompok [12].

### 2.1.3 Penelitian 3

- Judul penelitian : Aplikasi *K-Means* untuk Pengelompokan Mahasiswa Berdasarkan Nilai *Body Mass Index* (BMI) dan Ukuran Kerangka.
- Penulis : Tedy Rismawan dan Sri Kusumadewi  
Pada penelitian ini penulis mencoba membangun sebuah sistem untuk mengelompokkan data yang ada berdasarkan status gizi dan ukuran rangkanya dengan memasukkan parameter kondisi fisik dari orang tersebut. Dari hasil penelitian terhadap 20 data sampel diperoleh 3 kelompok mahasiswa berdasarkan nilai BMI dan ukuran rangka, yaitu: BMI normal dan kerangka besar, BMI obesitas sedang dan kerangka se-

dang, BMI obesitas berat dan kerangka kecil [11].

## 2.2 Dasar Teori

### 2.2.1 Instagram

*Instagram* adalah sebuah layanan aplikasi *mobile* yang memungkinkan penggunanya untuk mengabadikan serta berbagi berbagai momen kedalam bentuk foto ataupun video dengan mudah [9]. Layanan ini merupakan sebuah media sosial yang menghubungkan pengguna satu dengan lainnya. Beberapa kegiatan yang dapat dilakukan didalam *Instagram* adalah *like*, *comment* dan *follow*. Dimana kegiatan tersebut dapat diajabarkan sebagai berikut:

1. *Like* adalah suatu kegiatan dimana seseorang menyukai sebuah tampilan/postingan dengan menekan ikon *love*.
2. *comment* adalah suatu kegiatan dimana seseorang membubuhkan komentar pada kolom komentar pada sebuah postingan.
3. *follow* adalah kegiatan dimana seorang pengguna mengikuti pengguna yang lain, dengan cara menekan ikon *follow*.

Aplikasi *Instagram* juga memiliki layanan API (*Application Programming Interface*) yang mana dapat membantu pengambilan data-data seperti foto, *caption* bahkan waktu aktivitas *online* yang pada umumnya digunakan untuk menganalisa sebuah tren dan penelitian [7].

### 2.2.2 Crawler

Web Crawler adalah sebuah program yang berfungsi untuk melakukan penjelajahan dan pengambilan halaman-halaman web yang ada di internet [13]. Proses yang dilakukan dalam web crawler adalah seperti berikut:

1. Mengunduh halaman web.
2. Memparsing halaman yang didownload dan mengambil semua link.
3. Untuk setiap link yang diambil, ulangi proses.

Tahap pertama adalah melakukan pengambilan URL dan mengunduh halaman berdasarkan URL tersebut. Kemudian halaman yang telah diunduh disimpan kedalam database. Tahap kedua, web crawler melakukan parsing keseluruhan halaman yang telah diunduh dan mengambil link-link ke halaman lain[13]. Tiap link dalam halaman didefinisikan dengan sebuah penanda HTML seperti berikut:

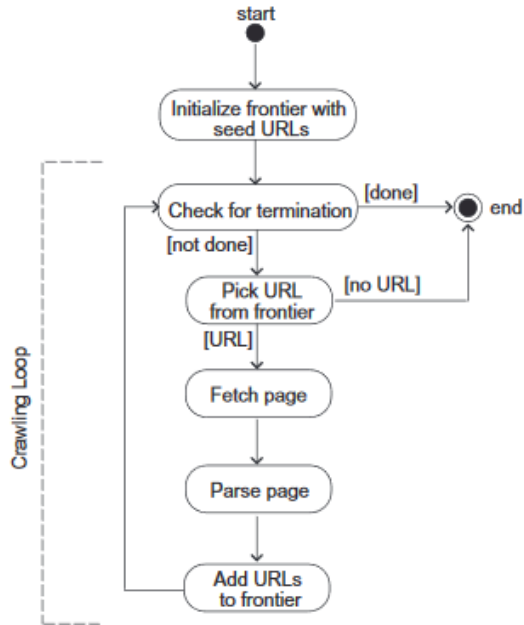
```
1 <a href="http://www.host.com/directory/file.html">link </a>
2
```

#### **Kode 2.1:** Potongan Kode Penanda Link

Setelah crawler mengambil link pada halaman, kemudian tiap link ditambahkan ke sebuah daftar untuk dicrawl. Tahap ketiga, mengulangi proses.

Berikut ini adalah proses web crawler berurutan:





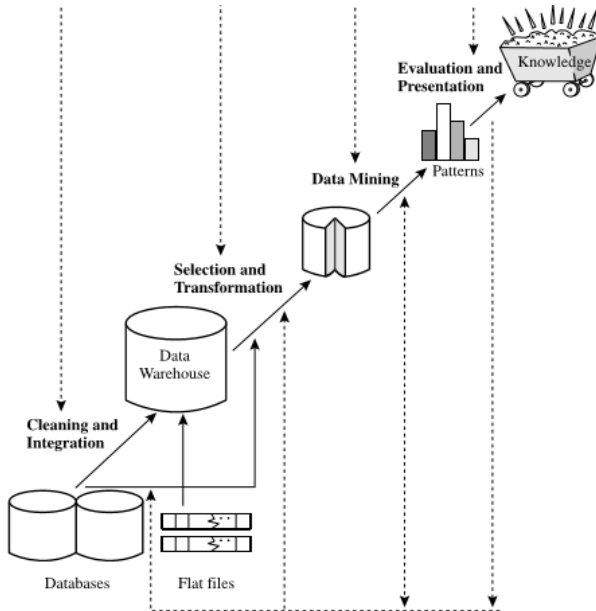
**Gambar 2.1:** Proses web crawler

[10]

### 2.2.3 Data Mining

Data Mining adalah suatu proses penggalian pengetahuan dari penyimpanan data yang besar yang bertujuan untuk menemukan suatu pola atau tren dari kumpulan data tersebut [5].

Tahapan-tahapan yang dilakukan dalam data mining adalah sebagai berikut:



**Gambar 2.2:** Proses data mining

[5]

Dalam melakukan proses data mining urutan tahapan yang dilakuk-  
an adalah:

- *Data cleaning* (Pembersihan data)  
Pembersihan data dilakukan untuk menghilangkan data yang mengganggu atau *noise* dan tidak tidak konsisten. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik.
- *Data integration* (Integrasi data)  
Integrasi data dilakukan untuk menyatukan data atau meng-

gabungkan data dari beberapa sumber.

- *Data selection* (Pemilihan data) Melakukan proses pemilihan data yang dianggap relevan dan membuang data yang tidak relevan dari sebuah database.
- *Data transformation* (Transformasi data) Mengubah data mentah kedalam suatu format data untuk memudahkan proses data mining.
- *Data mining* (Penggalian data) Melakukan proses penggalian data menggunakan metode cerdas yang diaplikasikan untuk mengekstraksi pola.
- *Pattern evaluation* (Evaluasi pola) Proses mengenali pola yang menarik dari hasil data mining.
- *Knowledge Presentation* (Penyajian pengetahuan) Melakukan penyajian / membuat visualisasi data untuk memudahkan pengguna dalam memahami pola data.

#### 2.2.4 Klasterisasi

Klasterisasi adalah proses mengelempokkan kumpulan obyek fisik atau abstrak kedalam suatu kelas atau kelompok yang memiliki karakteristik obyek yang mirip. Sebuah klaster merupakan kumpulan dari obyek yang memiliki kemiripan dalam satu kelompok klaster dan memiliki perbedaan dengan kumpulan obyek di klaster lainnya. Salah satu contoh metode klasterisasi yang umum adalah *K-Means clustering*. *K-Means* merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [1]. Data clustering menggunakan *K-Means* umumnya dilakukan dengan algoritma dasar seperti ber-

ikut [1]:

1. Tentukan jumlah kluster.
2. Alokasikan data ke dalam kluster secara random.
3. Hitung *centroid*/rata-rata dari data yang ada di masing-masing kluster.
4. Alokasikan masing-masing data ke *centroid*/rata-rata data terdekat.
5. Kembali ke tahap 3, apabila masih ada data yang berpindah kluster atau apabila perubahan nilai *centroid*, ada yang diatas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan diatas nilai *threshold* yang ditentukan.

Didalam klasterisasi terdapat beberapa tipe data yang digunakan yaitu:

1. Variabel berskala interval
2. Variabel Biner
3. Variabel Kategorikal, Ordinal dan Skala Rasio
4. Variabel Campuran
5. Obyek Vektor

### ***K-Modes Clustering***

K-Modes merupakan pengembangan dari algoritma *K-means clustering* untuk menangani data kategorik di mana *means* diganti oleh *modes*. K-Modes menggunakan *simple matching measure* dalam penentuan kesamaan dari suatu kluster.

Algoritma dalam *K-Modes clustering* adalah sebagai berikut:

1. Pilih mode awal "K".
2. Alokasikan sebuah obyek ke klaster yang mode nya paling dekat dengan klaster tsb, dengan menggunakan definisi pada nomor 1. *Update* mode klaster tiap setelah alokasi.
3. Setelah semua objek dialokasikan ke klaster masing-masing, lakukan pengujian ulang dengan mode yang baru dan *update* kembali klaster tersebut.
4. Ulangi langkah 2 dan 3 hingga tidak ada perubahan pada klaster. [5]

### 2.2.5 R-studio

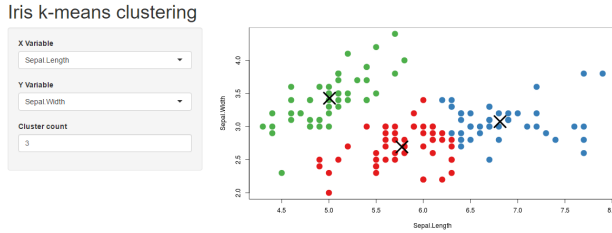
R-studio merupakan salah satu perangkat lunak untuk melakukan penggalian data / data mining yang dapat dijalankan sistem operasi Windows, Mac OS, dan Linux. R-studio merupakan perangkat lunak *open source* yang berguna untuk memudahkan pengguna dalam melakukan pengkodean berbahasa R karena lebih *user friendly* [6].

### 2.2.6 Rshiny

Shiny merupakan sebuah *open source package* dari R yang menyediakan *framework* untuk membangun web aplikasi berbahasa R yang elegan dan kuat. Shiny membantu untuk menganalisis web yang interaktif tanpa membutuhkan keahlian HTML, CSS atau JavaScript [6]. Beberapa tampilan yang dapat dibangun menggunakan Rshiny adalah sebagai berikut:

1. Dashboard  
Tampilan dashboard digunakan untuk merangkum tampilan / informasi penting dalam satu layar.





**Gambar 2.5:** Visualisasi kluster rshiny

## 2. Visualisasi kluster

Tampilan contoh visualisasi dari hasil klusterisasi data.

### 2.2.7 Klasifikasi Naive Bayes

Klasifikasi merupakan teknik dalam data mining yang dilakukan untuk melempokkan obyek berdasarkan ciri-ciri persamaan dan perbedaannya. Naive Bayes merupakan salah satu metode dalam melakukan klasifikasi, yang diperkenalkan oleh Thomas Bayes pada tahun 1702. Dalam sebuah kemungkinan bahwa  $X$  terjadi dengan adanya bukti  $Y$ , sehingga dapat ditulis dengan  $P(X|Y)$ .

Naive Bayes dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas.

Untuk memperkirakan nilai probabilitas bisa didapatkan dengan rumus dibawah ini [5]:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

(1)

Terdapat 2 fase tahapan dalam naive bayes, yaitu tahap *training*

dan tahap *testing*. Pada proses *training* akan dimasukkan data yang nantinya diproses sehingga mendapatkan obyek yang mencerminkan kelas yang berkaitan. Kemudian setelah proses *training* akan dilanjutkan dengan proses *testing*.

Dari proses *training* nantinya akan didapatkan nilai probabilitistik dari  $P(V_j)$  dan  $P(W_k|V_j)$ . Untuk menghitung perbandingan jumlah kelas dokumen  $P(V_j)$  digunakan rumus:

$$P(V_j) = \frac{|fd(V_j)|}{|D|} \quad (2)$$

$fd(V_j)$  adalah jumlah kata-kata di setiap kategori  $j$ , dan  $D$  adalah jumlah kelas dokumen untuk dijadikan model.

Kemudian setelah dilakukan perbandingan kelas selanjutnya adalah melakukan penghitungan indeks kelas dengan frekuensi kata kunci dengan rumus:

$$P(W_k|V_j) = \frac{f(W_k|V_j) + 1}{N + |W|} \quad (3)$$

$P(W_k|V_j)$  adalah jumlah kejadian dari kata  $W_k$  dalam kategori  $V_j$ . Sedangkan  $N$  adalah jumlah semua kata yang ada pada kategori  $V_j$ , dan  $W$  adalah jumlah kata unik dalam data *testing*. [5]



## **BAB 3**

### **METODOLOGI**

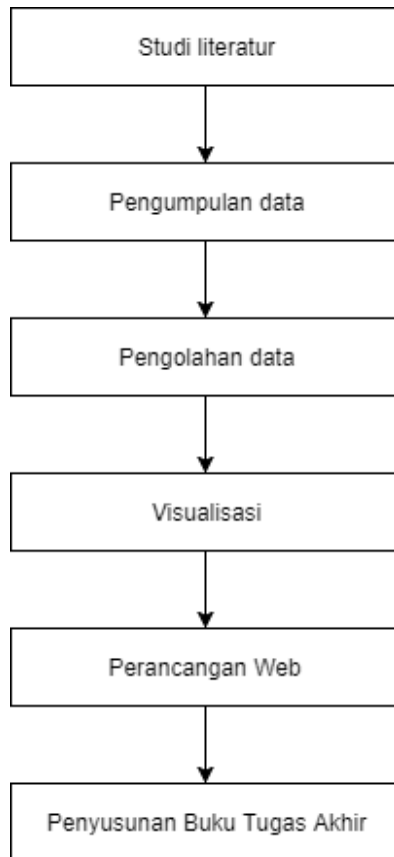
Pada bab metodologi akan menjelaskan bagaimana langkah pengerjaan tugas akhir dengan disertakan deskripsi dari setiap penjelasan untuk masing-masing tahapan beserta jadwal kegiatan pengerjaan tugas akhir.

#### **3.1 Tahapan pengerjaan tugas akhir**

Tahapan pelaksanaan tugas akhir yang menjelaskan mengenai proses pengerjaan gambar A.4

##### **3.1.1 Studi literatur**

Tahap awal yang akan dilakukan dalam pengerjaan tugas akhir ini adalah studi literatur. Studi literatur yang dilakukan adalah pembelajaran terhadap literatur terkait konsep-konsep serta metodologi yang didapatkan dari penelitian-penelitian yang telah ada sebelumnya yang didokumentasikan dalam bentuk buku, jurnal ilmiah dan artikel website. Beberapa literatur yang digunakan sebagai landasan dalam penelitian ini adalah mengenai teknik *Crawling*, *K-Modes Clustering*, dan aktivitas anak remaja di media sosial.



**Gambar 3.1:** Metodologi Penelitian

### 3.1.2 Pengumpulan Data

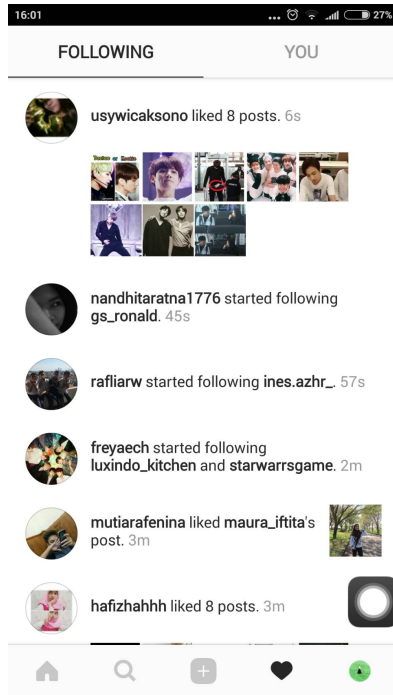
Pada tahap ini akan dilakukan proses pengumpulan data dimana terdapat 2 metode yang digunakan yakni survey dan crawl. Berikut ini adalah penjelasan dari masing-masing metode pengambilan data:

#### **Survey Data**

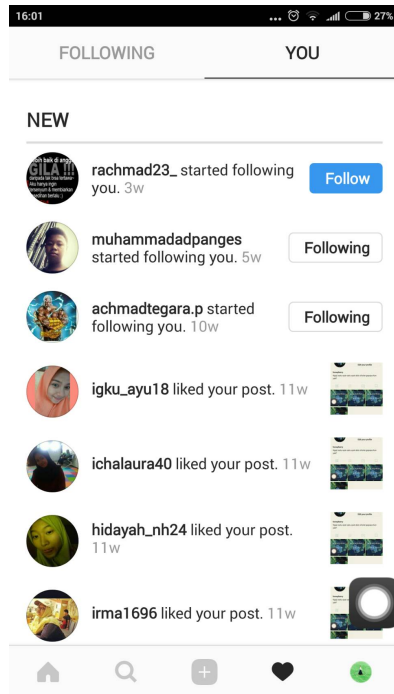
Metode survey adalah metode yang bertujuan untuk mendapatkan data/informasi mengenai akun media sosial dan aktivitas harian remaja dengan cara membuat sebuah lembaran formulir kuisioner yang akan diisi oleh responden. Pengambilan data survey didapatkan bersumber dari 2 kegiatan yaitu, pertama, Survey Edukasi Penggunaan Smartphone bagi Siswa SMP di Surabaya, kedua, Survey Mandiri. Kegiatan survey ditujukan kepada pelajar di 18 SMP di Surabaya.

#### ***Crawling***

Teknik Crawling adalah teknik yang digunakan untuk mengambil data secara online melalui layanan API dan link instagram/akun pelajar yang telah didapatkan dari proses survey sebelumnya. Data yang diambil pada proses ini adalah aktivitas tiap waktu yang dilakukan pengguna. Dalam aplikasi instagram aktivitas tersebut dapat dilihat pada halaman *following* (gambar 3.2) dan *you* (gambar 3.3)



**Gambar 3.2:** Halaman *following*



**Gambar 3.3:** Halaman *you*

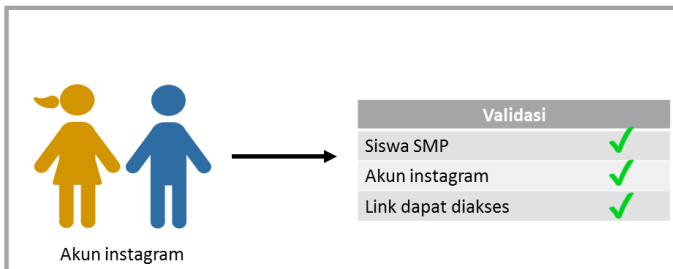
### 3.1.3 Pengolahan Data

Pada proses pengolahan data dilakukan beberapa tahapan yaitu validasi akun instagram, pemindahan data survey dari Ms.Excel ke database MySQL, identifikasi atribut data dan klasterisasi *K-Modes*.

## Pembersihan Data

Pembersihan data dilakukan untuk menghilangkan data yang tidak sesuai / tidak digunakan dalam kebutuhan penelitian. Proses ini dilakukan setelah mendapatkan data khususnya dari data hasil survey Edukasi Penggunaan Smartphone bagi Siswa SMP di Surabaya, seperti contoh pengambilan akun instagram siswa SMP.

Tahap ini bertujuan agar akun yang akan dianalisis merupakan akun valid (bukan akun palsu) selain itu proses ini juga penting untuk melakukan *crawling* data aktivitas online siswa. Beberapa hal yang perlu diperhatikan dalam validasi akun:



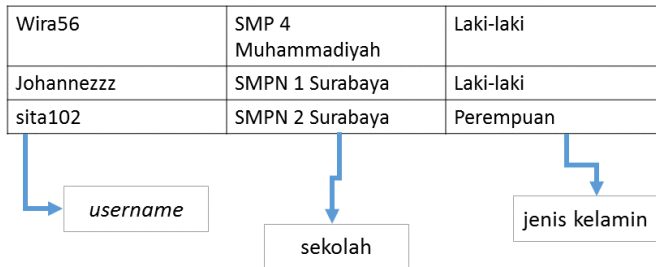
**Gambar 3.4:** Proses validasi akun

### Keterangan

1. Akun media sosial merupakan milik pelajar SMP
2. Akun yang dipilih adalah instagram
3. Akun aktif (bukan akun palsu) yakni dipastikan dengan cara membuka setiap *link/username* yang tercantum dalam formulir survey.

## Identifikasi Atribut

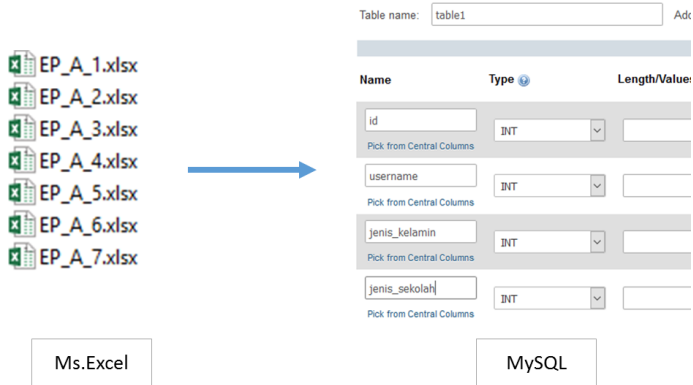
Data yang berhasil didapatkan dari proses pengumpulan data akan diidentifikasi jenis atributnya dan kemudian disusun sebelum dilakukan tahap kluster.



**Gambar 3.5:** Identifikasi atribut

## Pemindahan Data

Setelah akun instagram di validasi, langkah selanjutnya adalah memindahkan data-data survey dari format ms.Excel kedalam database MySQL untuk proses crawling data aktivitas *online* dan agar mempermudah pemrosesan data dengan *query*.



**Gambar 3.6:** Pemindahan data

## Klasterisasi

Untuk melakukan klasterisasi data berupa kategori penulis menggunakan metode *K-Modes Clustering*. Tahapan dalam melakukan klaster adalah sebagai berikut:

1. Pilih mode awal "K".
2. Alokasikan sebuah obyek ke klaster yang mode nya paling dekat dengan klaster tsb, dengan menggunakan definisi pada nomor 1. *Update* mode klaster tiap setelah alokasi.
3. Setelah semua objek dialokasikan ke klaster masing-masing, lakukan pengujian ulang dengan mode yang baru dan *update* kembali klaster tersebut.
4. Ulangi langkah 2 dan 3 hingga tidak ada perubahan pada



klaster. [5]

Dalam melakukan proses klasterisasi *K-Modes* penulis menggunakan aplikasi RStudio. Berikut merupakan perintah R untuk melakukan *K-Modes Clustering*:

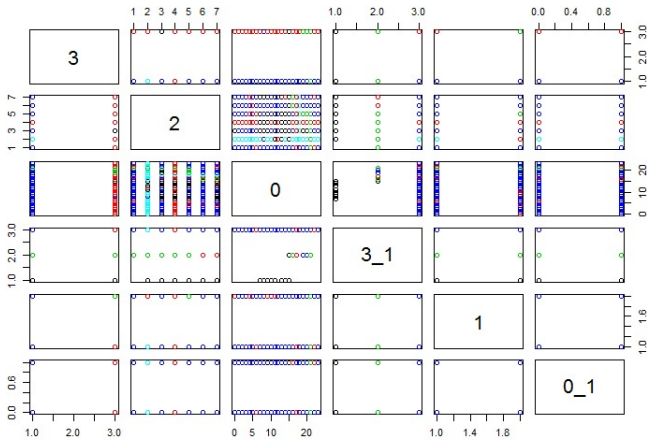
- `kmodes(data, modes, iter.max = 10, weighted = FALSE)`

Keterangan:

- *data* = data/data frame yang digunakan (berupa data kategori).
- *modes* = jumlah mode (inisiasi jumlah klaster awal).
- *iter.max* = jumlah maksimal melakukan perulangan proses klaster.
- *weighted* = Pencocokan jarak tiap klaster.

### 3.1.4 Visualisasi

Setelah dilakukan klasterisasi, langkah selanjutnya adalah memvisualisasikan hasil klaster kedalam sebuah tampilan plot. Visualisasi tersebut dapat dilakukan dengan menggunakan *library* yang sudah diinstal di aplikasi RStudio.



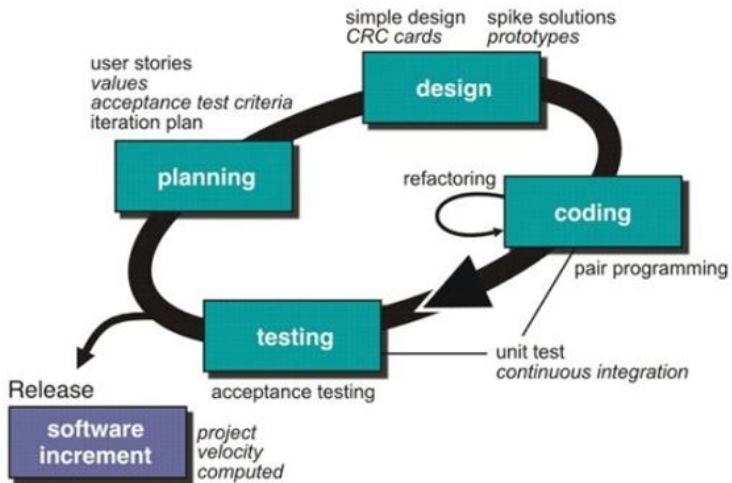
**Gambar 3.7:** Visualisasi plot kluster

## Analisis Data

Tahapan ini dilakukan setelah proses klusterisasi data divisualisasikan, dengan melihat visualisasi dan hasil kluster dapat analisis karakteristik dari setiap data. Dari analisis ini akan dijabarkan mengenai pola yang terbentuk dari proses pengolahan data.

### 3.1.5 Perancangan Web

Pada tahap ini akan dilakukan perancangan website yang digunakan untuk memvisualisasikan hasil kluster. Dalam merancang website, penulis melakukan teknik *extreme programming* dalam membuat rancang bangun websitenya. Berikut ini adalah tahapan dalam melaksanakan *Extreme programming* yang ditunjukkan Gambar 3.8:



**Gambar 3.8:** Tahapan *Extreme Programming* [4]

Berikut ini tahapan-tahapan yang dilakukan dalam melaksanakan *Extreme Programming*:

### 1. *Planning*

Tahap *planning* atau perencanaan merupakan tahap dimana dilakukannya penggalian kebutuhan. Pada perencanaan akan dibuat *User Stories* dimana akan menjelaskan mengenai *output*, fungsi dan fitur-fitur aplikasi yang akan dibangun.

Berikut merupakan daftar kebutuhan dalam rancang bangun perangkat lunak:

- (a) Menampilkan grafik waktu online
- (b) Menampilkan data waktu harian remaja, seperti: jam sekolah, jam belajar dan jam istirahat
- (c) Menampilkan aktivitas *online* dari tiap jenis kelamin, sekolah dan aktivitas.

## 2. *Design*

Tahap kedua adalah melakukan desain aplikasi mulai dari struktur database hingga *interface* atau antar muka dari halaman website. Tahap kedua adalah melakukan desain perangkat lunak. Beberapa kegiatan desain yang dibutuhkan adalah sebagai berikut:

- (a) Desain *database*
- (b) Desain *crawler*
- (c) Desain *interface*

## 3. *Coding*

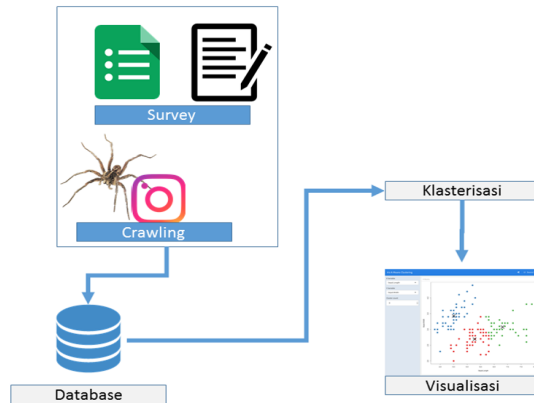
Selanjutnya adalah melakukan *coding*/pengkodean, dimana pengkodean yang akan dilakukan adalah seperti berikut:

- (a) *Crawler*  
Melakukan pengkodean *web crawler* untuk mengambil data yang dibutuhkan, yakni waktu *online*.
- (b) *Sistem*  
Melakukan pengkodean sistem dengan mengikuti metode *K-Means Clustering* yang berfungsi untuk mengcluster beberapa kategori data.
- (c) *Interface*  
Melakukan pengkodean *interface* atau antar muka website. Dimana antar muka utama yang ditawarkan adalah berupa grafik yang terdiri atas beberapa kategori *cluster*.

## 4. *Testing*

Tahap terakhir adalah melakukan *testing* untuk mengetahui kelayakan dari aplikasi/web yang telah diprogram.

### 3.1.6 Arsitektur Teknologi



**Gambar 3.9:** Visualisasi plot kluster

Gambar 3.9 merupakan bentuk aritektur / teknologi yang diterapkan dalam proses penelitian. Dimana terdapat beberapa komponen diantaranya:

1. Survey adalah teknik untuk pengambilan data secara *offline*.
2. Crawling adalah teknik untuk pengambilan data secara online, dimana dalam proses teknik *crawling* tersebut penulis menggunakan bahasa pemrograman php. Potongan kode 3.1 berikut ini merupakan *script* melakukan crawl data dengan bahasa php.

```

1 date_default_timezone_set('Asia/Jakarta');
2
3 $i = new \InstagramAPI\Instagram($debug);
4 $i->setUser($username, $password);
5
6 try {
7     $i->login($force=true);

```

```

8 } catch (Exception $e) {
9     echo 'something went wrong ' . $e->getMessage() . "\n";
10    exit(0);
11 }
12 try {
13     $helper = null;
14     $followers = [];
15
16     do {
17         if (is_null($helper)) {
18             $helper = $i->getFollowingRecentActivity();
19             foreach ($helper->stories as $row) {
20                 save_time( $row );
21                 # code ...
22             }
23
24             $helper = $i->getRecentActivity();
25             foreach ($helper->old_stories as $row) {
26                 save_time( $row );
27                 # code ...
28             }
29
30             echo json_encode($helper); //menampilkan json
31         } else {
32             $helper = $i->getSelfUserFollowers($helper->
33                 getNextMaxId());
34         }
35
36         // $followers = array_merge($followers, $helper->
37             getUsers());
38     } while (!is_null($helper->getNextMaxId()));
39
40     // echo "My followers: \n";
41     // foreach ($followers as $follower) {
42     //     echo '- ' . $follower->getUsername() . "\n";
43     // }
44 } catch (Exception $e) {
45     echo $e->getMessage();
46 }

```

### Kode 3.1: crawl data

3. Pengumpulan seluruh data mentah menggunakan database MySQL. Tujuan dari penggunaan database tersebut karena MySQL secara umum digunakan dalam penyimpanan data

dan tersedia di banyak penyedia layanan server/hostinger seperti BEON, Jagoan hosting, dan lain-lain. Dengan menggunakan MySQL pemrosesan data dapat dilakukan dengan hanya melakukan query untuk mengolah data serta bila ada data baru yang masuk dapat langsung diproses.

4. Klasterisasi merupakan teknik pengelompokan data dalam Data Mining. Dalam proses klasterisasi penulis menggunakan *K-Modes Clustering* yang diproses menggunakan aplikasi R-studio. R-studio sendiri merupakan aplikasi berbahasa R. Berikut merupakan contoh penerapan klasterisasi di R-studio(kode 3.2):

```
1 kmodes(data, modes, iter.max=10, weighted=FALSE)
2
```

**Kode 3.2:** K-Modes di R-studio

5. Visualisasi dalam menampilkan hasil klasterisasi adalah menggunakan R-shiny. R-shiny merupakan sebuah paket dari R-studio yang merupakan layanan berbahasa R yang sesuai untuk menampilkan hasil Data Mining. Penggunaan R-shiny adalah dengan melakukan pengkodean dan *compile* R di aplikasi R-studio. Tampilan yang dapat dibangun menggunakan R-shiny adalah dashboard, diagram, dan lain-lain.

### 3.1.7 Penyusunan Buku Tugas Akhir

Tahap terakhir adalah melakukan penyusunan Buku Tugas Akhir. Penulisan buku ini dilaksanakan bersamaan dengan tahapan metodologi yang telah disebutkan sebelumnya. Harapan dari selesainya Buku Tugas Akhir ini adalah agar dapat menjadi sumber referensi penelitian selanjutnya.

*Halaman ini sengaja dikosongkan*



## BAB 4

### PERANCANGAN

Pada bab ini akan dijelaskan rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian akan dilakukan.

#### 4.1 Pengumpulan data

Pada tahap ini akan dilakukan proses pengumpulan data. Untuk mendapatkan data, metode yang digunakan adalah survey dan *crawling* data. Berikut ini akan dijelaskan masing-masing metode pengumpulan data tersebut:

**Tabel 4.1:** Metode pengumpulan data

<b>Kegiatan</b>	<b>Periode</b>	<b>Keterangan</b>
Survey 1	3 - 21 Oktober 2016	Mendapatkan biodata dan akun/username remaja SMP
Survey 2	9 - 17 Mei 2017	Mengambil data jam harian - sekolah - belajar - istirahat
Crawling	23 April - 18 Mei 2017	Mendapatkan waktu online: - like - comment - follow

### 4.1.1 Survey Data

Survey dilakukan bertujuan untuk mendapatkan data/informasi mengenai akun media sosial instagram, jenis kelamin, jam sekolah, jam belajar dan jam istirahat para remaja di Surabaya.

Pengambilan data survey didapatkan bersumber dari 2 kegiatan survey, yaitu :

1. Survey Edukasi Penggunaan *Smartphone* bagi Siswa SMP di Surabaya.  
Survey edukasi penggunaan *smartphone* bagi siswa di Surabaya merupakan survey yang dilaksanakan oleh kelas Etika Profesi jurusan Sistem Informasi ITS pada tanggal 3 sampai dengan 21 Oktober 2016. Dari hasil survey ini akan diambil data berupa akun media sosial instagram pelajar di 18 SMP Surabaya sebagai representatif remaja di Surabaya.
2. Survey Mandiri  
Survey mandiri adalah survey yang dilaksanakan sendiri oleh penulis untuk mendapatkan data aktivitas harian remaja/pelajar di 18 SMP Surabaya (mengacu pada kunjungan survey Etika Profesi). Aktivitas yang dimaksud adalah waktu remaja dalam melaksanakan kegiatan sekolah, belajar dan beristirahat. Kegiatan ini dilaksanakan pada tanggal 9 sampai dengan 17 Mei 2017.

Berikut ini adalah daftar SMP (Sekolah Menengah Pertama) yang dikunjungi untuk dilakukan pengambilan data/survey:

**Tabel 4.2:** Daftar sekolah yang dikunjungi

No	Nama Sekolah
1	SMPN 1 Surabaya
2	SMPN 2 Surabaya
3	SMPN 4 Surabaya
4	SMPN 6 Surabaya
5	SMPN 12 Surabaya
6	SMPN 13 Surabaya
7	SMPN 15 Surabaya
8	SMPN 18 Surabaya
9	SMPN 19 Surabaya
10	SMPN 23 Surabaya
11	SMPN 29 Surabaya
12	SMPN 30 Surabaya
13	SMPN 35 Surabaya
14	SMPN 44 Surabaya
15	SMPN 45 Surabaya
16	SMP 4 Muhammadiyah
17	SMP 5 Muhammadiyah
18	SMP 9 Muhammadiyah

Dari daftar sekolah yang dikunjungi tersebut, didapatkan hasil survey seperti berikut:

1. Survey Edukasi Penggunaan *Smartphone* bagi Siswa di Surabaya.

Data yang didapatkan dari hasil survey ini meliputi jawaban pertanyaan seputar penggunaan media sosial dan juga akun media sosial yang dimiliki pelajar tersebut.

Gambar 4.1 berikut merupakan hasil dari data survey Edukasi Penggunaan *Smartphone* bagi Siswa di Surabaya.

The screenshot shows an Excel spreadsheet titled 'EP\_C\_3.xlsx'. The main data area is a table with columns labeled C3, C4, C5, and C6. Each column has a header 'Ket (Lai n-lain )' followed by a series of 1s and 0s. The 'ID IG' column contains names of students. The 'ID TW T' column contains a value '1'. The 'C7' column contains the text 'Tidak diperbolehkan' and 'mencari teman di lupa'.

	BAGIAN C												ID IG	ID TW T	C7										
	C3			C4					C5						C6										
Ket (Lai n-lain )	1	2	3	1	2	3	4	5	1	2	3	4	5	6	1	2	3	4	5	6	ID TW T	Ket (Lai n-lain )	C7		
3																									
4	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	farellaaa	0	Tidak diperbolehkan	
5	0	1	1	0	0	0	0	0	1	0	1	0	0	1	0	1	1	0	0	0	0	finamufarroha	0		
6	1				1	1									1	1						wardani_sna			
7	1				1	1	1	1							1	1	1					folad		folad1	
8	1				1	1									1	1	1					hesty_aprilia13			
9	1	1			1	1									1										
10			1																						
11	1	1			1	1	1								1	1						muhammad panges		mencari teman di	
12	1	1			1	1	1															lanoble_moo	1	1	
13	1				1																	navedaaa			
14	1				1	1																bbm	1		
																							bbm1	1	
																									melihat vidio memasak

Gambar 4.1: Proses validasi akun

Tabel 4.3 berikut merupakan potongan data akun media sosial instagram yang dimiliki oleh pelajar SMP:

**Tabel 4.3:** Potongan akun instagram dari hasil survey ke SMP

No	ID Instagram
1	amsaril
2	thean-rudaly
3	-
4	Helios-Line
5	Sastro Arto P
6	Princesa Angelica
7	-
8	-
9	-
10	aniviolashrti
11	-
12	-
13	-
14	-
15	fransiscanancy_

## 2. Survey Mandiri

Dalam melakukan survey mandiri, penulis menyusun formulir isian untuk mendapatkan informasi mengenai aktivitas pe-  
lajar di tiap jamnya. Pada tabel 4.4 berikut ini merupakan  
formulir yang digunakan dalam survey mandiri:

**Tabel 4.4:** Formulir survey aktivitas harian

Jam	Aktivitas	
	SMP Negeri	SMP Swasta
13:00 WIB	1	1
14:00 WIB	1	1
15:00 WIB	1	3
16:00 WIB	3	2

Dari survey yang telah dilakukan, data yang didapat adalah seperti berikut:

- Akun instagram sebanyak 579 akun, dan didapatkan 290 akun yang aktif.
- Data jam sekolah, belajar, istirahat dari 54 responden (3 responden tiap sekolah).

#### 4.1.2 Crawl Data

Teknik Crawling adalah teknik yang digunakan untuk mengambil data secara online melalui layanan API dan link instagram/akun pelajar yang telah didapatkan dari proses survey sebelumnya. Link/akun instagram yang digunakan adalah sebanyak 290 akun. Data yang akan diambil pada dasarnya merupakan log aktivitas pengguna yang berada di halaman Following dan You. Data yang berhasil terambil kemudian disimpan ke dalam database.

Berikut merupakan kode untuk melakukan *crawl data*

```

1
2 //Bagian disimpannya data ke dalam database
3 function save_time( $row ) {
4   global $db;
5   $waktu_online = date("Y-m-d H:i:s", $row->args->timestamp);
6   $waktu_online2 = date("Y-m-d H:i:s e", $row->args->
7     timestamp);
8   $id_akun = $row->args->profile_id;
9   $tipe_online = cek( $row->args->text );
10
11 $f = $db->f( 'akun', '*', 'WHERE id=?', $id_akun );
12 if ( ! $f ) $db->i( 'akun', 'id', array($id_akun) );
13
14 $f = $db->f( 'waktu_online', '*', 'where waktu_online = ? and
15   id_akun = ?', array( $waktu_online, $id_akun ) );
16 if ( ! $f ) $f = $db->i( 'waktu_online', 'id_akun',
17   waktu_online, tipe_online', array( $id_akun,
18   $waktu_online, $tipe_online) );

```

```

16 echo "\n" . ( $f ? 'berhasil' : 'gagal' ) . "\n\n";
17 echo "$waktu_online $waktu_online2 $sid_akun $stipe_online <
    br>";
18 }
19
20 //Bagian dimana dilakukannya pengambilan data dari
    halaman Following dan You dari instagram (crawling)
21 date_default_timezone_set('Asia/Jakarta');
22
23 $i = new \InstagramAPI\Instagram($debug);
24 $i->setUser($username , $password);
25
26 try {
27     $i->login($force=true);
28 } catch (Exception $e) {
29     echo 'something went wrong '. $e->getMessage() . "\n";
30     exit(0);
31 }
32 try {
33     $helper = null;
34     $followers = [];
35
36     do {
37         if ( is_null($helper) ) {
38             $helper = $i->getFollowingRecentActivity();
39             foreach ($helper->stories as $row ) {
40                 save_time( $row );
41                 # code...
42             }
43
44             $helper = $i->getRecentActivity();
45             foreach ($helper->old_stories as $row ) {
46                 save_time( $row );
47                 # code...
48             }
49
50             echo json_encode($helper); //menampilkan json
51         } else {
52             $helper = $i->getSelfUserFollowers($helper->
                getNextMaxId());
53         }
54
55         // $followers = array_merge($followers , $helper->getUsers
            ());
56     } while (! is_null($helper->getNextMaxId()));
57
58 // echo "My followers: \n";

```

```

59 // foreach ($followers as $follower) {
60 //     echo '- '. $follower->getUsername() . "\n";
61 // }
62 } catch (Exception $e) {
63     echo $e->getMessage();
64 }
65 }
66
67
68
69
70

```

**Kode 4.1:** potongan kode melakukan crawling

## 4.2 Pengolahan Data

Setelah melakukan pengumpulan data, maka tahap selanjutnya adalah melakukan pengolahan data.

### 4.2.1 Penggabungan Data

Pada tahap ini akan dilakukan penggabungan dan penyimpanan data mentah yang berasal dari survey dan crawling kedalam sebuah database. Database yang digunakan adalah MySQL.

**Tabel 4.5:** Data yang digabungkan

No	Data
1	Survey Edukasi Penggunaan Smartphone bagi Siswa SMP di Surabaya
2	Survey mandiri
3	Crawling



### 4.2.2 Data reduksi

Tujuan dari tahap adalah menampilkan atribut data menjadi lebih sedikit / menghilangkan data yang tidak dibutuhkan dalam klasterisasi. Data yang dihasilkan dalam tahap ini diproses dalam database MySQL. Atribut-atribut mentah tersebut terdapat didalam setiap tabel database berikut:

**Tabel 4.6:** Daftar tabel dalam database

No	Nama tabel
1	aktivitas
2	jam_aktivitas
3	time_akun
4	time_hari
5	time_waktu_online

Berikut ini merupakan daftar atribut dalam setiap tabel:

1. Tabel aktivitas

**Tabel 4.7:** Tabel aktivitas

Nama atribut	Tipe Data
id	integer
nama_aktivitas	varchar

2. Tabel jam aktivitas

**Tabel 4.8:** Tabel jam aktivitas

<b>Nama atribut</b>	<b>Tipe Data</b>
id	integer
id_aktivitas	integer
jam	integer
id_sekolah	integer
id_hari	integer

## 3. Tabel time akun

**Tabel 4.9:** Tabel time akun

<b>Nama atribut</b>	<b>Tipe Data</b>
id	integer
username	varchar
id_akun	varchar
gender	varchar
id_sekolah	integer

## 4. Tabel time hari

**Tabel 4.10:** Tabel time hari

<b>Nama atribut</b>	<b>Tipe Data</b>
id	integer
nama_hari	varchar

## 5. Tabel waktu online

**Tabel 4.11:** Tabel waktu online

Nama atribut	Tipe Data
id	integer
id_akun	varchar
waktu_online	datetime
tipe_online	integer

### 4.2.3 Proses Klasterisasi

Pada tahap ini akan dilakukan proses klasterisasi menggunakan metode *K-Modes Clustering*. Data yang diolah dalam proses ini berupa data kategorikal. Klasterisasi akan diimplementasikan menggunakan aplikasi R Studio, untuk lebih lengkapnya akan dijelaskan di BAB 5.

Dalam melakukan *K-Modes Clustering* di R yang perlu digunakan adalah library "klaR".

```

1  install.packages("klaR")
2  library(klaR)
3

```

#### Kode 4.2: library klaR

Kemudian untuk menjalankan perintah klasterisasi K-Modes perintah yang dijalankan adalah seperti berikut:

```

1  cobak <- kmodes(data, 3, iter.max = 10, weighted = FALSE)
2

```

#### Kode 4.3: K-Modes di R

Keterangan:

1. "cobak" adalah nama perintah
2. "kmodes" adalah proses klaster

3. "data" adalah *data set* atau *data frame*
4. "3" adalah inisiasi modes
5. "iter.max" adalah jumlah iterasi maksimal
6. "weighted" adalah jarak dari setiap data

### 4.3 Perancangan Visualisasi Data

Tahap ini adalah melakukan perancangan-perancangan antara lain adalah perancangan database, klaster dan visualisasi website.

#### Perancangan Database

Tahap ini akan dilakukan pemilihan atribut-atribut utama dari database yang mengandung data mentah dengan cara menjalankan *query sql*.

Query yang dilakukan bertujuan untuk mendapatkan atribut-atribut seperti berikut:

**Tabel 4.12:** Atribut hasil query

<b>Nama atribut</b>	<b>Nilai atribut</b>	<b>Keterangan</b>
id_akun	#####	id pengguna
tipe_online	1, 2, 3	1 = like 2 = comment 3 = follow
id_hari	1,2,3,4,5,6,7	1 = Minggu 2 = Senin 3 = Selasa 4 = Rabu 5 = Kamis 6 = Jumat 7 = Sabtu
jam	0 sampai dengan 23	pukul 0 s/d pukul 23
id_aktivitas	1,2,3	1 = Sekolah 2 = Belajar 3 = Istirahat
id_sekolah	1,2	1 = Negeri 2 = Swasta
gender	0,1	0 = Perempuan 1 = Laki-laki

## Perancangan Visualisasi Klaster

Visualisasi klaster akan menggunakan tampilan yang disediakan oleh R Shiny, dimana tampilan ini merupakan salah satu fitur R Studio yang dapat menampilkan hasil olahan data kedalam bentuk html.

## **Perancangan Website**

Pada tahap ini akan dilakukan perancangan visualisasi website. Dimana setelah data selesai diolah, akan dilakukan perancangan website agar informasi mudah dipahami dan dianalisis. Dalam merancang website, metode yang digunakan adalah extreme programming. Dalam website akan ada 2 fitur yaitu, pertama, untuk menampilkan traffic aktivitas pelajar di Instagram, kedua, visualisasi kluster melalui fasilitas R Shiny.

## BAB 5

### IMPLEMENTASI DAN VISUALISASI

Pada bab ini berisi tentang proses pengolahan data yang didapatkan dari tahap rancangan sebelumnya dan proses implementasi data mining dengan menggunakan metode klusterisasi serta pembahasan visualisasi.

#### 5.1 Lingkungan Implementasi

Pada bagian ini dibahas terkait lingkungan pengujian yang digunakan dalam implementasi tugas akhir terkait perangkat yang digunakan baik perangkat keras maupun perangkat lunak. Tabel 5.1 yang berisikan spesifikasi perangkat keras dan perangkat lunak untuk implementasi pada tugas akhir ini.

**Tabel 5.1:** Spesifikasi Perangkat Keras

Perangkat	Spesifikasi
Jenis	Laptop Asus A455L Series
Processor	Intel(R) Core (TM) i3-5005U CPU @ 2.00GHz
RAM	4.00 GB
Hard Disk Drive	500 GB

Kemudian untuk perangkat lunak yang digunakan dalam implementasi model ditunjukkan dalam tabel 5.2.

**Tabel 5.2:** Spesifikasi Perangkat Lunak

Nama Perangkat Lunak	Kegunaan dalam Implementasi
Xampp 3.2.2 dengan PHP 5.6.28	Webserver
RStudio	Tools data mining
Rshiny	Tools Visualisasi
Sublime Text 3	Text Editor
Mozilla Firefox	Web Browser

## 5.2 Crawling

Berikut ini merupakan potongan kode untuk melakukan crawling data.

```

1 <?php
2 //identifikasi aktivitas online
3 function cek( $str ) {
4     if ( strpos($str, 'liked') !== false ) $hasil = 1;
5     if ( strpos($str, 'commented') !== false ) $hasil = 2;
6     if ( strpos($str, 'following') !== false ) $hasil = 3;
7     return $hasil;
8 }

```

**Kode 5.1:** potongan kode crawl (1)

Potongan kode 5.1 merupakan potongan dari kode *crawling*. Bagian tersebut adalah fungsi untuk mengidentifikasi aktivitas *online* yang ada didapatkan dari proses *crawling*.

Fungsi tersebut memiliki alur yakni, mengecek aktivitas *online* yang diterima. Terdapat 3 jenis aktivitas yang diidentifikasi, kemudian jika aktivitas tersebut terambil maka akan diubah menjadi sebuah angka/symbol aktivitas. Berikut merupakan deskripsi dari setiap simbol:

- aktivitas *like* diubah menjadi angka 1



- aktivitas *comment* diubah menjadi angka 2
- aktivitas *follow* diubah menjadi angka 3

```

1
2 include '../vendor/autoload.php';
3 require '../src/Instagram.php';
4 require 'csc_pdo.php';
5
6 ini_set("log_errors", 1);
7 ini_set("error_log", "php-error.log");

```

**Kode 5.2:** potongan kode crawl (2)

Potongan kode 5.2 merupakan sebuah fungsi untuk memanggil *library* dari github instagram dan koneksi pada database MySQL.

```

1
2 $_db = new stdClass;
3 $_db->name = 'partmaps.teenstagram'; // Nama database
4 $_db->user = 'partmaps.alvin'; // username database
5 $_db->pass = 'Coba-123'; //
6
7 $db = new csc_pdo( $_db->user, $_db->pass, $_db->name );
8 $db->px = 'time_';

```

**Kode 5.3:** potongan kode crawl (3)

Potongan kode 5.3 merupakan sebuah fungsi untuk mengkoneksikan proses dengan database MySQL yang terpasang di *cloud server*. Kode tersebut terdiri dari

1. Nama database
2. Nama pengguna database
3. Kata sandi database
4. Membuat koneksi ke database
5. Mengidentifikasi nama tabel database dengan kata awal/*prefix* "time"

```

1
2 function save_time( $row ) {
3     global $db;
4     $waktu_online = date("Y-m-d H:i:s", $row->args->timestamp);

```

```

5  $waktu_online2 = date("Y-m-d H:i:s e", $row->args->
    timestamp);
6  $id_akun = $row->args->profile_id;
7  $tipe_online = cek( $row->args->text );
8
9  $f = $db->f( 'akun', '*', 'WHERE id=?', $id_akun );
10 if ( ! $f ) $db->i( 'akun', 'id', array($id_akun) );

```

**Kode 5.4:** potongan kode crawl (4)

Potongan kode 5.4 merupakan bagian untuk menyimpan data yang masuk berupa waktu aktivitas *online*. Data tersebut diambil dari halaman yang ada pada instagram pada bagian halaman "You" dan "Following". Data waktu tersebut dihitung dan diidentifikasi dari id akun pengguna instagram yang sudah disiapkan sebelumnya.

Potongan kode *crawl* diatas merupakan

```

1
2
3  $f = $db->f( 'waktu_online', '*', 'where waktu_online = ? and
    id_akun = ?', array( $waktu_online, $id_akun ) );
4  if ( !$f ) $f = $db->i( 'waktu_online', 'id_akun',
    waktu_online, tipe_online', array( $id_akun,
    $waktu_online, $tipe_online ) );
5  echo "\n" . ( $f ? 'berhasil' : 'gagal' ) . "\n\n";
6  echo "$waktu_online $waktu_online2 $id_akun $tipe_online <
    br>";
7
8 }

```

**Kode 5.5:** potongan kode crawl (5)

Potongan kode 5.5 merupakan bagian untuk memasukkan data waktu aktivitas *online* kedalam sebuah database dimana data waktu juga diidentifikasi dari id akun yang masuk. Jika proses berhasil maka akan ada tulisan "berhasil" pada hasil *crawling*. Jika gagal maka akan muncul kata "gagal".

```

1
2 /////// CONFIG /////
3 //akun instagram untuk menghubungkan ke tiap akun-akun

```

```

4 $username = 'rahmanhumaira25';
5 $password = '*****';
6 $debug = false;
7 // $debug = true;
8 ///////////////////////////////////////////////////

```

### Kode 5.6: potongan kode crawl (6)

Dalam melakukan proses *crawling*, penulis membuat sebuah akun bernama "rahmanhumaira25" yang digunakan untuk mengikuti dan mengetahui aktivitas *online* yang dilakukan oleh remaja dengan cara *memfollow* setiap akun yang telah disiapkan. Potongan kode 5.6.

```

1
2 date_default_timezone_set('Asia/Jakarta');
3
4 $i = new \InstagramAPI\Instagram($debug);
5 $i->setUser($username, $password);
6
7 try {
8     $i->login($force=true);
9 } catch (Exception $e) {
10     echo 'something went wrong '. $e->getMessage(). "\n";
11     exit(0);
12 }
13 try {
14     $helper = null;
15     $followers = [];
16
17     do {
18         if (is_null($helper)) {
19             $helper = $i->getFollowingRecentActivity();
20             foreach ($helper->stories as $row) {
21                 save_time($row);
22                 # code...
23             }
24
25             $helper = $i->getRecentActivity();
26             foreach ($helper->old_stories as $row) {
27                 save_time($row);
28                 # code...
29             }
30
31             echo json_encode($helper); //menampilkan json
32         } else {

```

```

33     $helper = $i->getSelfUserFollowers($helper->
34     getNextMaxId());
35 }
36 // $followers = array_merge($followers, $helper->getUsers
37 // ());
38 } while (!is_null($helper->getNextMaxId()));
39 // echo "My followers: \n";
40 // foreach ($followers as $follower) {
41 //     echo '- '. $follower->getUsername()."\n";
42 // }
43 } catch (Exception $e) {
44     echo $e->getMessage();
45 }
46
47
48

```

**Kode 5.7:** potongan kode crawl (7)

Potongan kode 5.7 merupakan bagian untuk:

1. Identifikasi zona waktu
2. Koneksi ke API Instagram
3. Menggunakan akun yang sudah disiapkan untuk proses *crawling* "rahmanhumaira25".
4. Masuk kedalam akun "rahmanhumaira25", kemudian mengambil data aktivitas *online* dari akun yang telah di ikuti sebelumnya.

### 5.3 Penjadwalan *Crawling*

Untuk mendapatkan data waktu aktivitas *online* setiap 1 jam, maka dilakukan penjadwalan perintah *crawling* menggunakan layanan *cron job* di *cloud server*. Perintah *cron job* tersebut adalah seperti berikut:

```
1  
2 0 * * * * php -q/home/partmaps/teenstagram.part-maps.com/  
3     examples/PaginationExample2.php  
4
```

### **Kode 5.8:** perintah cron job

Pada potongan kode 5.8, secara teknis perintah *cron job* diatur di pengaturan cPanel (*cloud server*). Pada potongan kode tersebut, waktu pengambilan data telah diatur menjadi setiap satu jam yang disimbolkan menjadi `0 * * * *`. Kemudian setelah mengatur waktu pengambilan data, selanjutnya adalah memasukkan perintah untuk mengambil data dengan cara memanggil *file* yang berisi perintah *crawl* yang sudah dibuat sebelumnya.

## **5.4 Query Atribut Klaster**

Untuk mempermudah dalam melakukan proses klusterisasi, maka data atribut perlu digabungkan.

remaja_smp jam_aktivitas	remaja_smp aktivitas
id : int(11)	id : int(11)
# id_aktivitas : int(11)	# nama_aktivitas : varchar(255)
# jam : int(11)	
# id_sekolah : int(11)	
# id_hari : int(11)	

remaja_smp time_akun	remaja_smp time_waktu_online
id : int(11)	id : int(11)
username : varchar(60)	id_akun : varchar(25)
id_akun : varchar(25)	waktu_online : datetime
gender : varchar(10)	# tipe_online : int(11)
# id_sekolah : int(11)	

remaja_smp time_hari
id : int(11)
nama_hari : varchar(11)

**Gambar 5.1:** Semua data yang tersimpan didalam database MySQL

Dari seluruh data dalam database tersebut akan diambil beberapa atribut yang dibutuhkan untuk klasterisasi.?? Berikut merupakan *query* untuk mengambil atribut yang diperlukan:

```

1
2
3 SELECT
4     w.id_akun ,
5     w.tipe_online ,
6     j.id_hari ,
7     j.jam ,
8     j.id_aktivitas ,
9     k.id_sekolah ,
10    IF (k.gender = 'Laki-laki' ,1,0) AS 'gender'
11
12
13 FROM 'time_waktu_online' w, 'jam_aktivitas' j, 'time_akun' k, '

```

```

14     aktivitas 'a
15     WHERE j.id_aktivitas = a.id
16     AND DAYOFWEEK(w.waktu_online) = j.id_hari
17     AND HOUR(w.waktu_online) = j.jam
18     AND w.id_akun = k.id_akun
19

```

**Kode 5.9:** query pemilihan atribut

Berikut merupakan pengertian dari setiap bagian query (kode 5.9):

1. Menampilkan id akun, tipe online, id hari, jam, id aktivitas, id sekolah dan gender.
2. Menentukan tabel tempat pengambilan atribut, yakni time waktu online, jam aktivitas, time akun dan aktivitas.
3. Setiap aktivitas diidentifikasi berdasarkan id akun.
4. Memisahkan atribut dalam tabel waktu menjadi dua atribut yaitu berdasarkan hari dan jam.

## 5.5 Menyambungkan database dengan R-studio

Tahap ini akan dilakukan penyambungan/pengkoneksian database MySQL dengan R-studio dengan tujuan jika terdapat data masuk maka secara langsung dapat diproses dengan R.

Berikut ini merupakan perintah untuk koneksi database MySQL dengan R-studio:

```

1 #install.packages("dbConnect", type="source")

```

**Kode 5.10:** koneksi database dengan R-studio

Pada potongan kode 5.10 koneksi database dengan R-studio diatas merupakan bagian untuk menginstal paket "dbConnect" yang berisi library untuk menghubungkan database.

```

1 library(RMySQL)
2 library(dbConnect)

```

**Kode 5.11:** koneksi database dengan R-studio

Potongan kode 5.11 merupakan pemanggilan *library* "RMySQL" dan *library* "dbConnect".

```

1 con = dbConnect(MySQL() ,
2 user = 'root', password = '',
3 dbname = 'remaja_smp',
4 hostname='localhost' )
5 dbListTables(con)

```

**Kode 5.12:** koneksi database dengan R-studio

Potongan kode 5.12 merupakan bagian untuk

```

1 #Membuat tabel di R dari MySQL
2 myQuery <- "SELECT
3     w.id_akun ,
4     w.tipe_online ,
5     j.id_hari ,
6     j.jam ,
7     j.id_aktivitas ,
8     k.id_sekolah ,
9     IF (k.gender = 'Laki-laki ',1,0) AS 'gender '
10
11
12     FROM 'time_waktu_online' w, 'jam_aktivitas' j, 'time_akun'
13     k, 'aktivitas' a
14
15     WHERE j.id_aktivitas = a.id
16     AND DAYOFWEEK(w.waktu_online) = j.id_hari
17     AND HOUR(w.waktu_online) = j.jam
18     AND w.id_akun = k.id_akun"
19 df <- dbGetQuery(con , myQuery)
20 str(df)
21

```

**Kode 5.13:** koneksi database dengan R-studio

Potongan kode query 5.13 merupakan query untuk menampilkan atribut yang diperlukan untuk klusterisasi.



## 5.6 Klasterisasi

Pada tahap ini akan dilakukan klasterisasi dengan metode K-Modes Clustering, yang dijalankan di aplikasi R-studio.

Beberapa tahap sebelum klasterisasi adalah sebagai berikut:

1. Membuat *data frame* dengan nama remaja

```
1   remaja <- data.frame(df)
2
3
```

### Kode 5.14: Data frame

Pada data frame awal remaja terdapat 7 atribut yaitu id akun, tipe online, id hari, jam, id aktivitas, id sekolah, gender.

2. Menghilangkan atribut id akun karena tidak dibutuhkan dalam proses klaster

```
1   #Menghilangkan kolom 1 (id akun)
2   remaja <- data.frame(remaja[-1])
3
```

### Kode 5.15: menghilangkan atribut id akun

3. Melakukan *K-Modes Clustering*

```
1
2   kr3 <- kmodes(remaja, 3, iter.max = 10, weihthed=FALSE)
3
4
```

### Kode 5.16: K-Modes

4. Menampilkan plot hasil klasterisasi

```
1
2   plot(remaja, col=ct3$cluster, pch = 19)
3
4
```

### Kode 5.17: plot klaster

## 5.7 Uji Coba

Pada bagian ini akan dilakukan percobaan klusterisasi dengan jumlah kluster secara acak yang bertujuan untuk mengukur jumlah kluster mana yang lebih baik. Percobaan yang dilakukan adalah menggunakan jumlah kluster yaitu 3, 4, 5, 6, dan 7.

### 5.7.1 K=3

Pada percobaan yang pertama dilakukan klusterisasi K-Modes dengan jumlah jumlah modes yaitu 3.

```

1 #klusterisasi dengan jumlah k = 3
2 kr3 <- kmodes(remaja , 3, iter.max = 10, weighted=FALSE)
3
4 #plot hasil kluster
5 plot(remaja , col=kr3$cluster , pch=19)
6
7 #membuat file csv untuk nilai withindiff
8 write.csv(kr3$withindiff , file = "kr3_withindiff.csv")
9
10 #membuat file csv untuk nilai size
11 write.csv(kr3$size , file = "kr3_size.csv")
12
13

```

**Kode 5.18:** uji coba klusterisasi dengan jumlah k

## 5.8 Klasifikasi

Pada tahap ini akan dilakukan proses klasifikasi dari data yang sudah memiliki nilai kluster. Sehingga data baru yang masuk dapat teridentifikasi akan masuk ke kluster x. Berikut ini adalah proses klasifikasi menggunakan metode Naive Bayes. Dalam melakukan klasifikasi, data di pecah menjadi 2 bagian dengan perbandingan

70 persen dan 30 persen. 70 persen merupakan data untuk proses *training* dan data 30 persen untuk data *testing*.

```

1 #penggunaan library
2 library(NLP)
3 library(e1071)
4
5 #pembagian data klasifikasi
6 remajatrain <- data.frame(kkr3)
7 train_bagi <- floor(0.7*nrow(remajatrain))
8 klasremaja <- sample(seq_len(nrow(remajatrain)), size =
      train_bagi)
9 train_remaja <- remajatrain[klasremaja,]
10 test_remaja <- remajatrain[-klasremaja,]
11
12 #data training
13 atr <- train_remaja[,-7]
14 btr <- as.character(train_remaja[,7])
15
16 #data testing
17 ate <- test_remaja[,-7]
18 bte <- as.character(test_remaja[,7])
19
20 #pembuatan model klasifikasi Naive Bayes
21 model <- naiveBayes(atr, factor(btr))
22
23 #melakukan prediksi
24 prediksi <- predict(model, ate)
25 aktual <- bte
26 bandingin <- cbind(prediksi, aktual)
27
28 #mengukur akurasi klasifikasi
29 akurasi <- 100*sum(prediksi==aktual)/length(prediksi)
30 akurasi
31
32

```

**Kode 5.19:** metode naive bayes untuk mengetahui persentase klasifikasi data

## 5.9 Visualisasi

Pada tahap ini dilakukan pembangunan visualisasi dengan menggunakan R-studio dengan *package* R-shiny.

```

1 #server.R
2 library(shiny)
3 library(shinydashboard)
4 library(RMySQL)
5 library(dbConnect)
6
7 con = dbConnect(MySQL(),
8 user = 'root',
9 password = '',
10 dbname = 'remaja_smp',
11 hostname = 'localhost')
12
13 myQuery <- "Select
14             w.id_akun,
15             w.tipe_online ,
16             j.id_hari ,
17             j.jam ,
18             i.id_aktivitas ,
19             k.id_sekolah ,
20             IF (k.gender ='Laki-laki ',1,0) AS 'gender '
21 FROM
22     'time_waktu_online 'w,
23     'jam_aktivitas 'j ,
24     'time_akun 'k,
25     'aktivitas 'a
26
27     WHERE j.id_aktivitas = a.id
28     AND DAYOFWEEK(w.waktu_online) = j.id_hari
29     AND HOUR(w.waktu_online) = j.jam
30     AND w.id_akun = k.id_akun"
31
32 df <- dbGetQuery(con , myQuery)
33 str(df)
34
35 remaja <- data.frame(df)
36 remaja <- remaja[-1]
37 library(klaR)
38
39 function(input , output , session) {
40
41     selectedData <- reactive({

```

```

42     remaja[, c(input$xcoll, input$ycol)]
43   })
44
45   clusters <- reactive({
46     kmodes(selectedData(), input$clusters)
47   })
48
49
50   output$plot1 <- renderPlot({
51     palette(c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3",
52             "#FF7F00" ))
53
54     par(mar = c(5.1, 4.1, 0, 1))
55     plot(selectedData(),
56          col = clusters()$cluster,
57          pch = 20, cex = 3)
58     points(clusters()$centers, pch = 4, cex = 4, lwd = 4)
59   })
60
61
62

```

### Kode 5.20: server.R

Potongan kode 5.20 merupakan kode untuk membuat fungsi visualisasi website. Kode tersebut terdiri dari:

1. Pemanggilan *library*.
2. Koneksi database MySQL dengan R-studio.
3. Query menampilkan atribut klasterisasi.
4. Pembuatan data frame.
5. Membuat fungsi *input data* atribut.
6. Melakukan proses klasterisasi.
7. Menampilkan plot dan juga memberikan warna setiap plot klaster.

```

1 #ui.R
2 library(shiny)
3 library(shinydashboard)
4
5 shinyUI(
6   dashboardPage(

```

```

7 dashboardHeader(title = "Teenstagram"),
8 dashboardSidebar(
9   sidebarMenu(
10    menuItem("Dashboard",
11     tabName = "dashboard",
12     icon = icon("dashboard")),
13    menuItem("Analysis",
14     tabItem = "analisis",
15     icon = icon("hourglass")),
16    menuItem("RawData",
17     tabName = "raw",
18     icon = icon("database"))
19  ),
20 ),
21 dashboardBody(
22   tabItems(
23     tabItem(tabName = "dashboard",
24      fluidRow(
25        pageWithSidebar(
26          headerPanel('k-modes clustering'),
27          sidebarPanel(
28            selectInput('xcol', 'X Variable', names(remaja)),
29            selectInput('ycol', 'Y Variable', names(remaja)),
30            selected=names(remaja)[[2]],
31            numericInput('clusters', 'Cluster count', 3,
32             min = 1, max = 5)
33          ),
34          mainPanel(
35            plotOutput('plot1')
36          )
37        )
38      )
39    )
40  )
41 )
42 )
43 )
44 ))
45
46
47

```

### Kode 5.21: ui.R

Potongan kode 5.21 merupakan kode untuk menampilkan plot hasil proses klusterisasi dan menginputkan atribut data.

Pada tahap menampilkan hasil visualisasi kedalam HTML. Tampilan yang dihasilkan dari R-shiny di masukkan kedalam kode berikut:

```
1 <li><a href="http://127.0.0.1:5348/">Analisis Kmodes Cluster<
  /a></li>
2
```

**Kode 5.22:** web teenstagram dari HTML

Selain menampilkan data dalam bentuk hasil klasterisasi, website juga dirancang untuk menampilkan grafik berupa *traffic online* aktivitas remaja di instagram yang dibuat dengan potongan kode berikut.

```
1
2 $_db = new stdClass;
3 $_db->name = 'remaja_smp';
4 $_db->user = 'root';
5 $_db->pass = ''; //
6 $db = new csc_pdo( $_db->user , $_db->pass , $_db->name );
```

**Kode 5.23:** kode PHP untuk grafik garis aktivitas online (1)

Potongan kode 5.23 merupakan bagian untuk terhubung ke database MySQL sehingga data dapat diolah dan ditampilkan.

```
1
2 $f = $db->r( 'time_waktu_online', 'hour( waktu_online ) as h
  , count( id ) as n', 'GROUP BY hour( waktu_online )' );
3 foreach ( $f as $r ) {
4 # code ...
5 $isi_satu[] = $r['n'];
6 $pukul = $r['h'];
7 $label[] = "'Pukul $pukul'";
8 }
9
10 $f = $db->r( 'time_waktu_online', 'hour( waktu_online ) as h
  , count( id ) as n', 'WHERE dayname(waktu_online) != ?
  AND dayname(waktu_online) != ? GROUP BY hour(
  waktu_online )', array( 'Saturday', 'Sunday' ) );
11 foreach ( $f as $r ) {
12 # code ...
13 $isi_duwa[] = $r['n'];
14 }
```

15  
16**Kode 5.24:** kode PHP untuk grafik garis aktivitas online (2)

Potongan kode 5.24 merupakan bagian untuk mengambil data yang ada dalam database, serta untuk menampilkan data yang diinginkan kedalam halaman website.

```

1 <div>
2 <canvas id="myChart" width="250" height="250"></canvas>
3
4 </div>
5
6 <script src="https://cdnjs.cloudflare.com/ajax/libs/Chart.js
7 /2.4.0/Chart.min.js"></script>
8 <script>
9 var ctx = document.getElementById("myChart");
10 var myLineChart = new Chart(ctx, {
11   type: 'line',
12   data: {
13     labels: [<?=implode( ' ', $label ) ?>],
14     datasets: [
15       {
16         label: '# Setiap Hari',
17         data: [<?=implode( ' ', $isi_satu ) ?>],
18         /* backgroundColor: [
19           //'rgba(255, 99, 132, 0.2)',          // merubah
20         warna
21           'pink',          // merubah warna
22         ],*/
23         borderColor: [
24           'rgba(255,99,132,1)',
25         ],
26         borderWidth: 1*/
27       },
28       {
29         label: '# Hari Sekolah',
30         data: [<?=implode( ' ', $isi_duwa ) ?>],
31         /* backgroundColor: [
32           //'rgba(255, 99, 132, 0.2)',          // merubah
33         warna
34           'lightblue',          // merubah warna
35         ],*/
36       }
37     ]
38   }
39 }
40 ]

```



```
35 },
36 options: {
37     responsive: true ,
38     maintainAspectRatio: false ,
39     /* scales: {
40         yAxes: [{
41             ticks: {
42                 beginAtZero: true
43             }
44         }]
45     }*/
46 } });
47 </script >
48
49
50
51
```

**Kode 5.25:** kode PHP untuk grafik garis aktivitas online (3)

Potongan kode 5.25 merupakan bagian untuk menampilkan data ke-dalam sebuah tampilan grafik garis. Dalam grafik garis juga dapat ditentukan antara hari aktif sekolah atau setiap hari.

*Halaman ini sengaja dikosongkan*

## **BAB 6**

### **HASIL DAN PEMBAHASAN**

Pada bab ini akan dijelaskan hasil dan pembahasan dari proses pengujian aplikasi.

#### **6.1 Hasil**

Pada bagian ini akan dijelaskan mengenai hasil uji coba aplikasi yang telah dilakukan pada bab 5.

##### **6.1.1 *Crawling***

Berikut ini merupakan hasil dari perintah crawling data waktu aktivitas online pengguna instagram.

2017-01-31 19:12:17 Asia/Jakarta 2049436713 1  
2017-01-31 19:11:53 Asia/Jakarta 2049436713 1  
2017-01-30 17:37:24 Asia/Jakarta 1912899283 3  
2017-01-29 14:07:11 Asia/Jakarta 4532334020 3  
2017-01-26 15:15:08 Asia/Jakarta 1555712578 2  
2017-01-26 15:14:27 Asia/Jakarta 1555712578 3  
2017-01-26 15:14:06 Asia/Jakarta 1555712578 2  
2017-01-21 23:55:18 Asia/Jakarta 12537 3  
2017-01-20 15:19:44 Asia/Jakarta 4482861641 3  
2017-01-19 21:46:40 Asia/Jakarta 4481724045 3  
2017-01-18 17:58:20 Asia/Jakarta 3782172993 3  
2017-01-16 18:16:47 Asia/Jakarta 1565540476 3  
2017-01-15 21:44:37 Asia/Jakarta 1503763156 3  
2017-01-14 23:04:07 Asia/Jakarta 3270518355 3  
2017-01-14 21:14:32 Asia/Jakarta 4463863736 3  
2017-01-13 05:08:37 Asia/Jakarta 3021499509 3  
2017-01-12 17:57:33 Asia/Jakarta 2300404077 3  
2017-01-11 14:02:09 Asia/Jakarta 1978759078 3  
2017-01-10 20:30:17 Asia/Jakarta 4016774347 3  
2017-01-10 17:04:28 Asia/Jakarta 2190396324 3  
2017-01-10 15:27:54 Asia/Jakarta 3093719780 3  
2017-01-09 16:00:50 Asia/Jakarta 1068811516 3

**Gambar 6.1:** Hasil crawling

Bagian dari setiap data pada gambar 6.1 adalah

1. Tanggal
2. Jam, menit, detik
3. Zona waktu
4. id akun instagram
5. Kode aktivitas online
  - 1 adalah *like*
  - 2 adalah *comment*
  - 3 adalah *follow*

### 6.1.2 Menyimpan Hasil *Crawling*

Data waktu aktivitas *online* yang berhasil di *crawling* dimasukkan kedalam database MySQL. Gambar 6.2









+ Options			id	id_akun	waktu_online	tipe_online
<input type="checkbox"/>	Edit  Copy  Delete	2	2081647444	2017-03-27 21:36:58	1	
<input type="checkbox"/>	Edit  Copy  Delete	3	1787992158	2017-03-27 21:36:38	3	
<input type="checkbox"/>	Edit  Copy  Delete	4	467862193	2017-03-27 21:36:16	1	
<input type="checkbox"/>	Edit  Copy  Delete	5	1559887778	2017-03-27 21:35:06	1	
<input type="checkbox"/>	Edit  Copy  Delete	6	1559887778	2017-03-27 21:34:06	1	
<input type="checkbox"/>	Edit  Copy  Delete	7	503916510	2017-03-27 21:32:35	1	
<input type="checkbox"/>	Edit  Copy  Delete	8	1768172441	2017-03-27 21:30:44	1	
<input type="checkbox"/>	Edit  Copy  Delete	9	1103271414	2017-03-27 21:30:16	1	
<input type="checkbox"/>	Edit  Copy  Delete	10	205350913	2017-03-27 21:29:52	1	
<input type="checkbox"/>	Edit  Copy  Delete	11	2147483647	2017-03-27 21:29:47	1	
<input type="checkbox"/>	Edit  Copy  Delete	12	2147483647	2017-03-27 21:29:39	1	
<input type="checkbox"/>	Edit  Copy  Delete	13	2147483647	2017-03-27 21:29:13	1	
<input type="checkbox"/>	Edit  Copy  Delete	14	2147483647	2017-03-27 21:29:03	1	
<input type="checkbox"/>	Edit  Copy  Delete	15	1998623873	2017-03-27 21:29:00	3	
<input type="checkbox"/>	Edit  Copy  Delete	16	2147483647	2017-03-27 21:28:36	1	
<input type="checkbox"/>	Edit  Copy  Delete	17	2147483647	2017-03-12 21:45:17	3	
<input type="checkbox"/>	Edit  Copy  Delete	18	2147483647	2017-03-12 08:33:48	1	
<input type="checkbox"/>	Edit  Copy  Delete	19	2147483647	2017-03-07 15:27:01	3	
<input type="checkbox"/>	Edit  Copy  Delete	20	2147483647	2017-02-22 19:58:17	3	
<input type="checkbox"/>	Edit  Copy  Delete	21	2049436713	2017-01-31 19:12:17	1	
<input checked="" type="checkbox"/>	Console  Copy  Delete	22	2049436713	2017-01-31 19:11:53	1	

**Gambar 6.2:** Hasil *crawling* di simpan di database

### 6.1.3 Penjadwalan *Crawling*

Berikut merupakan hasil perintah *cron job* yang digunakan untuk melakukan perintah *crawling* setiap 1 jam. Gambar 6.3

Cron <partmaps@enigma> php -q /home/partmaps/teenstagram.part-maps.com /examples/PaginationExample2.php Inbox x

	(Cron Daemon) berhasil 2017-07-04 09:00:10 2017-07-04 09:00:10 Asia/Jakarta 1534179120 1 <b...	Jul 3 (2 days ago) ☆
	(Cron Daemon) berhasil 2017-07-04 09:59:54 2017-07-04 09:59:54 Asia/Jakarta 1571212642 1 <b...	Jul 3 (2 days ago) ☆
	(Cron Daemon) berhasil 2017-07-04 11:00:00 2017-07-04 11:00:00 Asia/Jakarta 1491062223 1 <b...	Jul 3 (2 days ago) ☆
	(Cron Daemon) berhasil 2017-07-04 12:00:08 2017-07-04 12:00:08 Asia/Jakarta 2014311953 1 <b...	Jul 3 (2 days ago) ☆
	(Cron Daemon) berhasil 2017-07-04 13:00:05 2017-07-04 13:00:05 Asia/Jakarta 1689028087 1 <b...	Jul 3 (2 days ago) ☆
	(Cron Daemon) berhasil 2017-07-04 13:59:59 2017-07-04 13:59:59 Asia/Jakarta 2231544950 1 <b...	12:00 am (1 day ago) ☆
	(Cron Daemon) berhasil 2017-07-04 15:00:04 2017-07-04 15:00:04 Asia/Jakarta 1103271414 3 <b...	1:00 am (1 day ago) ☆
	(Cron Daemon) berhasil 2017-07-04 16:00:06 2017-07-04 16:00:06 Asia/Jakarta 1968811546 1 <b...	2:00 am (1 day ago) ☆

**Gambar 6.3:** Penjadwalan *crawling* setiap satu jam

### 6.1.4 Data terkumpul

Tabel 6.1 berikut merupakan data yang didapatkan dari metode *crawling* dan *survey*:

**Tabel 6.1:** Data mentah

<b>Kegiatan</b>	<b>Data</b>
Crawling	200209 aktivitas online
Survey Etika Profesi	290 akun instagram
Survey Aktivitas Harian	54 responden (3 responden/sekolah)

### 6.1.5 Hasil *query* pemilihan atribut

Tabel 6.2 berikut merupakan hasil *query* MySQL dalam menampilkan atribut yang diperlukan untuk proses klusterisasi.

**Tabel 6.2:** Gabungan data atribut

<b>id akun</b>	<b>tipe online</b>	<b>id hari</b>	<b>jam</b>	<b>id aktivitas</b>	<b>id sekolah</b>	<b>gender</b>
24147483647	3	2	0	3	1	0
24147483647	1	2	0	3	1	0
24147483647	1	2	0	3	1	0
24147483647	1	2	0	3	1	0
24147483647	3	2	0	3	1	0
24147483647	1	2	1	3	1	0
24147483647	3	2	1	3	1	0

### 6.1.6 Penyambungan MySQL dengan R-studio

Gambar 6.4 erikut ini merupakan tampilan bahwa database MySQL sudah tersambung dengan R-studio.

	tipe_online	id_hari	jam	id_aktivitas	id_sekolah	gender
1	3	2	0	3	1	0
2	1	2	0	3	1	0
3	1	2	0	3	1	0
4	1	2	0	3	1	0
5	1	2	0	3	1	0
6	3	2	0	3	1	0
7	1	2	0	3	1	0
8	1	2	0	3	1	0
9	1	2	0	3	1	0
10	1	2	0	3	1	0
11	3	2	0	3	1	0
12	1	2	0	3	1	0
13	1	2	0	3	1	0
14	1	2	0	3	1	0
15	1	2	0	3	1	0
16	3	2	0	3	1	0
17	1	2	0	3	1	0
18	1	2	0	3	1	0
19	1	2	0	3	1	0
20	1	2	0	3	1	0
21	3	2	0	3	1	0
22	3	2	1	3	1	0

Showing 1 to 23 of 38,008 entries

**Gambar 6.4:** Hasil crawling

Tabel 6.3 merupakan keterangan dari setiap nilai atribut:

**Tabel 6.3:** Data atribut

Atribut	Simbol	Keterangan
Tipe Online	1,2,3	- Like = 1 - Comment = 2 - Follow = 3



Hari	1,2,3,4,5,6,7	<ul style="list-style-type: none"> <li>- Minggu = 1</li> <li>- Senin = 2</li> <li>- Selasa = 3</li> <li>- Rabu = 4</li> <li>- Kamis = 5</li> <li>- Jumat = 6</li> <li>- Sabtu = 7</li> </ul>
Jam	Pukul 00:00 s/d 23:00 WIB	<ul style="list-style-type: none"> <li>- 00:00 = 0</li> <li>- 01:00 = 1</li> <li>- 02:00 = 2</li> <li>- 03:00 = 3</li> <li>- 04:00 = 4</li> <li>- 05:00 = 5</li> <li>- 06:00 = 6</li> <li>- 07:00 = 7</li> <li>- 08:00 = 8</li> <li>- 09:00 = 9</li> <li>- 10:00 = 10</li> <li>- 11:00 = 11</li> <li>- 12:00 = 12</li> <li>- 13:00 = 13</li> <li>- 14:00 = 14</li> <li>- 15:00 = 15</li> <li>- 16:00 = 16</li> <li>- 17:00 = 17</li> <li>- 18:00 = 18</li> <li>- 19:00 = 19</li> <li>- 20:00 = 20</li> <li>- 21:00 = 21</li> <li>- 22:00 = 22</li> <li>- 23:00 = 23</li> </ul>

Aktivitas	1,2,3	Aktivitas harian. - Jam sekolah = 1 - Jam belajar = 2 - Jam istirahat = 3
Jenis sekolah	1,2	- Negeri = 1 - Swasta = 2
Gender	0,1	Jenis kelamin - Perempuan = 0 - Laki-laki = 1

Tabel 6.4 merupakan daftar jumlah komponen tiap atribut:

**Tabel 6.4:** Jumlah komponen atribut

Atribut	Jumlah Komponen Atribut
Tipe Online	- Like = 22565 - Comment = 2 - Follow = 15443
Hari	- Minggu = 6016 - Senin = 3616 - Selasa = 10148 - Rabu = 5660 - Kamis = 3918 - Jumat = 4532 - Sabtu = 4118

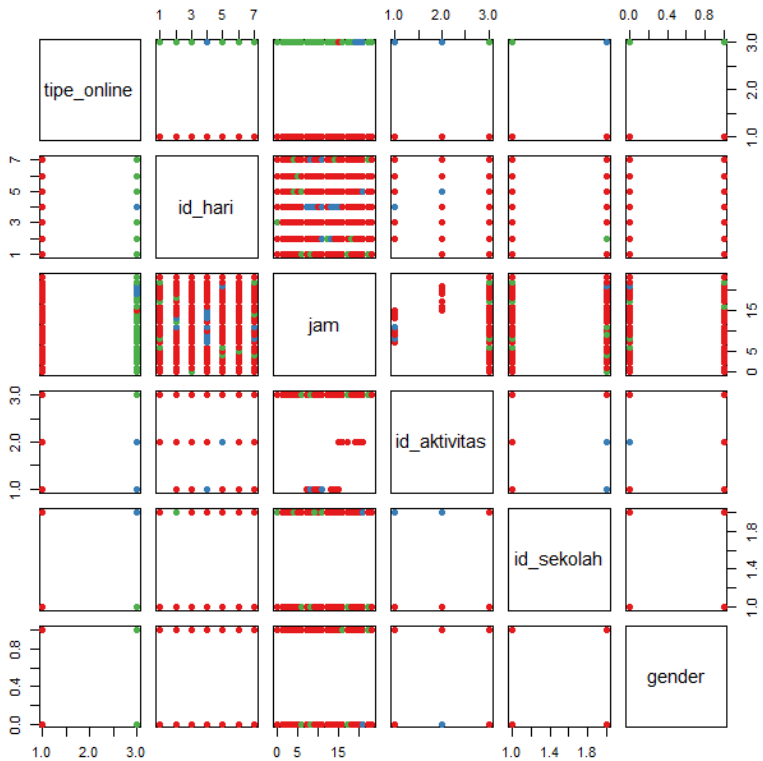
Jam	<ul style="list-style-type: none"> <li>- 00:00 = 700</li> <li>- 01:00 = 280</li> <li>- 02:00 = 192</li> <li>- 03:00 = 1164</li> <li>- 04:00 = 360</li> <li>- 05:00 = 1574</li> <li>- 06:00 = 724</li> <li>- 07:00 = 602</li> <li>- 08:00 = 2568</li> <li>- 09:00 = 550</li> <li>- 10:00 = 592</li> <li>- 11:00 = 646</li> <li>- 12:00 = 568</li> <li>- 13:00 = 618</li> <li>- 14:00 = 4022</li> <li>- 15:00 = 7794</li> <li>- 16:00 = 722</li> <li>- 17:00 = 2790</li> <li>- 18:00 = 622</li> <li>- 19:00 = 1636</li> <li>- 20:00 = 1512</li> <li>- 21:00 = 4188</li> <li>- 22:00 = 1754</li> <li>- 23:00 = 1830</li> </ul>
Aktivitas	<p>Aktivitas harian.</p> <ul style="list-style-type: none"> <li>- Jam sekolah = 10938</li> <li>- Jam belajar = 6838</li> <li>- Jam istirahat = 2037</li> </ul>
Jenis sekolah	<ul style="list-style-type: none"> <li>- Negeri = 35971</li> <li>- Swasta = 2037</li> </ul>

Gender	Jenis kelamin - Perempuan = 35371 - Laki-laki = 2637
--------	--

### 6.1.7 Klaterisasi

Gambar 6.5 merupakan tampilan plot hasil klasterisasi dengan jumlah  $k$  adalah 3.

1. Plot  $k=3$



**Gambar 6.5:** Hasil plot  $k=3$

## Uji Akurasi Kluster

Tabel 6.5 merupakan nilai *withindiff* setiap kluster. Nilai tersebut merupakan jarak obyek kluster ke setiap kluster.

1. Jumlah *withindiff* tiap Kluster

**Tabel 6.5:** Withindiff

<b>K</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
3	37416	19628	16443				
4	34939	15117	3668	16158			
5	27852	13292	7157	11441	4090		
6	15079	19855	10430	2821	3400	7515	
7	28604	11200	9082	4083	1433	2403	1759

Pada tabel 6.6 merupakan jumlah withindiff setiap klaster.

**Tabel 6.6:** Jumlah Withindiff

<b>K</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>Jumlah</b>
3	37416	19628	16443					<b>73487</b>
4	34939	15117	3668	16158				<b>69882</b>
5	27852	13292	7157	11441	4090			<b>63832</b>
6	15079	19855	10430	2821	3400	7515		<b>59100</b>
7	28604	11200	9082	4083	1433	2403	1759	<b>58564</b>

## 2. Jumlah *size* tiap Klaster

**Tabel 6.7:** Size

<b>K</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
3	16507	11193	10308				
4	16488	9658	1619	10243			
5	15703	7953	5507	6383	2456		
6	9682	11878	6543	1334	3340	5231	
7	16649	6871	6174	3515	957	1186	2656

**Tabel 6.8:** Total Size

<b>K</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>Jumlah</b>
3	16507	11193	10308					<b>38008</b>
4	16488	9658	1619	10243				<b>38008</b>
5	15703	7953	5507	6383	2456			<b>38008</b>
6	9682	11878	6543	1334	3340	5231		<b>38008</b>
7	16649	6871	6174	3515	957	1186	2656	<b>38008</b>

3. Akurasi Tahap berikut ini merupakan proses pengukuran akurasi klaster.

**Tabel 6.9:** Total Withindiff

	<b>K=3</b>	<b>K=4</b>	<b>K=5</b>	<b>K=6</b>	<b>K=7</b>
1	2.266675	2.119056	1.773674	1.557426	1.718061
2	1.753596	1.565231	1.671319	1.671578	1.630039
3	1.595169	2.265596	1.299619	1.59407	1.471007
4		1.577468	1.790734	2.114693	1.161593
5			1.665309	1.017964	1.497388
6				1.436628	2.026138
7					0.662274
<b>Rata-rata</b>	<b>1.871813</b>	<b>1.881838</b>	<b>1.640131</b>	<b>1.565393</b>	<b>1.452357</b>

### 6.1.8 Klasifikasi

Gambar 6.6 merupakan percobaan klasifikasi dengan hasil akurasi 82.43445

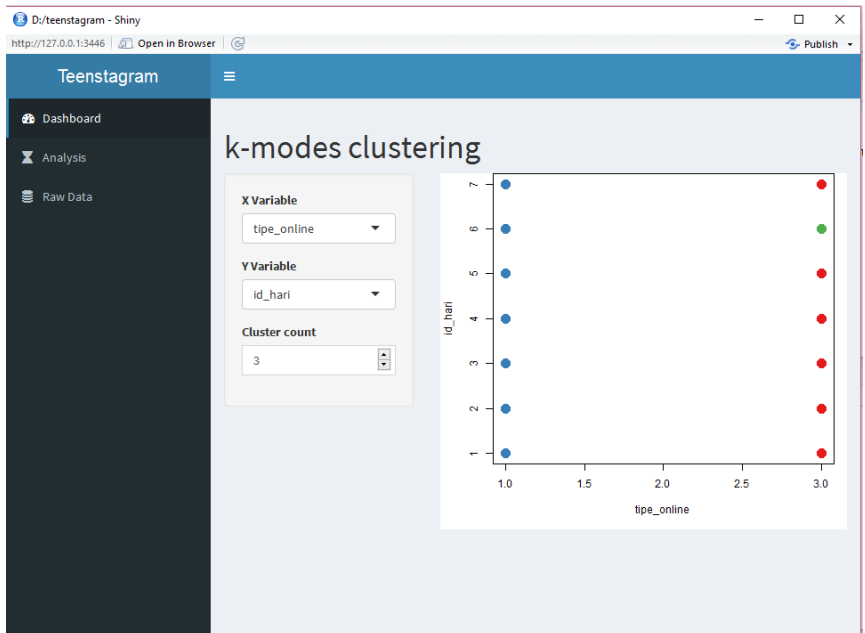
```
> remajatrain <- data.frame(kkr3)
> train_bagi <- floor(0.7*nrow(remajatrain)) #membagi data training 70%
> klasremaja <- sample(seq_len(nrow(remajatrain)), size = train_bagi) #index
> train_remaja <- remajatrain[klasremaja,]
> test_remaja <- remajatrain[-klasremaja,]
>
> atr <- train_remaja[,-7]
> btr <- as.character(train_remaja[,7])
>
> ate <- test_remaja[,-7]
> bte <- as.character(test_remaja[,7])
>
> model <- naiveBayes(atr, factor(btr))
>
> prediksi <- predict(model, ate)
> aktual <- bte
> bandingin <- cbind(prediksi, aktual)
> akurasi <- 100*sum(prediksi==aktual)/length(prediksi)
> akurasi
[1] 82.43445
```

**Gambar 6.6:** Hasil akurasi klasifikasi

### 6.1.9 Visualisasi

Gambar 6.7 berikut ini adalah tampilan utama dari website





**Gambar 6.7:** Hasil visualisasi Teenstagram

Gambar diatas merupakan tampilan dari hasil klasterisasi atribut menggunakan K-Modes. Dari gambar tersebut terdapat 3 warna yang mewakili jumlah jumlah klaster.

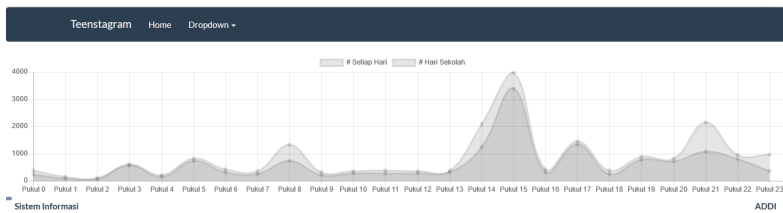
1. Biru klaster A
2. Merah klaster B
3. Hijau klaster C

Jika dicermati, visualisasi diatas dapat diambil informasi bahwa klaster A didominasi oleh aktivitas *like* yang muncul setiap hari (Minggu, Senin, Selasa, Rabu, Kamis, Jumat, Sabtu), kemudian klaster B didominasi oleh aktivitas *follow* pada hari Minggu, Senin, Selasa, Rabu, Kamis, dan Sabtu. Sedangkan klaster C terdiri dari

aktivitas *follow* pada hari jumat.

### 6.1.10 Visualisasi PHP

Berikut merupakan tampilan website dengan PHP yang berfungsi untuk menampilkan data *traffic* aktivitas *online* remaja di instagram.



**Gambar 6.8:** Visualisasi *traffic* aktivitas remaja di instagram

## 6.2 Pembahasan Hasil

Pada subbab ini akan dibahas dan disimpulkan hasil dari pengujian fungsional perangkat lunak.

### 6.2.1 Pengujian Fungsional

Pada proses uji fungsional didapatkan beberapa fungsi yang dapat berjalan adalah sebagai berikut:

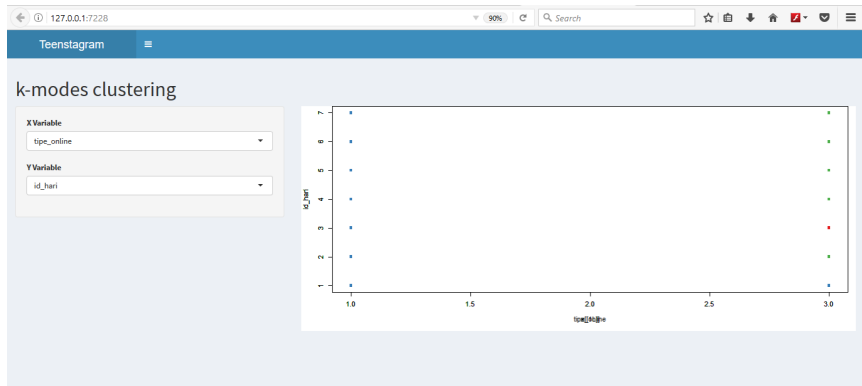
1. Fungsi Menampilkan Plot Hasil Klasterisasi

Tampilan dashboard dapat menampilkan hasil klasterisasi dalam tampilan plot. Tampilan plot terdiri dari varian warna

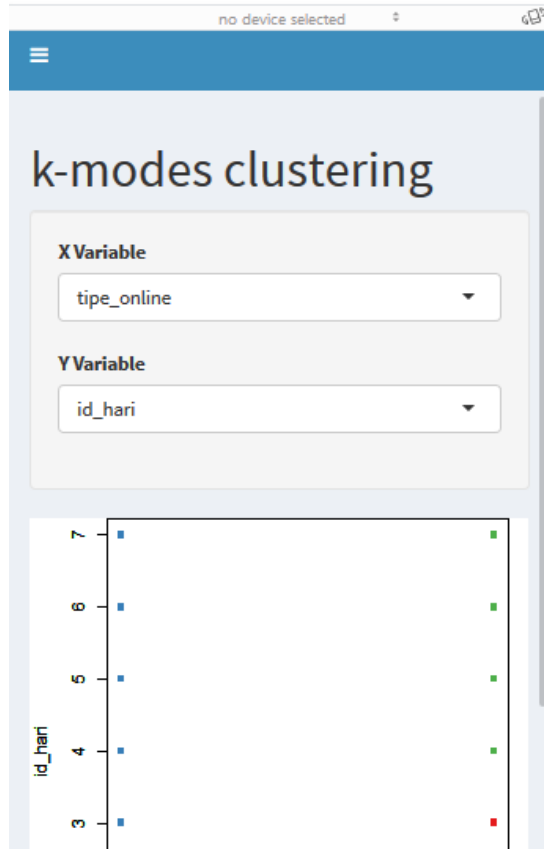
- sesuai dengan jumlah klaster yang dipilih.
2. Pengguna dapat memilih persilangan antara 2 atribut untuk melihat hasil plot dari klasterisasi. Atribut yang dimiliki oleh data sebanyak 6 atribut, yang dimana setiap atributnya dapat disilangkan dengan atribut yang lain untuk dilihat pola dari kedua atribut.
  3. Pengguna dapat memilih jumlah klaster untuk melihat perubahan plot. Jumlah klaster yang terdapat pada aplikasi secara berurutan terdiri 3,4,5,6 dan 7. Sehingga akan ada 6 versi tampilan visualisasi berdasarkan jumlah klasternya.

## 6.2.2 Pengujian non-Fungsional

1. Fungsi Crawling  
Crawling dapat berjalan setiap 1 jam. Data crawling yang berhasil diambil akan ditunjukkan dengan notifikasi yang masuk di email. Kemudian data crawling berhasil dimasukkan pada database.
2. Tampilan website  
Pengujian non-fungsional dari website yakni dengan cara mengakses website dalam tampilan komputer dan *handphone*. Berikut ini merupakan uji tampilan website yang diakses menggunakan komputer (gambar 6.9) dan menggunakan (*gambarhandphone*) 6.10



**Gambar 6.9:** Tampilan dari komputer



**Gambar 6.10:** Tampilan dari *handphone*

Dari hasil visualisasi yang ditunjukkan oleh gambar 6.9 dan 6.10, website dapat secara baik ditampilkan melalui media komputer maupun *handphone*.

*Halaman ini sengaja dikosongkan*

## BAB 7

### KESIMPULAN DAN SARAN

Pada bab ini akan dijelaskan kesimpulan dan saran dalam pengerjaan tugas akhir.

#### 7.1 Kesimpulan

Dari hasil penelitian

1. Dalam melakukan penelitian terkait visualisasi pola aktivitas remaja (siswa SMP), sumber data survey terdiri dari 2 metode yaitu survey edukasi penggunaan *smartphone* bagi siswa SMP di Surabaya dan survey mandiri yang dilakukan penulis untuk mendapatkan data waktu aktivitas harian.
2. Dari hasil survey didapatkan sebanyak 579 siswa mengisikan akun instagram dan setelah melalui tahap validasi, didapatkan 290 akun instagram yang valid/tidak palsu.
3. Atribut/variabel yang digunakan dalam penelitian ini terdiri dari 6 atribut. Yaitu tipe aktivitas online, hari, jam, aktivitas harian, jenis sekolah, dan jenis kelamin.
4. Dengan menggunakan teknik crawling, data aktivitas online siswa SMP dapat diambil secara berkala, yakni setiap 1 jam.
5. Rancang bangun perangkat lunak dapat menampilkan informasi pola aktivitas remaja di Instagram.
6. Tampilan website dapat secara baik diakses melalui media komputer maupun *handphone*.

## 7.2 Saran

Saran penulis untuk penelitian selanjutnya sebagai berikut:

1. Untuk penelitian kedepan guna mendapatkan data yang *realtime* diharapkan perangkat lunak di pasang di *cloud server* sehingga aplikasi terhubung dengan database yang selalu menerima data aktivitas online pengguna instagram.
2. Dilakukannya analisis yang lebih mendalam mengenai klasifikasi dan juga klasifikasi sehingga akurasi informasi lebih baik lagi.
3. Diharapkan data pengguna selalu di perbarui setiap tahun karena data yang diambil merupakan pengguna dengan status pelajar, dimana status tersebut bisa berubah setiap tahunnya.
4. Dalam mengetahui pola aktivitas remaja, atribut yang disarankan untuk diperlihatkan adalah mengenai waktu dan jenis aktivitas yang dilakukan oleh pengguna, seperti contoh: pada pukul 8:00 seharusnya seorang siswa berada di jam sekolah. Pada saat jam 8:00 itu pula apakah juga banyak ditemukan siswa yang aktif di instagram? Sehingga dari kasus tersebut dapat diambil suatu langkah tertentu dalam menyikapi pola aktivitas remaja.



## DAFTAR PUSTAKA

- [1] Yudi Agusta. K-means-penerapan, permasalahan dan metode terkait. *Jurnal Sistem dan Informatika*, 3(1):47–60, 2007.
- [2] APJII. Penetrasi dan perilaku pengguna internet indonesia. 2016.
- [3] beritasatu. Sulit mengontrol pengakses internet usia muda. link dapat diakses di <http://www.beritasatu.com/iptek/166929-tifatul-sembring-sulit-mengontrol-pengakses-internet-usia-muda.html>.
- [4] Ahmad Fatoni and Dhany Dwi. Rancang bangun sistem extreme programming sebagai metodologi pengembangan sistem. *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, 3(1), 2017.
- [5] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- [6] CO Handley Jr, DE Wilson, and AL Gardner. Introduction. *Demography and natural history of the common fruit bat, *Artibeus jamaicensis*, on Barro Colorado Island, Panama. Smithsonian Contributions to Zoology*, 511:1–173, 1991.
- [7] Yuheng Hu, Lydia Manikonda, Subbarao Kambhampati, et al. What we instagram: A first analysis of instagram photo content and user types. In *ICWSM*, 2014.
- [8] Zhexue Huang. A fast clustering algorithm to cluster very large categorical data sets in data mining. *DMKD*, 3(8):34–39, 1997.
- [9] Instagram. Instagram, 2016. link dapat diakses di <https://play.google.com/store/apps/details?id=com.instagram.android>.

- [10] Gautam Pant, Padmini Srinivasan, and Filippo Menczer. Crawling the web. In *Web Dynamics*, pages 153–177. Springer, 2004.
- [11] Tedy Rismawan and Sri Kusumadewi. Aplikasi kmeans untuk pengelompokkan mahasiswa berdasarkan nilai body mass index (bmi) & ukuran kerangka. In *Seminar Nasional Aplikasi Teknologi Informasi*, number 1907-5022, pages E43–E48, 2008.
- [12] Neva Satyahadewi, Muhlasah Novitasari Mara Ediyanto, et al. Pengklasifikasian karakteristik dengan metode k-means cluster analysis. *BIMASTER*, 2(02), 2013.
- [13] Eri Zuliarso and Khabib Mustofa. Crawling web berdasarkan ontology. *Jurnal Teknologi Informasi DINAMIK*, 14(2):105–112, 2009.

# LAMPIRAN A

## KODE

### A.1 Crawling

Berikut ini merupakan potongan kode untuk melakukan crawling data.

```
1 <?php
2 //identifikasi aktivitas online
3 function cek( $str ) {
4     if ( strpos( $str, 'liked' ) !== false ) $hasil = 1;
5     if ( strpos( $str, 'commented' ) !== false ) $hasil = 2;
6     if ( strpos( $str, 'following' ) !== false ) $hasil = 3;
7     return $hasil;
8 }
9
10 include '../vendor/autoload.php';
11 require '../src/Instagram.php';
12 require 'csc_pdo.php';
13
14 ini_set("log_errors", 1);
15 ini_set("error_log", "php-error.log");
16
17 $_db = new stdClass;
18 $_db->name = 'partmaps_teenstagram'; // Nama database
19 $_db->user = 'partmaps_alvin'; // username database
20 $_db->pass = 'Coba-123'; //
21
22 $db = new csc_pdo( $_db->user, $_db->pass, $_db->name );
23 $db->prefix = 'time_';
24
25 function save_time( $row ) {
26     global $db;
27     $waktu_online = date("Y-m-d H:i:s", $row->args->timestamp);
28     $waktu_online2 = date("Y-m-d H:i:s e", $row->args->
29         timestamp);
30     $sid_akun = $row->args->profile_id;
31     $stipe_online = cek( $row->args->text );
32
33     $f = $db->f( 'akun', '*', 'WHERE id=?', $sid_akun );
34     if ( ! $f ) $db->i( 'akun', 'id', array($sid_akun) );
```

```

34
35
36 $f = $db->f( 'waktu_online','*', 'where waktu_online = ? and
    id_akun = ?', array( $waktu_online, $id_akun ) );
37 if ( !$f ) $f = $db->i( 'waktu_online', 'id_akun',
    waktu_online , tipe_online', array( $id_akun ,
    $waktu_online , $tipe_online) );
38 echo "\n" . ( $f ? 'berhasil' : 'gagal' ) . "\n\n";
39 echo "$waktu_online $waktu_online2 $id_akun $tipe_online <
    br>";
40
41 }
42
43 /////// CONFIG ///////
44 //akun instagram untuk menghubungkan ke tiap akun-akun
45 $username = 'rahmanhumaira25';
46 $password = 'Share-123';
47 $debug = false;
48 // $debug = true;
49 // //////////////////////////////////////
50
51
52 date_default_timezone_set( 'Asia/Jakarta' );
53
54 $i = new \InstagramAPI\Instagram($debug);
55 $i->setUser($username, $password);
56
57 try {
58     $i->login($force=true);
59 } catch (Exception $e) {
60     echo 'something went wrong '. $e->getMessage() . "\n";
61     exit(0);
62 }
63 try {
64     $helper = null;
65     $followers = [];
66
67     do {
68         if ( is_null($helper) ) {
69             $helper = $i->getFollowingRecentActivity();
70             foreach ( $helper->stories as $row ) {
71                 save_time( $row );
72                 # code ...
73             }
74
75             $helper = $i->getRecentActivity();
76             foreach ( $helper->old_stories as $row ) {

```

```

77     save_time( $row );
78     # code ...
79 }
80
81     echo json_encode( $helper ); // menampilkan json
82 } else {
83     $helper = $i->getSelfUserFollowers( $helper->
getNextMaxId() );
84 }
85
86     // $followers = array_merge( $followers , $helper->getUsers
87     ());
88 } while ( ! is_null( $helper->getNextMaxId() ) );
89
90 // echo "My followers: \n";
91 // foreach ( $followers as $follower ) {
92 //     echo '- '. $follower->getUsername() . "\n";
93 // }
94 } catch ( Exception $e ) {
95     echo $e->getMessage();
96 }
97
98

```

**Kode A.1:** potongan kode crawl

## A.2 Penjadwalan *Crawling*

Untuk mendapatkan data waktu aktivitas *online* setiap 1 jam, maka dilakukan penjadwalan perintah *crawling* menggunakan layanan *cron job* di *cloud server*. Perintah *cron job* tersebut adalah seperti berikut:

```

1
2 0 * * * * php -q/home/partmaps/teenstagram.part-maps.com/
3     examples/PaginationExample2.php
4

```

**Kode A.2:** perintah *cron job*

### A.3 Query

Query untuk mengambil atribut yang diperlukan:

```

1
2
3 SELECT
4     w.id_akun ,
5     w.tipe_online ,
6     j.id_hari ,
7     j.jam ,
8     j.id_aktivitas ,
9     k.id_sekolah ,
10    IF (k.gender = 'Laki-laki',1,0) AS 'gender '
11
12
13 FROM 'time_waktu_online' w, 'jam_aktivitas' j, 'time_akun' k, '
    aktivitas' a
14 WHERE j.id_aktivitas = a.id
15 AND DAYOFWEEK(w.waktu_online) = j.id_hari
16 AND HOUR(w.waktu_online) = j.jam
17 AND w.id_akun = k.id_akun
18
19

```

**Kode A.3:** query pemilihan atribut

### A.4 Koneksi DB MySQL ke R-studio

Berikut ini merupakan perintah untuk koneksi database MySQL dengan R-studio:

```

1 #install.packages("dbConnect", type="source")
2
3 library(RMySQL)
4 library(dbConnect)
5
6 con = dbConnect(MySQL(),
7 user = 'root', password = '',
8 dbname = 'remaja_smp',
9 hostname='localhost')

```

```

10 dbListTables ( con )
11
12 #Membuat tabel di R dari MySQL
13 myQuery <- "SELECT
14     w.id_akun ,
15     w.tipe_online ,
16     j.id_hari ,
17     j.jam ,
18     j.id_aktivitas ,
19     k.id_sekolah ,
20     IF ( k.gender = 'Laki-laki ',1,0) AS 'gender '
21
22
23     FROM 'time_waktu_online 'w, 'jam_aktivitas 'j, 'time_akun '
24     k, 'aktivitas 'a
25
26     WHERE j.id_aktivitas = a.id
27     AND DAYOFWEEK(w.waktu_online) = j.id_hari
28     AND HOUR(w.waktu_online) = j.jam
29     AND w.id_akun = k.id_akun"
30 df <- dbGetQuery ( con , myQuery )
31 str ( df )
32

```

**Kode A.4:** koneksi database dengan R-studio

## A.5 Pengolahan di R-studio

Membuat *data frame* dengan nama remaja

```

1     remaja <- data.frame ( df )
2
3

```

**Kode A.5:** Data frame

Menghilangkan atribut id akun karena tidak dibutuhkan dalam proses kluster

```

1     #Menghilangkan kolom 1 (id akun)
2     remaja <- data.frame (remaja [ -1])

```

3

**Kode A.6:** menghilangkan atribut id akunMelakukan *K-Modes Clustering*

1

2 `kr3 <- kmodes(remaja, 3, iter.max = 10, weihthed=FALSE)`

3

4

**Kode A.7:** K-Modes

## Menampilkan plot hasil klusterisasi

1

2 `plot(remaja, col=ct3$cluster, pch = 19)`

3

4

**Kode A.8:** plot kluster

## Percobaan klasifikasi

```

1 #penggunaan library
2 library(NLP)
3 library(e1071)
4
5 #pembagian data klasifikasi
6 remajatrain <- data.frame(kkr3)
7 train_bagi <- floor(0.7*nrow(remajatrain))
8 klasremaja <- sample(seq_len(nrow(remajatrain)), size =
   train_bagi)
9 train_remaja <- remajatrain[klasremaja,]
10 test_remaja <- remajatrain[-klasremaja,]
11
12 #data training
13 atr <- train_remaja[,-7]
14 btr <- as.character(train_remaja[,7])
15
16 #data testing
17 ate <- train_remaja[,-7]
18 bte <- as.character(test_remaja[,7])
19
20 #pembuatan model klasifikasi Naive Bayes

```



```

21 model <- naiveBayes(atr, factor(btr))
22
23 #melakukan prediksi
24 prediksi <- predict(model, ate)
25 aktual <- bte
26 bandingin <- cbind(prediksi, aktual)
27
28 #mengukur akurasi klasifikasi
29 akurasi <- 100*sum(prediksi==aktual)/length(prediksi)
30 akurasi

```

### Kode A.9: klasifikasi

## A.6 Visualisasi

Pada tahap ini dilakukan pembangunan visualisasi dengan menggunakan R-studio dengan *package* R-shiny.

```

1 #server.R
2 library(shiny)
3 library(shinydashboard)
4 library(RMySQL)
5 library(dbConnect)
6
7 con = dbConnect(MySQL(),
8 user = 'root',
9 password = '',
10 dbname = 'remaja_smp',
11 hostname = 'localhost')
12
13 myQuery <- "Select
14             w.id_akun,
15             w.tipe_online,
16             j.id_hari,
17             j.jam,
18             i.id_aktivitas,
19             k.id_sekolah,
20             IF (k.gender = 'Laki-laki',1,0) AS 'gender'
21 FROM
22     'time_waktu_online' w,
23     'jam_aktivitas' j,
24     'time_akun' k,
25     'aktivitas' a

```

```

26
27     WHERE j.id_aktivitas = a.id
28     AND DAYOFWEEK(w.waktu_online) = j.id_hari
29     AND HOUR(w.waktu_online) = j.jam
30     AND w.id_akun = k.id_akun"
31
32 df <- dbGetQuery(con, myQuery)
33 str(df)
34
35 remaja <- data.frame(df)
36 remaja <- remaja[-1]
37 library(klaR)
38
39 function(input, output, session) {
40
41     selectedData <- reactive({
42         remaja[, c(input$xcol, input$ycol)]
43     })
44
45     clusters <- reactive({
46         kmodes(selectedData(), input$clusters)
47     })
48
49
50     output$plot1 <- renderPlot({
51         palette(c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3",
52             "FF7F00" ))
53
54         par(mar = c(5.1, 4.1, 0, 1))
55         plot(selectedData(),
56             col = clusters()$cluster,
57             pch = 20, cex = 3)
58         points(clusters()$centers, pch = 4, cex = 4, lwd = 4)
59     })
60
61
62

```

### Kode A.10: server.R

```

1 #ui.R
2 library(shiny)
3 library(shinydashboard)
4
5 shinyUI(
6     dashboardPage(

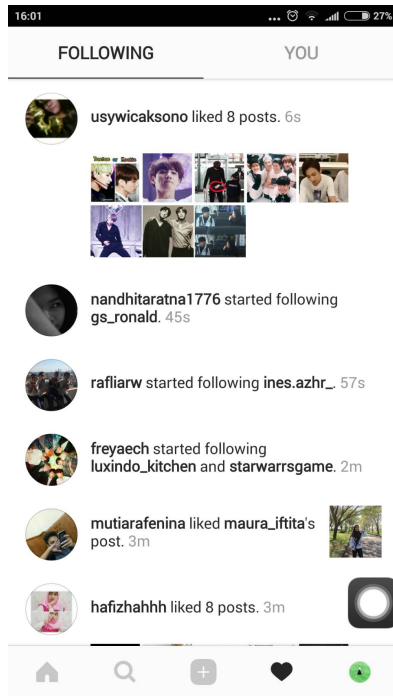
```

```

7 dashboardHeader(title = "Teenstagram"),
8 dashboardSidebar(
9   sidebarMenu(
10    menuItem("Dashboard",
11     tabName = "dashboard",
12     icon = icon("dashboard")),
13    menuItem("Analysis",
14     tabName = "analisis",
15     icon = icon("hourglass")),
16    menuItem("RawData",
17     tabName = "raw",
18     icon = icon("database"))
19  ),
20 ),
21 dashboardBody(
22   tabItems(
23     tabItem(tabName = "dashboard",
24      fluidRow(
25       pageWithSidebar(
26        headerPanel('k-modes clustering'),
27        sidebarPanel(
28         selectInput('xcol', 'X Variable', names(remaja)),
29         selectInput('ycol', 'Y Variable', names(remaja)),
30         selected=names(remaja)[[2]],
31         numericInput('clusters', 'Cluster count', 3,
32          min = 1, max = 5)
33        ),
34        mainPanel(
35         plotOutput('plot1')
36        )
37      )
38     )
39   ))
40
41
42 )
43 )
44 ))
45
46
47

```

**Kode A.11: ui.R**



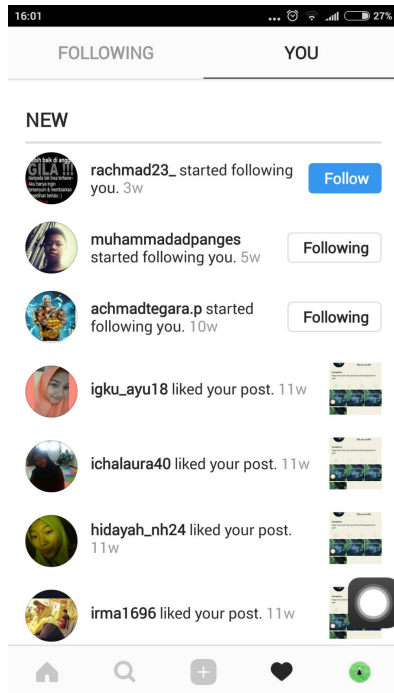
**Gambar A.1:** Halaman *following*

## A.7 Halaman Instagram

Berikut merupakan potongan halaman instagram yang merupakan informasi data untuk waktu aktivitas *online*

## A.8 Hasil Data Crawling

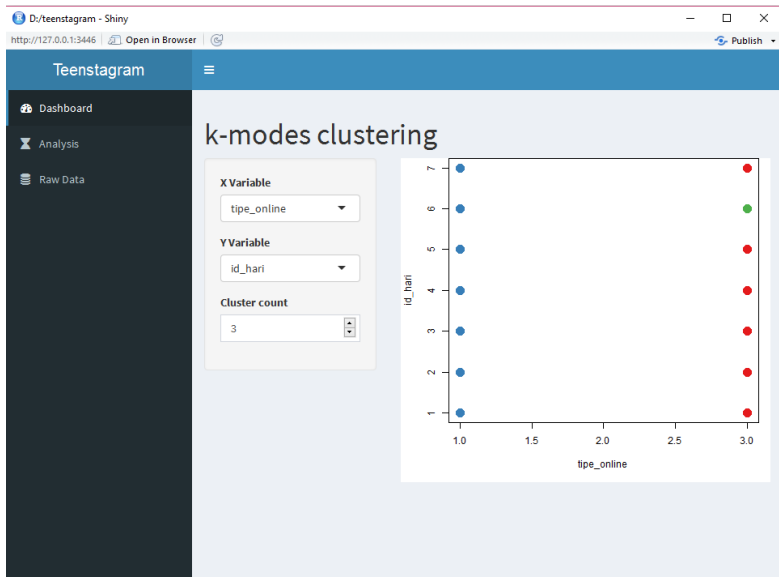
Bagian dari setiap data tersebut adalah



**Gambar A.2:** Halaman *you*

2017-01-31 19:12:17 Asia/Jakarta 2049436713 1  
2017-01-31 19:11:53 Asia/Jakarta 2049436713 1  
2017-01-30 17:37:24 Asia/Jakarta 1912899283 3  
2017-01-29 14:07:11 Asia/Jakarta 4532334020 3  
2017-01-26 15:15:08 Asia/Jakarta 1555712578 2  
2017-01-26 15:14:27 Asia/Jakarta 1555712578 3  
2017-01-26 15:14:06 Asia/Jakarta 1555712578 2  
2017-01-21 23:55:18 Asia/Jakarta 12537 3  
2017-01-20 15:19:44 Asia/Jakarta 4482861641 3  
2017-01-19 21:46:40 Asia/Jakarta 4481724045 3  
2017-01-18 17:58:20 Asia/Jakarta 3782172993 3  
2017-01-16 18:16:47 Asia/Jakarta 1565540476 3  
2017-01-15 21:44:37 Asia/Jakarta 1503763156 3  
2017-01-14 23:04:07 Asia/Jakarta 3270518355 3  
2017-01-14 21:14:32 Asia/Jakarta 4463863736 3  
2017-01-13 05:08:37 Asia/Jakarta 3021499509 3  
2017-01-12 17:57:33 Asia/Jakarta 2300404077 3  
2017-01-11 14:02:09 Asia/Jakarta 1978759078 3  
2017-01-10 20:30:17 Asia/Jakarta 4016774347 3  
2017-01-10 17:04:28 Asia/Jakarta 2190396324 3  
2017-01-10 15:27:54 Asia/Jakarta 3093719780 3  
2017-01-09 16:00:50 Asia/Jakarta 1060011546 3

**Gambar A.3:** Hasil crawling



**Gambar A.4:** Hasil visualisasi Teenstagram

1. Tanggal
2. Jam, menit, detik
3. Zona waktu
4. id akun instagram
5. Kode aktivitas online
  - 1 adalah *like*
  - 2 adalah *comment*
  - 3 adalah *follow*

### A.8.1 Visualisasi

Berikut ini adalah tampilan utama dari website

*Halaman ini sengaja dikosongkan*



## BIODATA PENULIS



Penulis lahir di Tulungagung pada tanggal 8 Januari 1995. Merupakan anak pertama dari 2 bersaudara dan telah menempuh pendidikan formal yaitu; SD Negeri Besuki 01 Tulungagung, SMP Negeri 1 Bandung Tulungagung, dan SMA Negeri 1 Boyolangu Tulungagung.

Pada tahun 2013 melanjutkan pendidikan di Departemen Sistem Informasi FTIF - Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5212100031. Selama menjadi mahasiswa penulis mengikuti kegiatan didalam dan diluar kampus seperti Himpunan Mahasiswa Sistem Informasi, TedxTuguPahlawan Surabaya, membangun UMKM persewaan alat musik decoustic, Menjadi Juara 2 dalam Lomba PIMNAS 29 kategori poster. Disamping aktif dalam kegiatan kemahasiswaan, penulis juga pernah menjadi asisten praktikum pada mata kuliah sistem Desain Basis Data.

Pada tahun keempat karena penulis tertarik dengan bidang desimnasi informasi, maka penulis mengambil bidang minat Laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI). Penulis dapat dihubungi melalui email [alvinrk8@gmail.com](mailto:alvinrk8@gmail.com).