PERJANJIAN PELAKSANAAN PENELITIAN
NOMOR: 1957-Int-DIR.PPKM/UNTAR/IX/2019

Pada hari ini, Selasa tanggal 29 Oktober tahun 2019 yang bertanda tangan di bawah ini:

1. N a m a : Jap Tji Beng., PhD.
   Kedudukan : Direktur Penelitian dan Pengabdian kepada Masyarakat
   Selanjutnya disebut PIHAK PERTAMA;

2. N a m a : Viny Christanti M, M.Kom
   Kedudukan : Peneliti Utama

bertindak untuk diri sendiri dan atas nama : 1 (satu) orang Peneliti Anggota :

a. Nama : Tri Sutrisno, S.Si., M.Sc
   Jabatan : Dosen Fakultas Teknologi Informasi

baik sendiri-sendiri maupun bersama-sama disebut PIHAK KEDUA, dengan ini sepakat untuk mengadakan perjanjian pelaksanaan penelitian sehubungan dengan telah disetujuinya usulan penelitian PIHAK KEDUA dengan judul:

"Aplikasi *Filtering* dan *Preprocessing Text* Data Komputer untuk *Mining* (Studi Kasus: Konten Web Series Iklan Youtube)"

Pasal 1

PIHAK PERTAMA untuk dan atas nama Direktorat Penelitian dan Pengabdian kepada Masyarakat Universitas Tarumanagara, bertindak selaku pemberi tugas kepada PIHAK KEDUA, untuk melaksanakan penelitian dengan kualifikasi: Penelitian Pemula Lanjutan

Pasal 2

PIHAK KEDUA menyatakan sanggup menyelesaikan tugas penelitian sebagaimana tercantum dalam usulan penelitian yang telah disetujui oleh PIHAK PERTAMA, selambat-lambatnya akhir Desember 2019

Pasal 3

PIHAK PERTAMA menyediakan biaya pelaksanaan penelitian bagi PIHAK KEDUA sebesar Rp. 12.200.000,- (dua belas juta dua ratus ribu rupiah) seperti tercantum dalam rencana dan rekapitulasi penggunaan biaya.

Pasal 4

(1) Penggunaan biaya penelitian yang diterima dari PIHAK PERTAMA adalah menjadi wewenang PIHAK KEDUA dengan catatan:
   a. tidak melampaui batas biaya tiap pos anggaran yang telah ditetapkan.
   b. peralatan/inventaris yang dibeli dengan anggaran biaya penelitian ini menjadi milik Direktorat.
(2) Daftar/inventaris sebagaimana tersebut pada Ayat (1) Huruf (b) selambat-lambatnya 1 (satu) bulan setelah penelitian selesai, telah dapat diterima oleh PIHAK PERTAMA

Pasal 5

(1) Biaya penelitian diserahkan oleh PIHAK PERTAMA kepada PIHAK KEDUA dalam 2 (dua) tahapan:
   a. tahap pertama sebesar 50 (lima puluh) persen dari keseluruhan biaya proyek penelitian dilakukan segera setelah naskah perjanjian proyek penelitian yang bersangkutan ditandatangani.
   b. tahap kedua sebesar 50 (lima puluh) persen dilakukan segera setelah laporan penelitian selesai diseminarkan serta menyerahkan laporan akhir proyek penelitian sebanyak 5 (lima) eksemplar hard copy, logbook 2 eksemplar, laporan keuangan 3 (tiga) eksemplar dan soft copy dalam bentuk CD serta draf artikel ilmiah sebanyak 1 (satu) eksemplar kepada Lembaga.

(2) PIHAK KEDUA berkewajiban menyerahkan ringkasan laporan penelitian yang ditulis dalam bentuk artikel ilmiah dilengkapi abstrak dalam bahasa Inggris yang diterima dalam seminar atau jurnal nasional/internasional (*peer reviewer*)

(3) Perjanjian ini berakhir setelah para pihak memenuhi semua kewajiban yang ditentukan Ayat (1) dan (2).

Pasal 6

(1) PIHAK PERTAMA berhak mempublikasikan ringkasan laporan penelitian yang dibuat PIHAK KEDUA ke dalam salah satu jurnal ilmiah yang terbit di lingkungan Universitas Tarumanagara.

(2) Hak cipta dan honorarium atas pemuatan ringkasan laporan penelitian sebagaimana dimaksud Ayat (1) dimiliki oleh PIHAK KEDUA.

(3) PIHAK KEDUA Berkewajiban membuat Poster penelitian yang sudah/sedang dilaksanakan, untuk dipamerkan pada saat kegiatan Research Week tahun terkait

(4) Penggandaan dan publikasi dalam bentuk apapun atas hasil penelitian hanya dapat dilakukan oleh PIHAK KEDUA setelah mendapat persetujuan tertulis dari PIHAK PERTAMA.

Pasal 7

(1) PIHAK KEDUA selain bertanggung jawab atas penyelesaian penelitian yang ditugaskan kepadanya, juga bertanggung jawab atas biaya yang telah diterimanya dari PIHAK PERTAMA.

(2) Apabila PIHAK KEDUA karena sesuatu hal terpaksa harus menyerahkan tanggung jawab kepada pihak lain, maka penyerahan tersebut harus sepengetahuan dan persetujuan tertulis dari PIHAK PERTAMA.

Pasal 8

PIHAK PERTAMA berhak menegur PIHAK KEDUA secara tertulis setiap saat, apabila dipandang PIHAK KEDUA telah melanggar isi perjanjian ini.

Pasal 9

(1) PIHAK KEDUA dapat meminta kepada PIHAK PERTAMA untuk memperpanjang batas waktu penyelesaian penelitian lebih daripada yang ditentukan Pasal 2, dengan menyertai alasan-alasannya.

(2) PIHAK PERTAMA berwenang untuk menolak permintaan PIHAK KEDUA sebagaimana dinyatakan Ayat (1).

(3) Perpanjangan batas waktu sebagaimana dinyatakan Pasal ini hanya dapat diberikan satu kali.

Pasal 10

PIHAK KEDUA wajib mengembalikan biaya yang telah diterimanya kepada PIHAK PERTAMA, apabila PIHAK KEDUA menghentikan secara sepihak pelaksanaan penelitian.

Pasal 11

(1) Apabila penyelesaian tugas penelitian sebagaimana dimaksud Pasal 5 tidak dilakukan tepat pada waktu yang telah ditentukan dalam Pasal 2 perjanjian ini, maka PIHAK KEDUA didenda sebesar 1 (satu) persen untuk setiap hari kelambatan dengan jumlah setinggi-tingginya sebesar sisa anggaran yang belum diterimanya dan/atau dikenakan sanksi administratif.
(2) Sanksi administratif sebagaimana dinyatakan Ayat (1) diberikan oleh Rektor atas rekomendasi Direktorat Penelitian dan Pengabdian kepada Masyarakat.

Pasal 12

Biaya penelitian yang diberikan oleh PIHAK PERTAMA sebagai akibat kesepakatan dalam perjanjian ini dibebankan kepada anggaran Universitas Tarumanagara.

Pasal 13

Resume dan laporan hasil penelitian dibuat PIHAK KEDUA dalam 5 (lima) rangkap disampaikan kepada PIHAK PERTAMA.

Pasal 14

(1) Apabila terjadi perselisihan menyangkut pelaksanaan perjanjian ini, kedua belah pihak sepakat untuk menyelesaikannya secara musyawarah.
(2) Apabila penyelesaian sebagaimana dinyatakan dalam Ayat (1) tidak tercapai, keputusan diserahkan kepada Pimpinan Universitas.
(3) Keputusan sebagaimana dimaksud dalam pasal ini bersifat final.

Pasal 15

Perjanjian ini berlaku sejak ditandatangani oleh kedua belah pihak dan dibuat rangkap 3 (tiga)

PIHAK PERTAMA	PIHAK KEDUA
(Jap Tji Beng, PhD)	(Viny Christanti M, M.Kom)

RENCANA PENGGUNAAN BIAYA
(Rp)

Rencana Penggunan Biaya	Jumlah
Honorarium	Rp. 3.600.000,-
Pelaksanaan penelitian	Rp 8.600.000,-

REKAPITULASI RENCANA PENGGUNAAN BIAYA
(Rp)

No.	Pos Anggaran	Tahap I	Tahap II	Jumlah
1.	Honorarium	1.800.000,-	1.800.000,-	3.600.000,-
2.	Pelaksanaan penelitian	4.300.000,-	4.300.000,-	8.600.000,-
	Jumlah	6.100.000,-	6.100.000,-	12.200.000,-

Jakarta, 2019

Peneliti,

(Viny Christanti M, M.Kom)

LAPORAN PENELITIAN YANG DIAJUKAN
KE LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

APLIKASI FILTERING DAN PREPROCESSING TEXT PADA DATA KOMENTAR UNTUK *TEXT MINING*

(Studi Kasus: Komentar Web Series Iklan Youtube)

Disusun oleh:
Ketua Tim
Viny Christanti M. 0317088203/10805002

Anggota:
Tri Sutrisno, S.Si., M.Sc. 0320018801/10816004

Program Studi Teknik Informatika
Fakultas Teknologi Informasi
Universitas Tarumanagara
Jakarta

2020

HALAMAN PENGESAHAN
LAPORAN PENELITIAN
Semester Ganjil / Tahun 2019-2020

1. Judul : Aplikasi Filtering dan Preprocessing Text pada Data Komentar untuk *Text Mining* (Studi Kasus: Komentar Web Series Iklan Youtube)

2. Ketua Tim
   a. Nama dan Gelar : Viny Christanti M., M.Kom
   b. NIDN/NIK : 10805002 / 0317088203
   c. Jabatan/Gol : Lektor
   d. Program Studi : Teknik Informatika
   e. Fakultas : Teknologi Informasi
   f. Bidang Keahlian : Information Retrieval
   g. Alamat Kantor : Kampus 1 Universitas Tarumanagara Gedung R Lt. 11 Jl. Letjen. S. Parman No. 1, Grogol – Jakarta Barat 11440
   h. Nomor HP/Tlp/Email : 08159287872 / 021-5671747
3. Anggota Tim Penelitian
   a. Jumlah Anggota : Dosen 1 orang
   b. Nama Anggota I/Keahlian : Tri Sutrisno, S.Si., M.Sc. / Statistik
   c. Nama Anggota II/Keahlian :
   d. Nama Anggota III/Keahlian : ……./……
   e. Jumlah Mahasiswa : 1 orang
   f. Nama Mahasiswa/NIM : Walda / 535160010
4. Lokasi Kegiatan Penelitian : FTI Untar
5. Luaran yang dihasilkan : Jurnal
6. Jangka Waktu Pelaksanaan : Juli-Desember
7. Biaya yang disetujui DPPM : Rp 15.000.000

Mengetahui,	Jakarta, 30 Januari 2020
Dekan Fakultas Teknologi Informasi	Ketua Tim Penelitian
Prof. Dyah E. Herwindiati, M.Si	Viny Christanti M., M.Kom
NIP: 10189013	NIP: 10805002

Menyetujui
Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat

Jap Tji Beng, Ph.D
NIDN/NIK: 0323085501 / 10381047

DAFTAR ISI

DAFTAR ISI iii
ABSTRAK iv
BAB I PENDAHULUAN 1
A. Latar Belakang 1
B. Urgensi penelitian 5
BAB II TINJAUAN PUSTAKA 8
A. *Text Analysis* dalam Information retrieval 8
B. *Pre-Processing* 8
C. *Lexical analysis* 9
D. Tokenisasi 10
E. Opini dalam Sosial Media 10
F. Pencarian *World Wide Web* 13
G. Struktur Dasar Dokumen Web HTML 14
BAB III TUJUAN DAN MANFAAT PENELITIAN 17
A. Tujuan Penelitian 17
B. Manfaat Penelitian 17
BAB IV METODE PENELITIAN 18
A. Metode Penelitian 18
B. Jadwal Kegiatan 18
C. Prosedur Pengumpulan Data 20
BAB V HASIL DAN PEMBAHASAN 22
A. Hasil Penelitian 22
B. Pembahasan Pengujian 34
B.1. Pengujian Scraping 35
B.2. Pengujian validasi dan kategorisasi 37
BAB VI KESIMPULAN DAN SARAN 40
A. Kesimpulan 40
B. Saran 40
DAFTAR PUSTAKA vi
Lampiran I Context Diagram vii
Lampiran II Rancangan Interface viii
Lampiran III Contoh Hasil Komentar xvii
Lampiran IV Karya Ilmiah xix
Lampiran V Manual Penggunaan Aplikasi Scraping xxv
Lampiran VI Log Book xxvi

ABSTRAK

Text Mining atau yang sering disebut *Text Analysis* adalah salah satu bidang pengolahan data berupa teks yang dilakukan secara otomatis. Dimana data teks tersebut diperoleh dari sosial media, email, tweets, review produk dan teks lainnya yang ada di dunia maya. Proses *Text Analysis* dapat meliputi ekstraksi, klasifikasi dan proses otomatis lainnya yang bertujuan untuk memperoleh informasi. Dalam melakukan *Text Analysis* maka perlu melalu beberapa tahap salah satunya adalah tahap pengumpulan data sebagai bahan mentah yang akan digunakan dalam proses berikutnya.

Setelah tahap pengumpulan data maka selanjutnya akan masuk ke dalam tahap filtering dan pra-pemrosesan. Tahap filtering merupakan tahap pembersihan data yang dikumpulkan. Data yang masuk dari dunia maya dapat terdiri dari berbagai macam hal yaitu seperti adanya gambar, tanda-tanda atau berbagai tulisan yang tidak sesuai dengan bahan yang diinginkan. Salah satu tahapan pada filtering adalah membuang tag HTML untuk mendapatkan isi penting yang diinginkan untuk diolah. Hasil dari tahapan ini akan digunakan pada tahapan pra-pemrosesan.

Pra-pemprosesan merupakan tahapan awal dalam mengolah data *input* sebelum memasuki proses tahapan utama dari proses *Text Analysis*. Pra-pemprosesan yang dapat terdiri dari beberapa tahapan yaitu *case folding*, penghilangan angka, tanda baca dan karakter selain huruf alphabet, tokenisasi. Dimana tahapan ini dapat disesuaikan dengan kebutuhan *Text Analysis*.

Perbedaan sumber data, jenis data, struktur teks dalam html dan tahapan pra-pemrosesan yang akan dianalisis membuat proses tersebut menjadi sangat penting dibangun sebelum memasuki tahapan analisis teks yang sesungguhnya. Oleh karena itu perlu dibuat sebuah aplikasi yang dapat digunakan untuk melakukan proses pengumpulan data, filtering dan pra pemrosesan. Dimana hasil aplikasi ini dapat disesuaikan dengan bermacam kebutuhan.

BAB I
PENDAHULUAN

A. Latar Belakang

Perkembangan sosial media di internet sudah sangat pesat. Sosial media digunakan untuk banyak kebutuhan, mulai dari kebutuhan berkomunikasi antar komunitas, memasang pengumuman, memasarkan produk hingga mengetahui kualitas film sudah dilakukan melalu media sosial. Saat ini pendapat publik di sosial media melalui internet sangat mempengaruhi berbagai aspek kehidupan. Misalnya pendapat publik terhadap seorang tokoh saja sangat besar pengaruhnya untuk membangun image mereka. Apakah publik menganggap tokoh tersebut sebagai tokoh yang baik atau tidak.

Begitu pula pengaruh terhadap sebuah produk. Saat ini pemasaran produk lebih banyak dilakukan di media sosial. Bentuk iklan yang dipasang di media sosial pun banyak macam, ada yang menggunakan tulisan dan gambar saja, video yang berisi deskripsi produk atau membuat sebuah web series mengenai produk. Iklan-iklan tersebut diletakan dalam berbagai media seperti twitter, blog, instagram, youtube dan lainnya.

Twitter adalah salah satu layanan mikroblog populer yang memungkinkan penggunanya berbagi informasi, melaporkan hal yang sedang terjadi, mempromosikan produk baru, dan berbagi berita terkini dalam bentuk pesan teks yang disebut tweets [1]. Sebuah tweets dapat mengandung link html, sebutan (*@mention*) dan *hashtag*. Pada gambar 1 dapat dilihat contoh sebuah *posting* tweets dari salah satu user resmi sebuah perusahaan.

Gambar 1 Contoh *Posting* Tweets.

Sumber: Carrefour Indonesia, Timeline http://www.twitter.com/Carrefour_ID diakses pada 15 Februari 2013

Perkembangan blog di Indonesia sudah sangat pesat. Secara umum, blog biasanya berisi pendapat atau opini blogger terhadap sesuatu. Dapat berupa opini terhadapa produk, tokoh publik, film atau sebuah kejadian. Sebuah blog yang relevan harus memenuhi dua

kriteria, yaitu relevan terhadap tema dan mengandung suatu opini atau komentar mengenai tema tersebut, baik positif atau negatif [2]. Sebagai contoh, ketika sebuah blog membahas mengenai “Twitter”, sebuah blog dapat mengandung opini mengenai produk *twitter* tersebut. Sebuah blog yang menceritakan mengenai pendapat penulis mengenai *twitter* yang sedang diminati masyarakat dapat berisi opini yang positif, negatif atau netral.

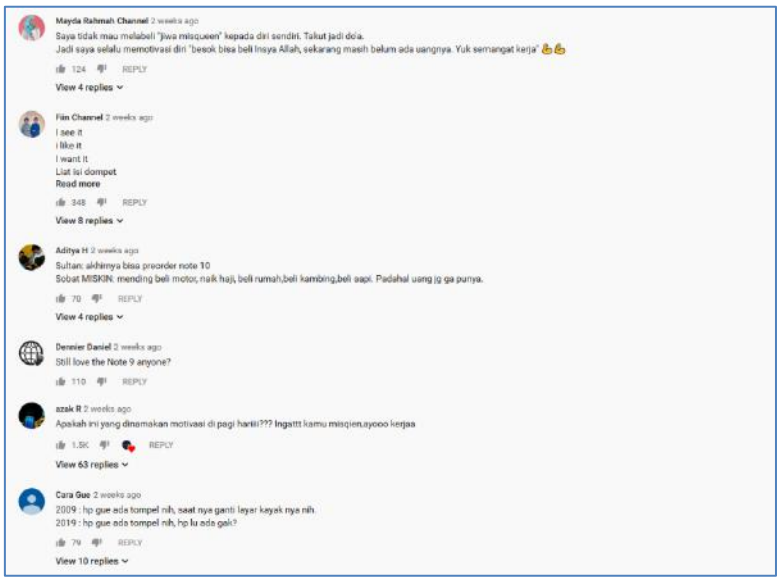
Tahun 2007 *blogger* Indonesia berjumlah sekitar 300.000 *blogger*, sedangkan pada tahun 2008 angka tersebut melonjak hingga mencapai > 600.000 *blogger*. Perkembangan pesat ini membuat Menteri Komunikasi dan Informatika Indonesia menetapkan tanggal 27 Oktober sebagai hari *blogger* nasional [3]. Sehingga saat ini blog menjadi salah satu topik penelitian yang sering dianalisis untuk memperoleh opini publik terhadap sesuatu untuk menjadi trend dalam kehidupan.

Youtube merupakan salah satu media untuk mengunggah video dan berbagi informasi dalam bentuk video [4]. Didalam youtube, pengguna dapat memberikan komentar terhadap video yang diunggahnya. Biasanya youtube digunakan sebagai media pemasaran produk dan mendapatkan opini melalui jumlah like atau komentar yang ditulis oleh para pembaca. Pada gambar 2 dapat dilihat video mengenai produk Samsung note 10.

Gambar 2 Video Samsung Note 10 di Youtube

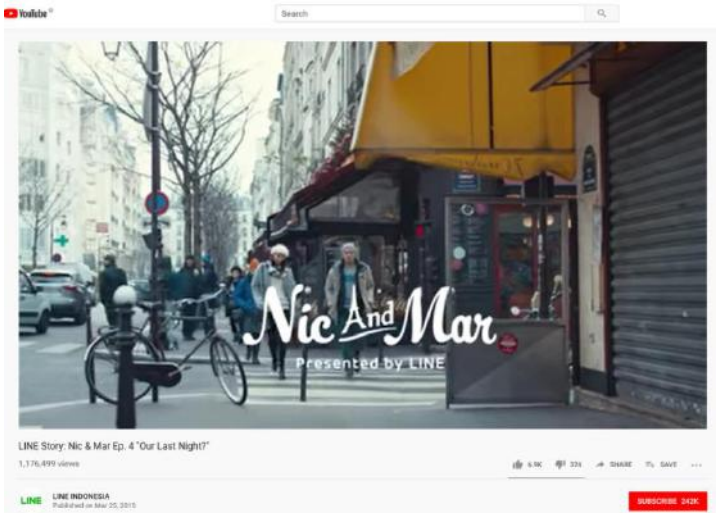
Video tersebut dibuat oleh pengguna yang mencoba smartphone tersebut dan memberikan review atau komentar mengenai smartphone tersebut melalui video. Pada gambar 2 dapat dilihat symbol tanda jempol yang merepresentasikan berapa jumlah orang yang menyukai dan yang tidak meyukainya. Pada video tersebut juga terdapat kolom komentar dimana pengguna lainnya dapat memberikan komentar terhadap video tersebut.

Pada gambar 3 dapat dilihat beberapa komentar terhadap produk tersebut. Ada yang menulisnya dengan Bahasa Inggris, ada juga yang menulisnya dengan Bahasa Indonesia.



Gambar 3 Komentar Samsung Note 10 di Youtube

Bentuk iklan lainnya pada youtube dapat berupa video mengenai produk yang dikeluarkan secara resmi oleh perusahaan atau berupa cerita bersambung yang secara terselubung memasarkan produk tersebut. Seperti pada gambar 4 dapat dilihat web series dari Nic and Mar yang dibuat oleh Line. Dimana pada film tersebut terdapat kolom komentar yang dapat menggambarkan pendapat publik mengenai film tersebut. Pendapat publik tersebut akan berpengaruh terhadap seesuatu seperti produk atau berapa lama hal tersebut akan bertahan [5]. Apakah akan mendapat respon positif atau negatif.



Gambar 4. Web Series Nic and Mar Line story di Youtube

Besarnya pengaruh opini publik di sosial media, membuat hal tersebut menjadi penelitian tersendiri. Dalam bidang Teknologi Informasi, masalah ini menjadi hal yang cukup menarik karena melibatkan pula bagaimana penggunaan komputer dalam membantu manusia untuk mendapatkan informasi mengenai pendapat yang ada di sosial media. Penelitian ini berkembang dengan nama opinion mining.

Opinion Mining merupakan sebuah bidang pembelajaran yang menganalisis opini, sentimen, penilaian, sikap, dan emosi terhadap suatu entitas seperti produk, pelayanan, organisasi, peristiwa, dan lain-lain [6]. Opini tersebut dapat berupa komentar yang ditulis dihalaman web, video atau suara. Secara garis besar, opini diklasifikasikan menjadi tiga kelas, yaitu positif, negatif, dan netral. Namun, dalam bidang lainnya opini dapat dikategorikan dalam beberapa macam kelas lainnya sesuai kebutuhan informasi yang akan diperoleh.

Opini dapat lebih mudah dianalisis dalam berupa teks dan jumlahnya pun lebih banyak diperoleh dibandingkan suara atau video lainnya. Oleh karena kegunaannya yang lebih beragam maka penelitian ini memfokuskan penelitian dalam *Text Analysis*. Proses pengumpulan dan pengolahan komentar yang berupa teks dari dunia maya memakan waktu yang cukup lama jika dilakukan secara manual dimana manusia membaca satu persatu dan merekapitulasinya. Oleh karena itu *Text Mining* atau sering disebut *Text Analysis* menjadi salah satu bidang yang banyak diteliti dalam bidang Teknologi Informasi. Dimana proses pengumpulan data, pengolahan data teks dilakukan secara otomatis sehingga diperoleh data-data yang dapat digunakan untuk memperoleh informasi dengan cepat dan mudah.

Data-data yang digunakan untuk melakukan *opinion mining* dalam bentuk teks biasanya diambil dari *blog* (*blogspot*, *wordpress*) dan juga *microblog* (*twitter, facebook, youtube)* [6]. Data tersebut diperoleh dari halaman web yang ada di dunia maya dimana bentuk setiap halaman web tersebut berbeda. Tidak semua informasi dalam halaman web tersebut dapat diambil untuk diolah menjadi data yang akan dianalisis. Pengguna harus memilih bagian mana yang akan diambil dan menjadi sumber data untuk *Text Analysis*.

Text Mining atau *Text Analysis* merupakan sebuah proses ekstraksi, klasifikasi untuk memperoleh informasi yang berguna untuk manusia [7]. Dimana proses *Text Analysis* melibatkan banyak bidang seperti information retrieval, information extraction, natural language processing, data minning dan lainnya. Didalam setiap proses tersebut juga akan melibatkan *lexical analysis*, anotasi, tagging dan lainnya yang berhubungan dengan proses mengenal bahasa.

Sebelum proses tersebut dapat dilakukan, tahapan yang paling pertama dilakukan adalah proses pengumpulan data, kemudian data yang diperoleh tidak dapat langsung digunakan namun harus melalui proses penyaringan karena bentuk data di dunia maya tidak dapat langsung diolah. Setelah data dikumpulkan dan disaring barulah dimulai tahap pertama proses *Text Analysis* yaitu proses *lexical analysis* dan kemudian tahap ekstraksi, klasifikasi, deteksi dan lainnya dapat dilakukan untuk memperoleh informasi.

B. Urgensi penelitian

Beragamnya sumber data yang akan diolah dari halaman web menjadi sebuah permasalahan sendiri dalam mengumpulkan web yang akan dianalisis. Perbedaan jenis setiap sumber halaman web membutuhkan perlakuan khusus dimana halaman web di dunia maya tersimpan dalam struktur HTML. Dokumen HTML disimpan dalam format teks reguler dan mengandung *tag-tag* yang memerintahkan *web browser* untuk mengeksekusi perintah-perintah yang dispesifikasikan. Struktur web antara twitter berbeda dengan blog atau youtube. Walaupun struktur HTML memiliki standar yang diatur oleh W3C (world wide web consortium), namun setiap informasi dari setiap web memiliki ciri khas tersendiri.

Blog berbahasa Indonesia juga memiliki karakteristik yang berbeda dari blog berbahasa Inggris. Perbedaan struktur bahasa Indonesia dan bahasa Inggris menyebabkan perlu adanya pemrosesan yang berbeda antara koleksi dokumen blog berbahasa Indonesia dengan blog berbahasa Inggris, terutama dalam proses pengindeksan (contohnya pemotongan kata berimbuhan). Kata-kata Indonesia dapat memiliki awalan (seperti me-, di-, dan pe-) dan akhiran (-an, -kan, -i) sedangkan bahasa Inggris hanya memiliki akhiran saja (-s, -es, dan –ness). Dalam perolehan teks yang akan dianalisis, kata-kata opini yang biasa digunakan pada bahasa Inggris berbeda dengan yang biasanya digunakan pada bahasa Indonesia.

Banyaknya kebutuhan *Text Analysis* yang dibutuhkan dalam bidang pemerintahan, penelitian, kebutuhan bisnis membuat proses *Text Analysis* membutuhkan peranan teknologi informasi. Aplikasi *Text Analysis* sudah banyak tersedia secara bebas ada yang berbayar ada yang gratis (Amenity Analytics, RapidMiner, Gate, NLTK). Namun aplikasi *Text Analysis* tersebut secara umum tersedia untuk menganalisis data yang sudah dikumpulkan, sehingga proses pengumpulan data dilakukan terpisah. Pada aplikasi *Text Analysis* yang sudah ada (RapidMiner), membutuhkan data yang sudah valid atau data yang sudah rapi untuk masuk ke dalam tahap analisis selanjutnya.

Secara nyata proses analisi teks membutuhkan pengumpulan data terlebih dahulu. Proses pengumpulan data ini tidak mungkin hanya dilakukan secara manual. Bagaimana

mengambil data komentar dari video youtube yang berbeda-beda. Bagaimana mengumpulkan data twitter dari akun yang berbeda-beda. Atau bagaimana mengumpulkan data dari blog yang memiliki karakteristik yang berbeda antar satu blog yang satu dengan blog lain. Oleh karena itu dibutuhkan aplikasi *Text Analysis* yang dilengkapi dengan modul pengumpulan data dari awal dalam bentuk data mentah.

Perbedaan karakteristik Bahasa juga mempengaruhi teks yang akan diambil dari setiap halaman web. Oleh karena itu dibutuhkan aplikasi yang dapat digunakan untuk memilih data teks mana yang mau diambil dari setiap sumber yang berbeda. Misalnya data dari twitter terdiri dari tweet dan retweet. Data dari blog terdapat opini atau pendapat dalam blog tersebut. Data dari youtube diambil dalam bentuk komentar yang ada pada kolom komentar. Pada penelitian ini difokuskan untuk mengambil data komentar dari youtube terkait web series yang terdapat di youtube.

Sebagai contoh untuk mengumpulkan data *Tweets* yang berupa *posting*-an *tweets* dalam *timeline* pengguna digunakan Twitter API Twitter OAuth PHP. Twitter OAuth PHP adalah teknik *crawling* berbasis html yang disediakan oleh Twitter API dalam bentuk script. Sehingga hanya dapat digunakan oleh peneliti yang mengerti pemrograman. Sedangkan apabila ingin digunakan oleh orang lain harus diseusaikan terlebih dahulu melalui pembuatan coding. Penelitian ini bertujuan untuk membangun aplikasi yang dapat memudahkan peneliti dalam mengumpulkan dokumen tanpa harus membuat program lagi.

Selain masalah pengumpulan data, masalah lainnya adalah proses validasi data. Komputer hanya membantu mengumpulkan data secara cepat, namun data yang dikumpulkan belum tentu semuanya relevan dengan kebutuhan yang diinginkan. Misalnnya pada saat peneliti ingin mengumpulkan opini publik mengenai sebuah produk maka peneliti membutuhkan kalimat dalam bentuk komentar.

Bebasnya komentar yang ditulis oleh para pembaca membuat kalimat yang muncul belum tentu merupakan komentar dari produk tersebut. Sehingga perlu dilakukan validasi terlebih dahulu sebelum data tersebut dianalisis dalam tahap selanjutnya. Validasi hanya bisa dilakukan oleh pakar seperti contoh manusia akan membaca komentar yang ditulis kemudian memutuskan apakah kalimat tersebut merupakan komentar atau hanya kalimat yang tidak berisi komentar.

Proses validasi akan lebih mudah dilakukan jika dilakukan melalui komputer. Pakar akan membaca kemudian memilih apakah komentar tersebut valid berupa kalimat komentar atau bukan kemudian memasukan dalam kategori yang disediakan. Misalnya peneliti membutuhkan 2 kategori opini yang perlu divalidasi oleh 3 validator maka aplikasi dapat

disesuaikan dan muncul form yang berisi kata-kata opini yang akan divalidasi oleh validator kemudian menyimpan hasil validasi. Sehingga peneliti dapat memperoleh data yang sudah tervalidasi untuk digunakan sebagai analisis text atau pembelajaran mesin berikutnya.

Apabila peneliti membutuhkan kategori dan data yang lain untuk divalidasi maka aplikasi dapat dengan mudah disesuaikan sehingga hasil validasi dapat disimpan, direkap dan kemudian dipakai untuk tahapan analisis berikutnya. Ssecara umu, setelah divalidasi maka data tersebut baru masuk ke aplikasi *Text Analysis* untuk diproses ke tahap selanjutnya untuk pembelajaran mesin.

Kebutuhan pengumpulan data, filtering data dan validasi data ini yang akan menjadi kebutuhan utama pada penelitian ini untuk membantu proses analisis teks selanjutnya. Dalam membangun aplikasi ini dibutuhkan untuk menganalisa struktur setiap ragam web, menganalisa bahasa yang digunakan, menganalisa bagaimana membangun mesin yang dapat disesuaikan berdasarkan data yang digunakan untuk memudahkan pakar dalam melakukan validasi data.

BAB II
TINJAUAN PUSTAKA

A. *Text Analysis* dalam Information retrieval

Information Retrieval merupakan bagian dari bidang ilmu komputer yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri [8]. Masalah utama dalam *information retrieval* secara umum adalah menemukan informasi dalam suatu teks. Dalam menemukan informasi dalam teks akan melibatkan *Text Analysis*. Proses *Text Analysis* melibatkan lingusitik, statistik dan machine learning. Hasil dari analisis teks dapat digunakan untuk memperoleh informasi sehingga memuaskan pengguna terhadap permasalahan kebutuhan informasi mereka.

Didalam informasion retrieval terdapat beberapa penelitian yang membutuhkan análisis teks. Seperti penelitian dalam *Blog track* yang merupakan sistem perolehan informasi berupa opini dari dokumen *blog*. *Web track* yang merupakan sistem perolehan informasi dari dokumen *web* yang dapat berupa data teks pada web. Masih banyak bidang penelitian di dalam *information retrieval*, namun pada penelitian ini difokuskan pada *Text Analysis* untuk memperoleh komentar atau opini dari pengguna pada web series di youtube.

B. *Pre-Processing*

Proses *Text Analysis* dalam *information retrieval*, biasanya suatu data atau dokumen akan melalui suatu tahapan sebelum pemrosesan utama. Tahapan awal tersebut disebut *pre-processing*. *Pre-processing* atau sering disebut juga tahap pra-pemrosesan merupakan tahap yang mempermudah dalam melakukan pengolahan data suatu dokumen pada proses berikutnya (seperti klasifikasi, ekstraksi dan lainnya). Terdapat 6 jenis pengolahan teks dalam tahap pra-pemrosesan, yaitu: [9]

1. *Lexical analysis*: merupakan proses mengubah semua komponen dalam teks menjadi sebuah *token*. Proses ini juga merupakan proses penghapusan tanda baca, perubahan huruf besar menjadi *lowercase*, dan penghapusan beberapa karakter khusus lainnya yang disesuaikan dengan kebutuhan sistem yang dirancang.
2. *Stopwords removal*: proses penghapusan kata-kata yang sering ditampilkan dalam dokumen seperti: *and, or, not* dan sebagainya.
3. *Stemming*: merupakan proses mengubah suatu kata bentukan menjadi kata dasar.

4. *POS tagging* atau *Part-Of-Speech tagging* adalah proses untuk menempatkan *part-of-speech*, seperti kata benda, kata kerja, kata ganti, preposisi, kata keterangan, atau kata sifat untuk setiap kata dalam dokumen.
5. *Construction of word*: merupakan tahap mencari kata yang memiliki sinonim dengan kata lainnya dalam satu atau banyak dokumen. Contohnya adalah kata ”sore” memiliki arti yang sama dengan ”senja”.
6. *Named Entities Recognition*: adalah proses pemberian entitas yang terdapat dalam suatu dokumen seperti nama orang, nama organisasi, nama tempat, tanggal, angka. Dalam penelitian ini digunakan Stanford Naned Entities Recognition

C. *Lexical analysis*

Lexical analysis menghapus karakter-karakter yang tidak diperlukan dari suatu dokumen, biasanya seperti *tag* HTML, tanda baca, angka, atau karakter lainnya [9]. Namun proses penghapusan tag HTML tidak dapat dilakukan dengan mudah karena bergantung dari karakter masing-masing web. Setelah membuang tag HTML tahap selanjutnya adalah menyusun data tersebut dalam bentuk struktur yang mudah dibaca dengan diberikan tanda-tanda khusus dan dikumpulkan dalam urutan tertentu. Oleh karena itu proses membuang tag HTML dilakukan secara terpisah.

Setelah melakukan penghapusan karakter yang dianggap tidak perlu, proses *lexical analysis* ini akan memperlakukan urutan karakter menjadi *token*. Setiap *token* (kata) akan dipisahkan dengan spasi. Proses ini juga merupakan proses perubahan huruf besar menjadi huruf kecil. Perubahan huruf besar menjadi huruf kecil ini berguna dalam melakukan perhitungan setiap *token*, karena apabila suatu dokumen memiliki kata yang sama, tetapi memiliki perbedaan dalam huruf besar dan huruf kecil, sistem akan menganggap kata tersebut adalah kata yang berbeda. Contohnya adalah, kata ”hijau” yang seharusnya memiliki arti yang sama dengan kata ”HIJAU”, tetapi sistem akan melihatnya sebagai kata yang berbeda karena perbedaan huruf besar dan huruf kecil tersebut.

Lexical analysis pada perancangan ini mengubah semua urutan karakter menjadi *token*, mengubah semua huruf besar menjadi huruf kecil, dan menghapus semua tanda baca kecuali ”.” (titik) dan ”-” (strip). Tanda baca titik tidak dihapus karena akan digunakan sebagai pemisah antara kalimat yang satu dengan kalimat lainnya dalam perhitungan *bobot kalimat*. Sedangkan tanda baca strip tidak dihapus karena pada bahasa Indonesia terdapat beberapa kata dasar yang menggunakan tanda baca strip. Contoh : ”lumba-lumba”. Bila tanda baca strip pada kata ”lumba-lumba” dihapus maka kata tersebut akan menjadi dua kata dasar

terpisah yaitu ”lumba” dan ”lumba”. Hal ini akan menyebabkan kata dasar tersebut kehilangan makna yang sesungguhnya dan membuat proses *indexing* menjadi tidak akurat.

D. Tokenisasi

Tokenisasi merupakan proses pemotongan urutan karakter menjadi potongan-potongan yang disebut dengan token [10]. Token merupakan urutan karakter pada sebuah dokumen yang dikelompokkan sebagai unit semantik untuk pemrosesan. Kesulitan yang timbul pada proses tokenisasi adalah proses pemotongan kata yang mengandung karakter khusus seperti alamat *email*, alamat IP, dan sebagainya. Selain itu, untuk bahasa yang tidak menggunakan huruf Latin, seperti bahasa Cina, Jepang, Korea, dan lain lain. Diperlukan teknik segmentasi yang tepat sebelum melakukan pengindeksan karena bahasa-bahasa tersebut ditulis tanpa menggunakan spasi diantara kata-kata [10].

Contoh setelah proses tokenisasi adalah sebagai berikut:

Input : Budi membeli baju, celana, dan dompet di mall.

Output :

Budi	membeli	baju	celana	dan	dompet	di	mall

Proses pengindeksan sederhana umumnya menggunakan kata atau kata yang sudah di-*stemming* (kata dasar). Namun, pengindeksan berdasarkan kata biasanya dibagi berdasarkan spasi juga memiliki kelemahan [9]. Nama (misalnya DKI Jakarta) atau frase (misalnya rumah sakit) bila diindeks berdasarkan kata akan mempersulit proses pencarian bila dibandingkan pengindeksan berdasarkan frase. Pengindeksan berdasarkan frase kemudian seringkali digunakan untuk meningkatkan efisiensi dan efektivitas, khususnya untuk pencarian dengan menggunakan kueri yang mengandung frase [9]. Sebagai contohnya, menemukan dokumen yang berisi “komisi pemberantasan korupsi” lebih baik daripada menemukan dokumen yag mengandung “komisi”, “pemberantasan”, dan “korupsi”.

E. Opini dalam Sosial Media

Opini adalah pendapat yang dikemukan oleh manusia. Bentuk opini dapat berupa kalimat tertulis, lisan, gambar, tindakan atau lainnya. Opini dapat disampaikan melalui berbagai media. Salah satu yang saat ini paling mudah digunakan adalah media sosial di dunia maya. Manusia dapat menyampaikan pendapat apapun dalam media sosial.

Perolehan opini dapat dilakukan berdasarkan tulisan dan opini baik positif atau negatif dapat dilihat pada penggunaan kata-kata. Daftar kata opini dapat disusun dengan membuat suatu daftar kata-kata yang menunjukkan opini positif atau opini negatif. Beberapa contoh kata yang menunjukkan opini antara lain "bagus", "jelek", "indah", "buruk", "mantap", "hancur", dan sebagainya. Kata-kata yang menunjukkan opini disusun ketika membaca komentar.

Blog merupakan aplikasi *web* yang memuat tulisan dalam bentuk *posting* yang dapat diperbaharui secara berkala. Perbedaan antara dokumen blog dan dokumen *web* terletak pada struktur dokumen blog biasanya memiliki opini dari para pembaca sehingga akan mempengaruhi hasil perolehan informasi. Informasi yang terkandung di dalam blog bervariasi, berisi mengenai opini peristiwa yang sedang hangat diperbincangkan [11].

Perolehan opini blog (*blog opinion retrieval*) merupakan perolehan informasi yang dikembangkan untuk pencarian opini pada dokumen *web* yang berbentuk blog. Penelitian yang telah dilakukan diantaranya adalah penelitian mesin pencari blog berbahasa Inggris yang banyak dilakukan oleh berbagai institusi pendidikan dari berbagai negara dan didaftarkan pada konferensi internasional TREC (*Text Retrieval Conference*) yaitu "*blog track"* yang telah dimulai sejak tahun 2006.

Beberapa mesin pencari *web* seperti *Google* menyediakan layanan pencari blog. Perbedaan mesin pencari blog *Google* dengan mesin pencari *web*, hanya terletak pada jenis koleksi dokumen yang digunakan, yaitu koleksi dokumen yang berupa blog [11]. Namun, mesin pencari blog *Google* masih menggunakan teknik yang sama seperti mesin pencari *web* pada umumnya, yang merupakan fitur yang paling penting dari dokumen blog tidak diperhatikan padahal mesin pencari blog harus mencari informasi yang merupakan opini.

Pada tahun 2006, konferensi TREC meluncurkan "*Blog Track*" untuk pertama kalinya yang merupakan penelitian yang dilakukan mengenai perolehan informasi pada dokumen blog. Fokus utama *blog track* adalah perolehan opini (*opinion retrieval*). Perkembangan blog di Indonesia yang berkembang dengan pesat mendukung adanya penelitian perolehan informasi blog berbahasa Indonesia.

Opini berbahasa Indonesia memiliki karakteristik yang berbeda dari opini berbahasa Inggris. Perbedaan struktur bahasa Indonesia dan bahasa Inggris menyebabkan perlu adanya pemrosesan yang berbeda antara koleksi komentar berbahasa Indonesia dengan komentar berbahasa Inggris. Dalam perolehan opini, kata-kata opini yang biasa digunakan pada bahasa Inggris berbeda dengan yang biasanya digunakan pada bahasa Indonesia.

Twitter merupakan salah satu mikrobloging populer yang menyediakan layanan kepada penggunanya untuk memposting status yang dikenal sebagai tweets [1]. Tweets memiliki keterbatasan sebanyak 140 karakter. Hal ini menyebabkan *posting* tweet seringkali mengalami kesalahan sintaks dalam penulisan karena para pengguna seringkali mempersingkat penulisan kata-kata untuk menyampaikan informasi mereka dalam sebuah tweet [12]. Informasi yang diberikan di dalam sebuah tweet, dapat berupa informasi cuaca, informasi lalu lintas, berita, gosip, promosi barang dan bahkan obrolan sesama pengguna Twitter.

Youtube merupakan salah satu media di web yang menampilkan dan berbagi video. Dalam youtube juga tersedia kolom komentar untuk menuliskan pendapat mengenai video yang dibagikan. Pada prinsipnya opini yang ada didalam youtube hamper serupa dengan opini pada blog. Dimana pada blog juga tersedia kolom komentar yang dapat digunakan oleh pengguna untuk menuliskan opini mengenai tulisan yang dicantumkan oleh penulis. Pada youtube layanan komentar yang terbatas panjang karakternya.

Layanan youtube menyediakan kolom untuk menuliskan komentar yang dapat berisi opini namun terkadang belum tentu opini yang dicantumkan ke dalam komentar tersebut. Pada kolom komentar tersebut, pembaca lainnya dapat mereply komentar tersebut untuk menunjukan apakah mereka setuju atau tidak, atau menambahkan komentar lainnya. Kalimat pada komentar ini pun dapat saja tidak berhubungan dan tidak saling mendukung. Ketidak konsistenan Bahasa, ketidak tepatan bentuk kalimat komentar inilah yang membutuhkan validasi dari manusia agar data yang masuk untuk dianalysis merupakan data yang tepat yaitu berupa komentar bukan lainnya.

Berbagai perbedaan perolehan opini dengan perolehan informasi secara umum sehingga dapat mempengaruhi data yang digunakan antara lain:

1. Perolehan opini merupakan sebuah proses pengurutan koleksi dokumen dan bukan pengurutan satu dokumen. Sebuah opini dapat dianggap sebagai satu dokumen besar yang mengandung banyak isi (*entries*) atau sebagai dokumen kecil yang mengandung satu isi (*entry*) [13].
2. Karakteristik kedua adalah struktur bahasa dari komentar. Tidak seperti koleksi dari sumber berita dimana setiap dokumen ditulis dengan singkat dan mengandung topik yang berhubungan pada tulisannya, sebuah tulisan komentar biasanya sedikit mengalami perbaikan dan cenderung bersifat tidak teratur. Pada penggunaan kata-kata komentar misalnya kata “sedap” ditulis menjadi “sedaaaap”, sehingga pendeteksian kata-kata opini menjadi lebih sulit. Selain itu halaman tulisan komentar seringkali

berisi komentar yang dapat memperpanjang tulisan tanpa menambahkan teks yang berhubungan dengan topik yang sedang dibahas. Koleksi komentar juga rentan pada isi *spam* dalam bentuk *spam blogs* (*splogs*) dan komentar yang tidak relevan [13].

3. Jenis kueri yang digunakan. Andrei Broder mengelompokkan kueri yang digunakan pada web menjadi [14]:
   a. *Navigational* : menemukan alamat situs tertentu yang pernah dikunjungi pengguna atau diyakini pengguna ada. Contoh kueri "Toshiba", yang memiliki target "www.toshiba.com".
   b. *Transactional* : kueri yang ditujukan untuk menemukan alamat situs dimana akan terjadi interaksi lebih lanjut setelahnya. Umumnya digunakan untuk berbelanja, mengunduh berbagai jenis berkas, mengakses basis data, menemukan server, dan lain lain.
   c. *Informational* : menemukan informasi yang diasumsikan tersedia di *web* dan tidak ada yang dilakukan selain membaca. Contoh kueri "Demam Berdarah", dimana informasi yang dinginkan mengenai penyakit demam berdarah.

F. Pencarian *World Wide Web*

Web Information Retrieval merupakan topik *Information Retrieval* yang ditujukan untuk menangani data dengan jumlah jutaan dokumen, dimana koleksi dokumen tersebut mewakili *world wide web* secara keseluruhan [9]. Dalam memperoleh dokumen web di dunia maya terdapat beberapa teknik yang telah digunakan oleh peneliti-peneliti sebelumnya dalam memperoleh informasi dari *web* antara lain sebagai berikut:

1. Perolehan informasi berdasarkan analisis struktur dokumen web dan target domain. Pengembalian halaman *web* berdasarkan isi halaman, target domain dan isi halaman, dan kombinasi dari judul halaman dan target domain [15].
2. Perolehan informasi berdasarkan analisis struktur dokumen web dan *anchor text*. Perolehan informasi dilakukan dengan analisis isi halaman, *link, URLs, dan Anchors* [16].
3. Perolehan informasi berdasarkan analisis struktur dokumen web. Perolehan informasi dilakukan dengan analisis struktur dokumen web dan *PageRank* [17].

Teknik tersebut akan mempengaruhi bagaimana sebuah keyword dipakai untuk mencari web yang akan dikumpulkan sebagai data dalam proses *Text Analysis*. Struktur web yang berbeda dari setiap halaman akan menyebabkan perhitungan kemiripan dokumen dipengaruhi dari setiap halaman yang dikumpulkan. Halaman web youtube berbeda dengan

blog, twitter dan media opini lainnya. Oleh karena itu setiap halaman web perlu diperlakukan secara khusus pada saat pengumpulan data dikumpulkan yang kemudian data tersebut dipakai sebagai dasar untuk pembelajaran mesin.

G. Struktur Dasar Dokumen Web HTML

HTML merupakan singkatan dari *HyperText Markup Language* adalah *script* untuk menyusun dokumen-dokumen Web [16]. Dokumen HTML disimpan dalam format teks reguler dan mengandung *tag-tag* yang memerintahkan *web browser* untuk mengeksekusi perintah-perintah yang dispesifikasikan. Struktur dasar dokumen HTML dapat dilihat pada gambar 5.

```
<html>
<head>
<title>Di sini Judul Dokumen HTML</title>
</head>
<body>
  Disini penulisan  informasi Web
</body>
</html>
```

Gambar 5 Struktur Dasar Dokumen HTML

Struktur HTML tersebut terdiri dari:

1. ***Tag*** adalah teks khusus (*markup*) berupa dua karakter "<" dan ">", sebagai contoh *<body>* adalah *tag* dengan nama *body*. Secara umum *tag* ditulis secara berpasangan, yang terdiri atas *tag* pembuka dan *tag* penutup (ditambahkan karakter "/" setelah karakter "<"), sebagai contoh *<body>* ini adalah *tag* pembuka isi dokumen HTML, dan *</body>* ini adalah tag penutup isi dokumen HTML.
2. ***Element*** terdiri atas tiga bagian, yaitu *tag* pembuka, isi, dan *tag* penutup. Sebagai contoh untuk menampilkan judul dokumen HTML pada *web browser* digunakan *element title*, di mana *<title>* ini adalah tag pembuka judul dokumen HTML. Di sini Judul Dokumen HTML ini adalah isi judul dokumen HTML. *</title>* ini adalah *tag* penutup judul dokumen HTML. *Tag-tag* yang ditulis secara berpasangan pada suatu *element* HTML, tidak boleh saling tumpang tindih dengan pasangan *tag-tag* lainnya. Contoh penulisan *tag-tag* yang benar pada gambar 6.

```
<p>
      <b>
  ..............
      </b>
</p>
```

Gambar 6 Contoh Penulisan *Tag*

3. ***Attribute*** mendefinisikan properti dari suatu *element* HTML, yang terdiri atas nama dan nilai. Secara umum nilai *attribute* harus berada dalam tanda petik satu atau dua. Sebagai contoh, untuk membuat warna teks menjadi kuning dan latar belakang halaman web menjadi hitam, penulisannya adalah *<body bgcolor="black" text="yellow">*. Penulisannya adalah sebagai berikut:

```
<TAG>
 nama-attr="nilai-attr"
 nama-attr="nilai-attr"
 .................
>
 .................

</TAG>
```

Gambar 7 Penulisan *Attribute*

4. ***Element HTML*** menyatakan pada *browser* bahwa dokumen Web yang digunakan adalah HTML. Sintaksnya sebagai berikut:

```
<html>
..........
</html>
```

Gambar 8 Sintaks HTML

5. ***Element HEAD***

   Merupakan kepala dari dokumen HTML. *Tag <head>* dan *tag </head>* terletak di antara tag <html> dan *tag </html>*.

```
<head>
...........
</head>
```

Gambar 9 Sintaks *HEAD*

6. ***Element TITLE*** merupakan judul dari dokumen HTML yang ditampilkan pada judul jendela *browser*. *Tag <title>* dan *tag </title>* terletak di antara *tag <head>* dan *tag </head>*.

```
<title>
.........

</title>
```

Gambar 10 Sintaks *TITLE*

7. ***Element BODY*** untuk menampilkan isi dokumen HTML. *Tag <body>* dan *tag </body>* terletak di bawah *tag <head>* dan *tag </head>*. *Element BODY* mempunyai *attribute-*

attribute yang menspesifikasikan khususnya warna dan latar belakang dokumen yang akan ditampilkan pada *browser*. *Attribute text* memberikan warna pada teks, *bgcolor* memberikan warna pada latar belakang dokumen HTML, *background* memberikan latar belakang dokumen HTML dalam bentuk gambar, *link* memberikan nilai warna untuk *link*, *alink* memberikan warna untuk *link* yang sedang aktif, *vlink* memberikan warna untuk *link* yang telah dikunjungi. Jika *attribute bgcolor* dan *background* keduanya dispesifikasikan maka *attribute background* yang akan digunakan, akan tetapi jika nilai *attribute background* (gambar) tidak ditemukan pada dokumen HTML maka *attribute bgcolor* yang akan digunakan.

```
<body text="v" bgcolor="w" background="uri" link="x"
alink="y" vlink="z">
.............
</body>
```

Gambar 11 Sintaks *BODY*

BAB III
TUJUAN DAN MANFAAT PENELITIAN

A. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun sebuah aplikasi yang digunakan secara dinamis untuk mengumpulkan dan membersihkan data teks dari web secara otomatis.
2. Membangun aplikasi *Text Analysis* yang dapat dikembangkan oleh pengguna untuk dapat digunakan sesuai kebutuhan masing-masing.
3. Mengumpulkan, membersihkan data komentar dari web series yang ada di youtube sebagai studi kasus.
4. Membuat sebuah aplikasi untuk membuat validasi yang dapat digunakan sebagai form untuk melakukan proses validasi data teks sesuai kebutuhan pengguna (Studi kasus melakukan validasi data komentar dari Youtube oleh responden)

B. Manfaat Penelitian

Penelitian ini dapat bermanfaat untuk beberapa bidang kegiatan seperti berikut:

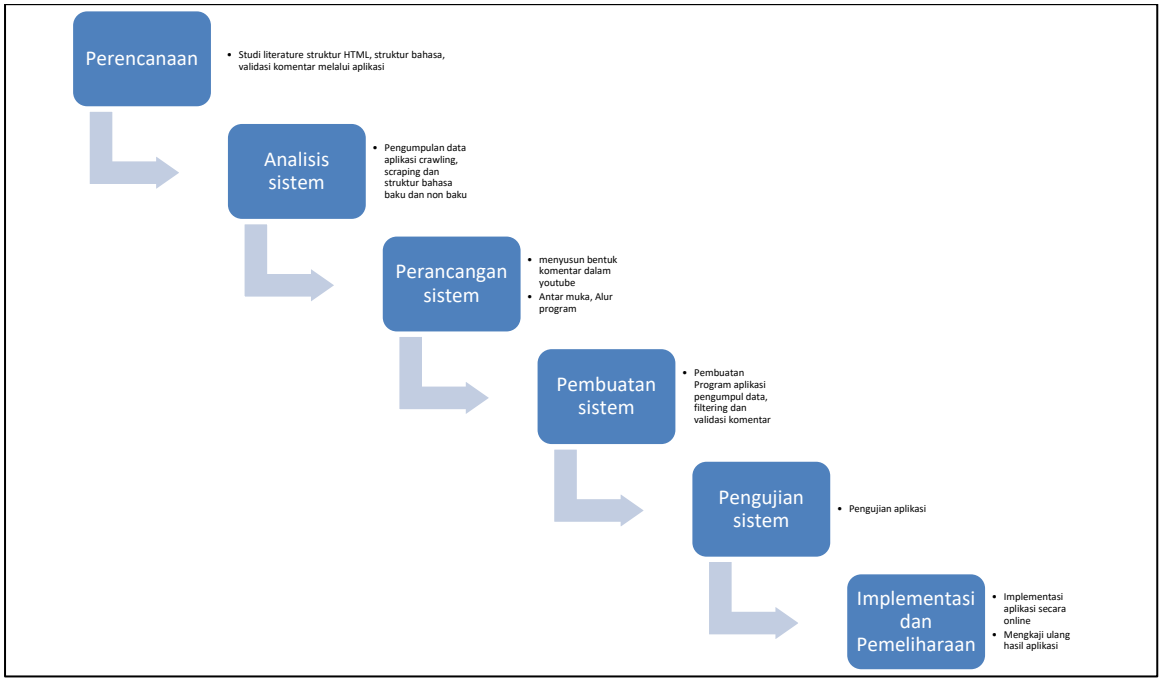
1. Pada bidang ilmu komunikasi membantu para penulis atau jurnalis untuk membahas mengenai hasil review suatu produk,
2. Pada bidang ilmu pemasaran membantu manajemen pemasaran untuk melihat bentuk pemasaran mana yang paling cocok untuk dilakukan,
3. Secara umum hasil dari analisa aplikasi ini dapat digunakan untuk merangkum opini masyarakat dengan cepat sehingga membantu untuk melihat bagaimana pendapat masyarakat tentang suatu film, produk atau iklan,
4. Pada bidang teknologi informasi dapat digunakan sebagai desain penelitian awal sehingga dapat dipakai sebagai alat memudahkan persiapan data untuk diproses dalam proses data mining selanjutnya

BAB IV
METODE PENELITIAN

A. Metode Penelitian

Metode pembuatan program ini menggunakan metode adalah waterfall pada metode System Development Life Cycle (SDLC) yang merupakan salah satu metode pembuatan program aplikasi. SDLC terdiri dari perencanaan, analisis sistem, perancangan sistem, pembuatan sistem, pengujian sistem oleh peneliti dan pemeliharaan.

Metode yang digunakan dalam penelitian ini adalah metode dalam web IR yang memperhatikan struktur web dalam melakukan *crawling* dan scraping. Metode selanjutnya adalah metode dalam filtering tag HTML dan *pre-processing.* Data disimpan dalam bentuk dokumen .xml yang kemudian dipilah untuk mengambil teks untuk menjalani proses *pre-processing.* Skema dari sistem dapat dilihat pada gambar 12.



Gambar 12. Tahapan Pelaksanaan Penelitian Tahap kedua

B. Jadwal Kegiatan

Pada tabel 2 dapat dilihat jadual pelaksanaan penelitian yang akan dilaksanakan mulai bulan Juli hingga Desember 2019.

Tabel 2. Jadwal penelitian

Keterangan	Juli	Agustus	September	Oktober	November	Desember
Studi literature						
Pengumpulan bentuk halaman komentar						
Analisis sistem						
Perancangan antar muka dan alur sistem						
Perancangan basis data						
Pembuatan program Crawling dan filtering halaman web						
Pembuatan program pra-pemrosesan halaman web						
Pengujian sistem						
Implementsi ke server						
Pengumpulan data komentar						
Evaluasi hasil filtering dan prapemrosesan						
Pembuatan laporan						

Tabel 3. Susunan tim penelitian

No.	Nama	NIDN	Fakultas / Prodi	Bidang Keahlian	Tugas	Alokasi Waktu (jam / minggu)
1	Viny Christanti M.	0317088203	FTI / Teknik Informatika	Information Retrieval	Menganalisa struktur web series di youtube dan membuat program	4
2	Tri Sutrisno, S.Si., M.Sc.	10816004	FTI / Teknik Informatika	Statistik	Merekapitulasi data validasi dan menganalisa secara statistik	4
Mahasiswa						
3	Walda	535160010	FTI / Teknik Informatika	Data Analisis	Dokumentasi dan administrasi	2
Pengguna						
4	Cokki S.E., M.M., Dr	10108011	FE / Manajemen	Marketing	Uji coba aplikasi	

Pada tabel 3 dapat dilihat Penelitian ini dilakukan oleh tim peneliti yang terdiri dari 2 orang dosen, 1 orang mahasiswa dan 1 orang pengguna penelitian dengan tugas sebagai berikut:

1. Ketua tim adalah salah satu dosen memiliki kepakaran dalam bidang Information Retrieval dan Natural Language Processing. Ketua tim bertugas merencanakan, menganalisis dan merancang kebutuhan sistem aplikasi filtering dan pra-pemrosesan teks.
2. Anggota tim pertama adalah pakar di implementasi statistik untuk menganalisis komentar dari youtube secara statistik dan menampilkannya dalam aplikasi.
3. Mahasiswa FTI 1 orang yang membantu dalam membuat dokumentasi dan menjadi bagian tugas akhir mahasiswa.
4. Penguji adalah pakar didalam bidang marketing untuk menganalisis komentar, perilaku konsumen dari youtube. Penguji bertugas melakukan uji coba aplikasi yang sudah jadi.
5. Pada penelitian ini juga dibantu oleh asisten peneliti yang merupakan alumni mahasiswa FTI yang bertugas untuk membantu membuat program, menyiapkan dan mengimplementasikan program ke server.

Penelitian dilakukan dengan menggunakan PC dan laptop sebagai perangkat keras. Sedangkan perangkat lunak yang digunakan adalah office, Java, Perl dan Html. Pada penelitian ini juga perlu menyediakan satu buah unit PC yang akan digunakan sebagai server untuk mengimplementasikan aplikasi sehingga aplikasi dapat digunakan secara online.

C. Prosedur Pengumpulan Data

Pada tahap awal, pengumpulan data akan dilakukan secara manual untuk mempelajari struktur komentar yang ada dalam web series youtube. Setelah struktur web dari komentar dipelajari, maka tahap selanjutnya adalah menyusun bentuk filtering yang dapat digunakan untuk membuang tag HTML yang tidak diperlukan. Setelah dikumpulkan maka dibuatlah program untuk mengumpulkan komentar dalam jumlah besar. Hasil dari data tersebut akan dirapikan dan disimpan dalam format yang dapat memudahkan tahapan *Text Analysis* selanjutnya.

Koleksi dokumen yang dibuat diambil dari dokumen awal yang di-*crawling* dengan hanya mengambil judul, isi dan *comment* dari dokumen awal. Koleksi dokumen memiliki format nomor dokumen <DOCNO>, judul <title>, isi dokumen <body>, dan opini pengunjung <comment> [5]. Komentar yang dijadikan contoh akan dilihat dari beberapa struktur komentar yang ada di Youtube. Bagaiman mengambil komentar utama dan komentar reply dari pengguna lainnya. Data yang akan digunakan adalah komentar web series yang ada di youtube dalam bentuk teks.

Aplikasi yang dibangun akan digunakan untuk mengumpulkan data dari link yang ada di youtube. Secara spesifik data yang akan dikumpulkan melalui aplikasi akan difilter dan diambil komentarnya saja. Sebagai contoh untuk mengumpulkan data *Tweets* yang berupa *posting*-an *tweets* dalam *timeline* pengguna digunakan Twitter API Twitter OAuth PHP. Sehingga hanya dapat digunakan oleh peneliti yang mengerti pemrograman.

Penyajian data untuk daftar kata umum disimpan dalam *file* stopword.xml. Penyimpanan dilakukan dengan memisahkan setiap kata dengan *tag string*. Tanda ini digunakan untuk mempermudah program dalam membaca setiap kata dalam *file* stopword.xml. Penyimpanan data untuk hasil *preprocessing* disimpan pada *file*. Setiap satu *file* kumpulan komentar mempunyai informasi yang akan disimpan. Setiap informasi dipisah berdasarkan *tag* xml. Informasi yang akan disimpan adalah kata dan kategori dari komentar.

BAB V
HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Penelitian ini telah menghasilkan aplikasi scrapping yang dapat mengambil komentar dari youtube kemudian diolah dan divalidasi. Aplikasi ini dapat diakses melalui nama link www.rukomindo.com. Aplikasi ini dapat diakses oleh 2 pengguna yaitu koordinator dan coder. Setiap pengguna yang akan menggunakan aplikasi ini akan dimulai dengan menu utama untuk login seperti pada gambar 13. Tampilan awal rukomindo adalah form login dan tidak ada *page* yang diperuntukkan bagi *guest* (user tanpa login). Ada 2(dua) tipe user pada rukomindo, yaitu koordinator atau admin dan coder. Pada gambar 13 dapat dilihat *field* username, *field* password, *button* untuk *submit* form dan *button* untuk *reset* form.

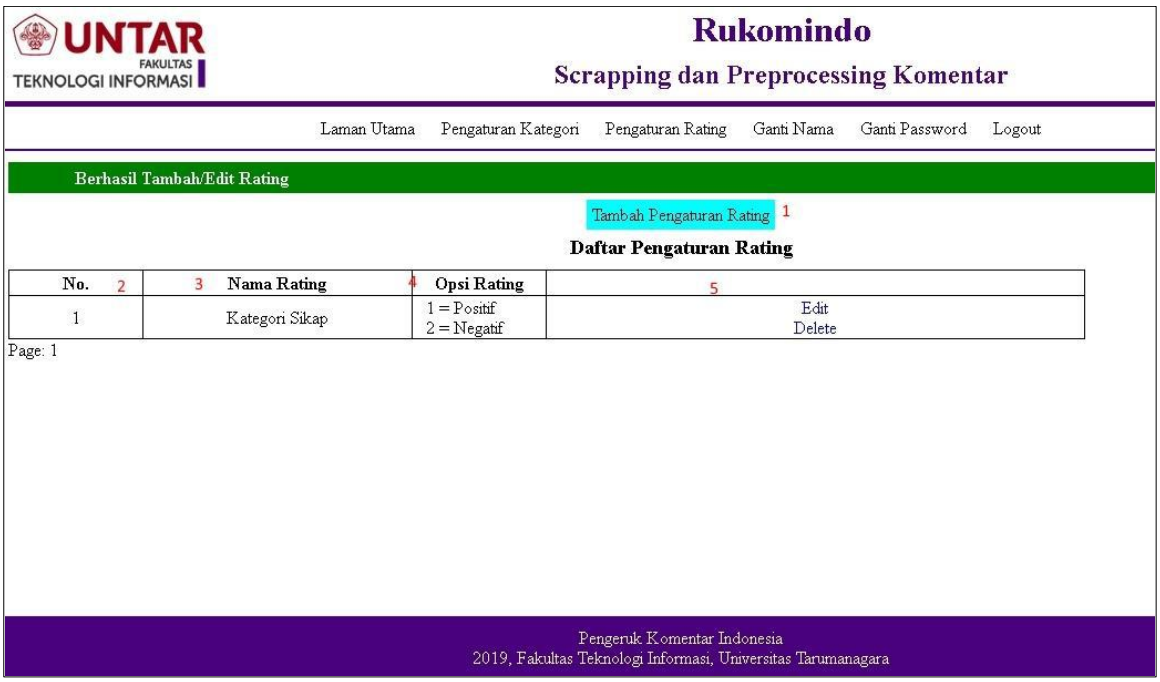
Gambar 13. Home Rukomindo

Koordinator berlaku sebagai admin yang dapat mengatur video apa yang akan diambil komentarnya. Menu koordinator terdiri dari 5 fungsi utama yaitu pengaturan pengujian, hasil pengujian, data coder, pengaturan kategori dan pengaturan rating. Menu untuk admin awalnya tidak ditampilkan semua. Beberapa menu baru ditampilkan setelah selesainya proses atau tahap tertentu. Tampilan menu admin adalah sebagai berikut:

1. Menu Home. Pada gambar 14 dapat dilihat tampilan home Admin

Gambar 14 Home Admin

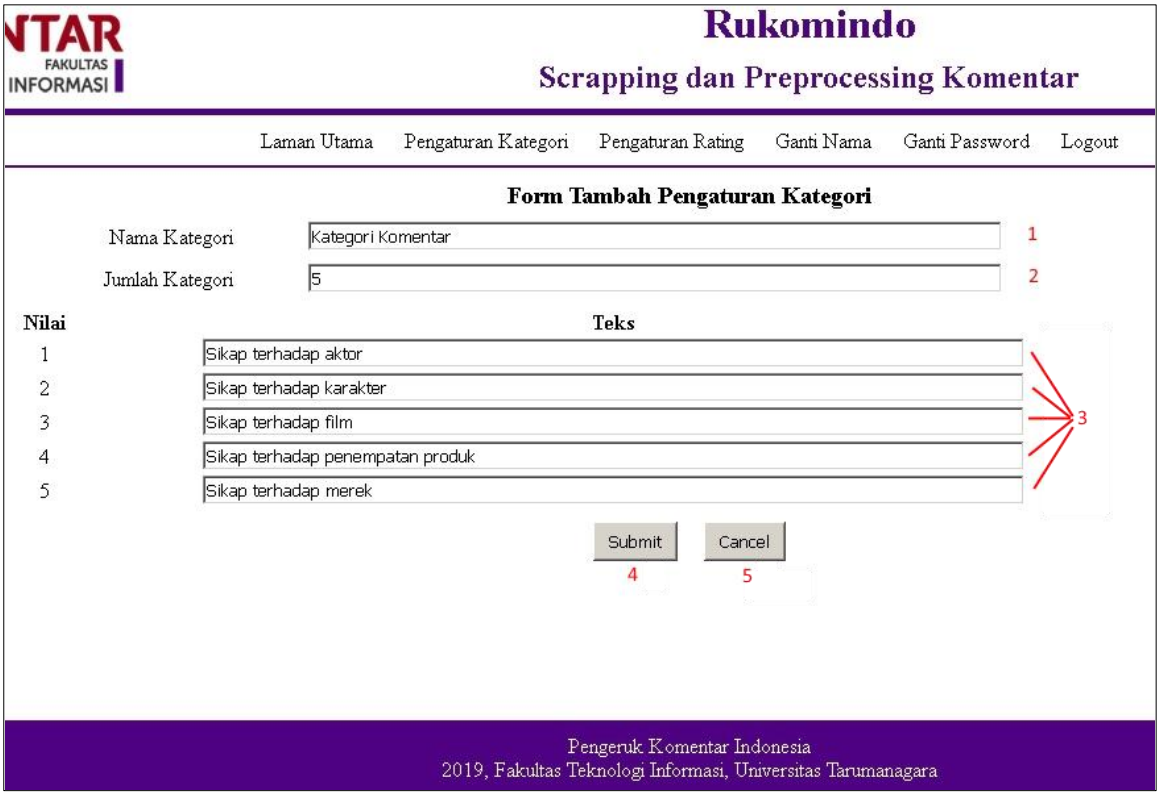
2. Menu **Pengaturan Rating**. Pada gambar 15 menampilkan daftar pilihan rating, yaitu pilihan yang diberikan pada coder atau admin saat memberikan rating pada setiap komentar. Koordinator dapat mengatur berapa banyak rating yang digunakan.



Gambar 15. Daftar Rating

3. Menu **Pengaturan Kategori**. Pada gambar 16 menampilkan daftar pilihan kategori, yaitu pilihan yang diberikan pada coder/admin saat menentukan kategori pada setiap komentar. Koordinator dapat mengatur berapa banyak kategori komentar yang digunakan seperti pada gambar 17 terdapat form untuk mengisi kategori komentarnya.

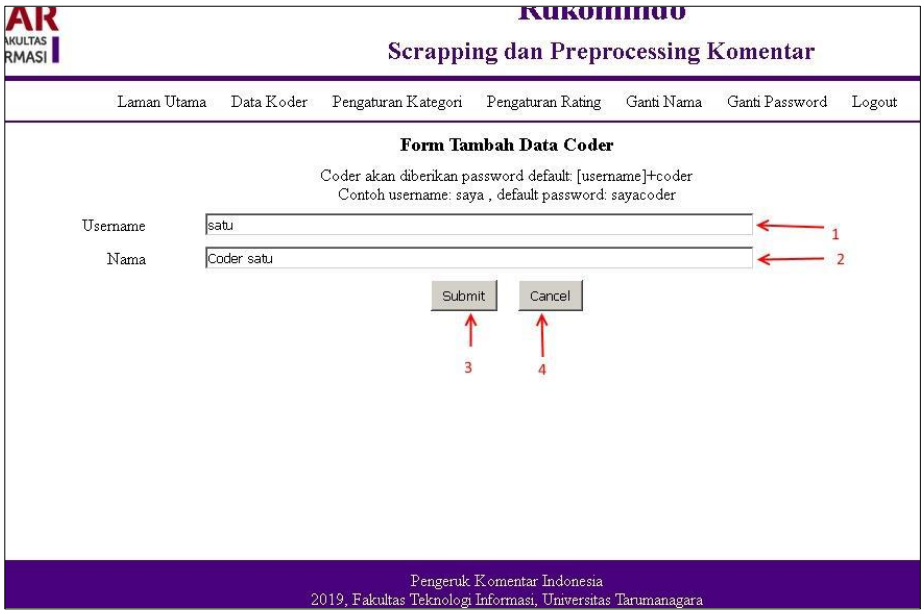
Gambar 16 Daftar Kategori



Gambar 17 Form Rating

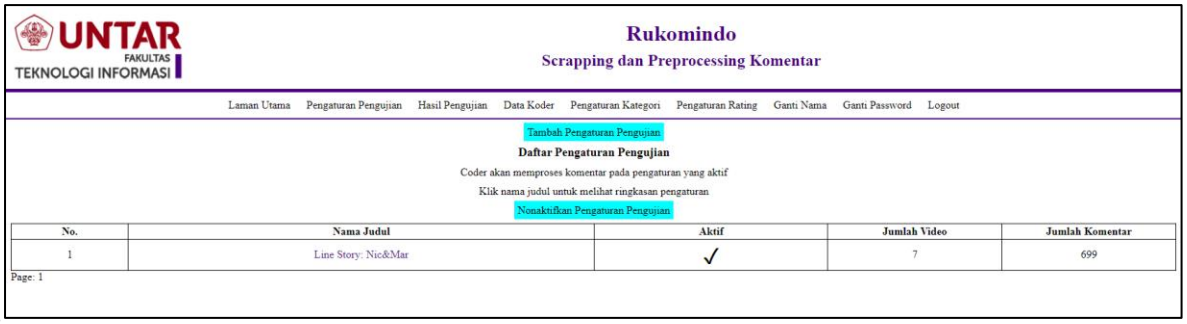
4. Menu **Data Coder**. Pada gambar 18 menampilkan daftar coder dimana pada menu ini koordinator dapat melihat nama-nama coder dan koordinator dapat menambah data coder seperti pada gambar 19. Pada Form Coder koordinator harus menginput username dan nama coder. Setiap coder akan diberi password *default*.

Gambar 18 Data Coder



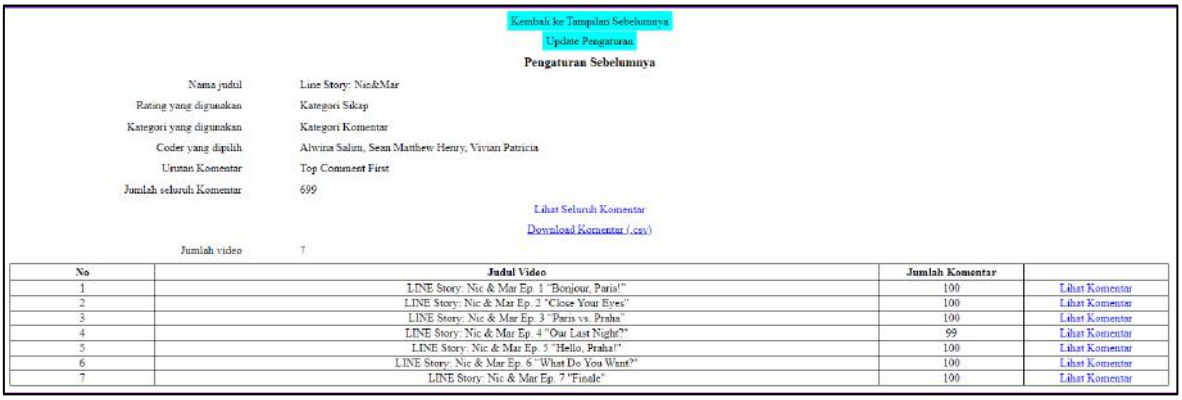
Gambar 19 Form Coder

5. Menu **Pengaturan Pengujian**. Pada halaman ini menampilkan daftar pengaturan pengujian. Pengaturan pengujian yang digunakan adalah pengaturan yang aktif, admin dapat mengubah pengaturan mana yang aktif. Contoh tampilan dapat dilihat pada Gambar 20.



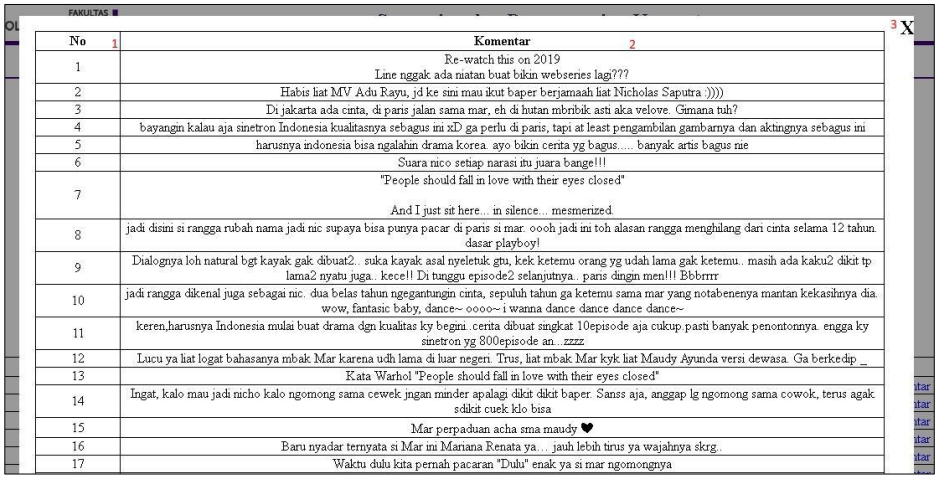
Gambar 20 Daftar Pengaturan Pengujian

Pada gambar 21 dapat dilihat **Rincian Pengaturan Pengujian** yang menampilkan informasi mengenai pengaturan pengujian tersebut, termasuk juga daftar komentar yang akan diproses oleh coder.



Gambar 21 Rincian Pengaturan Pengujian

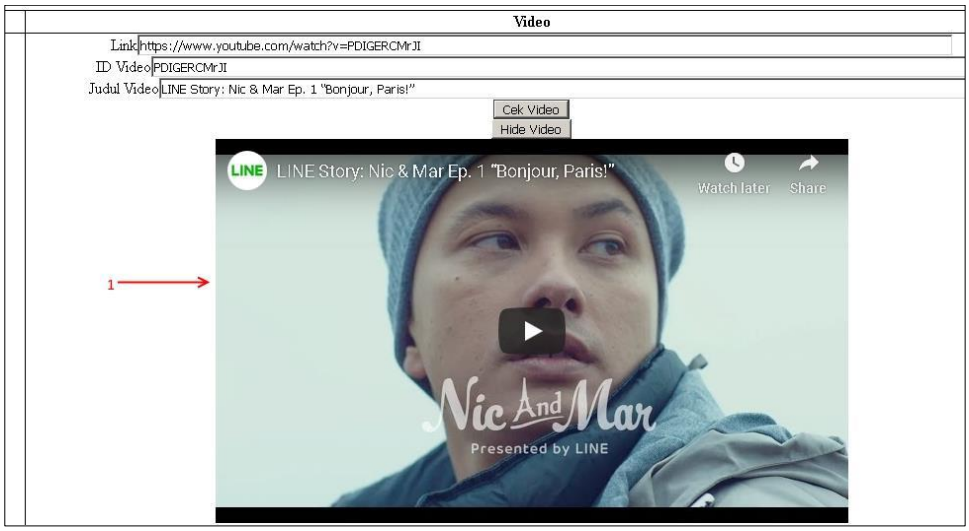
Hasil komentar dapat diklik pada tulisan lihat komentar hasil komentar dapat dilihat pada gambar 22.



No	Komentar
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie
6	Suara nico setiap narasi itu juara bange!!!
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!
9	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~
11	keren,harusnya Indonesia mulai buat drama dgn kualitas ky begini..cerita dibuat singkat 10episode aja cukup.pasti banyak penontonnya. engga ky sinetron yg 800episode an...zzzz
12	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _
13	Kata Warhol "People should fall in love with their eyes closed"
14	Ingat, kalo mau jadi nicho kalo ngomong sama cewek jngan minder apalagi dikit dikit baper. Sanss aja, anggap lg ngomong sama cowok, terus agak sdikit cuek klo bisa
15	Mar perpaduan acha sma maudy ❤
16	Baru nyadar ternyata si Mar ini Mariana Renata ya… jauh lebih tirus ya wajahnya skrg..
17	Waktu dulu kita pernah pacaran "Dulu" enak ya si mar ngomongnya

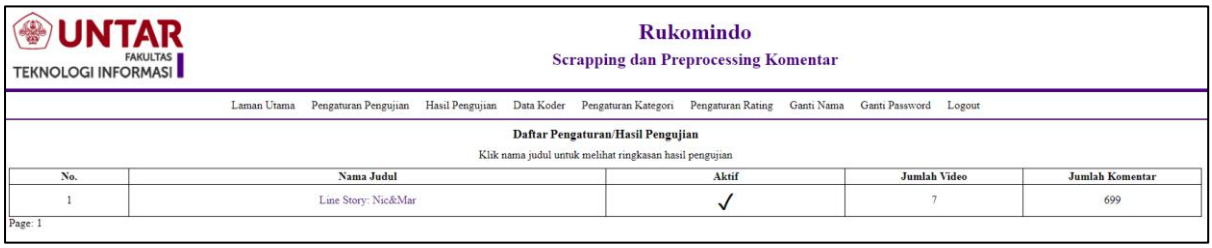
Gambar 22 Daftar Komentar

Pada pengaturan pengujian, koordinator juga dapat melihat video yang akan diambil link youtubenya kemudian di copy pada menu link seperti pada gambar 23.



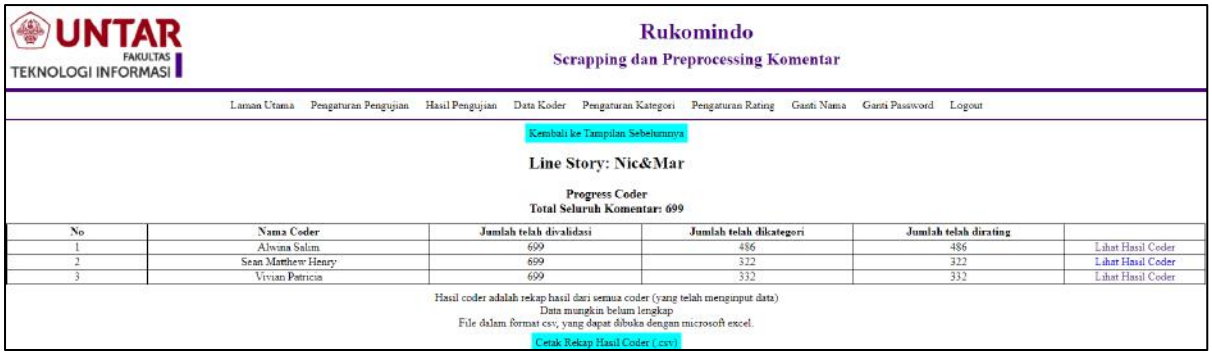
Gambar 23 Contoh Menampilkan Video Youtube

6. Menu **Daftar Hasil Pengujian**. Halaman ini menampilkan daftar pengujian yang ada. Pengujian aktif akan diberi penanda, hanya pengujian tersebut yang dapat diinput oleh admin. Admin hanya dapat melihat hasil dari pengujian lainnya Contoh tampilan dapat dilihat pada Gambar 24.



Gambar 24 Daftar Hasil Pengujian

Halaman **Rekap Pengujian Coder** pada gambar 25 adalah tampilan yang dapat memperlihatkan data berupa rekapan dari progress pengujian yang telah dilakukan semua coder. Apabila semua coder telah menyelesaikan pengujian, admin dapat menentukan hasil akhir atau kesimpulan untuk setiap komentar.



UNTAR FAKULTAS TEKNOLOGI INFORMASI

Rukomindo

Scrapping dan Preprocessing Komentar

Laman Utama | Pengaturan Pengujian | Hasil Pengujian | Data Koder | Pengaturan Kategori | Pengaturan Rating | Ganti Nama | Ganti Password | Logout

Kembali ke Tampilan Sebelumnya

Line Story: Nic&Mar

Progress Coder
Total Seluruh Komentar: 699

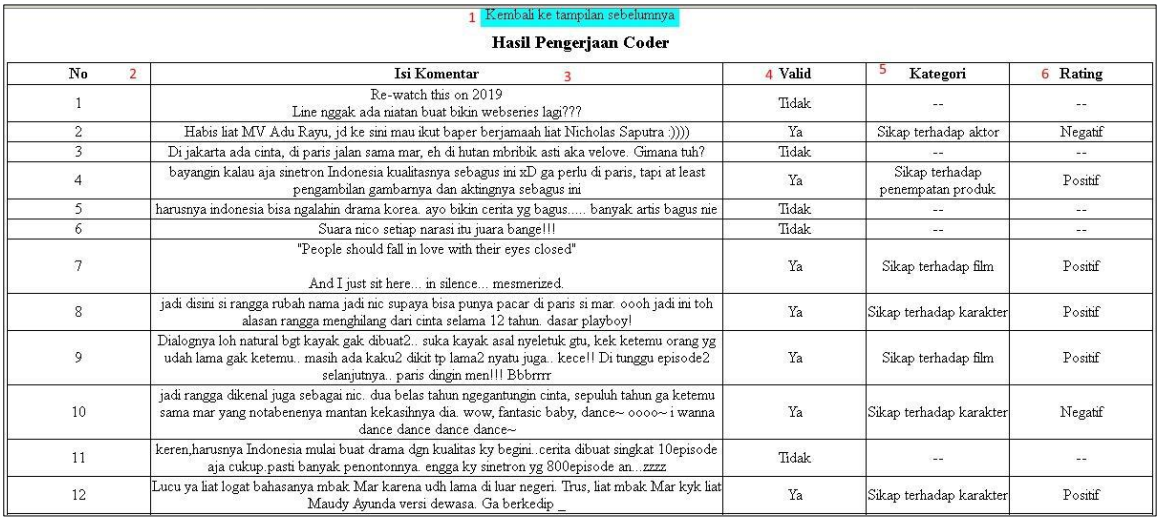
No	Nama Coder	Jumlah telah divalidasi	Jumlah telah dikategori	Jumlah telah dirating	
1	Alwina Salim	699	486	486	Lihat Hasil Coder
2	Sean Matthew Henry	699	322	322	Lihat Hasil Coder
3	Vivian Patricia	699	332	332	Lihat Hasil Coder

Hasil coder adalah rekap hasil dari semua coder (yang telah menginput data)
Data mungkin belum lengkap
File dalam format csv, yang dapat dibuka dengan microsoft excel.

Cetak Rekap Hasil Coder (.csv)

Gambar 25 Rekap Pengujian Coder

Halaman **Hasil Pengujian Coder** pada gambar 26, menampilkan hasil pengujian dari coder tersebut. Informasi yang ditampilkan adalah komentar valid/tidak, kategori, dan rating yang ditentukan oleh coder. Komentar yang tidak valid, tidak memiliki kategori dan rating.



1 Kembali ke tampilan sebelumnya

Hasil Pengerjaan Coder

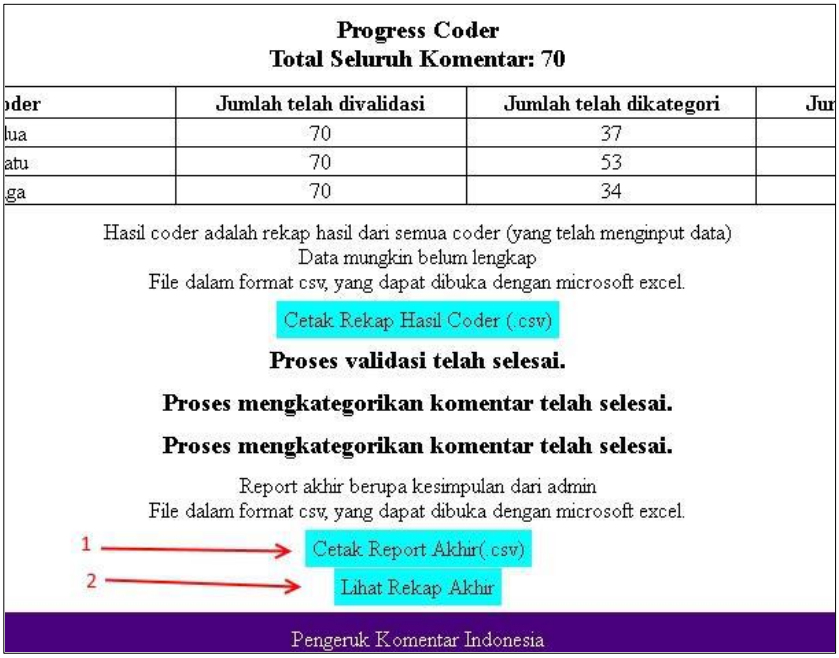
No 2	Isi Komentar 3	4 Valid	5 Kategori	6 Rating
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???	Tidak	--	--
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	Ya	Sikap terhadap aktor	Negatif
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?	Tidak	--	--
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	Ya	Sikap terhadap penempatan produk	Positif
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	Tidak	--	--
6	Suara nico setiap narasi itu juara bange!!!	Tidak	--	--
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	Ya	Sikap terhadap film	Positif
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	Ya	Sikap terhadap karakter	Positif
9	Dialognya loh natural bgt kayak gak dibuat2... suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. keee!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	Ya	Sikap terhadap film	Positif
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	Ya	Sikap terhadap karakter	Negatif
11	keren,harusnya Indonesia mulai buat drama dgn kualitas ky begini..cerita dibuat singkat 10episode aja cukup.pasti banyak penontonnya. engga ky sinetron yg 800episode an...zzzz	Tidak	--	--
12	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _	Ya	Sikap terhadap karakter	Positif

Gambar 26 Hasil Pengujian Coder

Pada halaman **Validasi Komentar** ini berisi form yang digunakan admin untuk menentukan apakah komentar tersebut valid atau tidak, setiap *page* hanya menampilkan 10 komentar. Admin dapat melihat jumlah coder yang memilih komentar valid dan tidak valid. Jika salah satu pilihan jumlahnya lebih besar, form akan otomatis memilih pilihan dengan jumlah terbanyak.

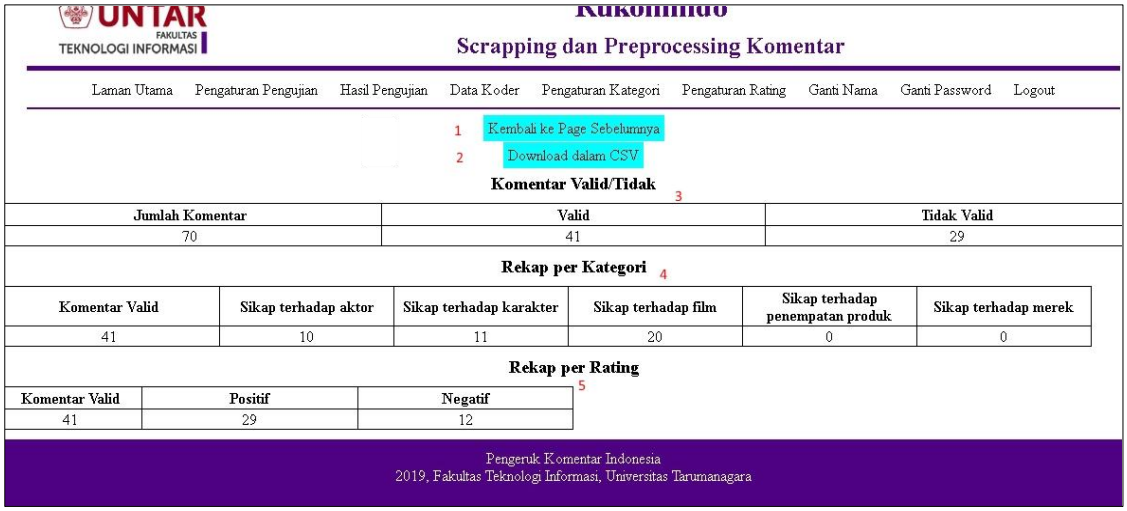
Setelah admin selesai menentukan kategori atau rating untuk semua komentar yang valid, admin akan diarahkan kembali ke tampilan rekap pengujian coder. Jika keduanya telah selesai, tampilan rekap pengujian akan seperti pada Gambar 27. Link untuk menentukan

kategori dan rating sudah tidak ada, diganti dengan link untuk mendapatkan rekap akhir dalam file csv (link nomor 1 pada Gambar 27) dan link untuk melihat rekap akhir (link nomor 2 pada Gambar 27).



Gambar 27 Tampilan Rekap Pengujian Coder Setelah Admin Selesai Menentukan Kategori dan Rating Komentar

Halaman **Rekap Hasil Akhir** ini berisi rekap dari input admin. Rekap ditampilkan dalam 3 (tiga) tabel, yaitu tabel komentar valid/tidak, tabel rekap per kategori, dan tabel rekap per rating Contoh tampilan dapat dilihat pada Gambar 28.



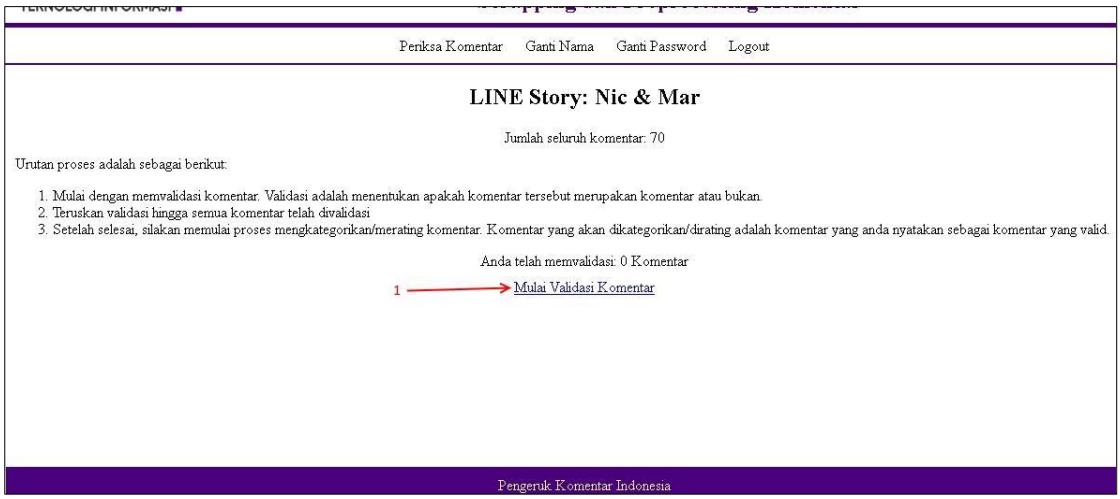
Gambar 28 Rekap Hasil Akhir

Menu untuk coder terdiri dari 3 fungsi utama yaitu periksa komentar, ganti nama dang anti password. Ini adalah tampilan awal untuk admin, seperti terlihat pada Gambar 29.

Gambar 29 Home Coder

Periksa Komentar

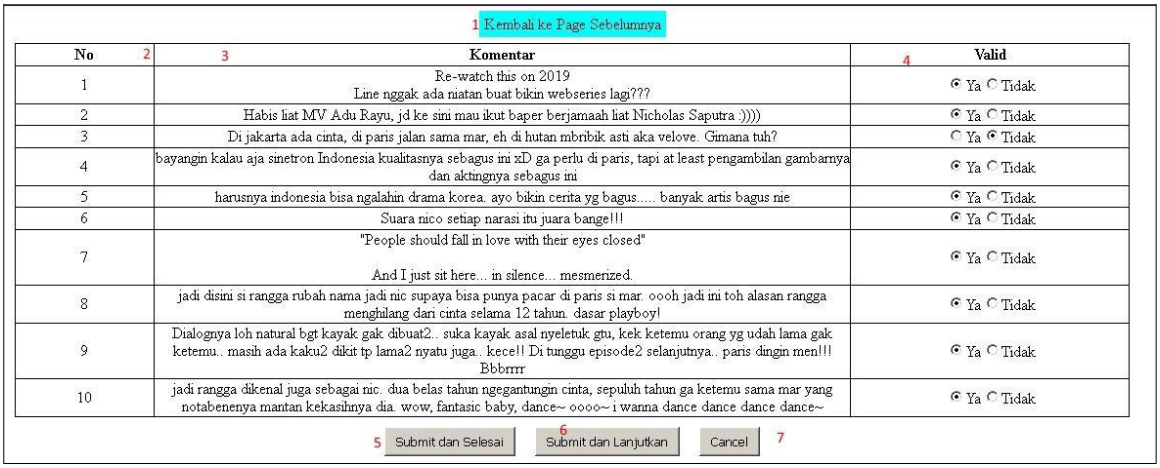
Ini adalah tampilan untuk memulai dan melihat progress pengujian, seperti terlihat pada Gambar 30. Pengujian diawali dengan memvalidasi komentar, yang dapat dilakukan dengan mengeklik link nomor 1 pada Gambar 30.



Gambar 30 Periksa Komentar

Form Validasi Komentar

Page ini berisi form yang digunakan coder untuk menentukan apakah komentar tersebut valid atau tidak, setiap *page* hanya menampilkan 10 komentar. Contoh tampilan dapat dilihat pada Gambar 31.

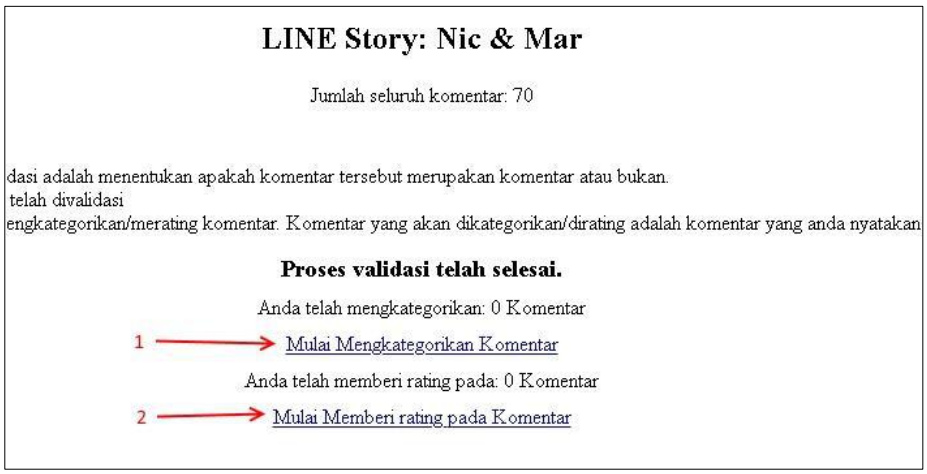

1 Kembali ke Page Sebelumnya

No 2	3 Komentar	4 Valid
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???	⦿ Ya ○ Tidak
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	⦿ Ya ○ Tidak
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?	○ Ya ⦿ Tidak
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	⦿ Ya ○ Tidak
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	⦿ Ya ○ Tidak
6	Suara nico setiap narasi itu juara bange!!!	⦿ Ya ○ Tidak
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	⦿ Ya ○ Tidak
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	⦿ Ya ○ Tidak
9	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	⦿ Ya ○ Tidak
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	⦿ Ya ○ Tidak

5 Submit dan Selesai 6 Submit dan Lanjutkan Cancel 7

Gambar 31 Form Validasi Komentar

Setelah coder selesai memvalidasi semua komentar, coder akan diarahkan kembali ke tampilan periksa komentar. Contoh tampilan dapat dilihat pada Gambar 32. Link untuk melakukan validasi sudah tidak ada, diganti dengan link untuk menentukan kategori komentar (link nomor 1 pada Gambar 32) dan rating komentar (link nomor 2 pada Gambar 32).



Gambar 32 Tampilan Periksa Komentar Setelah Coder Selesai Validasi Komentar

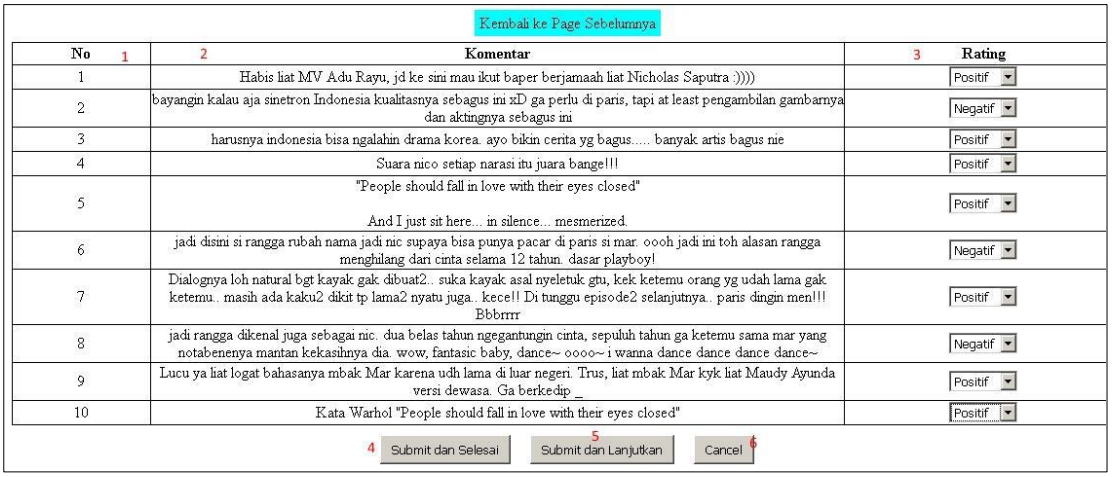
Form Kategori Komentar

Page ini berisi form yang digunakan coder untuk menentukan kategori untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Contoh tampilan dapat dilihat pada Gambar 33.

Gambar 33 Form Kategori Komentar

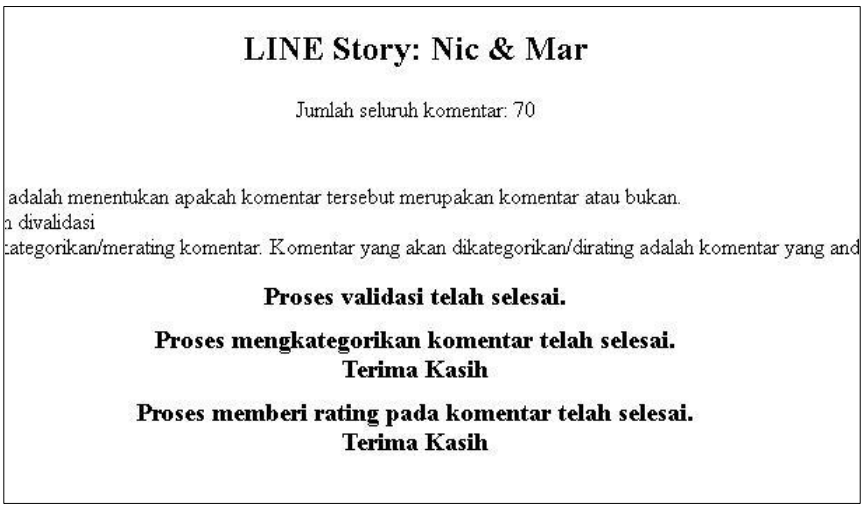
Form Rating Komentar

Page ini berisi form yang digunakan coder untuk menentukan rating untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Contoh tampilan dapat dilihat pada Gambar 34.



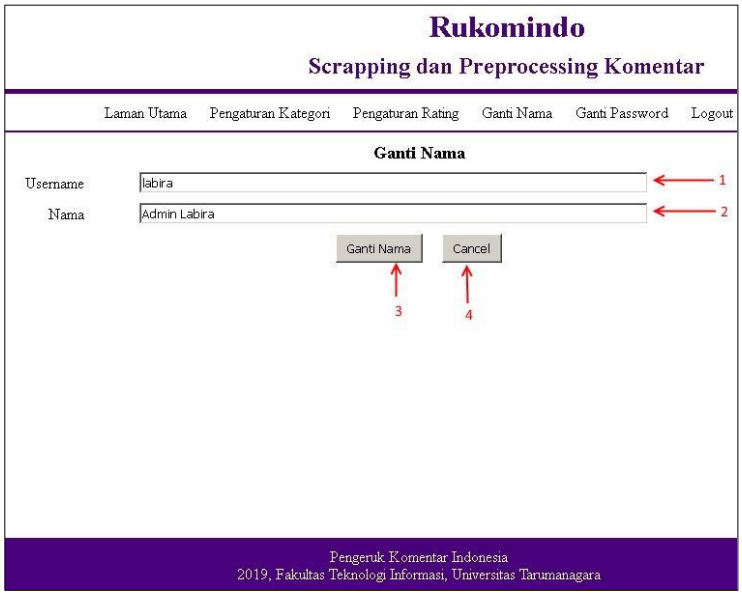
Gambar 34 Form Rating Komentar

Setelah coder selesai menentukan kategori atau rating untuk semua komentar yang valid, coder akan diarahkan kembali ke tampilan periksa komentar. Jika keduanya telah selesai, tampilan periksa komentar akan seperti pada Gambar 35. Terima kasih telah melakukan pengujian.



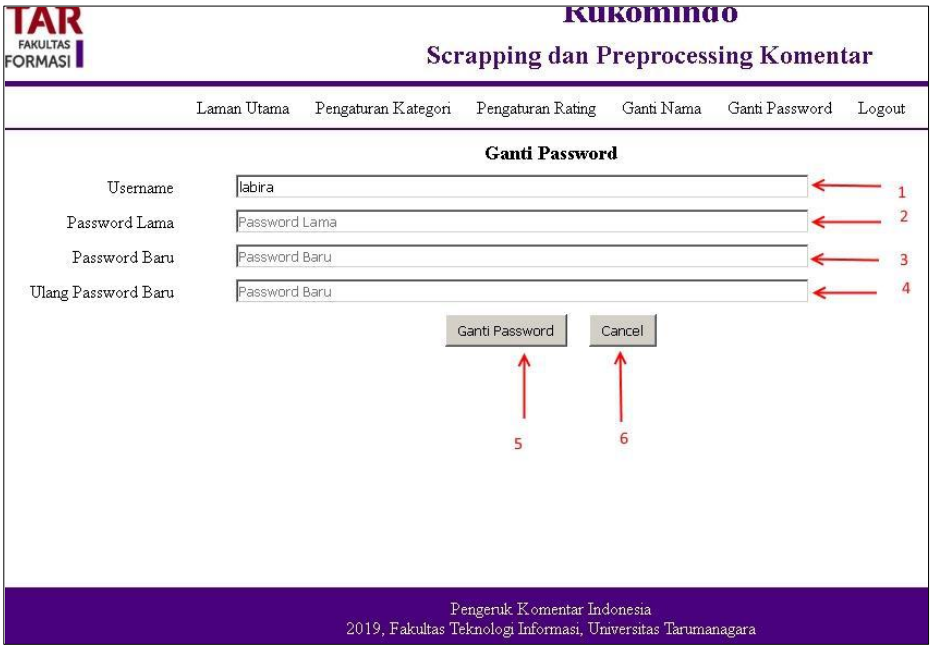
Gambar 35 Tampilan Periksa Komentar Setelah Coder Selesai Menentukan Kategori dan Rating Komentar

Fungsi lain pada aplikasi ini adalah ganti nama dan ganti password. Halaman **Ganti Nama** ini berisi form yang digunakan untuk mengganti nama user. Contoh tampilan dapat dilihat pada Gambar 36.



Gambar 36 Form Ganti Nama

Halaman **Ganti Password** ini berisi form yang digunakan untuk mengganti password user. Contoh tampilan dapat dilihat pada Gambar 37.



Gambar 37 Form Ganti Password

B. Pembahasan Pengujian

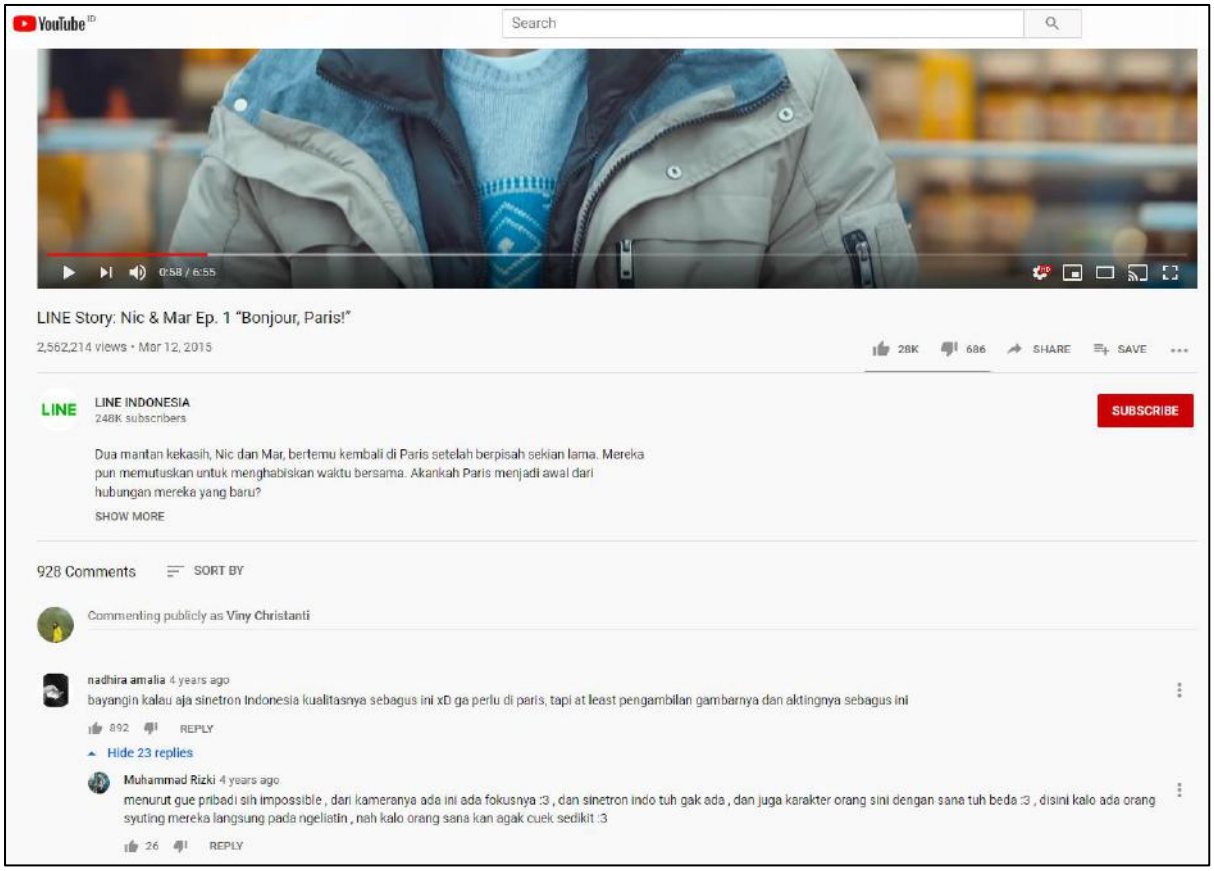
Penelitian ini diuji oleh Bapak Cokki dengan menggunakan link youtube Line Story: Nic and Marc. Pengujian ini menggunakan 7 series video dan 3 orang coder. Hasil dari pengujian ini menghasilkan 699 komentar. Rekapitulasi data komentar dapat dilihat pada tabel 4.

Tabel 4. Judul video dan jumlah komentar yang diambil

No	Judul Video	Jumlah Komentar
1	LINE Story: Nic & Mar Ep. 1 “Bonjour, Paris!”	100
2	LINE Story: Nic & Mar Ep. 2 "Close Your Eyes"	100
3	LINE Story: Nic & Mar Ep. 3 “Paris vs. Praha”	100
4	LINE Story: Nic & Mar Ep. 4 "Our Last Night?"	99
5	LINE Story: Nic & Mar Ep. 5 "Hello, Praha!"	100
6	LINE Story: Nic & Mar Ep. 6 “What Do You Want?"	100
7	LINE Story: Nic & Mar Ep. 7 "Finale"	100
	Total	699

B.1. Pengujian Scraping

Pada tabel 4 dapat dilihat komentar video nomor 4 hanya berjumlah 99 komentar. Komentar pada video “Our Last Night” hanya berisi 99 komentar utama sehingga tidak ada komentar lagi yang dapat diambil.



Gambar 38 Contoh komentar pada video nomor 1

Pada aplikasi ini komentar yang diambil dapat berupa komentar utama atau komentar dan “reply” dari user. Dimana jumlah komentar pada gambar 38 adalah jumlah komentar dan replynya yaitu 928. Pengujian yang dilakukan saat ini merupakan pengujian tanpa mengambil reply sehingga yang dihitung adalah hanya komentar utama. Pada tabel 5 dapat dilihat komentra utama adalah komentar dari nadhira amalia dan terdapat 23 reply. Pada hasil pengujian ini hanya akan mengambil komentar utama yaitu yang berada di kolom pertama.

Tabel 5 Contoh komentar utama dan reply

Komentar Utama	Reply dari komentar Nadhira Amalia
nadhira amalia4 years ago bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	 Muhammad Rizki4 years ago

892 REPLY Hide 23 replies	menurut gue pribadi sih impossible , dari kameranya ada ini ada fokusnya :3 , dan sinetron indo tuh gak ada , dan juga karakter orang sini dengan sana tuh beda :3 , disini kalo ada orang syuting mereka langsung pada ngeliatin , nah kalo orang sana kan agak cuek sedikit :3 26 REPLY Reza Dwi Saputro4 years ago Betuuul , sinetron indo pengambilan gambarnya buruk pake banget. Keliatannya engga natural ckck mungkin itu karena sekarang jamannya sinetron stripping kali ya padahal dulu sinetron 2000an ga gini gini banget kaya sekarang ueeeek. Jadi klo stripping Ceritanya gak berkualitas terus pengambilan gambar seadanya ckck 28 REPLY nadhira amalia4 years ago @reza rizky ga ada yg impossible kok. Masalahnya punya niat ga? Keliatan bgt sutradara sinetron2 disini tuh pada ga niat bikinnya. Asal jadi doang. Plot gajelas, gambarnya ancur, akting juga ga jelas x(sama tv thailand aja kita kalah jauh.. 28 REPLY

Hasil dari scraping komentar dari Youtube gambar 38 dapat dilihat pada web dan dapat juga disimpan dalam bentuk file excel. Tabel 6 adalah contoh hasil komentar utama tanpa reply yang berhasil diambil oleh penguji. Dapat dilihat pada komentar nomor 6 adalah kalimat komentar hasil scraping yang sama dengan komentar yang tercantum pada Youtube gambar 38.

Tabel 6 Contoh hasil komentar yang berhasil di scraping

No	Komentar
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))
3	no one: literally no one: Indonesian people when in Europe: "try to use American accent"
4	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?
5	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.
6	**bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini**
7	Suara nico setiap narasi itu juara bange!!!

8	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie
9	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!
10	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr

B.2. Pengujian validasi dan kategorisasi

Tahapan selanjutnya adalah proses validasi yang dilakukan oleh coder. Proses ini dilakukan untuk memvalidasi setiap komentar yang masuk apakah berupa kalimat komentar atau kalimat yang bukan merupakan pernyataan. Koordinator scraping dapat mengatur berapa jumlah coder yang akan melakukan validasi. Pada tabel 7 dapat dilihat pengujian ini mengatur 3 orang coder yang akan memvalidasai data komentar yang berhasil di scraping. Komentar tersebut akan divalidasi kemudian dikateogrikan dan dirating.

Tabel 7 Jumlah data coder dan komentar yang divalidasi, dikategori dan dirating

No	Nama Coder	Jumlah telah divalidasi	Jumlah telah dikategori	Jumlah telah dirating
1	Alwina Salim	699	486	486
2	Sean Matthew Henry	699	322	322
3	Vivian Patricia	699	332	332

Pada pengujian ini kategori yang digunakan adalah 5 kategori yaitu kategori terhadap aktor, merk, film, karakter dan penempatan produk. Kategori ini dapat diubah dan diatur di menu pengaturan kategori seperti pada gambar 39. Rating yang digunakan pada penelitian ini adalah 2 rating yaitu positif dan negative. Pada aplikasi ini rating juga dapat diubah atau ditambah seperti pada gambar 40.

Gambar 39 Pengaturan Kategori

Gambar 40 Pengaturan Rating

Hasil komentar yang sudah divalidasi, diberi kategori dan diberi rating terdapat pada tabel 8. (Hasil lengkap contoh 20 komentar dapat dilihat pada tabel 9 lampiran 3).

Tabel 8 Hasil Komentar yang sudah divalidasi

No	Isi Komentar	Valid	Kategori	Rating
5	"People should fall in love with their eyes closed"	Ya	Sikap terhadap karakter	Positif
	And I just sit here... in silence... mesmerized.			
6	**bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini**	**Ya**	**Sikap terhadap film**	**Positif**
7	Suara nico setiap narasi itu juara bange!!!	Ya	Sikap terhadap aktor	Positif

Seluruh pengambilan keputusan komentar yang divalidasi akan dihitung berdasarkan voting. Apabila dari 3 coder, 2 memilih valid dan 1 memilih tidak maka hasil keputusan validasi adalah ya. Hal tersebut akan berlaku sama untuk pemberian kategori dan rating. Apabila terjadi jumlah yang sama pada voting maka keputusan terakhir akan diambil oleh koordinator. Pada menu koordinator akan disediakan menu untuk memutuskan komentar yang hasil votingnya seri.

Aplikasi ini juga akan menghasilkan rekapitulasi jumlah komentar. Pada tabel 9, 10 dan 11 dapat dilihat rekapitulasi akhir dari ketiga coder yang sudah melakukan validasi, pengkategorian dan pemberian rating. Pada tabel 9 dapat dilihat dari 699 total seluruh komentar yang masuk hanya ada 382 komentar yang dianggap merupakan sebuah pernyataan. Sedangkan sisanya merupakan kalimat yang dianggap bukan pernyataan. Contoh kalimat yang dianggap bukan sebagai komentar terhadap film tersebut adalah “Agustus 2019 Meluncuuurr kesinii” atau “hoaaaam –“. Kalimat tersebut dianggap sebagai kalimat yang tidak berhubungan atau bukan sebuah pernyataan yang mengomentari film yang dimaksud.

Tabel 9 Jumlah komentar yang valid dan tidak

jumlah komentar	699
valid	382
tidak valid	317

Hasil dari komentar yang sudah divalidasi akan diberikan kategori sesuai dengan kategori yang telah diatur. Coder akan dengan mudah diberi menu yang dapat digunakan untuk menentukan kategori dari setiap komentar yang masuk. Setiap halaman akan memuat 10 komentar agar tidak membuat mata pembaca jenuh. Hasil rekapitulasi dari komentar valid yang sudah diberi kategori dapat dilihat pada tabel 10.

Tabel 10 Jumlah komentar valid yang sudah diberi kategori

komentar valid	382
Sikap terhadap aktor	81
Sikap terhadap karakter	57
Sikap terhadap film	214
Sikap terhadap penempatan produk	3
Sikap terhadap merek	27

Setelah komentar valid mendapatkan kategori maka selanjutnya komentar tersebut akan diberi rating apakah komentar tersebut memiliki rating positif atau negative terhadap sikap yang sudah ditentukan sebelumnya. Seperti pada kalimat berikut yang merupakan kalimat yang positif terhadap sikap film.

6	**bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini**	**Ya**	**Sikap terhadap film**	**Positif**

Rekapitulasi dari hasil komentar valid yang sudah diberi rating dapat dilihat pada tabel 11. Bedasarkan hasil rekapitulasi, jumlah komentar yang berating positif lebih banyak terhadapa film tersebut. Hasil jumlah rekapitulasi ini memudahkan bagi perusahaan atau bagi pengguna untuk mendapatkan rekomendasi apa yang akan dilakukan selanjutnya.

Tabel 11 Jumlah komentar yang valid yang telah diberi rating

Komentar valid	382
Positif	303
Negatif	79

BAB VI
KESIMPULAN DAN SARAN

A. Kesimpulan

Hasil pembuatan aplikasi ini sudah sampai kepada penarikan komentar dari Youtube, pengaturan berapa jumlah orang yang akan memvalidasi dan tampilan aplikasi yang dapat digunakan oleh coder untuk melakukan validasi. Aplikasi ini mengumpulkan data yang kemudian akan digunakan dalam tahap analisa selanjutnya. Dimana aplikasi ini berguna sebagai aplikasi pengumpulan data tahap pra-pemrosesan data mining, analisis semantic dan proses analisis lainnya.

Penelitian ini telah menghasilkan kesimpulan sebagai berikut:

1. Aplikasi scraping sudah berhasil mengumpulkan 699 komentar utama dari 7 link Youtube
2. Aplikasi dapat menghasilkan komentar yang sudah bersih data teks dari web secara otomatis.
3. Koordinator dapat mengatur berapa banyak video yang akan diambil, berapa banyak komentar yang akan diambil dan mengatur komentar diambil lengkap dengan reply atau tidak.
4. Aplikasi ini telah digunakan oleh 3 coder untuk melakukan validasi dimana setiap coder akan mendapatkan akun untuk login ke aplikasi dan memberikan validasi.
5. Setiap coder memberikan validasi, kemudian memilih kategori dari 5 kategori pada setiap komentar serta memberikan rating positif atau negate pada setiap komentar.
6. Koordinator dapat mengatur jumlah coder, jumlah kategori dan jumlah rating yang dilakukan pada setiap komentar.
7. Hasil seluruh komentar dapat disimpan menjadi sebuah file data yang kemudian dapat digunakan dalam tahap analisis selanjutnya.

B. Saran

Hasil penelitian ini telah menghasilkan aplikasi scraping yang telah diuji coba oleh Fakultas Ekonomi Program Studi Manajemen. Dalam tahap selanjutnya komentar yang

diharapkan dapat discraping adalah berita, komentar dari website seperti blogspot, situs resmi film seperti IMDB dan lainnya.

Saran selanjutnya adalah penelitian ini dapat langsung diintegrasikan dengan aplikasi data mining lainnya sehingga data komentar yang telah diperoleh dapat dilanjutkan dengan proses teks pemrosesan selanjutnya seperti stemmer, stopword removal, tagging, parsing dan lainnya. Penelitian selanjutnya penelitian ini akan menambahkan beberapa sumber pengambilan data sehingga dapat digabungkan dengan komentar dari Youtube.

DAFTAR PUSTAKA

[1] Lunden, Ingrid. "Analyst: Twitter passed 500m users in June 2012, 140m of them in US; Jakarta 'Biggest Tweeting'city." *TechCrunch RSS* 30 (2012).

[2] Lee, Yeha, Seung-Hoon Na, Jungi Kim, Sand-Hyob Nam, Hun-young Jng, and Jong-Hyeok Lee. *Kle at trec 2008 blog track: Blog post and feed retrieval*. POHANG UNIV OF SCIENCE AND TECHNOLOGY (SOUTH KOREA), 2008.

[3] Enda Nasution, Pesta Blogger, http://goblogmedia.com, 7 Februari 2010

[4] Hopkins, Jim. "Surprise! There's a third YouTube co-founder." *USA Today* 11, no. 10 (2006).

[5] Pike, Steven. *Destination marketing*. Routledge, 2012.

[6] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5, no. 1 (2012): 1-167.

[7] Hobbs, Jerry R., Donald E. Walker, and Robert A. Amsler. "Natural language access to structured text." In *Proceedings of the 9th conference on Computational linguistics-Volume 1*, pp. 127-132. Academia Praha, 1982.

[8] Mustaqim, Taufik. "Analisis dan Implementasi Penerapan Algoritma Genetika pada Information Retrieval." *Tugas Akhir. Bandung: IT Telkom* (2009).

[9] Ribeiro-Neto, Berthier, and Ricardo Baeza-Yates. "Modern information retrieval." *Addison-Wesley* 4 (1999): 107-109.

[10] Manning, Christoper D. "SchützeHinrich (1999) Foundations of statistical natural language processing."

[11] Vania, Clara (2009), *Perolehan Dokumen Blog Berbahasa Indonesia*, Jakarta : Tesis Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Indonesia

[12] Sriram, Bharath, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. "Short text classification in twitter to improve information filtering." In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pp. 841-842. ACM, 2010.

[13] Elsas, Jonathan L., Jaime Arguello, Jamie Callan, and Jaime G. Carbonell. "Retrieval and feedback models for blog feed search." In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 347-354. ACM, 2008.

[14] Broder, Andrei. "A taxonomy of web search." In *ACM Sigir forum*, vol. 36, no. 2, pp. 3-10. ACM, 2002.

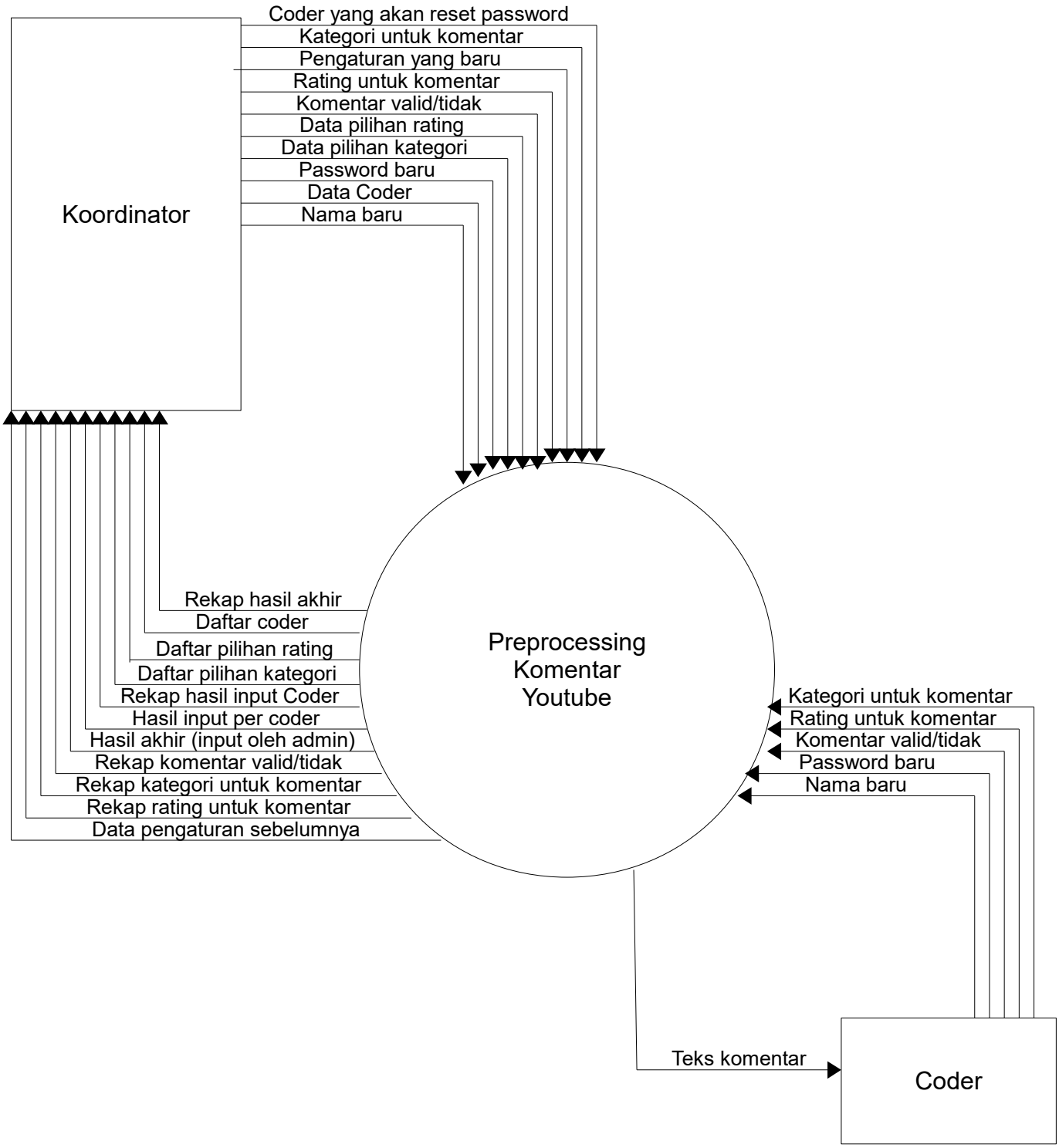
[15] Adriani, Mirna, and Rama Pandugita. "Using the Web information structure for retrieving Web pages." In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pp. 892-897. Springer, Berlin, Heidelberg, 2005.

[16] Westerveld, Thijs, Wessel Kraaij, and Djoerd Hiemstra. "Retrieving web pages using content, links, urls and anchors." In *TREC*, vol. 1, pp. 663-672. 2001.

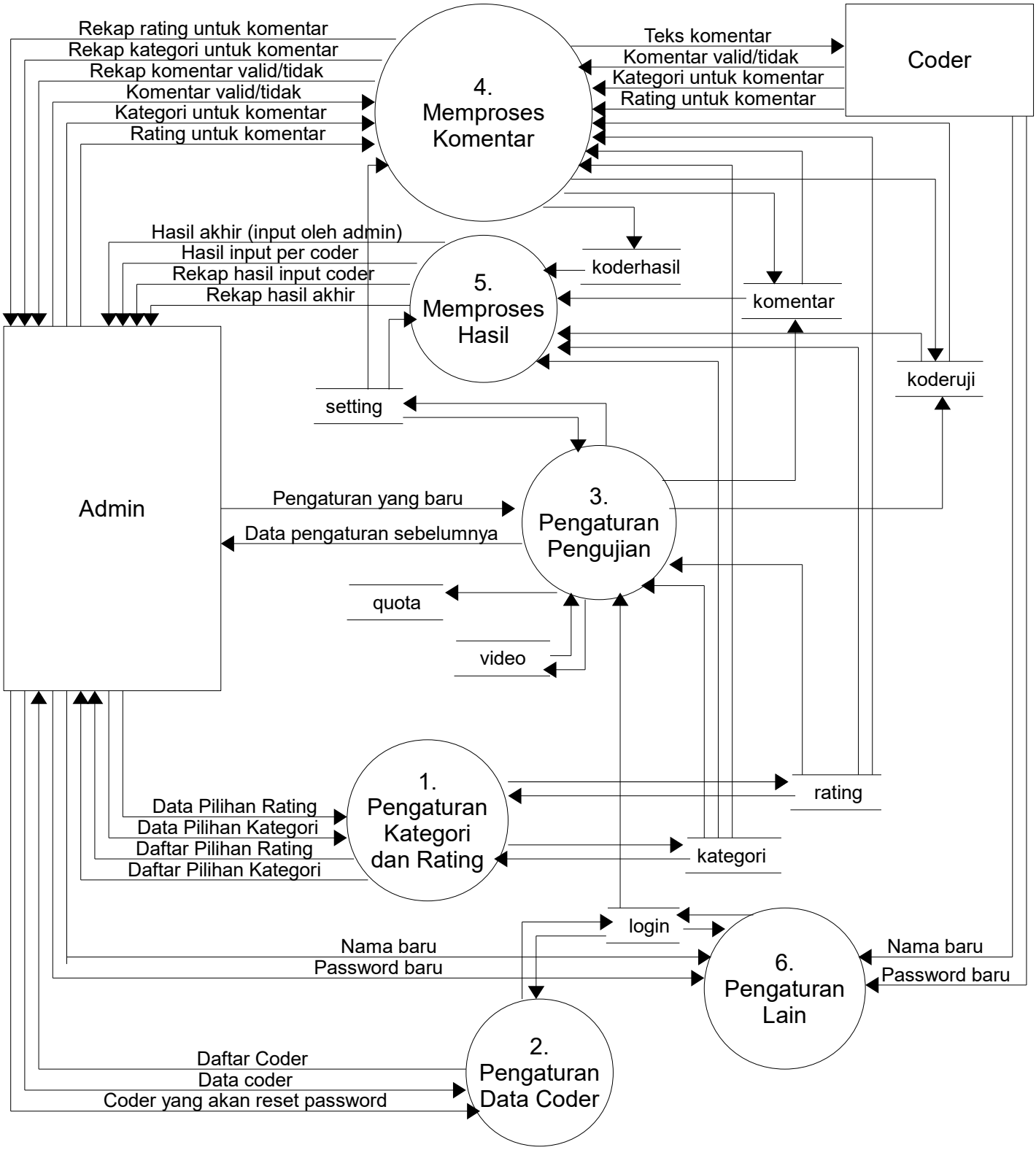
[17] Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab, 1999.

Lampiran I
Context Diagram

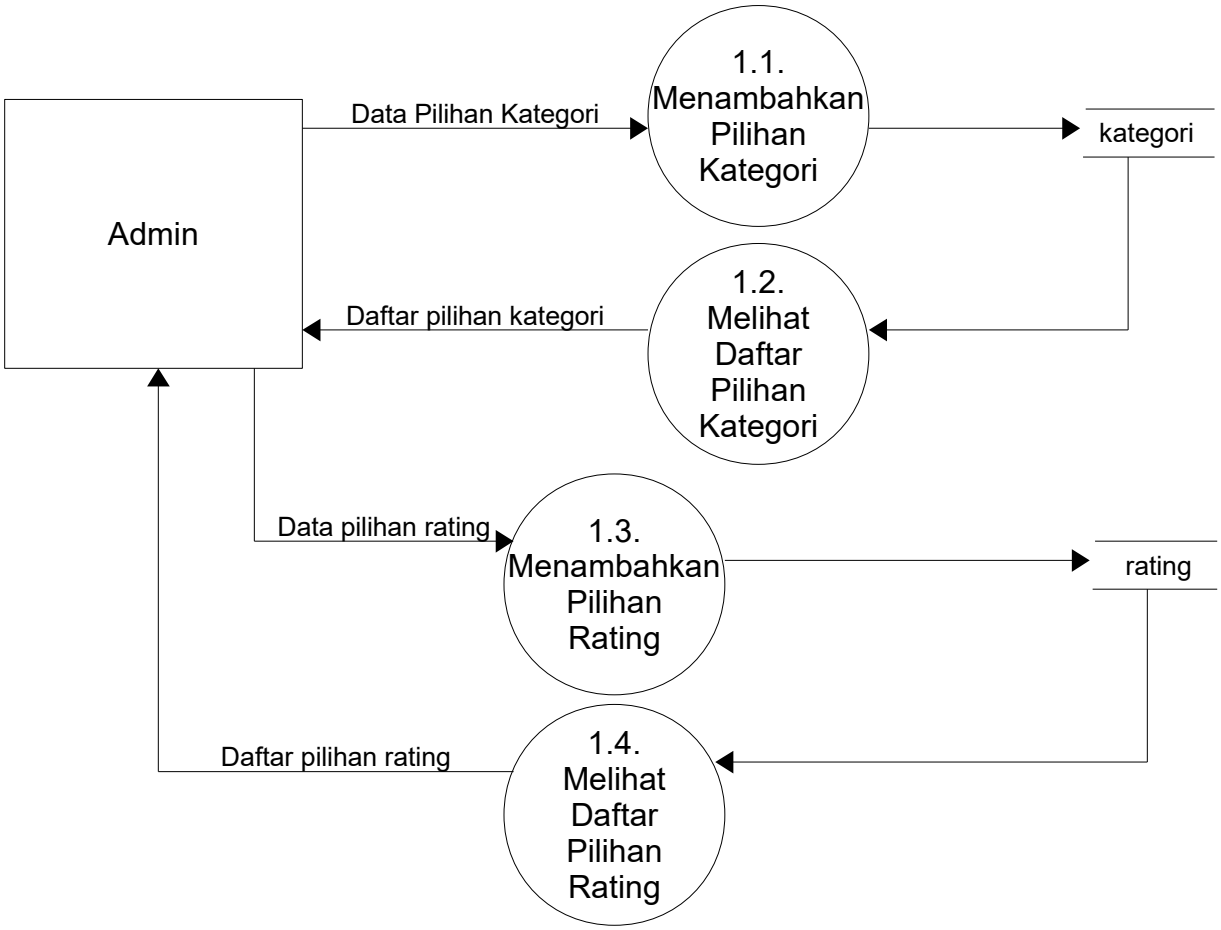
CONTEXT DIAGRAM



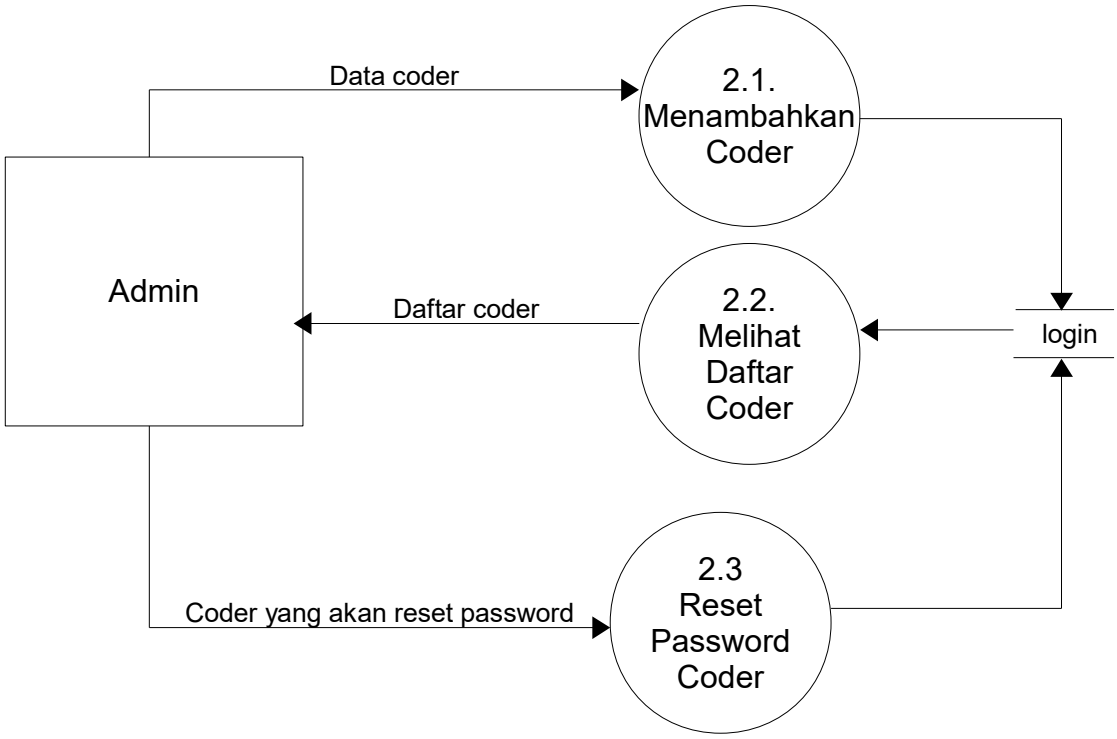
DFD Level 1


Rekap rating untuk komentar
Rekap kategori untuk komentar
Rekap komentar valid/tidak
Komentar valid/tidak
Kategori untuk komentar
Rating untuk komentar
4. Memproses Komentar
Teks komentar
Komentar valid/tidak
Kategori untuk komentar
Rating untuk komentar
Coder
Hasil akhir (input oleh admin)
Hasil input per coder
Rekap hasil input coder
Rekap hasil akhir
5. Memproses Hasil
koderhasil
komentar
koderuji
setting
Admin
Pengaturan yang baru
Data pengaturan sebelumnya
3. Pengaturan Pengujian
quota
video
1. Pengaturan Kategori dan Rating
Data Pilihan Rating
Data Pilihan Kategori
Daftar Pilihan Rating
Daftar Pilihan Kategori
rating
kategori
login
Nama baru
Password baru
6. Pengaturan Lain
Nama baru
Password baru
2. Pengaturan Data Coder
Daftar Coder
Data coder
Coder yang akan reset password

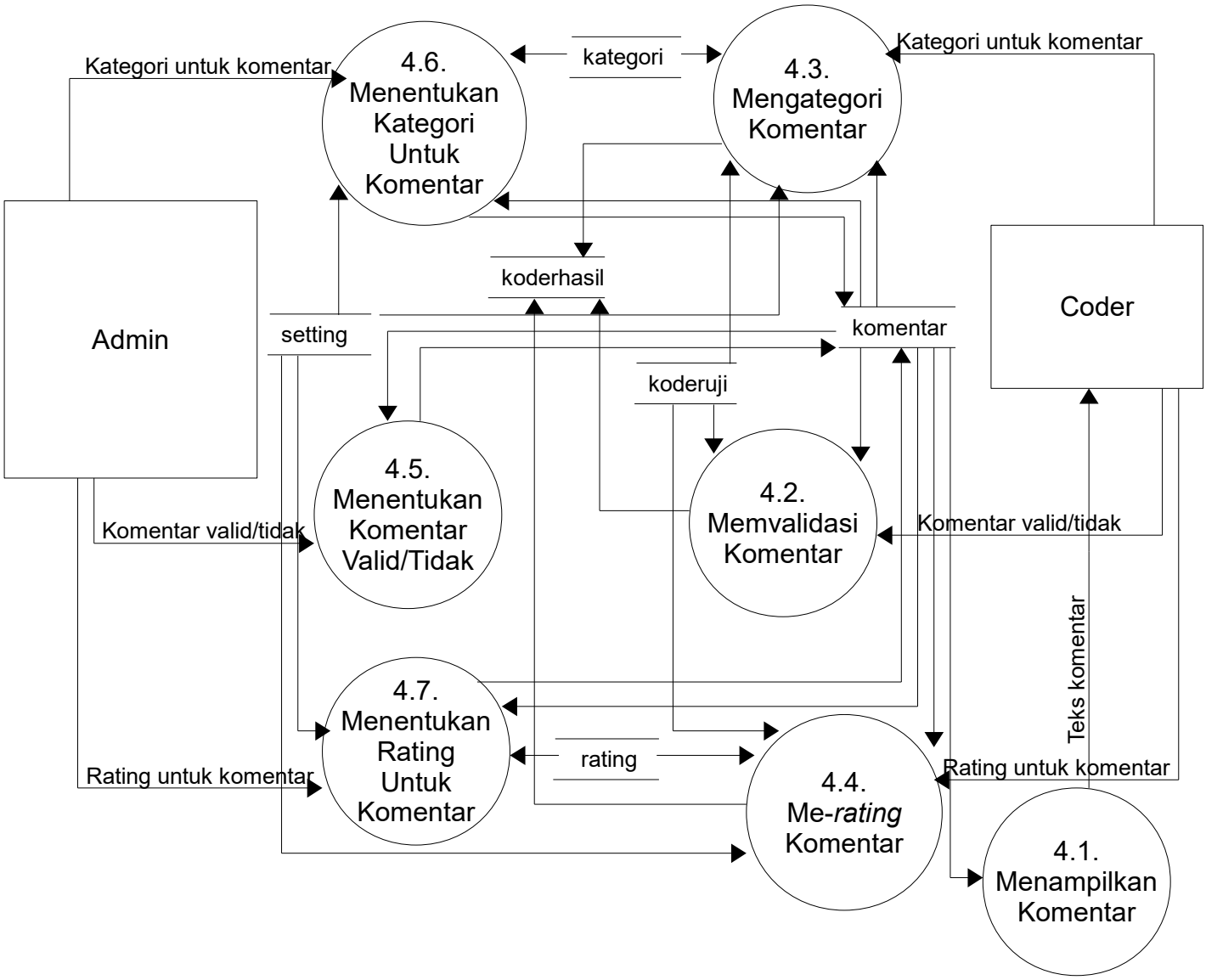
DFD Level 2 Proses 1



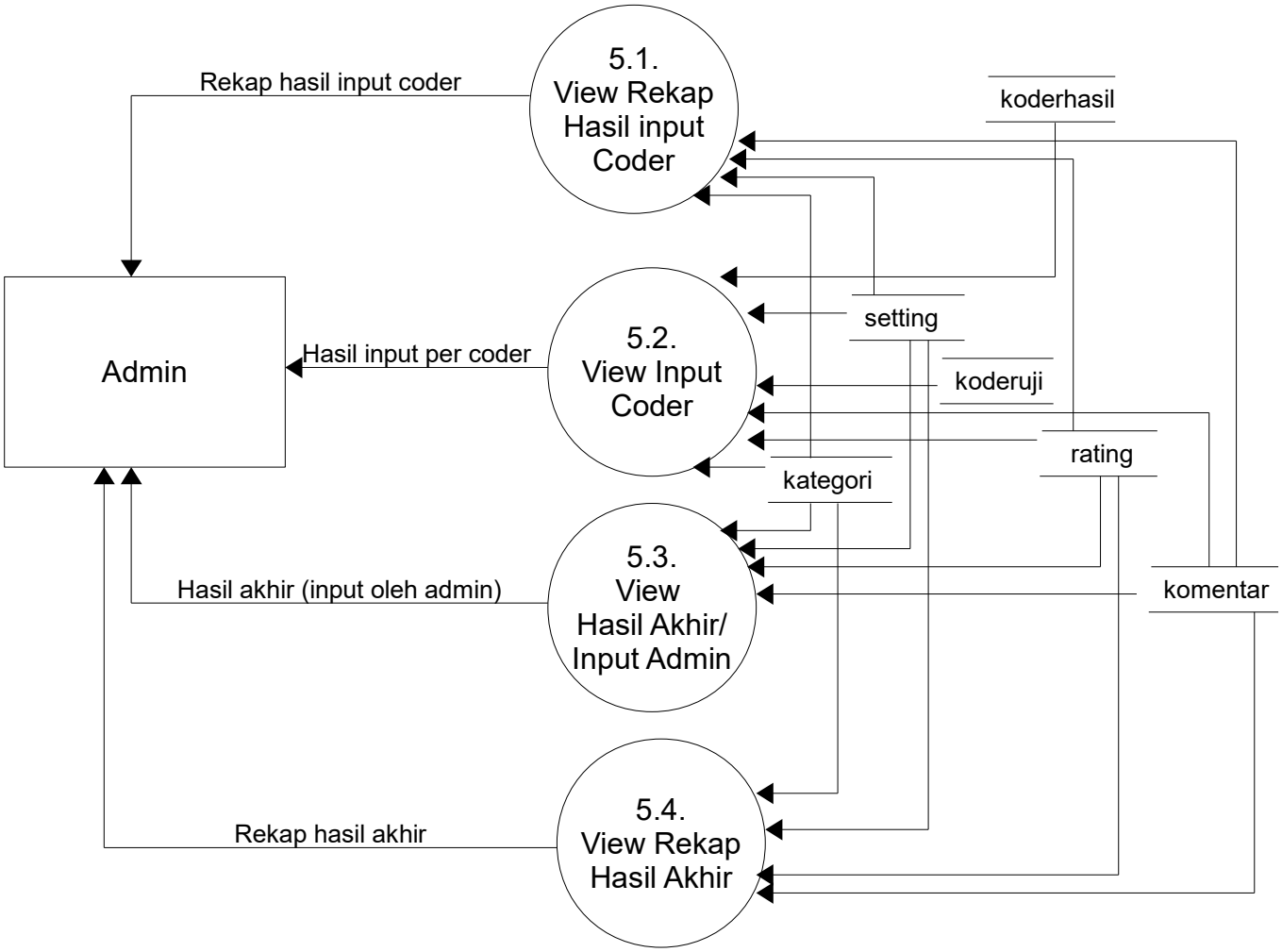
DFD Level 2 Proses 2



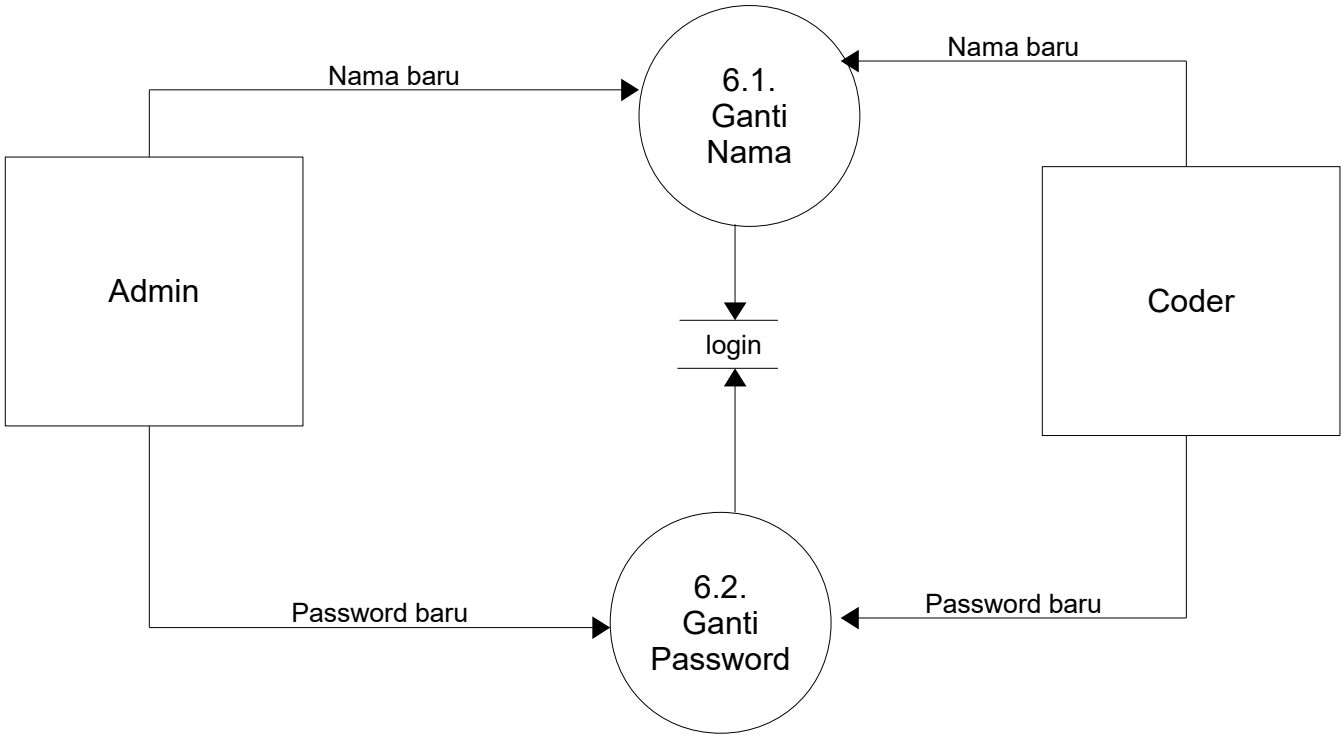
DFD Level 2 Proses 4

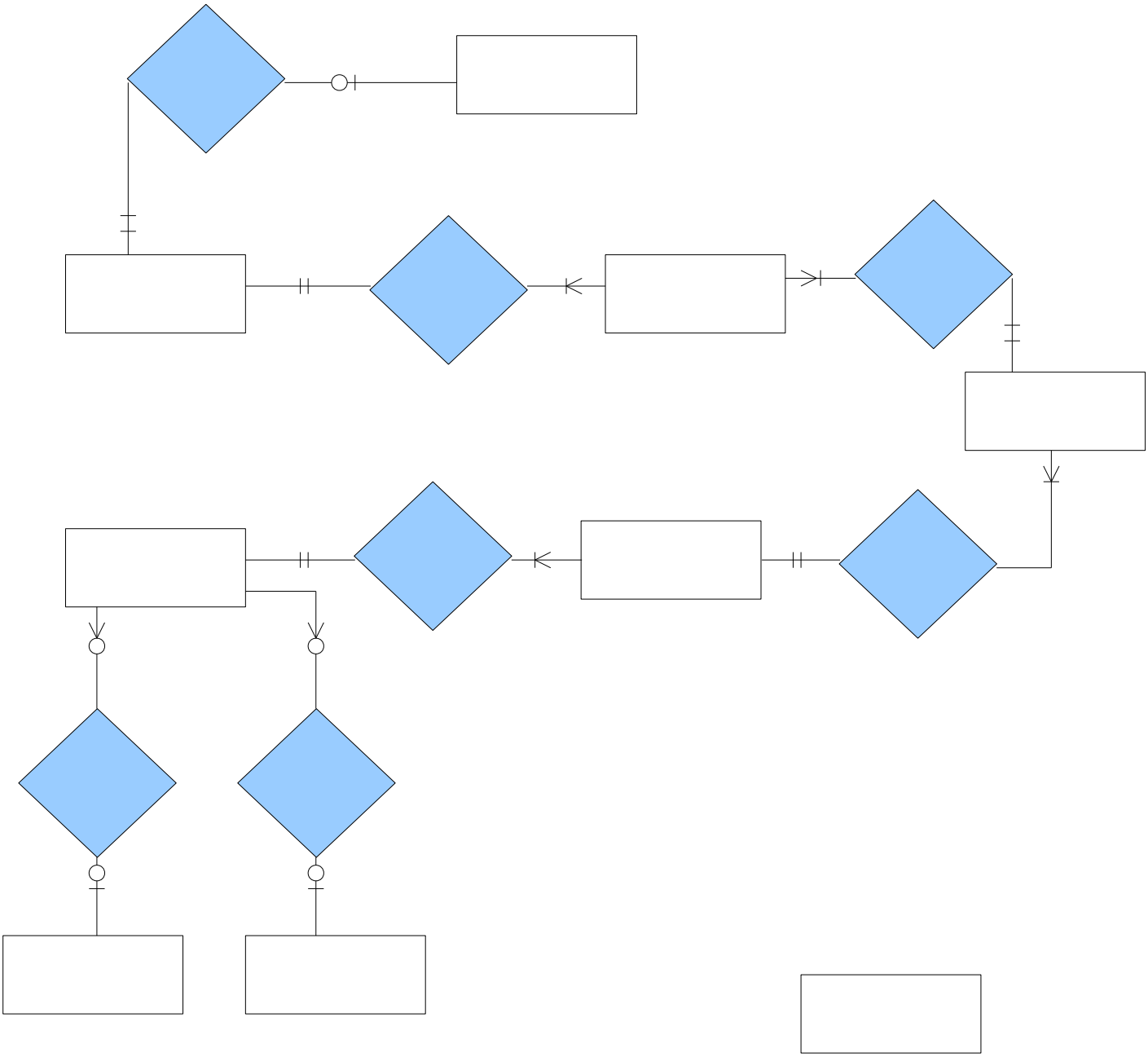


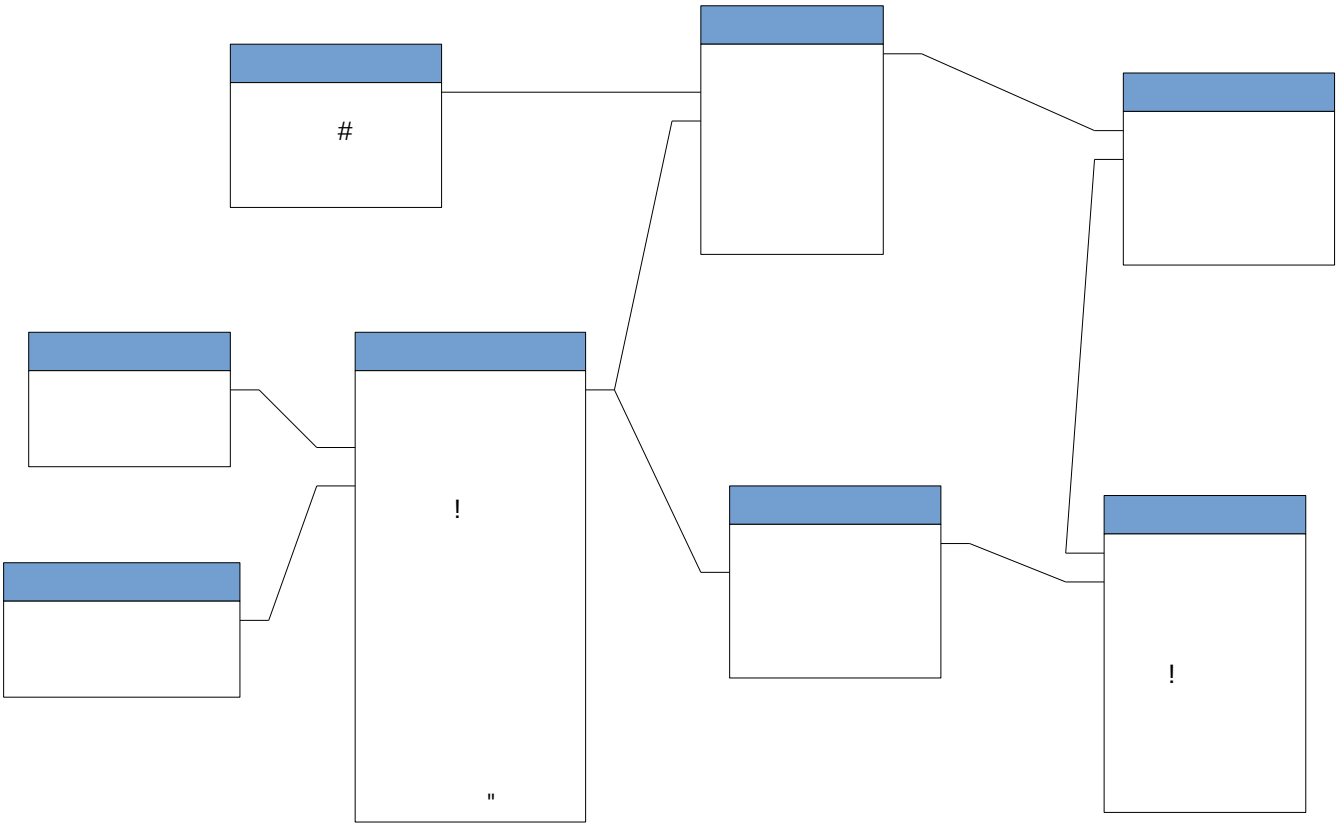
DFD Level 2 Proses 5

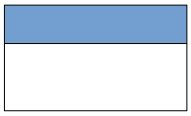


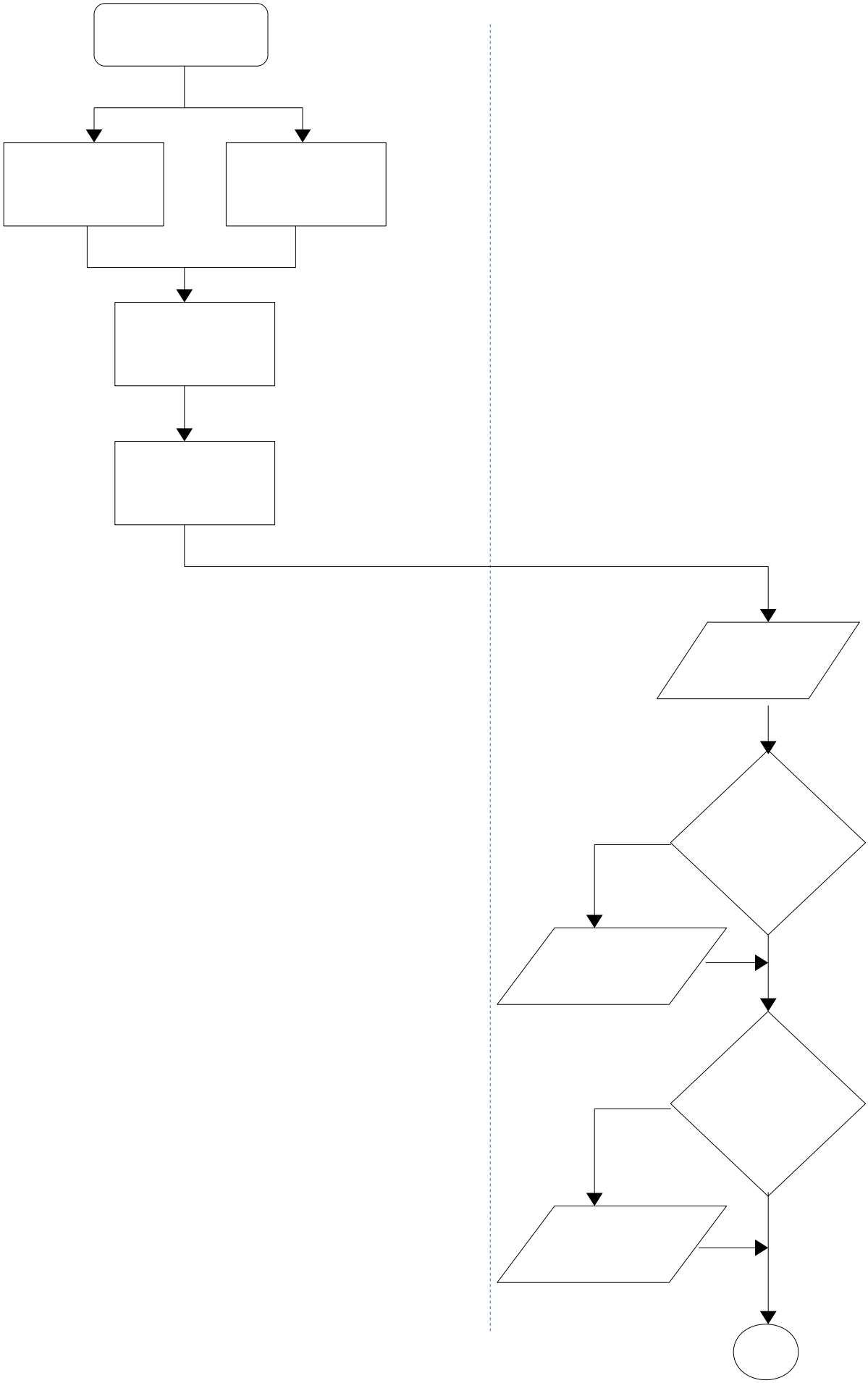
DFD Level 2 Proses 6

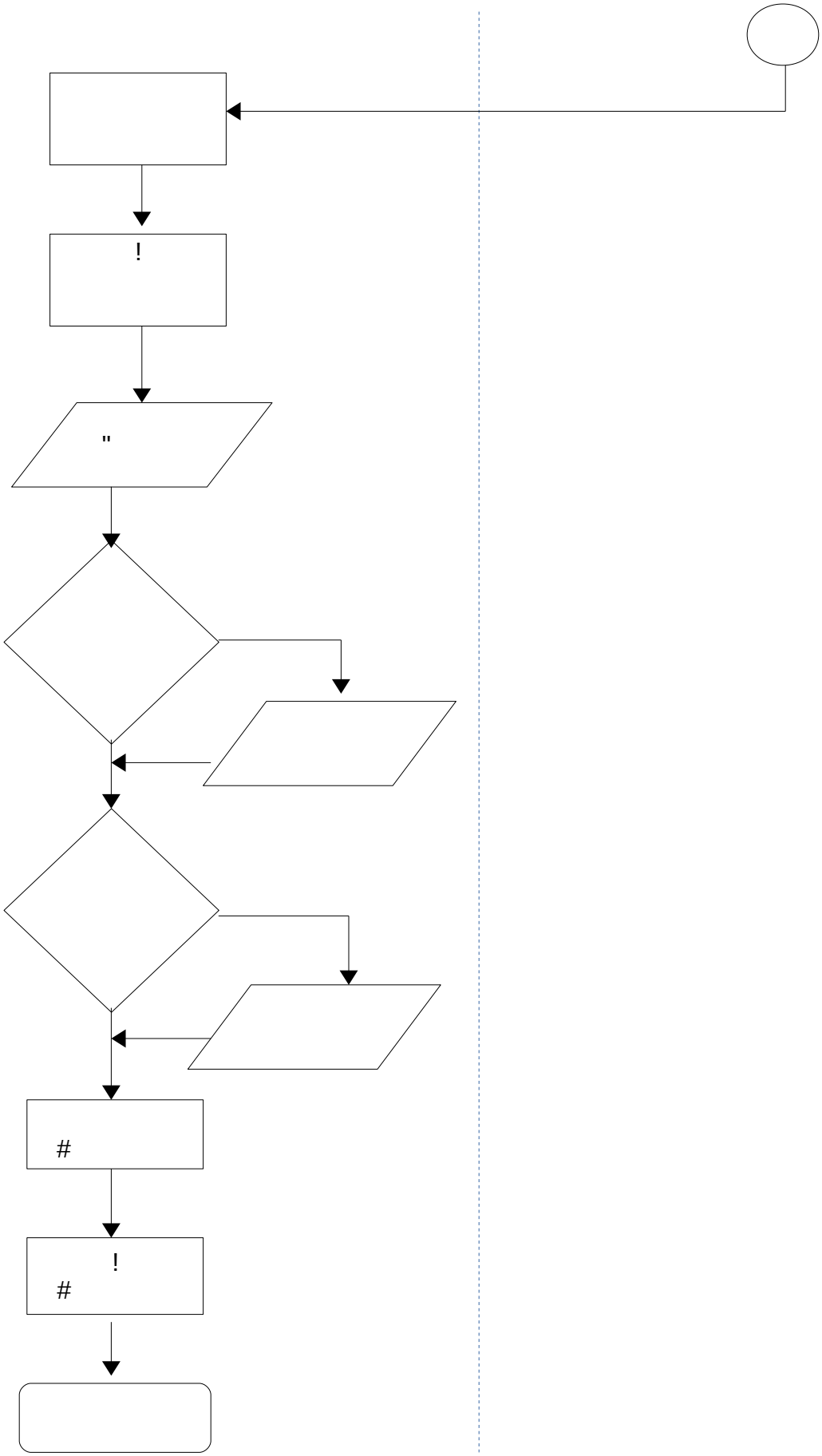












Spesifikasi Tabel

A. Tabel table_kategori

Tabel untuk menyimpan data pilihan kategori

No	Nama Kolom	Tipe	Keterangan
1	**id_kategori**	char(3)	Kode untuk pilihan kategori: K + 2 digit angka
2	nama_kategori	varchar(100)	Nama untuk pilihan kategori
3	teks_kategori	json	Daftar pilihan kategori

B. Tabel table_koderhasil

Tabel untuk menyimpan hasil input dari para koder

No	Nama Kolom	Tipe	Keterangan
1	**no_koder**	char(9)	Nomor koder
2	**no_komentar**	char(7)	Nomor komentar
3	valid	tinyint	Penanda, 0:bukan komentar, 1:komentar
4	rating	int	Nomor rating yang diberikan
5	kategori	int	Nomor kategori yang diberikan

C. Tabel table_koderuji

Tabel untuk menyimpan data koder yang dilibatkan dalam tahap pengujian

No	Nama Kolom	Tipe	Keterangan
1	**no_koder**	int	Nomor untuk koder
2	koder	varchar(20)	Username koder
3	id_setting	int	Id untuk pengaturan pengujian
4	prog_valid	int	Progress validasi komentar
5	prog_kategori	int	Progress penentuan kategori komentar
6	prog_rating	int	Progress penentuan rating komentar
7	count_valid	int	Jumlah komentar yang dinyatakan valid oleh koder

D. Tabel table_komentar

Tabel untuk menyimpan data komentar yang digunakan untuk pengujian

No	Nama Kolom	Tipe	Keterangan
1	**no_komentar**	int	Nomor untuk komentar
2	no_video	int	Nomor untuk video
3	idkomenyoutube	varchar(255)	ID komentar youtube
4	parentkomen	varchar(255)	ID komentar dari parent komentar (menandakan ini

No	Nama Kolom	Tipe	Keterangan
			adalah komentar balasan)
5	textkomen	text	Isi komentar
6	jmllike	int	Jumlah like untuk komentar
7	rating_akhir	int	Rating yang diberikan oleh admin
8	kategori_akhir	int	Kategori yang diberikan oleh admin
9	valid_akhir	tinyint	Komentar valid atau tidak yang diberikan oleh admin

E. Tabel table_quota

Tabel untuk menghitung kuota penggunaan api youtube

No	Nama Kolom	Tipe	Keterangan
1	**tgl_use**	date	Tanggal pemanggilan api
2	quota	int	Jumlah kuota yang digunakan

F. Tabel table_rating

Tabel untuk menyimpan data pilihan rating

No	Nama Kolom	Tipe	Keterangan
1	**id_rating**	char(3)	Kode untuk pilihan rating: R + 2 digit angka
2	nama_rating	varchar(100)	Nama untuk pilihan rating
3	teks_rating	json	Daftar pilihan rating

G. Tabel table_setting

Tabel untuk menyimpan pengaturan pengujian

No	Nama Kolom	Tipe	Keterangan
1	**id**	int	Id untuk pengaturan pengujian
2	nama	varchar(255)	Nama/Judul untuk pengaturan pengujian
3	opsirating	char(3)	Kode untuk pilihan rating yang akan digunakan
4	opsikategori	char(3)	Kode untuk pilihan kategori yang akan digunakan
5	opsimaxkomentar	int	Jumlah komentar utama yang diambil dari setiap video
6	jml_video	int	Jumlah video yang menjadi sumber komentar
7	jml_komentar	int	Jumlah komentar yang terambil
8	listkoder	json	Daftar koder yang melakukan pengujian
9	opsiorder	varchar(20)	Opsi urutan komentar yang diambil dari youtube.
10	cek_valid	int	Progress validasi komentar oleh admin
11	cek_kategori	int	Progress penentuan kategori komentar oleh admin

No	Nama Kolom	Tipe	Keterangan
12	cek_rating	int	Progress penentuan rating komentar oleh admin
13	count_valid	int	Jumlah komentar yang dinyakan valid oleh admin
14	ambil_reply	tinyint	0: tidak ambil reply, 1: ambil reply
15	aktif	tinyint	Penanda pengujian yang digunakan untuk coder

H. Tabel table_video

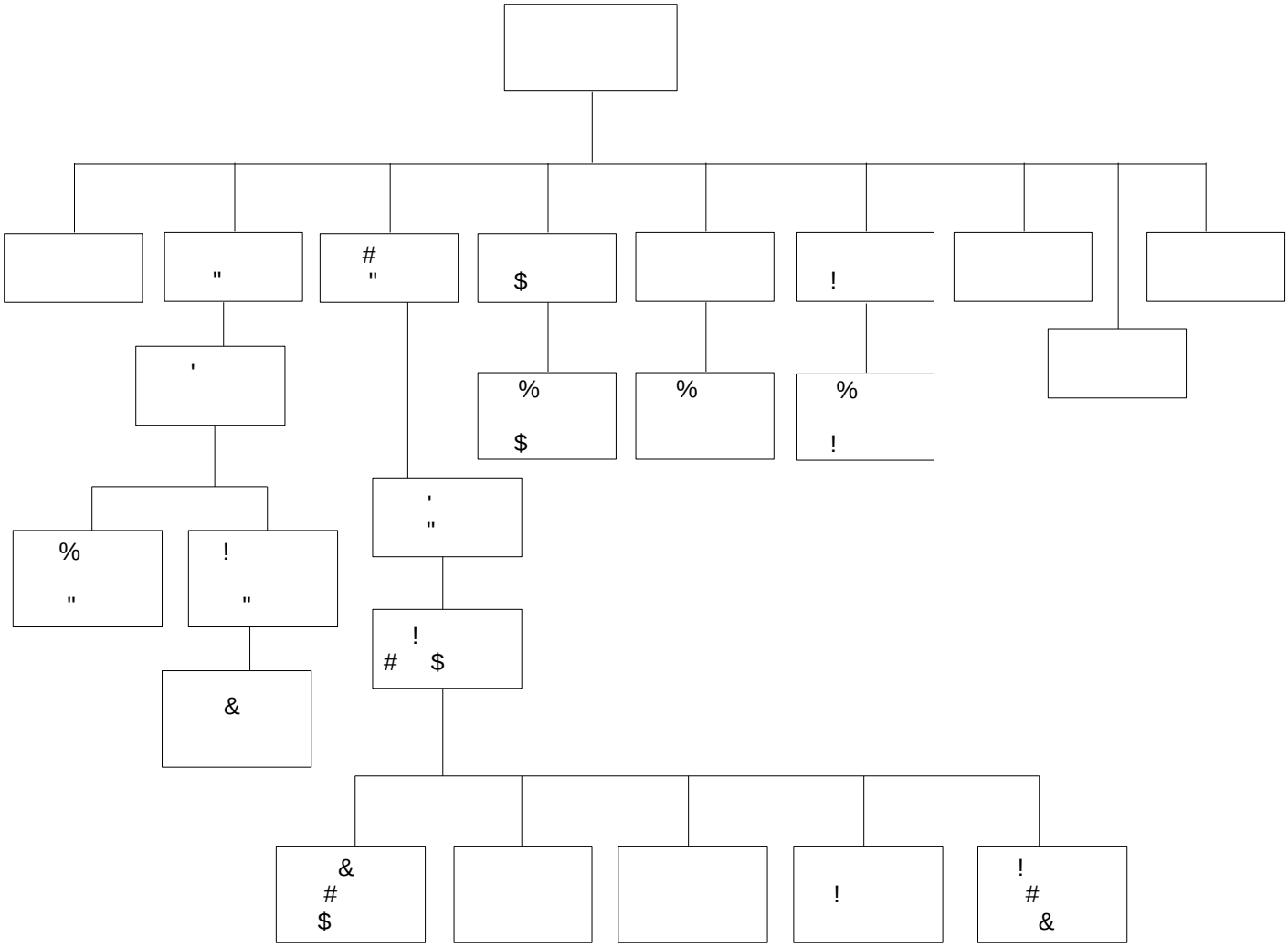
Tabel untuk menyimpan data video yang digunakan saat pengujian

No	Nama Kolom	Tipe	Keterangan
1	**no_video**	int	Nomor untuk video
2	id_setting	int	Id pengaturan pengujian
3	idvideoyoutube	varchar(255)	Id video dari youtube
4	judulvideo	varchar(255)	Judul video
5	jumlahkomen	int	Jumlah komentar yang terambil dari video tersebut

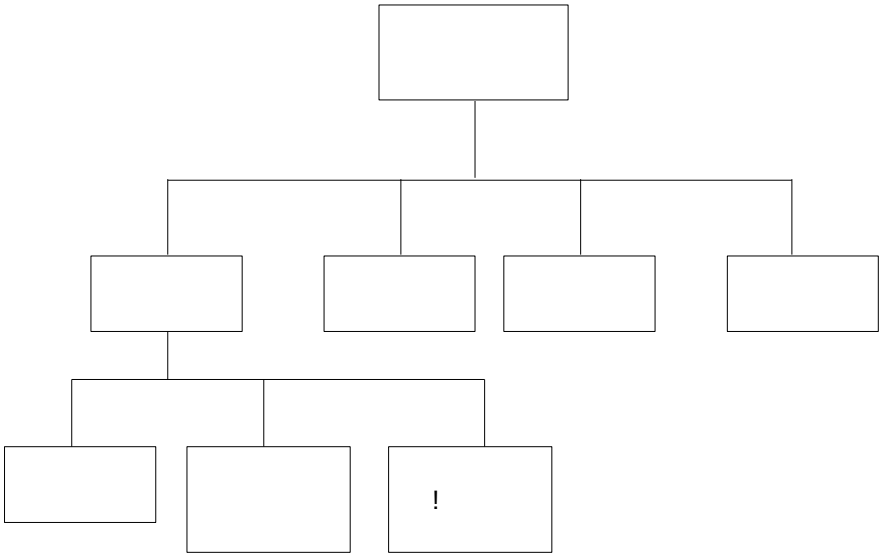
I. Tabel tbl_login

Tabel untuk menyimpan data untuk login user

No	Nama Kolom	Tipe	Keterangan
1	**username**	varchar(20)	Username untuk login
2	password	varchar(255)	Password untuk login
3	tipe	tinyint	1: admin, 2: koder
4	nama	varchar(255)	Nama user



$

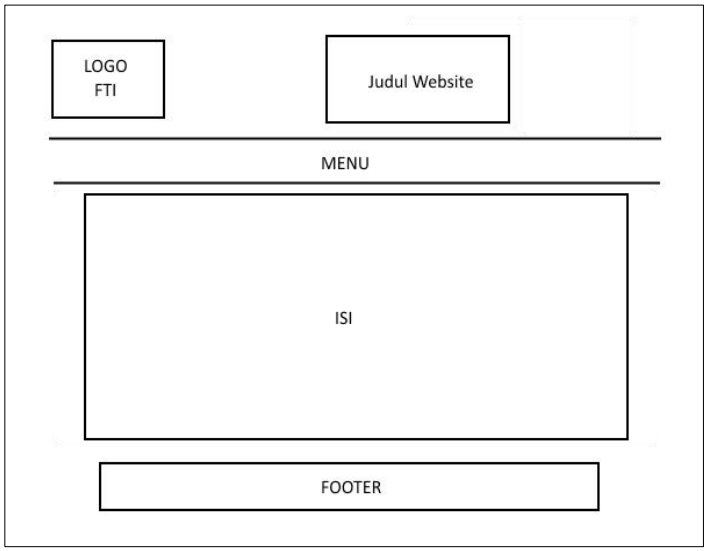


Lampiran II
Rancangan Interface

Rancangan Umum

Rancangan setiap *page* dapat dilihat pada Gambar 1, yang terdiri atas:

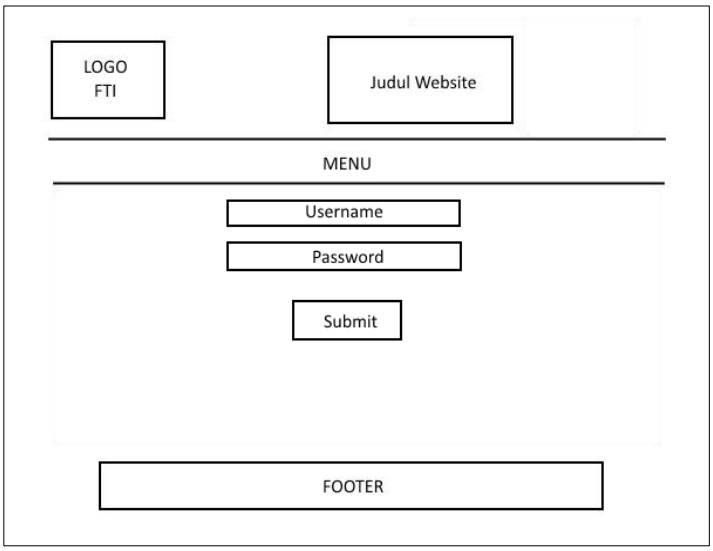
1. Logo FTI Untar
2. Judul Website
3. Menu, disesuaikan dengan tipe akun yang digunakan
4. Isi dari setiap *page*
5. Footer



Gambar 1 Rancangan Umum Semua *Page*

Home

Home dari website ini berupa form login seperti terlihat pada Gambar 2, karena tidak ada *page* yang diperuntukkan bagi *guest*. Form ini terdiri atas 2 *field*, yaitu username dan password, serta 1 *button* untuk submit.

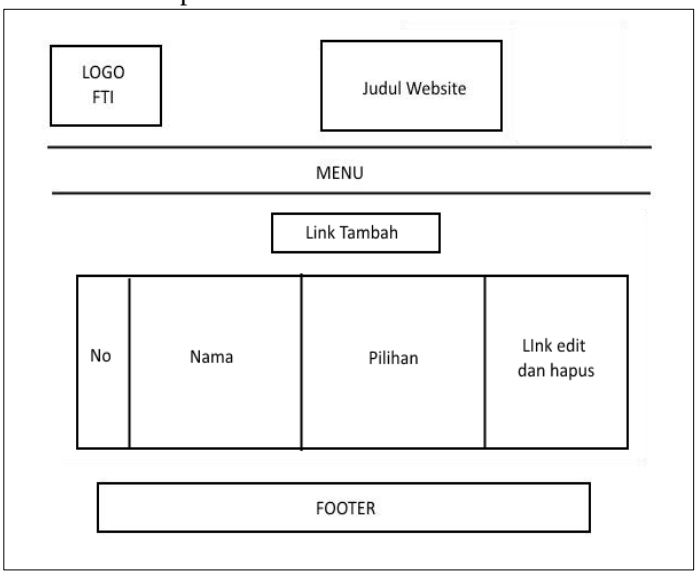


Gambar 2 Rancangan Home

Daftar Pilihan Kategori dan Rating

Rancangan berikut digunakan pada dua tampilan, yaitu daftar pilihan kategori dan daftar pilihan rating. Pada tampilan ini terdapat sebuah link untuk menambahkan data baru, serta tabel yang berisi data pilihan kategori/rating yang telah diinput, seperti terlihat pada Gambar 3. Tabel terdiri atas empat kolom, yaitu:

1. Nomor
2. Nama, yaitu judul yang diberikan untuk pilihan kategori/rating
3. Pilihan, berisi pilihan kategori/rating
4. Link, untuk edit/ubah dan hapus data

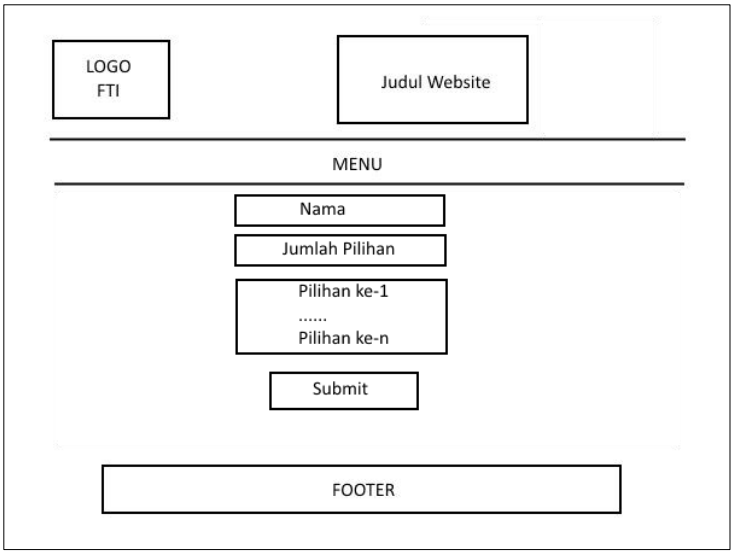


Gambar 3 Rancangan Daftar Pilihan Kategori dan Rating

Form Pilihan Kategori dan Rating

Rancangan berikut digunakan pada dua tampilan, yaitu form pilihan kategori dan form pilihan rating. Form ini digunakan untuk menambah atau mengubah data pilhan kategori/rating. Rancangan dapat dilihat pada Gambar 4. Form ini terdiri atas:

1. *Field* nama, untuk pilihan kategori/rating
2. *Field* jumlah pilihan, jumlah kategori/rating yang dapat dipilih
3. *Field* untuk input teks pilihan kategori/rating, *field* ini akan berjumlah sebanyak isi *field* jumlah pilihan
4. *Button* submit

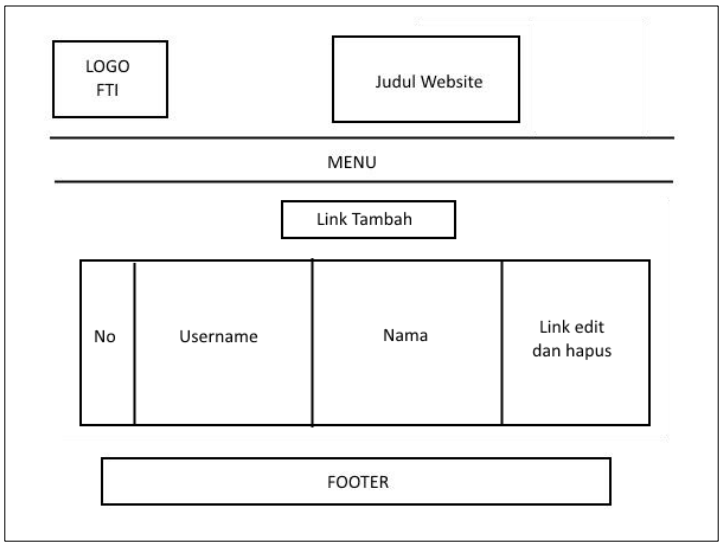


Gambar 4 Rancangan Form Pilihan Kategori dan Rating

Daftar Coder

Pada tampilan ini terdapat sebuah link untuk menambahkan data baru, serta tabel yang berisi data coder yang telah diinput, seperti terlihat pada Gambar 5. Tabel terdiri atas empat kolom, yaitu:

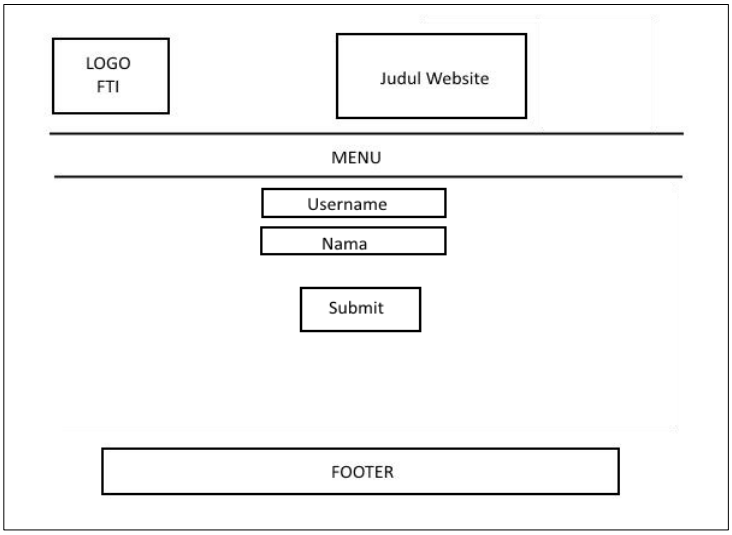
1. Nomor
2. Username
3. Nama
4. Link, untuk edit/ubah dan hapus data



Gambar 5 Rancangan Daftar Coder

Form Coder

Form ini digunakan untuk menambah atau mengubah data coder. Rancangan dapat dilihat pada Gambar 6. Form ini hanya terdiri atas 2 *field*, yaitu username (yang digunakan untuk login) dan nama. Password akan dihasilkan secara default oleh sistem.



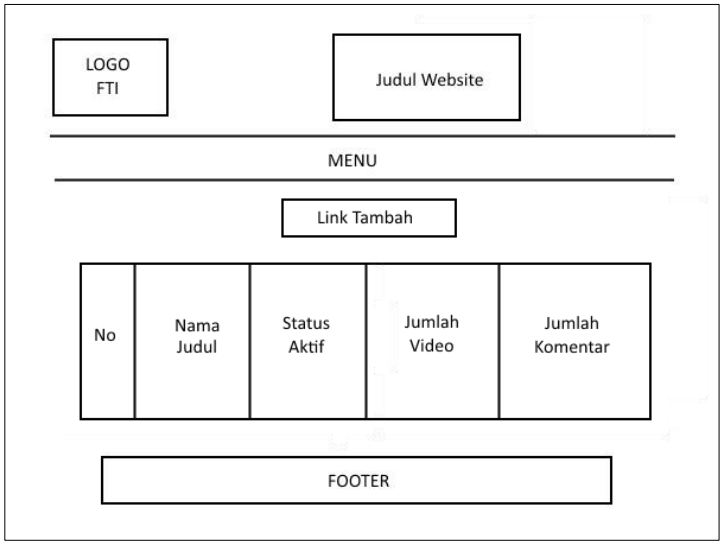
Gambar 6 Rancangan Form Coder

Daftar Pengujian

Rancangan ini digunakan pada daftar pengaturan pengujian. Pada tampilan ini terdapat sebuah link untuk menambahkan data baru, serta tabel yang berisi data pengaturan pengujian yang telah diinput, seperti terlihat pada Gambar 7. Tabel terdiri atas lima kolom, yaitu:

1. Nomor

2. Nama judul, klik untuk melihat rincian pengaturan pengujian
3. Status aktif, menandakan pengujian mana yang akan dilakukan oleh coder
4. Jumlah video, yang menjadi sumber komentar
5. Jumlah komentar, yang akan diproses oleh coder.

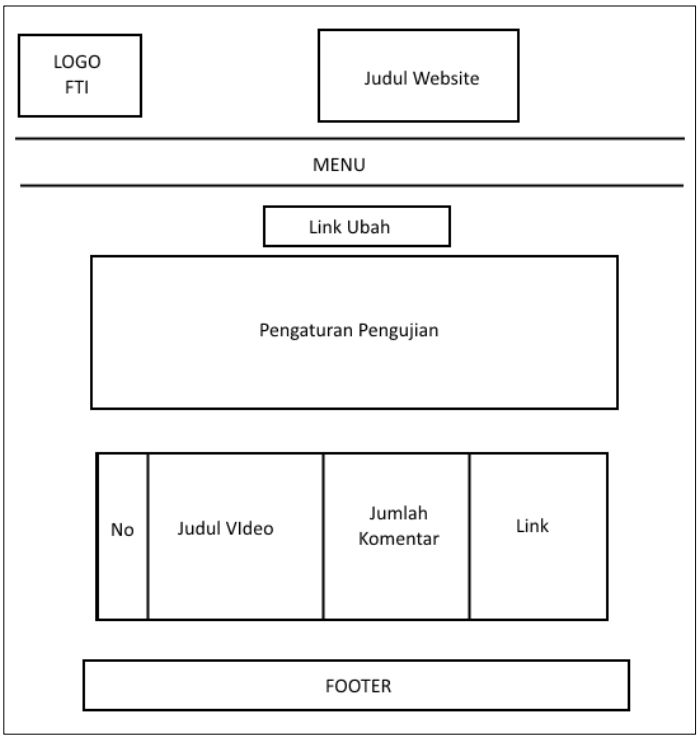


Gambar 7 Rancangan Daftar Pengaturan Pengujian

Rincian Pengaturan Pengujian

Data yang ditampilkan pada tampilan ini adalah rincian dari pengaturan pengujian tersebut. Selain data pengaturan, juga terdapat link untuk mengubah pengaturan tersebut serta tabel yang berisi data terkait video yang menjadi sumber komentar, seperti terlihat pada Gambar 8. Tabel terdiri atas empat kolom, yaitu:

1. Nomor
2. Judul video, judul video youtube
3. Jumlah komentar, jumlah komentar yang diambil dari setiap video
4. Link, untuk melihat seluruh komentar yang diambil dari setiap video.

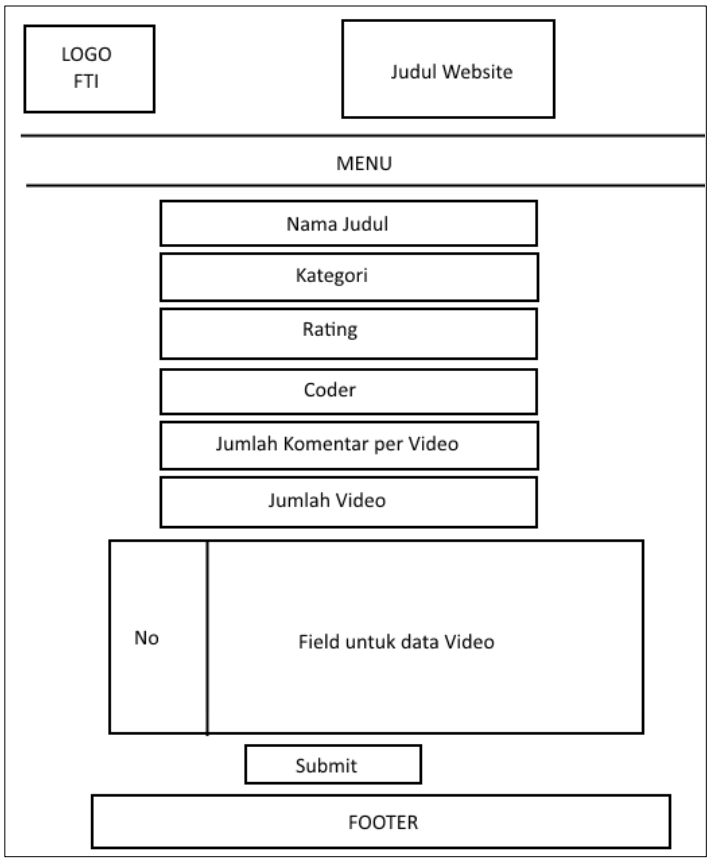


Gambar 8 Rancangan Rincian Pengaturan Pengujian

Form Pengaturan Pengujian

Form ini digunakan untuk menambah atau mengubah data pilhan pengaturan pengujian. Rancangan dapat dilihat pada Gambar 9. Form ini terdiri atas:

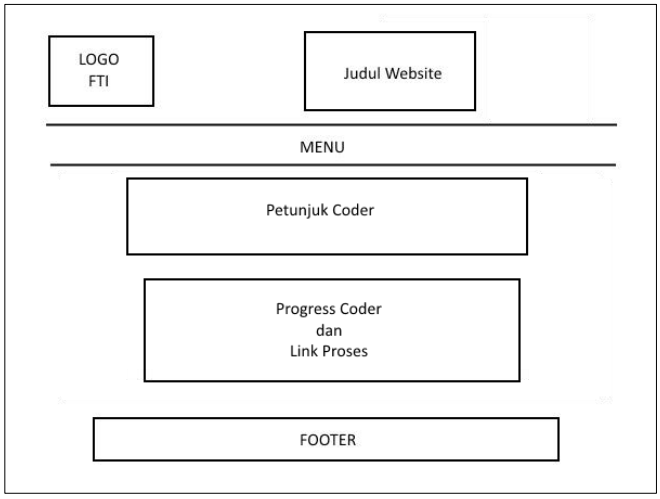
1. *Field* nama judul, yaitu nama yang diberikan untuk pengaturan tersebut
2. *Field* kategori, pilihan kategori yang akan digunakan pada pengujian
3. *Field* rating, pilihan rating yang akan digunakan pada pengujian
4. *Field* coder, untuk menentukan coder yang ikut tahap pengujian
5. *Field* jumlah komentar per video, menandakan jumlah komentar yang akan diambil dari setiap video
6. *Field* jumlah video, yaitu jumlah video yang akan diambil komentarnya
7. Tabel untuk menginput data terkait video seperti url atau id video youtube.
8. *Button* submit



Gambar 9 Rancangan Form Pengaturan Pengujian

Tampilan Awal Pengujian Coder

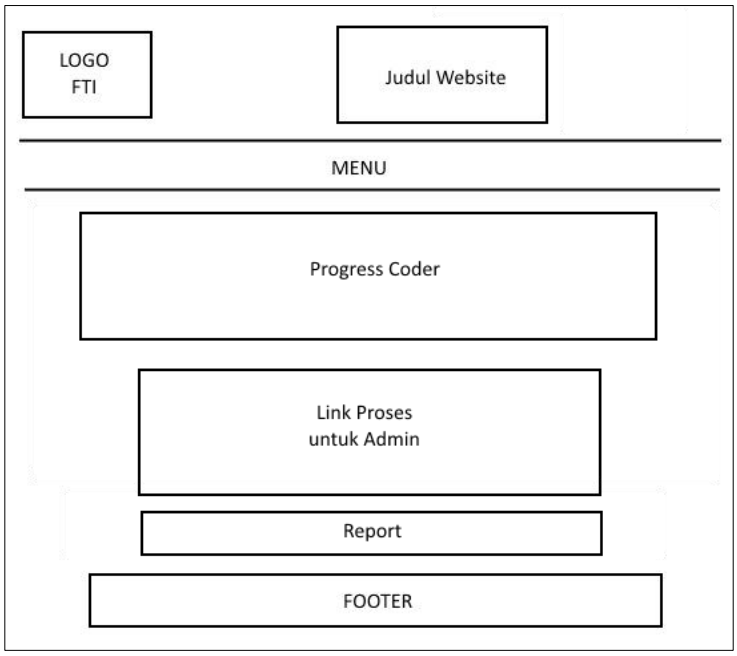
Ini adalah tampilan untuk coder sebelum/setelah melakukan pengujian. Rancangan dapat dilihat pada Gambar 10. Form ini terdiri atas petunjuk untuk coder dan link untuk memulai proses pengujian. Ada 3 (tiga) tahap, yaitu validasi komentar, menentukan kategori komentar, dan menentukan rating komentar.



Gambar 10 Rancangan Tampilan Awal Pengujian Coder

Progress Coder

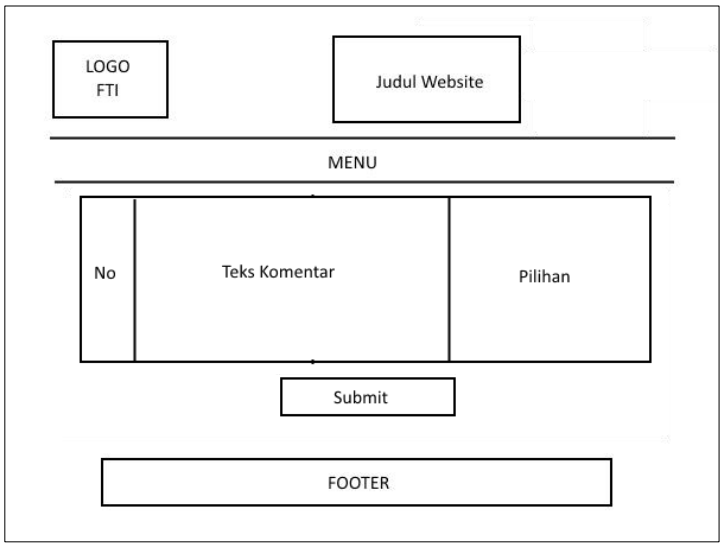
Ini adalah tampilan pada admin untuk melihat progress pengujian yang dilakukan oleh coder. Rancangan dapat dilihat pada Gambar 11. Selain melihat progress pengujian, admin juga akan memproses setiap komentar berdasarkan hasil input dari para coder.



Gambar 11 Rancangan Progress Coder

Form Proses Komentar

Ini adalah form yang digunakan oleh coder dan admin untuk memproses komentar, yang terdiri atas validasi komentar, menentukan kategori komentar, dan menentukan rating komentar. Rancangan dapat dilihat pada Gambar 12. Sejumlah komentar akan ditampilkan pada tabel, lalu coder atau admin menentukan pilihan valid/tidak, kategori, atau rating untuk setiap komentar.



Gambar 12 Rancangan Form Proses Komentar

Lampiran III
Contoh Hasil Komentar

No	Isi Komentar	Valid	Kategori	Rating
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???	Ya	Sikap terhadap merek	Positif
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	Ya	Sikap terhadap aktor	Positif
3	no one: literally no one: Indonesian people when in Europe: "try to use American accent"	Ya	Sikap terhadap aktor	Negatif
4	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?	Ya	Sikap terhadap aktor	Negatif
5	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	Ya	Sikap terhadap karakter	Positif
6	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	Ya	Sikap terhadap film	Positif
7	Suara nico setiap narasi itu juara bange!!!	Ya	Sikap terhadap aktor	Positif
8	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	Ya	Sikap terhadap film	Positif
9	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	Ya	Sikap terhadap aktor	Negatif
10	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	Ya	Sikap terhadap karakter	Positif
11	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	Ya	Sikap terhadap aktor	Positif
12	buat yang nyari laguna ada nie gw Marion Grace - When You Are Near	Ya	Sikap terhadap film	Positif
13	Aduh. Duo ini. Nic & Mar. After Janji Joni ketemuan lagi di sini. Great job, LINE! Bawa lagi figur2 yang idola banget di tahun 2000an. Definitely will subscribe dan nonton terus episode2nya.	Ya	Sikap terhadap aktor	Positif
14	Line klo bikin mini drama selalu memanjakan mata dan telinga pengen terus nonton lagi lagi dan lagi. Good job Line.	Ya	Sikap terhadap merek	Positif

No	Isi Komentar	Valid	Kategori	Rating
	Keep up good work.			
15	mar, terimakasih kehangatan yang kamu berikah, dari secangkir kopi, dan sepasang sarung tangan, njirr :3	Ya	Sikap terhadap karakter	Positif
16	ditunguin cinta...malah asik sama mar y di paris LOL	Ya	Sikap terhadap aktor	Negatif
17	jadi dia ninggalin cinta gara - gara si Mar ini :D	Ya	Sikap terhadap aktor	Negatif
18	Proyek nekad. Budget besar. Eksekusi mantab. Semoga dewi keberuntungan berpihak . Penasaran, agency yg buat iklannya siapa ya?	Ya	Sikap terhadap film	Positif
19	Nic : Mar... Terimakasih buat kehangatan yang telah kamu berikan Me : *dlm hati.. Sama-sama	Ya	Sikap terhadap karakter	Positif
20	OMG just watch this in 2019!!!! Kemana aja w!???!???!	Tidak	--	--

Lampiran IV
Karya Ilmiah

COMMENTS SCRAPING APPLICATION FOR REVIEW YOUTUBE CONTENT

Viny Christanti M., Walda, Tri Sutrisno*
Computer Science Department, Faculty of Information Technology, Universitas Tarumanagara

* viny@untar.ac.id

Abstract. YouTube is one of social media that allows people to upload, view, and share videos through website. YouTube's mission statement is to give everyone a voice and show them the world. Through YouTube, people can express their voices or show their creativity effortlessly. YouTube provides service for users to give opinions through comment section. The comments of users can be used to decide either the video has good ratings or not. Web Scraping can be used to take the comments of users from comment section. Web Scraping is a method to take some information from website by extracting data through tags in html. By inspecting code the YouTube website, the information of comments is contained in 'ytd-comment-thread-renderer'. Information of comments is divided into reviewer and content of comments. Reviewer's information is contained in id 'author-text' and the content is in 'content-text'.

1. Introduction

In the new era of globalization, business's competitions are likely increased, both in national and international. It affects the company to think about how to promote products or services in creative and innovative way. One of the ways to promote products or services is promoting some contents through social media.

YouTube is one of the social media that is popular nowadays. Lembaga riset pasar statista predicts that the users of YouTube will reach around 1.8 billion in 2021 [1]. YouTube's mission statement is to give everyone a voice and show them the world. Through YouTube, people can express their voices or show their creativity effortlessly. People can involve indirectly in YouTube by viewing the videos, giving comments, and sharing the videos to another people. By the advantages of YouTube, company decide to do promotion through YouTube. The promotions are shared by video that has a good content inside.

Company can get the feedback for users by seeing the comments of users in the comment section. The comments can be used to decide either the content of video has good ratings of not. Checking the comments one by one in the comment section will take a long time, so there is a need to scrap all of the comments and use it to do some predictions. Scraping the comments in YouTube can be done by web scraping method. Web Scraping is a method to take some information from website by extracting data through tags in HTML. Web Scraping can be done by inspecting the code of YouTube and see position of the information that will be taken from it. The reviewer and the comments from You Tube will be extracted from web.

Anand, Kedar, and Shweta [2] use web scraping method to mine information from different and unstructured websites and transform it into a comprehensible structure like spreadsheets, database, or

comma-separated values (CSV) file. The data such as item pricing, stock pricing, reports, market pricing, and product details can be used to take effective decisions in business process.

Ahmat, Leon, and Suryayusra [3] use web scraping to make an application for searching scientific article. The application is run by inputting the keywords and the machine will search the information through some portals, such as Garuda, ISJD, and Google Scholar. The machine will show the information if the keywords are matching with the data in database. If the machine did not find the keywords then it will show “the information are not found”

Susan Li [4] uses web scraping to get hotel reviews in Trip Advisor. The reviews can be used to do a sentiment analysis. Sentiment analysis can decide the voice of customer materials that comes from reviews or survey responses. Mahek, Ricky, and Nishil [5] uses web scraping to get reviews of different mobile company such as Apple, Samsung, and Oneplus. The reviews are likely consist of “not Satisfied and satisfied” reviews. The reviews are classified based on keywords, ratings, and emotions that decide the final review score of the products.

In this research we build a system that is used dynamically to scrap and clean text data from the YouTube automatically. The system is equipped with a validation function that can be used as a form to carry out the comment validation process. The result of this application is to prepare comment data for analysis in the opinion mining process.

2. Methodology

The method in this research is the waterfall method in the System Development Life Cycle (SDLC) which is one of the methods of making application programs. SDLC consists of planning, system analysis, system design, system development, system testing by researchers and maintenance. The stages used in this research are the stages in text processing from web that pay attention to the web structure in crawling and scraping.

2.1. YouTube

YouTube is a media for uploading videos and sharing information in the form of videos [6]. YouTube provides a commentary column for comments on shared videos. YouTube is often used as a place to market products and get opinions through the number of likes or comments written by readers. In the Figure 1 we can see video about Samsung note 10 and at figure 2 we can see review from people.

Figure 1. Video about product

Figure 2. Review product from reader

YouTube service provides a column for writing comments that can contain opinions. But sometimes, the word is not an opinion but only meaningless term. In the comments column, other readers can reply the comment to indicate whether they agree or not, or add other comments. The sentence in this comment can also not be related and not mutually supportive. Language inconsistency, inaccuracy in the form of the comment sentence requires validation from humans so that the data entered for analysis is the right data in the form of comments rather than others

In principle, opinions on YouTube are almost similar to opinions on blogs. Where on the blog there is also a comment column that can be used by users to write opinions about the writings listed by the author. On YouTube the comment service is limited in character length. An opinion can be

considered as one large document that contains many contents or as a small document that contain one content [7].

Unlike collections from news sources where each document is written briefly and contains topics related to the writing, a comment post is usually slightly improved and tends to be irregular. In the use of comment words for example the word "sedap" is written as "sedaaap", so that the detection of opinion words becomes more difficult. In addition the comments writing page often contains comments that can extend the writing without adding text related to the topic being discussed. Comments collection is also vulnerable to spam content in the form of spam blogs (splogs) and irrelevant comments [7].

2.2. Web HTML Structure

HTML is an abbreviation of Hypertext Markup Language is a script for compiling Web documents [16]. HTML documents are stored in regular text format and contain tags that instruct the web browser to execute the specified commands. The basic structure of an HTML document can be seen in Figure 3. The structure HTML consists of Tag, Element and Attribute.

```
<html>
<head>
<title>Di sini Judul Dokumen HTML</title>
</head>
<body>
  Disini penulisan  informasi Web
</body>
</html>
```

Figure 3. HTML Structure

On YouTube it also consists of existing structures in HTML. In figure 4, we can see comments that are visible on the web. But when the source is seen the form consists of html structure. So we must create application that only takes comments without taking other data that is not needed. Some of comments not only consist of letters but can contain images. The sentence “ya ampun senyumnya nic itu sumpah ya bikin meleleh.... ” consist of emoticon so the comment must be cleaned up before it can be processed in opinion mining.

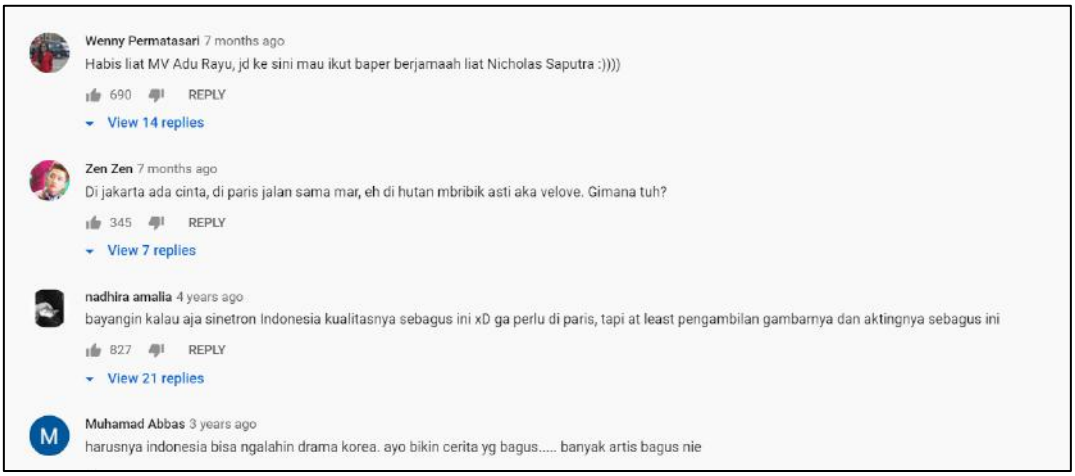


Figure 4. Example of YouTube comment

3. Result

This research makes a comment scraping program on YouTube by using the PHP and HTML programming languages. This name of this application is “Rukomindo” (Pengeruk Komentar Indonesia) equipped with a validation function. The goal of this application is scrap the comment, clean the comment, validation the comment and prepare comment data for analysis. The program

consists of 5 main modules namely scraping module, comment results, validation process, and summary of comments validation.

At figure 5, we can see interface of scraping module. At scraping module, we only copy paste the link of YouTube video that we will scrap. We can see the video and the application will automatically take out the comment. User can arrange what they want to scrap, comment only or comment with all replies.

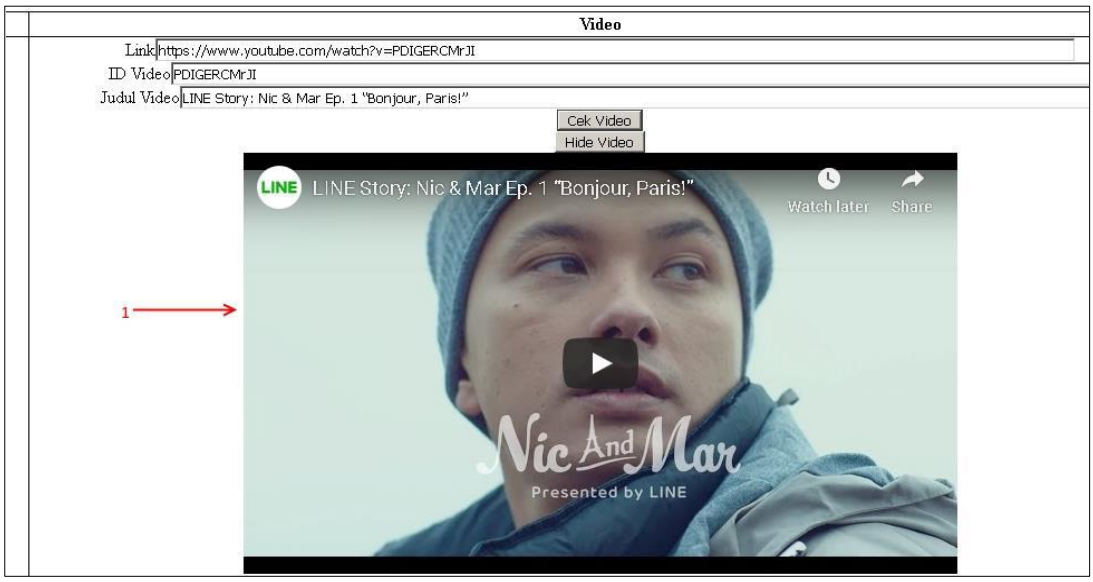


Figure 5. Interface Scraping Module

After put the YouTube link, next step is see the comment result. We can see the comment result at figure 6. There are lists comment that already represent in table format. We can validation the comment and prepare for analysis step. We only save the comment without the name ID who wrote of comment.

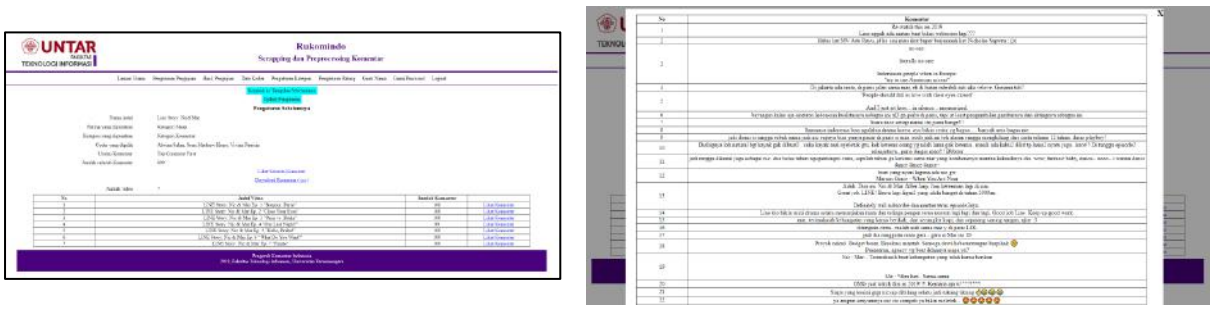


Figure 6. Interface Comment Result

At figure 7 we can see interface of validation process. We can arrange how many validators will give voting for each comment. In the end user will make decision is the comment is opinion or not. User will prepare the data separate the sentence that not consists of statement or opinion. Only opinion will be used for sentiment analysis.

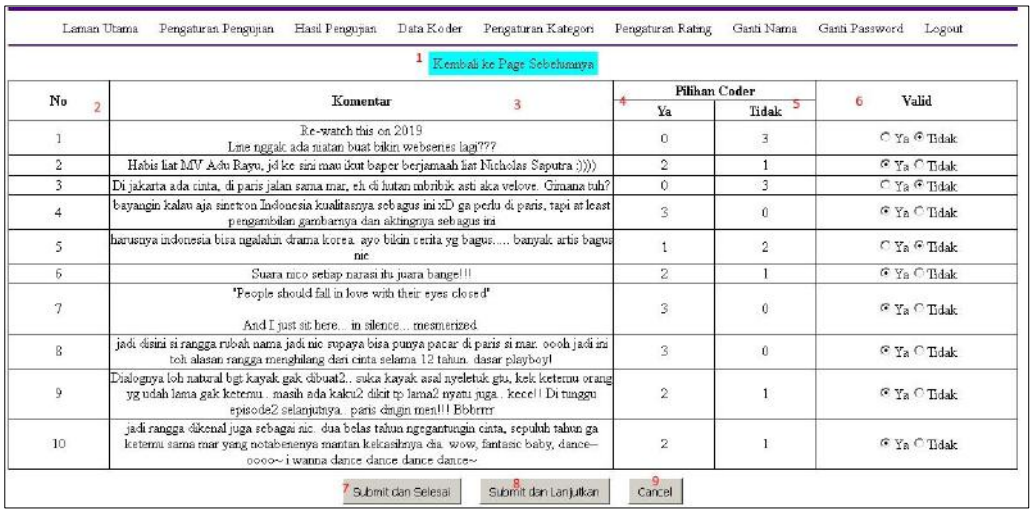


Figure 7. Interface validation process

After validation process, we can see at figure 8 the summary of validation process. We can download recapitulation from validation process. So we can prepare the data for another process like sentiment analysis or mining process.

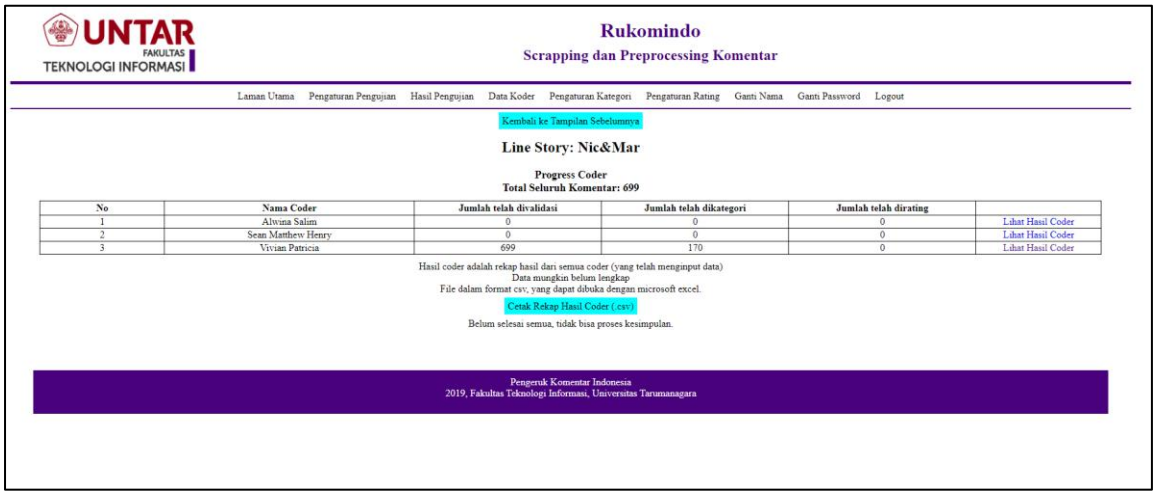


Figure 8. Interface Summary of Comment Validation

4. Conclusion

In this research a scraping application has been built to obtain comment data on YouTube. This application can be used to pull comments from a video and provide validation for these comments. So the results of comments that have been collected can be used for mining or further analysis. This application helps users in taking comments from YouTube movies just by filling in the links from the desired YouTube.

This application has been used by the management team who wants to process the film data of Nic and Mar 7 episodes and obtained 699 comments from the video. The research team from the management faculty has also tried to provide validation for the comments in the film.

For further research we will try to create scraping for another social media such as Twitter, Facebook, Instagram and etc. After create scraping, this application must develop for pre-processing step, like separate reply and real comment, separate which negative and positive comment or save the account id to evaluate hoax or non-hoax comment.

Acknowledgments

We would like to gratitude to DPPM Untar, for funding this research. Thank you to research assistant that carries out this research. And we would like to thank to Dr. Cokki from Economics Faculty Untar who used our application for their research.

References

[1] Translation from Praditya, Diaz. (2018). 3 Fakta menarik dari Riset Google tentang Perkembangan Youtube di Indonesia, https://id.techinasia.com/fakta-perkembangan-youtube-di-indonesia, Accessed on November 2019, 16th .

[2] Saurkar, Anand V; Pathare, Kedar G; Gode Shweta A. (2018). An Overview On Web Scraping Techniques and Tools. International Journal on Future Revolution in Computer Science & Communication Engineering Volume 4.

[3] Translation from Josi, Ahmat; Abdillah, Leon A.; Suryayusra. (2014). Penerapan Web Scraping pada Mesin Pencari Ilmiah. Bina Darma University.

[4] Susan Li. (2018). Web Scraping TripAdvisor, Text Mining and Sentiment Analysis for Hotel Reviews. https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333. Accessed on November 2019, 16th.

[5] Merchant, Mahek; Shah, Nishil; P.Boominathan. (2016). Sentiment Analysis of Web Scraped Product Reviews using Hadoop. IJRASET Volume 4 Issue VIII

[6] Hopkins, Jim. "Surprise! There's a third YouTube co-founder." *USA Today* 11, no. 10 (2006).

[7] Elsas, Jonathan L., Jaime Arguello, Jamie Callan, and Jaime G. Carbonell. "Retrieval and feedback models for blog feed search." In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 347-354. ACM, 2008.

Lampiran V
Manual Penggunaan Aplikasi Scraping

PANDUAN MANUAL

Aplikasi Filtering dan Preprocessing Text pada Data Komentar Youtube

Disusun oleh:
Ketua Tim
Viny Christanti M. 0317088203/10805002

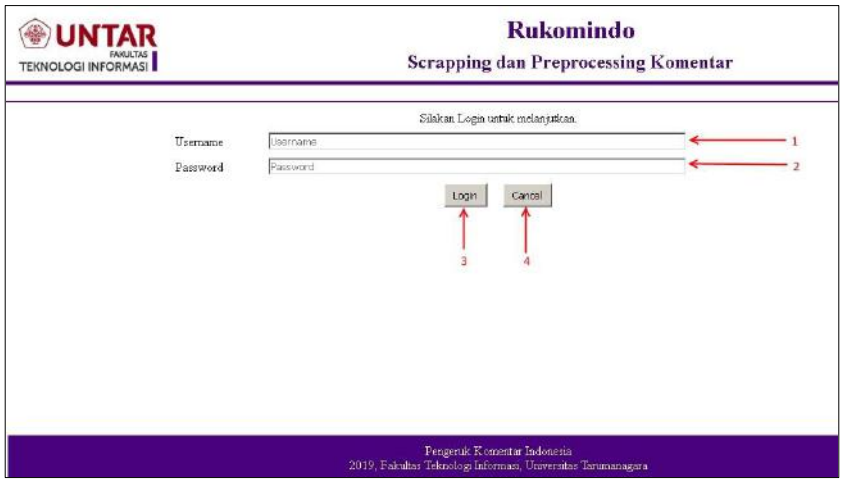
Anggota:
Tri Sutrisno, S.Si., M.Sc. 0320018801/10816004
Freddy Kurniawan

Fakultas Teknologi Informasi
Universitas Tarumanagara
Jakarta
Januari 2020

A. Home

Tampilan awal rukomindo adalah form login seperti pada Gambar 1. Tidak ada *page* yang diperuntukkan bagi *guest* (user tanpa login). Ada 2(dua) tipe user pada rukomindo, yaitu admin dan coder. Berikut adalah keterangan untuk Gambar 1:

1. *Field* username
2. *Field* password
3. *Button* untuk *submit* form
4. *Button* untuk *reset* form



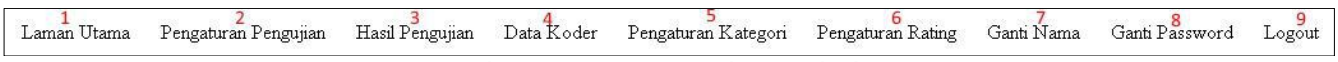
Gambar 1. Home Rukomindo

B. Admin

Menu Admin

Menu untuk admin awalnya tidak ditampilkan semua. Beberapa menu baru ditampilkan setelah selesainya proses/tahap tertentu. Menu lengkap dapat diilihat pada Gambar 2. Berikut adalah keterangan untuk Gambar 2:

1. Menu home admin
2. Menu daftar pengaturan pengujian. Menu ini berkaitan dengan pengaturan untuk pengujian yang akan dilakukan koder, diantaranya video youtube dan komentar dari setiap video.
3. Menu daftar hasil pengujian. Menu ini berkaitan dengan hasil yang telah dilakukan oleh coder dan admin.
4. Menu data coder. Menu ini untuk melihat data coder yang pernah dibuat.
5. Menu pengaturan kategori. Menu ini untuk melihat data pilihan kategori.
6. Menu pengaturan rating. Menu ini untuk melihat data pilihan rating.
7. Menu ganti nama. Menu ini untuk mengganti nama akun.
8. Menu ganti password. Menu ini untuk mengganti password akun.
9. Klik untuk logout



Gambar 2 Menu Lengkap Admin

Home Admin

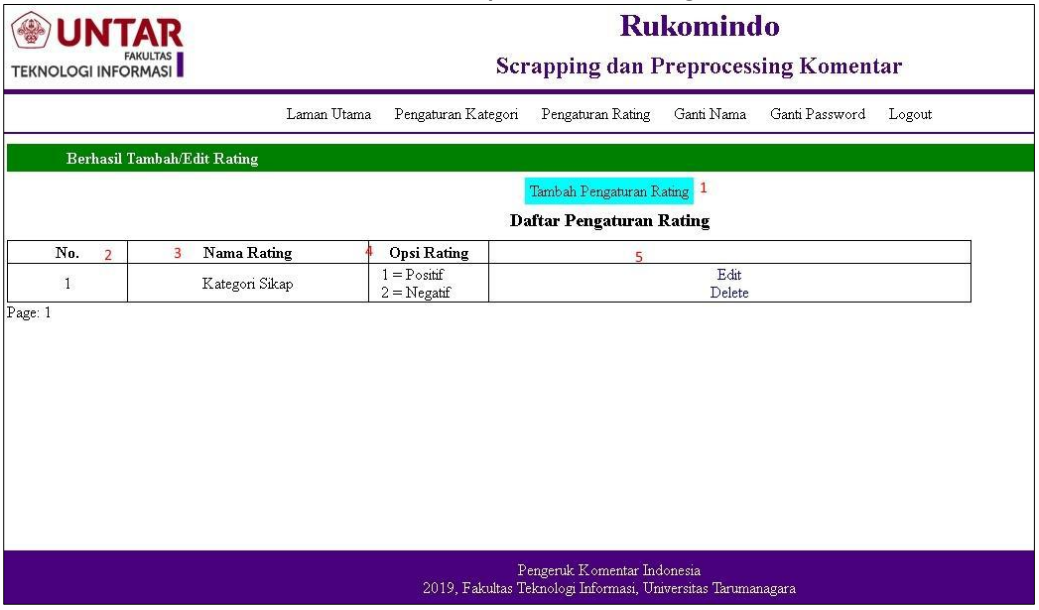
Ini adalah tampilan awal untuk admin. Pada tampilan ini terdapat urutan proses yang harus dilakukan oleh admin, seperti terlihat pada Gambar 3.

Gambar 3 Home Admin

Pengaturan Rating

Page ini menampilkan daftar pilihan rating, yaitu pilihan yang diberikan pada coder/admin saat memberikan rating pada setiap komentar. Contoh tampilan dapat dilihat pada Gambar 4. Berikut adalah keterangan dari Gambar 4:

1. Link menuju form rating.
2. Kolom nomor
3. Kolom nama rating, nama ini digunakan pada form pengaturan pengujian.
4. Kolom opsi rating, kolom ini berisi opsi/pilihan untuk rating
5. Kolom berisi link untuk ubah atau hapus data rating.

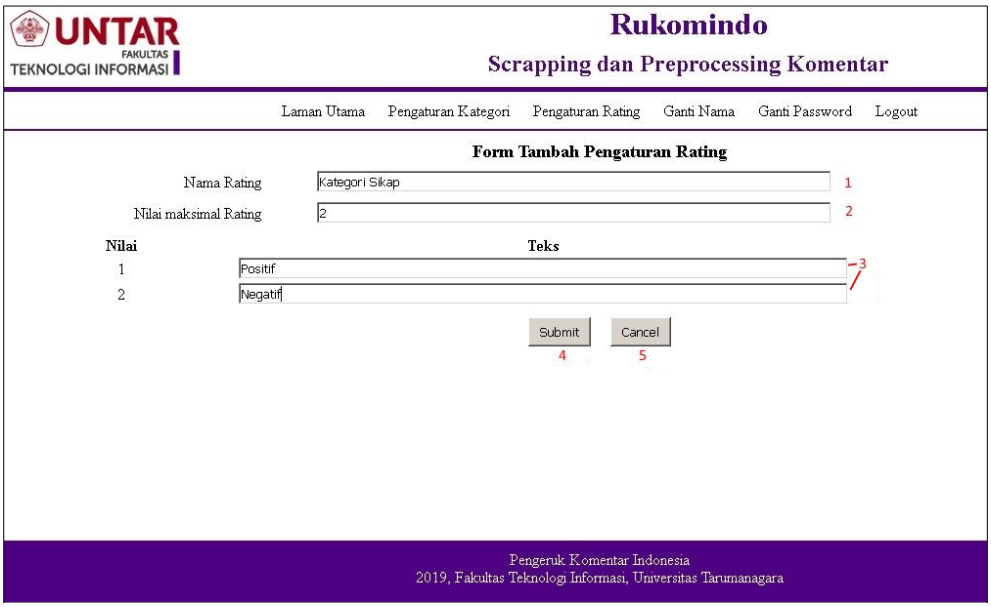


Gambar 4. Daftar Rating

Form Rating

Page ini berisi form untuk input data rating. Admin harus menginput nama rating dan menentukan jumlah opsi. Setelah jumlah opsi diubah, admin dapat menginputkan teks untuk setiap pilihan rating. Contoh tampilan dapat dilihat pada Gambar 5. Berikut adalah keterangan dari Gambar 5:

1. *Field* nama rating, nama ini digunakan pada form pengaturan pengujian.
2. *Field* jumlah pilihan rating
3. *Field* nama salah satu pilihan rating
4. *Button* untuk *submit* form
5. *Button* untuk *reset* form

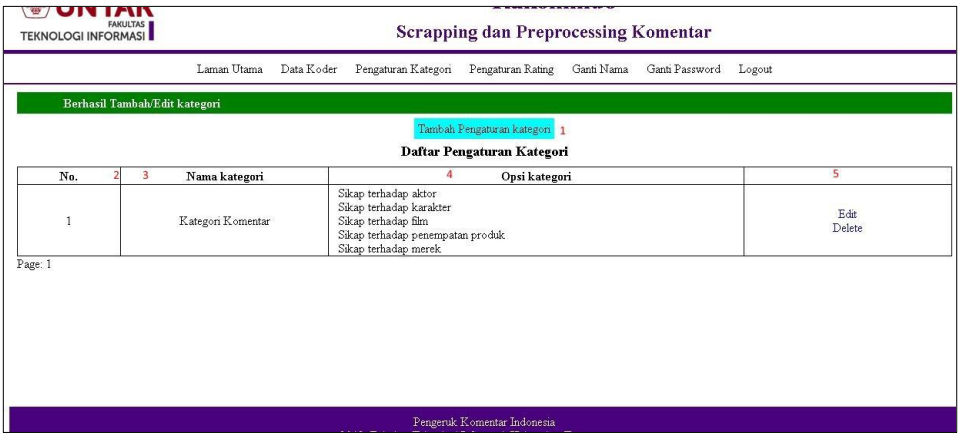


Gambar 5 Form Rating

Pengaturan Kategori

Page ini menampilkan daftar pilihan kategori, yaitu pilihan yang diberikan pada coder/admin saat menentukan kategori pada setiap komentar. Contoh tampilan dapat dilihat pada Gambar 6. Berikut adalah keterangan dari Gambar 6:

1. Link menuju form kategori.
2. Kolom nomor
3. Kolom nama kategori, nama ini digunakan pada form pengaturan pengujian.
4. Kolom opsi kategori, kolom ini berisi opsi/pilihan untuk kategori
5. Kolom berisi link untuk ubah atau hapus data kategori.



Gambar 6 Daftar Kategori

Form Kategori

Page ini berisi form untuk input data kategori. Admin harus menginput nama kategori dan menentukan jumlah pilihan kategori. Setelah jumlah pilihan kategori diubah, admin dapat menginputkan teks untuk setiap pilihan kategori. Contoh tampilan dapat dilihat pada Gambar 7. Berikut adalah keterangan dari Gambar 7:

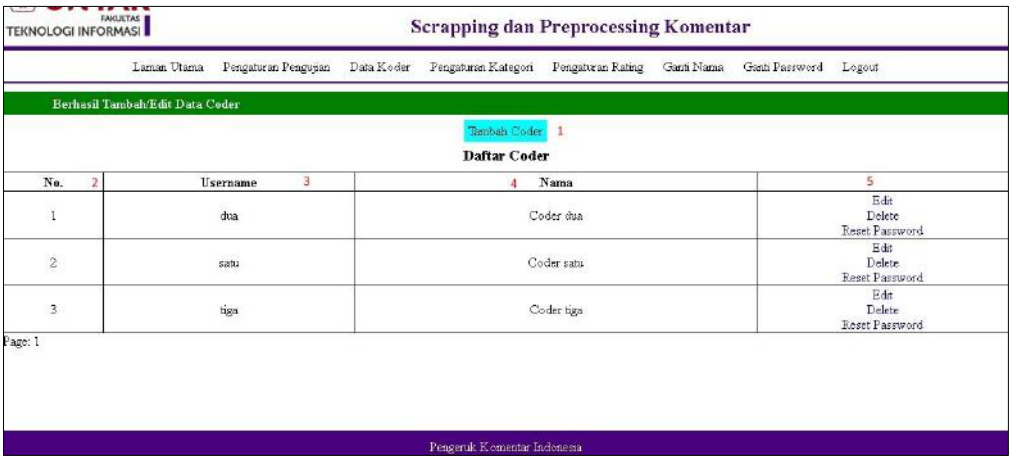
1. *Field* nama kategori, nama ini digunakan pada form pengaturan pengujian.
2. *Field* jumlah pilihan kategori
3. *Field* nama salah satu pilihan rating
4. *Button* untuk *submit* form
5. *Button* untuk *reset* form

Gambar 7 Form Rating

Data Coder

Page ini menampilkan daftar coder. Contoh tampilan dapat dilihat pada Gambar 8. Berikut adalah keterangan dari Gambar 8:

1. Link menuju form coder.
2. Kolom nomor
3. Kolom username, yang akan digunakan untuk login
4. Kolom nama coder.
5. Kolom berisi link untuk ubah, hapus, dan reset password coder.

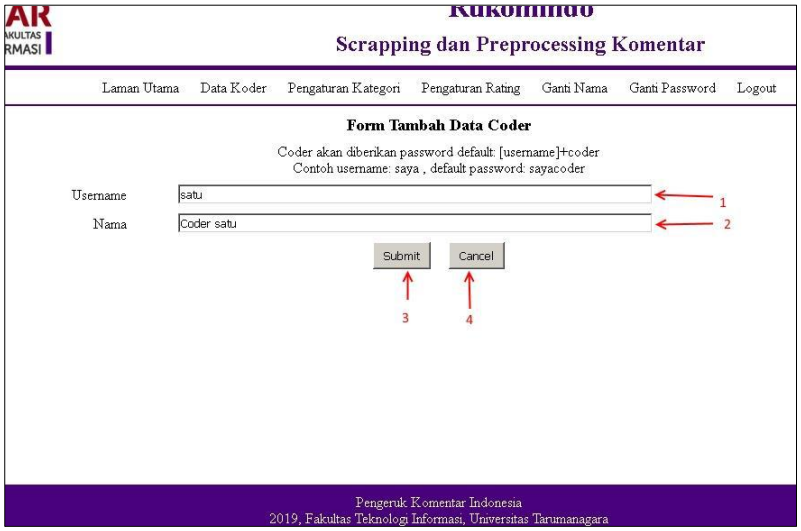


Gambar 8 Data Coder

Form Coder

Page ini berisi form untuk input data coder. Admin harus menginput username dan nama coder. Setiap coder akan diberi password *default*. Contoh tampilan dapat dilihat pada Gambar 9. Berikut adalah keterangan dari Gambar 9:

1. *Field* username.
2. *Field* nama coder
3. *Button* untuk *submit* form
4. *Button* untuk *reset* form

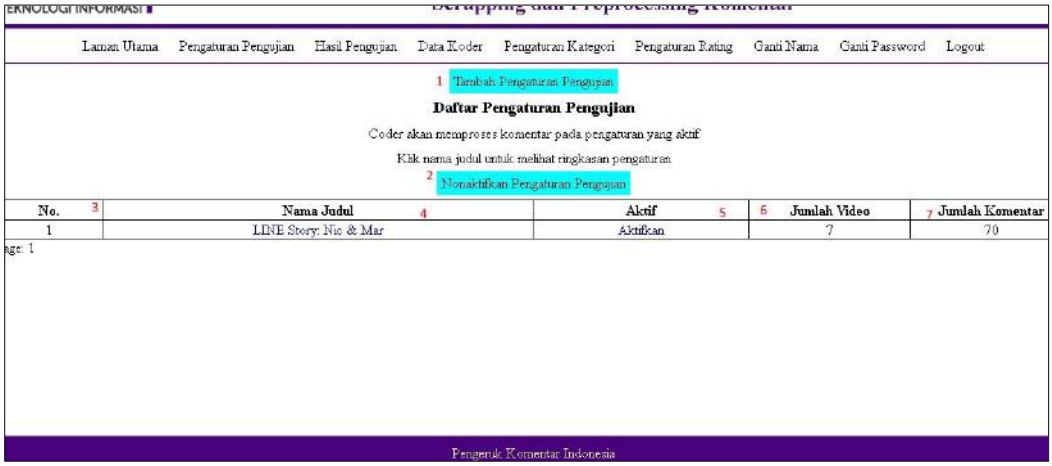


Gambar 9 Form Coder

Daftar Pengaturan Pengujian

Page ini menampilkan daftar pengaturan pengujian. Pengaturan pengujian yang digunakan adalah pengaturan yang aktif, admin dapat mengubah pengaturan mana yang aktif. Contoh tampilan dapat dilihat pada Gambar 10. Berikut adalah keterangan dari Gambar 10:

1. Link menuju form pengaturan pengujian.
2. Link untuk menonaktifkan pengaturan aktif.
3. Kolom nomor
4. Kolom nama judul pengaturan, klik teks pada kolom ini untuk melihat rincian dari pengaturan tersebut.
5. Kolom penanda pengaturan aktif, klik link/tulisan Aktifkan untuk mengaktifkan pengaturan tersebut
6. Kolom jumlah video dalam pengaturan tersebut.
7. Kolom jumlah komentar dalam pengaturan tersebut.

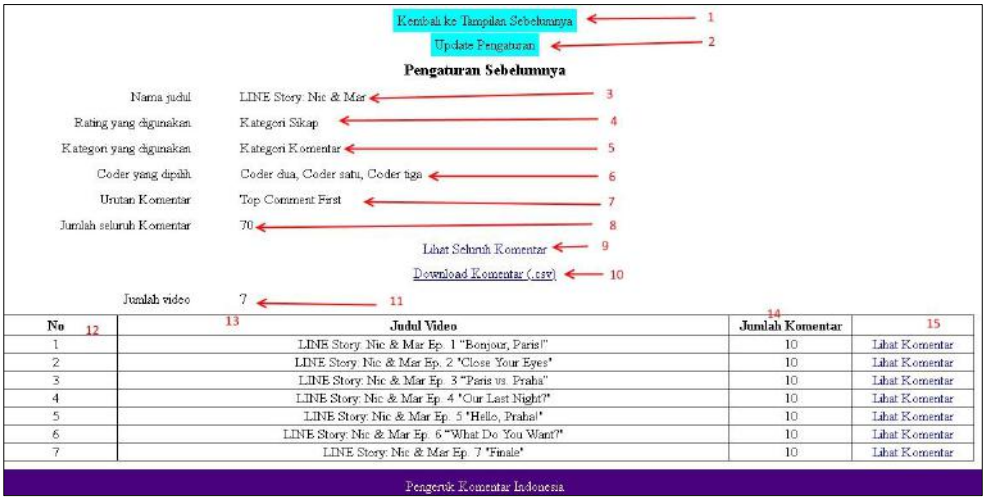


Gambar 10 Daftar Pengaturan Pengujian

Rincian Pengaturan Pengujian

Page ini menampilkan informasi mengenai pengaturan pengujian tersebut, termasuk juga daftar komentar yang akan diproses oleh coder. Contoh tampilan dapat dilihat pada Gambar 11. Berikut adalah keterangan dari Gambar 11:

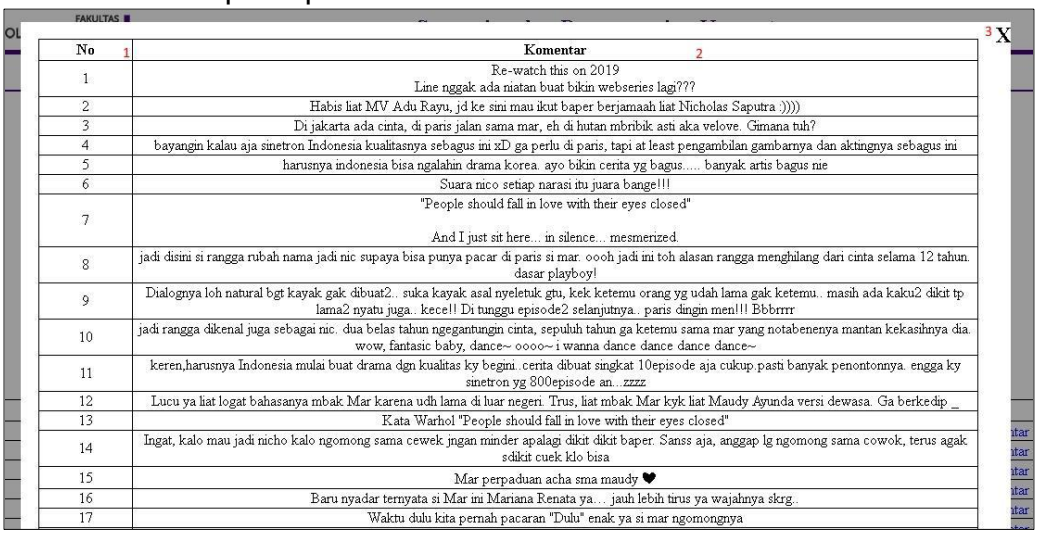
1. Link menuju daftar pengaturan pengujian
2. Link untuk ubah pengaturan pengujian. Pastikan pengaturan sudah benar saat coder telah melakukan pengujian, karena data terkait pengujian tersebut akan dihapus setiap kali mengubah pengaturan.
3. Nama judul untuk pengujian.
4. Nama rating yang digunakan untuk pengujian
5. Nama kategori yang digunakan untuk pengujian
6. Daftar coder yang akan melakukan pengujian
7. Urutan komentar yang diambil oleh API youtube.
8. Jumlah seluruh komentar pada pengujian tersebut.
9. Klik untuk melihat seluruh komentar. Contoh tampilan dapat dilihat pada Gambar 12.
10. Download file csv berisi seluruh komentar pada pengujian.
11. Jumlah video yang komentarnya diambil untuk pengujian
12. Kolom nomor
13. Kolom judul video youtube.
14. Kolom jumlah komentar yang diambil dari video tersebut
15. Kolom berisi link untuk melihat komentar per video.



Gambar 11 Rincian Pengaturan Pengujian

Berikut adalah keterangan dari Gambar 12:

1. Kolom nomor
2. Kolom teks komentar
3. Klik untuk menutup tampilan



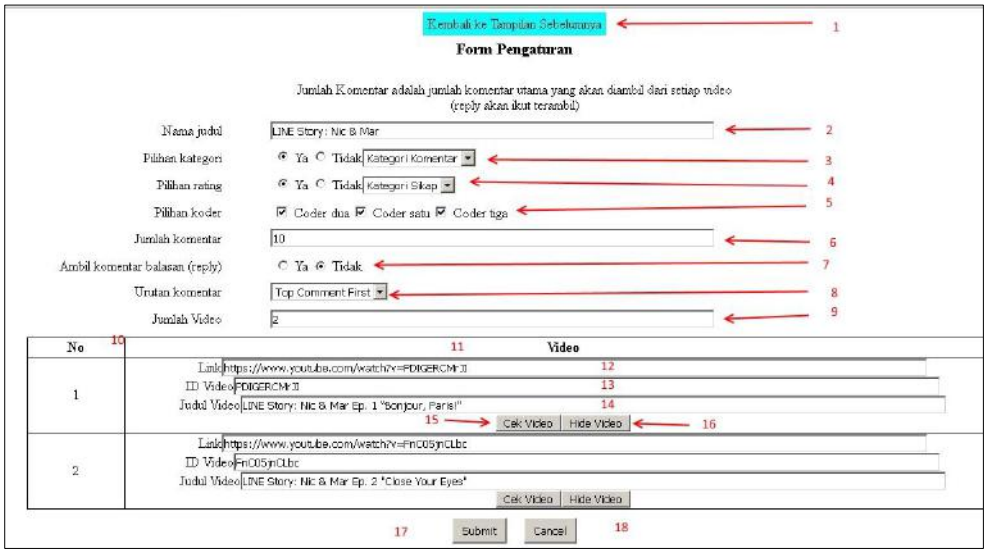
No 1	Komentar 2
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie
6	Suara nico setiap narasi itu juara bange!!!
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!
9	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~
11	keren,harusnya Indonesia mulai buat drama dgn kualitas ky begini..cerita dibuat singkat 10episode aja cukup.pasti banyak penontonnya. engga ky sinetron yg 800episode an...zzzz
12	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _
13	Kata Warhol "People should fall in love with their eyes closed"
14	Ingat, kalo mau jadi nicho kalo ngomong sama cewek jngan minder apalagi dikit dikit baper. Sanss aja, anggap lg ngomong sama cowok, terus agak sdikit cuek klo bisa
15	Mar perpaduan acha sma maudy ❤
16	Baru nyadar ternyata si Mar ini Mariana Renata ya... jauh lebih tirus ya wajahnya skrg..
17	Waktu dulu kita pernah pacaran "Dulu" enak ya si mar ngomongnya

Gambar 12 Daftar Komentar

Form Pengaturan Pengujian

Page ini berisi form untuk mengatur pengujian yang dilakukan. Admin dapat menentukan apakah coder akan menentukan kategori serta rating. Pengaturan lain seperti jumlah komentar, ambil komentar balasan, dan urutan komentar digunakan pada saat mengambil komentar menggunakan API youtube. Admin harus mengisikan link berupa url video youtube, untuk mendapatkan ID dan judul video. Contoh tampilan dapat dilihat pada Gambar 13. Berikut adalah keterangan dari Gambar 13:

1. Link menuju rincian pengaturan pengujian
2. *Field* nama judul untuk pengaturan pengujian.
3. *Field* pilihan kategori. Jika ya, coder diharuskan menentukan kategori komentar
4. *Field* pilihan rating. Jika ya, coder diharuskan menentukan rating komentar
5. *Field* pilihan coder. Admin menentukan coder mana yang akan melakukan pengujian
6. *Field* untuk menentukan jumlah komentar utama yang diambil dari setiap video.
7. *Field* ambil komentar balasan. Jika ya, komentar balasan akan ikut diambil dan jumlah komentar yang diambil mungkin melebihi isi *field* 6. Jika tidak, hanya komentar utama yang diambil dari setiap video
8. *Field* urutan komentar. API youtube memberikan 2 (dua) pilihan, yaitu *top comment* (inilah pilihan *default* pada youtube) dan *newest* (terbaru).
9. *Field* jumlah video yang akan diambil komentarnya.
10. Kolom nomor
11. Kolom berisi *field* untuk setiap video
12. *Field* yang akan diisi url video youtube. Setiap kali user selesai menginput *field* ini, lalu memindahkan fokus ke *field* lain, akan muncul pesan seperti pada Gambar 14.
13. *Field* yang akan terisi id video youtube
14. *Field* yang akan terisi judul video youtube
15. *Button* untuk menampilkan video sesuai isi *field* 12. Tampilan seperti pada Gambar 15. Video yang dimaksud ditunjuk oleh nomor 1 pada Gambar 15.
16. *Button* untuk menghilangkan video tersebut dari form.
17. *Button* untuk *submit* form
18. *Button* untuk *reset* form



Gambar 13 Form Pengaturan Pengujian

Gambar 14 Pesan Memeriksa Url Youtube

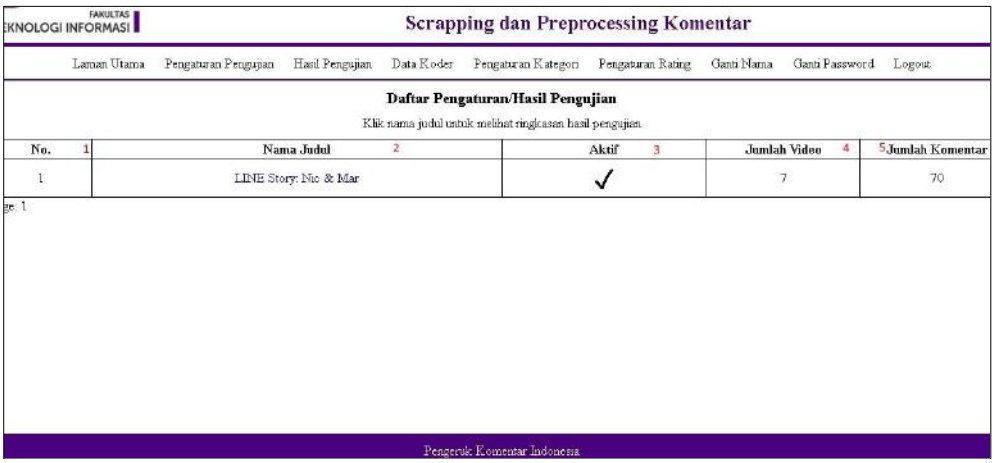
Gambar 15 Contoh Menampilkan Video Youtube

Daftar Hasil Pengujian

Page ini menampilkan daftar pengujian yang ada. Pengujian aktif akan diberi penanda, hanya pengujian tersebut yang dapat diinput oleh admin. Admin hanya dapat melihat hasil dari pengujian lainnya Contoh tampilan dapat dilihat pada Gambar 16. Berikut adalah keterangan dari Gambar 16:

1. Kolom nomor
2. Kolom nama judul pengaturan pengujian. Klik nama judul untuk melihat rekap hasil pengujian dari para coder.
3. Kolom penanda pengujian aktif

4. Kolom jumlah video dalam pengujian tersebut
5. Kolom jumlah komentar dalam pengujian tersebut



Gambar 16 Daftar Hasil Pengujian

Rekap Pengujian Coder

Page ini menampilkan data berupa rekapan dari progress pengujian yang telah dilakukan semua coder. Apabila semua coder telah menyelesaikan pengujian, admin dapat menentukan hasil akhir/kesimpulan untuk setiap komentar. Contoh tampilan dapat dilihat pada Gambar 17. Berikut adalah keterangan dari Gambar 17:

1. Link untuk kembali ke daftar hasil pengujian
2. Kolom nomor
3. Kolom nama coder.
4. Jumlah komentar yang telah divalidasi setiap coder.
5. Jumlah komentar yang telah dikategori oleh setiap coder.
6. Jumlah komentar yang telah dirating oleh setiap coder.
7. Kolom berisi link untuk melihat hasil pengujian setiap coder.
8. Link untuk mendapatkan file csv berisi rekapan dari hasil pengujian semua coder.
9. Link untuk menentukan validasi untuk setiap komentar, melalui form validasi komentar. Link ini hanya ditampilkan ketika semua coder telah selesai melakukan pengujian.

Gambar 17 Rekap Pengujian Coder

Hasil Pengujian Coder

Page ini menampilkan hasil pengujian dari coder tersebut. Informasi yang ditampilkan adalah komentar valid/tidak, kategori, dan rating yang ditentukan oleh coder. Komentar yang tidak valid, tidak memiliki kategori dan rating. Contoh tampilan dapat dilihat pada Gambar 18. Berikut adalah keterangan dari Gambar 18:

1. Link untuk kembali ke rekap pengujian coder
2. Kolom nomor
3. Kolom teks komentar
4. Kolom komentar valid atau tidak
5. Kolom kategori untuk komentar, jika komentar tersebut valid.
6. Kolom rating untuk komentar, jika komentar tersebut valid

1 Kembali ke tampilan sebelumnya

Hasil Pengerjaan Coder

No 2	Isi Komentar 3	4 Valid	5 Kategori	6 Rating
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???	Tidak	--	--
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	Ya	Sikap terhadap aktor	Negatif
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?	Tidak	--	--
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	Ya	Sikap terhadap penempatan produk	Positif
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	Tidak	--	--
6	Suara nico setiap narasi itu juara bange!!!	Tidak	--	--
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	Ya	Sikap terhadap film	Positif
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	Ya	Sikap terhadap karakter	Positif
9	Dialognya loh natural bgt kayak gak dibuat2... suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga... kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	Ya	Sikap terhadap film	Positif
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	Ya	Sikap terhadap karakter	Negatif
11	keren,harusnya Indonesia mulai buat drama dgn kualitas ky begini..cerita dibuat singkat 10episode aja cukup.pasti banyak penontonnya. engga ky sinetron yg 800episode an...zzzz	Tidak	--	--
12	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _	Ya	Sikap terhadap karakter	Positif

Gambar 18 Hasil Pengujian Coder

Form Validasi Komentar

Page ini berisi form yang digunakan admin untuk menentukan apakah komentar tersebut valid atau tidak, setiap *page* hanya menampilkan 10 komentar. Admin dapat melihat jumlah coder yang memilih komentar valid dan tidak valid. Jika salah satu pilihan jumlahnya lebih besar, form akan otomatis memilih pilihan dengan jumlah terbanyak. Contoh tampilan dapat dilihat pada Gambar 19. Berikut adalah keterangan dari Gambar 19:

1. Link untuk kembali ke rekap pengujian coder
2. Kolom nomor
3. Kolom teks komentar
4. Kolom berisi jumlah coder yang menentukan komentar tersebut valid.
5. Kolom berisi jumlah coder yang menentukan komentar tersebut tidak valid.
6. Kolom berisi *field* untuk menentukan komentar tersebut valid atau tidak.
7. *Button* untuk *submit* dan kembali ke rekap pengujian coder
8. *Button* untuk *submit* dan melanjutkan validasi ke komentar berikutnya.
9. *Button* untuk *reset* form.

Laman Utama Pengaturan Pengujian Hasil Pengujian Data Koder Pengaturan Kategori Pengaturan Rating Ganti Nama Ganti Password Logout

1 Kembali ke Page Sebelumnya

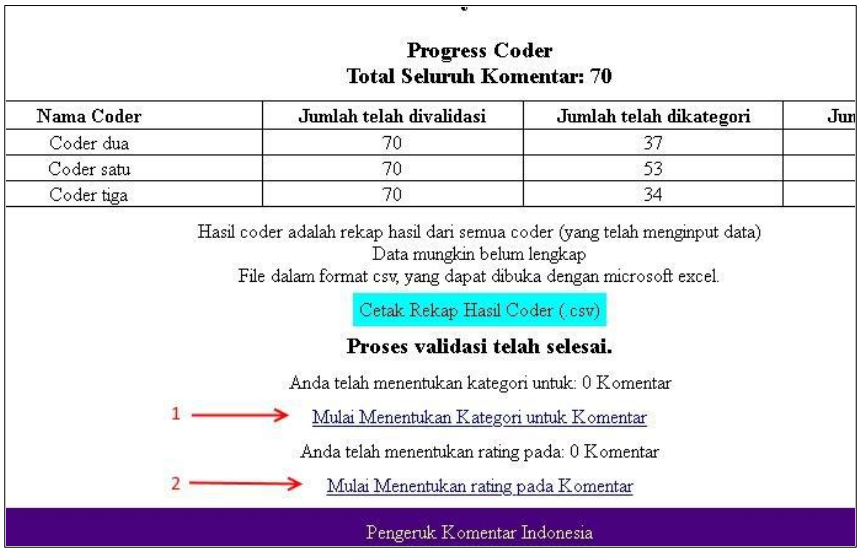
No 2	Komentar 3	4 Pilihan Coder Ya	5 Pilihan Coder Tidak	6 Valid
1	Re-watch this on 2019 Line nggak ada niatan buat bikin webseries lagi???	0	3	○ Ya ◉ Tidak
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	2	1	◉ Ya ○ Tidak
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbribik asti aka velove. Gimana tuh?	0	3	○ Ya ◉ Tidak
4	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	3	0	◉ Ya ○ Tidak
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	1	2	○ Ya ◉ Tidak
6	Suara nico setiap narasi itu juara bange!!!	2	1	◉ Ya ○ Tidak
7	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	3	0	◉ Ya ○ Tidak
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	3	0	◉ Ya ○ Tidak
9	Dialognya loh natural bgt kayak gak dibuat2... suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga... kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	2	1	◉ Ya ○ Tidak
10	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	2	1	◉ Ya ○ Tidak

7 Submit dan Selesai 8 Submit dan Lanjutkan 9 Cancel

Gambar 19 Form Validasi Komentar Admin

Setelah admin selesai memvalidasi semua komentar, admin akan diarahkan kembali ke tampilan rekap pengujian coder. Contoh tampilan dapat dilihat pada Gambar 20. Link

untuk melakukan validasi sudah tidak ada, diganti dengan link untuk menentukan kategori komentar (link nomor 1 pada Gambar 20) dan rating komentar (link nomor 2 pada Gambar 20).



Progress Coder
Total Seluruh Komentar: 70

Nama Coder	Jumlah telah divalidasi	Jumlah telah dikategori	Jum
Coder dua	70	37	
Coder satu	70	53	
Coder tiga	70	34	

Hasil coder adalah rekap hasil dari semua coder (yang telah menginput data)
Data mungkin belum lengkap
File dalam format csv, yang dapat dibuka dengan microsoft excel.

Cetak Rekap Hasil Coder (.csv)

Proses validasi telah selesai.

Anda telah menentukan kategori untuk: 0 Komentar

1 → Mulai Menentukan Kategori untuk Komentar

Anda telah menentukan rating pada: 0 Komentar

2 → Mulai Menentukan rating pada Komentar

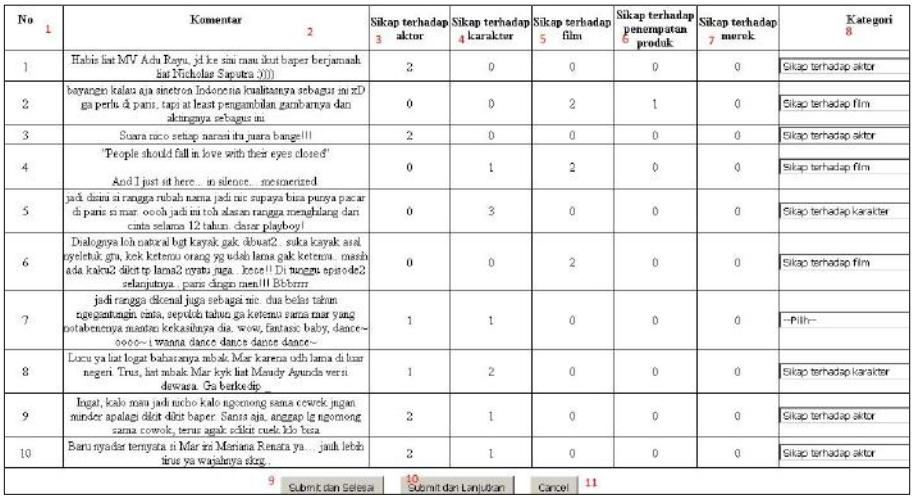
Pengeruk Komentar Indonesia

Gambar 20 Tampilan Rekap Pengujian Coder Setelah Admin Selesai Validasi Komentar

Form Kategori Komentar

Page ini berisi form yang digunakan admin untuk menentukan kategori untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Admin dapat melihat jumlah coder yang memilih setiap kategori. Jika salah satu pilihan jumlahnya lebih besar, form akan otomatis memilih kategori dengan jumlah terbanyak. Contoh tampilan dapat dilihat pada Gambar 21. Berikut adalah keterangan dari Gambar 21:

1. Kolom nomor
2. Kolom teks komentar
3. Kolom berisi jumlah coder yang memilih kategori pertama
4. Kolom berisi jumlah coder yang memilih kategori kedua
5. Kolom berisi jumlah coder yang memilih kategori ketiga
6. Kolom berisi jumlah coder yang memilih kategori keempat
7. Kolom berisi jumlah coder yang memilih kategori kelima
8. Kolom berisi *field* untuk menentukan kategori komentar
9. *Button* untuk *submit* dan kembali ke rekap pengujian coder
10. *Button* untuk *submit* dan melanjutkan proses ke komentar berikutnya.
11. *Button* untuk *reset* form.



No 1	Komentar 2	Sikap terhadap aktor 3	Sikap terhadap karakter 4	Sikap terhadap film 5	Sikap terhadap penempatan produk 6	Sikap terhadap merek 7	Kategori 8
1	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	2	0	0	0	0	Sikap terhadap aktor
2	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	0	0	2	1	0	Sikap terhadap film
3	Suara nico sehap narasi itu juara bange!!!	2	0	0	0	0	Sikap terhadap aktor
4	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	0	1	2	0	0	Sikap terhadap film
5	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghalang dari cinta selama 12 tahun. dasar playboy!	0	3	0	0	0	Sikap terhadap karakter
6	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. keee!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	0	0	2	0	0	Sikap terhadap film
7	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepotoh tahun ga ketemu sama mar yang potabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	1	1	0	0	0	--Pilih--
8	Lucu ya liat logat bahasanya nbak Mar karena udh lama di luar negeri. Trus, liat nbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip...	1	2	0	0	0	Sikap terhadap karakter
9	Ingat, kalo mau jadi nicho kalo ngomong sama cewek jngan minder apalagi dikit dikit baper. Santai aja, anggap lg ngomong sama cowok, terus agak edikit cuek klo bisa.	2	1	0	0	0	Sikap terhadap aktor
10	Baru nyadar ternyata si Mar ini Mariana Renata ya... jauh lebih tirus ya wajahnya skrg..	2	1	0	0	0	Sikap terhadap aktor

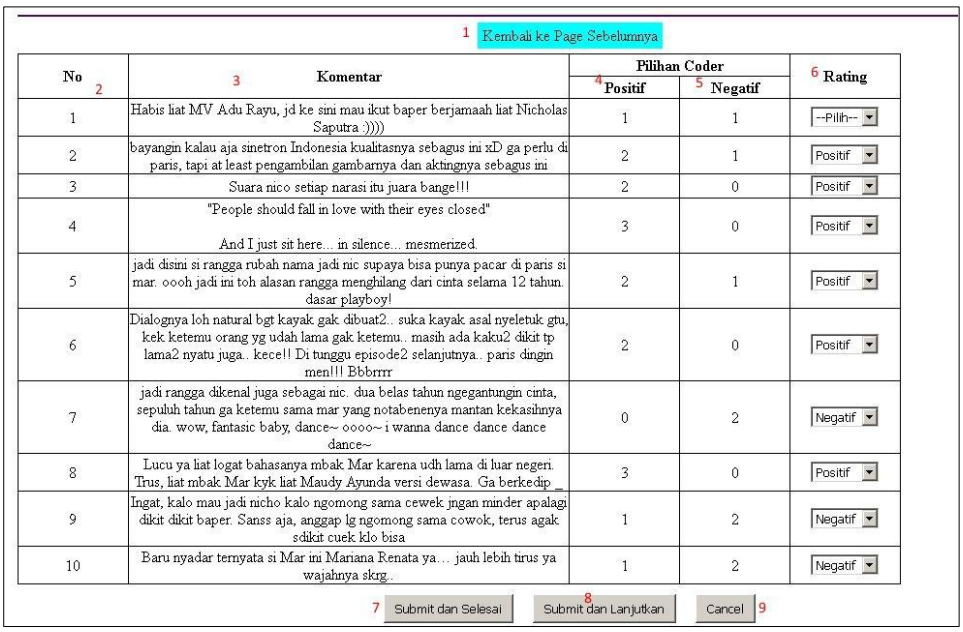
9 Submit dan Selesai | 10 Submit dan Lanjutkan | Cancel 11

Gambar 21 Form Kategori Komentar

Form Rating Komentar

Page ini berisi form yang digunakan admin untuk menentukan rating untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Admin dapat melihat jumlah coder yang memilih setiap rating. Jika salah satu pilihan jumlahnya lebih besar, form akan otomatis memilih rating dengan jumlah terbanyak. Contoh tampilan dapat dilihat pada Gambar 22. Berikut adalah keterangan dari Gambar 22:

1. Link untuk kembali ke rekap pengujian coder
2. Kolom nomor
3. Kolom teks komentar
4. Kolom berisi jumlah coder yang memilih rating pertama
5. Kolom berisi jumlah coder yang memilih rating kedua
6. Kolom berisi *field* untuk menentukan rating komentar
7. *Button* untuk *submit* dan kembali ke rekap pengujian coder
8. *Button* untuk *submit* dan melanjutkan proses ke komentar berikutnya.
9. *Button* untuk *reset* form.



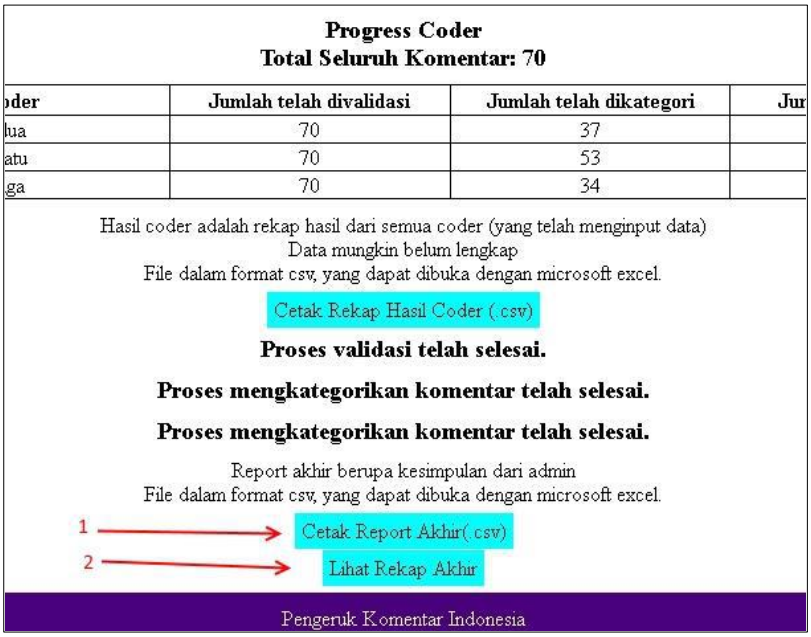
1 Kembali ke Page Sebelumnya

No 2	3 Komentar	Pilihan Coder		6 Rating
		4 Positif	5 Negatif	
1	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	1	1	--Pilih--
2	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	2	1	Positif
3	Suara nico setiap narasi itu juara bange!!!	2	0	Positif
4	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	3	0	Positif
5	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	2	1	Positif
6	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	2	0	Positif
7	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	0	2	Negatif
8	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _	3	0	Positif
9	Ingat, kalo mau jadi nicho kalo ngomong sama cewek jngan minder apalagi dikit dikit baper. Sanss aja, anggap lg ngomong sama cowok, terus agak sdikit cuek klo bisa	1	2	Negatif
10	Baru nyadar ternyata si Mar ini Mariana Renata ya... jauh lebih tirus ya wajahnya skrg..	1	2	Negatif

7 Submit dan Selesai 8 Submit dan Lanjutkan Cancel 9

Gambar 22 Form Rating Komentar

Setelah admin selesai menentukan kategori atau rating untuk semua komentar yang valid, admin akan diarahkan kembali ke tampilan rekap pengujian coder. Jika keduanya telah selesai, tampilan rekap pengujian akan seperti pada Gambar 23. Link untuk menentukan kategori dan rating sudah tidak ada, diganti dengan link untuk mendapatkan rekap akhir dalam file csv (link nomor 1 pada Gambar 23) dan link untuk melihat rekap akhir (link nomor 2 pada Gambar 23).



Progress Coder
Total Seluruh Komentar: 70

der	Jumlah telah divalidasi	Jumlah telah dikategori	Jur
lua	70	37	
atu	70	53	
ga	70	34	

Hasil coder adalah rekap hasil dari semua coder (yang telah menginput data)
Data mungkin belum lengkap
File dalam format csv, yang dapat dibuka dengan microsoft excel.

Cetak Rekap Hasil Coder (.csv)

Proses validasi telah selesai.

Proses mengkategorikan komentar telah selesai.

Proses mengkategorikan komentar telah selesai.

Report akhir berupa kesimpulan dari admin
File dalam format csv, yang dapat dibuka dengan microsoft excel.

1 Cetak Report Akhir(.csv)
2 Lihat Rekap Akhir

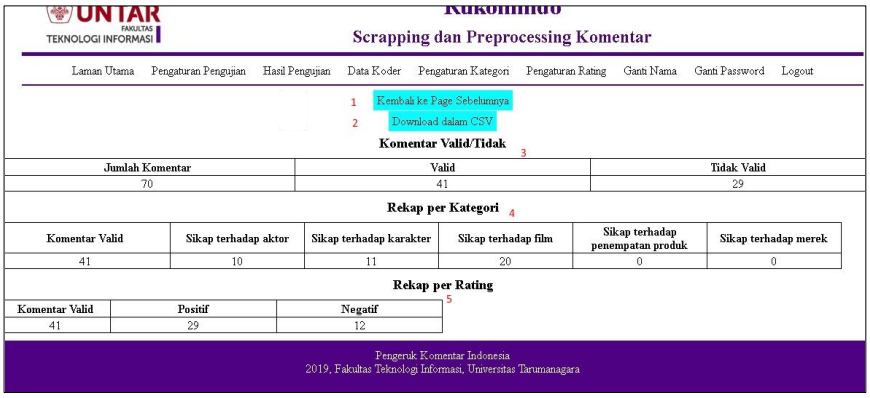
Pengeruk Komentar Indonesia

Gambar 23 Tampilan Rekap Pengujian Coder Setelah Admin Selesai Menentukan Kategori dan Rating Komentar

Rekap Hasil Akhir

Page ini berisi rekap dari input admin. Rekap ditampilkan dalam 3 (tiga) tabel, yaitu tabel komentar valid/tidak, tabel rekap per kategori, dan tabel rekap per rating Contoh tampilan dapat dilihat pada Gambar 24. Berikut adalah keterangan dari Gambar 24:

1. Link untuk kembali ke rekap pengujian coder
2. Link untuk dapat rekap dalam file csv
3. Tabel berisi jumlah komentar valid dan tidak valid
4. Tabel berisi jumlah komentar yang termasuk pada setiap kategori
5. Tabel berisi jumlah komentar yang termasuk pada setiap ratig



UNTAR FAKULTAS TEKNOLOGI INFORMASI

Scrapping dan Preprocessing Komentar

Laman Utama | Pengaturan Pengujian | Hasil Pengujian | Data Koder | Pengaturan Kategori | Pengaturan Rating | Ganti Nama | Ganti Password | Logout

1 Kembali ke Page Sebelumnya
2 Download dalam CSV

Komentar Valid/Tidak 3

Jumlah Komentar	Valid	Tidak Valid
70	41	29

Rekap per Kategori 4

Komentar Valid	Sikap terhadap aktor	Sikap terhadap karakter	Sikap terhadap film	Sikap terhadap penempatan produk	Sikap terhadap merek
41	10	11	20	0	0

Rekap per Rating 5

Komentar Valid	Positif	Negatif
41	29	12

Pengeruk Komentar Indonesia
2019, Fakultas Teknologi Informasi, Universitas Tarumanagara

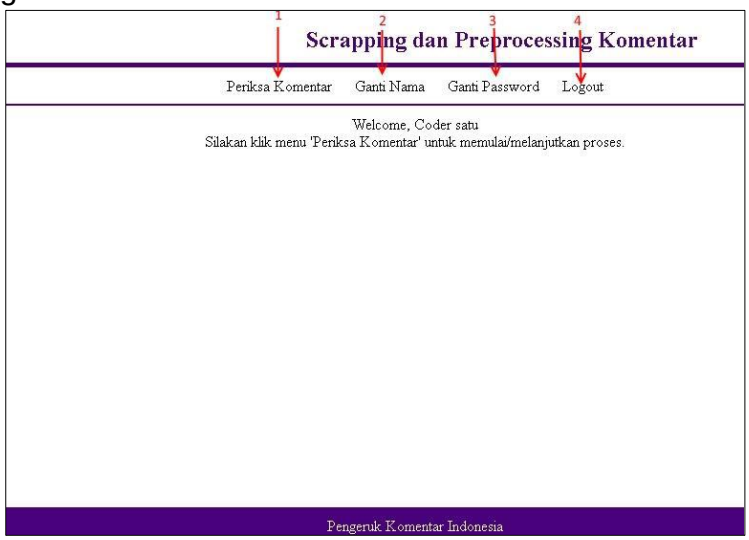
Gambar 24 Rekap Hasil Akhir

C. Coder

Home Coder

Ini adalah tampilan awal untuk admin, seperti terlihat pada Gambar 25. Berikut adalah penjelasan untuk menu Coder pada Gambar 25:

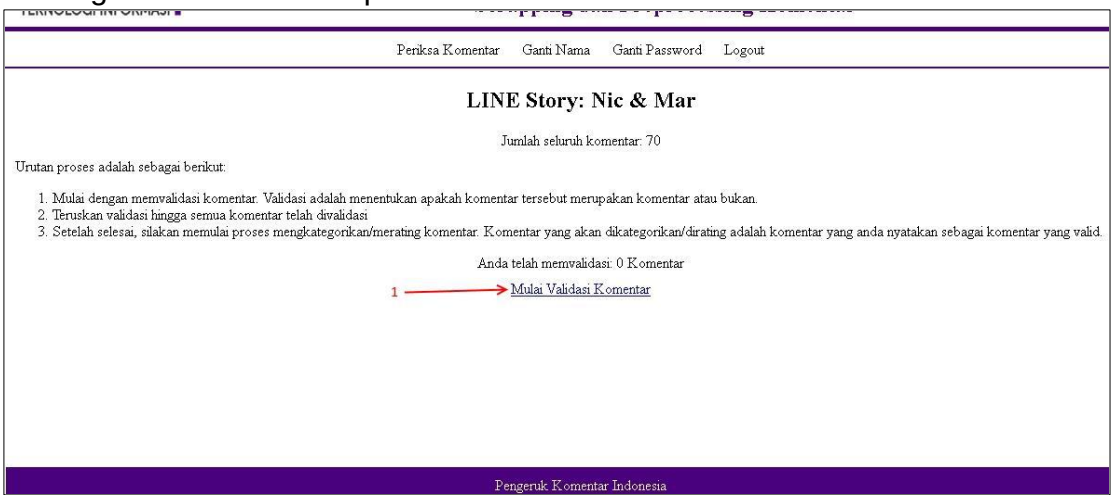
1. Menu periksa komentar. Menu untuk melihat progress pengujian coder tersebut.
2. Menu ganti nama. Menu ini untuk mengganti nama akun.
3. Menu ganti password. Menu ini untuk mengganti password akun.
4. Klik untuk logout



Gambar 25 Home Coder

Periksa Komentar

Ini adalah tampilan untuk memulai dan melihat progress pengujian, seperti terlihat pada Gambar 26. Pengujian diawali dengan memvalidasi komentar, yang dapat dilakukan dengan mengeklik link nomor 1 pada Gambar 26.



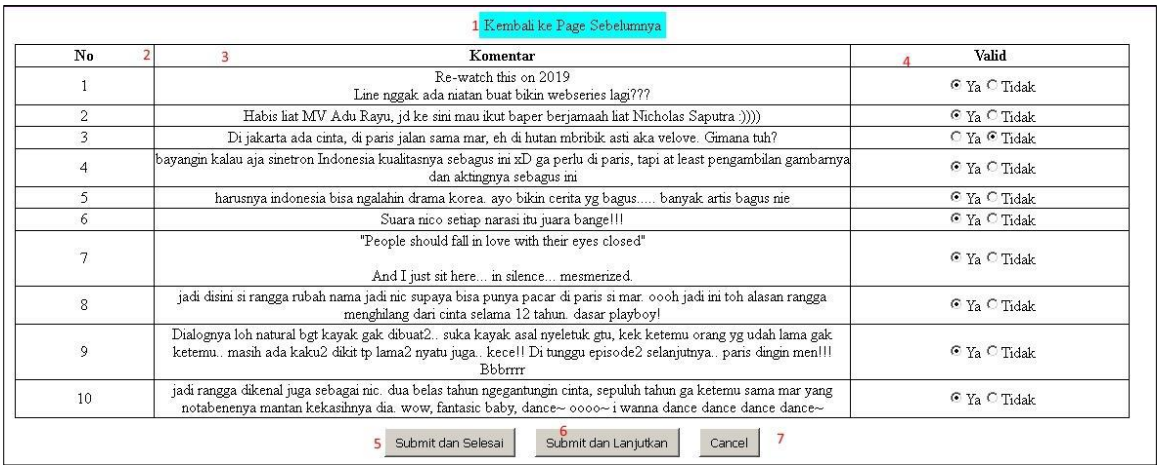
Gambar 26 Periksa Komentar

Form Validasi Komentar

Page ini berisi form yang digunakan coder untuk menentukan apakah komentar tersebut valid atau tidak, setiap *page* hanya menampilkan 10 komentar. Contoh tampilan dapat dilihat pada Gambar 27. Berikut adalah keterangan dari Gambar 27:

1. Link untuk kembali ke periksa komentar
2. Kolom nomor
3. Kolom teks komentar
4. Kolom berisi *field* untuk menentukan komentar tersebut valid atau tidak.
5. *Button* untuk *submit* dan kembali ke periksa komentar

6. *Button* untuk *submit* dan melanjutkan validasi ke komentar berikutnya.
7. *Button* untuk *reset* form.



Gambar 27 Form Validasi Komentar

Setelah coder selesai memvalidasi semua komentar, coder akan diarahkan kembali ke tampilan periksa komentar. Contoh tampilan dapat dilihat pada Gambar 28. Link untuk melakukan validasi sudah tidak ada, diganti dengan link untuk menentukan kategori komentar (link nomor 1 pada Gambar 28) dan rating komentar (link nomor 2 pada Gambar 28).

Gambar 28 Tampilan Periksa Komentar Setelah Coder Selesai Validasi Komentar

Form Kategori Komentar

Page ini berisi form yang digunakan coder untuk menentukan kategori untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Contoh tampilan dapat dilihat pada Gambar 29. Berikut adalah keterangan dari Gambar 29:

1. Kolom nomor
2. Kolom teks komentar
3. Kolom berisi *field* untuk menentukan kategori komentar
4. *Button* untuk *submit* dan kembali ke periksa komentar
5. *Button* untuk *submit* dan melanjutkan proses ke komentar berikutnya.
6. *Button* untuk *reset* form.

Periksa Komentar Ganti Nama Ganti Password Logout

Kembali ke Page Sebelumnya

No 1	2 Komentar	3 Kategori
1	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	Sikap terhadap aktor
2	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	Sikap terhadap film
3	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	Sikap terhadap film
4	Suara nico setiap narasi itu juara bange!!!	Sikap terhadap aktor
5	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	Sikap terhadap karakter
6	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	Sikap terhadap karakter
7	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	Sikap terhadap film
8	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	Sikap terhadap aktor
9	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _	Sikap terhadap aktor
10	Kata Warhol "People should fall in love with their eyes closed"	Sikap terhadap karakter

4 Submit dan Selesai 5 Submit dan Lanjutkan Cancel 6

Gambar 29 Form Kategori Komentar

Form Rating Komentar

Page ini berisi form yang digunakan coder untuk menentukan rating untuk setiap komentar, setiap *page* hanya menampilkan 10 komentar. Komentar yang ditampilkan hanya komentar yang dipilih sebagai komentar yang valid. Contoh tampilan dapat dilihat pada Gambar 30. Berikut adalah keterangan dari Gambar 30:

1. Kolom nomor
2. Kolom teks komentar
3. Kolom berisi *field* untuk menentukan rating komentar
4. *Button* untuk *submit* dan kembali ke periksa komentar
5. *Button* untuk *submit* dan melanjutkan proses ke komentar berikutnya.
6. *Button* untuk *reset* form.

Kembali ke Page Sebelumnya

No 1	2 Komentar	3 Rating
1	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :))))	Positif
2	bayangin kalau aja sinetron Indonesia kualitasnya sebagus ini xD ga perlu di paris, tapi at least pengambilan gambarnya dan aktingnya sebagus ini	Negatif
3	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	Positif
4	Suara nico setiap narasi itu juara bange!!!	Positif
5	"People should fall in love with their eyes closed" And I just sit here... in silence... mesmerized.	Positif
6	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oooh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar playboy!	Negatif
7	Dialognya loh natural bgt kayak gak dibuat2.. suka kayak asal nyeletuk gtu, kek ketemu orang yg udah lama gak ketemu.. masih ada kaku2 dikit tp lama2 nyatu juga.. kece!! Di tunggu episode2 selanjutnya.. paris dingin men!!! Bbbrrrr	Positif
8	jadi rangga dikenal juga sebagai nic. dua belas tahun ngegantungin cinta, sepuluh tahun ga ketemu sama mar yang notabenenya mantan kekasihnya dia. wow, fantasic baby, dance~ oooo~ i wanna dance dance dance dance~	Negatif
9	Lucu ya liat logat bahasanya mbak Mar karena udh lama di luar negeri. Trus, liat mbak Mar kyk liat Maudy Ayunda versi dewasa. Ga berkedip _	Positif
10	Kata Warhol "People should fall in love with their eyes closed"	Positif

4 Submit dan Selesai 5 Submit dan Lanjutkan Cancel 6

Gambar 30 Form Rating Komentar

Setelah coder selesai menentukan kategori atau rating untuk semua komentar yang valid, coder akan diarahkan kembali ke tampilan periksa komentar. Jika keduanya telah selesai, tampilan periksa komentar akan seperti pada Gambar 31. Terima kasih telah melakukan pengujian.

LINE Story: Nic & Mar

Jumlah seluruh komentar: 70

adalah menentukan apakah komentar tersebut merupakan komentar atau bukan.
n divalidasi
ategorikan/merating komentar. Komentar yang akan dikategorikan/dirating adalah komentar yang and

Proses validasi telah selesai.

Proses mengkategorikan komentar telah selesai.
Terima Kasih

Proses memberi rating pada komentar telah selesai.
Terima Kasih

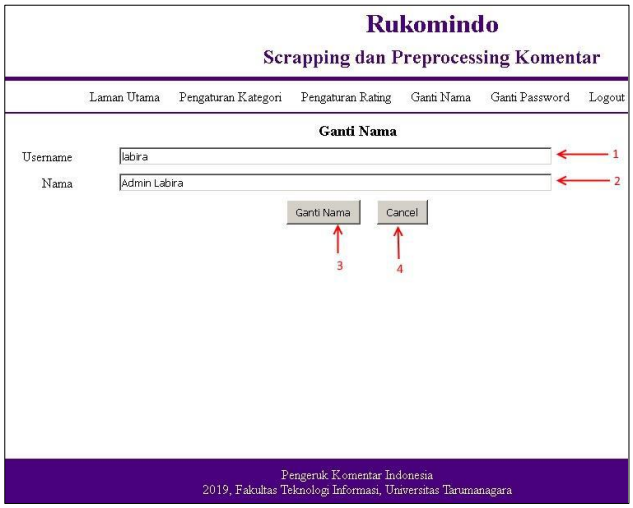
Gambar 31 Tampilan Periksa Komentar Setelah Coder Selesai Menentukan Kategori dan Rating Komentar

D. Fungsi Umum

Ganti Nama

Page ini berisi form yang digunakan untuk mengganti nama user. Contoh tampilan dapat dilihat pada Gambar 32. Berikut adalah keterangan dari Gambar 32:

1. *Field* yang otomatis terisi username dari user.
2. *Field* untuk nama user
3. *Button* untuk *submit* form.
4. *Button* untuk *reset* form.

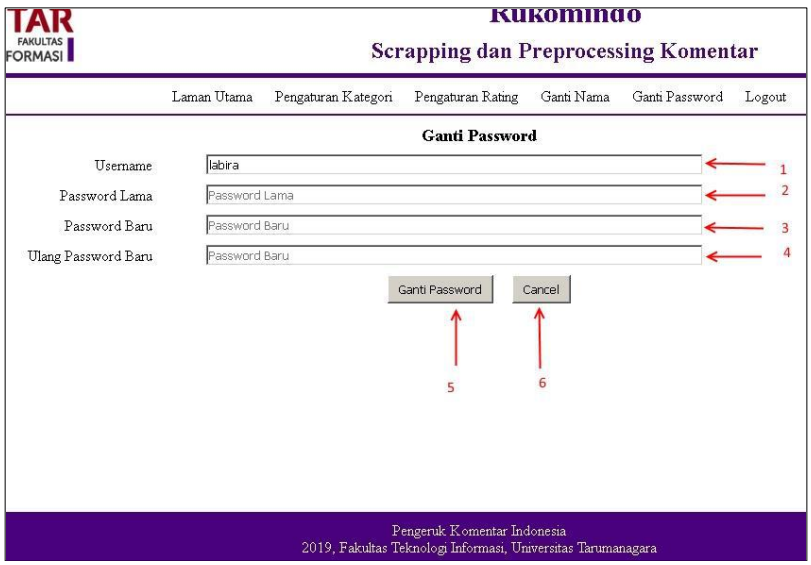


Gambar 32 Form Ganti Nama

Ganti Password

Page ini berisi form yang digunakan untuk mengganti password user. Contoh tampilan dapat dilihat pada Gambar 33. Berikut adalah keterangan dari Gambar 33:

1. *Field* yang otomatis terisi username dari user.
2. *Field* untuk password lama
3. *Field* untuk password baru
4. *Field* untuk ulang password baru.
5. *Button* untuk *submit* form.
6. *Button* untuk *reset* form.



Gambar 33 Form Ganti Password

Lampiran VI
Log Book

LOG BOOK PENELITIAN YANG DIAJUKAN
KE DIREKTORAT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS TARUMANAGARA

Aplikasi Filtering dan Preprocessing Text pada Data Komentar untuk *Text Mining* (Studi Kasus: Komentar Web Series Iklan Youtube)

Diusulkan oleh:
Ketua Tim
Viny Christanti M. 0317088203/10805002

Anggota:
Tri Sutrisno, S.Si., M.Sc. 0320018801/10816004

Fakultas Teknologi Informasi
Universitas Tarumanagara
Jakarta

2020

LOG BOOK PENELITIAN
Aplikasi Filtering dan Preprocessing Text pada

Data Komentar untuk *Text Mining*

(Studi Kasus: Komentar Web Series Iklan Youtube)
(Agustus 2019-Januari 2020)

1. Peneliti Utama : Viny Christanti M.
2. Peneliti Anggorta : Tri Sutrisno
3. Mahasiswa : Walda Vivian / 535160010
4. Riset Asisten : Freddy Kurniawan (Asisten Teknik Informatika)

No	**Keterangan Penelitian**	**Peneliti /Anggota peneliti**
1.	Studi literature	Viny, Tri Sutrisno, Walda
2.	Analisa kebutuhan dengan pakar	Viny, Cokky (FE)
3.	Analisis sistem	Viny, Tri Sutrisno
4.	Perancangan antar muka dan alur sistem	Viny, Freddy
5.	Perancangan basis data	Viny, Freddy
6.	Pembuatan program Crawling dan filtering halaman web	Viny, Freddy, Walda
7.	Pembuatan program pra-pemrosesan halaman web	Viny, Freddy, Walda
8.	Pengujian sistem	Viny, Freddy, Walda, Cokky
9.	Implementsi ke server	Viny, Freddy
10.	8 November 219 Pembelian domain rukomindo.com	Viny, Freddy
11.	Pengumpulan data komentar	Viny, Cokky, Coder 3 orang
12.	Evaluasi hasil filtering dan prapemrosesan	Viny, Freddy,

No	Keterangan Penelitian	Peneliti /Anggota peneliti
		Walda, Cokky
13.	Pembuatan laporan kemajuan	Viny, Tri Sutrisno
14.	Pembuatan draft karya ilmiah	Viny
15.	Presentasi karya ilmiah ke TICATE 21-22 November 2019	Viny
16.	Revisi paper TICATE	Viny
17.	Pembuatan laporan akhir	Viny, Tri Sutrisno

Mengetahui,

Ketua Penelitian

(Viny Christanti M.)