



**IMPLEMENTASI TEKNOLOGI *TEXT MINING*  
DAN *WEB SCRAPING* UNTUK MENGUMPULKAN  
INFORMASI PADA *WEBSITE* PELATIHAN  
KETERAMPILAN KHUSUS PEREMPUAN**

**LAPORAN SKRIPSI**

**NISYA YUNITA**

**4616010015**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK NEGERI JAKARTA**

**2020**



**IMPLEMENTASI TEKNOLOGI *TEXT MINING*  
DAN *WEB SCRAPING* UNTUK MENGUMPULKAN  
INFORMASI PADA *WEBSITE* PELATIHAN  
KETERAMPILAN KHUSUS PEREMPUAN**

**LAPORAN SKRIPSI**

**Dibuat untuk Melengkapi Syarat-Syarat yang Diperlukan untuk  
Memperoleh Diploma Empat Politeknik**

**NISYA YUNITA**

**4616010015**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK NEGERI JAKARTA**

**2020**

## HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Nisya Yunita  
NPM : 4616010015  
Tanggal : 11 Agustus 2020  
Tanda Tangan :

POLITEKNIK  
NEGERI  
JAKARTA



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta





© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

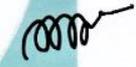
LEMBAR PENGESAHAN

Skripsi diajukan oleh :

Nama : Nisya Yunita  
NIM : 4616010015  
Program Studi : Teknik Informatika  
Judul Skripsi : Implementasi Teknologi *Text Mining* dan *Web Scraping* Untuk Mengumpulkan Informasi Pada *Website* Pelatihan Keterampilan Khusus Perempuan

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Selasa, Tanggal 14, Bulan Juli, Tahun 2020, dan dinyatakan **LULUS**.

Disahkan oleh

Pembimbing I : Risna Sari, S.Kom., M.T.I. (  )  
Pembimbing II : (  )  
Penguji I : Eriya, S.Kom., M.T. (  )  
Penguji II : Asep Taufik Muharram, S.Kom., M.Kom. (  )  
Penguji III : Drs. Refirman, M.Kom. (  )

POLITEKNIK  
NEGERI  
JAKARTA

Mengetahui,

Jurusan Teknik Informatika dan Komputer

Ketua

  
Mauldy Laya, S.Kom., M.Kom.  
NIP. 197802112009121003



## KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Allah SWT, karena atas berkat dan limpahan rahmat-Nya, penulis dapat menyelesaikan Skripsi ini. Penulisan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Terapan Politeknik.

Pembahasan dari penulisan skripsi ini mengenai pembuatan *website* tentang pengumpulan informasi mengenai pelatihan keterampilan khusus perempuan guna memudahkan kaum perempuan dapat mendapatkan informasi mengenai pelatihan keterampilan pada satu sumber informasi.

Dalam penyusunan Skripsi ini, penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, sulit bagi penulis untuk menyelesaikan Skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

- a. Ibu Risna Sari, S.Kom., M.T.I. selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan Skripsi ini;
- b. Bapak Muhammad Ihsan, selaku Asisten Deputi Kesetaraan Gender Bidang Ekonomi Kementerian Pemberdayaan Perempuan dan Perlindungan Anak yang telah banyak membantu dalam usaha memperoleh data yang penulis perlukan;
- c. Orang tua dan keluarga penulis yang memberikan bantuan dukungan melalui doa, moral, dan material;
- d. Partner, sahabat, dan teman-teman yang telah mendukung dan memberi semangat selama pengerjaan Skripsi.

Akhir kata, penulis berharap Allah SWT membalas segala kebaikan kepada pihak yang telah membantu. Semoga Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 09 April 2020

Penulis



## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Nisya Yunita  
NIM : 4616010015  
Program Studi : Teknik Informatika  
Jurusan : Teknik Informatika dan Komputer  
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Implementasi Teknologi *Text Mining* dan *Web Scraping* Untuk Mengumpulkan Informasi Pada *Website* Pelatihan Keterampilan Khusus Perempuan beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Politeknik Negeri Jakarta berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada tanggal: 09 April 2020

Yang menyatakan

(Nisya Yunita)

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## Implementasi Teknologi *Text Mining* dan *Web Scraping* Untuk Mengumpulkan Informasi Pada *Website* Pelatihan Keterampilan Khusus Perempuan

### *Abstrak*

Mencari informasi melalui internet sangatlah mudah, dengan waktu yang relatif singkat bisa mendapatkan berbagai macam informasi yang diinginkan. Banyaknya sumber Informasi yang ada dan beragam akan menyebabkan timbulnya *information overload*. *Information overload* adalah kondisi dimana seseorang sulit dalam menerima, memahami, dan mengolah informasi dikarenakan sumber informasi yang berlebih. *Information overload* salah satunya terjadi pada informasi mengenai pelatihan keterampilan perempuan yang dibutuhkan untuk meningkatkan kualitas diri. Dengan banyaknya laman web yang menyajikan informasi mengenai pelatihan keterampilan, tidaklah mudah untuk perempuan dapat mendapatkan informasi yang diinginkan karena harus mengakses sebanyak mungkin laman web guna mendapatkan informasi yang cocok. Penelitian ini bertujuan untuk membuat sistem pengumpulan informasi mengenai pelatihan keterampilan khusus perempuan dengan mengimplementasikan teknologi *text mining* dan *web scraping* dan menggabungkannya ke dalam satu laman informasi sehingga mempermudah kaum perempuan dalam mendapatkan informasi mengenai pelatihan keterampilan. Dari eksperimen yang sudah dilakukan, *text mining* dan *web scraping* berhasil diimplementasikan, dengan adanya laman web yang mengintegrasikan informasi dari laman-laman web yang lain, perempuan mendapatkan informasi mengenai pelatihan keterampilan pada satu sumber informasi.

**Kata kunci:** Keterampilan, Pelatihan, Perempuan, *Text Mining*, *Web Scraping*.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PERNYATAAN ORISINALITAS .....	iii
LEMBAR PENGESAHAN .....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI .....	vi
<i>Abstrak</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL .....	xii
DAFTAR LAMPIRAN .....	xiii
<b>BAB I.....</b>	<b>1</b>
<b>PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat.....	3
1.5 Metode Pelaksanaan Skripsi.....	4
1.5.1 Metode Pengumpulan Data.....	4
1.5.2 Metode Pengembangan Sistem .....	4
<b>BAB II .....</b>	<b>6</b>
<b>TINJAUAN PUSTAKA .....</b>	<b>6</b>
2.1 Web Scraping .....	6
2.2 Text Mining .....	8
2.3 Pelatihan .....	11
2.4 Keterampilan .....	12
2.5 Website .....	12
2.6 Bootstrap .....	13
2.7 Python.....	13
2.8 Flask .....	14
2.9 Penelitian Sejenis .....	14

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

<b>BAB III.....</b>	<b>16</b>
<b>PERENCANAAN DAN REALISASI.....</b>	<b>16</b>
3.1    Deskripsi Program Aplikasi .....	16
3.1.1    Analisis Kebutuhan .....	17
3.2    Cara Kerja Pembuatan Aplikasi .....	18
3.3    Cara Kerja Program Aplikasi .....	21
3.4    Rancangan Program Aplikasi .....	22
3.4.1    Use Case Diagram.....	22
3.4.2    Activity Diagram.....	23
3.4.3    Sequence Diagram .....	28
3.4.4    Class Diagram.....	31
3.4.5    Desain Tabel.....	32
3.4.6    Mockup Aplikasi.....	33
3.4.7    Realisasi Program Aplikasi.....	35
<b>BAB IV .....</b>	<b>54</b>
<b>PEMBAHASAN .....</b>	<b>54</b>
4.1    Pengujian .....	54
4.1.1    Deskripsi Pengujian .....	54
4.1.2    Prosedur Pengujian .....	54
4.1.3    Data Hasil Pengujian.....	55
4.2    Analisis Data/Evaluasi .....	69
<b>BAB V.....</b>	<b>72</b>
<b>PENUTUP.....</b>	<b>72</b>
5.1    Kesimpulan.....	72
5.2    Saran.....	73
<b>DAFTAR PUSTAKA .....</b>	<b>74</b>



## DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi svm .....	10
Gambar 3. 1 Flowchart pengumpulan informasi menggunakan text mining dan web scraping .....	18
Gambar 3. 2 Flowchart modul ‘Membersihkan informasi website melalui tahapan text pre-processing’ .....	19
Gambar 3. 3 Flowchart modul ‘Menentukan suatu website masuk ke dalam kelas tertentu melalui tahapan text classification menggunakan metode SVM’ .....	20
Gambar 3. 4 Flowchart program pada website pelatihan keterampilan khusus perempuan .....	21
Gambar 3. 5 Use case diagram pada website pelatihan keterampilan khusus perempuan .....	22
Gambar 3. 6 Activity diagram mencari pelatihan .....	24
Gambar 3. 7 Activity diagram memilih kategori pelatihan .....	25
Gambar 3. 8 Activity diagram melihat detail informasi pelatihan .....	25
Gambar 3. 9 Activity diagram login admin .....	26
Gambar 3. 10 Activity diagram mengelola data informasi .....	27
Gambar 3. 11 Sequence diagram login admin .....	28
Gambar 3. 12 Sequence diagram mengelola data informasi .....	29
Gambar 3. 13 Sequence diagram melihat detail pelatihan makanan .....	29
Gambar 3. 14 Sequence diagram melihat detail pelatihan kerajinan tangan .....	30
Gambar 3. 15 Sequence diagram pencarian pelatihan .....	30
Gambar 3. 16 Class diagram aplikasi .....	31
Gambar 3. 17 Mock up login admin .....	33
Gambar 3. 18 Mock up halaman admin .....	33
Gambar 3. 19 Mock up halaman utama .....	34
Gambar 3. 20 Mock up detail informasi pelatihan .....	34
Gambar 3. 21 Tampilan login admin .....	36
Gambar 3. 22 Tokenisasi .....	37
Gambar 3. 23 Stopwords removal .....	38
Gambar 3. 24 Stopwords_pelatihan .....	39
Gambar 3. 25 Stemming dan Lemmatization .....	40
Gambar 3. 26 Persiapan data klasifikasi .....	41
Gambar 3. 27 Mengubah kata menjadi matriks .....	42
Gambar 3. 28 Tampilan halaman admin .....	45
Gambar 3. 29 Tampilan halaman website mengenai pelatihan keterampilan .....	46
Gambar 3. 30 Tampilan halaman website bukan mengenai pelatihan keterampilan .....	46
Gambar 3. 31 Tampilan halaman bukan url .....	46

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Gambar 3. 32 Tampilan hasil scraping dengan tanggal .....	48
Gambar 3. 33 Tampilan hasil scraping tanpa tanggal .....	49
Gambar 3. 34 Tampilan form update data .....	50
Gambar 3. 35 Tampilan berhasil mempebarui data .....	50
Gambar 3. 36 Tampilan halaman detail informasi pelatihan makanan.....	51
Gambar 3. 37 Tampilan halaman detail informasi pelatihan kerajinan tangan.....	52
Gambar 3. 38 Tampilan halaman pencarian pelatihan .....	53
Gambar 3. 39 Hasil kuisisioner data usia .....	65
Gambar 3. 40 Hasil kuisisioner data status.....	66
Gambar 3. 41 Hasil kuisisioner data status pekerjaan .....	66
Gambar 3. 42 Hasil kuisisioner pertanyaan 1 .....	66
Gambar 3. 43 Hasil kuisisioner pertanyaan 2 .....	67
Gambar 3. 44 Hasil kuisisioner pertanyaan 3 .....	67
Gambar 3. 45 Hasil kuisisioner pertanyaan 4 .....	67
Gambar 3. 46 Hasil kuisisioner pertanyaan 5 .....	68
Gambar 3. 47 Hasil kuisisioner pertanyaan 6 .....	68
Gambar 3. 48 Hasil kuisisioner pertanyaan 7 .....	68

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**POLITEKNIK  
NEGERI  
JAKARTA**



## DAFTAR TABEL

Tabel 1. Desain Tabel Admin .....	32
Tabel 2. Desain Tabel Articles.....	32
Tabel 3. Prosedur pengujian black box testing .....	54
Tabel 4. Pengujian login admin .....	55
Tabel 5. Pengujian halaman admin .....	56
Tabel 6 Pengujian url website mengenai pelatihan keterampilan.....	57
Tabel 7. Pengujian halaman detail informasi pelatihan kategori makanan.....	61
Tabel 8. Pengujian halaman detail informasi pelatihan kategori kerajinan tangan.....	61
Tabel 9. Pengujian halaman pencarian pelatihan .....	62
Tabel 10. Demografi user 1.....	62
Tabel 11. Halaman detail informasi pelatihan kategori makanan pada user 1.....	63
Tabel 12. Halaman detail informasi pelatihan kategori kerajinan pada user 1 .....	63
Tabel 13. Pencarian pelatihan pada user .....	63
Tabel 14. Demografi user 2.....	64
Tabel 15. Halaman detail informasi pelatihan kategori makanan pada user 2.....	64
Tabel 16. Halaman detail informasi pelatihan kategori kerajinan pada user 2 .....	64
Tabel 17. Pencarian pelatihan pada user 2 .....	65

**POLITEKNIK  
NEGERI  
JAKARTA**

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## DAFTAR LAMPIRAN

Lampiran 1. Transkrip Wawancara.....	78
Lampiran 2. Dokumentasi Pengujian.....	81





- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Era globalisasi, manusia diharuskan memiliki keterampilan untuk memenuhi kebutuhan hidupnya. Manusia memerlukan keterampilan agar mengetahui potensi diri yang dimiliki dan dapat berwirausaha agar tidak bergantung kepada orang lain terutama kaum perempuan yang bergantung pada laki-laki. Badan Pusat Statistik menyatakan tingkat pengangguran terbuka pada perempuan tahun 2015-2018 berkisar 5—6,5% (BPS, 2019). Data tersebut terjadi karena perempuan biasanya tidak boleh dalam mencari nafkah dan lebih kepada mengurus rumah tangga saja dalam artian memang kultural sebagai perempuan dan karena perempuan mempunyai informasi yang lebih minim daripada laki-laki (Ihsan, 2020). Padahal kaum perempuan memiliki potensi untuk meningkatkan kualitas dirinya sebagai manusia. Dapat dilihat dari perubahan cara berfikir dan bersikap yang mencerminkan adanya kesadaran akan kemampuan atau identitas dirinya. Selain itu, kekuatan pembangunan ekonomi akan optimal apabila penduduk termasuk kaum perempuan dapat berperan dalam membangun suatu daerah melalui kemampuan karena mampu mengembangkan potensi dirinya (Telaumbanua, 2018). Kaum perempuan dapat mengembangkan potensi dirinya dengan memberikan pelatihan keterampilan yang dapat mendorong kemandirian dan meningkatkan kesadaran dan kualitas diri.

Pelatihan dapat didefinisikan sebagai pengajaran atau pemberian pengalaman kepada seseorang untuk mengembangkan tingkah laku (pengetahuan, *skill*/keterampilan, dan sikap) agar mencapai sesuatu yang diinginkan (Robinson dalam Hidayat, 2016). Berbekal pelatihan keterampilan, perempuan memungkinkan dalam memperoleh akses untuk mendapat sumber mata pencaharian yang layak, memberikan peluang untuk memperluas ilmu pengetahuan dan pengalaman, dan mengurangi tingkat stres dalam memenuhi kebutuhan hidup. Kementerian Pemberdayaan Perempuan dan Perlindungan Anak telah mengadakan model pengembangan Industri Rumahan yang telah dilaksanakan selama 3 tahun



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

semenjak tahun 2016 akhir – 2019 akhir yang dimana pelatihan keterampilan diadakan di berbagai daerah. Namun, penyampaian adanya informasi pelatihan biasanya disampaikan melalui *whatsapp* atau pendamping yang merupakan warga untuk diberitau akan adanya pelatihan. Sehingga informasi yang diberikan belum secara merata. Maka dari itu, salah satu langkah pertama yang dapat dilakukan untuk mewujudkan hal tersebut adalah aktif mencari dan mengumpulkan berbagai informasi dari seseorang atau kelompok yang berkompeten dan berpengalaman mengenai pola pemberdayaan perempuan. Tujuannya agar diketahui kegiatan yang dilakukan untuk mengisi waktu dan menghasilkan tenaga yang terampil (Rahayu, 2019).

Saat ini mengumpulkan informasi dari seseorang atau kelompok tidaklah cukup, sangat dimungkinkan seseorang mencari informasi melalui internet karena sangatlah mudah, dengan waktu yang relatif singkat bisa mendapatkan berbagai informasi yang diinginkan, seperti sumber informasi pelatihan. Namun, mencari informasi mengenai pelatihan keterampilan tidaklah mudah karena banyaknya sumber informasi yang ada dan beragam. Informasi dengan jumlah yang sangat banyak dan beragam akan menyebabkan timbulnya *information overload*. *Information overload* adalah banyaknya informasi yang diterima oleh manusia sehingga sulit untuk mengolahnya (Priyanto & Ma'arif, 2018). Dengan adanya *information overload*, perlu menyatukan berbagai informasi dari berbagai sumber yang dibentuk menjadi satu sumber informasi. Maka dari itu, perlu memilih informasi yang relevan dari berbagai sumber informasi pelatihan keterampilan untuk dimasukkan ke dalam satu laman informasi.

Dalam penelitian ini akan dikembangkan pengumpulan informasi dari *website* yang tersedia di internet dengan memanfaatkan teknologi *text mining* yang kemudian akan di ambil informasinya secara otomatis menggunakan teknologi *web scraping*. Saat ini banyak *web scraping tools* yang tersedia namun dalam proses pengambilan informasinya masih secara manual seperti contohnya pada *web scraper* perlu membuat *sitemap* terlebih dahulu untuk menavigasi *website*, kurangnya fitur-fitur otomatis, dan terbatas dalam penggunaannya sehingga dapat menghasilkan data yang tidak sesuai ekspektasi. Sehingga pada penelitian ini menggunakan teknologi *text mining* untuk mengetahui apakah *website* yang akan diambil informasinya

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

adalah *website* mengenai pelatihan keterampilan atau tidak, teknologi *web scraping* untuk mengambil informasi yang sudah ditentukan secara otomatis dan menggabungkannya ke dalam satu laman informasi sehingga mempermudah kaum perempuan dalam mendapatkan informasi mengenai pelatihan keterampilan.

### 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka perumusan masalah dalam pembuatan sistem ini adalah:

Bagaimana teknologi *text mining* dan *web scraping* diimplementasikan untuk mengumpulkan informasi pada *website* pelatihan keterampilan khusus perempuan?

### 1.3 Batasan Masalah

Batasan masalah dalam pembuatan sistem ini adalah:

- a. Sistem ini hanya diperuntukkan kaum perempuan.
- b. Sistem ini sebagai sarana informasi kepada kaum perempuan dalam mencari informasi pelatihan keterampilan.
- c. Kategori pelatihan keterampilan yang digunakan adalah data dari hasil wawancara dengan Asisten Deputi Kesetaraan Gender Bidang Ekonomi Kementrian Pemberdayaan Perempuan dan Perlindungan Anak.
- d. Sistem ini mengeksplorasi *website* yang memberikan informasi mengenai pelatihan keterampilan dengan mengumpulkan informasi yang ada pada *website* menggunakan teknologi *text mining* dengan dua tahapan yakni *text pre-processing* dan *text classification* metode SVM (*Support Vector Machine*) dan *web scraping* menggunakan pustaka *beautifulsoup*.
- e. Sistem ini melakukan pengambilan informasi yang sudah ditentukan yaitu judul, tanggal, dan deskripsi.

### 1.4 Tujuan dan Manfaat

Adapun tujuan dan manfaat selama melakukan penelitian, sebagai berikut:

Tujuan dari pembuatan sistem ini adalah membuat sistem pengumpulan informasi mengenai pelatihan keterampilan khusus perempuan pada *website* dengan

menggunakan teknologi *text mining* dan *web scraping*. Adapun manfaat yang dapat diperoleh dari pembuatan sistem ini adalah:

- a. Kaum perempuan dapat mendapatkan informasi mengenai pelatihan keterampilan pada satu sumber informasi.
- b. Kaum perempuan dapat menemukan pelatihan keterampilan yang cocok berdasarkan informasi yang didapatkan.
- c. Kaum perempuan dapat mengetahui kegiatan yang dilakukan mengenai pelatihan keterampilan.

## 1.5 Metode Pelaksanaan Skripsi

### 1.5.1 Metode Pengumpulan Data

#### a. Wawancara

Wawancara adalah proses pengumpulan data dengan cara interaksi baik secara langsung maupun jarak jauh untuk menggali informasi tertentu yang dibahas. Wawancara dilakukan dengan mengajukan beberapa pertanyaan tentang pelatihan keterampilan untuk perempuan.

#### b. Dokumentasi

Dokumentasi adalah proses dalam pengumpulan data berdasarkan pencatatan berbagai sumber informasi. Dengan adanya dokumentasi, penulis mendapatkan informasi mengenai pelatihan keterampilan untuk perempuan.

#### c. Studi Literatur

Penulis mengumpulkan data dan mempelajari serta mencari informasi yang dibutuhkan, yang berkaitan dengan masalah yang akan dibahas, sebagai acuan dalam pembuatan dan pengembangan sistem, dan penyusunan laporan.

### 1.5.2 Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan adalah metode *waterfall*. Metode *waterfall* merupakan model pengembangan perangkat lunak yang sistematis dan sekuensial. (Sasmito, 2017).

Berikut adalah tahapan proses dalam metode *waterfall* sebagai berikut:



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumuk dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### a. *Requirement Analysis and Definition*

Pada tahap ini, menganalisis untuk menspesifikasikan kebutuhan perangkat lunak yang didefinisikan secara jelas sehingga dapat mudah dipahami, seperti *input* dan *output* sistem, proses/analisa yang berjalan pada sistem, dan penggunaan basis data.

### b. *System and Software Design*

Tahapan perancangan sistem adalah perancangan aplikasi yang akan dibangun, merancang arsitektur sistem berdasarkan hasil dari tahap analisis kebutuhan. Dimana perancangan sistem ini dengan menggunakan *software* Adobe XD dan menggunakan UML (*Unified Modelling Language*) dalam memodelkan alur sistem yang akan dibuat.

### c. *Implementation*

Pada tahap ini merupakan proses dimana keseluruhan desain sistem yang telah disusun direalisasikan sebagai serangkaian program. Menggunakan teknologi *web scraping* dan *text mining* untuk memperoleh informasi dan bahasa pemrograman yang digunakan dalam pembuatan *website* adalah HTML, CSS, Javascript, disertai dengan menggunakan *framework bootstrap*, dan menggunakan bahasa pemrograman python dengan *framework flask* sebagai *back-end* sistem.

### d. *Testing*

Pada tahap ini, sistem yang sudah dibuat dilakukan pengujian untuk memastikan apakah sudah berjalan sesuai dengan kebutuhan perangkat lunak atau tidak dan sesuai dengan yang dirancang.

### e. *Maintenance*

*Maintenance* pada sistem seperti proses perbaikan sistem apabila ditemukan adanya kesalahan atau *bug* yang tidak ditemukan pada tahap *testing*.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta  
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Web Scraping

*Web scraping* adalah teknik untuk mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari *web scraper* adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam web yang baru. *Web scraping* berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Dalam definisi lain, *Web scraping* adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa laman web yang dibangun dengan bahasa *markup* seperti HTML atau XHTML yang bertujuan untuk mengambil informasi dari halaman tersebut baik secara keseluruhan atau sebagian untuk digunakan bagi kepentingan lain. Secara umum, ada empat tahapan dalam penggunaan *web scraping* untuk mengambil data secara otomatis dari sebuah laman web sebagai berikut (Priyanto & Ma'arif, 2018):

1. Mempelajari dokumen HTML dari *website* yang akan diambil informasinya untuk tag HTML yang mengapit informasi yang akan diambil.
2. Menelusuri mekanisme navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat.
3. Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan.
4. Informasi yang didapat dari langkah 3 disimpan dalam format data tertentu.

Dari apa yang dinyatakan, *web scraping* adalah proses pengambilan informasi dari *website* yang telah ditentukan untuk digunakan dalam kepentingan lain yang dikumpulkan dalam *website* yang baru.

Semakin meningkatnya teknologi, terdapat banyak *web scraping tools* dalam proses pengambilan informasi dari *website* secara instan. Untuk mendukung pemahaman pembaca mengenai parameter-parameter yang perlu diperhatikan dalam tahapan pengambilan informasi dari sebuah laman web. Dalam hal ini, penulis



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

membandingkan tiga macam *web scraping* yang sudah ada. Berikut ini analisis perbandingan *web scraping* yaitu *Web Scraper*, *Data Scraper*, dan *Parsehub*.

### 1. *Web Scraper*

*Web Scraper* merupakan *web scraping tool* dalam bentuk ekstensi chrome. Dimana ekstensi ini dapat membuat *sitemap* bagaimana sebuah web dijelajah dan apa yang seharusnya diambil. Dengan menggunakan *sitemap*, *Web Scraper* akan menavigasi situs dan mengekstrak semua data yang diinginkan. Data yang di *scraping* dapat di *export* sebagai CSV. (webscraper.io, 2019)

### 2. *Data Scraper*

*Data Scraper* mengesktrak data dari halaman web HTML dan mengimpornya ke dalam *spreadsheet* Microsoft Excel. *Data Scraper* adalah *web scraping tool* dalam bentuk ekstensi chrome yang dapat mengeskpor halaman web ke file XLS, CSV, XLSX atau TSV. Pengguna dapat menggunakan *Data Scraper* secara gratis dengan diberikannya 500 halaman *scraping* setiap bulan. (data-miner.io, 2019). Dimana *Data Scraper* memiliki 7 tahapan dalam proses pengambilan informasi yaitu

- Start* untuk memilih jenis halaman seperti *list page* atau *detail page*.
- Rows* untuk menentukan lokasi atau tempat untuk data yang akan diambil.
- Columns* untuk informasi apa saja yang ingin diambil.
- Navigator* untuk mendeteksi halaman yang akan di *scraping*.
- Actions* apabila membutuhkan suatu operasi seperti penjumlahan atau pengurangan.
- Javascript*.
- Save* untuk menyimpan dan menentukan berapa halaman yang di *scraping* (berlaku dalam *list page*).

### 3. *Parsehub*

*Parsehub* adalah aplikasi yang dapat menganalisis dan mengambil data dari situs web dan mengubahnya menjadi data yang bermakna. *Parsehub* menggunakan *machine learning* untuk mengenali dokumen yang paling rumit dan menghasilkan file *output* dalam JSON, CSV, Google Sheets, atau melalui API. Memiliki fitur canggih seperti *pagination*, *pop-up*, *navigation*, dan *scrolling pages* yang tidak

#### Hak Cipta :

- Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
- Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



terbatas. Pengguna bahkan dapat memvisualisasikan data dari *Parsehub* ke *Tableau*. Versi gratis *Parsehub* memiliki batas 5 proyek dengan 200 halaman per *scraping*. (Parsehub, 2019)

## 2.2 Text Mining

*Text mining* adalah proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu. *Text mining* memiliki tugas yang lebih kompleks karena melibatkan data teks yang sifatnya tidak terstruktur. *Text mining* merupakan bidang multidisiplin yang melibatkan analisis teks, ekstraksi informasi, *clustering*, kategorisasi, visualisasi, teknologi basis data, *machine learning*, dan *data mining* (Ayani, Pratiwi, & Muhandi, 2019). Dalam arti lain, *Text mining* adalah sebuah teknik/pendekatan algoritmik berbasis komputer untuk mendapatkan suatu pengetahuan baru yang tersembunyi dari sekumpulan teks. *Text mining* merupakan bagian dari keilmuan *information retrieval* (temu balik informasi) yang bekerja pada data bertipe teks yang cenderung tidak terstruktur. (Priyanto & Ma'arif, 2018). Dalam penelitian ini, proses *text mining* terbagi menjadi dua tahapan, yakni *text pre-processing* dan *text classification*.

### 1. *Text pre-processing*

Tahapan ini adalah tahapan yang berfungsi untuk membersihkan teks sebelum diolah lebih lanjut. Data teks mentah yang tidak terstruktur memiliki cukup banyak *noise* seperti tanda baca, angka, imbuhan, karakter-karakter khusus, *slang word*, dan lain sebagainya. Dalam tahapan ini, data teks tersebut dibersihkan sehingga tersisa bentuk dasarnya saja untuk keperluan analisis teks lebih lanjut. Ada empat tahapan *text pre-processing* yang dilakukan sebagai berikut:

1. Tokenisasi, yaitu pemotongan kalimat berdasarkan kata yang menyusunnya.
2. *Stopword removal*, yaitu tahapan membuang kata-kata yang tidak berpengaruh terhadap proses klasifikasi seperti kata depan, kata sambung, dan lain sebagainya.
3. *Case folding*, yaitu tahapan untuk menyeragamkan bentuk huruf menjadi huruf besar atau huruf kecil.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4. *Stemming* dan *Lemmatization*, yaitu proses mengubah kata ke dalam bentuk dasarnya.

2. *Text Classification*

*Text Classification* sangat dibutuhkan dalam era digitalisasi, jumlah produksi data teks saat ini sangat besar dan pertumbuhannya sangat cepat. Klasifikasi teks dapat didefinisikan sebagai proses untuk menentukan suatu dokumen teks ke dalam suatu kelas tertentu (Hidayatullah & Ma'arif, 2016). Untuk melakukan proses klasifikasi teks menggunakan metode *machine learning* yang salah satu diantaranya adalah SVM (*Support Vector Machine*). SVM (*Support Vector Machine*) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik yang dikembangkan oleh Vapnik, Boser dan Guyon pada tahun 1992 (Christianini dalam dsn, 2018). SVM merupakan salah satu metode regresi atau pengklasifikasi data berdasarkan data-data sebelumnya dan pemodelannya disupervisi terlebih dahulu. SVM termasuk kedalam jenis klasifikator yang biner, linier dan non probabilistik. SVM menggunakan *decision boundary* (batas keputusan) yang akan menentukan klasifikasi dari data-data *training*, sehingga dapat dibentuk sebuah model linier atau *hyperplane* yang paling optimal untuk mengklasifikasi data (Mutawalli, Zaen, & Bagye, 2019).

SVM adalah metode yang paling sering digunakan dalam klasifikasi karena memiliki hasil akurasi yang paling baik yang dimana bertujuan untuk mencari *hyperplane* dengan nilai jarak maksimum dengan menggunakan persamaan sebagai berikut

$$f(x) = \text{sign}(w^T x + b) \quad (1)$$

dimana:

$w^T$  = parameter bobot

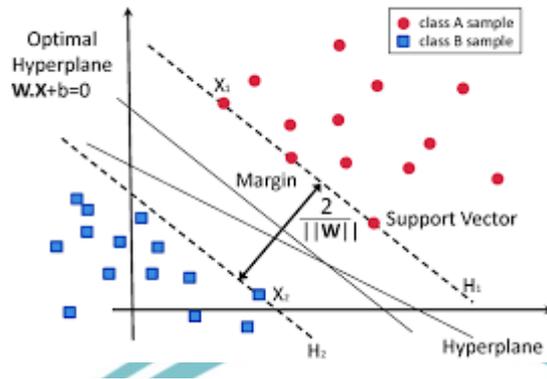
$x$  = data vektor

$b$  = parameter bias



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 2. 1 Ilustrasi svm

(Sumber: 2019, medium.com)

Pada gambar diatas merupakan ilustrasi SVM yang dimana posisi *hyperplane* berada di tengah-tengah antara dua kelas, artinya jarak antara *hyperplane* dengan objek-objek data berbeda dengan kelas yang berdekatan yang diberi tanda kotak biru dan bulat merah. Dalam SVM objek data terluar yang paling dekat dengan *hyperplane* disebut *support vector* yang paling sulit di klasifikasikan karena posisi yang hampir tumpang tindih (*overlap*) dengan kelas lain. Sehingga *support vector* inilah yang diperhitungkan untuk menemukan *hyperplane* yang paling optimal oleh SVM. Dimana  $\frac{2}{\|w\|}$  adalah margin yang merupakan jarak terdekat antara *hyperplane* dan data dari masing-masing kelas sehingga  $\frac{2}{\|w\|}$  adalah nilai jarak maksimum. Apabila diberikan data *training*  $\{(x_i, y_i) \mid x_i \in X^n, y_i \in \{-1, +1\}, i = 1, 2, 3, \dots, N\}$ , maka masalah optimasi SVM adalah

$$\min. \frac{1}{2} \|w\|^2 \quad (2)$$

dengan syarat

$$y_i [w_i^T x_i + b] \geq 1 \quad (3)$$

dimana  $y_i$  adalah kelas data ke-i dan  $x_i$  adalah vektor dari data ke-i.

Apabila terdapat toleransi kesalahan dalam klasifikasi, maka masalah optimasi SVM adalah

$$\min. \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (4)$$

dengan syarat

$$y_i[w_i^T x_i + b] \geq 1 - \xi_i \quad (5)$$

$$\xi_i \geq 0 \quad (6)$$

dimana  $\xi_i$  adalah variabel slack dan parameter  $C$  untuk mengontrol *trade off* antara margin (jarak) dan toleransi kesalahan pada klasifikasi. Persamaan (4), (5), (6) disebut SVM *soft margin* (Samsudiney, 2019).

Persamaan (4), (5), (6) diselesaikan menggunakan pemrograman kuadratik yang dimana pada penelitian ini menggunakan bantuan program *Python* dengan *Scikit-Learn package*. Dalam mengklasifikasikan data baru, SVM di uji dengan formula *hyperplane* pada persamaan (1).

Dari apa yang dinyatakan, *text mining* adalah teknik algoritmik dalam menganalisis teks untuk mengekstrak informasi yang sifatnya tidak terstruktur. Dimana penulis membagi proses *text mining* menjadi dua tahap yaitu *text pre-processing* dan *text classification* yang menggunakan metode SVM (*Support Vector Machine*).

### 2.3 Pelatihan

Telah banyak tokoh mendefinisikan pelatihan. Salah satu diantaranya dikemukakan oleh Greenberg. Greenberg mengemukakan pelatihan adalah suatu proses dimana orang – orang secara sistematis memperoleh dan meningkatkan keterampilan dan kemampuan yang dibutuhkan untuk meningkatkan kinerja mereka (Greenberg dalam Pratiwi, 2016). Definisi lain menjabarkan, Pelatihan adalah aktivitas yang didesain untuk meningkatkan performa individual orang lain pada area tertentu secara spesifik (Malcolm dalam Huda, Sulistiowati, & Maulana, 2016). Disamping itu pelatihan juga didefinisikan sebagai sebuah konsep program yang bertujuan meningkatkan pengetahuan dan keterampilan seseorang (sasaran didik), berkembang sangat pesat, dan modern. Perkembangan model pelatihan (*capacity building, empowering* dan *training*) saat ini tidak hanya terjadi pada dunia usaha, akan tetapi pada lembaga-lembaga profesional tertentu. Model pelatihan berkembang pesat sesuai dengan kebutuhan belajar, proses belajar (proses edukatif), asesmen, sasaran, dan tantangan lainnya (Kamal dalam Febriana, 2016).



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Dengan demikian, pelatihan adalah aktivitas seperti pengajaran kepada seseorang untuk meningkatkan pengetahuan, keterampilan, pengalaman yang dapat membantu seseorang mencapai tujuan yang diharapkan sesuai dengan kriteria yang diberikan.

#### 2.4 Keterampilan

Pada hakikatnya keterampilan adalah suatu ilmu yang diberikan kepada manusia, kemampuan manusia dalam mengembangkan keterampilan yang dimiliki memang tidak mudah, perlu mempelajari, perlu menggali agar lebih terampil. Keterampilan merupakan ilmu yang secara lahir ada didalam diri manusia dan perlunya dipelajari secara mendalam dengan mengembangkan keterampilan yang dimiliki. Sedangkan yang dikemukakan oleh Nadler mendefinisikan keterampilan adalah kegiatan yang memerlukan praktek atau dapat diartikan sebagai pengembangan dari aktivitas (Ummatin, 2019)

Dengan demikian, keterampilan adalah kemampuan yang dapat dipelajari, dikembangkan, dan dilatih dari aktivitas yang dilakukan sehingga dapat membuat sesuatu menjadi lebih bermakna karena hasil dari kreatifitas yang dimiliki.

#### 2.5 Website

*Website* merupakan istilah yang sudah tidak asing lagi. Secara umum *website* dapat diartikan sebagai sebuah halaman yang tersedia dalam sebuah server yang dapat diakses menggunakan jaringan internet, didalamnya berisi bermacam-macam informasi dari suatu konten tertentu. Sebuah halaman web umumnya dibuat melalui serangkaian *plain text* yang dikenal dengan istilah HTML (*Hyper Text Markup Language*) atau XHTML (*eXtensible HyperText Markup Language*) (Herliana & Rasyid, 2016). Definisi lain menjabarkan *website* adalah sekumpulan halaman yang terdiri atas beberapa laman yang berisi informasi dalam bentuk data digital, baik berupa teks, gambar, video, audio, dan animasi lainnya yang disediakan melalui jalur koneksi internet (Hanafri, Triono, & Luthfiudin, 2018).

Dengan demikian, *website* adalah halaman yang tersedia yang dapat diakses melalui koneksi internet, baik berupa teks, gambar, video, audio, dan lainnya. Didalam *website* terdapat serangkaian bahasa standar seperti HTML (*Hyper Text Markup*



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

*Language*) untuk membuat halaman web, CSS (*Cascading Style Sheet*) untuk memperindah tampilan web, dan Javascript untuk fungsionalitas halaman web.

### 2.6 Bootstrap

Bootstrap merupakan kerangka paling populer dalam desain HTML, CSS, dan Javascript untuk mengembangkan web yang responsif, dan kompatibel dengan perangkat *mobile* (Septiawan, Kridalukmana, & Windasari, 2016). Dalam kesempatan lain, Eko mendefinisikan Bootstrap merupakan salah satu *framework* HTML, CSS, dan Javascript yang digunakan untuk membuat *website* yang bersifat *responsive layout* berdasarkan ukuran *viewport* dari *device* pengaksesnya, mulai dari *smartphone*, *tablet*, maupun layar PC (Eko dalam Sanjaya & Hesinto, 2017). Dalam arti yang lain, Bootstrap adalah sebuah *framework* yang dibuat dengan menggunakan bahasa dari HTML dan CSS, namun juga menyediakan efek javascript yang dibangun dengan menggunakan *jquery*. Bootstrap telah menyediakan kumpulan komponen *class interface* dasar yang telah dirancang sedemikian rupa untuk menciptakan tampilan yang menarik, bersih dan ringan. Selain itu, bootstrap juga memiliki fitur *grid* yang berfungsi untuk mengatur *layout* yang bisa digunakan dengan sangat mudah dan cepat (Sanjaya & Hesinto, 2017).

Dengan demikian, Bootstrap adalah *framework* HTML, CSS, dan Javascript yang menyediakan *class interface* untuk mengembangkan tampilan web yang menarik dan memudahkan dalam mengatur *layout website* sehingga membuat tampilan web responsif.

### 2.7 Python

Python adalah satu dari bahasa pemrograman tingkat tinggi yang bersifat interpretatif, interaktif, *object-oriented*, dan dapat beroperasi di hampir semua platform seperti Linux, Windows, Mac, dan platform lainnya. Python adalah salah satu bahasa pemrograman tingkat tinggi yang mudah dipelajari karena sintaks yang jelas dan elegan, yang dikombinasikan dengan penggunaan modul-modul yang mempunyai struktur data tingkat tinggi, efisien, dan siap langsung digunakan (Rosmala dan Dwipa dalam Sulistiyanti, Setyawan, Sivam, & Purwiyanti, 2019). Dalam arti yang lain, Python adalah bahasa pemrograman tingkat tinggi dengan

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



sistem yang dinamis, sintaks Python yang sederhana dan mudah dipelajari berfokus pada keterbacaan kode. Python mendukung modul dan paket, yang mendorong modularitas program. (Pramono, Nurdiansyah, & Hidayatullifa, 2019).

Dari apa yang dinyatakan, Python adalah salah satu bahasa pemrograman interpretatif yang menekankan pada keterbacaan kode sehingga mudah dipelajari karena sintaks yang jelas dan mendukung penggunaan modul dan paket yang mendorong modularitas program.

## 2.8 Flask

Flask adalah salah satu *web framework* dengan menggunakan bahasa python. Flask ini termasuk dalam *microframework* karena tidak membutuhkan library tambahan maupun aplikasi tertentu. Flask mendukung ekstensi yang dapat menambahkan fitur aplikasi seolah-olah mereka diimplementasikan dalam Flask itu sendiri. Ekstensi ada untuk pemetaan obyek-relasional, validasi formulir, penanganan unggahan, berbagai teknologi *open authentication* dan beberapa alat terkait *framework* (Clinton & Sengkey, 2019). Flask adalah *web microframework* berbasis Python. Flask memiliki fungsi-fungsi yang cocok diimplementasikan untuk program yang terbatas secara *energy* dan *memory*. Meskipun termasuk dalam kategori *framework* yang ringan, flask memiliki fungsi yang bisa dikembangkan sesuai kebutuhan. Flask juga memiliki integrasi yang baik dengan database melalui SQLAlchemy. Hal ini akan mengoptimalkan aplikasi yang juga menggukan database sebagai tempat penyimpanan data (Ferdianto, Yahya, & Dewi, 2018).

Dengan demikian, Flask adalah *web microframework* berbasis Python yang ringan sehingga tidak membutuhkan *library* tambahan dan memiliki fungsi yang bisa dikembangkan sesuai kebutuhan seolah-olah diimplementasikan dalam Flask sendiri.

## 2.9 Penelitian Sejenis

Penelitian sejenis bertujuan untuk mendapatkan bahan perbandingan dan acuan. Selain itu, untuk menghindari terjadinya persamaan dengan penelitian ini. Maka dalam tinjauan pustaka ini penulis mencantumkan hasil-hasil penelitian terdahulu sebagai berikut:

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

1. Penelitian Vivensius Mitra, Herry Sujaini, dan Arif Bijakasana Putra Negara (2017)

Penelitian Vivensius Mitra, Herry Sujaini, dan Arif Bijakasana Putra Negara dengan judul “Rancang Bangun Aplikasi *Web Scraping* untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM”. Penelitian ini bertujuan untuk mengumpulkan korpus paralel Bahasa Indonesia dan Inggris (Mitra, Sujaini, & Putra Negara, 2017).

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa, dokumen korpus paralel dapat dihasilkan melalui proses *scraping* dengan metode HTML DOM dengan *website* yang diambil datanya adalah *website* berita dua bahasa. Dimana HTML DOM adalah standar untuk bagaimana mendapatkan, mengubah, menambah, atau menghapus elemen HTML. Penelitian ini menghasilkan korpus parallel Bahasa Indonesia-Inggris sebanyak 38.712 pasang korpus paralel. Hasil akhir dari penelitian ini di bentuk dalam format .txt

2. Penelitian Dhita Deviacita Ayani, Helen Sasty Pratiwi, Hafiz Muhardi (2019)

Penelitian Dhita Deviacita Ayani, Helen Sasty Pratiwi, Hafiz Muhardi dengan judul “Implementasi *Web Scraping* untuk Pengambilan Data pada Situs *Marketplace*”. Penelitian ini bertujuan untuk melakukan pencarian di beberapa situs *marketplace* dan menampilkan hasil pencarian secara bersamaan (Ayani, Pratiwi, & Muhardi, 2019).

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa, pengambilan data pada situs *marketplace* menggunakan *web scraping* dapat menampilkan hasil sesuai kebutuhan, seperti nama produk, harga, jumlah terjual, rating, dan tombol detail. Setiap pencarian produk dapat memilih pilihan *marketplace* yang disediakan. Tombol detail pada hasil pencarian produk akan masuk ke halaman web asli produk yang dipilih.



## BAB III

### PERENCANAAN DAN REALISASI

#### 3.1 Deskripsi Program Aplikasi

Sistem yang akan dibangun adalah implementasi *text mining* dan *web scraping* untuk mengumpulkan informasi pada *website* pelatihan keterampilan khusus perempuan.

Dimana proses awal pembuatan yang dilakukan adalah pencarian *website* yang membahas tentang informasi pelatihan keterampilan. Setelah menemukan *website* yang telah ditentukan akan di lakukan proses *text mining* yang terdiri atas dua tahapan yaitu *text pre-processing* dengan melalui tahapan pemotongan kalimat berdasarkan kata yang menyusunnya disebut juga tahap tokenisasi, membuang kata-kata yang tidak berpengaruh terhadap proses klasifikasi, menyeragamkan bentuk huruf menjadi huruf kecil, dan proses mengubah kata ke dalam bentuk dasarnya. Setelah tahap *text pre-processing* selesai dilanjutkan dengan tahapan *text classification* untuk menentukan suatu *website* masuk ke dalam kelas tertentu. Dimana dalam proses *text classification* ini menggunakan metode SVM (*Support Vector Machine*) untuk mendapatkan nilai akurasi sehingga *website* mengenali URL yang akan dimasukkan, masuk ke dalam *website* yang membahas tentang informasi pelatihan keterampilan atau bukan. Jika ya, akan diambil informasinya menggunakan *web scraping*. Dalam menggunakan *web scraping* terlebih dahulu harus mengetahui struktur HTML dari *website* untuk menentukan *tag* HTML yang mana informasi seperti judul, tanggal, deskripsi akan diambil. Setelah sistem mengenali bahwa url yang di masukkan adalah *website* yang membahas tentang informasi pelatihan keterampilan, sistem akan menampilkan dan menyimpan hasil *scraping* dan admin dapat mengklik *button* edit untuk menginput data yang kosong dan menentukan kategori yang kemudian admin dapat menyimpan pembaruan data.

Hasil dari implementasi *text mining* dan *web scraping* akan disimpan ke *database* secara otomatis. Dalam *website* pelatihan keterampilan khusus perempuan, pengguna dapat melakukan pencarian informasi, memilih kategori pelatihan, dan melihat detail informasi pelatihan.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

### 3.1.1 Analisis Kebutuhan

Pada tahap ini, menganalisis untuk menspesifikasikan kebutuhan baik dari kebutuhan fungsional dan kebutuhan pengguna pada *website* pelatihan keterampilan khusus perempuan.

#### 1. Analisis Pengguna

Analisis pengguna untuk menentukan aktor yang terlibat dalam sistem yang akan dibangun. Terdapat dua aktor yang terlibat dalam *website* pelatihan keterampilan khusus perempuan, yaitu Pengguna dan Admin. Pengguna dapat melakukan pencarian informasi, memilih kategori pelatihan, dan melihat detail informasi pelatihan. Admin dapat melakukan *login* dan mengelola data informasi.

#### 2. Analisis Kebutuhan Fungsional dan Non-Fungsional

Analisis kebutuhan fungsional dan non-fungsional merupakan analisis yang dibutuhkan untuk menentukan spesifikasi dan proses pada sistem. Berikut adalah kebutuhan fungsional dan non-fungsional dari *website* pelatihan keterampilan khusus perempuan:

##### a. Kebutuhan Fungsional

1. Sistem hanya diperuntukkan oleh kaum perempuan.
2. Admin dapat melakukan *login* sesuai dengan hak akses pada sistem.
3. Sistem menentukan *URL* yang dimasukkan adalah mengenai pelatihan keterampilan atau bukan.
4. Sistem dapat mengambil informasi judul, tanggal, dan deskripsi pada *website* mengenai pelatihan keterampilan.
5. Admin dapat mengelola data informasi *website* yang diambil informasinya.
6. Pengguna dapat melakukan pencarian pelatihan
7. Pengguna dapat memilih kategori pelatihan.
8. Pengguna dapat melihat detail informasi pelatihan.

##### b. Kebutuhan Non-Fungsional

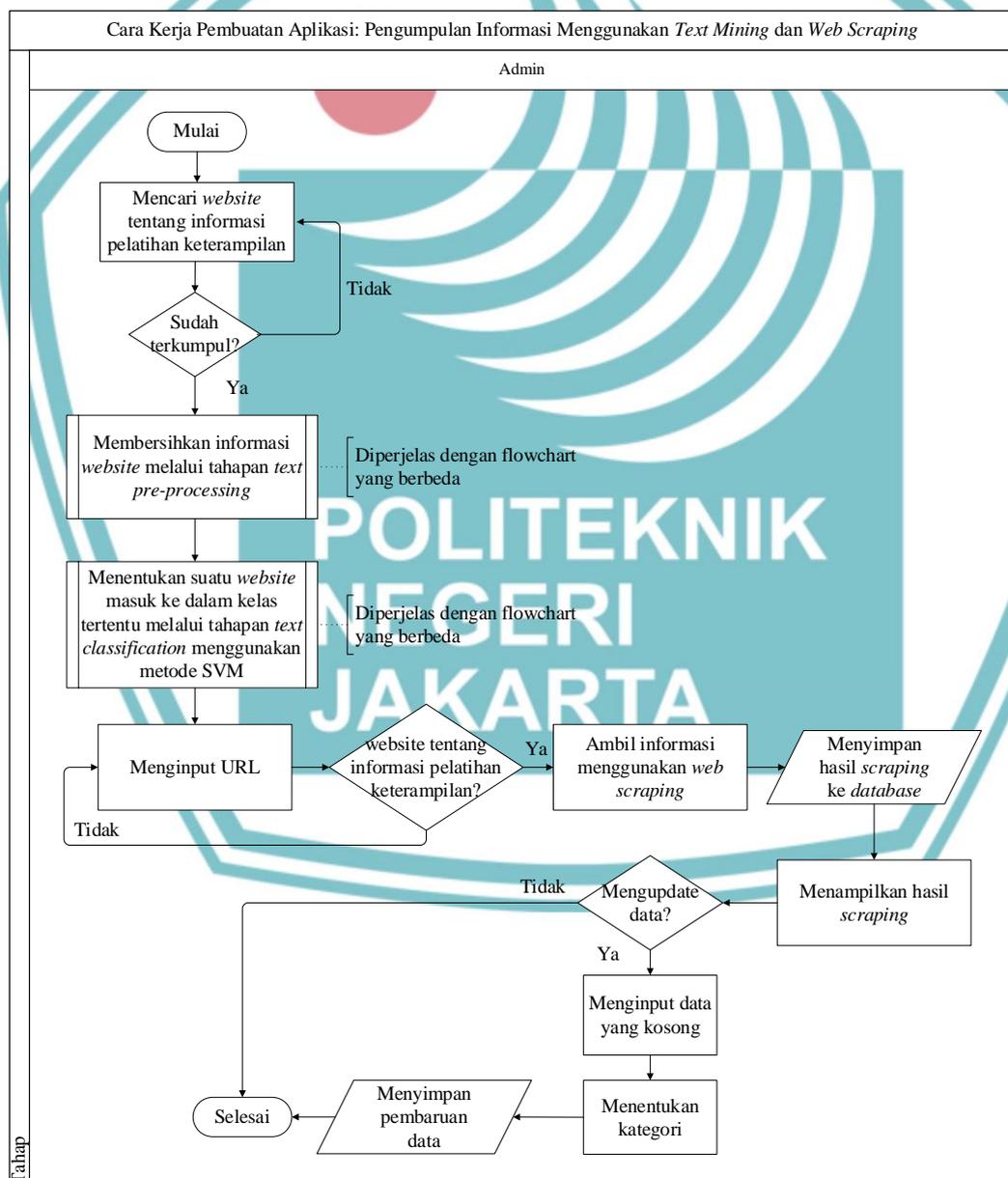
1. Sistem dapat diakses menggunakan *browser* seperti Google Chrome, Internet Explorer, dan Mozilla Firefox.
2. Sistem menggunakan *database* MySQL



### 3.2 Cara Kerja Pembuatan Aplikasi

Untuk alur pembuatan aplikasi dapat diketahui dengan *swimlane* diagram. Berikut adalah alur pembuatan sistem dalam implementasi teknologi *text mining* dan *web scraping* yang dimana dalam alur pembuatan sistem akan terbagi menjadi satu diagram secara keseluruhan dan dua diagram untuk menjelaskan secara detail masing-masing modul dalam pengumpulan informasi menggunakan teknologi *text mining* dan *web scraping*.

#### 1. Alur pembuatan sistem secara keseluruhan



Gambar 3. 1 Flowchart pengumpulan informasi menggunakan text mining dan web scraping

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

*Flowchart* pada gambar 3.1 adalah proses dimana pengumpulan informasi menggunakan *text mining* dan *web scraping* menggunakan *swim lane* diagram yang dimana admin akan mencari *website* terlebih dahulu tentang informasi pelatihan yang akan diambil informasinya. Apabila sudah terkumpul, admin akan membersihkan informasi *website* melalui tahapan *text pre-processing* dan dilanjutkan dengan menentukan suatu *website* masuk ke dalam kelas tertentu melalui tahapan *text classification* menggunakan metode SVM (*Support Vector Machine*). Setelah semua modul telah selesai, admin menginput URL untuk mengenali URL yang dimasukkan, masuk ke dalam *website* tentang informasi pelatihan keterampilan atau bukan. Jika ya, sistem akan mengambil dan menyimpan informasi *website* menggunakan *web scraping*, menampilkan hasil *scraping*, dan admin dapat mengupdate data seperti menginput data yang kosong dan menentukan kategori yang kemudian admin dapat menyimpan pembaruan data.

2. Alur pembuatan sistem pada modul ‘Membersihkan informasi *website* melalui tahapan *text pre-processing*’



Gambar 3. 2 Flowchart modul ‘Membersihkan informasi website melalui tahapan text pre-processing’

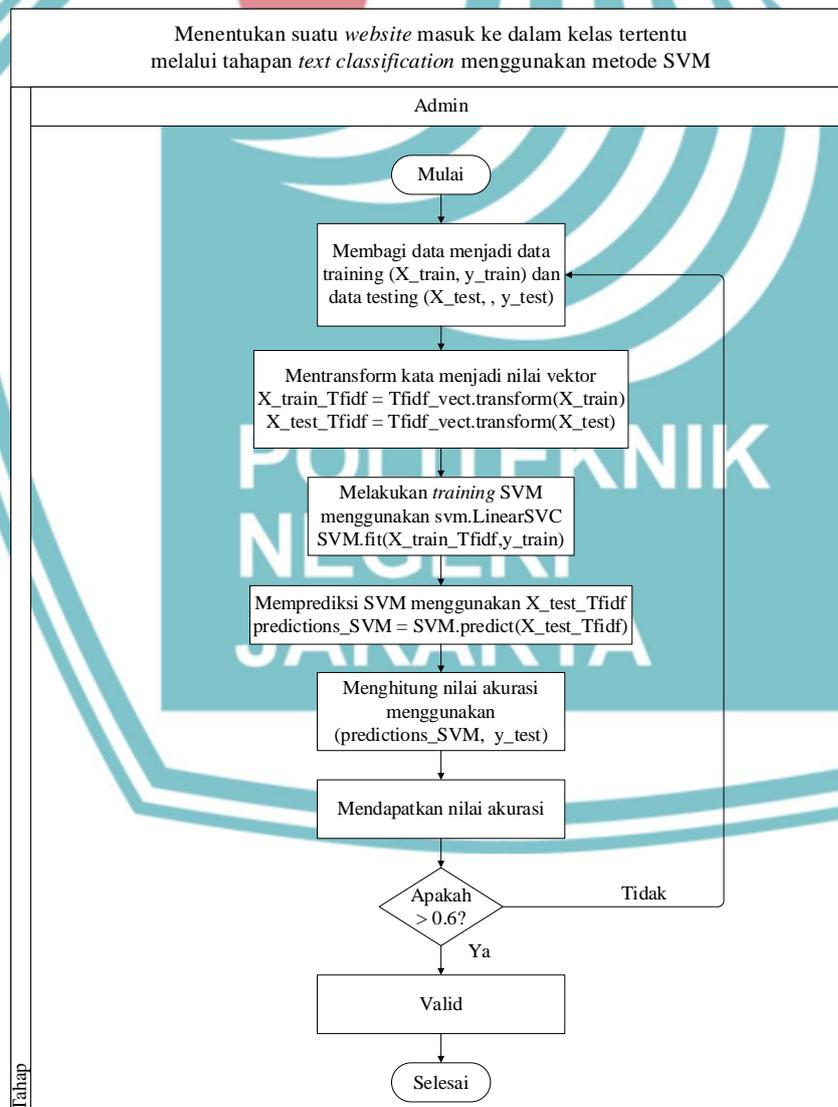
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Flowchart modul ‘membersihkan informasi *website* melalui tahapan *text pre-processing*’ dimana informasi akan di bersihkan dengan langkah awal mempersiapkan teks dokumen *website* yang telah terkumpul, kemudian adanya pemotongan kalimat berdasarkan kata yang menyusunnya disebut juga tahap tokenisasi, setelah itu membuang kata-kata yang tidak berpengaruh terhadap proses klasifikasi seperti kata sambung, kata depan, dan lain-lain, berikutnya adalah menyeragamkan bentuk huruf menjadi huruf kecil, dan terakhir mengubah kata ke dalam bentuk dasarnya.

3. Alur pembuatan sistem pada modul ‘Menentukan suatu *website* masuk ke dalam kelas tertentu melalui tahapan *text classification* menggunakan metode SVM’



Gambar 3. 3 Flowchart modul ‘Menentukan suatu website masuk ke dalam kelas tertentu melalui tahapan *text classification* menggunakan metode SVM’

#### Hak Cipta :

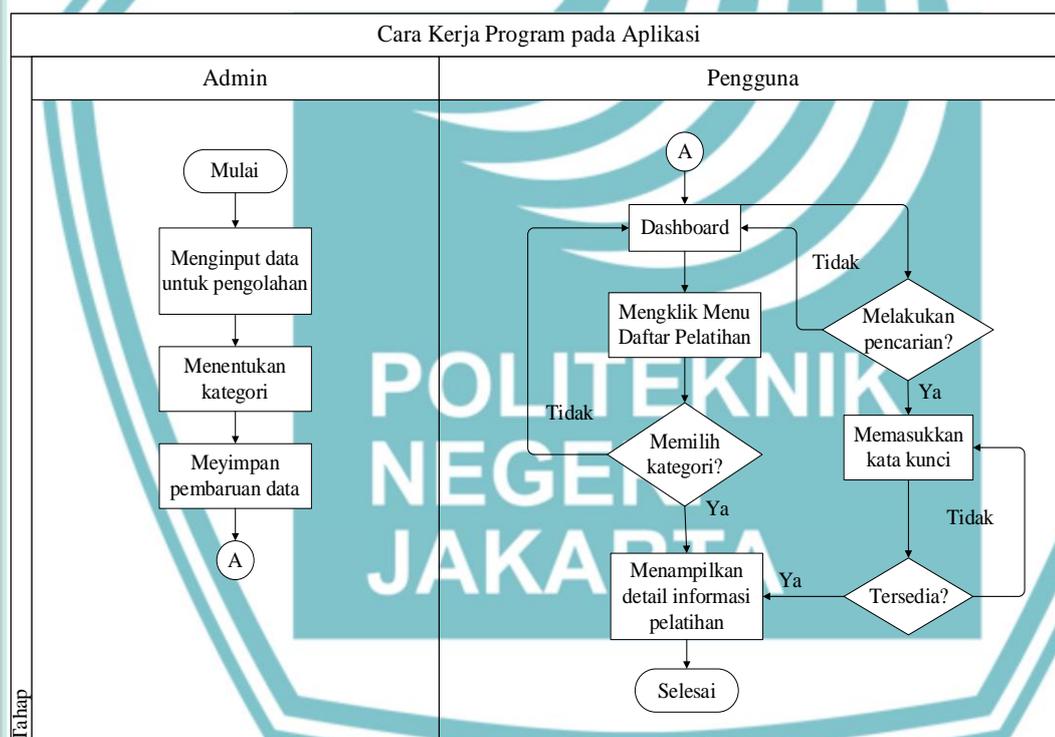
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta


**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Pada *flowchart* modul ‘menentukan suatu *website* masuk ke dalam kelas tertentu melalui tahapan *text classification* menggunakan metode SVM’ dimana setelah membersihkan informasi melalui *text pre-processing* bagi data menjadi data *training* (X\_train, y\_train) dan data *testing* (X\_test y\_test). Dikarenakan kata masih dalam berbentuk teks maka perlu mentransform menjadi angka berupa nilai vektor. Setelah itu, dilakukan proses *training* SVM menggunakan svm.LinearSVC yang mampu melakukan klasifikasi biner dan multi-*class* pada dataset dan memprediksi SVM menggunakan X\_test yang sudah di transform. Kemudian, hitung nilai akurasi menggunakan prediksi SVM dan y\_test. Apabila nilai akurasi yang didapatkan > 0,6 maka nilai akurasi valid dan jika tidak valid, maka ulangi proses.

### 3.3 Cara Kerja Program Aplikasi



Gambar 3. 4 Flowchart program pada website pelatihan keterampilan khusus perempuan

Pada *flowchart* diatas, admin menginput data untuk pengolahan, menentukan kategori, dan menyimpan pembaruan data. Sehingga pada saat ditampilkan di *website* pelatihan keterampilan khusus perempuan, pengguna dapat melihat halaman *dashboard* yang dimana pengguna dapat klik menu ‘Daftar Pelatihan’ dan memilih kategori yang akan menampilkan detail informasi pelatihan atas kategori yang dipilih. Apabila pengguna ingin melakukan pencarian, pengguna terlebih



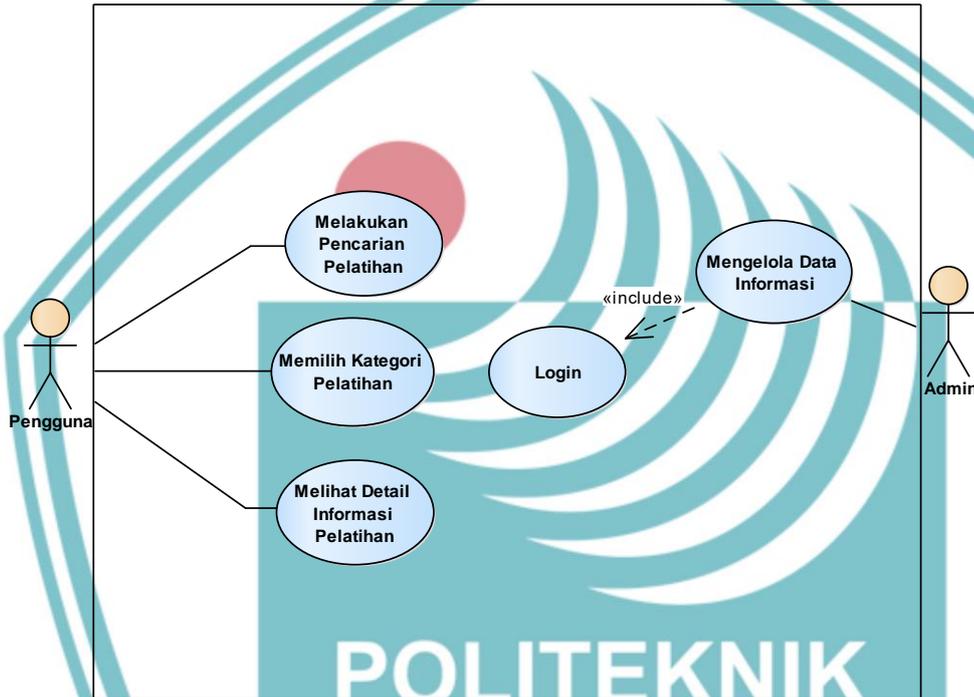
**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dahulu memasukkan kata kunci yang ingin dicari, apabila tersedia maka akan menampilkan detail informasi pelatihan, apabila tidak tersedia maka akan muncul *alert* untuk pengguna dapat memasukkan kata kunci yang lain.

### 3.4 Rancangan Program Aplikasi

#### 3.4.1 Use Case Diagram



Gambar 3. 5 Use case diagram pada website pelatihan keterampilan khusus perempuan

Diagram use case pada *website* pelatihan keterampilan khusus perempuan dapat dilihat pada gambar diatas. Dimana sistem ini melibatkan dua aktor pengguna sistem, yaitu Pengguna dan Admin.

1. Pengguna adalah aktor yang dapat melihat informasi pelatihan keterampilan hasil implementasi *text mining* dan *web scraping*.
  - a. *Use case* ‘Melakukan pencarian pelatihan’, pengguna dapat melakukan pencarian pelatihan berdasarkan kata kunci yang diketahuinya. Apabila kata kunci tersedia, sistem akan menampilkan daftar informasi pelatihan yang tersedia dan pengguna dapat melihat informasi pelatihan serta dapat melihat detail dari pelatihan yang dipilih. Apabila kata kunci tidak tersedia, sistem



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

akan menampilkan *alert* untuk pengguna dapat mencari kata kunci yang lain.

- b. *Use case* ‘Memilih kategori pelatihan’, pengguna dapat memilih kategori pelatihan yang tersedia, ketika pengguna memilih kategori pelatihan yang diinginkan akan menampilkan daftar informasi pelatihan yang tersedia dan pengguna dapat melihat informasi pelatihan serta dapat melihat detail dari pelatihan yang dipilih.
  - c. *Use case* ‘Melihat detail pelatihan’, pengguna dapat melihat detail informasi pelatihan yang dipilih, dimana pengguna dapat menemukan informasi yang diinginkan dari informasi pelatihan yang dipilih.
2. Admin adalah aktor yang mengelola informasi hasil implementasi *web scraping* dan *text mining*.
    - a. *Use case* ‘Login’ dengan menginput *username* dan *password* yang sudah di *setting* admin. Apabila *username* benar dan *password* salah maka admin gagal dalam memasuki laman web, apabila *username* salah dan *password* benar maka admin gagal dalam memasuki laman web, apabila keduanya salah maka admin gagal dalam memasuki laman web, dan apabila keduanya benar maka admin berhasil dalam memasuki laman web.
    - b. *Use case* ‘Mengelola data informasi’, admin mengelola data informasi apakah informasi yang ada pada *website* sudah sesuai dengan implementasi *web scraping* dan *text mining* atau tidak. Apabila belum sesuai maka admin dapat memperbaikinya.

### 3.4.2 Activity Diagram

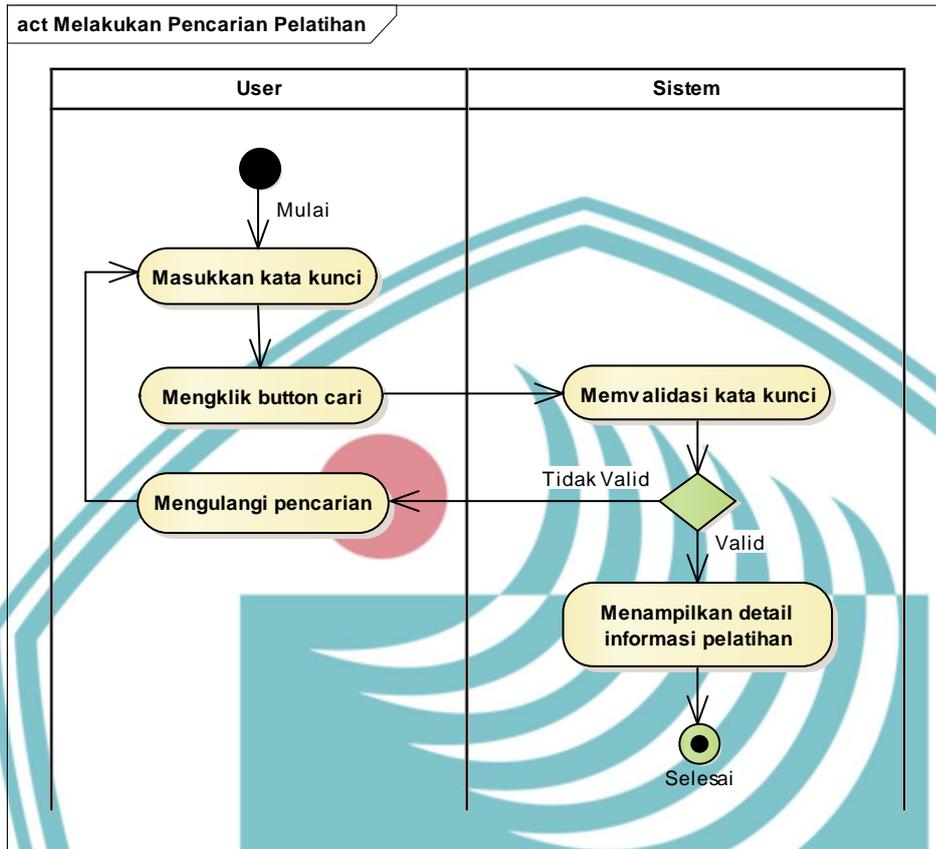
Berikut ini akan menjelaskan mengenai semua *activity* diagram pada sistem yang di buat.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

1. Diagram *Activity* Mencari Pelatihan



Gambar 3. 6 Activity diagram mencari pelatihan

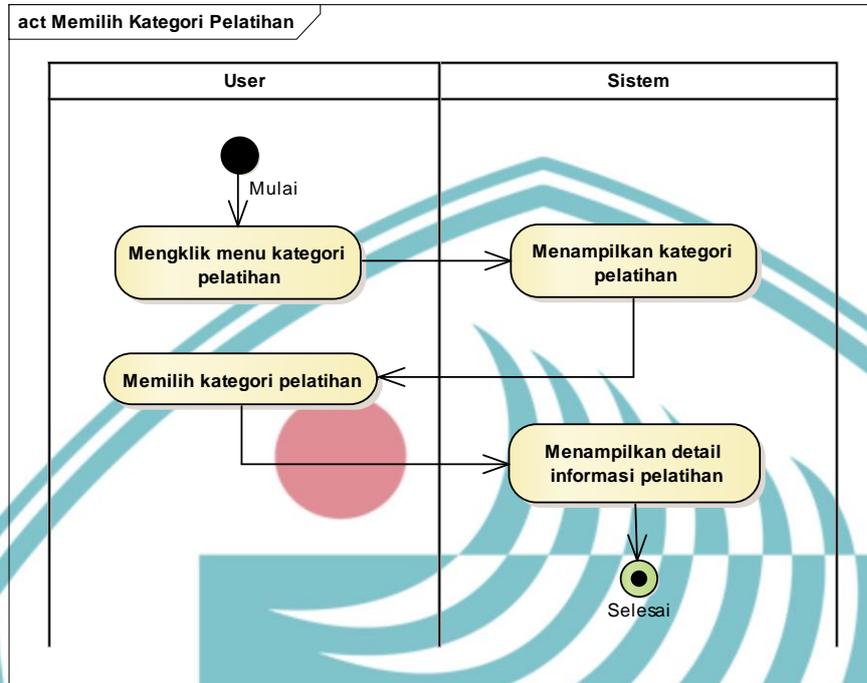
Pada gambar di atas merupakan aktivitas mencari pelatihan yang dilakukan oleh *user*. *User* akan memasukkan kata kunci yang diketahuinya dan mengklik *button* cari dan sistem akan melakukan validasi apakah kata kunci yang di masukkan valid atau tidak. Jika valid maka sistem akan menampilkan detail informasi pelatihan dan jika tidak valid maka *user* akan mengulangi pencarian dengan memasukkan kembali kata kunci.

2. Diagram *Activity* Memilih Kategori Pelatihan

Pada gambar 3.7 di bawah ini merupakan aktivitas memilih kategori pelatihan yang dilakukan oleh *user*. Aktivitas ini dimulai dengan *user* mengklik menu kategori pelatihan dan sistem akan menampilkan kategori pelatihan yang tersedia. Setelah itu, *user* akan memilih kategori pelatihan yang di inginkan dan sistem akan

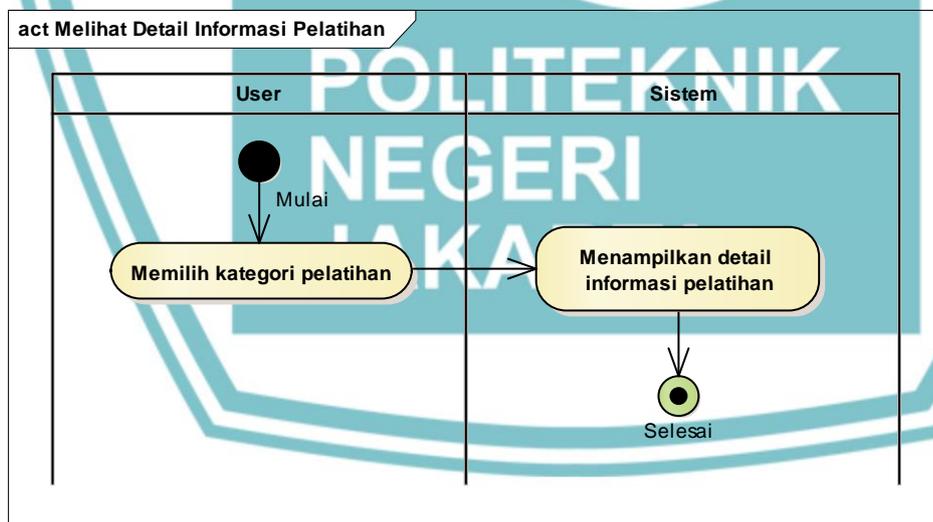


menampilkan detail informasi pelatihan dari kategori pelatihan yang dipilih oleh *user*.



Gambar 3. 7 Activity diagram memilih kategori pelatihan

### 3. Diagram *Activity* Melihat Detail Informasi Pelatihan



Gambar 3. 8 Activity diagram melihat detail informasi pelatihan

Pada gambar 3.8 merupakan aktivitas melihat detail informasi pelatihan yang dilakukan oleh *user*. Dimana aktivitas ini dimulai dengan *user* yang memilih kategori pelatihan dan sistem akan menampilkan detail informasi pelatihan sesuai dengan kategori yang dipilih oleh *user*.

#### Hak Cipta :

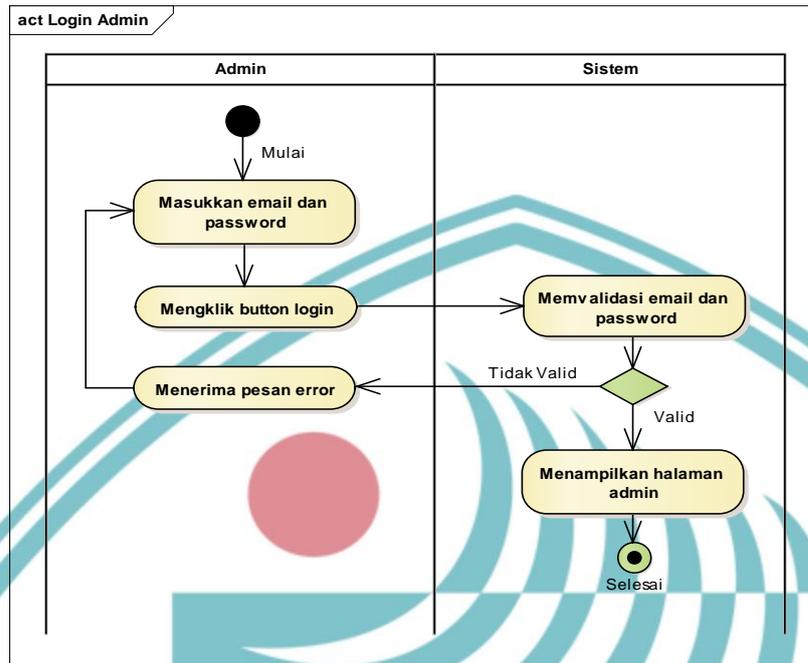
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4. Diagram *Activity Login Admin*



Gambar 3. 9 Activity diagram login admin

Pada gambar 3.9 merupakan aktivitas login yang dilakukan oleh admin. Admin akan diminta untuk memasukkan email dan password untuk melakukan login dan setelah admin mengklik button login, sistem akan melakukan validasi apakah email dan password yang dimasukkan sudah benar. Jika benar maka admin akan masuk ke halaman admin dan jika salah maka admin menerima pesan error dan diminta untuk memasukkan kembali email dan password.

5. Diagram *Activity Mengelola Data Informasi*

Gambar 3.10 merupakan aktivitas proses kelola data informasi yang dilakukan oleh admin. Di awali dengan admin memasukkan url dan mengklik *button get*. Kemudian sistem akan memvalidasi url yang di masukkan oleh admin. Jika tidak valid, admin akan menerima pesan error dan kembali memasukkan url. Jika valid sistem akan menampilkan elemen html dan admin dapat mengklik *button scrape* dan simpan. Setelah itu, sistem akan menampilkan dan menyimpan hasil *scraping*.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Dimana admin dapat mengupdate data seperti menginput data yang kosong dan menentukan kategori yang kemudian admin menyimpan pembaruan data.



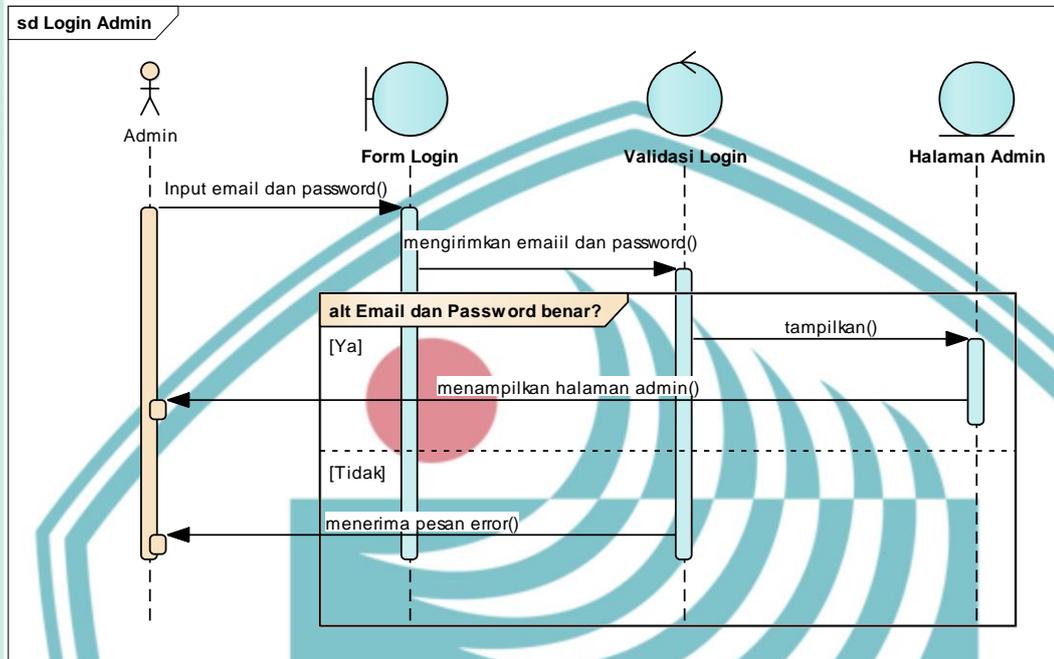
Gambar 3. 10 Activity diagram mengelola data informasi

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### 3.4.3 Sequence Diagram

#### 1. Sequence Diagram Login Admin



Gambar 3. 11 Sequence diagram login admin

Gambar 3.11 merupakan *sequence* diagram dari admin ketika melakukan login ke sistem. Dimana admin menginput *email* dan *password* yang sudah terdaftar. Sistem akan mevalidasi *login* admin. Jika data valid, maka langsung menampilkan halaman admin dan jika data tidak valid, admin akan menerima pesan *error*.

#### 2. Sequence Diagram Mengelola Data Informasi

Pada gambar 3.12 merupakan *sequence* diagram dari admin ketika mengelola data informasi. Dimana admin memasukkan url yang ingin diambil informasinya. Sistem akan melakukan validasi url apakah url tersebut *website* mengenai pelatihan keterampilan atau bukan. Jika ya, maka akan melakukan proses *scraping* yang kemudian akan disimpan dan menampilkan hasil *scraping*. Setelah itu, admin dapat melakukan *update* data seperti menginput data yang kosong dan menentukan kategori. Sistem akan melakukan proses pembaruan data dan data berhasil diperbarui ketika kembali pada halaman admin. Jika tidak, admin menerima pesan *error* bahwa url yang dimasukkan bukan *website* mengenai pelatihan keterampilan.

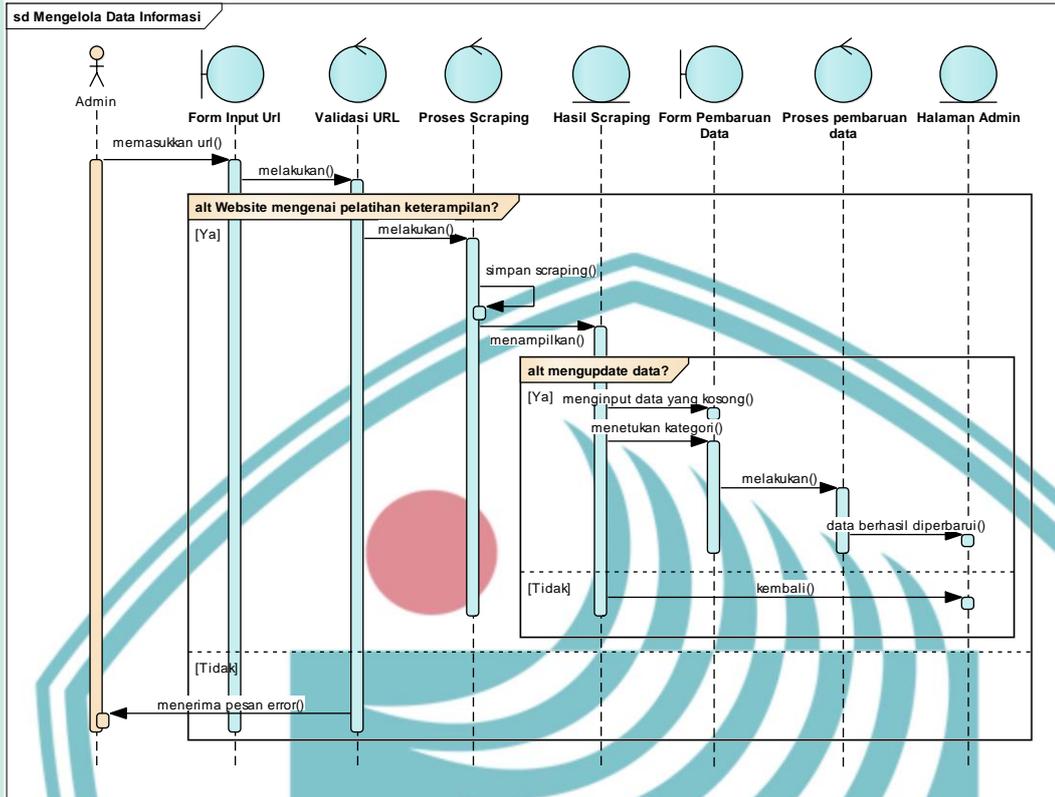
#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



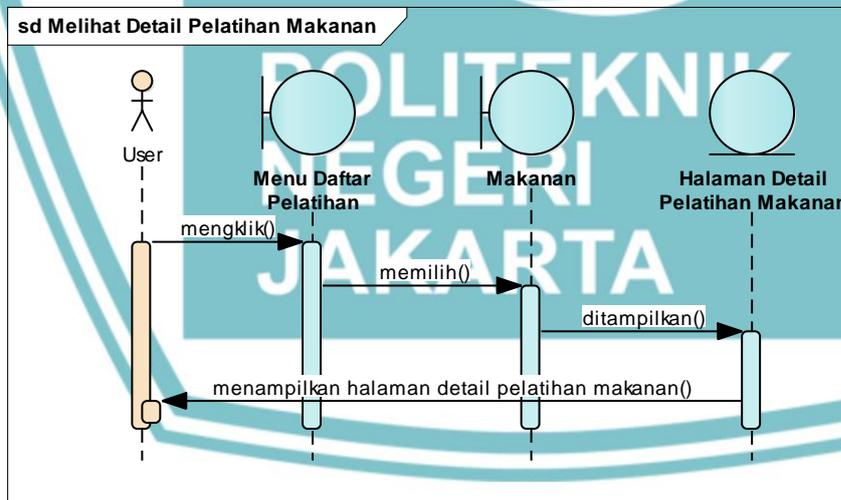
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 3. 12 Sequence diagram mengelola data informasi

3. Sequence Diagram Melihat Detail Pelatihan Makanan

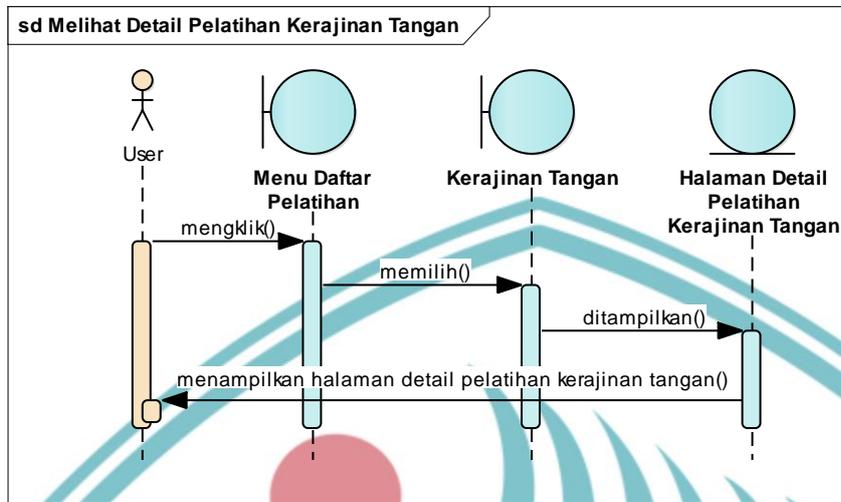


Gambar 3. 13 Sequence diagram melihat detail pelatihan makanan

Gambar 3.13 adalah *sequence* diagram dari *user* ketika melihat detail pelatihan makanan. Dimana *user* mengklik menu daftar pelatihan dan memilih kategori makanan setelah itu *user* akan ditampilkan halaman detail pelatihan keterampilan kategori makanan.



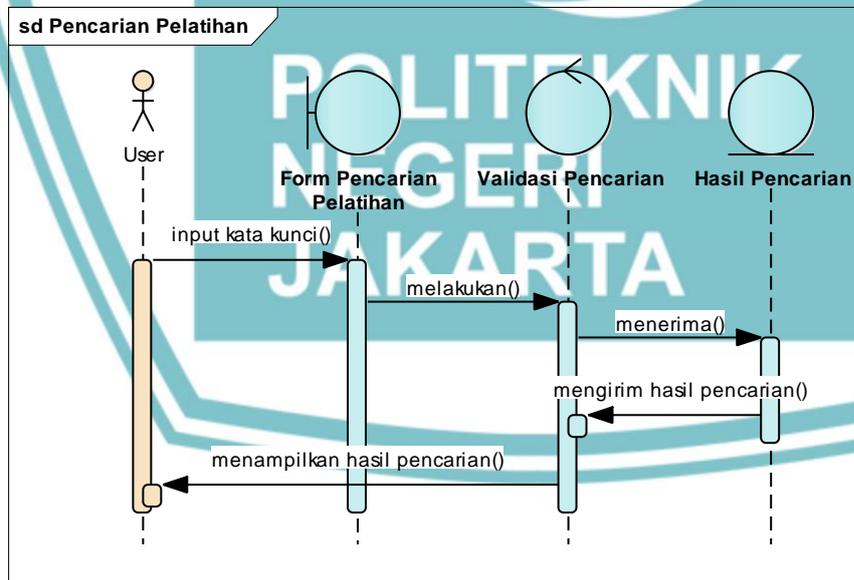
#### 4. Sequence Diagram Melihat Detail Pelatihan Kerajinan Tangan



Gambar 3. 14 Sequence diagram melihat detail pelatihan kerajinan tangan

Gambar 3.14 adalah *sequence* diagram dari *user* ketika melihat detail pelatihan kerajinan tangan. Dimana *user* mengklik menu daftar pelatihan dan memilih kategori kerajinan tangan setelah itu *user* akan ditampilkan halaman detail pelatihan keterampilan kategori kerajinan tangan.

#### 5. Sequence Diagram Pencarian Pelatihan



Gambar 3. 15 Sequence diagram pencarian pelatihan

Gambar 3.15 adalah *sequence* diagram dari *user* ketika melakukan pencarian pelatihan Dimana *user* menginput kata kunci pada form pencarian yang kemudian

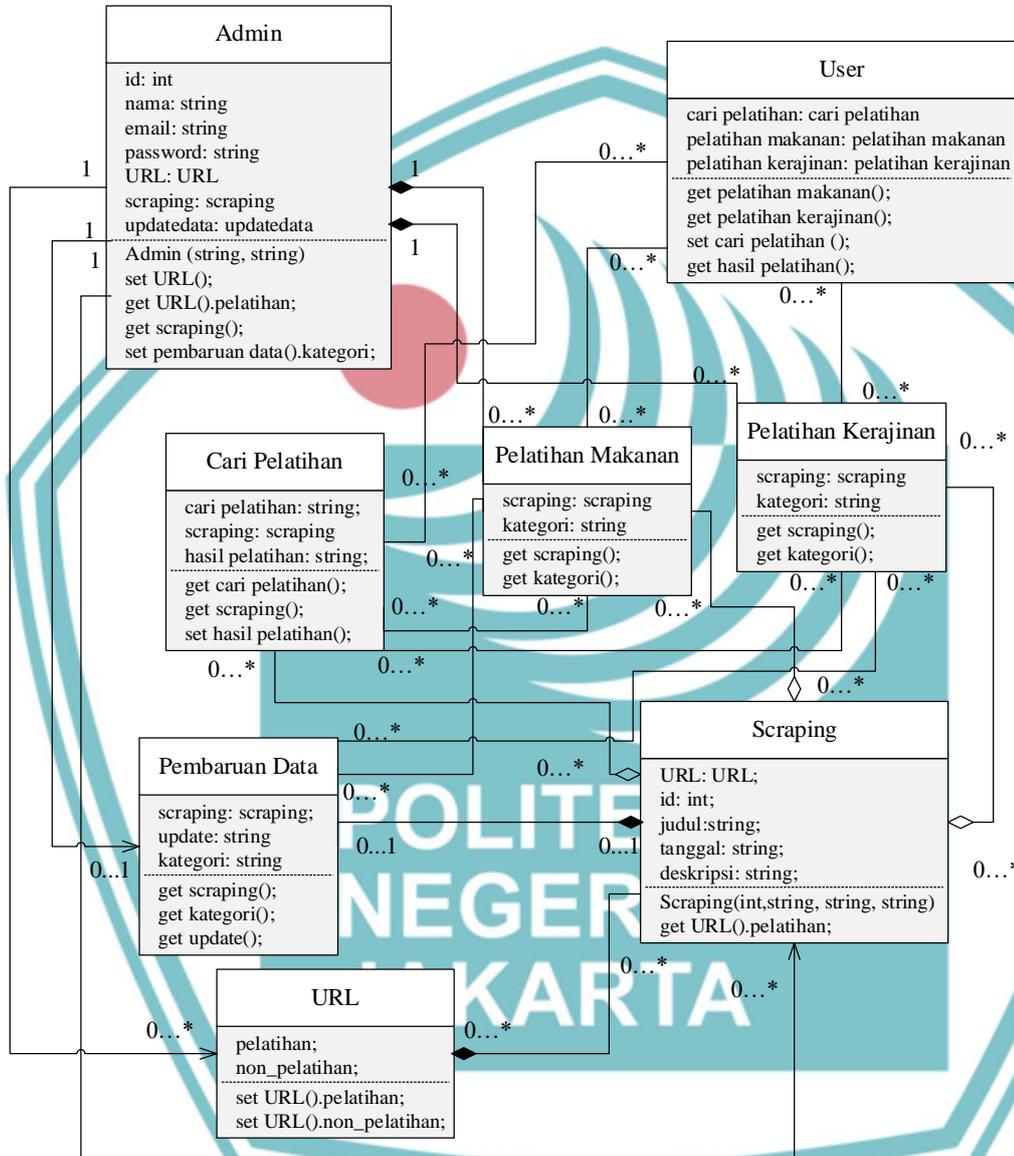
#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



sistem akan mevalidasi kata kunci yang diinput dan mengirim hasil pencarian yang diinput oleh *user* dan menampilkan hasil pencarian pada *user*.

3.4.4 Class Diagram



Gambar 3. 16 Class diagram aplikasi

Gambar 3.16 merupakan *class* diagram pada aplikasi *website* pelatihan keterampilan khusus perempuan. Pada bagian ini, *class* diagram yang dipaparkan terkait dengan hubungan pada setiap *class*, yaitu *class* Admin, *class* User, *class* URL, *class* Scraping, *class* Pembaruan Data, *class* Pelatihan Makanan, *class* Pelatihan Kerajinan, dan *class* Cari Pelatihan.

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### 3.4.5 Desain Tabel

Berdasarkan *class* diagram yang telah dibuat, maka adanya desain tabel yang sesuai dengan yang berada di dalam *database*. Desain tabel ini untuk memudahkan admin dalam mengetahui data apa saja yang disimpan. Berikut adalah desain tabel pada *website* pelatihan keterampilan khusus perempuan.

#### 1. Tabel Admin

Pada desain tabel Admin, id merupakan *primary key* dari tabel admin yang dimana menggunakan *auto\_increment* untuk secara otomatis id pada tabel admin dapat menambahkan nilainya sendiri tanpa perlu diinput.

Tabel 1. Desain Tabel Admin

No	Nama	Tipe Data	Length	Keterangan	Extra
1.	id	int	11	<i>Primary Key</i>	auto_increment
2.	name	varchar	100	Not Null	
3.	email	varchar	100	Not Null	
4.	password	varchar	100	Not Null	

#### 2. Tabel Articles

Pada desain tabel Articles, id merupakan *primary key* dari tabel articles yang dimana menggunakan *auto\_increment* untuk secara otomatis id pada tabel articles dapat menambahkan nilainya sendiri tanpa perlu diinput. Kategori merupakan kolom ketika data yang ditambahkan apakah masuk ke dalam kategori makanan atau kerajinan tangan pada saat pembaruan data (*update*).

Tabel 2. Desain Tabel Articles

No	Nama	Tipe Data	Length	Keterangan	Extra
1.	id	int	11	<i>Primary Key</i>	auto_increment
2.	title	varchar	300	Null	
3.	date	varchar	100	Null	
4.	description	varchar	5000	Null	
5.	kategori	varchar	50	Null	

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

### 3.4.6 Mockup Aplikasi

#### 1. Halaman Login Admin

Gambar 3. 17 Mock up login admin

Gambar 3.17 merupakan tampilan *login* admin yang berisikan *form input email* dan *password*. Pada tampilan ini admin memasukkan *email* dan *password* lalu mengklik tombol *submit*.

#### 2. Halaman Admin

Gambar 3. 18 Mock up halaman admin

Gambar 3.18 merupakan tampilan halaman admin, dimana pada halaman ini berisikan *form input url* dan *button scrape*.

#### 3. Halaman Utama

Gambar 3.19 merupakan tampilan halaman utama pada *website* pelatihan keterampilan khusus perempuan. Pada halaman ini berisikan *slider* yang memberikan informasi singkat mengenai *website*, *form input* dan tombol pencarian, dan dua *card* mengenai kategori pelatihan.



CARI BERDASARKAN KATEGORI



Makanan

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

[Daftar Pelathan](#)



Kerajinan Tangan

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

[Daftar Pelathan](#)

Gambar 3. 19 Mock up halaman utama

4. Halaman Detail Informasi Pelatihan

**Masakan Nusantara**

22/012020

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas sem lectus, molestie ut eleifend eu, tempus ac urna. Morbi pretium nunc at augue tincidunt varius. .

**Produk Olahan Makanan**

13/10/2019

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas sem lectus, molestie ut eleifend eu, tempus ac urna. Morbi pretium nunc at augue tincidunt varius. .

**Pelatihan Keterampilan Menjahit**

06/07/2019

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas sem lectus, molestie ut eleifend eu, tempus ac urna. Morbi pretium nunc at augue tincidunt varius. .

Gambar 3. 20 Mock up detail informasi pelatihan

Gambar 3.20 merupakan tampilan halaman detail informasi pelatihan yang berisikan judul, tanggal, dan deskripsi dari informasi pelatihan yang tersedia.

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritikan atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### 3.4.7 Realisasi Program Aplikasi

Realisasi program aplikasi dijelaskan menggunakan implementasi kode yang merupakan realisasi keseluruhan desain sistem yang di susun dalam serangkaian program. Implementasi kode *website* pelatihan keterampilan khusus perempuan menggunakan bahasa pemrograman Python dengan *framework* Flask sebagai *back-end system* dan HTML, CSS, Javascript, dan *framework* Bootstrap sebagai pembuatan *website* bagian *front-end*.

#### 1. Fungsi Login

```
@app.route('/login', methods=["GET", "POST"])
def login():

    if request.method == 'POST':
        email = request.form['email']
        password = request.form['password'].encode('utf-8')

        cur = mysql.connection.cursor(MySQLdb.cursors.DictCursor)
        cur.execute("SELECT * FROM admin WHERE email=%s", (email,))
        admin = cur.fetchone()
        cur.close()

        if admin:
            if bcrypt.hashpw(password,
admin['password'].encode('utf-8')) == admin['password'].encode('utf-8'):

                session['logged_in'] = True
                session['id'] = admin['id']
                session['email'] = admin['email']
                # flash("Selamat Datang Admin!", 'success')
                return render_template('home_admin.html')
            else:
                flash("Email atau Password Salah!", 'danger')
                return render_template('login.html')
        else:
            flash("Email atau Password Salah!", 'danger')
            return render_template('login.html')
    else:
        # flash("Anda Belum Login!", 'info')
        return render_template('login.html')
```

Pada kode diatas dimana sistem mengirimkan *email* dan *password* yang telah di masukkan oleh admin dan akan memeriksa apakah data yang dimasukkan sesuai dengan yang ada di *database* jika benar maka sistem akan menampilkan halaman admin (*home\_admin.html*) dan jika email atau password salah akan menerima pesan *error* 'Email atau Password Salah!'.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 3. 21 Tampilan login admin

Pada halaman login, admin mengisi *email* dan *password* yang sudah terdaftar di *database* lalu klik tombol *submit* untuk dapat masuk ke halaman admin.

## 2. Fungsi *Text Mining*

### 1. Fungsi *Text Pre-Processing*

#### a. Tokenisasi

```
import pandas as pd
import numpy as np
import json

data = pd.read_json("kategori_dataset.json")
data.head()

import nltk
nltk.download('punkt')

from nltk import sent_tokenize
data["short_desc_tokenized"] = data.short_desc.apply(sent_tokenize)

from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')
data["short_desc_tokenized"] = data.short_desc_tokenized.apply(lambda x: [tokenizer.tokenize(split_kalimat) for split_kalimat in x])
data

data["short_desc_tokenized"] = data.short_desc_tokenized.apply(lambda x: [[split_word.lower() for split_word in kalimat] for kalimat in x])
data
```


**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```
import pandas as pd
import numpy as np
import json

data = pd.read_json("kategori_dataset.json")

data.head()
```

	kategori	link	judul	short_desc
0	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Keterampilan Produk Olahan Makanan	pelaksanaan pelatihan keterampilan produk olah...
1	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Ekonomi Produktif Bagi Perempuan	pelatihan pemberdayaan dan peningkatan keteram...
2	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Pembuatan Manggleg	pelatihan peningkatan kualitas olahan pangan
3	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Peningkatan Kualitas Produk Olahan	pelatihan peningkatan kualitas produk olahan m...
4	pelatihan	https://sulselprov.go.id/welcome/post/1000-per...	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...

```
import nltk

nltk.download('punkt')

[nltk_data] Downloading package punkt to C:\Users\Wisya
[nltk_data] Yunita\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!

True

from nltk import sent_tokenize
data["short_desc_tokenized"] = data.short_desc.apply(sent_tokenize)

from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')
data["short_desc_tokenized"] = data.short_desc_tokenized.apply(lambda x: [tokenizer.tokenize(split_kalimat) for split_kalimat in x])
```

	kategori	link	judul	short_desc	short_desc_tokenized
0	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Keterampilan Produk Olahan Makanan	pelaksanaan pelatihan keterampilan produk olah...	[[pelaksanaan, pelatihan, keterampilan, produk...
1	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Ekonomi Produktif Bagi Perempuan	pelatihan pemberdayaan dan peningkatan keteram...	[[pelatihan, pemberdayaan, dan, peningkatan, k...
2	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Pembuatan Manggleg	pelatihan peningkatan kualitas olahan pangan	[[pelatihan, peningkatan, kualitas, olahan, pa...
3	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Peningkatan Kualitas Produk Olahan	pelatihan peningkatan kualitas produk olahan m...	[[pelatihan, peningkatan, kualitas, produk, ol...
4	pelatihan	https://sulselprov.go.id/welcome/post/1000-per...	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...	[[pelatihan, ini, memberikan, keterampilan, ba...
...	...	...	...	...	...
96	non_pelatihan	https://www.cnnindonesia.com/internasional/202...	Produsen Medis Buang Alat Swab usai Dikunjungi...	keputusan untuk membuang alat swab karena bany...	[[keputusan, untuk, membuang, alat, swab, kare...
97	non_pelatihan	https://www.merdeka.com/jateng/insecure-adalah...	Insecure adalah Perasaan Tidak Aman yang Norma...	ketika seseorang merasa tidak nyaman dengan di...	[[ketika, seseorang, merasa, tidak, nyaman, de...
98	non_pelatihan	https://www.merdeka.com/jakarta/mrt-sediakan-r...	MRT Sediakan Ruangan Isolasi untuk Penumpang D...	yang menunjukkan gejala akan langsung di arahk...	[[yang, menunjukkan, gejala, akan, langsung, d...

Gambar 3. 22 Tokenisasi

Pada kode diatas untuk memulai tahapan *text pre-processing* bagian tokenisasi perlu mengimport beberapa *library python* terlebih dahulu seperti *pandas*, *numpy*, dan *json*. Dimana penulis mempersiapkan dokumen pencarian *website* mengenai informasi keterampilan dengan format *.json* yang berisikan 101 data yang akan digunakan untuk proses *text pre-processing*. Setelah *pandas* membaca format *.json* yang telah ditampilkan pada gambar 3.22, import *nltk* untuk mendukung proses pengolahan bahasa natural seperti *tokenization*, *stopword*, *case folding*, *stemming* dan *lemmatization*, serta *classification*. Setelah itu download *nltk.punkt* untuk membagi teks menjadi daftar kalimat.



Dimana dari nltk akan meng-import `sent_tokenize` untuk dilakukannya pemotongan kalimat berdasarkan kata yang menyusunnya dan data yang diambil adalah `'short_desc'` yaitu deskripsi singkat mengenai isi *website* yang akan di latih menjadi `'short_desc_tokenized'` yaitu deskripsi singkat pada setiap *website* telah dipotong menjadi per kata apabila telah berhasil memotong kalimat. Setelah itu dari nltk.tokenize meng-import `RegexpTokenizer` (`r'\w+'`) untuk memisahkan kalimat menjadi kata-kata tanpa punctuation menggunakan *regular expression*. Sehingga `'short_desc_tokenized'` akan di potong atau dipisahkan kalimat menjadi kata-kata dan mengubah bentuk kata-kata menjadi huruf kecil semua atau disebut dengan *case folding*.

#### b. Stopwords removal

```
from nltk.corpus import stopwords

data["text"] = data.short_desc_tokenized.apply(lambda x:
[[split_word for split_word in kalimat if split_word not in
stopwords.words('stopwords_pelatihan')] for kalimat in x])
data
```

	kategori	link	judul	short_desc	short_desc_tokenized	text
0	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Keterampilan Produk Olahan Makanan	pelaksanaan pelatihan keterampilan produk olah...	[[pelaksanaan, keterampilan, produk...	[[pelaksanaan, pelatihan, keterampilan, produk...
1	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Ekonomi Produktif Bagi Perempuan	pelatihan pemberdayaan dan peningkatan keterampilan...	[[pelatihan, pemberdayaan, dan, peningkatan, k...	[[pelatihan, pemberdayaan, peningkatan, ketera...
2	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Pembuatan Manggleng	pelatihan peningkatan kualitas olahan pangan	[[pelatihan, peningkatan, kualitas, olahan, pa...	[[pelatihan, peningkatan, kualitas, olahan, pa...
3	pelatihan	https://www.giripanggung-tepus.desa.id/first/a...	Pelatihan Peningkatan Kualitas Produk Olahan	pelatihan peningkatan kualitas produk olahan m...	[[pelatihan, peningkatan, kualitas, produk, ol...	[[pelatihan, peningkatan, kualitas, produk, ol...
4	pelatihan	https://sulselprov.go.id/welcome/post/1000-per...	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...	[[pelatihan, ini, memberikan, keterampilan, ba...	[[pelatihan, keterampilan, perempuan]]
...	...	...	...	...	...	...
96	non_pelatihan	https://www.cnnindonesia.com/internasional/202...	Produsen Medis Buang Alat Swab usai Dikunjungi...	keputusan untuk membuang alat swab karena bany...	[[keputusan, untuk, membuang, alat, swab, kare...	[[keputusan, alat, swab, banyaknya, ...

Gambar 3. 23 Stopwords removal

Pada kode diatas, tahapan *text-pre-processing* bagian *stopwords removal*. Setelah tahapan tokenisasi, *stopwords removal* untuk membuang kata-kata yang tidak berpengaruh terhadap proses klasifikasi sehingga telah disiapkan file bernama `'stopwords_pelatihan'` untuk kata-kata yang memiliki kata depan, kata

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

sambung, kata yang sering muncul namun tidak memiliki makna akan dihilangkan seperti pada gambar 3.24.

```

stopwords_pelatihan - Notepad
File Edit Format View Help
ialah
ibarat
ibaratkan
ibaratnya
ibu
ikut
ingat
ingat-ingat
ingin
inginkan
inginkan
ini
inikah
inilah
itu
itukah
itulah
jadi
jadilah
jadinya
jangan
janganlah
janganlah
jauh
  
```

Gambar 3. 24 Stopwords\_pelatihan

Sehingga 'short\_desc\_tokenized' akan di potong kata-kata yang memiliki kata depan, kata sambung, kata yang sering muncul, dan hanya tersisa kata yang tidak ada pada stopwords\_pelatihan menjadi 'text'.

c. *Stemming dan Lemmatization*

```

from nltk.stem import WordNetLemmatizer
lm = WordNetLemmatizer()

data["text"] = data.text.apply(lambda x:
[[lm.lemmatize(split_word) for split_word in words] for
words in x])
data

data["joined_text"] = data["text"].apply(lambda x: ["
".join(sent) for sent in x])
data

data["joined_text"] = data["joined_text"].apply(" ".join)
data

data.to_csv("preprocessed_pelatihan2.csv")
  
```



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```

from nltk.stem import WordNetLemmatizer
lm = WordNetLemmatizer()

data["text"] = data.text.apply(lambda x: [[lm.lemmatize(split_word) for split_word in words] for words in x])
data

```

kategori	link	judul	short_desc	short_desc_tokenized	text	
0	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Keterampilan Produk Olahsan Makanan	pelaksanaan pelatihan keterampilan produk olah...	[[pelaksanaan, pelatihan, keterampilan, produk...]]	[[pelaksanaan, pelatihan, keterampilan, produk...]]
1	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Ekonomi Produktif Bagi Perempuan	pelatihan pemberdayaan dan peningkatan keterampilan...	[[pelatihan, pemberdayaan, dan, peningkatan, k...]]	[[pelatihan, pemberdayaan, peningkatan, ketera...]]
2	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Pembuatan Manggleg	pelatihan peningkatan kualitas olahan pangan	[[pelatihan, peningkatan, kualitas, olahan, pa...]]	[[pelatihan, peningkatan, kualitas, olahan, pa...]]
3	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Peningkatan Kualitas Produk Olahsan	pelatihan peningkatan kualitas produk olahsan m...	[[pelatihan, peningkatan, kualitas, produk, ol...]]	[[pelatihan, peningkatan, kualitas, produk, ol...]]
4	pelatihan	<a href="https://sulselprov.go.id/welcome/post/1000-per...">https://sulselprov.go.id/welcome/post/1000-per...</a>	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...	[[pelatihan, ini, memberikan, keterampilan, ba...]]	[[pelatihan, keterampilan, perempuan]]
...	...	...	...	...	...	...
96	non_pelatihan	<a href="https://www.cnnindonesia.com/internasional/202...">https://www.cnnindonesia.com/internasional/202...</a>	Produsen Medis Buang Alat Swab usai Dikunjungi...	keputusan untuk membuang alat swab karena bany...	[[keputusan, untuk, membuang, alat, swab, kare...]]	[[keputusan, membuang, alat, swab, banyaknya]]
97	non_pelatihan	<a href="https://www.merdeka.com/jateng/insecure-adalah...">https://www.merdeka.com/jateng/insecure-adalah...</a>	Insecure adalah Perasaan Tidak Aman yang Norma...	ketika seseorang merasa tidak nyaman dengan di...	[[ketika, seseorang, merasa, tidak, nyaman, de...]]	[[nyaman, situasi]]

```

data["joined_text"] = data["text"].apply(lambda x: [" ".join(sent) for sent in x])
data

```

kategori	link	judul	short_desc	short_desc_tokenized	text	joined_text	
0	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Keterampilan Produk Olahsan Makanan	pelaksanaan pelatihan keterampilan produk olah...	[[pelaksanaan, pelatihan, keterampilan, produk...]]	[[pelaksanaan, pelatihan, keterampilan, produk...]]	[[pelaksanaan, pelatihan, keterampilan, produk ola...]]
1	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Ekonomi Produktif Bagi Perempuan	pemberdayaan dan peningkatan keterampilan...	[[pelatihan, pemberdayaan, dan, peningkatan, k...]]	[[pelatihan, pemberdayaan, peningkatan, ketera...]]	[[pelatihan, pemberdayaan, peningkatan, keterampilan...]]
2	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Pembuatan Manggleg	pelatihan peningkatan kualitas olahan pangan	[[pelatihan, peningkatan, kualitas, olahan, pa...]]	[[pelatihan, peningkatan, kualitas, olahan, pa...]]	[[pelatihan, peningkatan, kualitas olahan pangan]]
3	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Peningkatan Kualitas Produk Olahsan	pelatihan peningkatan kualitas produk olahsan m...	[[pelatihan, peningkatan, kualitas, produk, ol...]]	[[pelatihan, peningkatan, kualitas, produk, ol...]]	[[pelatihan, peningkatan, kualitas produk olahsan ...]]
4	pelatihan	<a href="https://sulselprov.go.id/welcome/post/1000-per...">https://sulselprov.go.id/welcome/post/1000-per...</a>	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...	[[pelatihan, ini, memberikan, keterampilan, ba...]]	[[pelatihan, keterampilan, perempuan]]	[[pelatihan, keterampilan perempuan]]
...	...	...	...	...	...	...	...
96	non_pelatihan	<a href="https://www.cnnindonesia.com/internasional/202...">https://www.cnnindonesia.com/internasional/202...</a>	Produsen Medis Buang Alat Swab usai Dikunjungi...	keputusan untuk membuang alat swab karena bany...	[[keputusan, untuk, membuang, alat, swab, kare...]]	[[keputusan, membuang, alat, swab, banyaknya, ...]]	[[keputusan, membuang alat swab banyaknya orang ...]]

```

data["joined_text"] = data["joined_text"].apply(" ".join)
data

```

kategori	link	judul	short_desc	short_desc_tokenized	text	joined_text	
0	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Keterampilan Produk Olahsan Makanan	pelaksanaan pelatihan keterampilan produk olah...	[[pelaksanaan, pelatihan, keterampilan, produk...]]	[[pelaksanaan, pelatihan, keterampilan, produk...]]	[[pelaksanaan, pelatihan, keterampilan, produk olah...]]
1	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Ekonomi Produktif Bagi Perempuan	pemberdayaan dan peningkatan keterampilan...	[[pelatihan, pemberdayaan, dan, peningkatan, k...]]	[[pelatihan, pemberdayaan, peningkatan, ketera...]]	[[pelatihan, pemberdayaan, peningkatan, keterampilan...]]
2	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Pembuatan Manggleg	pelatihan peningkatan kualitas olahan pangan	[[pelatihan, peningkatan, kualitas, olahan, pa...]]	[[pelatihan, peningkatan, kualitas, olahan, pa...]]	[[pelatihan, peningkatan, kualitas olahan pangan]]
3	pelatihan	<a href="https://www.giripanggung-tepus.desa.id/first/a...">https://www.giripanggung-tepus.desa.id/first/a...</a>	Pelatihan Peningkatan Kualitas Produk Olahsan	pelatihan peningkatan kualitas produk olahsan m...	[[pelatihan, peningkatan, kualitas, produk, ol...]]	[[pelatihan, peningkatan, kualitas, produk, ol...]]	[[pelatihan, peningkatan, kualitas produk olahsan m...]]
4	pelatihan	<a href="https://sulselprov.go.id/welcome/post/1000-per...">https://sulselprov.go.id/welcome/post/1000-per...</a>	1000 Perempuan Dapatkan Pelatihan Keterampilan...	pelatihan ini memberikan keterampilan bagi per...	[[pelatihan, ini, memberikan, keterampilan, ba...]]	[[pelatihan, keterampilan, perempuan]]	[[pelatihan, keterampilan perempuan]]
...	...	...	...	...	...	...	...
96	non_pelatihan	<a href="https://www.cnnindonesia.com/internasional/202...">https://www.cnnindonesia.com/internasional/202...</a>	Produsen Medis Buang Alat Swab usai Dikunjungi...	keputusan untuk membuang alat swab karena bany...	[[keputusan, untuk, membuang, alat, swab, kare...]]	[[keputusan, membuang, alat, swab, banyaknya, ...]]	[[keputusan, membuang alat swab banyaknya orang r...]]
			Insecure adalah Perasaan Tidak	ketika seseorang merasa tidak	[[ketika, seseorang, merasa, tidak, nyaman, de...]]	[[nyaman	

Gambar 3. 25 Stemming dan Lemmatization


**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Pada kode diatas, tahapan *text-pre-processing* bagian *stemming* dan *lemmatization*. Setelah dilakukannya tahapan *stopwords removal*, kata-kata yang tersisa akan diubah menjadi bentuk kata dasarnya. Dari `nlTK.stem` meng-import `WorldNetLemmatizer` untuk mengembalikan kata menjadi bentuk dasarnya. Sehingga data 'text' akan di lematisasi menjadi kata dasarnya yang dapat dilihat pada gambar 3.25 kata 'nyaman' dan 'situasi' adalah kata dasar. Setelah itu kata-kata dasar yang tersisa digabungkan menjadi 'joined\_text' yang masih terdapat tanda koma (,) dan kurung [ ], yang kemudian di hilangkannya tanda koma (,) kurung [ ], dan disimpan ke dalam format .csv.

## 2. Text Classification

```
from sklearn import model_selection
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test =
model_selection.train_test_split(data['joined_text'],data['
kategori'],test_size=0.4)
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn import model_selection
X_train, X_test, y_train, y_test = model_selection.train_test_split(data['joined_text'],data['kategori'],test_size=0.4)
```

```
print (X_train.shape, y_train.shape)
print (X_test.shape, y_test.shape)
```

```
(60,) (60,)
(41,) (41,)
```

Gambar 3. 26 Persiapan data klasifikasi

Pada kode diatas, merupakan tahapan *text classification* yang dimana menggunakan *library* `sklearn` untuk memulai pengklasifikasian menggunakan *machine learning* dengan meng-import `model_selection` dan *sublibrary* dari `model_selection` yaitu `train_test_split`. Dimana akan terbagi menjadi 2 bagian yaitu *training set* dan *testing set*. `X_train` sebagai data X yang ingin kita latih, `X_test` sebagai data X yang ingin kita tes, `y_train` adalah variabel dependen yang dilatih, dan `y_test` adalah variabel dependen yang diuji. Dimana data 'joined\_text' sebagai variabel independen dan 'kategori' sebagai variabel dependen yang dibagi untuk di test sebesar 40%. Sehingga dapat dilihat pada gambar 3.26 bahwa dari 101 data yang ada dengan data testing 40% maka akan menampilkan (60) (60) untuk `X_train` dan `y_train` dan (41) (41) untuk `X_test` dan `y_test`.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```

from sklearn import linear_model
from sklearn.feature_extraction.text import TfidfVectorizer
Tfidf_vect = TfidfVectorizer(max_features=5000)
Tfidf_vect.fit(data['joined_text'])
X_train_Tfidf = Tfidf_vect.transform(X_train)
X_test_Tfidf = Tfidf_vect.transform(X_test)

import pickle
filename = "vectorizer.sav"
with open(filename, 'wb') as handle:
    pickle.dump(Tfidf_vect, handle)

from sklearn.preprocessing import LabelEncoder
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)

filename = "encoder.sav"
with open(filename, 'wb') as handle:
    pickle.dump(Encoder, handle)

```

```

from sklearn import linear_model

from sklearn.feature_extraction.text import TfidfVectorizer
Tfidf_vect = TfidfVectorizer(max_features=5000)
Tfidf_vect.fit(data['joined_text'])
X_train_Tfidf = Tfidf_vect.transform(X_train)
X_test_Tfidf = Tfidf_vect.transform(X_test)

import pickle
filename = "vectorizer.sav"
with open(filename, 'wb') as handle:
    pickle.dump(Tfidf_vect, handle)

from sklearn.preprocessing import LabelEncoder
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)

filename = "encoder.sav"
with open(filename, 'wb') as handle:
    pickle.dump(Encoder, handle)

```

```

print(Tfidf_vect.vocabulary_)
{'pelaksanaan': 267, 'pelatihan': 269, 'keterampilan': 164, 'produk': 309, 'olahan': 252, 'makanan': 190, 'pemberdayaan': 271, 'perempuan': 295, 'peningkatan': 286, 'pembuatan': 274, 'kue': 178, 'kualitas': 177, 'pangan': 259, 'kelompok': 148, 'belajar': 32, 'merajut': 239, 'ta': 366, 'kegiatan': 146, 'berharap': 36, 'penyuluhan': 293, 'peserta': 300, 'memiliki': 210, 'mengikut i': 225, 'kerajinan': 158, 'tangan': 367, 'pendidikan': 280, 'bidang': 47, 'perlindungan': 297, 'anak': 14, 'idah': 115, 'syahi dah': 365, 'bekerjasama': 31, 'lembaga': 185, 'menggelar': 224, 'handycraft': 104, 'warga': 404, 'program': 310, 'pembinaan': 2

```

```

print(X_train_Tfidf)
(0, 295) 0.1675667306508102
(0, 269) 0.12189296634422674
(0, 265) 0.37162470889097854
(0, 254) 0.28884167835329705
(0, 210) 0.30258003824552937
(0, 140) 0.30258003824552937
(0, 128) 0.37162470889097854
(0, 78) 0.37162470889097854
(0, 77) 0.37162470889097854
(0, 3) 0.37162470889097854
(1, 397) 0.3399759559434501
(1, 201) 0.3399759559434501
(1, 121) 0.3399759559434501
(1, 89) 0.3399759559434501
(1, 60) 0.3399759559434501
(1, 59) 0.5536227021013714
(1, 56) 0.3399759559434501
(2, 394) 0.4472135954999579
(2, 236) 0.4472135954999579
(2, 167) 0.4472135954999579
(2, 134) 0.4472135954999579
(2, 49) 0.4472135954999579
(3, 391) 0.46821361002124223
(3, 263) 0.5101555368678837
(3, 101) 0.5101555368678837
:
:
(57, 269) 0.14867091120674605
(57, 193) 0.45326474328090166
(58, 411) 0.2656700654559485

```

Gambar 3. 27 Mengubah kata menjadi matriks



Selanjutnya, dari sklearn meng-import `linear_model` dan `feature_extraction.text` untuk mengekstrak fitur dalam format yang didukung oleh algoritma *machine learning* dari dataset yang ada yaitu format teks dan meng-import `TfidfVectorizer` untuk menciptakan matriks pada semua kata dan skor di dalam dokumen. Sehingga akan mengubah kata-kata yang ada pada data `'joined_text'` menjadi matriks. Dimana `X_train_Tfidf` adalah transformasi `X_train` menjadi matriks dan `X_test_Tfidf` adalah transformasi `X_test` menjadi matriks. Dengan meng-import `pickle` akan membaca filename `vectorizer.sav` untuk menyimpan matriks `Tfidf_vect`. Selanjutnya meng-import `LabelEncoder` untuk menemukan semua kelas dan memberikan masing-masing nilai antara 0 dan 1 dan digunakan pada `y_train` dan `y_test`. Dengan filename `encoder.sav` untuk menyimpan nilai label. Sehingga dapat dilihat pada gambar 3.27 bahwa setiap kata yang ada memiliki nilai nya masing-masing seperti `'pelaksanaan': 267`, `'pelatihan': 269` dan apabila kita melihat data `X_train ('joined_text')`, data yang telah memiliki nilai pada setiap katanya akan mendapatkan nilai matriks seperti pada data ke 0 terdapat kata `'pelatihan'` dengan nilainya yaitu 269 mendapatkan nilai matriks yaitu 0.12189296634422674.

```
from sklearn import svm
from sklearn.metrics import accuracy_score

# Pengklasifikasian Algoritma SVM
# Sesuai dengan dataset pelatihan pada classifier
SVM = svm.LinearSVC(C=1.0, class_weight='balanced',
multi_class='ovr', random_state=40, max_iter=10000)
SVM.fit(X_train_Tfidf,y_train)
# prediksi label pada set data validasi
predictions_SVM = SVM.predict(X_test_Tfidf)
# Menggunakan fungsi accuracy_score untuk mendapatkan nilai
akurasi
print("SVM Accuracy Score ->
",accuracy_score(predictions_SVM, y_test))
```

Setelah menjadikan `'joined_text'` menjadi matriks dan `'kategori'` menjadi label maka selanjutnya pengklasifikasi menggunakan SVM (*Support Vector Machine*) untuk mendapatkan nilai akurasi dari dataset yang di telah dilatih apakah valid atau tidak. Menggunakan model `Linear.SVC` dengan parameternya adalah (`C=1.0`, `class_weight= 'balanced'`, `multi_class= 'ovr'`, `random_state = 40`, `max_iter = 10000`) dan fit modelnya adalah `X_train_Tfidf`

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



dan `y_train`. Prediksi svm yang digunakan adalah `X_test_Tfidf`. Sehingga akan didapatkan nilai akurasi dengan menggunakan prediksi svm dan `y_test`. Dimana pada penelitian ini didapatkan nilai akurasi sebesar 0.975609756097561.

```
from sklearn.externals import joblib

result = open("training_skills.pkl", "wb")

joblib.dump(SVM, result)

result.close
```

Selanjutnya menggunakan `sklearn.externals` dengan meng-import `joblib` untuk menyimpan dan memuat data model ke `.pkl` dengan nilai yang dimasukkan adalah hasil SVM. Dimana 'training\_skills.pkl' akan digunakan pada saat menentukan website yang dimasukkan adalah website mengenai pelatihan keterampilan atau bukan.

```
#ML Packages
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.utils._joblib import joblib

@app.route('/scrape')
def scrape():

    data = pd.read_csv("data/preprocessed_pelatihan2.csv")
    data_X = data.joined_text
    data_Y = data.kategori

    corpus = data_X
    cv = CountVectorizer()
    X = cv.fit_transform(corpus)

    result = open("models/training_skills.pkl", "rb")
    clf = joblib.load(result)

    url = request.args.get('url')
    try:
        response = requests.get(url)
        content = BeautifulSoup(response.text,
' lxml').prettify()
    except:
        flash('Tidak Dapat Mengambil URL "%s"' % url,
'danger')
        content = ''

    datas = [url]
    vect = cv.transform(datas).toarray()
    my_scrape = clf.predict(vect)
    return render_template("home_admin.html", content =
content, scraping = my_scrape)
```

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Pada kode diatas, merupakan implementasi *text mining* ke dalam *website*. Dimana menggunakan *library* `sklearn.feature_extraction.text` dengan meng-import `CountVectorizer` untuk menjalankan fungsi matriks dan `sklearn.utils.joblib` dengan meng-import `joblib` untuk memuat data model. Selanjutnya pada fungsi `scrape()` akan di implementasikan hasil preprocessing yang dibentuk dalam format `.csv` dimana membagi `data_X` adalah `'joined_text'` dan `data_Y` adalah `'kategori'` setelah itu `data_X` akan di transform menjadi nilai matriks dan memuat model yang telah di latih dalam bentuk `.pkl` untuk digunakan. Input yang dimasukkan adalah url, sehingga ketika url di masukkan sistem akan melakukan *exception handling* Jika url, sistem akan melakukan *scraping* dengan menggunakan `BeautifulSoup`. Jika bukan url yang dimasukkan maka akan menampilkan error `'Tidak dapat mengambil url'`. Dimana menginisialisasi `datas` adalah url yang kemudian akan di transform menjadi array untuk url dapat di prediksi apakah url yang dimasukkan adalah *website* mengenai pelatihan keterampilan atau bukan sesuai dengan yang ada pada `training_skills.pkl` dan akan di tampilkan di `home_admin.html`.



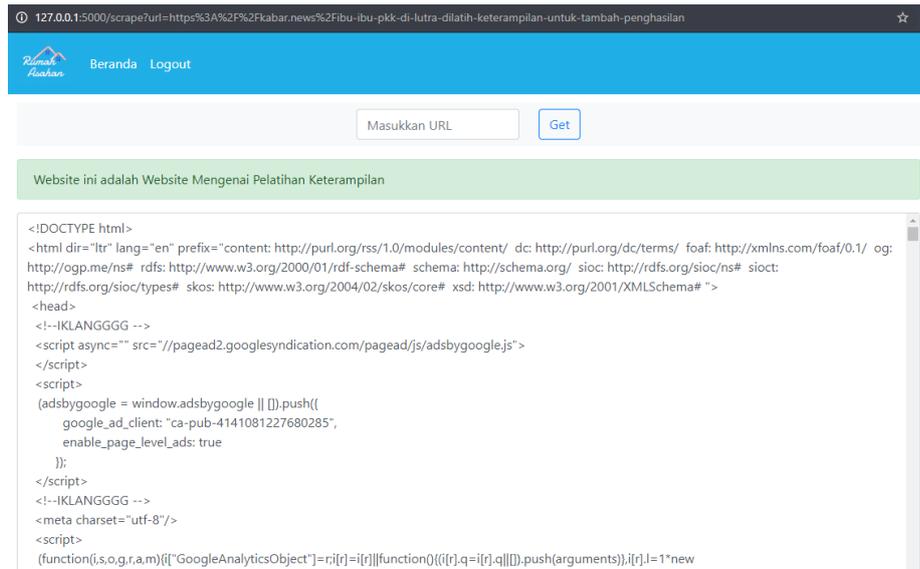
Gambar 3. 28 Tampilan halaman admin

Pada halaman admin, admin dapat memasukkan url melalui form input dan klik *button* `get` untuk mengetahui apakah yang dimasukkan benar adalah url dan apakah url tersebut *website* mengenai pelatihan keterampilan atau bukan.



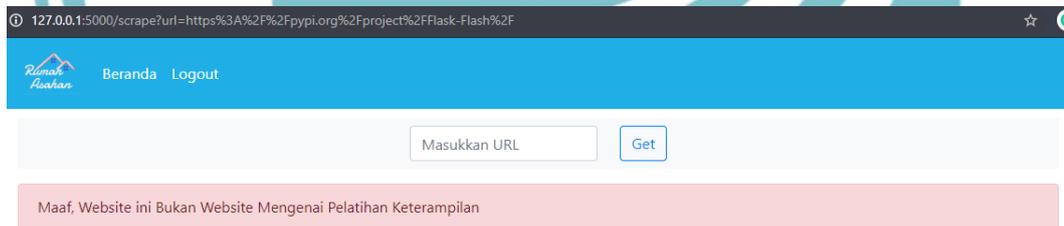
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



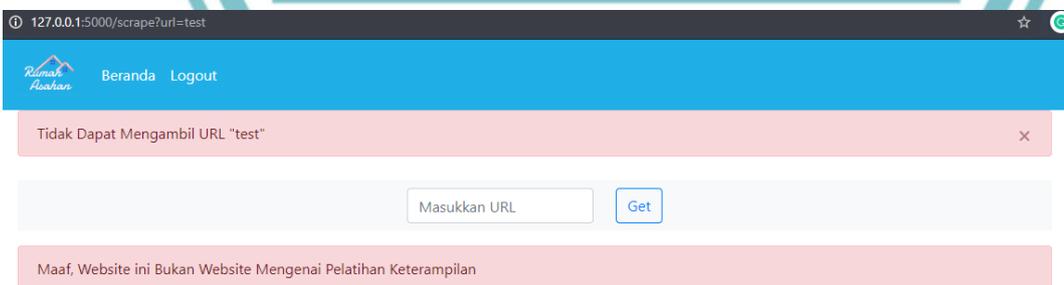
Gambar 3. 29 Tampilan halaman website mengenai pelatihan keterampilan

Pada halaman 3.29 menandakan bahwa url yang dimasukkan benar adalah url dan masuk ke dalam *website* mengenai pelatihan keterampilan. Dimana url tersebut akan melakukan *scraping* dan menampilkan elemen html.



Gambar 3. 30 Tampilan halaman website bukan mengenai pelatihan keterampilan

Pada gambar 3.30 merupakan bahwa url yang dimasukkan benar adalah url tetapi bukan masuk ke dalam *website* mengenai pelatihan keterampilan.



Gambar 3. 31 Tampilan halaman bukan url

Pada gambar 3.31 menandakan bahwa yang dimasukkan bukan url.



### 3. Fungsi *Web Scraping*

```

@app.route("/results")
def results():
    args = []
    results = []
    final_result = []

    response = requests.get(request.args.get('url'))
    content = BeautifulSoup(response.text, 'lxml')

    try:
        titles = content.find_all('title')
        result_title = titles[0].text.strip()

        time = content.find_all('time')
        result_time = time[0].text.strip()

        cp = content.find_all('p')
        count = len(cp)
        for index in range(0, count):
            p = content.find_all('p')[index].text.strip()
            results.append(p)
            final_result = " ".join(results)

    except:
        titles = content.find_all('title')
        result_title = titles[0].text.strip()

        time = content.find_all('time')
        result_time = ''

        cp = content.find_all('p')
        count = len(cp)
        for index in range(0, count):
            p = content.find_all('p')[index].text.strip()
            results.append(p)
            final_result = " ".join(results)

    title = result_title
    date = result_time
    description = final_result

    cur = mysql.connection.cursor()
    cur.execute("INSERT INTO articles (title, date,
description) VALUES (%s,%s,%s)", (title, date, description,))
    mysql.connection.commit()

    flash("Data sudah tersimpan! Klik Edit apabila ada data
yang ingin diubah", 'info')
    return render_template("results.html",
result_title=result_title, result_time=result_time,
p=final_result)

```

Pada kode diatas, dimana setelah mengetahui url yang dimasukkan adalah *website* mengenai pelatihan keterampilan dan telah melakukan *scraping* pada *website* yang

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

diinput maka akan di lakukan implementasi *web scraping* yang akan di ambil informasinya seperti judul, tanggal, deskripsi dengan masing-masing tag yang di ambil adalah <title>, <time>, dan <p>. Menggunakan *exception handling* apabila pada saat di lakukan *scraping* ada *website* yang tidak memiliki <time> sehingga hanya menampilkan hasil *scraping* dari judul dan deskripsi. Untuk deskripsi dilakukan perulangan sehingga tidak hanya satu <p> saja melainkan semua <p> yang ada pada *website* yang di masukkan akan di ambil informasinya. Hasil dari pengambilan informasi menggunakan *web scraping* dengan *library* BeautifulSoup akan disimpan ke *database* sesuai dengan kolom yang telah disediakan seperti judul, tanggal, dan deskripsi. Hasil dari implementasi *web scraping* akan di tampilkan pada halaman `results.html`.

judul	tanggal	deskripsi	action
Pelatihan Ketrampilan Memasak Berbahan Dasar Jagung untuk Pemberdayaan Perempuan Salatiga – Dinas Pemberdayaan Perempuan dan Pelindungan Anak	9 Mei 2018	Salatiga – Dalam rangka melanjutkan Program Pemberdayaan Perempuan di Kota Salatiga, Pada hari Senin s/d Rabu, tanggal 7 s/d 9 Mei 2018, Dinas Pemberdayaan Perempuan dan Pelindungan Anak Kota Salatiga kembali menyelenggarakan kegiatan Pelatihan Ketrampilan Memasak Berbahan dasar Jagung yang sebelumnya juga menyelenggarakan kegiatan serupa yaitu Pelatihan ketrampilan memasak dengan bahan dasar dari ubi pada tanggal 24 s/d 26 April 2018 yang lalu. Kegiatan yang di laksanakan di rumah dinas Walikota Salatiga tersebut diikuti oleh perwakilan perempuan dari keluarga kurang mampu sebanyak 100 orang. Dengan mendatangkan narasumber yang sama dengan kegiatan sebelumnya yaitu Kristin Anggriani, Founder dari Brownies Ndeso Kris Kris. Kali ini menu yang dipilih tidak kalah unik untuk diolah yaitu Krokot, Crunchy cookies dan Bakpao berbahan dasar jagung yang bercitarasa tinggi. Dalam acara pembukaan kegiatan pelatihan Kepala Dinas Pemberdayaan Perempuan dan Pelindungan Anak Kota Salatiga, Dra. Sri Satuti, MM. kembali menegaskan tujuan kegiatan pelatihan memasak tersebut masyarakat perempuan dari kurang mampu di Kota Salatiga dapat lebih mandiri dalam meningkatkan taraf hidup keluarganya dan dapat memberikan kontribusi dalam meningkatkan pendapatan keluarga dari ketrampilan yang di dapat dari kegiatan pelatihan tersebut. Alamat email Anda tidak akan dipublikasikan. Ruas yang wajib ditandai * Komentar Nama * Email * Situs Web Simpan nama, email, dan situs web saya pada peramban ini untuk komentar saya berikutnya.	Edit

Gambar 3. 32 Tampilan hasil scraping dengan tanggal

Pada gambar diatas merupakan tampilan hasil scraping pada *website* yang memiliki <time> sehingga menampilkan keseluruhan informasi yang telah di tentukan yaitu judul, tanggal, dan deskripsi. Dimana hasil *scraping* sudah tersimpan ke dalam *database*. Di berikan *alert* untuk admin dapat memperbarui data *website* tanpa mengubah isinya, dikarenakan *website* ini memiliki <time> maka perlu di ubah format tanggal menjadi (dd/mm/yyyy) dan menentukan kategori apakah *website* ini masuk ke dalam kategori makanan atau kerajinan tangan dengan mengklik *button* edit.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

judul	tanggal	deskripsi	action
Ibu-ibu PKK di Lutra Dilatih Keterampilan Untuk Tambah Penghasilan - kabar.news		KABAR.NEWS, Masamba - Tim Penggerak PKK Kabupaten Luwu Utara (Lutra) mengasa keterampilan perempuan di Lutra melalui pelatihan membuat hantaran, di Aula Hotel Remaja Lutra, Jumat (3/8/2018). Pelatihan itu dipimpin langsung Ketua Tim Penggerak PKK Lutra, Enny Tahar beserta Tim Pokja II PKK Lutra. Dikesempatan itu Enny Tahar mengatakn kegiatan pelatihan digelar untuk meningkatkan pengetahuan ibu - ibu PKK. "Kegiatan ini dilaksanakan dengan tujuan untuk meningkatkan pengetahuan dan keterampilan ibu-ibu tim penggerak PKK dan diharapkan ibu-ibu yang mengikuti kegiatan ini dapat memperaktekan di tempatnya masing-masing," kata Enny Tahar melalui rilis Humas dan Protokol Lutra kepada KABAR.NEWS, Jumat (3/8/2018). Menurut dia, selain menambah pengetahuan ibu - ibu PKK juga diharap pelatihan keterampilan tersebut menambah pendapatan ekonomi. "Bukan hanya itu keterampilan ini kalau dikelola dengan manajemen yang baik dan profesional dapat meningkatkan ekonomi produktif yang tentunya meningkatkan pendapatan masyarakat khususnya ibu-ibu," kata Enny. Sementara Ketua Tim Pokja II PKK Lutra, Candarayani Aspar menjelaskan bahwa kegiatan tersebut diikuti 50 peserta yang terdiri dari pengurus PKK tingkat kabupaten dan Pengurus PKK Kecamatan se - Kabupaten Lutra. "Kami berharap pelaksanaan kegiatan ini sebagai indikator output dalam menghasilkan indikator outcome seperti peningkatan produksi kerajinan yang dapat menambah pendapatan keluarga dan kita dapat mencapai indikator sasaran yaitu mengsejahterakan masyarakat Luwu Utara," tandasnya.	Edit

Gambar 3. 33 Tampilan hasil scraping tanpa tanggal

Pada gambar diatas merupakan tampilan hasil scraping pada *website* yang tidak memiliki `<time>` sehingga hanya menampilkan judul dan deskripsi. Dimana hasil *scraping* sudah tersimpan ke dalam *database*. Di berikan *alert* untuk admin dapat memperbarui data *website* tanpa mengubah isinya. Dikarenakan *website* ini tidak memiliki `<time>` maka admin dapat mengklik *button* edit untuk admin dapat menginput tanggal sesuai dengan *website* aslinya dan diinput dengan formar (dd/mm/yyyy) dan menentukan kategori apakah *website* ini masuk ke dalam kategori makanan atau kerajinan tangan.

#### 4. Fungsi *Update*

```
@app.route('/update', methods=["POST", "GET"])
def update():

    if request.method == 'POST':
        date = request.form.get('date')
        kategori = request.form.get('kategori')

        cur = mysql.connection.cursor()
        cur.execute("UPDATE articles SET date=%s, kategori=%s
order by id desc limit 1", (date, kategori))
        flash("Data berhasil di Update!", 'success')
        mysql.connection.commit()

    print (date)
    print (kategori)
    return redirect(url_for("home_admin"))
```



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Pada kode di atas merupakan fungsi *update* untuk admin dapat memperbarui data atau menginput data yang kosong seperti tanggal dan menentukan kategori *website*. Dimana fungsi *update* dijalankan apabila admin mengklik *button* edit dan menampilkan tanggal yang telah tersimpan di *database* baik ada maupun tidak sesuai dengan id yang terakhir masuk sehingga admin dapat memperbarui atau menginputnya dan admin dapat menentukan kategori pada *website* yang telah di ambil informasinya masuk ke dalam kategori pelatihan keterampilan makanan atau kerajinan tangan. Data yang telah diperbarui akan disimpan ke *database* menampilkan *alert* bahwa ‘Data berhasil di update!’ dan kembali pada halaman *home\_admin.html*.

Gambar 3. 34 Tampilan form update data

Pada gambar 3.34 adalah tampilan form *update* data. Dimana *website* yang diambil informasinya ada pada gambar 3.32 yang menampilkan keseluruhan informasi yang telah di tentukan yaitu judul, tanggal, dan deskripsi. Sehingga ketika admin mengklik *button* edit akan menampilkan tanggal yang telah tersimpan di *database* yang di ubah formatnya menjadi (dd/mm/yyyy) dan admin menentukan kategori *website* ini adalah makanan.

Gambar 3. 35 Tampilan berhasil mempebarui data

Pada gambar 3.35 menandakan bahwa data telah berhasil di *update* dan tersimpan di *database*.

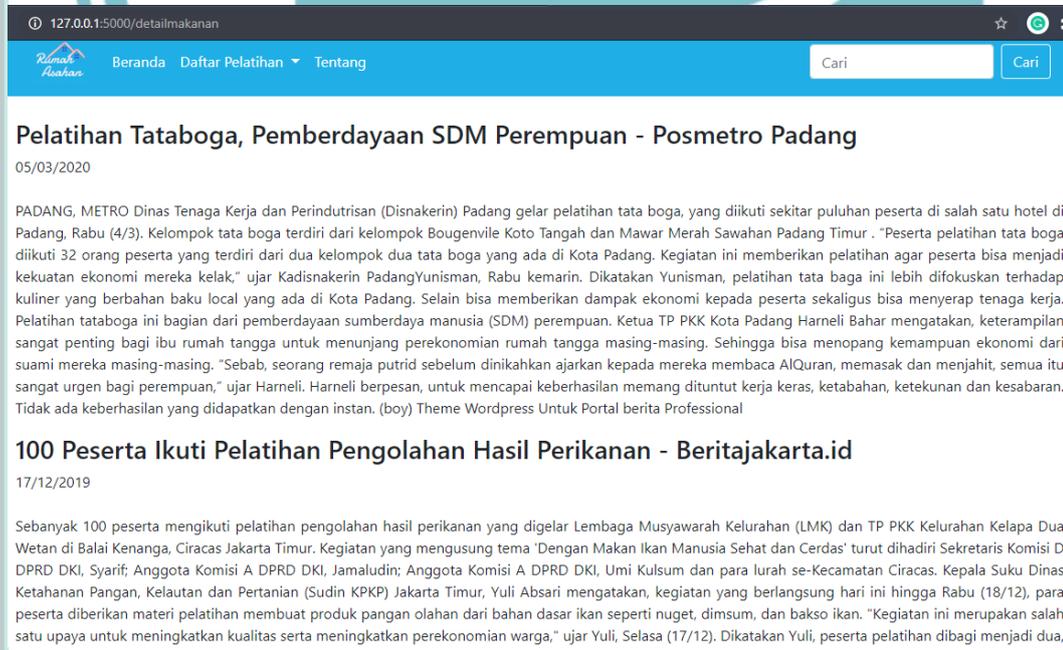


## 5. Fungsi Melihat Detail Informasi Pelatihan

```
@app.route('/detailmakanan')
def detail_makanan():
    cur = mysql.connection.cursor()
    cur.execute('''SELECT title, date, description FROM articles
WHERE kategori='makanan' ''')
    rv = cur.fetchall()
    return render_template("detailmakan.html", value=rv)

@app.route('/detailkerajinan')
def detail_kerajinan():
    cur = mysql.connection.cursor()
    cur.execute('''SELECT title, date, description FROM articles
WHERE kategori='kerajinan' ''')
    rv = cur.fetchall()
    return render_template("detailkerajinan.html", value=rv)
```

Pada kode di atas merupakan fungsi melihat detail informasi pelatihan sesuai dengan kategorinya masing-masing. Pada detail makanan mengeksekusi judul, tanggal, dan deskripsi pada table articles dengan kategori ‘makanan’ dan ditampilkan pada halaman detailmakan.html. Pada detail kerajinan mengeksekusi judul, tanggal, dan deskripsi pada table articles dengan kategori ‘kerajinan’ dan ditampilkan pada halaman detailkerajinan.html.



Gambar 3. 36 Tampilan halaman detail informasi pelatihan makanan

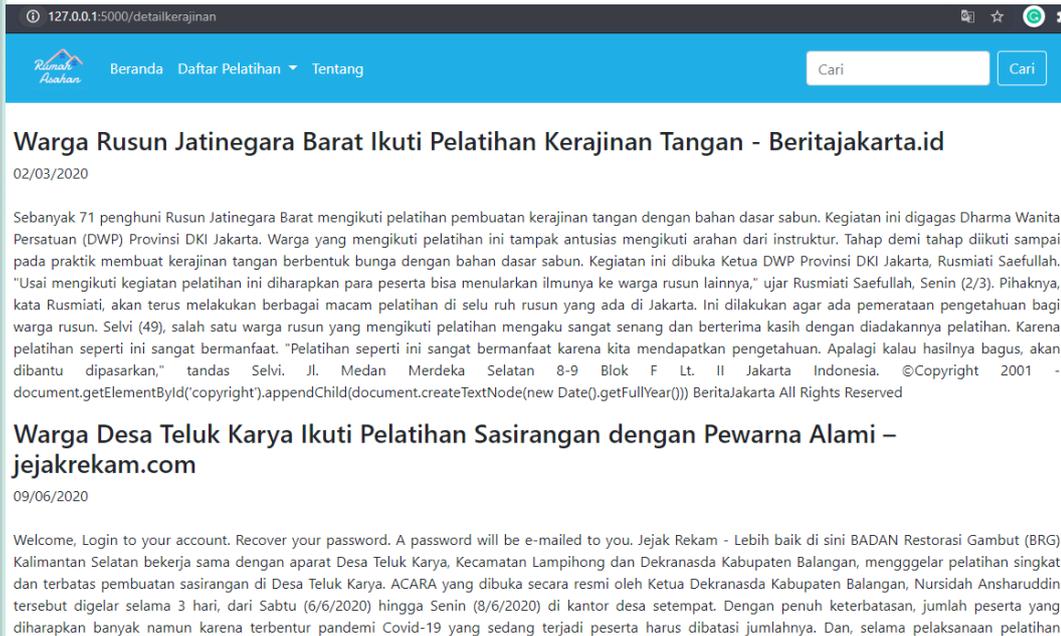
Gambar 3.36 merupakan tampilan halaman detail informasi pelatihan makanan yang telah berhasil diambil informasinya menggunakan *text mining* dan *web scraping*.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 3. 37 Tampilan halaman detail informasi pelatihan kerajinan tangan

Gambar 3.37 merupakan tampilan halaman detail informasi pelatihan kerajinan tangan yang telah berhasil diambil informasinya menggunakan *text mining* dan *web scraping*.

## 6. Fungsi Pencarian Pelatihan

```
@app.route('/search', methods=['GET', 'POST'])
def search():
    conn = mysql.connection
    cursor = conn.cursor()

    if request.method == "POST":
        article = "%" + request.form['article'] + "%"
        cursor.execute("SELECT title, date, description FROM
articles WHERE title LIKE %s OR description LIKE %s", (article,
article))
        conn.commit()
        data = cursor.fetchall()

    if len(data) == 0 and article == 'all':
        cursor.execute("SELECT title, date, description FROM
articles")
        conn.commit()
        data = cursor.fetchall()
    return render_template("search.html", data=data)
return render_template("search.html")
```

Pada kode diatas merupakan fungsi pencarian pelatihan. Apabila *user* menginput kata kunci yang diinginkan, sistem akan menampilkan judul, tanggal, dan deskripsi



dari table articles yang dimana menyesuaikan judul atau deskripsi dari hasil input *user*. Hasilnya akan di tampilkan pada halaman search.html.

The screenshot shows a search interface with a blue header. The search bar contains the text 'singkong'. Below the search bar, two search results are listed. The first result is titled 'Pelatihan Peningkatan Kualitas Produk Olahan - Website Giripanggung' with a date of 10/10/2019. The second result is titled 'Pelatihan Boga PPUEP di Kecamatan Ngaliyan - Dinas Pemberdayaan Perempuan dan Perlindungan Anak Kota Semarang' with a date of 19/08/2019. The background of the screenshot features a large, semi-transparent watermark of the Politeknik Negeri Jakarta logo.

Gambar 3. 38 Tampilan halaman pencarian pelatihan

Pada gambar 3.38 adalah tampilan pencarian pelatihan. Sebagai contohnya, *user* melakukan pencarian dengan kata kunci 'singkong' setelah *user* mengklik *button* cari maka sistem akan menampilkan hasil pencarian pelatihan berupa 2 artikel yang membahas mengenai kata kunci 'singkong'.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## BAB IV

### PEMBAHASAN

#### 4.1 Pengujian

##### 4.1.1 Deskripsi Pengujian

Setelah aplikasi selesai dibuat, tahapan selanjutnya adalah pengujian aplikasi. Proses pengujian pada aplikasi ini terbagi menjadi dua tahap pengujian, pengujian *black box testing* (fungsional) dan pengujian metode SVM (*Support Vector Machne*). Proses pengujian pada aplikasi ini bertujuan untuk mengetahui *bug* atau masalah yang ada pada sistem dan memperbaiki *bug* tersebut agar sistem yang dibuat dapat berjalan sesuai dengan kebutuhan pengguna.

##### 4.1.2 Prosedur Pengujian

Prosedur pengujian pada *website* Pelatihan Keterampilan Khusus Perempuan menggunakan *black box testing* dan pengujian metode SVM (*Support Vector Machine*). *Black box testing* adalah pengujian pada fungsionalitas yang ada dari setiap bagian di dalam sistem yang dibuat. Bertujuan untuk memastikan sistem yang telah dibuat sudah berjalan sesuai dengan harapan atau tidak.

Tabel 3. Prosedur pengujian black box testing

No	Item Uji	Detail Pengujian	Jenis Pengujian
1.	Login Admin	Melakukan proses login pada Admin	<i>Black Box</i>
2.	Halaman Admin	Melakukan proses input URL	<i>Black Box</i>
		Melakukan proses <i>scraping</i> dan simpan	<i>Black Box</i>
		Melihat hasil <i>scraping</i>	<i>Black Box</i>
		Memperbarui data hasil <i>scraping</i>	<i>Black Box</i>
3.	Pengujian URL	URL <i>website</i> mengenai pelatihan keterampilan	<i>Black Box</i>



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4.	Halaman Detail Informasi Pelatihan Keterampilan Kategori Makanan	Melihat detail informasi pelatihan keterampilan kategori 'Makanan'	<i>Black Box</i>
5.	Halaman Detail Informasi Pelatihan Keterampilan Kategori Kerajinan Tangan	Melihat detail informasi pelatihan keterampilan kategori 'Kerajinan Tangan'	<i>Black Box</i>
6.	Pencarian Pelatihan	Melakukan proses pencarian pelatihan	<i>Black Box</i>

#### 4.1.3 Data Hasil Pengujian

##### 1. *Alpha Testing*

*Alpha Testing* dilakukan beberapa kali pada masa pengembangan. Pengujian ini merupakan proses pengecekan pertama dimana dalam proses tersebut aplikasi sudah jadi dengan fiturnya menggunakan *black box testing*.

##### a. Pengujian Login Admin

Tabel 4. Pengujian login admin

Skenario Pengujian (Data Normal)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	
Admin memasukkan data <i>email</i> dan <i>password</i> Email: admin@gmail.com Password: admin123456	Sistem akan menampilkan halaman admin dari <i>website</i>	Sistem dapat masuk ke halaman admin setelah berhasil login	Valid
Skenario Pengujian (Data Salah)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Admin memasukkan data <i>email</i> dan <i>password</i> yang tidak terdaftar di <i>database</i>	Sistem akan menampilkan <i>alert</i> ‘kesalahan’	Sistem menampilkan <i>alert</i> ‘Email dan Password Salah’	Valid
Admin memasukkan data <i>email</i> dan tidak memasukkan <i>password</i>	Sistem akan menampilkan <i>alert</i> ‘kesalahan’	Sistem menampilkan <i>alert</i> pengisian pada <i>password</i>	Valid
Admin tidak memasukkan data <i>email</i> dan memasukkan <i>password</i>	Sistem akan menampilkan <i>alert</i> ‘kesalahan’	Sistem menampilkan <i>alert</i> pengisian pada <i>email</i>	Valid
Tidak memasukkan data <i>email</i> dan <i>password</i>	Sistem akan menampilkan <i>alert</i> ‘kesalahan’	Sistem menampilkan <i>alert</i> pengisian <i>email</i> dan <i>password</i>	Valid

b. Pengujian Halaman Admin

Tabel 5. Pengujian halaman admin

Skenario Pengujian (Data Normal)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	
Admin memasukkan data url <i>website</i> mengenai informasi pelatihan keterampilan	Sistem akan menampilkan <i>alert</i> bahwa url yang dimasukkan adalah <i>website</i> mengenai pelatihan keterampilan	Sistem menampilkan <i>alert</i> hijau bahwa benar url yang dimasukkan adalah <i>website</i> mengenai pelatihan keterampilan	Valid
Admin melihat elemen html pada url <i>website</i> pelatihan keterampilan	Sistem akan meng- <i>scraping</i> url <i>website</i> mengenai pelatihan keterampilan dengan	Sistem menampilkan elemen html pada <i>website</i> yang berisi informasi pelatihan keterampilan	Valid



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

	menampilkan elemen html		
Admin mengklik <i>button</i> scrape dan simpan untuk melihat hasil <i>scraping</i>	Sistem akan menyimpan dan menampilkan hasil <i>scraping</i> yang sudah ditentukan	Sistem menampilkan <i>alert</i> bahwa data telah tersimpan dan menampilkan hasil <i>scraping</i> yang telah ditentukan	Valid
Admin mengklik <i>button</i> edit dan mengisi form <i>update data</i>	Sistem akan menampilkan form <i>update data</i> dan kembali pada halaman admin saat data sudah berhasil diperbarui	Sistem dapat menampilkan form <i>update data</i> dan kembali pada halaman admin dengan <i>alert</i> bahwa data telah berhasil diperbarui	Valid
<b>Skenario Pengujian (Data Salah)</b>			<b>Kesimpulan</b>
<b>Data Masukkan</b>	<b>Hasil yang diharapkan</b>	<b>Hasil Pengujian</b>	
Admin memasukkan data url <i>website</i> bukan mengenai informasi pelatihan keterampilan	Sistem akan menampilkan <i>alert</i> bahwa url yang dimasukkan adalah bukan <i>website</i> mengenai pelatihan keterampilan	Sistem menampilkan <i>alert</i> merah bahwa url yang dimasukkan bukan <i>website</i> mengenai pelatihan keterampilan	Valid

### c. Pengujian Url *Website* Mengenai Pelatihan Keterampilan

Tabel 6 Pengujian url website mengenai pelatihan keterampilan

Skenario Pengujian				Kesimpulan
No	Kategori	Laman Pengujian		
		URL	Website mengenai Pelatihan Keterampilan	



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

1.	Makanan	<a href="https://posmetropadang.co.id/pelatihan-tataboga-pemberdayaan-sdm-perempuan/">https://posmetropadang.co.id/pelatihan-tataboga-pemberdayaan-sdm-perempuan/</a>	Sesuai	Valid
2.	Makanan	<a href="http://www.beritajakarta.id/read/75287/100-peserta-ikuti-pelatihan-pengolahan-hasil-perikanan">http://www.beritajakarta.id/read/75287/100-peserta-ikuti-pelatihan-pengolahan-hasil-perikanan</a>	Sesuai	Valid
3.	Kerajinan tangan	<a href="http://www.beritajakarta.id/read/77520/warga-rusun-jatinegara-barat-ikuti-pelatihan-kerajinan-tangan#.XrdnL2gzbDc">http://www.beritajakarta.id/read/77520/warga-rusun-jatinegara-barat-ikuti-pelatihan-kerajinan-tangan#.XrdnL2gzbDc</a>	Sesuai	Valid
4.	Kerajinan tangan	<a href="https://jejakrekam.com/2020/06/09/warga-desa-teluk-karya-ikuti-pelatihan-sasirangan-dengan-pewarna-alami/">https://jejakrekam.com/2020/06/09/warga-desa-teluk-karya-ikuti-pelatihan-sasirangan-dengan-pewarna-alami/</a>	Sesuai	Valid
5.	Makanan	<a href="https://infopublik.solokkota.go.id/tingkatkan-hasil-olahan-dinas-pertanian-adakan-pelatihan-pembuatan-kripik-ikan/">https://infopublik.solokkota.go.id/tingkatkan-hasil-olahan-dinas-pertanian-adakan-pelatihan-pembuatan-kripik-ikan/</a>	Sesuai	Valid
6.	Kerajinan tangan	<a href="https://infopublik.solokkota.go.id/pekka-kota-solok-ikuti-pelatihan-menjahit/">https://infopublik.solokkota.go.id/pekka-kota-solok-ikuti-pelatihan-menjahit/</a>	Sesuai	Valid
7.	Makanan	<a href="https://matabanua.co.id/2019/10/30/disnaker-beri-pelatihan-membuat-kue-dan-kerupuk-bagi-perempuan-desa-parigi-kacil/">https://matabanua.co.id/2019/10/30/disnaker-beri-pelatihan-membuat-kue-dan-kerupuk-bagi-perempuan-desa-parigi-kacil/</a>	Sesuai	Valid
8.	Kerajinan tangan	<a href="https://suryaradio.com/2020/02/17/tingkatkan-perekonomian-pemkab-kediri-beri-pelatihan-ketrampilan-kerajinan-tangan-para-ibu-ibu/">https://suryaradio.com/2020/02/17/tingkatkan-perekonomian-pemkab-kediri-beri-pelatihan-ketrampilan-kerajinan-tangan-para-ibu-ibu/</a>	Sesuai	Valid



### © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

9.	Makanan	<a href="https://www.giripanggung-tepus.desa.id/first/artikel/393-Pelatihan-Peningkatan-Kualitas-Produk-Olahan">https://www.giripanggung-tepus.desa.id/first/artikel/393-Pelatihan-Peningkatan-Kualitas-Produk-Olahan</a>	Sesuai	Valid
10.	Makanan	<a href="https://dp3a.semarangkota.go.id/blog/post/pelatihan-boga-kecamatan-tugu-dan-kecamatan-mijen">https://dp3a.semarangkota.go.id/blog/post/pelatihan-boga-kecamatan-tugu-dan-kecamatan-mijen</a>	Sesuai	Valid
11.	Kerajinan tangan	<a href="https://sulsel.sulsatsu.com/2017/10/17/sulsel/gowatamapan/tim-penggerak-pkk-gowa-gelar-pelatihan-pengolahan-limbah.html">https://sulsel.sulsatsu.com/2017/10/17/sulsel/gowatamapan/tim-penggerak-pkk-gowa-gelar-pelatihan-pengolahan-limbah.html</a>	Sesuai	Valid
12.	Makanan	<a href="http://dp3a.semarangkota.go.id/blog/post/pelatihan-kewirausahaan-dinas-pemberdayaan-perempuan-dan-perlindungan-anak">http://dp3a.semarangkota.go.id/blog/post/pelatihan-kewirausahaan-dinas-pemberdayaan-perempuan-dan-perlindungan-anak</a>	Sesuai	Valid
13.	Makanan	<a href="https://www.giripanggung-tepus.desa.id/first/artikel/248-Pelatihan-ekonomi-Produktif-Bagi-Perempuan">https://www.giripanggung-tepus.desa.id/first/artikel/248-Pelatihan-ekonomi-Produktif-Bagi-Perempuan</a>	Sesuai	Valid
14.	Kerajinan tangan	<a href="https://rgol.id/idah-motivasi-ibu-ibu-sulap-daun-jagung-jadi-kerajinan-tangan/">https://rgol.id/idah-motivasi-ibu-ibu-sulap-daun-jagung-jadi-kerajinan-tangan/</a>	Sesuai	Valid
15.	Makanan	<a href="https://jejakrekam.com/2019/08/27/lk3-berdayakan-kaum-perempuan-kelurahan-tanjung-pagar/">https://jejakrekam.com/2019/08/27/lk3-berdayakan-kaum-perempuan-kelurahan-tanjung-pagar/</a>	Sesuai	Valid
16.	Kerajinan tangan	<a href="https://posmetropadang.co.id/perempuan-kepala-keluarga-harus-terampil/">https://posmetropadang.co.id/perempuan-kepala-keluarga-harus-terampil/</a>	Sesuai	Valid



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

17.	Kerajinan tangan	<a href="https://posmetropadang.co.id/dp3ap2kb-kota-payakumbuh-beripelatihan-menjahit/">https://posmetropadang.co.id/dp3ap2kb-kota-payakumbuh-beripelatihan-menjahit/</a>	Sesuai	Valid
18.	Kerajinan tangan	<a href="https://jejakrekam.com/2019/10/19/pelatihan-sasirangan-pewarna-alam-dan-modifikasi-anyaman-bakul-di-alalak-utara/">https://jejakrekam.com/2019/10/19/pelatihan-sasirangan-pewarna-alam-dan-modifikasi-anyaman-bakul-di-alalak-utara/</a>	Sesuai	Valid
19.	Kerajinan tangan	<a href="https://posmetropadang.co.id/warga-kampung-kb-diberipelatihan-membatik/">https://posmetropadang.co.id/warga-kampung-kb-diberipelatihan-membatik/</a>	Sesuai	Valid
20.	Kerajinan tangan	<a href="https://matabanua.co.id/2019/12/20/perempuan-loktabat-utara-dilatih-menjahit/">https://matabanua.co.id/2019/12/20/perempuan-loktabat-utara-dilatih-menjahit/</a>	Sesuai	Valid
21.	Tidak Ada	<a href="https://news.detik.com/berita/d-5109183/jokowi-kini-rapat-virtual-lagi-ini-penjelasan-istana?tag_from=wp_nhl_3">https://news.detik.com/berita/d-5109183/jokowi-kini-rapat-virtual-lagi-ini-penjelasan-istana?tag_from=wp_nhl_3</a>	Tidak Sesuai (menampilkan <i>alert</i> bahwa bukan <i>website</i> mengenai pelatihan keterampilan)	Valid
22.	Tidak Ada	<a href="https://www.liputan6.com/news/read/4276831/cegah-covid-19-anies-terapkan-sistem-ganjil-genap-di-pasar-tradisional?medium=Headline&amp;campaign=Headline_click_1">https://www.liputan6.com/news/read/4276831/cegah-covid-19-anies-terapkan-sistem-ganjil-genap-di-pasar-tradisional?medium=Headline&amp;campaign=Headline_click_1</a>	Tidak Sesuai (menampilkan <i>alert</i> bahwa bukan <i>website</i> mengenai pelatihan keterampilan)	Valid
23.	Tidak Ada	<a href="https://www.liputan6.com/tekn/read/4269681/biar-lebih-kompetitif-6-tipe-hero-mobile-legends-ini-wajib-player-miliki">https://www.liputan6.com/tekn/read/4269681/biar-lebih-kompetitif-6-tipe-hero-mobile-legends-ini-wajib-player-miliki</a>	Tidak Sesuai (menampilkan <i>alert</i> bahwa bukan <i>website</i> mengenai pelatihan keterampilan)	Valid



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

24.	Tidak Ada	<a href="https://www.cnnindonesia.com/hiburan/20200611175700-234-512375/agensi-buka-suara-soal-rumor-song-joong-ki-kencani-pengacara">https://www.cnnindonesia.com/hiburan/20200611175700-234-512375/agensi-buka-suara-soal-rumor-song-joong-ki-kencani-pengacara</a>	Tidak Sesuai (menampilkan <i>alert</i> bahwa bukan <i>website</i> mengenai pelatihan keterampilan)	Valid
25.	Tidak Ada	<a href="https://kumparan.com/kumparan-news/foto-polisi-di-bali-pakai-kostum-celuluk-saat-sosialisasi-pencegahan-corona-1tOtwzh6ANM">https://kumparan.com/kumparan-news/foto-polisi-di-bali-pakai-kostum-celuluk-saat-sosialisasi-pencegahan-corona-1tOtwzh6ANM</a>	Tidak Sesuai (menampilkan <i>alert</i> bahwa bukan <i>website</i> mengenai pelatihan keterampilan)	Valid

d. Pengujian Halaman Detail Informasi Pelatihan Keterampilan Kategori Makanan

Tabel 7. Pengujian halaman detail informasi pelatihan kategori makanan

Skenario Pengujian (Data Normal)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	
Memilih menu daftar pelatihan kategori makanan	Sistem akan menampilkan detail informasi pelatihan keterampilan kategori makanan	Sistem menampilkan detail informasi pelatihan keterampilan kategori makanan	Valid

e. Pengujian Halaman Detail Informasi Pelatihan Keterampilan Kategori Kerajinan Tangan

Tabel 8. Pengujian halaman detail informasi pelatihan kategori kerajinan tangan

Skenario Pengujian (Data Normal)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	
Memilih menu daftar pelatihan kategori kerajinan tangan	Sistem akan menampilkan detail informasi pelatihan	Sistem menampilkan detail informasi pelatihan keterampilan	Valid



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

	keterampilan kategori kerajinan tangan	kategori kerajinan tangan	
--	--	---------------------------	--

f. Pengujian Pencarian Pelatihan

Tabel 9. Pengujian halaman pencarian pelatihan

Skenario Pengujian (Data Normal)			Kesimpulan
Data Masukkan	Hasil yang diharapkan	Hasil Pengujian	
Memasukkan data berupa kata kunci yang di ketahui atau yang di inginkan	Sistem akan menampilkan hasil pencarian yang sesuai dengan kata kunci yang di masukkan	Sistem menampilkan hasil pencarian yang sesuai dengan kata kunci yang di masukkan	Valid

Hasil pengujian dari *alpha testing* menunjukkan bahwa *website* yang sudah dibangun sudah memenuhi persyaratan fungsional. Akan tetapi, pada prosesnya masih memungkinkan untuk terjadi kesalahan. Secara fungsionalitas sistem yang telah di bangun sudah dapat menghasilkan keluaran yang di harapkan.

2. *Beta Testing*

*Beta testing* merupakan tahap lanjutan dari *alpha testing* yang dimana dilakukan secara objektif dengan di uji secara langsung kepada *user*. Pengujian ini dilakukan pada sekitar kawasan rumah penulis menemui perempuan yang sudah berumah tangga dengan mengikuti skenario pada item pengujian.

Tabel 10. Demografi user 1

Demografi User 1	
Jenis Kelamin	Perempuan
Usia	38 Tahun
Status	Menikah
Status Pekerjaan	Mengurus Rumah Tangga

1. Halaman Detail Informasi Pelatihan Keterampilan Kategori Makanan

Aksi : Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Hasil yang diharapkan : Menampilkan detail informasi pelatihan keterampilan kategori 'Makanan'

Status : Berhasil/Gagal

Penguji : User 1

Tabel 11. Halaman detail informasi pelatihan kategori makanan pada user 1

Aksi	Hasil yang diharapkan	Status
Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'	Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'	Berhasil

2. Halaman Detail Informasi Pelatihan Keterampilan Kategori Kerajinan Tangan

Aksi : Mengklik menu 'Daftar Pelatihan' kategori 'Kerajinan Tangan'

Hasil yang diharapkan : Menampilkan detail informasi pelatihan keterampilan kategori 'Kerajinan Tangan'

Status : Berhasil/Gagal

Penguji : User 1

Tabel 12. Halaman detail informasi pelatihan kategori kerajinan pada user 1

Aksi	Hasil yang diharapkan	Status
Mengklik menu 'Daftar Pelatihan' kategori 'Kerajinan Tangan'	Mengklik menu 'Daftar Pelatihan' kategori 'Kerajinan Tangan'	Berhasil

3. Pencarian Pelatihan

Aksi : Menginput kata kunci yang di ketahui atau yang diinginkan

Hasil yang diharapkan : Menampilkan hasil pencarian sesuai dengan kata kunci yang diinput

Status : Berhasil/Gagal

Penguji : User 1

Tabel 13. Pencarian pelatihan pada user

Aksi	Hasil yang diharapkan	Status
------	-----------------------	--------



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Menginput kata kunci yang di ketahui atau yang diinginkan	Menampilkan hasil pencarian sesuai dengan kata kunci yang diinput	Berhasil
---	---	----------

Tabel 14. Demografi user 2

Demografi User 2	
Jenis Kelamin	Perempuan
Usia	32 Tahun
Status	Menikah
Status Pekerjaan	Sudah Bekerja

1. Halaman Detail Informasi Pelatihan Keterampilan Kategori Makanan

- Aksi : Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'
- Hasil yang diharapkan : Menampilkan detail informasi pelatihan keterampilan kategori 'Makanan'
- Status : Berhasil/Gagal
- Penguji : User 2

Tabel 15. Halaman detail informasi pelatihan kategori makanan pada user 2

Aksi	Hasil yang diharapkan	Status
Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'	Mengklik menu 'Daftar Pelatihan' kategori 'Makanan'	Berhasil

2. Halaman Detail Informasi Pelatihan Keterampilan Kategori Kerajinan Tangan

- Aksi : Mengklik menu 'Daftar Pelatihan' kategori 'Kerajinan Tangan'
- Hasil yang diharapkan : Menampilkan detail informasi pelatihan keterampilan kategori 'Kerajinan Tangan'
- Status : Berhasil/Gagal
- Penguji : User 2

Tabel 16. Halaman detail informasi pelatihan kategori kerajinan pada user 2

Aksi	Hasil yang diharapkan	Status
------	-----------------------	--------



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Mengklik menu ‘Daftar Pelatihan’ kategori ‘Kerajinan Tangan’	Mengklik menu ‘Daftar Pelatihan’ kategori ‘Kerajinan Tangan’	Berhasil
--	--	----------

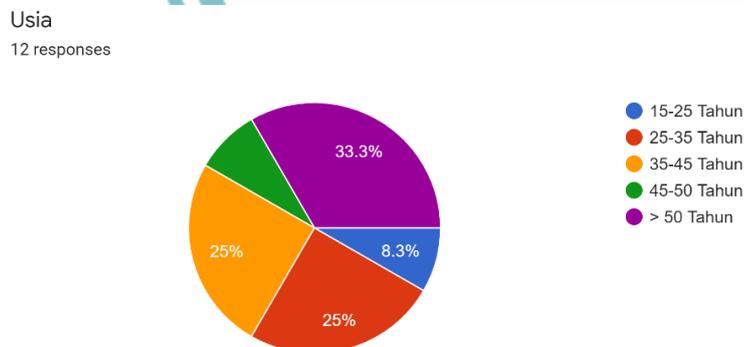
3. Pencarian Pelatihan

Aksi	: Menginput kata kunci yang di ketahui atau yang diinginkan
Hasil yang diharapkan	: Menampilkan hasil pencarian sesuai dengan kata kunci yang diinput
Status	: Berhasil/Gagal
Penguji	: User 2

Tabel 17. Pencarian pelatihan pada user 2

Aksi	Hasil yang diharapkan	Status
Menginput kata kunci yang di ketahui atau yang diinginkan	Menampilkan hasil pencarian sesuai dengan kata kunci yang diinput	Berhasil

Setelah melakukan pengujian ke dua rumah terdekat maka untuk mengetahui apakah *website* ini sesuai dengan kebutuhan kaum perempuan dan memudahkan kaum perempuan dalam mencari informasi mengenai pelatihan keterampilan maka dilakukan survei dengan bantuan video *website* yang dibuat dan menggunakan kuisisioner yang terdiri dari 10 pertanyaan (3 pertanyaan data diri dan 7 pertanyaan kuisisioner). Kuisisioner ini ditujukan kepada kaum perempuan. Adapun hasil dari data kuisisioner yang telah dilakukan sebagai berikut



Gambar 3. 39 Hasil kuisisioner data usia



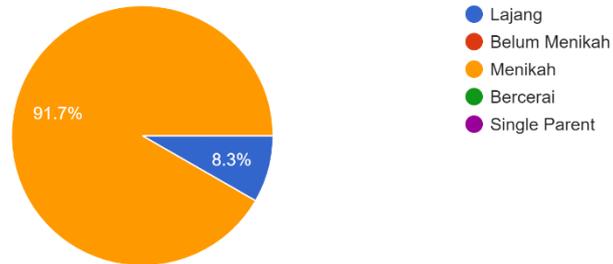
## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penerbitan karya ilmiah, penerbitan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

### Status

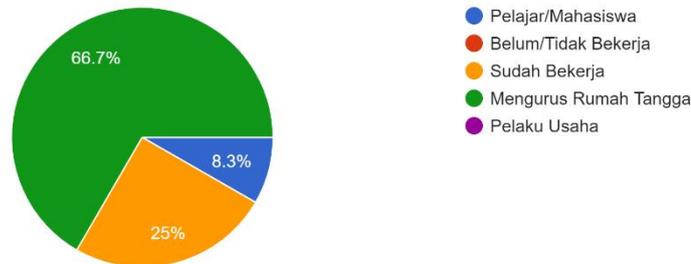
12 responses



Gambar 3. 40 Hasil kuisiner data status

### Status Pekerjaan

12 responses

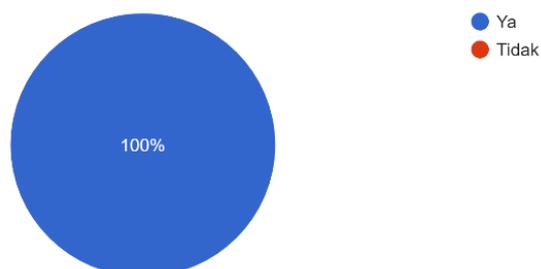


Gambar 3. 41 Hasil kuisiner data status pekerjaan

1. Dari video yang sudah di lihat, apakah fitur beranda dapat menampilkan informasi yang singkat, jelas, dan mudah dimengerti?

Dari video yang sudah di lihat, apakah fitur beranda dapat menampilkan informasi yang singkat, jelas, dan mudah dimengerti?

12 responses



Gambar 3. 42 Hasil kuisiner pertanyaan 1



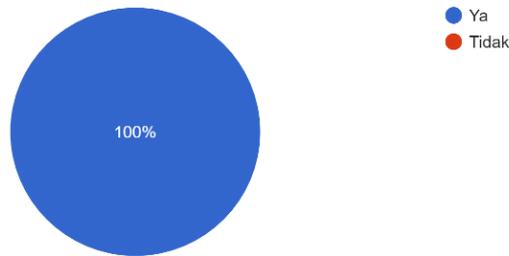
**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

2. Dari video yang sudah di lihat, apakah fitur daftar pelatihan dengan kategori 'Makanan' dapat menampilkan informasi pelatihan keterampilan seputar Makanan?

Dari video yang sudah di lihat, apakah fitur daftar pelatihan dengan kategori 'Makanan' dapat menampilkan informasi pelatihan keterampilan seputar Makanan?

12 responses

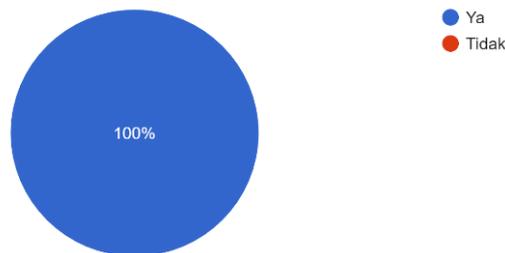


Gambar 3. 43 Hasil kuisisioner pertanyaan 2

3. Dari video yang sudah di lihat, apakah fitur daftar pelatihan dengan kategori 'Kerajinan Tangan' dapat menampilkan informasi pelatihan keterampilan seputar Kerajinan Tangan?

Dari video yang sudah di lihat, apakah fitur daftar pelatihan dengan kategori 'Kerajinan Tangan' dapat menampilkan informasi pelatihan keterampilan seputar Kerajinan Tangan?

12 responses

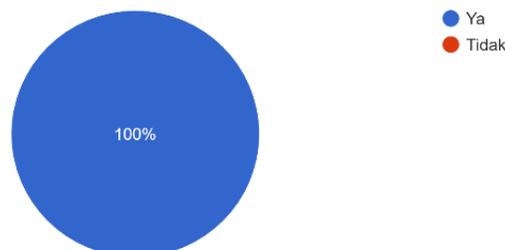


Gambar 3. 44 Hasil kuisisioner pertanyaan 3

4. Dari video yang sudah di lihat, apakah fitur pencarian dapat menampilkan hasil pencarian sesuai dengan kata kunci yang di inginkan?

Dari video yang sudah di lihat, apakah fitur pencarian dapat menampilkan hasil pencarian sesuai dengan kata kunci yang di inginkan?

12 responses



Gambar 3. 45 Hasil kuisisioner pertanyaan 4



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

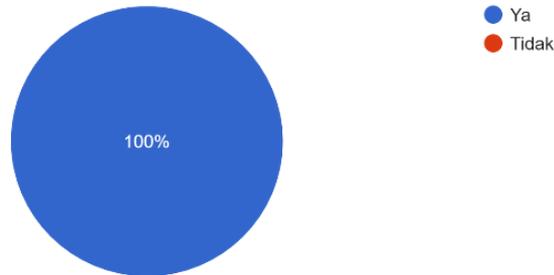
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

5. Apakah website Rumah Asahan mudah untuk digunakan oleh kaum perempuan?

Apakah website Rumah Asahan mudah untuk digunakan oleh kaum perempuan?

12 responses

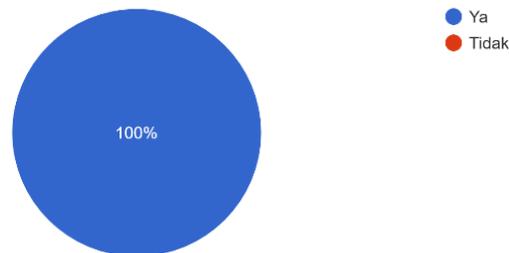


Gambar 3. 46 Hasil kuisisioner pertanyaan 5

6. Apakah website Rumah Asahan bermanfaat dalam mencari dan mengumpulkan berbagai informasi seputar pelatihan keterampilan?

Apakah website Rumah Asahan bermanfaat dalam mencari dan mengumpulkan berbagai informasi seputar pelatihan keterampilan?

12 responses

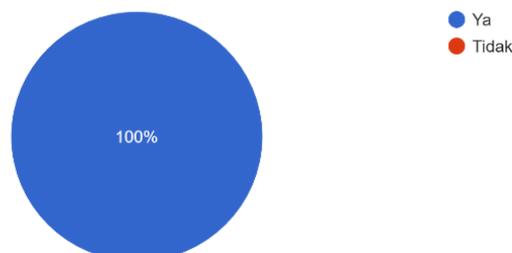


Gambar 3. 47 Hasil kuisisioner pertanyaan 6

7. Apakah website Rumah Asahan membantu kaum perempuan dalam meningkatkan sebuah keterampilan yang belum atau sudah dimiliki?

Apakah website Rumah Asahan membantu kaum perempuan dalam meningkatkan sebuah keterampilan yang belum atau sudah dimiliki?

12 responses



Gambar 3. 48 Hasil kuisisioner pertanyaan 7



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

### 3. Pengujian Metode SVM

Pada pengujian SVM pada data *website* mengenai pelatihan keterampilan ini menggunakan *dataset* untuk menentukan url yang dimasukkan adalah *website* mengenai pelatihan keterampilan atau bukan yang berjumlah 101 data. Dimana dalam menjalankan metode SVM dalam bentuk program menggunakan nilai parameter (C=1.0, class\_weight= 'balanced', multi\_class= 'ovr', random\_state = 40, max\_iter = 10000) dan menggunakan prediksi svm yang didapatkan dari nilai matriks X\_test\_Tfidf. Sehingga akan didapatkan nilai akurasi dengan menggunakan prediksi svm dan y\_test. Dimana pada penelitian ini didapatkan nilai akurasi sebesar 0.975609756097561 atau 97,56% dan dinyatakan valid.

#### 4.2 Analisis Data/Evaluasi

Analisis data ini dilakukan berdasarkan hasil *scraping* yang menampilkan informasi berupa judul, tanggal, dan deskripsi. Dimana dari beberapa *website* yang diambil informasinya menggunakan *web scraping* ternyata tidak semua *website* memiliki tag html yang sama. Sehingga perlu adanya pembaruan data untuk menginput data yang kosong seperti tanggal.

Berdasarkan hasil dari pengujian *alpha testing* yang telah dilakukan perlu dihitung persentase keberhasilan untuk membuktikan fungsionalitas sistem sudah berjalan sesuai dengan *requirement* yang telah ditentukan. Untuk menghitung persentase keberhasilan, bandingkan jumlah skenario yang berhasil dengan seluruh jumlah skenario yang diuji.

$$\text{Persentase Keberhasilan} = \frac{\text{jumlah skenario berhasil}}{\text{jumlah skenario}} \times 100 \%$$

*Alpha testing* yang dilakukan oleh penulis memiliki 6 skenario pengujian. Dimana seluruh pengujian yang dilakukan menunjukkan bahwa seluruh skenario yang diuji telah sesuai dengan yang diharapkan. Berikut adalah persentase keberhasilan pada pengujian *alpha testing*.

$$\text{Persentase Keberhasilan} = \frac{6}{6} \times 100 \% = 100 \%$$



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Dimana persentase keberhasilan pada pengujian *alpha testing* adalah 100%. Sehingga dapat disimpulkan bahwa fungsionalitas sistem berjalan sesuai dengan *requirement* yang telah ditentukan.

Pada pengujian sistem Url *Website* Mengenai Pelatihan Keterampilan, penulis melakukan pengujian pada 20 url *website* yang merupakan *website* mengenai pelatihan keterampilan. Sehingga perhitungan persentase keberhasilan pada setiap url yang dimasukkan adalah *website* mengenai pelatihan keterampilan, yaitu

$$\text{Persentase Keberhasilan} = \frac{20}{20} \times 100 \% = 100 \%$$

Dimana persentase keberhasilan pada pengujian sistem url *website* mengenai pelatihan keterampilan adalah 100% yang menunjukkan bahwa pengujian sistem telah berjalan dengan baik yang dimana sistem telah mengetahui url yang dimasukkan adalah *website* mengenai pelatihan keterampilan atau bukan.

Berdasarkan hasil dari pengujian *beta testing* yang telah dilakukan pada sekitar kawasan rumah penulis dan hasil kuisioner menyatakan bahwa

1. Dengan menggunakan *website* ini kaum perempuan dapat menemukan pelatihan keterampilan yang cocok berdasarkan informasi yang didapatkan karena dari kedua *user* yang ditemukan lebih tertarik kepada pelatihan keterampilan kategori kerajinan tangan.
2. Dengan menggunakan *website* ini kaum perempuan dapat mengetahui kegiatan apa saja yang dilakukan mengenai pelatihan keterampilan di berbagai daerah karena dari kedua *user* yang ditemukan baru mengetahui ada pelatihan keterampilan yang diadakan di masing-masing daerahnya dan mereka terbantu akan adanya informasi mengenai pelatihan keterampilan.
3. Dengan menggunakan *website* ini kedua *user* yang ditemukan mempunyai kesempatan untuk memulai mengembangkan keterampilan masing-masing karena pada dasarnya kedua *user* yang ditemukan pernah memiliki dasar dalam membuat bahan mentah menjadi sebuah produk.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4. Dari hasil kuisioner yang dilakukan melalui *google form* terdapat 12 responden yang memiliki usia lebih dari 50 tahun sebanyak 33%, usia sekitar 25-35 dan 35-45 tahun sebanyak 25%, status yang dimana 92% telah menikah, dan status pekerjaan yang dimana 67% mengurus rumah tangga. Sehingga dapat dinyatakan bahwa responden telah sesuai dengan data yang didapatkan pada saat wawancara.
5. Dari hasil kuisioner yang dilakukan melalui *google form* 12 responden menyatakan bahwa fitur beranda, fitur daftar pelatihan baik kategori makanan maupun kerajinan tangan, dan fitur pencarian pelatihan dapat berjalan dengan baik dan berhasil. Dimana *website* pelatihan keterampilan khusus perempuan (Rumah Asahan) ini mudah digunakan oleh kaum perempuan, bermanfaat dalam mencari dan mengumpulkan berbagai informasi seputar pelatihan keterampilan, dan membantu kaum perempuan dalam meningkatkan sebuah keterampilan yang belum atau sudah dimiliki.

POLITEKNIK  
NEGERI  
JAKARTA



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Penelitian dari “Implementasi Teknologi *Text Mining* dan *Web Scraping* untuk Mengumpulkan Informasi pada *Website* Pelatihan Keterampilan Khusus Perempuan” telah berhasil dilakukan. Aplikasi *website* yang dibangun dapat memudahkan kaum perempuan mendapatkan informasi mengenai pelatihan keterampilan pada satu sumber informasi, dapat menemukan pelatihan keterampilan yang cocok berdasarkan informasi yang didapatkan, dapat mengetahui kegiatan yang dilakukan mengenai pelatihan keterampilan. Fitur-fitur yang ada pada *website* adalah sebagai berikut:

1. Dengan adanya halaman admin, membantu admin untuk mengelola data informasi *website* mengenai pelatihan keterampilan sehingga admin dapat mengetahui terlebih dahulu mana yang *website* mengenai pelatihan keterampilan atau bukan dengan memasukkan url yang telah dilatih menggunakan *text mining* metode SVM (*Support Vector Machine*) dan informasi yang dapat di ambil menggunakan *web scraping*.
2. Fitur Daftar Pelatihan membantu kaum perempuan dalam memilih kategori pelatihan yang tersedia.
3. Halaman detail informasi pelatihan membantu kaum perempuan melihat detail informasi pelatihan sesuai dengan kategori yang dipilih, dimana kaum perempuan dapat menemukan informasi yang diinginkan.
4. Fitur pencarian pelatihan membantu kaum perempuan dalam melakukan pencarian pelatihan berdasarkan kata kunci yang diketahui atau yang diinginkan.
5. Nilai akurasi yang didapatkan dalam mengklasifikasikan url *website* apakah masuk ke dalam *website* pelatihan keterampilan atau bukan menggunakan SVM adalah 97,56%



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### 5.2 Saran

Aplikasi *Website* Pelatihan Keterampilan Khusus Perempuan (Rumah Asahan) ini masih memiliki banyak kekurangan dan dapat dilakukan perbaikan. Berikut ini beberapa saran terkait dengan pengembangan aplikasi selanjutnya, yaitu:

1. Setiap data yang telah di *scraping* dapat lebih di bersihkan lagi walaupun berada di dalam tag html yang sama.
2. Dapat melakukan pengambilan informasi lebih banyak lagi untuk mendukung informasi yang diambil lebih lengkap.
3. Penambahan fitur yang berhubungan dengan sumber *website* asli yang di ambil informasinya.



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## DAFTAR PUSTAKA

- Ayani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi*, 257-262.
- BPS, B. P. (2019, Oktober 7). *Tingkat Pengangguran Terbuka Menurut Jenis Kelamin (Persen), 2015 - 2018*. Retrieved from bps.go.id: <https://www.bps.go.id/dynamictable/2019/10/07/1690/tingkat-pengangguran-terbuka-menurut-jenis-kelamin-persen-2015---2018.html>
- Clinton, R. M., & Sengkey, R. (2019). Purwarupa Sistem Daftar Pelanggaran Lalulintas Berbasis Mini-Komputer Raspberry Pi. *Jurnal Teknik Elektro dan Komputer*, 181-192.
- data-miner.io. (2019, March 8). <https://chrome.google.com/webstore/detail/data-scraping-easy-web-scr>. Retrieved from Data Scraper - Easy Web Scraping: <https://chrome.google.com/webstore/detail/data-scraping-easy-web-scr/nndknepjnlbdbbepjfgmncbggmopgden>
- dsn, C. (2018, November 16). *Algoritma Support Vector Machine (SVM)*. Retrieved from <https://cahyadsn.phpindonesia.id>: <https://cahyadsn.phpindonesia.id/extra/svm.php>
- Febriana, R. (2016). Identifikasi Komponen Model Pelatihan Pedagogi Untuk Meningkatkan Profesionalitas Calon Guru Kejuruan. *Jurnal Pendidikan Teknologi dan Kejuruan*, 79-89.
- Ferdianto, F. F., Yahya, W., & Dewi, R. K. (2018). Pengembangan Sistem Monitoring Aktivitas Jaringan pada Mikrokomputer. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 768-775.
- Hanafri, M. I., Triono, & Luthfiudin, I. (2018). Rancang Bangun Sistem Monitoring Kehadiran Dosen Berbasis Web Pada STMIK Bina Sarana Global. *Jurnal Sisfotek Global*, 8(1), 81-86.
- Herliana, A., & Rasyid, P. M. (2016). Sistem Informasi Monitoring Pengembangan Software Pada Tahap Development Berbasis Web. *Jurnal Informatika*, III(1), 41-50.
- Hidayat, D. (2016). Dampak Pelatihan Keterampilan Hidup (Life Skills) Montir Otomotif terhadap Kesempatan Kerja dan Pendapatan Warga Belajar Dampak Pelatihan Keterampilan Hidup (Life Skills) Montir Otomotif terhadap Kesempatan Kerja dan Pendapatan Warga Belajar. *Jurnal Ilmiah VISI PPTK PAUDNI*, 81-98.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

- Hidayatullah, A. F., & Ma'arif, M. R. (2016). Penerapan Text Mining dalam Klasifikasi Judul Skripsi. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2016* (pp. A-33-A-36). Yogyakarta: Universitas Islam Indonesia.
- Huda, M., Sulistiowati, & Maulana, Y. M. (2016). Rancang Bangun Aplikasi Training Needs Analysis Berdasarkan Kompetensi Pada PT.Tunggal Djaja Indah. *JSIKA*, 1-8.
- Ihsan, M. (2020, Maret 19). Pelatihan Keterampilan Khusus Perempuan. (N. Yunita, Interviewer)
- Mitra, V., Sujaini, H., & Putra Negara, A. B. (2017). Rancang Bangun Aplikasi Web Scraping untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)* , 36-41.
- Mutawalli, L., Zaen, M. T., & Bagye, W. (2019). Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto). *JIRE (Jurnal Informatika & Rekayasa Elektronika)* , 43-51.
- Parsehub. (2019, Januari 1). <https://www.scrapehero.com/>. Retrieved from Best Free and Paid Web Scraping Tools and Software: <https://www.scrapehero.com/top-free-and-paid-web-scraping-tools-and-software/>
- Pramono, N. A., Nurdiansyah, M. H., & Hidayatullifa, D. Z. (2019). Rancang Bangun Pembuatan Sistem Pengiriman Sensor Secara Real Time Menggunakan Python dan Raspberry Pi . *Risalah Fisika*, 43-46.
- Pratiwi, S. (2016). Rancangan Model Pelatihan Sumber Daya Manusia Berbasis E-Training Dalam Rangka Implementasi Learning Organization (Organisasi Pembelajar). *Jurnal Manajemen Pendidikan*, 1234-1243.
- Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scraping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi Laman Web Tentang Hidroponik. *Indonesian Journal of Information Systems (IJIS)*, 25-33.
- Rahayu, A. (2019). Penanggulangan Pengangguran Dengan Pelatihan Keterampilan Menjahit. *Jurnal Pendidikan Luar Sekolah*, 38-45.
- Samsudiney. (2019, Juli 25). *Penjelasan Sederhana tentang Apa Itu SVM?* Retrieved from [medium.com: https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02](https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02)
- Sanjaya, R., & Hesinto, S. (2017). Rancang Bangun Website Profil Hotel Agung Prabumulih Menggunakan Framework Bootstrap. *Jurnal Teknologi dan Informasi*, 57-64.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

- Sasmito, G. W. (2017). Penerapan Metode Waterfall Pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 6-12.
- Septiawan, T. D., Kridalukmana, R., & Windasari, I. P. (2016). Perancangan dan Implementasi Presensi Digital Guru dan Karyawan SMA Negeri 9 Semarang. *Jurnal Teknologi dan Sistem Komputer*, 215-222.
- Sulistiyan, S. R., Setyawan, F. A., Sivam, K., & Purwiyanti, S. (2019). Alat Identifikasi Jenis Daging dengan Pengolahan Citra Digital Menggunakan Python 2.7 dan OpenCV Berbasis Raspberry Pi 3. *Jurnal Matrix*, 54-60.
- Telaumbanua, M. (2018). Peran Ibu Rumah Tangga Dalam Meningkatkan Kesejahteraan Keluarga. *Sosio Informa*, 418-436.
- Ummatin, K. (2019, September). [www.seputarpengertian.co.id](http://www.seputarpengertian.co.id). Retrieved from Pengertian Keterampilan, Menurut Para Ahli dan Macamnya: <https://www.seputarpengertian.co.id/2019/09/keterampilan.html>
- Utari, D. S. (2019, Juli 9). *Machine Learning Support Vector Machine (SVM)*. Retrieved from medium.com: <https://medium.com/@16611030/machine-learning-support-vector-machine-svm-2a2376c53dae>
- webscraper.io. (2019, August 23). <https://chrome.google.com/webstore/detail/web-scraper>. Retrieved from Web Scraper: <https://chrome.google.com/webstore/detail/web-scraper/jnhgnonknehpejjnehehlkklplmbmhn?hl=en>

POLITEKNIK  
NEGERI  
JAKARTA

## DAFTAR RIWAYAT HIDUP PENULIS

Nisya Yunita



Lulusan dari SDN 09 Pagi Cibubur tahun 2010, SMPN 258 Jakarta Timur pada tahun 2013, dan SMAN 106 Jakarta pada tahun 2016. Saat ini sedang menempuh pendidikan Diploma IV Program Studi Teknik Informatika, Jurusan Teknik Informatika dan Komputer Politeknik Negeri Jakarta.

© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## Lampiran 1. Transkrip Wawancara

### Data Diri Responden

Nama : Muhammad Ihsan  
 Jabatan : Asisten Deputi Kesetaraan Gender Bidang Ekonomi  
 Kementerian Pemberdayaan Perempuan dan Perlindungan Anak

### Pertanyaan dan Jawaban Wawancara

1. Perempuan mampu untuk mengembangkan potensi dirinya melalui pelatihan, lalu bagaimana kaum perempuan mendapatkan informasi pelatihan untuk perempuan dapat bisa mengembangkan potensinya?

Jawab: Jika dibilang mampu, pasti mampu. Tapi jika bagaimana mendapatkan informasi pelatihan biasanya disampaikan melalui *whatsapp* atau pendamping yang merupakan warga untuk diberitahu akan adanya pelatihan. Jika pelatihan tersebut akan diadakan lagi, informasi secara langsung disampaikan pada saat pelatihan.

2. Badan Pusat Statistik menyatakan tingkat pengangguran terbuka pada perempuan tahun 2015-2018 berkisar 5—6,5%. Menurut Pak Ihsan apa yang menyebabkan tingkat pengangguran terbuka pada perempuan memiliki persentase yang cukup tinggi?

Jawab: Hal tersebut terjadi karena perempuan biasanya tidak boleh dalam mencari nafkah dan lebih kepada mengurus rumah tangga saja dalam artian memang kultural sebagai perempuan dan karena perempuan mempunyai informasi yang lebih minim daripada laki-laki.

3. Pelatihan pelatihan apa saja yang dibutuhkan untuk perempuan?

Jawab: Pelatihan yang dilaksanakan biasanya pelatihan tentang usahanya sendiri untuk produk yang dihasilkan dapat lebih baik lagi dan mempunyai nilai tambah baik dari segi kemasan dan *branding*.

4. Apa saja kategori untuk pelatihan keterampilan perempuan?

Jawab: Biasanya lebih kepada makanan seperti bahan mentah atau bahan baku yang menjadi produk yang punya nilai, seperti halnya bahan baku singkong yang dapat diubah menjadi keripik. Selain itu, kerajinan tangan seperti tenun.



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

(lanjutan)

5. Apakah ada target berapa banyak perempuan yang dapat mengikuti pelatihan keterampilan dan usia yang boleh mengikuti pelatihan keterampilan?

Jawab: Disini saya jelaskan terlebih dahulu bahwa pelatihan keterampilan ini disebut sebagai model pengembangan Industri Rumahan yang telah dilaksanakan selama 3 tahun semenjak tahun 2016 akhir – 2019 akhir yang ditujukan kepada pelaku usaha yang usahanya dilakukan di rumah. Dimana kami tidak memasang target tapi yang telah mengikuti pelatihan ini adalah 3.764 pelaku usaha yang tersebar di 21 kabupaten/kota se-Indonesia dengan rata-rata usia 30-50 tahun.

6. Bagaimana KemenPPPA mengajak kaum perempuan untuk mengikuti sebuah pelatihan keterampilan?

Jawab: Dengan memberikan motivasi yang dimana dengan adanya pelatihan keterampilan dapat meningkatkan pendapatan keluarga.

7. Informasi seperti apa yang diberikan atau disampaikan pada saat pelatihan keterampilan?

Jawab: Informasi yang diberikan biasanya bagaimana pelaku usaha dapat mengembangkan usaha lebih baik lagi seperti cara membuat makanan atau kerajinan untuk bisa menjadi nilai tambah, memberikan informasi terhadap sumber daya usaha ekonomi produktif, pemberian sarana prasarana/pelatihan, pendampingan, pemasaran bisa dalam bentuk kemasan maupun *branding*, dan cara manajemen keuangan sederhana agar adanya peningkatan pendapatan hasil usahanya.

8. Apakah KemenPPPA melakukan kerjasama dengan pihak lain untuk mengadakan pelatihan keterampilan?

Jawab: Untuk pengembangan Industri Rumahan berhasil, KemenPPPA tidak sendiri sehingga adanya kerjasama dengan Pemerintah daerah, Lembaga Masyarakat, dan *Non-Governmental Organization*.

## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



(lanjutan)

9. Berapa persen kaum perempuan yang telah mengikuti pelatihan keterampilan? Berapa persen kaum perempuan yang masih butuh adanya pelatihan keterampilan? Apakah KemenPPPA memiliki data-data yang sudah ikut pelatihan?

Jawab: Untuk persentase tidak dapat dipersenkan karena kami disini hanya memodelkan pengembangan. Untuk data ada, tapi dikarenakan saya tidak memegang data tersebut karna data ada pada tim saya dan saat ini sedang melakukan *work from home*.

10. Apakah Kementerian Pemberdayaan Perempuan dan Perlindungan Anak mempunyai *website* khusus untuk pemberdayaan perempuan atau *website* yang dapat membantu dalam mencari informasi pelatihan? Jika ada, apakah boleh diberitahu untuk link seputar pencarian informasi pelatihan?

Jawab: Ada, kami beri nama SISIR yaitu media informasi industri rumahan yang merupakan *website* untuk komunikasi antar pelaku usaha yang dimana masih dalam tahap pengembangan dan belum *launching* secara resmi.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

POLITEKNIK  
NEGERI  
JAKARTA

## Lampiran 2. Dokumentasi Pengujian



### © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

