



**RANCANG BANGUN APLIKASI PERSENTASE PEMBELIAN
PRODUK PT. INSURTECH TECHNOLOGIES INDONESIA
MENGUNAKAN METODE *WEB SCRAPING* DAN
ALGORITMA CART**

TUGAS AKHIR

Nama : M. Marcelino Hambali

NIM : 41518210010

Pembimbing : Nenden Siti Fatonah, Dr,
M.Kom

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021**



UNIVERSITAS
MERCU BUANA

**RANCANG BANGUN APLIKASI PERSENTASE PEMBELIAN
PRODUK PT. INSURTECH TECHNOLOGIES INDONESIA
MENGUNAKAN METODE *WEB SCRAPING* DAN
ALGORITMA CART**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

UNIVERSITAS

MERCU BUANA

Oleh:

Nama : M. Marcelino Hambali

NIM : 41518210010

Pembimbing : Nenden Siti Fatonah, Dr,
M.Kom

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021**

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518210010

Nama : M. Marcelino Hambali

Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 15 Februari 2022




M. Marcelino Hambali

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : M. Marcelino Hambali
NIM : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Februari 2022



M. Marcelino Hambali

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : M. Marcelino Hambali
NIM : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	√
		Jurnal Nasional Terakreditasi	√		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: JEPIN (Jurnal Edukasi dan Penelitian Informatika)			
	ISSN	: eISSN: 2548-9364 / pISSN: 2460-0741			
	Link Jurnal	: https://jurnal.untan.ac.id/index.php/jepin/index			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA


Nenden Siti Fatonah, Dr, M.Kom

Jakarta, 15 Februari 2022


M. Marcelino Hambali

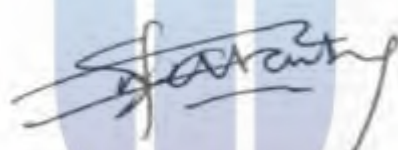
LEMBAR PERSETUJUAN

Nama Mahasiswa : M. Marcelino Hambali
NIM : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian
Produk PT. Insurtech Technologies Indonesia
Menggunakan Metode *Web Scraping* Dan
Algoritma CART

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 15 Februari 2022

Menyetujui,



Nenden Siti Fatonah, Dr. M.Kom
Dosen Pembimbing


UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : M. Marcelino Hambali
Nama : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 15 Februari 2022



Dwi Anindyani Rochman, S.T, M.Ti

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : M. Marcelino Hambali
Nama : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 21 Maret 2022



Umniy Salamah, ST., MMSI

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : M. Marcelino Hambali
Nama : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 15 Februari 2022



Dwiki Jatikusumo, S.Kom, M.Kom

UNIVERSITAS
MERCU BUANA

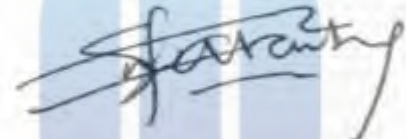
LEMBAR PENGESAHAN

NIM : M. Marcelino Hambali
Nama : 41518210010
Judul Tugas Akhir : Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode *Web Scraping* Dan Algoritma CART

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 15 Februari 2022

Menyetujui,



Nenden Siti Fatonah, Dr, M.Kom
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Emil R. Kaburuan, Ph.D.)
Ka. Prodi Teknik Informatika

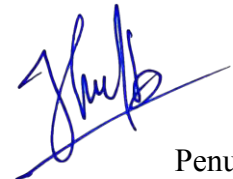
KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah SWT, atas berkat dan rahmatnya penulis dapat melakukan dan menulis laporan penelitian Tugas Akhir yang berjudul “RANCANG BANGUN APLIKASI PERSENTASE PEMBELIAN PRODUK PT. INSURTECH TECHNOLOGIES INDONESIA MENGGUNAKAN METODE *WEB SCRAPING* DAN ALGORITMA CART”. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari pihak terkait, penelitian ini tidak Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Chief Product Officer PT. Insurtech Technologies Indonesia (Aman), **Mr. Le Khanh An.**
2. Chief Executive Officer PT. Insurtech Technologies Indonesia (Aman), **Mr. Steven Tannason.**
3. Bapak **Herry Derajad Wijaya, S.Kom., M.M.**, selaku Kaprodi Teknik Informatika.
4. Ibu **Nenden Siti Fatonah**, selaku dosen pembimbing program kerja praktek.
5. Serta rekan-rekan di PT. Insurtech Technologies Indonesia (Aman).

Akhir kata, penulis berharap sebagai bentuk perbaikan, penulis terbuka pada saran dan masukan dari pembaca.

Jakarta, 20 November 2021



Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS.....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN	v
LEMBAR PENGESAHAN	ix
ABSTRAK	x
ABSTRACT	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	7
BAB 1. LITERATUR REVIEW.....	8
BAB 2. ANALISIS DAN PERANCANGAN.....	22
BAB 3. SOURCE CODE	28
BAB 4. DATASET.....	49
BAB 5. TAHAPAN EKSPERIMEN	52
BAB 6. HASIL SEMUA EKSPERIMEN.....	56
DAFTAR PUSTAKA	59
LAMPIRAN DOKUMEN HAKI.....	62
LAMPIRAN KORESPONDENSI	63

NASKAH JURNAL



JEPIN

(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. 1
No. 1
Desember
2021

Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode Web Scraping Dan Algoritma CART

M.Marcelino Hambali^{*1}, Nenden Siti Fatonah^{**2}

**Jurusan Teknik Informatika Fakultas Ilmu Komputer, Universitas Mercu Buana*

Jalan Meruya Jakarta Barat

141518210010@student.mercubuana.ac.id

2 nendenfatollah@gmail.com

Abstrak— Sebagai upaya menjaring pangsa pasar tentunya diperlukan data yang tepat dan cocok digunakan sebagai target penjualan. Data tersebut dapat diperoleh dari berbagai sumber dengan menggunakan berbagai metode. Di era internet ini, semua data telah tersebar luas dan dipublikasikan di jaringan internet yang tentunya dapat dimanfaatkan oleh PT. Insurtech Technologies Indonesia sebagai data acuan penentu pangsa pasar. Oleh karena itu, pembuatan sistem informasi mengenai persentase pembelian suatu produk berbasis website sangat diperlukan sebagai sarana untuk memperluas informasi pangsa pasar PT. Insurtech Technologies Indonesia. Dalam rangka memenuhi kebutuhan tersebut, maka diperlukan aplikasi berbasis web yang dapat menentukan persentase pembelian produk. Aplikasi web akan dibuat menggunakan framework Django dan dikombinasikan dengan menggunakan library Selenium serta BeautifulSoup. Hasil dari penelitian ini adalah berupa tabel yang berisi informasi mengenai persentase pembelian produk yang dapat ditraining menggunakan algoritma lain..

PT. Insurtech Technologies Indonesia merupakan start-up teknologi yang berjalan pada bidang asuransi yang berbasis di Indonesia. Perusahaan ini didirikan oleh para pemimpin teknologi yang berpengalaman dalam membangun dan menskalakan produk perangkat lunak untuk organisasi dengan pertumbuhan tinggi seperti Google, Traveloka, ByteDance, dan Oriente. Misinya adalah memberdayakan bisnis untuk meningkatkan pengalaman tunjangan karyawan pada suatu perusahaan dengan satu platform sentral yang mudah digunakan. Platform digital modern PT. Insurtech Technologies Indonesia membantu CFO (Chief Financial Officer) dan CHRO (Chief Human Resource Officer) dalam mencari, membeli, dan mengelola asuransi kesehatan untuk karyawan mereka agar lebih mudah.

Dalam rangka menambah calon klien dan mencari pangsa pasar yang cocok, PT. Insurtech Technologies Indonesia memerlukan sebuah aplikasi yang dapat menemukan data perusahaan yang berpotensi menjadi klien dan persentase pembelian produknya oleh perusahaan yang didata. Selanjutnya, data akan dianalisis oleh stakeholder.

Sebagai upaya menjaring pangsa pasar tentunya diperlukan data yang tepat dan cocok digunakan sebagai target penjualan. Data tersebut dapat diperoleh dari berbagai sumber dengan menggunakan berbagai metode. Di era internet ini, semua data telah tersebar luas dan dipublikasikan di jaringan internet yang tentunya dapat dimanfaatkan oleh PT. Insurtech Technologies Indonesia sebagai data acuan penentu pangsa pasar. Dengan memanfaatkan teknologi yang berkembang saat ini, tentunya akan membuat proses pencarian dan pengumpulan data menjadi lebih mudah dan cepat. Salah satu metode yang terkenal dalam pengumpulan dan pengekstrasian adalah web scraping.

Kata kunci— Web Scraping, Selenium, Algoritma CART, Beautiful Soup

1. PENDAHULUAN

Sejak terjadinya pandemi COVID-19 timbul ketidakseimbangan pada sektor ekonomi di Indonesia. Start-up berada di garis depan guncangan ekonomi yang disebabkan oleh pandemi COVID-19. Penerapan pembatasan sosial berskala besar (PSBB) serta penguncian serentak (lockdown) telah menghentikan aktivitas ekonomi secara tiba-tiba, yang mengakibatkan penurunan permintaan dan mengganggu rantai pasokan di seluruh dunia [5]. Untuk dapat menjaring pangsa pasar dan konsumen memang sudah selayaknya start-up memahami peranan dunia digital atau penjualan dengan memanfaatkan teknologi, seperti yang telah dilakukan oleh PT. Insurtech Technologies Indonesia.

Submitted 20-10-2021; Revised dd mm yy; Accepted dd mm yy

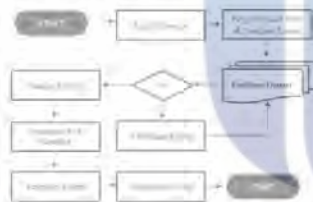
1

Web Scraping merupakan metode penting yang digunakan untuk mengekstraksi informasi tidak terstruktur dari website kemudian mengganti informasi itu menjadi terstruktur. Web scraping juga dikenal sebagai web data extraction, web harvesting atau screen scraping yang merupakan salah satu wujud dari data mining [9]. Hasil yang didapat adalah data yang terstruktur yang dapat disimpan kedalam database, spreadsheets atau comma-separated values (CSV).

Oleh karena itu, pembuatan sistem informasi mengenai persentase pembelian suatu produk berbasis website sangat diperlukan sebagai sarana untuk memperluas informasi pangsa pasar PT. Insurtech Technologies Indonesia. Dalam rangka memenuhi kebutuhan tersebut, maka diperlukan aplikasi berbasis web yang dapat menentukan persentase pembelian produk. Aplikasi web akan dibuat menggunakan framework Django dan dikombinasikan dengan menggunakan library Selenium serta BeautifulSoup. Hasil yang diperoleh dari aplikasi ini akan berformat comma-separated values (CSV) yang dapat disimpan ke database yang kemudian dapat dipergunakan sebagai bahan analisis pada aplikasi utama PT. Insurtech Technologies Indonesia (Aman) yang berbasis web.

II. METODOLOGI PENELITIAN

Secara umum, metodologi yang dilakukan pada penelitian ini terbagi menjadi dua bagian yaitu bagian diagram pengiriman dan diagram penerimaan, diagram pengiriman dapat dilihat pada Gambar 1.



Gambar. 1 Diagram alir metodologi penelitian (pengiriman data gambar)

Penelitian diawali dengan studi literatur terhadap penelitian terdahulu yang juga mengadopsi beberapa metode yang serupa yaitu Web Scraping dan Algoritma CART. Setelah dilakukan peninjauan dan analisis dari penelitian terdahulu penelitian dilanjutkan dengan pengumpulan data yang akan digunakan dalam sistem nanti. Pada saat penerapan sistem peneliti melakukan perancangan dan breakdown sistem terhadap kebutuhan yang diminta oleh PT. Insurtech Technologies Indonesia. Ada beberapa kebutuhan utama yang harus dipenuhi seperti mendapatkan persentase pembelian produk, informasi mengenai data perusahaan dan kategorisasi sebaran dari setiap perusahaan. Perancangan sistem adalah fase yang

terjadi setelah analisis sistem dan persyaratan sistem dilakukan. Perancangan sistem pada penelitian ini menggunakan pendekatan model perancangan data flow diagram (DFD) yang meliputi diagram konteks dan DFD level 0 sehingga diperoleh hasil sebagai berikut: Diagram Kontekstual-Perancangan Sistem dan Login DFD Level 0 adalah proses memasukkan sebuah layanan online yang meliputi nama dan password. Saat ini, halaman login dapat dilakukan di website atau halaman aplikasi sistem dapat mulai masuk secara teknis dalam berbagai cara, dimulai dengan penggunaan email dan nomor ponsel, atau melalui media sosial yang terintegrasi ke dalam sistem. Login melibatkan proses akses bagi pengguna untuk mengakses sistem komputer. Registrasi digunakan untuk mengelola proses identifikasi. *Pengambilan gambar*

Data yang diangkat dalam penelitian ini adalah kumpulan informasi dari setiap perusahaan yang terdaftar situs web Kementerian Perindustrian Republik Indonesia (KEMENPERIN). Sumber data yang digunakan dapat diakses secara umum pada situs web resmi KEMENPERIN melalui tautan berikut <https://kemenperin.go.id/direktori-perusahaan>. Informasi yang diambil dari tautan tersebut adalah nama perusahaan dan jenis komoditi. Kedua informasi tersebut nantinya akan dijadikan data primer yang dipergunakan untuk mencari parameter.

Pada penelitian ini, parameter akan berperan sebagai penentu persentase pembelian produk PT. Insurtech Technologies Indonesia oleh perusahaan-perusahaan yang telah dicatat sebelumnya. Adapun parameter yang diminta adalah jumlah karyawan (headcount), jumlah artikel berdasarkan kata kunci (keyword) dan pendanaan (funding). Parameter tersebut akan dicari menggunakan data primer pada situs-situs web yang ada di Internet. Parameter akan digambarkan pada Tabel dibawah ini.

No.	Parameter	Unit	Contoh Sistem
1.	Registron (RBC)	$R = \frac{RBC}{RBC + RBC} \times 100\%$ $R = \frac{RBC}{RBC + RBC} \times 100\%$ $R = \frac{RBC}{RBC + RBC} \times 100\%$	gambarkan data perusahaan
2.	keyword (KEY)	$K = \frac{KEY}{KEY + KEY} \times 100\%$ $K = \frac{KEY}{KEY + KEY} \times 100\%$ $K = \frac{KEY}{KEY + KEY} \times 100\%$	gambarkan data perusahaan
3.	Funding (FUN)	$F = \frac{FUN}{FUN + FUN} \times 100\%$ $F = \frac{FUN}{FUN + FUN} \times 100\%$ $F = \frac{FUN}{FUN + FUN} \times 100\%$	gambarkan data perusahaan

Setelah parameter dan data yang diperiksa telah valid dan sesuai dengan metode serta tujuan maka akan dilanjutkan dengan merancang struktur web app yang akan dibuat. Metode pengembangan aplikasi yang diterapkan dalam pembuatan web app ini adalah metode Agile Scrum.

Dataset yang akan digunakan dalam penelitian adalah dataset berbentuk table yang diambil dari website resmi Kementerian Perindustrian Indonesia (KEMENPERIN) yang berisi kumpulan informasi mengenai nama-nama perusahaan atau industry yang telah terdaftar resmi di

Indonesia. Berdasarkan sumber yang dipilih, website KEMENPERIN memiliki 33.882 data dan informasi yang dapat diolah seperti gambar 2 berikut



Gambar. 2 Dataset

Data tersebut akan diambil menggunakan metode web scraping dan diolah lebih lanjut menggunakan beberapa algoritma yang dibuat. Hasil akan dimasukan kedalam database dan akan ditampilkan didalam menu History. Sistem akan menghasilkan informasi dalam format .CSV yang berisi informasi-informasi terkait dengan persentase pembelian produk yang dijual oleh PT. Insurtech Technologies Indonesia. Selanjutnya penelitian dilanjut dengan proses training informasi yang telah kita dapat menggunakan algoritma CART. Data yang telah discraping dan disaring berdasarkan parameter akan berbentuk table seperti table 3 berikut.

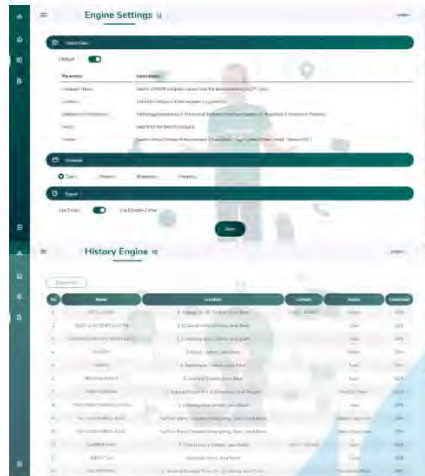
No	Name	Loc	Contact	Sector	Likelihood
1.	KOTA LUDANG	Jl. Kaligepi 11	0231 - 208	Seluhun	32%
2.	"BUDI" (CAP JEM)	Jl. Kh Zainal		Soun	52%
3.	"CHANDRA KIRAI"	Jl. Ki Baleduq		Soun	52%
4.	"GAOS P"	Jl. Buyul, Ciri		Soun	32%
5.	"HASSAN"	Jl. Bandenga		Soun	32%
6.	"IRAWAN/IRWAN"	Jl. Leliung, 1		Soun	32%
7.	"PRIMA/PERINA"	Jl. Suparjo R		Me Dan Si	32%
8.	"ROCHYADI/ DAI"	Jl. Ki Bledug		Soun	32%
9.	"SA" SUGIH AREN	Kp. Pasir Bar		Bahan Sap	32%
10.	"SA" SUGIH AREN	Kp. Pasir Bar		Bahan Sap	42%
11.	"SUMBER SARI"	Jl. Desa Low	0231 - 51k	Soun	32%
12.	(NENG TIA)	Kp. Citorek, G		Cimol	32%
13.	101 PRINTING	Jl. Simpang 2		Perontakar	32%
14.	13 PUTRA, PD	Jl. Bandunga		Garam Kiri	32%
15.	1388 PT	Jl. Yos Sudar		Ikan Kering	32%
16.	160 GOURMET CI	Wijaya 2 No,		Roti Dan K	52%
17.	2 M ADVERTISING	Jl. Ari Rahm		Papan Rek	32%
18.	2 SIDE (ARYANI)	Desa Karang		Pakaian Ja	32%
19.	3 KENCANA CV	Kawasan Ind		Timar	32%
20.	3 SINTERNATION	Kapak Kence		Alat Ukur F	32%
21.	3D NETWORK INI	Jl. Letten S P		Penyedia A	32%

Data yang selesai akan diekspor menjadi format csv akan di training menggunakan algoritma CART yang dimana setiap header atau field yang ada dalam tabel informasi akan dijadikan data independen kecuali Likelihood yang akan dijadikan data dependensi. Hasil training akan berisi pohon keputusan yang dapat menggambarkan penyebaran persentase pembelian berdasarkan jenis komoditi dan daerah. Penelitian ditutup dengan membuat laporan hasil yang dikeluarkan web app. Laporan diharapkan dapat berguna sebagai parameter pengambilan keputusan Stakeholder diwaktu yang akan datang.

III. HASIL DAN ANALISA

Pengembangan web app dimulai setelah metode pengembangan agile dideklarasikan. Langkah awal adalah pembuatan mock-up atau desain UI yang disesuaikan dengan kebutuhan sistem. Adapun hasil UI dari web app digambarkan pada Gambar .

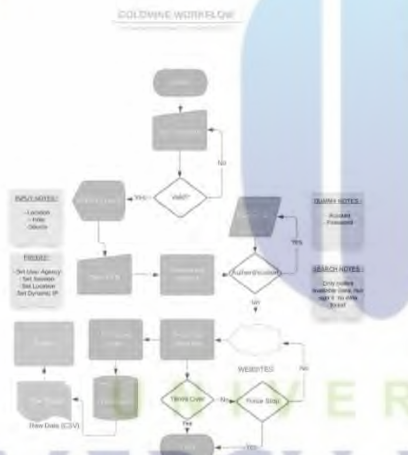




```

@csrf_exempt=1,4,4
@csrf_exempt=2,2,1
@csrf_exempt=2,5
@csrf_exempt=1,5,5
@csrf_exempt=4,9,3
@csrf_exempt=2020,12,5
@csrf_exempt=1,14,5
@csrf_exempt=4,8,8
@csrf_exempt=4,4,4
@csrf_exempt=2,9,2
@csrf_exempt=8,4,8
@csrf_exempt=3,4,8
@csrf_exempt=4,8,8
@csrf_exempt=8,4,1
@csrf_exempt=3,1,7
@csrf_exempt=3,4,8
@csrf_exempt=3,0,12
@csrf_exempt=2,10
@csrf_exempt=5,7,9
@csrf_exempt=9,70
@csrf_exempt=1,5,2
@csrf_exempt=6,2
@csrf_exempt=6,1
@csrf_exempt=3,1,8
  
```

Sistem aplikasi dibangun menggunakan framework Django versi 3.17 dimana konsep arsitektur yang dimiliki framework ini adalah MVT. Django didasarkan pada arsitektur MVT (Model-View-Template). MVT adalah pola desain perangkat lunak untuk mengembangkan aplikasi web. Model akan bertindak sebagai representasi dari data. Model juga mengatur seluruh struktur data logis di balik aplikasi dan akan mewakili database. View adalah pengontrol dari template dan routing seluruh aktifitas yang dilakukan pengguna yang dilihat di browser saat merender situs web. View biasanya berisi algoritma atau logic yang dapat mengubah element dalam HTML, CSS, Javascript dan memiliki ekstensi file .py.



Dalam membangun sistem, peneliti menggunakan beberapa teknologi yang nantinya akan diimplementasikan. Dependensi dan library yang digunakan dalam sistem adalah seperti gambar berikut.

```

Project_Goldmine-main
├── __pycache__
├── migrations
├── static
├── templates
├── __init__.py
├── admin.py
├── apps.py
├── models.py
├── test.py
├── urls.py
├── views.py
  
```

Selain itu peneliti juga menggunakan dependensi Virtual Environment yang akan membungkus seluruh aplikasi dan menjaga lingkungan dari aplikasi atau dependensi lain yang tidak diimplementasikan kedalam sistem. Virtual Environment digunakan untuk memudahkan peneliti dalam tahap development, peneliti hanya perlu mengaktifkan Virtual Environment dan menjalankan Django tanpa khawatir akan perbedaan versi ataupun perbedaan lainya saat dipindahkan kedalam lingkungan atau wadah yang berbeda.

```

.venv
├── bin
├── lib/python3.10/site-packages
├── .gitignore
├── pyvenv.cfg
  
```


Dalam menangkap dan memperoleh seluruh data yang akan digunakan dan diolah didalam sistem, peneliti menggunakan library open-source bernama beautifulsoup dan selenium. BeautifulSoup digunakan untuk melakukan scraping terhadap data yang berada disetiap elemen pada HTML didalam suatu website. BeautifulSoup memudahkan untuk menginjeksi, mencari dan mendapatkan informasi spesifik baik tunggal ataupun secara masal dalam suatu bagian. BeautifulSoup akan berjalan secara otomatis menggunakan algoritma yang telah dibuat. Library Selenium diperlukan untuk menjalankan beautifulsoup secara synchronous dan sesuai dengan line code yang telah dibuat. Selenium akan menjalankan chrome webdriver baik secara background ataupun tidak. Peneliti memecah algoritma dalam beberapa file untuk mencari data menggunakan library BeautifulSoup dan selenium. File pertama yang dibuat adalah scraperFun.py yang berfungsi sebagai scripting pencari seed funding berdasarkan parameter yang telah ditentukan sebelumnya seperti gambar dibawah ini.



File kedua yang peneliti buat adalah scraperHead.py yang berfungsi sebagai pencari data terkait dengan Head Count atau jumlah karyawan yang ada pada suatu perusahaan. Algoritma diatur untuk mencari data didalam website linkedin dan menyaring data yang didapat sesuai dengan parameter yang ditentukan. File scraperHead.py

File ketiga yang dibuat adalah scraperKey.py dimana file ini dibuat untuk mencari keyword atau kata kunci yang berada didalam suatu artikel. Sebelum menjalankan algoritma, peneliti melakukan prompt terhadap user untuk mendownload webdriver-manager untuk melakukan scraping data secara background menggunakan webdriver yang ditambah argument -headless. Kemudian algoritma dilanjutkan dengan melakukan verifikasi dan penyaringan terhadap data yang didapat sesuai dengan parameter dan informasi yang dibutuhkan

File scraping terakhir yang peneliti buat adalah scraperName.py dimana peneliti akan menggunakan algoritma ini untuk mencari data utama yang diambil dari website resmi Kementerian Perindustrian Indonesia (KEMENPERIN). Algoritma dalam file ini merupakan algoritma pertama yang akan dijalankan setelah user menekan tombol scrape now. File scraperName.py mengambil data berupa nama perusahaan yang terdaftar

didalam table yang tersedia. Algoritma yang ada pada file scraperName.py

IV. KESIMPULAN

Penulis mengambil kesimpulan dari hasil penelitian yang dilakukan, telah dihasilkan data atau informasi berupa table yang berisi kumpulan nama perusahaan dan persentase kebolehdajian untuk membeli produk yang dijual oleh PT. Insurtech Technologies Indonesia. Penelitian ini berfokus terhadap pengembangan web aplikasi yang dapat mengambil data sesuai dengan parameter yang telah didefinisikan sebelumnya dan menggunakan teknologi web scraping. Teknologi yang dipakai yakni BeautifulSoup, Selenium dan Algoritma CART

REFERENSI

- [1] I, Worldstats, "Top 20 Countries With The Highest Number of Internet Users," Internet Worldstats, 2018. [Online]. Sumber: <https://www.internetworldstats.com/top20.html>. [Diakses : 20-Apr-2021].
- [2] Yudhanto, Yudha. Information Technology Business Start-Up. Elex Media komputindo, 2019.
- [3] Reis, Eric. "The lean startup." New York: Crown Business 27 (2011).
- [4] Li, Huaqin, and Jinhwan Hong. "Factors influencing consumers' online repurchasing behavior: A review and research agenda." *IBusiness* 5,04 (2013): 161.
- [5] Thaha, Abdurrahman Firdaus. "Dampak covid-19 terhadap UMKM di Indonesia." *BRAND Jurnal Ilmiah Manajemen Pemasaran* 2.1 (2020): 147.
- [6] Awali, Husni. "Urgensi Pemanfaatan E-Marketing Pada Keberlangsungan Umkm Di Kota Pekalongan Di Tengah Dampak Covid-19." *Balanca: Jurnal Ekonomi Dan Bisnis Islam* 2.1 (2020): 1-14.
- [7] Kosasi, Sandy. "Pembuatan sistem informasi penjualan berbasis web untuk memperluas pangsa pasar." *Prosiding Snatif* (2014): 225-232.
- [8] Maulana, Afrizal Aziz, A. Susanto, and D. P. Kusumaningrum. "Rancang Bangun Web Scraping Pada Marketplace di Indonesia." *JOINS (Journal Inf. Syst., vol. 4, no. 1, pp. 41-53 (2019).*
- [9] Saurkar, Anand V., Kedar G. Pathare, and Shweta A. Gode. "An overview on web scraping techniques and tools." *International Journal on Future Revolution in Computer Science & Communication Engineering* 4.4 (2018): 363-367.
- [10] Flores, Veronica Ambassador, Putri Agung Permatasari, and Lie Jasa. "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword." *Majalah Ilmiah Teknologi Elektro* 19.2: 157-162.
- [11] Santoso, Anthony Justinus Andjarwirawan, and Anita Nathania Purbowo. "Penerapan Sistem Channel Management untuk Aplikasi Manajemen Ecommerce pada Pemilik Toko Online Multi Channel." *Jurnal Infra* 7.1 (2019): 197-202.
- [12] Utami, Ema, et al. "K-Nearest Neighbor and Naive Bayes Classifier Comparison for Individual Character Classification on Twitter." *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*. IEEE, 2020.
- [13] Ariyanto, R., C. Rachmad, and A. R. Syulistyo. "Automatically collect alumni data on social media." *IOP Conference Series: Materials Science and Engineering*. Vol. 732, No. 1. IOP Publishing, 2020.
- [14] Siregar, Alan Bangun, Efori Buulolo, and Permanan Ginting. "Pemanfaatan Algoritma Classification And Regression Tress

- (Cart) Untuk Memprediksi Omset Spanduk Pada CV. Moeha advertising." KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) 1.1 (2017).
- [15] Edy Winarno, S. T., M. Eng, and Ali Zaki. Pemrograman Web Berbasis Html 5, php, dan Javascript. Elex Media Komputindo, 2014.
- [16] Kuhlman, Dave. "A Python Book: Beginning Python, Advanced Python, and Python Exercises". Section 1.1. 2012.
- [17] "About Python". Python Software Foundation., second section "Fans of Python use the phrase "batteries included" to describe the standard library, which covers everything from asynchronous processing to zip files.", Retrieved 24 April 2012.
- [18] Zhao, Bo. "Web scraping." Encyclopedia of big data (2017): 1-3.
- [19] Richardson, Leonard. "Beautiful soup documentation." Dosegljivo: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. [Diakses pada: 20 April 2021] (2007).
- [20] Rubio, Daniel. Beginning Django. Apress., 2017.
- [21] Mohri, dkk. Foundations of machine learning. Cambridge: MIT Press. 2012.
- [22] Loh, Wei - Yin. "Classification and regression trees." Wiley interdisciplinary reviews: data mining and knowledge discovery 1.1 (2011): 14-23.
- [23] De'ath, Glenn, and Katharina E. Fabricius. "Classification and regression trees: a powerful yet simple technique for ecological data analysis." Ecology 81.11 (2000): 3178-3192.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan bagian kelengkapan dari artikel jurnal dengan judul “Rancang Bangun Aplikasi Persentase Pembelian Produk PT. Insurtech Technologies Indonesia Menggunakan Metode Web Scraping dan Algoritma Cart”. Dalam kertas kerja ini terdapat semua material hasil penelitian dari tugas akhir yang tidak disertakan dan dimuat dalam artikel jurnal. Pada kertas kerja ini, berisikan: literatur review, penggunaan dataset, lampiran source code, dan juga hasil dari keseluruhan eksperimen yang dijalankan.

