

Analisis Hubungan Sentimen Pelanggan dengan Harga Apartemen Termahal pada Website Pararius

Christofer Miko Lee

00000034222

Program Studi Sistem Informasi
Universitas Multimedia Nusantara
Banten, Indonesia

christofer.miko@student.umn.ac.id

Thendy Rhenaldy

00000033857

Program Studi Sistem Informasi
Universitas Multimedia Nusantara
Banten, Indonesia

t_hendy.rhenaldy@student.umn.ac.id

Nicholas

00000034616

Program Studi Sistem Informasi
Universitas Multimedia Nusantara
Banten, Indonesia

nicholas2@student.umn.ac.id

Lexand Ripai

Silaban 00000041354

Program Studi Sistem Informasi
Universitas Multimedia Nusantara
Banten, Indonesia

lexand.ripai@student.umn.ac.id

Willibrordus Bayu

00000034000

Program Studi Sistem Informasi
Universitas Multimedia Nusantara
Banten, Indonesia

willi.brordus@student.umn.ac.id

Abstract— Apartemen merupakan sebuah tempat tinggal yang dimana letaknya berada pada sebuah bangunan bertingkat. Apartemen sendiri juga memiliki asal usul berasal dari bahasa inggris “apartement”, yang memiliki definisi rumah petak atau flat. Dalam penelitian ini, data yang digunakan merupakan data Perusahaan Pararius yang berisi data tentang rental apartemen di Den Haag, Belanda. Pencarian data dilakukan dengan mengambil data dari suatu website dan kemudian melakukan scraping data menggunakan software “Jupyter Notebook” dengan bahasa pemrograman Python. Setelah itu, data kemudian melewati tahapan extract kedalam format csv. Setelah melakukan tahap extraction, korelasi data akan di cek. Kemudian tahapan text processing dengan model “Random Forest” akan dijalankan untuk menghilangkan huruf “@” pada kolom review, yang kemudian akan dilanjutkan dengan tahapan regresi linear. Langkah terakhir setelah melakukan regresi linear adalah mengecek fitur penting yang digunakan untuk memprediksi apartemen yang memiliki harga paling mahal pada website Pararius.

Kata kunci — Apartemen, Software, Python, Scraping, Jupyter Notebook

I. LATAR BELAKANG

Seperti yang kita ketahui saat ini, tidak sedikit orang memiliki banyak properti di suatu wilayah. Mulai dari tanah, rumah, ruko, gudang, hingga apartemen. Dari beberapa properti tersebut, apartemen merupakan salah satu opsi kedua setelah rumah yang biasanya dibeli oleh orang untuk dijadikan sebagai investasi. Investasi sendiri merupakan proses penempatan sejumlah dana, yang dimana proses penempatan tersebut diharapkan dapat memberikan keuntungan yang banyak di masa mendatang. Salah satu jenis investasi yang sudah tidak asing lagi namun sering didengar adalah investasi jenis properti. Investasi jenis ini juga tentunya dapat menghasilkan keuntungan yang banyak di masa mendatang mengingat pertumbuhan penduduk yang akan terus meningkat sehingga menyebabkan banyaknya permintaan tempat tinggal. Apartemen merupakan sebuah tempat tinggal yang dimana letaknya berada pada sebuah bangunan bertingkat yang telah

menyediakan berbagai fasilitas seperti layaknya rumah pribadi seperti kamar tidur, kamar mandi, tempat duduk, sofa, dan lain sebagainya dalam satu kamar. Apartemen sendiri juga memiliki asal usul berasal dari bahasa inggris *apartemen*, yang memiliki definisi rumah

petak atau flat. Dalam segi status kepemilikan, apartemen juga memiliki beberapa kategori seperti :

1. Apartement Individual (*condominium*)
2. Apartement Milik Bersama (*cooperative*)
3. Apartement Sewa (*rental*)

Pada penelitian ini, kami akan mengambil salah satu kategori status kepemilikan yaitu “Apartement Sewa (*rental*)”, yang dimana penelitian ini akan membahas mengenai hubungan sentiment pelanggan dengan harga apartement termahal yang terletak di Den Haag, Belanda. Sentimen disini merupakan sebuah metode analisis yang merupakan salah satu bagian dari ilmu pembelajaran Text Mining. Untuk mendapatkan suatu keputusan yang valid terhadap suatu permasalahan, maka diperlukannya analisis sentiment. Analisis Sentiment sendiri bertujuan untuk menganalisis suatu polaritas yang akan dibagi menjadi beberapa kategori seperti sentiment positif, negatif, dan netral.

II. STUDI LITERATUR

Pada bab ini memaparkan studi literatur yang digunakan dalam penelitian ini..

A. Web Scraping

Web Scraping merupakan teknik yang digunakan untuk mengambil data atau informasi dari sebuah website secara otomatis. *Web Scraping* bertujuan untuk mencari data tertentu untuk kemudian dikumpulkan dalam satu platform baru. manfaat dari web scraping adalah agar informasi yang dikeruk dapat lebih fokus sehingga lebih mudah dalam mencari suatu data tertentu, dan data yang dikeruk juga memiliki ukuran data yang bervariasi.

B. Data Extraction

Merupakan proses ekstraksi data dari sumber data, dan dimasukkan ke dalam platform penampungan/penyimpanan. Data yang disimpan merupakan data yang terstruktur dan tidak terstruktur.

C. Uji Korelasi

Uji korelasi adalah teknik pengujian dari analisis data yang memiliki fungsi untuk melihat hubungan kesamaan atau korelasi antara variabel (x) dan variabel (y) pada data terkait. Dengan memahami variabel independen dan dependen yang ada, hal ini akan mendukung pengaruh dari hasil pengujian data statistik, khususnya pada variabel yang bersifat kuantitatif. Untuk pengujian yang dilakukan pada penelitian ini menggunakan analisis korelasi pada *Matriks Heatmap* yang menggambarkan hubungan antara data yang diperoleh.

- Matriks Heatmap

Analisis menggunakan Matriks Heatmap adalah proses menganalisa dengan menggunakan hasil data interaksi dari visualisasi pemetaan dengan komposisi warna yang berbeda - beda. Dengan menggunakan Matriks Heatmap peneliti dapat mengetahui semakin tinggi angka pada suatu kelompok data, maka perolehan komposisi warna yang ada pada matriks akan semakin gelap dan begitu juga sebaliknya.

D. Text Processing

- Case Folding

Case folding adalah salah satu bentuk *text preprocessing* yang paling sederhana dan efektif meskipun sering diabaikan. Tujuan dari *case folding* untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima.

- Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token diartikan sebagai "kata" meskipun *tokenize* juga dapat dilakukan pada paragraf maupun kalimat.

- Filtering(Stopword)

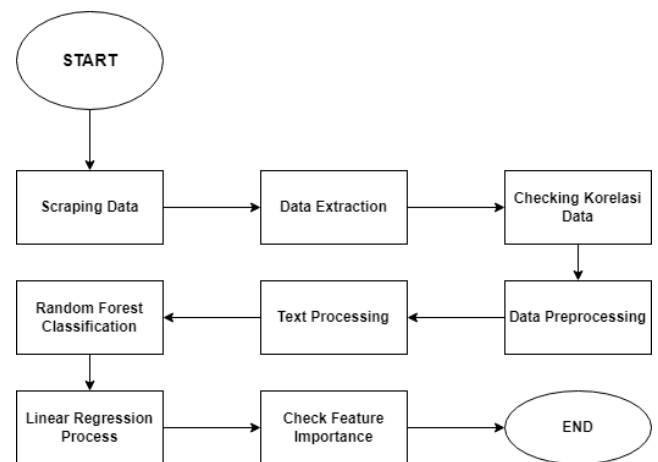
Filtering adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).

- Stemming

Stemming adalah proses menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata (*root word*).

III. METODOLOGI

BERIKUT MERUPAKAN METODOLOGI PENELITIAN :



Gambar 3.1 Alur Penelitian

A. Scraping Data

Data Scraping atau Web Scraping diambil dari halaman website:

<https://www.pararius.com/apartments/den-haag?ac=1> Setelah mendapatkan data dari website Pararius, maka di export dalam bentuk csv.

B. Data Extraction

Data Extraction merupakan langkah dari alur penelitian yang mengambil data dari file .csv yang telah dikeluarkan saat tahap data scraping dengan *library pandas*

C. Checking Korelasi Data

Korelasi Data antara kolom Area dan Kolom Price. Jika berwarna terang maka indikasi korelasi yang rendah. Jika berwarna semakin gelap maka indikasi korelasinya semakin baik.

D. Data Preprocessing

Data preprocessing merupakan salah satu teknik dalam merubah data mentah kedalam bentuk format yang lebih efisien dan bermanfaat. Preprocessing dilakukan karena terdapat data mentah yang tidak lengkap dan format yang kurang konsisten. Maka dari itu, preprocessing dilakukan demi meningkatkan kelengkapan dan akurasi data

yang lebih tersaring.

E. Text Processing

- Pengubahan dengan menghilangkan huruf “@” pada kolom review.
- Pengubahan huruf kapital menjadi huruf kecil.
- Pemisahan teks setelah tanda baca (.)
- Mengambil kata-kata penting (filtering)
- Stemming membuang kata depan sehingga hanya menyisakan kata dasar

F. Melakukan Random Forest Classification

Melakukan klasifikasi sentimen dari Review Hotelnya. Mencari tau Sentiment dari hasil review berapa banyak yang positif, negative, dan netral.

G. Melakukan Regresi Linear

Regresi linear adalah salah satu jenis analisis dengan menghubungkan 2 antar variabel yang terikat. Hal tersebut digunakan untuk menghubungkan 2 variabel dari target prediksinya yaitu harga rental apartment dari Pararius

G. Check Feature Importance

Pada Feature Importance, melakukan prediksi terhadap harga unit dari suatu apartemen yang paling mahal beserta nama, lokasi, dan lain-lain.

IV. HASIL DAN PEMBAHASAN

Berdasarkan dari data apartemen yang telah dilakukan analisis, hasil yang diperoleh adalah sebagai berikut :

A. Web Scrapping

	Id	Date	Tweet
0	1455814114111016962	2021-11-03 08:28:27	"8 op de 10 starters ervaren de huidige ##woni...
1	1448564109549649920	2021-10-14 08:19:31	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...
2	1420647251282649088	2021-07-29 07:27:53	Starters opgelet! 📈 Help jij ons met het in ka...
3	1415224609914462208	2021-07-14 08:20:15	De gemiddelde m2 prijs van #huurwoningen in de...
4	1395383581254488068	2021-05-20 14:19:05	Deze maand in @Vastgoedactueel: waarom het van...

	Name	Location	Area	Price
0	Apartment Damasstraat	2522 TJ Den Haag (Laakwartier-Oos...	75	1175
1	Apartment Doornstraat	2584 AL Den Haag (Geuzenkwartier) ...	26	775
2	Apartment Frankenslag	2582 HW Den Haag (Statenkwartier) ...	60	1050
3	House Zeestraat	2518 DH Den Haag (Willemspark) ...	110	1950
4	Apartment Laan van Meerdervoort	2563 BJ Den Haag (Valkenboskwartie...	37	895

Gambar 4.1 Data Twitter dan webiste hasil web scrapping

Pada tahap *web scrapping*, dilakukan proses ekstraksi data melalui Twitter dan Website Pararius. Dalam ekstraksi data melalui Twitter menggunakan API Twitter, sedangkan ekstraksi data melalui website

dengan library BeautifulSoup untuk mengambil data melalui atribut HTML. Pada ekstrasi data melalui Twitter, periode yang diambil pada tahun 2019-07-01 sampai dengan 2021-11-03. Data yang diambil dari Twitter dan Website Pararius terdapat 30 row dengan 4 variabel.

B. Text Processing (Remove Emoji)

	Tweet
0	"8 op de 10 starters ervaren de huidige ##woni...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...
2	Starters opgelet! 📈 Help jij ons met het in ka...
3	De gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...

Gambar 4.2 Text Processing

	Tweet	No_Emoji
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...

Gambar 4.3 Removal Emoji

	Tweet	No_Emoji	Tweet_Punctuation
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...

Gambar 4.4 Removal Punctuation

	Tweet	No_Emoji	Tweet_Punctuation	Tweet_Clean
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...

Gambar 4.5 Removal HTML Attributes

	Tweet	No_Emoji	Tweet_Punctuation	Tweet_Clean	Tweet_Lower
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	vrije sector huurprijzen weer in de lift! de...
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...

Gambar 4.6 Lower Casing

	Tweet	No_Emoji	Tweet_Punctuation	Tweet_Clean	Tweet_Lower	Tweet_Released
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	vrije sector huurprijzen weer in de lift! de...	vrije sector huurprijzen weer in de lift! de...
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...

Gambar 4.7 Tokenization

[id, data]	Tweet	No_Emoji	Tweet_Punctuation	Tweet_Clean	Tweet_Lower	Tweet_Released	No_Stopwords
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	vrije sector huurprijzen weer in de lift! de...	vrije sector huurprijzen weer in de lift! de...	
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...	
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...	
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...	

Gambar 4.8 Removal Stopwords

	Tweet	No_Emoji	Tweet_Punctuation	Tweet_Clean	Tweet_Lower	Tweet_Released	No_Stopwords
0	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige ##woni...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	"8 op de 10 starters ervaren de huidige woning...	
1	🏠 Vrije sector huurprijzen weer in de lift! 📈 ...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	Vrije sector huurprijzen weer in de lift! De...	vrije sector huurprijzen weer in de lift! de...	vrije sector huurprijzen weer in de lift! de...	
2	Starters opgelet! 📈 Help jij ons met het in ka...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	Starters opgelet! Help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...	starters opgelet! help jij ons met het in kaa...	
3	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	De gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...	de gemiddelde m2 prijs van #huurwoningen in de...	
4	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	Deze maand in @Vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...	deze maand in @vastgoedactueel: waarom het van...	

Gambar 4.9 Stemming

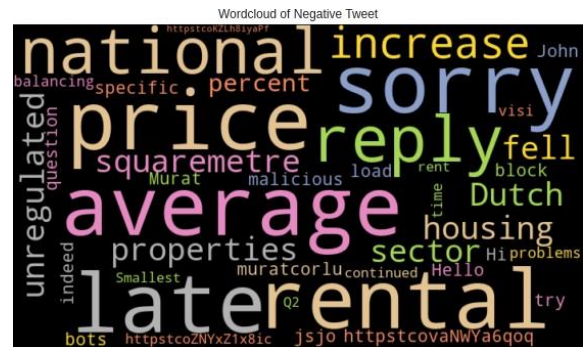
[illegible]

Gambar 4.10 Lemmatization

Pada tahap text preprocessing ini, kami melakukan pengolahan data berdasarkan kalimat, yang dimana pada tahap ini melakukan modifikasi terhadap suatu kalimat dengan menghilangkan emoji-emoji atau atribut yang tidak dibutuhkan oleh suatu data. Setelah itu, kami juga menghilangkan tanda baca atau simbol yang terdapat didalam data tersebut yang dikenal dengan “Removal Punctuation”. Dalam pemrosesan data ini juga, peneliti telah menghilangkan atribut HTML dalam data yang diambil setelah proses web scraping. Kemudian data ini juga dilakukan perubahan pada huruf suatu kalimat, seperti menentukan suatu kalimat kapital atau non-kapital. Pada data ini, peneliti juga melakukan Tokenization yang dimana proses pembagian text yang sebelumnya berupa kalimat menjadi suatu bagian tertentu. Proses pemetaan bentuk dari suatu kata menjadi bentuk kata dasarnya juga dilakukan pada tahap ini, yang dimana tahap tersebut disebut sebagai tahap “Stemming”. Dan pada tahap yang terakhir terdapat tahap “Lemmatization” yang dimana pada tahap ini peneliti mengubah data ke bentuk kata dasarnya.



Gambar 4.13 Visualize Neutral of Sentiment Tweet

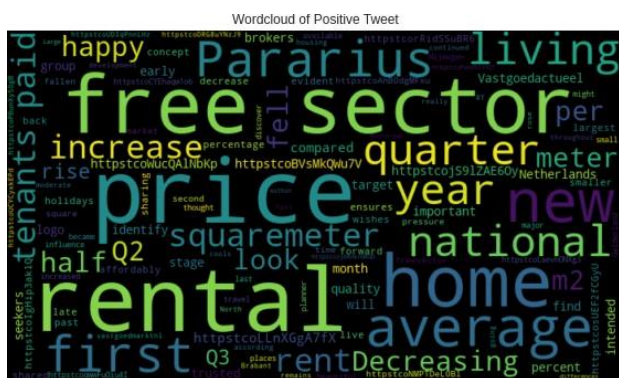


4.14 Visualize Negative of Sentiment Tweet

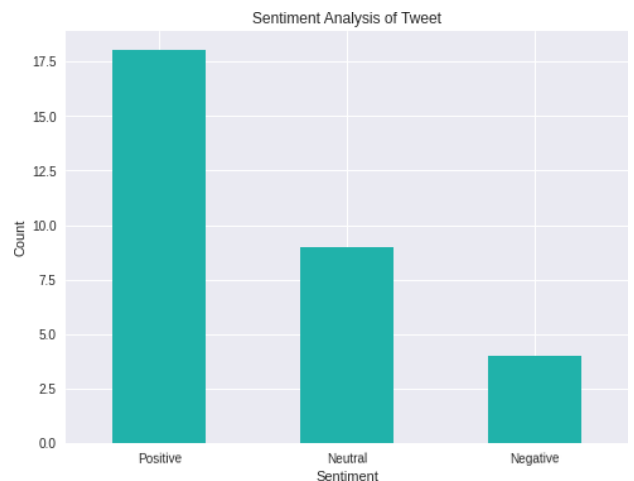
C. Sentiment Analysis

	Tweet_Clean	Subjectivity	Polarity	Sentiment
0	"8 op de 10 starters ervaren de huidige woning...	0.0	0.0	Neutral
1	Vrije sector huurprijzen weer in de lift De ...	0.0	0.0	Neutral
2	Starters opgelet Help jij ons met het in kaart...	0.0	0.0	Neutral
3	De gemiddelde m2 prijs van huurwoningen in de ...	0.0	0.0	Neutral
4	Deze maand in Vastgoedactueel waarom het van b...	0.0	0.0	Neutral

Gambar 4.11 Sentiment Analysis



Gambar 4.12 Visualize Positive of Sentiment Tweet



Gambar 4.15 Visualize Sentiment

Pada tahap ini, kami melakukan suatu analisis terhadap sentimen suatu data, yang dimana hasil sentimen dari website pararius didominasi pada sentiment positive dibandingkan dengan neutral dan negative.

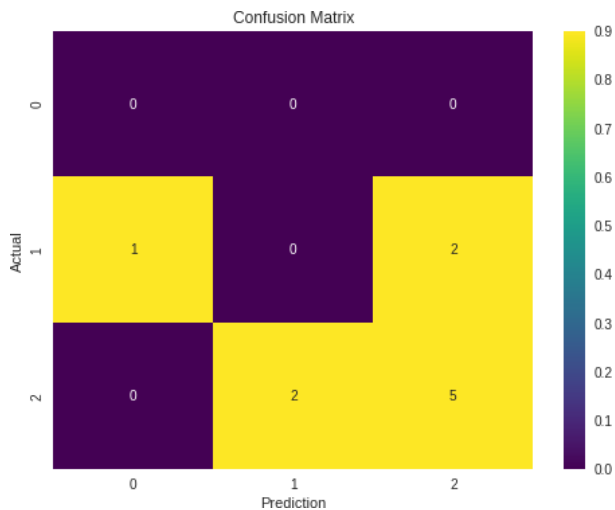
D. Random Forest

Random Forest merupakan gabungan dari tree (masing-masing) yang baik kemudian dikombinasikan menjadi sebuah model. Random Forest akan bergantung kepada nilai vector (random) yang akan didistribusikan kepada semua pohon yang masing-masing decision tree memiliki kedalaman yang maximum.

```
Accuracy Score : 80 %
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 100 out of 100 | elapsed: 0.1s finished
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 100 out of 100 | elapsed: 0.0s finished
```

Gambar 4.16 Random Forest Accuracy Score

Random Forest Accuracy Score ini merupakan akurasi score dari model data training menunjukkan angka akurasi yang tinggi yaitu 0.80 atau 80%.



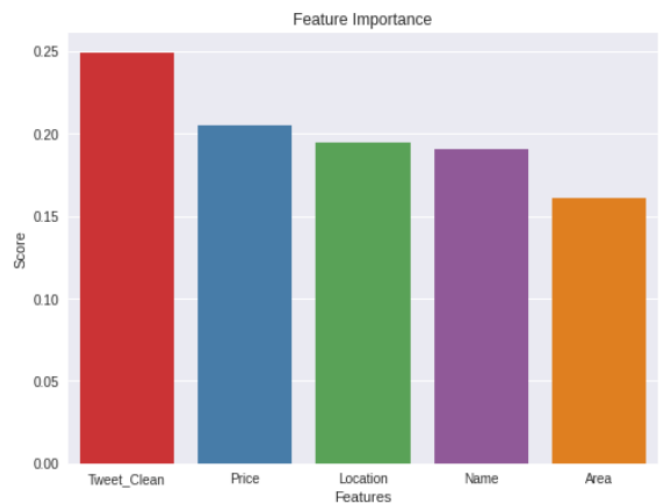
Gambar 4.17 Confusion Matrix

Confusion Matrix merupakan tabel 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Namun pada hasil Confusion Matrix menunjukkan hasilnya cuman satu atau tidak ada korelasi atau sentiment nya netral.

	precision	recall	f1-score	support
Negative	0.00	0.00	0.00	0
Neutral	1.00	0.33	0.50	3
Positive	0.88	1.00	0.93	7
accuracy			0.80	10
macro avg	0.62	0.44	0.48	10
weighted avg	0.91	0.80	0.80	10

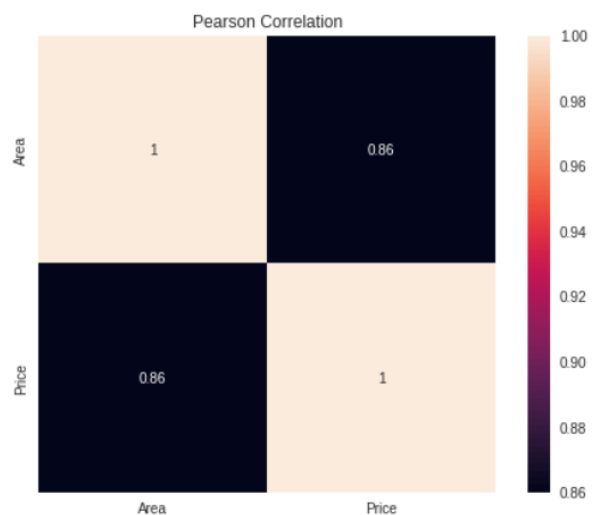
Gambar 4.18 Classification Report

Laporan klasifikasi memiliki manfaat untuk mengukur kualitas prediksi dari sebuah algoritma klasifikasi. Pada hasil presisi, recall, f1-score dan support lebih banyak sentimen positif dengan rata - ratanya 80%



Gambar 4.19 Visualize Feature Importance

Pada gambar 4.19, *feature importance* merupakan hasil dari prediksi model Random Forest. Berdasarkan hasil, dalam klasifikasi sentiment variabel yang memungkinkan untuk menentukan hasil tersebut adalah variabel tweet clean.



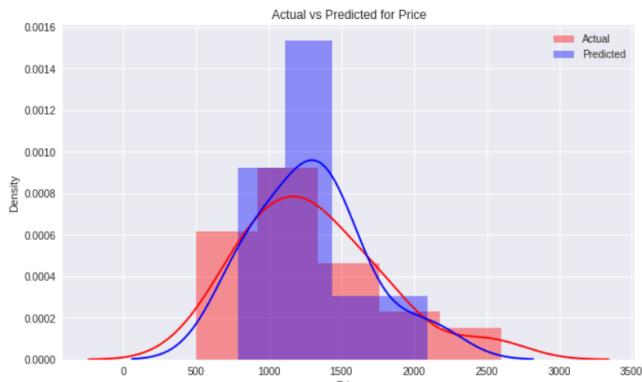
Gambar 4.20 Heatmap Correlation

Pada gambar 4.20 Terdapat correlation range yang menunjukkan angka 1 dikarenakan kotak tersebut saling berkorelasi terhadap variabel yang sama. Sedangkan values 0.86 menunjukkan adanya negatively correlated antara kedua variabel.

```
CPU times: user 4.47 ms, sys: 617 µs, total: 5.09 ms
Wall time: 28 ms
LinearRegression()
```

Gambar 4.21 Build Model

Pada gambar 4.21 terdapat suatu proses permodelan yang dimana pada hasil tersebut menunjukkan bahwa terdapat beberapa informasi yang ditunjukkan, seperti durasi CPU, Wall Time, dan *Linear Regression Score*. Pada perhitungan regresi linear antara X_{test} dan y_{test} .



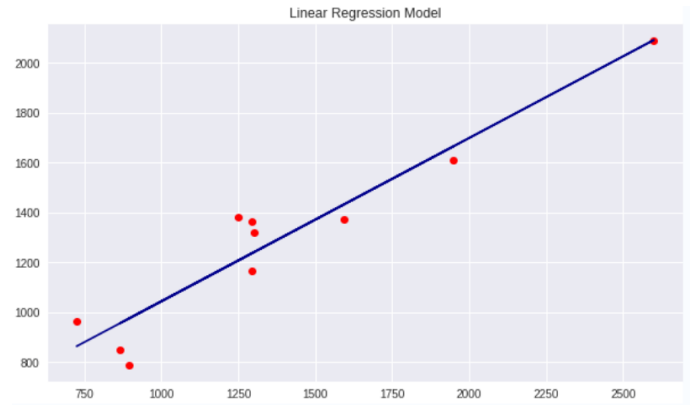
Gambar 4.22 Actual vs Predicted for Price

Pada gambar 4.22, merupakan hasil analisis dari harga yang sebenarnya dan hasil prediksi, dimana kepadatannya lebih tinggi yang harga prediksinya. Hal tersebut berarti hasil prediksi akan meningkat secara signifikan

```
Mean Absolute Error : 178.28
Mean Squared Error : 53190.47
Root Mean Squared Error : 13.352153384379616
R-Squared : 0.8118257339292616
```

Gambar 4.23 Check MAE, MSE, RMSE & R-Squared

Pada gambar 4.23 kami melakukan suatu pengecekan terhadap MAE (Mean Absolute Error), MSE (Mean Squared Error), dan RMSE (Root Mean Squared Error). Pada proses pengecekan ini, kami memperoleh nilai MAE sekitar **178.28**, MSE sebesar **53190.47**, RMSE sebesar **13.35** dan R-Squared sebesar **0.81**



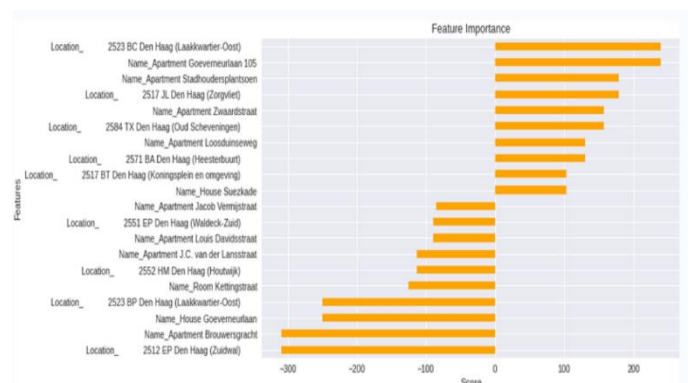
Gambar 4.24 Linear Regression Model

Pada gambar 4.24, merupakan hasil dari model linear regression dari hubungan antar variabel sangat positif menguat, dimana beberapa titik merah mendekati garis biru yang melintang ke atas kanan.



Gambar 4.25 Distribution of Price

Pada gambar 4.25 merupakan hasil distribusi dari harga rental apartment Pararius dimana dari hasil pada gambar diatas, distribusi nya *skew*.



Gambar 4.26 Feature Importance

Pada gambar 4.26, merupakan feature importance dari hasil model linear regression. Hasil analisis menunjukkan harga apartemen termahal di Belanda pada website Pararius

V. KESIMPULAN

Berdasarkan dari penelitian kami, terdapat suatu kesimpulan dimana kami menggunakan data dari Pararius yang diperoleh dengan melakukan web scraping. Berdasarkan dari hasil analisis sentimen, komentar pada tweet pararius menunjukkan bahwa sentimen “positive” lebih mendominasi dari pada sentimen neutral dan negative. Kemudian klasifikasi dari Random Forest menunjukkan nilai akurasi sebesar 80% yang dimana nilai tersebut sangat baik untuk memprediksi sentimen dari data Rental Apartement.. Kemudian hasil regresi linear dari data menunjukkan nilai akurasinya sebesar 81% dimana nilai tersebut cukup baik untuk memprediksi data dari rental apartemen pada Website Pararius.

REFERENCES

- [1] B. H. Mahendra, Adiwijaya, and U. N. Wisesty, “Kategorisasi Berita Multi-Label Berbahasa Indonesia Menggunakan Algoritma Random Forest,” *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9030–9041, 2019.
- [2] Y. W. Syaifudin, A. F. Syafiandini, and H. R. Prasadana, “Aplikasi Pencarian Penjualan Laptop Menggunakan Teknologi Web Scraping,” *J. Inform. Polinema*, vol. 4, no. 4, p. 246, 2018, doi: 10.33795/jip.v4i4.214.
- [3] N. Lihayati, R. E. Pawening, and M. Furqan, “Klasifikasi Jenis Daging Berdasarkan Tekstur Menggunakan Metode Gray Level Coocurent Matrix,” *Pros. SENTIA*, vol. 8, no. 1994, pp. 305–310, 2016.
- [4] A. Darono, “Web Data Extraction Dalam Analitika Data Audit: Pengembangan Artefak Teknologi Dalam Perspektif Design Science Research,” *Teknika*, vol. 9, no. 2, pp. 97–105, 2020, doi: 10.34148/teknika.v9i2.283.
- [5] F. Komunikasi, U. M. Surakarta, J. A. Yani, and T. Pos, “Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest,” *J. Tek. Elektro*, vol. 9, no. 1, pp. 24–29, 2017, doi: 10.15294/jte.v9i1.10452.
- [6] A. I. Permatasari and W. F. Mahmudy, “Pemodelan Regresi Linear dalam Konsumsi Kwh Listrik di Kota Batu Menggunakan Algoritma Genetika,” *DORO*

Kelompok 5 Work Role :

- **Christofer Miko Lee (00000034222)** : Metodologi, Tinjauan Pustaka, Check Feature Importance, Regresi Linear.
- **Lexand Ripai Silaban (00000041354)** : Determine *tinjauan Pustaka*, web scraping & Data extraction.
- **Nicholas (00000034616)** : Research Definitons, Creating Chapter 2 *Tinjauan Pustaka*, Help Subject Definitions and Explanations Task in Group.
- **Thendy Rhenaldy (00000033857)** : Research Format IEEE, Abstrak, Pendahuluan, References, Hasil dan Pembahasan, (Mendeley APA Style), Redraw Alur Penelitian.
- **Willibrordus Bayu (00000034000)** : Text Processing, Random Forest, analysis, and evaluation of data & Metodologi Penelitian, hasil dan pembahasan.

