

Penerapan Text Mining Untuk Visualisasi Fluktuasi Harga Komoditas Pangan (Studi Kasus Twitter, Liputan 6, Detik)

Rakha Beavis Luckyano ¹⁾ Valentinus Roby Hananto ²⁾ Vivine Nurcahyawati ³⁾

Program Studi/Jurusan Sistem Informasi
Institut Bisnis dan Informatika Stikom Surabaya
Jl. Raya Kedung Baruk 98 Surabaya, 60298

Email : 1)beavisrakha@gmail.com, 2)valentinus@stikom.edu, 3)vivine@stikom.edu

Abstract: The fluctuation of food commodity prices will be worsened as the after effect of the global warming. Nowadays, to collect food commodity price, we could search the unstructured data on the online news sites or social media sites which still going to take a lot of time to manually process. This research aim to build a desktop application which will implement text mining technique and jaro-winkler algorithm to process the data received from the source and visualized it on line diagram. Based on the accuracy testing result, the application recieved 75% accuracy

Keywords: Textmining, Python, Jaro-Winkler, Food Commodity, Visualization

Gagalnya produksi komoditas pangan dapat menyebabkan fluktuasi dan meroketnya harga dari komoditas pangan. Meski sudah tersedia alternatif media sosial dan media berita online untuk mengawasi fluktuasi harga tersebut, tetapi data yang ada tidak terstruktur, serta harus dibaca dan dipahami terlebih dahulu.

Penelitian ini bertujuan untuk membangun aplikasi berbasis desktop yang menerapkan teknik text mining dan algoritma Jaro-Winkler kepada Twitter, Liputan 6, dan Detik.com untuk memperoleh dan memvisualisasikan data harga komoditas pangan. penulis juga akan mengambil data pembandingan dari situs Pusat Informasi Harga Pangan Strategis Nasional (PIHPS). Aplikasi akan ditulis dengan bahasa pemrograman python dan data *scrapping* akan dilakukan menggunakan Beautiful Soup.

Objek penelitian adalah 3 komoditas dengan tingkat konsumsi tertinggi per-kategori pada tahun 2016 yang disebutkan didalam buku Statistik Pertanian 2017

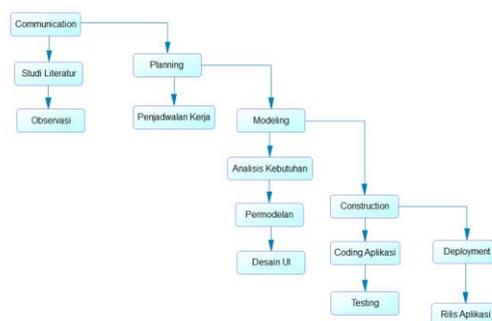
Menurut Lutz (2014) Python adalah bahasa pemrograman *open source* yang memiliki banyak manfaat dan memiliki dukungan untuk struktur pemrograman fungsional dan procedural.

Beberapa penelitian terdahulu yang digunakan sebagai referensi dalam penyusunan penelitian ini antara lain adalah ; penelitian (Saifulloh, 2017) yang menerapkan *natural language processing* untuk menjawab secara otomatis pertanyaan dari calon pelanggan, penelitian (Wardani, 2019) yang menerapkan

web *scrapping* dan metode *naïve bayes* untuk menganalisis sentimen dan melakukan pemeringkatan situs belanja online di Indonesia, Penelitian (Kressa, 2019) yang membangun aplikasi yang menggunakan web *scrapping* dan algoritma jaro-winkler untuk mencari beasiswa di internet.

METODE

Metode yang digunakan dalam pengembangan aplikasi dalam penelitian ini adalah Waterfall. Menurut Pressman(2015) System Development Life Cycle (SDLC) Waterfall adalah metode pengembangan sistem informasi terstruktur yang lebih dikenal juga dengan nama *classic life cycle*. Berikut merupakan gambar langkah-langkah pengembangan SDLC Waterfall



Gambar 1 SDLC Waterfall

Setelah data diekstrak dari sumber data akan mengalami proses *preprocessing* yang terdiri dari beberapa proses berikut.

1. Membersihkan Tag
Proses ini membersihkan tag yang ada didalam string data
2. Case Folding
Proses ini mengubah seluruh teks menjadi huruf kecil
3. Tokenizing
Proses ini memotong data string berdasarkan kata penyusunnya
4. Filtering
Tahap mengambil kata penting dari string

Selanjutnya data akan diseleksi menggunakan algoritma Jaro-Winkler. Menurut Kurniawati, Sulistyono dan Sazali (2010) algoritma Jaro-Winkler adalah salah satu varian dari jaro distance metric yang biasa dimanfaatkan mengukur kemiripan dari dua string. Berikut ini merupakan rumus yang digunakan untuk menghitung Jaro-Winkler distance. Rumus pertama digunakan untuk menghitung jaro distance (dj) antara dua string S1 dan S2 dengan memperhatikan jumlah karakter yang sama persis (m). Untuk lebih jelasnya dapat dilihat dibawah ini:

$$dj = 1/3 (m/S1 + m/S2 + ((m-t)/m)) \tag{1}$$

Keterangan:

- dj =Jaro distance
- m =Jumlah karakter yang sama persis
- S1 =Panjang string 1
- S2 =Panjang string 2
- t =jumlah transposisi

Setelah jarak (dj) antara 2 string ditemukan, langkah selanjutnya adalah menghitung Jaro-Winkler distance (dw). Hal ini dilakukan dengan memperhatikan panjang karakter yang sama sebelum ditemukan ketidaksamaan (l) dan konstanta scalling factor (p) yang menurut Winkler adalah 0,1. Adapun rumus perhitungannya adalah :

$$dw = dj + (l p (1 - dj)) \tag{2}$$

Keterangan:

- dw =Jaro-Winkler distance
- dj = Jaro distance
- l =Panjang karakter yang sama sebelum ditemukan perbedaan
- p = scaling factor (nilai selalu 0,1)

Didalam penerapan algoritma ini, penulis akan memakai modul yang bernama pyjarowinkler yang pada situs dokumentasinya di Pypi.org(2016) dikembangkan menggunakan rumus yang sama seperti diatas.

HASIL DAN PEMBAHASAN

Berikut merupakan langkah-langkah yang dilaksanakan serta hasil dari penelitian ini.

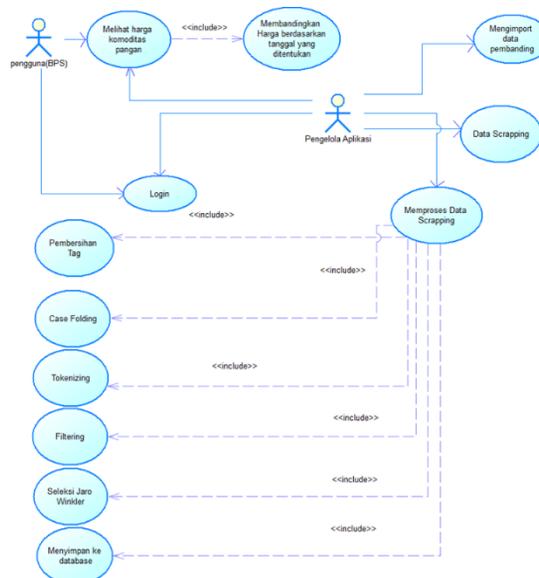
Modeling

Pada tahap ini dilakukan permodelan yang akan dibagi kembali menjadi beberapa tahap sebagai berikut

Analisis Kebutuhan

Untuk menggambarkan kebutuhan pengguna, digambarkan usecase diagram yang dapat dilihat pada gambar 2. Dari gambar diatas dapat disusun kebutuhan fungsional aplikasi yaitu

1. Fungsi Text Mining
Aplikasi dapat melakukan proses text mining yang terdiri dari ; ekstraksi data, pembersihan tag, case folding, tokenizing, filtering, seleksi jaro-winkler, penyimpanan data ke database, import data pembandingan
2. Fungsi Visualisasi Data
Aplikasi dapat memvisualisasikan data yang disimpan pada database.



Gambar 2 Use Case Diagram

Selain itu juga disusun kebutuhan non-fungsional sebagai berikut :

Adapun kebutuhan non-fungsional dari sistem adalah :

- 1 Security (Keamanan)
Fitur login dan logout menggunakan username dan password
- 2 Response Time

Waktu yang dilakukan untuk visualisasi data harga komoditas pangan tidak lebih dari 1 menit

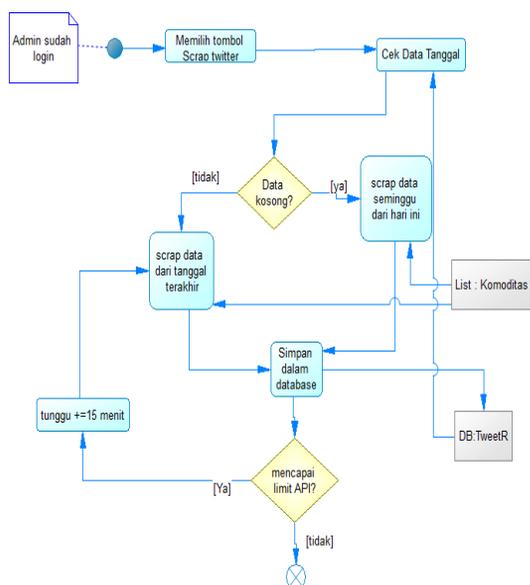
Permodelan

Pada tahap ini disusun permodelan sistem dari aplikasi. Untuk fungsi *scrapping* data akan dibagi menjadi *scrapping* web dan *scrapping* Twitter karena proses yang sedikit berbeda. Tahap pertama adalah penyusunan beberapa activity diagram. Activity diagram disusun untuk menggambarkan aktivitas-aktivitas yang terjadi didalam sistem. Activity diagram dibuat sesuai dengan fungsi yang telah digambarkan pada *use case*, dan yang telah disebutkan pada analisis kebutuhan fungsional. Adapun *activity* diagram yang dibuat dapat dilihat pada tabel berikut.

Tabel 1 Activity Diagram

No	Activity Diagram
1	Login
2	Scrapping web
3	Scrapping Twitter
4	Web processing
5	Tweet processing
6	Import data pembandingan
7	Lihat harga

Berikut merupakan salah satu contoh dari *activity* diagram yaitu *activity* diagram dari fungsi *scrapping* twitter yang dilakukan oleh admin dengan memilih tombol scrap twitter. Untuk lebih jelasnya dapat dilihat pada gambar dibawah ini.



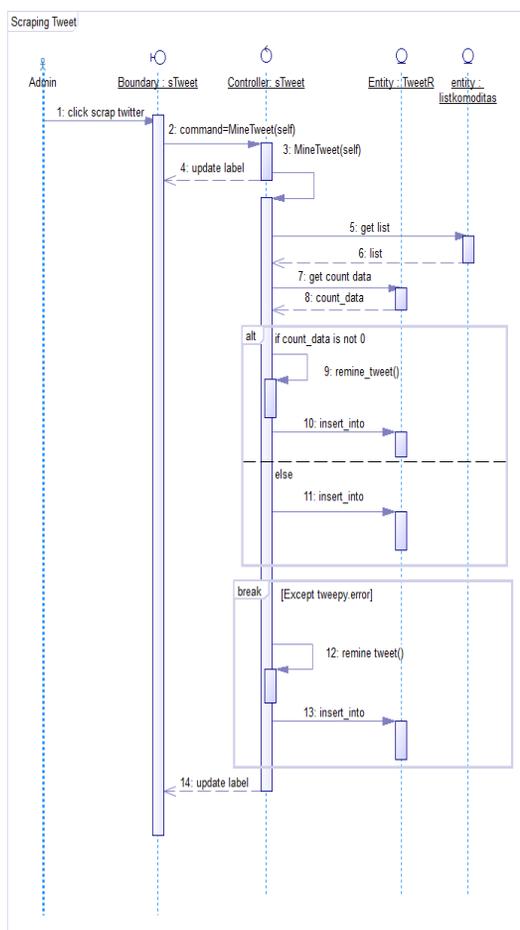
Gambar 3 Activity Diagram Scrapping Twitter

Selanjutnya disusun *sequence* diagram dari *activity* diagram yang sudah dibuat yang dapat dilihat pada tabel dibawah ini.

Tabel 2 Sequence Diagram

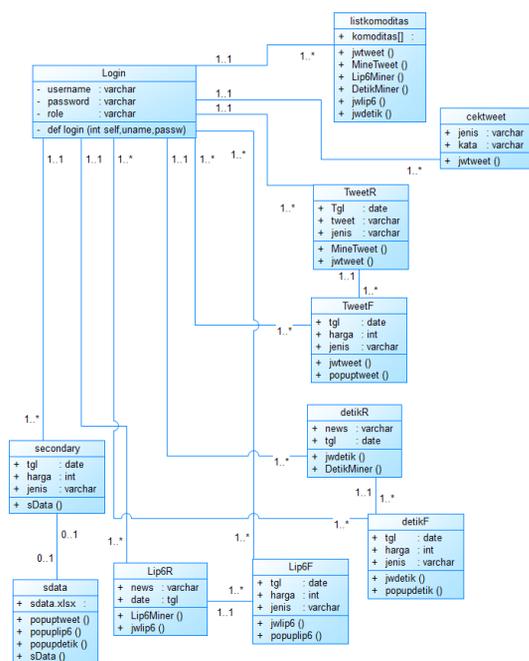
No	Sequence Diagram
1	Login
2	Scrapping web
3	Scrapping Twitter
4	Web processing
5	Tweet processing
6	Import data pembandingan
7	Lihat harga

Salah satu Contoh dari *sequence* diagram yang dibuat merupakan *sequence* diagram yang *scrapping* twitter yang digambarkan pada gambar berikut ini.



Gambar 4 Sequence Diagram Login

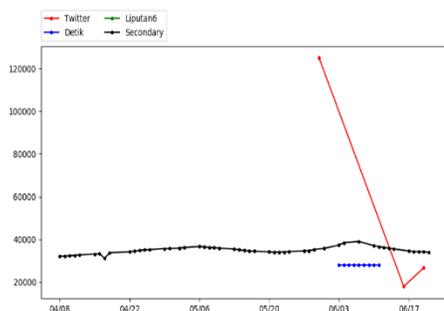
Terakhir adalah penyusunan *class* diagram yang dapat dilihat pada gambar berikut ini



Gambar 5 Class Diagram

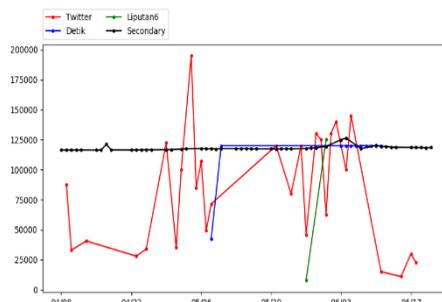
Evaluasi

Dalam sumber data perbandingan, tidak semua komoditas tersedia datanya. Adapun data yang disediakan antara lain adalah daging ayam ras, daging sapi, dan telur ayam ras. Dibawah ini adalah perbandingan antara data daging ayam ras yang didapat dari data perbandingan dan data yang didapat dari aplikasi persumbernya.



Gambar 6. Visualisasi Harga Daging Ayam Ras

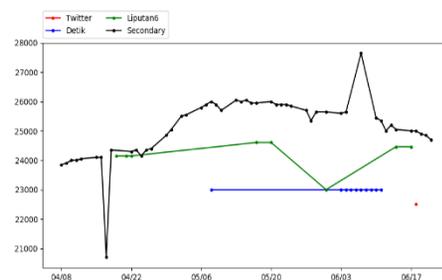
Selanjutnya adalah perbandingan antara data daging sapi yang didapat dari data perbandingan dan data yang didapat dari aplikasi persumbernya



Gambar 7 Visualisasi Harga Daging Sapi

Terakhir merupakan perbandingan antara data telur ayam ras yang didapat dari data perbandingan dan data yang didapat dari aplikasi persumbernya yang dapat dilihat pada gambar 8.

Seperti yang dapat dilihat pada gambar, Twitter memang mempunyai data lebih banyak tapi harga yang terseleksi lebih tidak akurat karena banyaknya fluktuasi yang terlalu tajam dibandingkan data dari media berita online. Hal ini didukung oleh penulisan media sosial yang lebih tidak teratur sehingga memerlukan kamus pengecekan “cek_tweet” pada database yang lebih luas untuk meningkatkan keakuratan seleksi. “cek_tweet” digunakan untuk menampung kata-kata yang membuat sebuah tweet dianggap tidak relevan. Contoh adalah tweet pada jenis komoditas jeruk yang berbunyi “Dijual Jus Jeruk Rp.1000”. Kata “Jus” akan dimasukkan kedalam database “cek_tweet” sehingga tweet tersebut secara otomatis dianggap tidak relevan.



Gambar 8. Visualisasi Harga Telur Ayam Ras

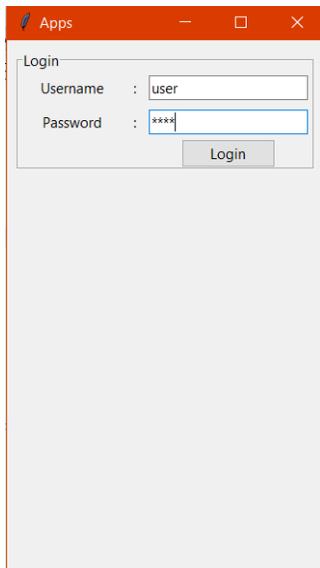
Memang Data yang didapat dari media berita online meskipun lebih stabil tetapi tetap saja lebih sedikit dari media sosial online, hal ini juga dipengaruhi dekatnya waktu pemilu saat pengambilan data sehingga banyak berita yang berisi ekonomi politik dibandingkan berita yang mengandung data harga komoditas.

Untuk mengetahui rata rata harga setiap komoditas yang didapat pada setiap sumbernya termasuk sumber data pembanding dapat dilihat pada gambar berikut ini.

Tabel 2. Rata-Rata Harga Komoditas

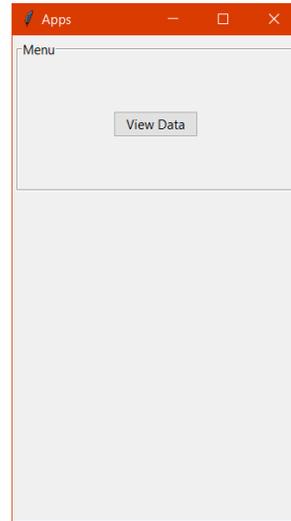
No	Nama Komoditas	Nama Sumber	Rata-Rata Harga	Selisih Harga Dengan Data Pembanding
1	Daging Sapi	Pembanding	117.949	
		Twitter	82.260	35.689
		Detik	112.963	4.986
		Liputan6	66.400	51.549
2	Daging Ayam Ras	Pembanding	35.102	
		Twitter	56.583	-21.481
		Detik	28.000	7.102
		Liputan6	-	35.102
3	Telur Ayam Ras	Pembanding	25.123	
		Twitter	22.500	2.623
		Detik	23.000	2.123
		Liputan6	24.197	926

Selanjutnya adalah penggambaran implementasi *user interface* (UI) dari aplikasi. Yang pertama adalah UI *login*. Terdapat dua *role*(peran) didalam proses *login* yaitu adalah *user* dan admin. Halaman yang dibuka setelah proses login akan tergantung dari *role* yang dideteksi saat proses login. Berikut ini merupakan implementasi UI dari proses *login*



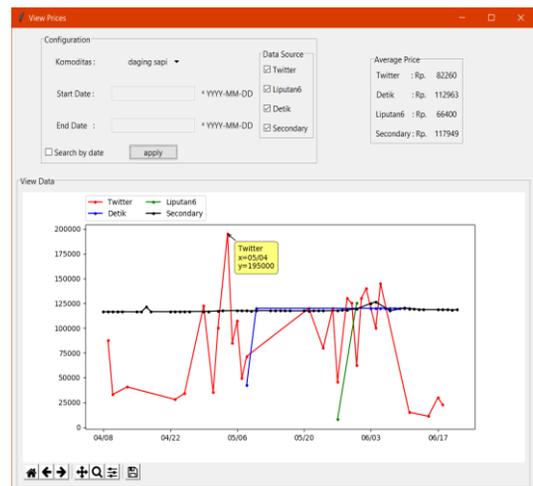
Gambar 9. UI Login

Selanjutnya merupakan halaman utama pengguna yang dapat dilihat pada gambar dibawah ini.



Gambar 10. UI Halaman Pengguna

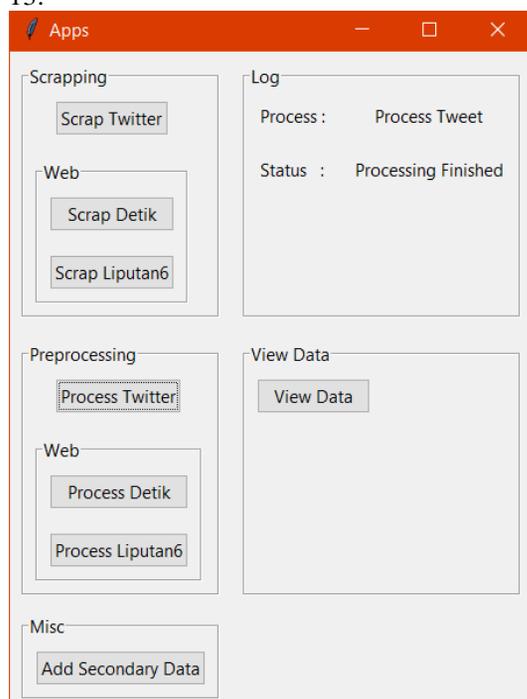
Selanjutnya adalah implementasi dari halaman lihat data yang akan dapat diakses dari halaman utama pengguna dan halaman utama admin. Keduanya memiliki tampilan dan fungsi yang sama. Contoh pada gambar 12 tersebut memperlihatkan data dengan komoditas daging sapi dari semua sumber dan melakukan klik pada diagram garis pada tanggal 2019-05-04. Halaman ini juga akan diakses oleh admin dari halaman utama admin. Disini pengguna atau admin dapat memilih komoditas yang ingin dilihat. Pengguna / admin juga dapat mencentang *search by date* untuk mengaktifkan *entry start date* dan *end date* untuk memasukkan tanggal yang diinginkan. Selain itu detail harga pada pada diagram dapat dilihat dengan melakukan klik pada lingkaran yang ada di diagram. Untuk lebih jelasnya dapat dilihat pada gambar 12 dibawah ini.



Gambar 11. UI Lihat Harga

Didalam label frame *View Data* terdapat kanvas tempat diagram garis digambarkan nantinya. Selain itu dibawah kanvas terdapat *navigation bar* yang berfungsi untuk membantu menavigasi diagram garis. Adapun penjelasan ikon dari kanan ke kiri adalah; ikon untuk menyimpan diagram garis sebagai gambar, ikon untuk mengatur lebar dan tinggi diagram, ikon untuk *zoom*, ikon untuk menavigasi menggunakan mouse, ikon *redo* ke tampilan sebelumnya, ikon *undo* ke tampilan sebelumnya, tombol kembali ke tampilan awal.

Selanjutnya adalah implementasi tampilan admin yang dapat dilihat pada gambar 13.



Gambar 12. UI Halaman Admin

Pada label *frame Scraping* adalah beberapa tombol untuk mengambil data dari sumber yang tertulis pada masing masing tombol. Untuk label *frame Preprocessing* berisi beberapa tombol untuk memproses data mentah yang diambil sebelumnya, juga dibagi sesuai sumbernya. Lalu pada label *frame Misc* berisi tombol untuk meng-*import* data pembandingan. Pada label *frame View Data* terdapat rangkaian label yang akan memberitahu pengguna proses yang sedang berlangsung dan status dari proses tersebut. Semua proses didalam button pada halaman ini kecuali *button view data*, akan ditampilkan pada label *frame Log*. Jika proses

dimulai akan muncul tulisan “[nama proses],please wait” pada status, jika selesai tulisan itu akan menjadi “[nama proses],process completed”. Terakhir adalah label *frame view data* yang akan memunculkan popup untuk melihat harga komoditas.

Selanjutnya adalah proses pengujian untuk menentukan tingkat akurasi dari aplikasi. Uji akurasi akan dibagi menjadi 2 yaitu uji Twitter dan media berita online. Dibawah ini merupakan uji akurasi Twitter.

Tabel 3. Uji Akurasi Twitter

No	Data Tweet	Hasil Pengujian
1.	'nya untuk Wilayah Kec.warungkondang\nDi Jual Daging Sapi untuk d antar H-1 Lebaran /Kg Rp.120.000\n*Bonus 1 Liter Minyak Goreng (Berlaku kelipatan).\nMinat Inbo'	Benar
2	b'Harga daging sapi meroket sampai Rp. 150.000/kg pun sangat menyentuh... \nHiksss... \xf0\x9f\x98\xad https://t.co/KKOioxNyeN'	Benar
3	b'Harga Daging Sapi Bertahan Rp 140.000 per Kilogram di Pasar Baru Bekasi - Warta Kota https://t.co/jzqwKHhhuJ'	Benar
4	b'Hai, Family Midi!\nHari ini kamu bisa TEBUS MURAH INDOFOOD Kecap Manis hanya Rp 5.000! Cukup dengan membeli daging sapi atau ayam menggunakan kartu Ponta di Alfamidi. Jangan sampai terlewatkan ya!\n#AyokeAlfamidi https://t.co/r7kO6M0gJa'	Benar
5	b'\xf0\x9f\x8f\xa1 SOTO SUMUR\n\xf0\x9f\x94\xb0 DEPAN KELURAHAN SETABELAN BANJARSARI\n\xe2\x8f\xb0 BUKA JAM 06.00-13.00	Salah

	WIB\n.\n\x0\x9f\x92\xb0 SOTO DAGING SAPI = Rp. 11.000,- \n\n#Kulinersolo https://t.co/rt4xGBXZly'	
6	b'Tahu bakso khas Malang, asli daging sapi\x0\x9f\x8d\x96 Rp 15.000,-/ pack \nReady di Lovren Gallery Cepu, sebelah barat SMP Negeri 4 Cepu. Buka pukul 13.00-18.00 WIB\n\nOrder by WA, chat saya di 082140409344 \nTerimakasih \x0\x9f\x98\x8d'	Salah
7	b'Hai, Family Midi!\nHari ini kamu bisa TEBUS MURAH INDOFOOD Kecap Manis hanya Rp 5.000! Cukup dengan membeli daging sapi atau ayam menggunakan kartu Ponta di Alfamidi. Jangan sampai terlewatkan ya!\n#AyokeAlfamidi https://t.co/r7kO6M0gJa'	Benar
8	b'rga daging sapi di Aceh melonjak drastis pada hari meugang atau dua hari jelang Ramadan. Di beberapa wilayah, harganya mencapai Rp 200.000/kg. https://t.co/ACDpNqo0Y'b' rga daging sapi di Aceh melonjak drastis pada hari meugang atau dua hari jelang Ramadan. Di beberapa wilayah, harganya mencapai Rp 200.000/kg. https://t.co/ACDpNqo0Y'	Benar
9	b'mbal Pecel Sudah ke Malaysia, Daging Sapi Jatim Menyusul ke Jepang https://t.co/WTkizp4LFG https://t.co/msYAenTS8'	Benar
10	b'Posting lagi saudara...\nMonggoh yg minat ciloknya..\nCilok daging sapi+ayam\nHarga Rp,30.000/pack \nIsi per	Salah

pack 100 bji..\nPO/DO/COD lokasi \nPurwodadi,capang\nWA 0822 4528 9393'	
---	--

Dari tabel uji akurasi Twitter diatas dapat dilihat dari 10 data yang diuji terdapat 3 data yang salah sehingga dari data tersebut dapat dihitung tingkat akurasi penyeleksian harga komoditas dari sumber Twitter adalah 7/10 atau jika dikonversikan ke persen adalah 70%.

Selanjutnya adalah pengujian akurasi penyeleksian harga komoditas dari sumber Media Berita online

Tabel 4. Uji Akurasi Media Berita Online.

No	Data Berita	Hasil Seleksi
1	Amran mengungkapkan daging sapi yang akan diimpor memiliki kualitas medium dengan rata-rata harga US\$ 3 per kilogram (kg) atau setara dengan Rp 42.600 (kurs Rp 14.200)	Benar
2	daging sapi beku Rp 78.00 per kilogram (kg), daging kerbau Rp 60.000 per 900 gram, daging ayam Rp 30.000 per kg	Salah
3	Harga ini sudah mengalami kenaikan Rp 1.000/kg sejak sebelum Ramadhan. Tetapi harga ini masih di bawah HAP telur ayam ras Rp 23.000/kg.	Salah
4	Adapun harga daging sapi Rp 120.000/kg dan daging ayam ras Rp 28.000/kg ekor hidup, baik di Pasar Pusat maupun di Pasar Cik Puan kota Pekanbaru.	Benar
5	Sedangkan untuk telur bebek dan telur puyuh masing-masing ia banderol seharga Rp 2.500 per butir	Benar
6	Pedagang telur lain di Toko Adam juga menjual telur ayam negeri di harga Rp 24 ribu per kg dan telur ayam kampung Rp 2.500 per kg	Benar
7	Tercatat harga daging sapi lokal mencapai Rp 130.000/kg.	Benar
8	Berdasarkan info pangan DKI Jakarta, harga normal telur ayam ras atau telur ayam negeri sebesar Rp 24.148 per kg.	Benar

9	Telur puyuh standar harganya Rp 29.000 per kg," ungkap dia.	Benar
10	Adapun, mengacu pada info pangan DKI Jakarta hari ini maka rerata harga jual telur ayam ras di pasar tradisional saat ini ialah sebesar Rp 24.462 per kg.	Benar

Dari tabel uji akurasi Twitter diatas dapat dilihat dari 10 data yang diuji terdapat 2 data yang salah sehingga dari data tersebut dapat dihitung tingkat akurasi penyeleksian harga komoditas dari sumber Twitter adalah 8/10 atau jika dikonversikan ke persen adalah 80%.Berikut ini adalah detail tingkat akurasi serta total rata-rata dari kedua pengujian yang dilakukan.

Tabel 5. Detail Total Akurasi Pengujian

No	Sumber	Akurasi
1	Twitter	70%
2	Media Berita Online	80%
Total Akurasi Rata-Rata		75%

SIMPULAN

Aplikasi dapat memvisualisasikan fluktuasi harga komoditas pangan.Penerapan seleksi data harga komoditas menggunakan algoritma Jaro-Winkler dalam aplikasi masih belum optimal.Dari hasil uji akurasi terlihat bahwa tingkat akurasi aplikasi adalah 75%.

RUJUKAN

- Kurniawati, A., Puspitodjati, S., & Rahman, S. (2010). *mplementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia*. Depok: Universitas Gunadarma.
- Lutz, M. (2014). *Python Pocket Reference, Fifth Edition*. Sebastopol: O'Reilly.
- Pressman, R. (2015). *Pendekatan Praktisi Buku 1*. Yogyakarta: Penerbit Andi.
- Ratte, J. B. (2016, Maret 23). *pyjarowinkler .PyPI*. Retrieved from Python Software Foundation:
<https://pypi.org/project/pyjarowinkler/>
- Waryanto, B., & Susanti, A. A. (2017). *Statistik Pertanian Agricultural Statistic 2017*. Jakarta: Pusat Data dan Informasi

Pertanian Kementerian Pertanian Republik Indonesia.

Saifullo, E. A. (2017). *Implementasi Natural Language Processing Untuk Mengurangi Risiko Terabaikannya Calon Pelanggan Pada PT. Shafira Tour & Travel*. Surabaya: Institut Bisnis dan Informatika Stikom Surabaya.

Wardani, F. K. (2019). *Analisis Sentimen Untuk Pemeringkatan Popularitas Situs Belanja Online Di Indonesia Menggunakan Metode Naive Bayes(Studi Kasus Data Sekunder)*. Surabaya: Institut Bisnis dan Informatika Stikom Surabaya.

Kressa, A. V. (2019). *Rancang Bangun Portal Beasiswa Berbasis Web Menggunakan Metode Jaro-Winkler*. Surabaya: Institut Bisnis dan Informatika Stikom Surabaya.