

Rekomendasi Pengambilan Tindakan dalam Menjual Barang Menggunakan Metode K-Means Clustering (Studi Kasus : Permainan Warframe)

Kukuh Rahingga Permadi¹, Jondri²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹kukuhrahinggapermadi@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id

Abstrak

Warframe merupakan permainan daring gratis dengan sistem perdagangan *item* permainan yang memungkinkan pemain melakukan transaksi dengan pemain lain. Ketika menjual *item* permainan, pemain harus mencari nilai *item* untuk menentukan tindakan yang menghasilkan keuntungan paling besar. Akan tetapi nilai *item* terus berubah sesuai dengan kondisi pasar, sehingga pencarian nilai dan penentuan tindakan yang paling menguntungkan sulit dilakukan. Saat ini pemain mencari nilai *item* secara manual dengan membandingkan atau memperkirakan nilainya sehingga keuntungan yang didapatkan kurang maksimal. Oleh sebab itu penelitian ini dilakukan untuk membuat sistem yang memberikan rekomendasi tindakan ketika menjual *item* yang diharapkan dapat membantu pemain dalam menentukan tindakan terhadap *item* dan mendapat keuntungan yang mungkin paling besar. Rekomendasi tindakan diambil berdasarkan hasil pengelompokan menggunakan metode K-Means, ekstraksi dataset penelitian dari situs web menggunakan teknik *web scraping*, dan *data editing* untuk mengolah data sebelum proses pengelompokan. Pada penelitian ini dilakukan beberapa pengujian dengan memodifikasi jarak hari data latih dan jarak tren harga untuk mendapatkan model yang paling akurat. Proses pengujian menggunakan data dua hari untuk jarak data latih dan sembilan hari untuk tren harga memberikan hasil dengan akurasi terbaik yaitu 89,59108% dengan *silhouette coefficient* 0,386854. Hasil menunjukkan bahwa metode pengelompokan k-means memberikan rekomendasi tindakan yang cukup akurat menggunakan data statistik penjualan.

Kata kunci: rekomendasi, *web scraping*, pengelompokan, *data editing*, k-means, *silhouette analyst*

Abstract

Warframe is a free online game with in-game item trading system that allows players to make transactions with other players. When selling in-game items, players must find the value of the item to determine the action that gives the highest profit. However, the value of items keeps changing according to market conditions, so finding the value and determining the most beneficial actions is difficult to do. Today, players search for item values manually by comparing or estimating items values so that the benefits obtained are not maximal. Therefore this research was conducted to create a system that provides recommendations for actions when selling items that are expected to help players determine the action of the item and get the greatest possible profit. Recommended actions are taken based on the results of grouping using the K-Means method, extraction of research datasets from websites using web scraping techniques, and data editing to process data before the grouping process. In this research, several tests were carried out by modifying the distance of the training data day and the distance of the price trend to get the most accurate model. The testing process uses two days of data for days and nine days for price trends to provide results with the best accuracy, namely 89.59108% with silhouette coefficient 0.386854. The results show that the k-means grouping method provides recommendations for actions that are quite accurate using sales statistics.

Keyword: recommendation, web scraping, clustering, data editing, k-means, silhouette analyst

1. Pendahuluan

Pada bagian ini berisi empat sub-bagian yaitu: Latar Belakang, Topik dan Batasannya, Tujuan, dan Organisasi Tulisan. Di bawah ini akan dijelaskan dari masing-masing sub-bagian tersebut.

Latar Belakang

Warframe merupakan permainan daring gratis dengan sistem perdagangan *item* permainan baik transaksi mikro antara pemain dengan pengembang atau jual beli seorang pemain dengan pemain lain [1]. Transaksi mikro dengan pengembang menggunakan mata uang asli (Dolar) untuk membeli item permainan atau mata uang permainan (platinum), mata uang permainan digunakan untuk transaksi antar pemain atau dijual kepada pemain lain [2]. *Item* permainan dapat dijual menjadi platinum atau dihancurkan menjadi material mentah yang mana dapat ditukarkan dengan *item* lain. Pada bulan Mei 2019, jumlah pemain warframe kurang lebih mencapai 50

juta pemain dengan nilai transaksi *item* permainan sekitar Rp500 juta rupiah per hari [3]. Pemain yang menjual *item* permainan akan mencari nilai dari *item* dan menentukan tindakan apa yang memberikan keuntungan paling besar terhadap *item* tersebut. Tindakan yang diambil dapat berupa dijual secara langsung menjadi mata uang permainan (platinum), dijadikan material mentah (ducat), atau disimpan karena tren harga *item* sedang naik.

Tindakan yang diambil ketika menjual *item* dapat mempengaruhi keuntungan yang didapatkan oleh pemain. Saat ini pemain yang akan menjual *item* harus mencari nilai setiap *item* secara manual dan memperkirakan keuntungan dengan membandingkan harga platinum, harga mentah, dan kemungkinan harga. Akan tetapi nilai setiap *item* berbeda dan terus berubah sesuai dengan kondisi pasar sehingga pemain kesulitan dalam menentukan tindakan untuk mendapatkan keuntungan yang mungkin paling besar. Dalam penelitian ini, penulis membuat sebuah sistem yang memberikan rekomendasi tindakan dalam menjual *item* permainan yang diharapkan dapat membantu pemain Warframe dalam menentukan tindakan ketika akan menjual *item* dan mendapatkan keuntungan yang mungkin merupakan keuntungan paling besar.

Berdasarkan uraian permasalahan diatas, untuk mendapatkan rekomendasi tindakan terhadap *item* yang dijual, penulis pengelompokan *item* ke dalam tindakan tertentu menggunakan metode k-means karena dataset yang digunakan tidak memiliki label (*Unsupervised Learning*) [4] [5] [6]. Dataset yang digunakan merupakan data dari situs web yang memuat informasi dasar setiap *item*, statistik penjualan selama 90 hari, dan harga *item* yang sedang aktif dijual ketika melakukan ekstraksi data. Data diambil menggunakan teknik *web scraping* yang merupakan proses ekstraksi data dari sebuah web [7]. *Data editing* digunakan untuk merubah data mentah yang didapatkan dari *scraping web* menjadi nilai yang dapat digunakan untuk proses pengelompokan k-means [8] [4]. *Silhouette* merupakan metode yang digunakan untuk mengukur kemiripan setiap objek di dalam kelompoknya dibandingkan dengan kelompok lain [6]. *Confusion Matrix* digunakan untuk mengukur akurasi hasil pengelompokan dengan kelas sebenarnya [9]. Penelitian ini menggunakan rentang hari data latih dan tren harga yang berbeda untuk melihat pengaruh perubahan tersebut terhadap akurasi pengelompokan yang didapatkan.

Topik dan Batasannya

Pengambilan tindakan ketika menjual *item* permainan dapat mempengaruhi keuntungan yang didapatkan pemain. Tindakan yang harus diambil untuk setiap *item* terus berubah sesuai dengan kondisi pasar dan menyebabkan pengambilan tindakan untuk mendapatkan keuntungan paling besar sulit dilakukan. Penelitian ini mengangkat permasalahan terkait bagaimana membangun sebuah sistem yang memberikan rekomendasi tindakan ketika menjual *item* permainan yang mungkin menghasilkan keuntungan terbesar berdasarkan kondisi pasar saat itu. Hasil rekomendasi tindakan yang didapatkan dipengaruhi oleh keruntutan data, jarak hari data latih dan jarak tren harga, *preprocessing*, atau metode pengelompokan yang digunakan. Titik fokus pada penelitian ini adalah pengaruh penggunaan jarak hari data latih dan jarak tren harga yang berbeda terhadap hasil pengelompokan yang didapatkan. Penelitian ini menggunakan teknik *web scraping* untuk mengekstrak data dari situs web, *data editing* untuk mengolah data agar dapat digunakan untuk proses pengelompokan, metode pengelompokan yang digunakan adalah k-means, mengukur akurasi hasil pengelompokan menggunakan *confusion matrix*, dan *silhouette* untuk mengukur seberapa baik kelompok yang dibentuk.

Pada penelitian ini terdapat beberapa hal yang dijadikan sebagai batasan masalah. Kelas yang digunakan terdiri dari tiga kelas yaitu jual platinum, jual mentah, dan tahan. Jumlah data yang digunakan adalah data statistik penjualan 90 hari terakhir dengan total data sebanyak 24.210 data tetapi akan ada kemungkinan kurang dari 24.210 karena terdapat *item* yang dalam sehari tidak memiliki penjualan. Data penjualan yang sedang aktif ketika melakukan ekstraksi data dari web sebanyak 269 data. Hasil keluaran dari sistem memiliki tiga kelas, tetapi ketika melakukan pengujian untuk mendapatkan akurasi, data dengan kelas tahan akan dianggap sebagai kelas jual platinum, karena kelas tahan merupakan kelas yang nantinya dijual menjadi platinum dan penelitian ini belum bisa mendapatkan data aktual kelas tahan secara matematis.

Tujuan

Tujuan dari penelitian ini adalah membangun sebuah sistem rekomendasi pengambilan tindakan dalam menjual *item* permainan menggunakan metode pengelompokan k-means. Menganalisis pengaruh penggunaan jarak hari data latih dan jarak tren harga yang berbeda terhadap hasil pengelompokan yang didapatkan. Serta ketepatan hasil rekomendasi tindakan yang didapatkan menggunakan metode k-means dalam menentukan tindakan yang paling menguntungkan.

Organisasi Tulisan

Selanjutnya pada bagian II akan dijelaskan tentang penelitian yang telah dilakukan sebelumnya dan dasar teori terkait dengan penelitian ini. Keseluruhan metodologi penelitian dan sistem yang dibangun akan dijelaskan pada bagian III. Hasil yang didapatkan dan evaluasi terhadap hasil akan dijelaskan di bagian IV. Terakhir pada bagian V akan menjelaskan kesimpulan penelitian dan saran untuk penelitian kedepannya.

2. Studi Terkait

Pada penelitian ini, pengelompokan menggunakan metode yang sebelumnya telah banyak digunakan oleh peneliti lain dalam menyelesaikan berbagai masalah yang ada. Pengelompokan digunakan untuk mengelompokkan data ke dalam beberapa kelas yang ditentukan sesuai dengan karakteristik data tersebut, seperti pengelompokan siswa, segmentasi warna, dan sebagainya [10] [11]. Dalam proses pengelompokan, terkadang data yang digunakan merupakan data mentah yang nilainya harus diproses terlebih dahulu agar data dapat digunakan oleh sistem pengelompokan. Penelitian yang menggunakan pengelompokan seringkali melakukan *preprocessing* untuk mengubah nilai yang agar dapat digunakan, seperti pembobotan, *text preprocessing*, perubahan dimensi, dan sebagainya [11] [12] [10]. Pada proses ini data yang ada diproses sesuai dengan aturan yang ditentukan oleh peneliti yang nantinya digunakan untuk melatih dan menguji model pengelompokan yang digunakan.

Sebelum masuk ke dalam proses *preprocessing*, data yang digunakan dalam penelitian harus dikumpulkan terlebih dahulu dari situs web atau sumber fisik. Data yang dikumpulkan dapat berupa data yang telah diproses ataupun data mentah yang ada di internet. Terdapat teknik yang dapat digunakan untuk mengumpulkan data seperti pada penelitian [13] yaitu *web scraping*.

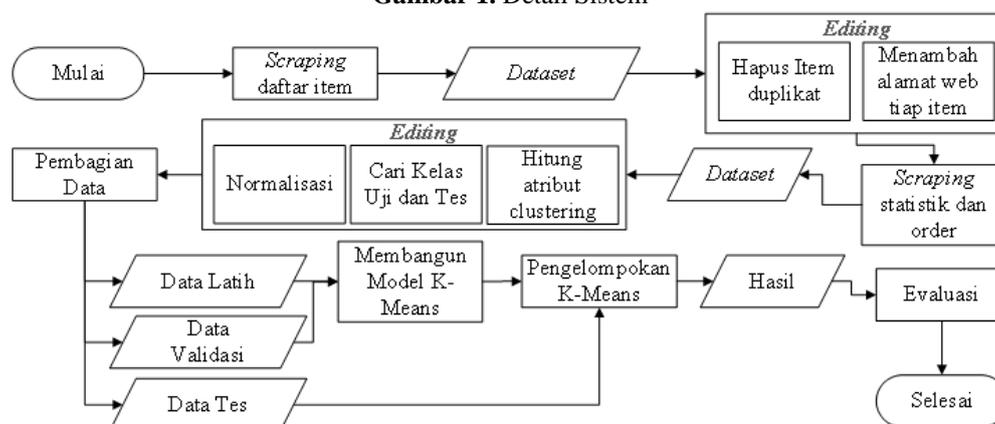
Pada penelitian [11], terdapat data non numerik sehingga peneliti mengubah data tersebut menjadi data numerik dan memberikan pembobotan untuk setiap data. Akurasi pengelompokan penjurusan siswa yang didapatkan pada penelitian tersebut adalah 90,58% dengan menggunakan metode k-means untuk pengelompokannya. Dimana K-means merupakan metode yang banyak digunakan karena fleksibel dan dapat digunakan untuk mengelompokkan data numerik [5] [11]. Walaupun hasil dari pengelompokan yang didapatkan kurang optimal dan performansi yang kurang berdasarkan penelitian yang dilakukan oleh [12] [14] yang diakibatkan dari inisialisasi *cluster* secara acak. Untuk mengatasi masalah tersebut, penelitian yang dilakukan oleh [11] [12] melakukan pengulangan proses pengelompokan untuk mendapat inisialisasi *cluster* yang paling optimal. Walaupun begitu *cluster* yang didapatkan belum tentu merepresentasikan suatu *cluster* yang baik, dengan demikian nilai *purity cluster* yang terbentuk belum mendekati satu [12].

Pengukuran akurasi merupakan pengukuran tingkat kesesuaian kelas yang diramalkan menggunakan klasifikasi dengan kelas asli yang telah diketahui. Penelitian [11] menggunakan *Confusion Matrix* untuk mengukur kesesuaian kelas ramalan yang didapatkan. Untuk mengukur keoptimalan kelompok yang terbentuk penelitian [12] menggunakan *Silhouette Coefficient* untuk mendapatkan nilai seberapa optimal setiap kelompok yang terbentuk. Semakin tinggi akurasi yang didapatkan maka kelas ramalan yang didapatkan sesuai dengan kelas asli, semakin tinggi *Silhouette coefficient* maka semakin optimal kelompok yang terbentuk.

3. Sistem yang Dibangun

Pada penelitian tugas akhir ini, peneliti membangun sebuah sistem yang memberikan rekomendasi tindakan yang mungkin paling menguntungkan ketika menjual *item* permainan. Keluaran yang dihasilkan dari sistem merupakan tiga kelas tindakan terhadap *item* yaitu jual platinum, jual mentah, atau tahan. Secara garis besar proses dibagi menjadi beberapa tahap yaitu data *scraping*, data *editing*, pengelompokan, dan evaluasi hasil. Detail sistem yang dibangun digambarkan dalam bentuk *flowchart* dapat dilihat pada Gambar 1 dibawah.

Gambar 1. Detail Sistem



Tahap awal pada penelitian ini adalah pengumpulan daftar *item* permainan yang terdaftar di dalam permainan. Daftar *item* yang dikumpulkan digunakan sebagai acuan ketika melakukan ekstraksi data statistik dan pasar. Ekstraksi daftar *item* dari situs web https://warframe.fandom.com/wiki/Void_Relic/ByRelic dilakukan secara otomatis menggunakan teknik *web scraping* menggunakan library “*PHP Simple HTML DOM Parser*”. Data yang dikumpulkan berupa *void relic*, nama *item* dan harga mentah dari setiap *item*. Representasi data dapat dilihat pada Tabel 1.

Tabel 1. Representasi daftar item

Void Relic	Nama Item	Harga Mentah
Lith A1	Akstiletto Prime Blueprint	100
Lith A2	Akstiletto Prime Blueprint	100
Lith A1	Saryn Prime Neuroptics Blueprint	45

Table 1 merupakan contoh dari hasil ekstraksi data yang didapatkan pada tahap *scraping* daftar *item*. Dapat dilihat jika pada hasil *scraping* masih terdapat data duplikasi dan belum didefinisikan alamat web *market*.

Setelah proses *scraping* daftar *item*, tahap selanjutnya adalah *editing* data daftar *item*. Proses *editing* bertujuan untuk menghilangkan data duplikasi dan menambahkan alamat web *market* dari setiap *item*. Menghilangkan data duplikat dilakukan terhadap *item* dengan nama yang sama, sehingga tidak ada lebih dari satu *item* dengan nama yang sama. Penambahan web *market* merupakan tahap penambahan alamat dari setiap *item* berdasarkan nama *item*, alamat yang ditambahkan akan merujuk ke situs *market* dari setiap *item* yang berisi penjualan aktif dan statistik penjualan. Penulis melakukan pengecekan setiap alamat secara manual untuk menghindari kesalahan *scraping*, karena tidak semua alamat hasil konversi nama *item* merujuk ke alamat market yang sesuai. Hasil data *editing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil editing daftar item

Void Relic	Nama Item	Harga Mentah	Alamat Market
Lith A1	Akstiletto Prime Blueprint	100	https://warframe.market/items/akstiletto_prime_blueprint
Lith A1	Saryn Prime Neuroptics Blueprint	45	https://warframe.market/items/saryn_prime_neuroptics_blueprint

Tabel 2 merupakan data hasil *editing* daftar *item* yang telah dihilangkan duplikasi nama *item* dan telah ditambah dengan alamat web *market* *item*.

Tahap selanjutnya adalah mengumpulkan data dari situs *market* menggunakan alamat yang telah didefinisikan pada proses sebelumnya. Dataset dikumpulkan dari situs <https://warframe.market> secara otomatis menggunakan teknik *scraping* dengan library yang sama dengan proses sebelumnya. Data yang dikumpulkan berupa statistik setiap *item* per-hari selama 90 hari terakhir dan rata-rata harga pasar setiap *item* yang aktif ketika melakukan ekstraksi. Data statistik yang dikumpulkan berupa data penjualan per-hari yaitu waktu, harga membuka, harga penutup, rata-rata harga, harga minimal, harga maksimal, dan volume sebanyak 90 hari untuk setiap *item*, lihat tabel 3. Data pasar yang dikumpulkan adalah harga dan jumlah *item* yang sedang aktif dijual kemudian langsung dihitung rata-rata harga masing-masing *item* lihat tabel 4.

Tabel 3. Dataset statistik

Nama Item	hari	harga pembuka	harga penutup	rata-rata harga	harga minimal	harga maksimal	volume
Akstiletto Prime Blueprint	0	13	19	15	10	20	17
Akstiletto Prime Blueprint	1	19	21	20	17	23	19

Tabel 3 merupakan data hasil *scraping* dari web *market*. Data diatas merupakan contoh data hasil *scraping* dari statistik yang mana data tersebut merupakan data dalam satu hari.

Tabel 4. Dataset Order

Nama Item	Rata-rata Harga
Akstiletto Prime Blueprint	24,14
Saryn Prime Neuroptics Blueprint	18,42

Tabel 4 dataset tes yang didapatkan merupakan rata-rata harga dari *item* yang sedang aktif dijual ketika melakukan ekstraksi data dari situs web.

Setelah semua data statistik dan pasar setiap *item* dikumpulkan, kemudian masuk ke proses data *editing*. Proses data *editing* yang pertama adalah menghitung atribut yang digunakan untuk proses pengelompokan. Atribut pertama adalah rata-rata harga untuk statistik yang mana langsung menggunakan rata-rata harga hasil ekstraksi data statistik. Rata-rata harga untuk *item* pasar juga telah dihitung ketika proses ekstraksi data. Atribut selanjutnya adalah nilai mentah dari *item*, Atribut ini berisi nilai material mentah dari sebuah *item* dengan harga tertentu. Untuk menghitung nilai tersebut menggunakan persamaan berikut.

$$\text{nilai mentah} = \frac{\text{Harga Mentah}}{\text{Harga platinum}} \quad (1)$$

Berdasarkan persamaan 1, *nilai mentah* merupakan nilai mentah dari *item* untuk setiap satu platinum, dengan membagi harga mentah dari *item* dengan harga platinum. Terakhir atribut tren harga dari *item* pada rentang hari

tertentu, dimana tren merupakan indikator kondisi harga berdasarkan harga dan volume *item* yang terjual [15]. Tren harga didapatkan dengan menghitung statistik harian harga dan volume *item* selama beberapa hari sebelumnya. Berikut persamaan untuk menghitung trend harga.

$$PVT_i = PVT_{i-1} + volume_i \times \frac{penutup_i - penutup_{i-1}}{penutup_{i-1}} \quad (2)$$

Berdasarkan persamaan 2, PVT_i merupakan tren harga pada hari ke- i , PVT_{i-1} adalah PVT di hari sebelumnya, $volume_i$ merupakan volume *item* terjual, $penutupan_i$ merupakan harga penutupan pada hari ke- i , dan $penutupan_{i-1}$ adalah harga penutupan di hari sebelumnya.

Data uji dan tes yang digunakan untuk mengukur keakuratan model ditambah dengan kelas asli yang mungkin memberikan keuntungan paling besar. Pemberian kelas dilakukan secara manual dengan membandingkan keuntungan ketika item dijual secara langsung atau dijadikan material mentah. Kelas asli yang didapatkan digunakan pada proses pengujian untuk mengukur akurasi dari sistem yang dibangun.

Tahap terakhir pada proses *editing* adalah normalisasi, normalisasi data dilakukan untuk menyamakan satuan ukur yang berbeda dari setiap atribut agar dapat digunakan untuk proses pengelompokan. Metode yang digunakan pada penelitian ini adalah *Min-Max* yang merupakan metode normalisasi dengan melakukan transformasi linear terhadap data asli [8]. Metode normalisasi *Min-Max* mengubah nilai rata-rata harga, nilai mentah dan trend harga pada rentang tertentu. Berikut adalah persamaan normalisasi *Min-Max*.

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (new_max_A - new_min_A) + new_min_A \quad (3)$$

Berdasarkan persamaan 3, v_i merupakan data yang akan ditransformasikan menjadi v'_i yang merupakan hasil dari normalisasi, \min_A dan \max_A adalah nilai minimal dan maksimal dari keseluruhan data dari atribut v_i , dan new_min_A dan new_max_A merupakan nilai minimal dan maksimal dari jarak nilai baru dari atribut v_i .

Data yang dihasilkan dari proses *editing* merupakan data yang digunakan untuk proses pengelompokan. Data hasil proses *editing* dapat dilihat pada tabel 5 untuk data latih, dan tabel 6 untuk data tes.

Tabel 5. Data Latih Hasil Editing

Nama Item	Hari	Harga	Nilai Mentah	Tren Harga	Tindakan
Akstiletto Prime Blueprint	90	6.63	48.427	22.073	Jual Mentah
Akstiletto Prime Blueprint	89	6.1	52.24	23.21	-

Tabel 5 merupakan atribut yang telah dinormalisasi dan digunakan untuk proses pembentukan model. Dataset uji telah memiliki kelas untuk dibandingkan dengan hasil yang didapatkan dari proses pengelompokan.

Tabel 6. Data Tes Hasil Editing

Nama Item	Harga	Nilai Mentah	Tren Harga	Tindakan
Akstiletto Prime Blueprint	6.539	43.502	18.497	Jual Mentah
Saryn Prime Neuroptics Blueprint	16.68	9.634	66.24	Jual Platinum

Tabel 6 data yang akan digunakan untuk proses pengujian model dan telah memiliki data aktual yang merupakan tindakan paling menguntungkan.

Proses selanjutnya adalah pembagian data latih, uji, dan tes. Data latih terdiri dari data statistik yang dibagi sesuai dengan waktu dan urutannya digunakan untuk membentuk model pengelompokan. Hari terakhir pada data latih akan digunakan untuk data uji, model yang telah terbentuk menggunakan data latih diuji akurasinya dalam pengelompokan berdasarkan kelas asli yang didapatkan pada proses sebelumnya. Data tes terdiri dari data *item* yang sedang aktif dijual ketika pengumpulan data dan digunakan untuk menguji model terhadap data *item* yang sedang aktif dijual.

Pada penelitian ini, data dikelompokkan menggunakan metode k-means. Secara garis besar metode k-means mengelompokkan data yang belum memiliki kelas ke dalam sebuah kelas berdasarkan karakteristik yang ada pada setiap data. Pengelompokan menggunakan jarak antara data dengan *centroid* yang terus diperbarui sampai *centroid* berada di titik tengah sebuah kelompok data. Inisialisasi *centroid* awal mengambil salah satu data latih secara acak karena kelas yang dikelompokkan belum diketahui, kemudian *centroid* diperbarui menggunakan data hari ke satu, setelah mendapatkan *centroid* maka masuk ke hari selanjutnya hingga hari terakhir secara bergiliran dimana satu hari memiliki 269 data dari masing-masing *item* pada hari tersebut. Model yang terbentuk dari metode k-means adalah *centroid* dari setiap kelas, yaitu titik tengah dari kelompok yang telah terbentuk. Kemudian data uji akan dimasukkan ke dalam kelas berdasarkan jarak data dengan *centroid* terdekat. *Euclidean distance* digunakan untuk menghitung jarak antara dua titik dan digunakan untuk menghitung jarak antara *centroid* dan data, berikut persamaannya [16].

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2} \quad (4)$$

Berdasarkan persamaan 4, $d(p, q)$ merupakan jarak antara titik p dan q dengan p berada pada titik koordinat $p_1, p_2,$ dan p_3 terhadap titik q pada koordinat $q_1, q_2,$ dan q_3 .

Tahap terakhir pada penelitian ini adalah evaluasi yaitu untuk menganalisis hasil yang didapatkan dari sistem yang dibangun. Penelitian ini menggunakan dua metode untuk mengevaluasi hasil yang didapatkan. Pertama untuk mengukur akurasi atau kesesuaian hasil pengelompokan dengan kelas sebenarnya menggunakan *confusion matrix*.

Tabel 7. Confusion Matrix

Kelas Data		Kelas Asli	
		Positif	Negatif
Kelas Klasifikasi	Positif	True Positives (TP)	False Positives (FP)
	Negatif	False Negatives (FN)	True Negatives (TN)
Total		Positif (P)	Negatif (N)

Tabel 5 merupakan *Confusion Matrix* dimana nilai TP, TN, FP, dan FN digunakan dalam mengukur akurasi dari hasil pengelompokan. Berikut adalah persamaan untuk menghitung akurasi [17].

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Pada persamaan 5, TP merupakan jumlah data positif yang terklasifikasi dengan benar, TN jumlah data negatif yang terklasifikasi dengan benar, FP jumlah data positif yang terklasifikasi dengan salah, FN jumlah data negatif yang terklasifikasi dengan salah.

Kemudian untuk mengukur keakuratan kelompok yang dibentuk oleh metode pengelompokan, penelitian ini menggunakan metode *silhouette Coefficient*. *Silhouette* mengukur seberapa dekat jarak sebuah data dengan data lain di kelompoknya dan seberapa jauh data dengan data lain di kelompok lain yang paling dekat. Hasil yang didapatkan dari metode ini adalah *silhouette coefficient* dari -1 sampai 1 yang menunjukkan seberapa akurat kelompok yang dihasilkan. Berikut persamaan yang digunakan.

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (6)$$

$S(i)$ adalah *silhouette coefficient* dari hasil pengelompokan dapat dilihat pada Persamaan 6, $b(i)$ merupakan nilai antara data i dengan data kelompok lain yang terdekat, dan $a(i)$ merupakan nilai antara data i dengan data lain di kelas yang sama dengan data i .

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (7)$$

Persamaan 7, $a(i)$ merupakan nilai data i yang didapatkan dari $\sum_{j \in C_i, i \neq j} d(i, j)$ rata-rata jarak data i dan j yang mana j merupakan data kelas di dalam *cluster* C_i tetapi $|C_i| - 1$ yaitu tidak $d(i, i)$.

$$b(i) = \min_{i \neq j} \frac{1}{|C_j|} \sum_{j \in C_j} d(i, j) \quad (8)$$

$b(i)$ merupakan nilai data i yang didapatkan dari $\sum_{j \in C_j} d(i, j)$ rata-rata jarak data i dan j dimana j merupakan anggota kelompok C_j dan $\min_{i \neq j}$ diambil kelas yang paling dekat dengan kelompok data i , detail persamaan pada Persamaan 6,.

4. Evaluasi

Pada bagian ini berisi Hasil Pengujian dan Analisis Hasil Pengujian. Di bawah ini akan dijelaskan dari masing-masing sub-bagian tersebut.

4.1 Hasil Pengujian

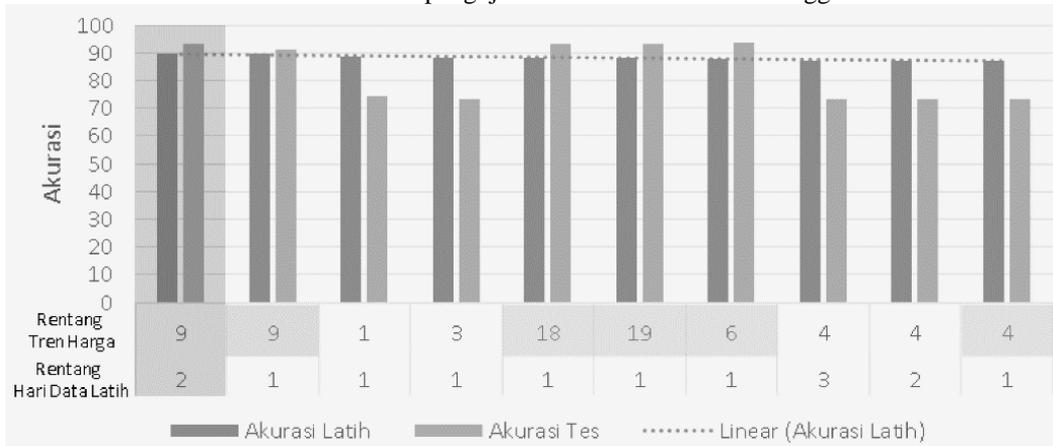
Penelitian ini menggunakan data statistik penjualan *item* dan rata-rata harga *item* yang sedang aktif ketika melakukan proses ekstraksi data dari situs web. Data pengujian terdiri dari 269 jenis *item* dengan statistik untuk masing-masing *item* sebanyak 90 hari. Rata-rata harga *item* yang sedang aktif berjumlah 269 untuk masing-masing jenis *item*.

Data hasil *scraping* diperiksa secara manual untuk mencari data yang hasil *scraping*-nya salah, kosong, atau tidak sesuai dengan sumber data. Kemudian memproses data dengan *data editing* untuk mendapatkan nilai yang digunakan dalam proses pengelompokan, karena metode k-means menggunakan titik yang dihasilkan dari atribut yang didapatkan, sehingga persebaran data lebih seimbang.

Penelitian ini melakukan beberapa pengujian menggunakan rentang hari yang berbeda untuk melatih model dan rentang tren harga yang berbeda pula. Setiap pengujian rentang yang berbeda sistem melakukan

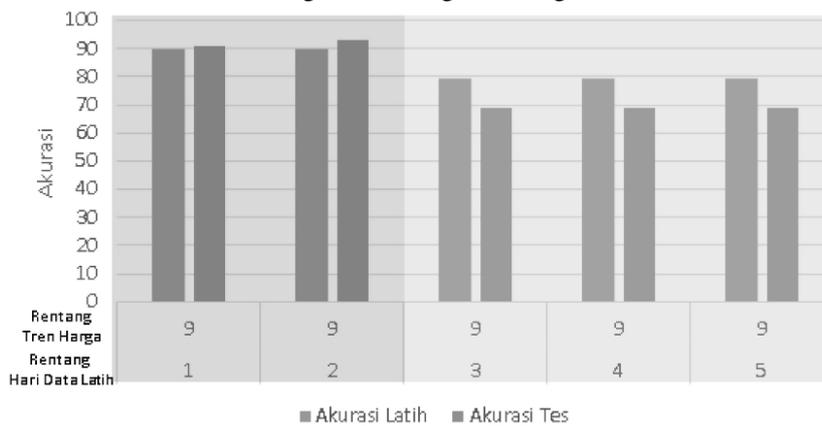
pembentukan model sebanyak 250 kali dan menyimpan model dengan hasil hasil akurasi yang paling tinggi. Hal ini dilakukan karena inisialisasi awal *centroid* dilakukan secara acak, sehingga hasil yang didapatkan tidak pasti. Hasil pengujian dari sistem yang dibangun menggunakan metode pengelompokan k-means kemudian dievaluasi menggunakan *confusion matrix* dan *silhouette*.

Gambar 2. Hasil pengujian terurut dari akurasi tertinggi



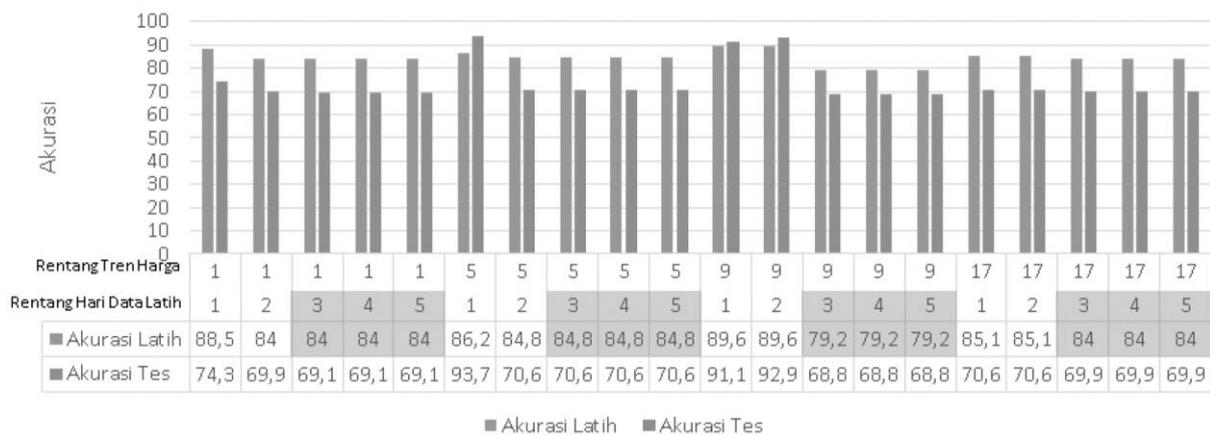
Gambar 2 menunjukkan terurut dari akurasi tertinggi. Rentang data latih 2 hari dan tren harga 9 hari mendapatkan akurasi tertinggi, rentang data latih 1 hari dan tren harga 9, 18, 19, 6, dan 4 mendapatkan akurasi yang cukup tinggi walaupun memiliki tren harga yang berbeda cukup jauh.

Gambar 3. Tren harga 9 hari dengan rentang data latih berbeda



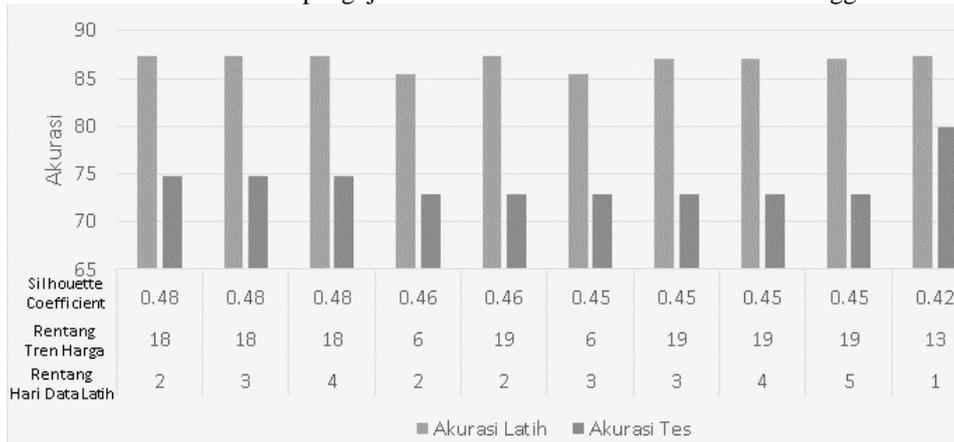
Rentang tren harga 9 dengan rentang data latih berbeda pada Gambar 3. Menunjukkan akurasi menurun ketika menggunakan rentang data latih lebih dari dua hari.

Gambar 4. Beberapa pengujian dengan hasil yang memiliki pola yang sama



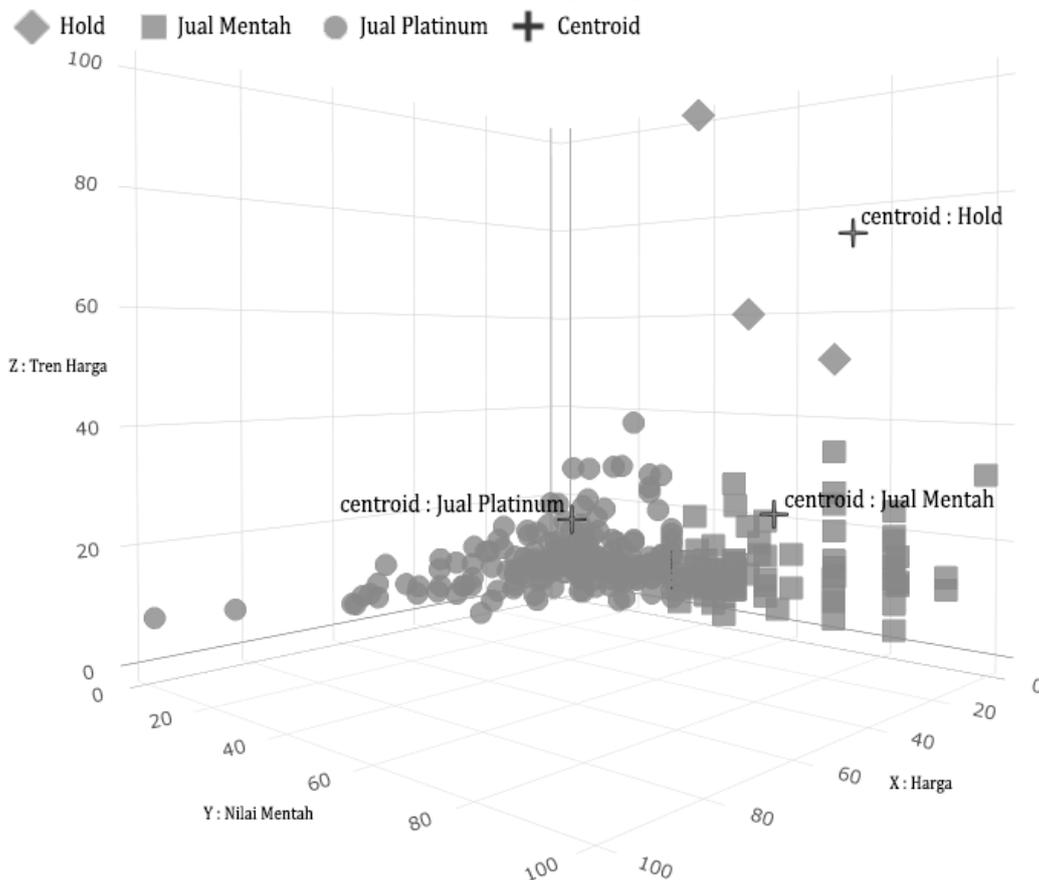
Beberapa hasil pengujian pada Gambar 4 memiliki pola yang hampir mirip. Beberapa sampel pengujian memiliki akurasi yang menurun pada rentang data latih lebih dari 2 hari dan cenderung menghasilkan nilai akurasi yang sama sampai hari ke lima.

Gambar 5. Hasil pengujian terurut dari silhouette coefficient tertinggi



Hasil pengujian terurut dari *silhouette coefficient* tertinggi pada Gambar 5. Terlihat bahwa akurasi latih yang didapatkan lebih rendah daripada pengujian lain dan akurasi tes lebih rendah lagi.

Gambar 6 Scatter Plot Hasil Pengelompokan Data Validasi



Hasil pengelompokan yang ditunjukkan pada *scatter plot* gambar 6, menunjukkan data telah dikelompokkan sesuai dengan karakteristik pengelompokan walaupun data belum terkelompok dengan baik.

4.2 Analisis Hasil Pengujian

Hasil pada gambar 2 rentang tren harga 9 hari dengan rentang data latih 1 hari menghasilkan akurasi tertinggi karena model yang terbentuk sebagian besar menghasilkan rekomendasi tindakan yang sesuai dengan kelas aktual. Pada rentang data latih 1 hari dan tren harga berbeda cukup jauh yaitu 9, 18, 19, 6, dan 4 hari menghasilkan akurasi yang cukup tinggi, hal tersebut terjadi karena model yang terbentuk memiliki pola yang hampir sama. Retang tren harga yang digunakan tidak terlalu berpengaruh pada model yang dibentuk dan tetap memberikan akurasi yang cukup tinggi pada rentang tertentu. Sedangkan untuk rentang data latih yang menghasilkan akurasi tinggi sebagian besar memiliki rentang rendah, dari 1 hari hingga 2 hari. Hal tersebut akan dipastika pada analisis selanjutnya terhadap gambar 3 dan 4.

Rentang data latih 1 dan 2 hari memberikan akurasi yang cukup tinggi tetapi setelah 2 hari dan seterusnya akurasi yang didapatkan menurun yang ditunjukkan pada gambar 3. Penurunan akurasi tersebut diakibatkan oleh titik awal centroid yang dihasilkan oleh hari yang lebih dari 2 hari berada pada titik yang berbeda dan menghasilkan pengelompokan yang kurang sesuai dengan kelompok aktual. Semakin jauh rentang data latih yang digunakan untuk pembentukan model, maka semakin rendah akurasi yang didapatkan. Karena data yang digunakan akan semakin tidak sesuai dengan kondisi terbaru.

Pada gambar 4 hasil yang didapatkan dari beberapa pengujian menunjukkan pola yang hampir mirip, yaitu rentang data latih setelah 2 hari menghasilkan akurasi yang lebih rendah. Penyebab dari penurunan akurasi tersebut karena data yang digunakan semakin jauh dari kondisi data yang sekarang, sehingga hasil pengelompokan yang didapatkan tidak seakurat rentang data latih 1 dan 2 hari.

Silhouette coefficient tertinggi yang ditunjukkan pada gambar 5 memiliki akurasi data latih dan data uji yang lebih rendah. walaupun kelompok yang terbentuk merupakan kelompok yang paling optimal, tetapi kelompok tersebut tidak menghasilkan akurasi yang terbaik. Hal tersebut diakibatkan oleh proses *preprocessing* yang kurang sempurna untuk menghasilkan persebaran data yang paling optimal juga memberikan akurasi yang terbaik.

Dari hasil *scatter plot* yang ditunjukkan pada gambar 6, data telah terkelompok kedalam centroid yang sesuai dengan karakteristiknya, sumbu x, y, dan z menunjukkan bahwa tindakan yang dikelompokkan sesuai dengan variabel yang sesuai dengan tindakan tersebut. Walaupun data masih sangat tersebar dan belum terkelompok dengan baik karena proses *preprocessing* yang belum optimal, tetapi hasil pengelompokan sudah menghasilkan akurasi yang cukup baik pada *Silhouette coefficient* tertentu.

5. Kesimpulan

Dari hasil penelitian dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa rentang jarak tren harga tidak memiliki pola tertentu dan akan menghasilkan akurasi yang cukup tinggi ketika model yang terbentuk sesuai. Jarak data latih akan menghasilkan akurasi yang paling tinggi pada rentang kurang dari 2 hari dan akan menurun pada hari ke 3 dan seterusnya. *Silhouette score* yang didapatkan masih kurang mendekati 1 karena metode *preprocessing* terhadap data yang dikumpulkan masih sangat sederhana. Hal ini membuat kelompok yang dihasilkan masih tersebar dan kurang berpusat pada *cluster* kelasnya sehingga menghasilkan akurasi yang cukup rendah pada *Silhouette score* tertinggi.

Metode k-means dapat digunakan untuk mengelompokkan data *item* ke dalam kelas tindakan tertentu untuk mendapatkan keuntungan yang mungkin paling besar menggunakan data statistik penjualan. Berdasarkan hasil pengujian dengan parameter yang menghasilkan akurasi paling tinggi yaitu rentang hari selama dua hari dan rentang trend harga selama Sembilan hari, didapatkan hasil akurasi data latih sebesar 89,59108% dengan *silhouette coefficient* 0,386854 dan untuk data tes akurasi yang didapatkan sebesar 92,9368% dengan *silhouette coefficient* 0,348637.

Saran untuk penelitian selanjutnya yang didasarkan pada penelitian ini yaitu, *preprocessing* terhadap data yang dikumpulkan menggunakan pendekatan yang berbeda agar pengelompokan data benar-benar terbagi sesuai dengan karakteristik masing-masing item. Pengujian menggunakan metode klasifikasi yang berbeda untuk mendapatkan akurasi tindakan yang lebih akurat. Menambahkan metode prediksi pada *preprocessing* untuk memperkirakan nilai atau harga pada hari ekstraksi, karena rekomendasi tindakan yang diambil masih berdasarkan data hari sebelumnya.

Daftar Pustaka

- [1] D. O'Dwyer, "The Story of Digital Extremes (Warframe Doc Part 1) - YouTube," Noclip, 19 Maret 2018. [Online]. Available: <https://www.youtube.com/watch?v=UOE6528pwFc&t=409s>. [Accessed 30 Maret 2019].
- [2] B. Kuchera, "How Warframe built an ethical free-to-play economy - Polygon," Polygon, 8 Juli 2018. [Online]. Available: <https://www.polygon.com/2018/1/2/16830328/warframe-free-to-play-f2p-platinum>. [Accessed 30 Maret 2019].
- [3] D. EXTREMES, "WARFRAME TURNS SIX! | Digital Extremes," DIGITAL EXTREMES, 26 March 2019. [Online]. Available: <https://www.digitalextremes.com/news/2019/03/warframe-turns-six>. [Accessed 1 May 2019].
- [4] J. Wu, *Advances in K-means Clustering: A Data Mining Thinking*, Heidelberg: Springer, 2012.
- [5] A. S. a. S. Vishwanathan, *Introduction to Machine Learning*, Cambridge: Cambridge University Press, 2008.
- [6] M. B. A.-. Zoubi and M. a. Rawi, "An Efficient Approach for Computing Silhouette Coefficients," *Journal of Computer Science*, vol. IV, no. 3, pp. 252-255, 2008.

- [7] J. Ward, Instant PHP Web Scraping, Birmingham: Packt Publishing Ltd., 2013.
- [8] J. Han, M. Kamber and J. Pei, Data Mining Concept and Technique, Waltham: Elsevier Inc, 2012.
- [9] C. Sammut and G. I. Webb, Encyclopedia of Machine Learning, Boston: Springer, 2010.
- [10] I. A. Kesuma, Herman and Munawir, "Penerapan Metode Klaster K-Means Pada Segmentasi Warna Citra Digital dengan Pemrograman R," *Prosiding Seminar Nasional Inovasi dan Teknologi Informasi*, vol. I, no. 3, pp. 427-430, 2016.
- [11] C. Purnamaningsih, R. Saptono and A. Aziz, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA," *JURNAL ITSMART*, vol. III, no. 1, pp. 27-33, 2014.
- [12] R. Handoyo, R. R. M and S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means Pada Pengelompokan Dokumen," *JSM STMIK Mikroskil*, vol. XV, no. 2, pp. 73-82, 2014.
- [13] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network," *Indonesian Journal of Computing and Cybernetics Systems*, vol. XI, no. 2, pp. 187-198, 2017.
- [14] R. D. Kusumah, B. Warsito and M. A. Mukid, "Perbandingan Metode K-Means Dan Self Organizing Map (Studi Kasus: Pengelompokan Kabupaten/Kota Di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia 2015)," *JURNAL GAUSSIAN*, vol. VI, no. 3, pp. 429-437, 2017.
- [15] D. G. Hawkin, "Modified Volume-Price Trend Indicator," Technical Analysis Inc, 2008.
- [16] R. Larson, Elementary Linear Algebra (7th ed.), Boston: Brooks/Cole, Cengage Learning, 2013.
- [17] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. XLV, no. 4, pp. 427-437, 2009.