

Penerapan Teknik Web Scraping pada Penjualan Artisan Tea di Shopee

Sylvia Candra Kurniawan*

Universitas Atma Jaya Yogyakarta; email: 225612165@students.uajy.ac.id

* Corresponding author

Abstrak

Teh adalah salah satu bahan yang lazim di pasaran. Penggemar teh mulai banyak dilihat dari beberapa kedai teh yang buka dan menjual khusus berbagai produk teh. Teh juga memiliki bermacam – macam khasiat menurut jenisnya. Artisan Tea merupakan teh yang umumnya dibuat dengan tangan dan dalam jumlah yang kecil oleh pengrajin teh menurut pengetahuan dan kemampuan pengrajin teh. Jenis Artisan Tea mulai dijual di pasaran terutama secara online. Beberapa toko telah menjual berbagai macam jenisnya. Jenis teh yang diminati oleh konsumen perlu diketahui untuk dapat dilakukan analisis mengenai teh dengan peminat terbanyak dan menyediakan stok lebih banyak untuk penjualan teh dengan jenis yang paling banyak laku terjual. Menggunakan teknik Web Scraping, dapat diketahui jenis teh dan jumlah penjualannya berdasarkan beberapa toko yang menjual teh. Analisis data dilakukan menggunakan R Studio untuk mengambil data yang tepat untuk dilakukan analisis. Penjualan teh dapat dilihat dan dilakukan analisis untuk mengetahui jenis teh yang paling banyak digemari oleh pembeli online. Berdasarkan hasil analisis 10 jenis teh yang memiliki penjualan terbanyak secara berurutan adalah Matcha, Jenis teh lain – lain, Earl Grey, Sencha, Chamomile, Teh Hitam, Ocha, Green Tea, Melati, dan Rosela. 5 berat produk dengan penjualan tertinggi secara berurutan adalah produk dengan berat produk 10 gr, 50g, 15gr, 40gr, dan 100gr. Berdasarkan hasil analisis tidak ditemukan hubungan antara berat produk dengan jumlah penjualan. Disarankan toko teh untuk menyediakan stok lebih banyak untuk 10 jenis teh dengan penjualan tertinggi dan tidak mengabaikan teh jenis lain-lain. Toko teh juga dapat mempertimbangkan mengemas teh dalam kemasan sample pack maupun kemasan besar yang disesuaikan dengan tujuan dan target toko.

Kata Kunci: Artisan Tea, Supply, Web Scraping, R Studio

Abstract

[Application of Web Scraping Techniques to Sales of Artisan Tea at Shopee] Tea is one of the most common ingredients in the market. Tea fans began to be seen from several tea shops that opened and specifically sold various tea products. Tea also has various benefits according to its type. Artisan tea is tea that is generally made by hand and in small quantities by tea craftsmen according to the knowledge and skills of tea craftsmen. This type of Artisan Tea is starting to be sold in the market, especially online. Several stores have sold various types. The types of tea that are in demand by consumers need to be known in order to be able to do an analysis of the teas with the most demand and to provide more stock for the sale of the types of tea that sell the most. Using Web Scraping technique, it can be known the type of tea and the amount of sales based on several shops that sell tea. Data analysis was performed using R Studio to retrieve the right data for analysis. Sales of tea can be seen and analyzed to find out which types of tea are most popular with online buyers. Based on the results of the analysis, the 10 types of tea that have the highest sales in order are Matcha, Other types of tea, Earl Grey, Sencha, Chamomile, Black Tea, Ocha, Green Tea, Jasmine, and Rosela. The 5 product weights with the highest sales in order are products with product weights of 10 grams, 50 grams, 15 grams, 40 grams, and 100 grams. Based on the analysis results, there is no relationship between product weight and total sales. It is recommended that tea shops provide more stock for the

10 types of tea with the highest sales and do not neglect other types of tea. Tea shops can also consider packaging tea in sample packs or large packs that are customized to the store's goals and targets.

Keywords: Artisan Tea, Supply, Web Scraping, R Studio

Kelompok BoK yang bersesuaian dengan artikel: *Engineering Management*

Saran format untuk mensitasi artikel ini:

Kurniawan, S.C. (2023). Penerapan Teknik *Web Scraping* pada Penjualan Artisan Tea di Shopee. *Prosiding Seminar Nasional Teknik Industri (SENASTI) 2023*, 730-737.

1. Pendahuluan

Teh mulai dijual secara *online* di berbagai toko yang menyediakan berbagai jenis teh yang digemari oleh masyarakat. Menurut Hossain dan Arefin (2018), *Artisan Tea* merupakan teh yang umumnya dibuat dengan tangan oleh pengrajin teh dalam jumlah yang kecil menurut pengetahuan dan kemampuan pengrajin teh. *Artisan* menurut Lingham et al. (2022) merupakan hal yang secara terus menerus dirancang kembali seiring dengan adanya inovasi-inovasi baru yang digunakan dan dihubungkan dengan tempat, produk, dan pembuatnya. Saat ini, penjualan *Artisan Tea* secara *online* dipilih oleh masyarakat dan antara 1 toko dengan lain terdapat persaingan penjualan teh. Toko yang menjual *Artisan Tea* tersebut telah memiliki data transaksi yang menunjukkan jenis teh yang laku dipasaran dan jumlah produk terjual secara *online*. Data tersebut dapat diolah untuk mengetahui jumlah stok yang dibutuhkan oleh toko tersebut agar jumlah stok tidak berkekurangan maupun stok berlebihan. Saat ini belum ditemukan adanya analisis pengendalian stok *Artisan Tea* untuk penjualan secara online melalui *e-commerce* yang tersedia di Indonesia.

Analisis untuk mengetahui jenis teh yang paling banyak dibeli oleh konsumen dapat menggunakan teknik *Web Scraping* untuk mengambil data di pasaran. Data diambil dari berbagai toko yang ada secara online. Data dapat diolah untuk kemudian dianalisis dan dibandingkan antar satu toko dengan yang lainnya. Jenis teh yang sudah diketahui dapat digunakan untuk menyediakan stok lebih banyak berdasarkan data penjualan. Menurut (H., Kasereka, 2020), *Web Scraping* merupakan suatu teknik berupa pengumpulan data melalui situs web secara otomatis. Data tersebut dapat digunakan untuk meningkatkan efisiensi bisnis *e-commerce* dan pemasaran serta mengembangkan promosi agar lebih efektif.

Menurut Lawson (2015), Peningkatan pesat jumlah data disebabkan oleh popularitas internet. Oleh karena itu diperlukan adanya *Web Scraping* yang mencari dan memindai setiap halaman web, mengekstraksi informasi, dan Menyusun indeks yang dapat dicari. Menurut Mooney et al. (2015), *Web Scraping* juga dikenal sebagai *Web Extraction* yang merupakan sebuah teknik untuk mengekstrak data dari *World Wide Web (WWW)* dan data disimpan pada sebuah sistem atau *database* untuk pengambilan atau dilakukan analisis. Pada saat ini, sejumlah besar data terus dihasilkan di *WWW* dan *Web Scraping* dikenal sebagai salah satu teknik yang efisien dan kuat untuk mengumpulkan data besar.

Web Scraping digunakan untuk mengambil data 3 toko *artisan tea* dari *e-commerce* Shopee. Data yang diambil adalah jenis produk atau nama produk *artisan tea* dan jumlah terjual.

Menurut (Almaqbal dkk, 2019), Data merupakan bahan baku yang diperlukan organisasi untuk membuat keputusan yang lebih baik. Saat ini, Sebagian besar data dapat ditemukan di internet.

World Wide Web (WWW) merupakan sistem yang memungkinkan pengguna untuk mengakses data yang tersedia di web. WWW merupakan arsitektur *client-server* artinya adalah data diakses dari server oleh klien. Data yang terdapat di WWW dirancang sebagai kerangka kerja konten besar, tidak langsung, dan terdistribusi. Kerangka kerja menggunakan hiperteks untuk saling menghubungkan antar dokumen. HTML dan HTTP merupakan protokol yang digunakan untuk mengambil dan menampilkan dokumen yang tersambung di WWW. WWW membuat informasi dari seluruh dunia dapat diakses. Pada WWW terdapat *Web Crawler* yang digunakan untuk membuat indeks dari web. Merupakan suatu program yang digunakan untuk memindai web agar dapat mengambil dan menemukan suatu informasi.

2. Metode

Data penjualan *artisan tea* diperoleh secara online melalui *Web Scraping* dari link yang berasal dari Shopee dengan melakukan studi literasi dan analisis mengenai toko yang memiliki penjualan *artisan tea* yang tinggi. Data penjualan diperoleh dari toko Haveltea, Heizl, dan Rumah Seduh. Data teh yang diambil adalah *artisan tea* dengan berbagai jenis tehnya. Data yang diambil adalah nama produk dan jumlah terjual dari produk tersebut. Nama produk biasanya mencantumkan juga berat produk. Menurut Hillen (2019), *Web Scraping* merupakan teknik yang digunakan untuk mengumpulkan data dari situs web secara otomatis. Teknik digunakan untuk berbagai tujuan termasuk penelitian ilmiah, analisis bisnis, dan pemantauan pasar. *Web Scraping* memungkinkan pengumpulan data jumlah besar dengan biaya yang minimal.

Data yang diterima kemudian dilakukan pembersihan data untuk menghilangkan data yang tidak termasuk kedalam *artisan tea* seperti alat seduh, kantong teh dan lain sebagainya. Data yang hanya terdiri dari jenis teh kemudian dipisahkan data berat dan jenis teh dari setiap nama. Pencarian berat dapat ditemukan dari angka yang tertera pada nama produk ataupun berdasarkan jenis kemasan produk. Pembersihan data dilakukan dengan menggunakan Microsoft Excel. Analisis dilakukan dengan menggunakan Microsoft Excel yang dilanjutkan dengan R Studio.

Data yang telah dibersihkan selanjutnya dianalisa jenis tehnya. Dicari data total penjualan (gram) untuk menyamakan jumlah produk terjual dengan produk yang memiliki beragam variasi berat. Data dengan jenis teh yang sama dijumlahkan total penjualan (gram) untuk mengetahui jenis teh dengan penjualan tertinggi. Analisis dilakukan dengan membuat grafik diagram batang total penjualan (gram) berdasarkan jenis tehnya.

Data yang telah dibersihkan selanjutnya dikelompokkan berdasarkan berat dan jumlah penjualan dijumlahkan untuk berat yang sama. Data teh jenis lain -lain divisualisasikan secara terpisah untuk melihat banyaknya data jenis teh lain-lain yang dapat dipertimbangkan untuk distok. Visualisasi data jenis teh lain-lain menggunakan grafik *tree map*.

Tree Map merupakan representasi visual dari data hierarkis. Data hierarkis merupakan data yang memiliki struktur seperti pohon. *Tree Map* menampilkan data hierarkis dengan bentuk persegi panjang untuk mewakili data. Ukuran persegi panjang dapat digunakan untuk mewakili nilai data dan warna serta bentuk dari persegi panjang dapat digunakan untuk mengurutkan atau mengkategorikan data persegi panjang, dan bentuk dari persegi panjang untuk mengetahui kategori atau urutan data. (T., Yasufumi., 2021)

Menurut Shah., K., dkk (2020), *Tree Map* adalah alat visualisasi yang dapat digunakan untuk menggambarkan data besar menjadi bagian yang lebih kecil. *Tree Map* digunakan untuk menunjukkan hubungan antara bagian data yang berbeda dan digunakan untuk mengidentifikasi tren dan pola dalam data.

Data selanjutnya dilakukan analisis korelasi dan analisis regresi untuk mengetahui hubungan antara jumlah penjualan dan berat produk. Menurut Chauhan (2015) analisis korelasi digunakan untuk mengetahui hubungan linear antara dua variabel dan mengukur seberapa besar suatu variabel mempengaruhi variabel lainnya.

Menurut Senthilnathan (2019) analisis korelasi menghasilkan nilai dengan rentang -1 sampai dengan +1. Menurut Gogtay dan Thatte (2017), koefisien korelasi sebesar 0,5 sampai dengan 1 berarti terdapat hubungan yang kuat antara kedua variabel. Koefisien korelasi lebih besar dari 0 dan lebih kecil dari 0,5 berarti terdapat hubungan yang lemah antar kedua variabel. Koefisien korelasi sebesar 0 menunjukkan tidak adanya hubungan antara kedua variabel.

Analisis regresi menurut Chatterjee dan Hadi (2013) adalah metode yang secara konseptual digunakan untuk mengetahui hubungan fungsional antar variabel. Menurut Kurniawan (2016) pada analisis regresi diperlukan penentuan variabel yang terikat dan variabel bebas untuk menunjukkan tingkat ketergantungan antar variabel. Menurut Sarstedt dan Mooi (2018), bila *p*-value lebih kecil dari 0,05 secara umum dapat disimpulkan bahwa variabel independen tertentu memiliki hubungan signifikan dengan variabel dependen.

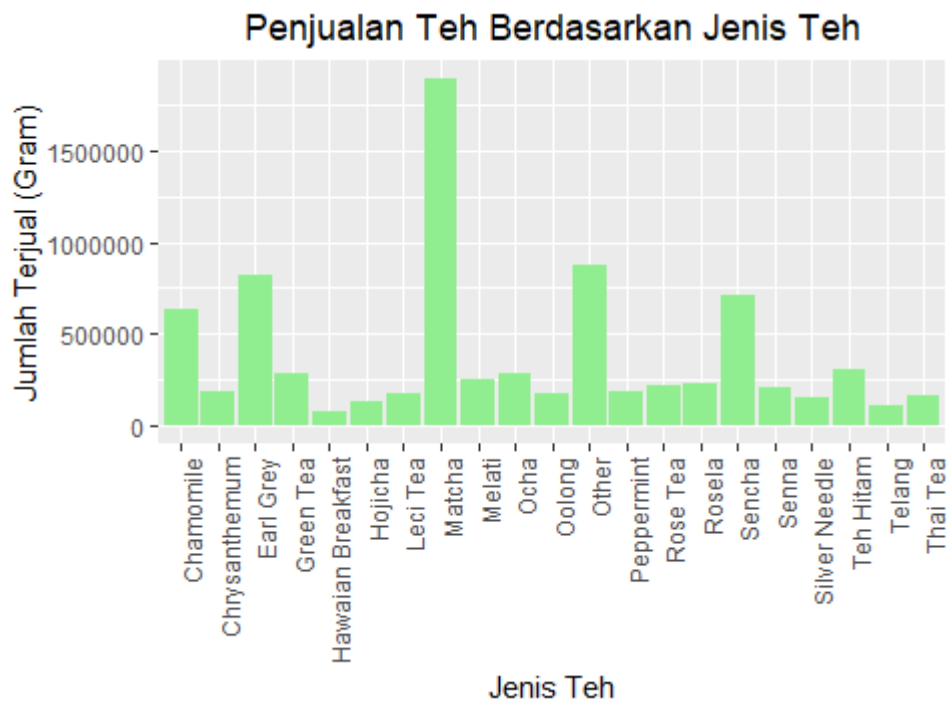
3. Hasil dan Pembahasan

Hasil pengelompokan 10 teh penjualan tertinggi beserta jumlah penjualannya (gram) dan persentase 10 teh penjualan tertinggi dapat dilihat pada tabel 1.

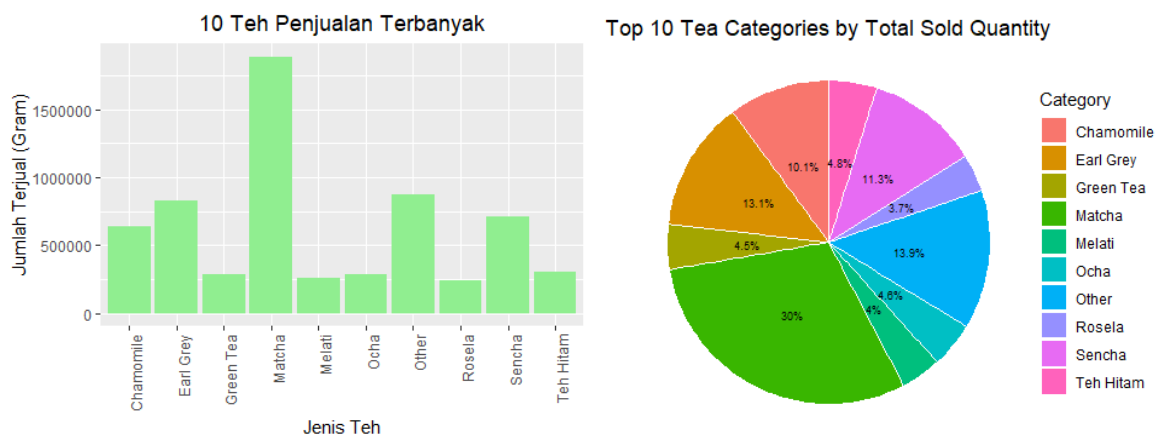
Tabel 1. Pengelompokan 10 teh dengan penjualan tertinggi.

Kategori	Total Terjual (Gram)	Persentase (%)
<i>Matcha</i>	1890800	30.2
<i>Earl Grey</i>	824660	13.2
<i>Other</i>	823733	13.2
<i>Sencha</i>	710200	11.4
<i>Chamomile</i>	639218	10.2
<i>Teh Hitam</i>	303653	4.9
<i>Ocha</i>	287500	4.6
<i>Green Tea</i>	281809.5	4.5
<i>Melati</i>	254689	4.1
<i>Rosela</i>	235941	3.8

Diagram batang digunakan untuk menunjukkan secara visual produk mana yang memiliki penjualan (gram) tertinggi. Grafik banyaknya penjualan (gram) terhadap berat produk dapat dilihat pada gambar 1. Dibuat diagram khusus untuk teh dalam 10 peringkat teratas dengan tujuan mempermudah melihat perbandingan banyaknya produk terjual dengan 10 peringkat teratas. Grafik banyaknya penjualan (gram) berdasarkan 10 jenis teh terbanyak dapat dilihat pada gambar 2. Grafik batang digunakan untuk membandingkan penjualan dan kisaran jumlah produk terjual, sedangkan grafik lingkaran digunakan untuk membandingkan besarnya penjualan teh dalam bentuk persentasi.

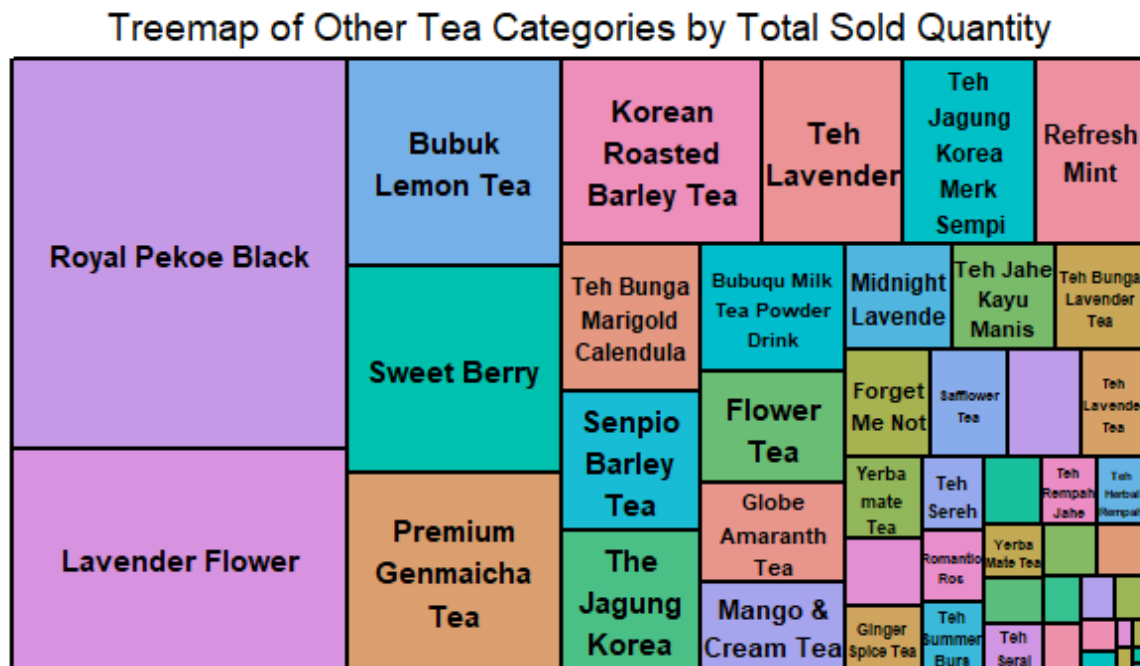


Gambar 1. Grafik Penjualan Teh Berdasarkan Jenis Teh



Gambar 2. Grafik Penjualan 10 Teh Penjualan Terbanyak

Jenis teh lain-lain dibuat dalam bentuk grafik treemap untuk memperlihatkan jenis teh dengan penjualan yang tinggi dan dapat dipertimbangkan untuk distok, sedangkan produk dengan besar blok yang kecil dianggap merupakan teh khusus toko tertentu atau teh dengan peminat yang lebih kecil sehingga blok dengan nama yang tidak terlihat namanya tidak dipertimbangkan. Grafik untuk teh jenis lain-lain dapat dilihat pada gambar 3.

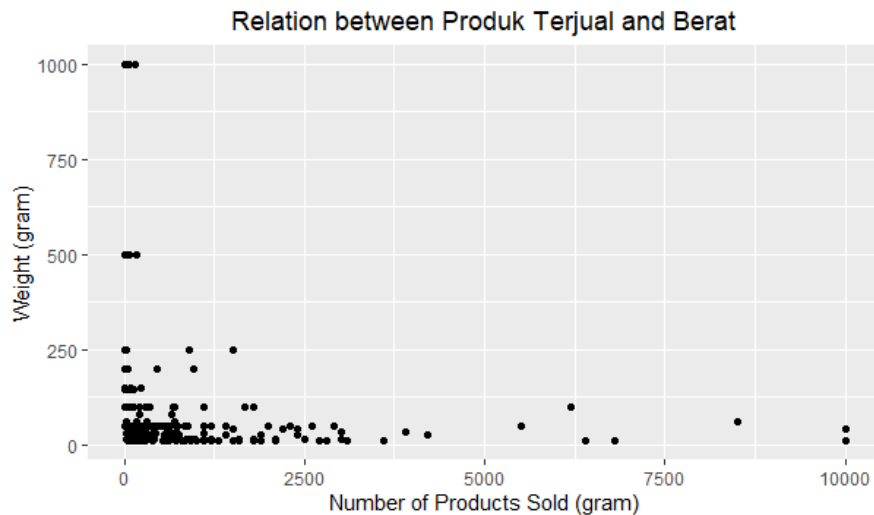


Gambar 3. Grafik Treemap Teh Jenis Lain-Lain

Berdasarkan hasil analisis, direkomendasikan untuk mempersiapkan stok yang lebih banyak untuk 10 jenis teh dengan total penjualan tertinggi dengan tidak mengabaikan teh jenis lain-lainnya. The jenis lain-lainnya yang dapat dipertimbangkan untuk di stok adalah the dengan luas area blok yang besar pada treemap. Hal ini dapat disesuaikan dengan karakteristik konsumen dari toko tersebut.

Analisis dilakukan untuk mengetahui hubungan antara berat kemasan dengan jumlah penjualan. Dilakukan analisis korelasi dan regresi antara berat kemasan yang sama dengan total jumlah penjualan. Didapatkan hasil korelasi sebesar -0,3 yang berarti hubungan antara berat kemasan dengan jumlah penjualan tidak kuat. Didapatkan hasil *P*-value pada analisis regresi sebesar 0,212 dimana *p*-value lebih besar dari 0,05 sehingga disimpulkan tidak ada bukti kuat untuk mengatakan bahwa berat kemasan memiliki pengaruh terhadap jumlah penjualan.

Analisis juga dilakukan untuk mengetahui hubungan antara berat the dengan jumlah total penjualan. Digunakan *scatter plot* untuk mengetahui hubungan antara berat the dengan jumlah total penjualan. Grafik *scatter plot* dapat dilihat pada gambar 4.



Gambar 4. Scatter plot produk terjual dan berat the

Berdasarkan hasil analisis scatter plot, produk dengan berat yang lebih rendah dari 125 gram memiliki jumlah penjualan yang lebih besar dibandingkan dengan produk dengan berat yang tinggi. Hal ini dianggap masuk akal karena konsumen e-commerce dapat berupa pemilik tea shop maupun pelanggan umum. Pelanggan yang membeli produk dengan berat yang tinggi akan melakukan pembelian yang lebih jarang dibandingkan dengan pelanggan yang membeli produk dengan berat yang rendah.

Toko dapat membuat variasi berat produk sesuai dengan tujuan dan target toko. Bila toko ingin meningkatkan jangkauan produk, toko dapat mengemas teh dengan berat yang rendah atau adanya kemasan sample pack sehingga pelanggan tidak ragu untuk mencoba jenis teh yang tersedia dan memperluas jangkauan toko. Apabila pelanggan menyukai jenis teh tersebut, pelanggan dapat membeli produk dengan berat yang lebih tinggi. Bila toko menargetkan konsumen dengan pembelian dengan berat yang tinggi (diasumsikan untuk pengusaha bisnis) maka toko dapat menjual produk dengan berat yang tinggi.

4. Kesimpulan

Analisis dilakukan menggunakan teknik *Web Scraping* dan menggunakan *software* Excel dan R Studio. Analisis hanya dilakukan untuk *Artisan Tea* pada *e-commerce* Shopee dan pada 3 toko yaitu Haveltea, Heizl, dan Rumah Seduh. 10 jenis teh dengan penjualan terbanyak dari ketiga toko *online* secara berurutan adalah Matcha, Jenis teh lain – lain, Earl Grey, Sencha, Chamomile, Teh Hitam, Ocha, Green Tea, Melati, dan Rosela.

Produk dengan berat yang lebih rendah dari 125 gram memiliki jumlah penjualan yang lebih besar dibandingkan dengan produk berat tinggi. Konsumen *e-commerce* terbagi menjadi 2 yaitu pelanggan umum dan pemilik *tea shop*. Pada konsumen *e-commerce*, pelanggan yang membeli produk dengan berat yang lebih tinggi akan melakukan pembelian yang lebih jarang dibandingkan dengan pelanggan yang membeli produk dengan berat yang rendah.

Rekomendasi yang diberikan adalah toko dapat mempersiapkan stok lebih banyak untuk 10 jenis teh dengan total penjualan paling tinggi dan tidak mengabaikan teh jenis lain – lain. Penyesuaian dapat dilakukan dengan mengetahui karakteristik konsumen dari toko tersebut. Toko dapat membuat variasi berat produk yang sesuai dengan tujuan dan target toko. Memperluas jangkauan produk dapat dilakukan dengan mengemas teh dengan berat

yang lebih rendah atau dengan adanya kemasan *sample pack* sehingga pelanggan dapat mencoba terlebih dahulu membeli berbagai jenis teh dalam kemasan kecil. Apabila toko menargetkan konsumen dengan pembelian jumlah besar maka toko dapat menjual produk dengan kemasan berat yang lebih tinggi dengan asumsi pembelian dilakukan oleh pengusaha bisnis atau pemilik *tea shop*.

Dimasa depan, penelitian dapat dilanjutkan untuk menganalisis jumlah data yang lebih banyak dan diketahui berapa jumlah stok jenis teh yang dibutuhkan agar stok tercukupi.

5. Daftar Pustaka

- Almaqbal, I. S., Al Khufairi, F. M., Khan, M. S., Bhat, A. Z., Ahmed, I. (2019). *Web scrapping: data extraction from websites*. Journal of Student Research.
- Chatterjee, S., & Hadi, A. S. (2013). *Regression analysis by example (4th ed.)*. John Wiley & Sons.
- Chauhan, D.A. K. (2015). *Research analytics*. Wiley Global Education Asia.
- Gogtay, N. J., & Thatte, U. M. (2017). *Principle of correlation analysis*. Journal of The Association of Physicians of India, Vol. 65. 78-81.
- Hillen, J. (2019). *Web Scraping for Food Price Research*. British Food Journal, 121, 3350–3361.
- Hossain, M. I., & Arefin, M. R. (2018). *Developing quality white tea (artisan tea) manufacturing protocol by using different clone with their different combinations in respect of bangladesh*. European Academic Research, VI(S). 1352.
- Jadeja, M., & Shah, K. (2020). *Tree-Map: A visualization tool for large data*. Semantic Scholar.
- Kasereka, H. (2020). *Importance of web scraping in e-commerce and e-marketing*. SSRN Electronic Journal.
- Lawson, J. M., & Schreck, G. (2015). *Web Scraping: a review of techniques and applications*. Journal of Data and Information Quality, 7(2), 117-138.
- Lingham, S., Hill, I., & Manning, L. (2022). *Artisan food production: What makes food 'artisan'?* (pp. 101-117). Springer
- Mehedi Shamrat, F. M. J., Tasnim, Z., Rahman, A. K. M. S., Nobel, N. I., & Hossain, S. A. (2020). *An effective implementation of web crawling technology to retrieve data from the world wide web (www)*. International Journal of Scientific & Technology Research, 9(1), 1252.
- Mooney, S. J., Westreich, D. J., & El-Sayed, A. M. (2015). *Web scraping for big data analytics: a survey*. Climatic Change, 140(3), 413-422.
- Sarstedt, M., & Mooi, E. (2018). *Regression Analysis*. Springer Texts in Business and Economics, 209–256. https://doi.org/10.1007/978-3-662-56707-4_7
- Senthilnathan, S. (2019). *Usefulness of Correlation Analysis*. *Social science research network*. <http://dx.doi.org/10.2139/ssrn.3416918>
- T., Yasufumi., T., Yuna., M., Yoshiyuki., S., Hiroki. (2021). *Treemap-based cluster visualization and its application to text data analysis*. Tokyo Metropolitan University.