Jurnal Ilmiah Teknologi Informasi Asia
Vol.17, No.1, Tahun 2023
ISSN: 2580-8397 (O); 0852-730X (P)

Klasterisasi Tingkat Penjualan pada Startup Panak.id dengan Algoritma K-Means

Sutha Wijaya Harjono[1], Nengah Widya Utami[2], I Gusti Agung Pramesti Dwi Putri[3]
[1,2,3]STMIK Primakara
[1]suthawijaya@outlook.com, [2]widya@primakara.ac.id, [3]pramesti@primakara.ac.id

ABSTRAK. Pengolahan data penjualan merupakan salah satu cara bagi perusahaan dalam menunjang pengambilan keputusan ke depannya. Salah satu cara mengolah data penjualan yaitu melakukan klasterisasi dengan teknik *data mining* yang hadir sebagai solusi perusahan untuk bisa mengolah data penjualan yang dimiliki dengan melakukan proses yang menghasilkan pengetahuan baru. Pengetahuan yang didapatkan dari data-data yang dimiliki biasa disebut sebagai *Knowledge Discovery in Database* (KDD) yang terdiri dari langkah-langkah seperti; pemilihan data, pembersihan data, pencarian pola (*data mining*), dan interpretasi. Pada penelitian ini, penulis akan membahas teknik dan aplikasi data mining menggunakan algoritma K-Means yang diaplikasikasikan dengan bahasa pemrograman Python. Penelitian dilakukan di perusahaan startup Panak.id dengan mengakses data penjualan produk olahan ternak pada periode Juli 2021 – Juli 2022. Hasil penelitian menunjukkan tingkat penjualan produk didominasi oleh produk-produk yang masuk ke dalam kategori kurang laris. Hal ini menunjukkan bahwa dari banyaknya jenis produk yang ditawarkan oleh Panak.id, masih banyak produk yang kurang diminati oleh masyarakat. Dari pengetahuan tersebut, pihak perusahaan dapat mengambil keputusan terkait strategi penjualan maupun inovasi pada produk-produk yang kurang laris dengan harapan dapat lebih meningkatkan penjualan. Selain itu, pihak perusahaan juga dapat menjadikan pengetahuan ini sebagai prediksi dalam hal pengontrolan stok agar produk-produk yang kurang laku dapat lebih terawasi produksinya untuk menghindari penumpukan stok di gudang.
Kata Kunci: Algoritma K-Means, *data mining*, data penjualan, KDD, klasterisasi

ABSTRACT. *Sales data processing is one way for companies to support future decision-making. One way to process sales data is to cluster with data mining techniques that are present as a company solution to be able to process sales data that is owned by carrying out processes that generate new knowledge. Knowledge obtained from owned data is commonly referred to as Knowledge Discovery in Database (KDD) which consists of steps such as; data selection, data cleaning, pattern search (data mining), and interpretation. In this study, the author will discuss data mining techniques and applications using the K-Means algorithm which is applied in the Python programming language. The research was conducted at the startup company Panak.id by accessing data on sales of processed livestock products for the period July 2021 – July 2022. The results showed that the level of product sales was dominated by products that were in the less-selling category. This shows the many types of products offered by Panak.id, there are still many products that are of little interest to the public. From this knowledge, the company can make decisions related to sales strategies and innovations in products that are not in demand with the hope of further increasing sales. In addition, the company can also use this knowledge as a prediction in terms of stock control so that products that are not selling well can be monitored more closely to avoid stock buildup in warehouses.*
Keywords: *Clustering, data mining, KDD, K-Means algorithm, sales data.*

PENDAHULUAN

Perkembangan teknologi informasi yang saat ini semakin pesat, mendorong para pemilik perusahaan untuk tetap bisa bersaing dalam mengembangkan bisnisnya agar dapat bertahan dalam persaingan. Salah satu persaingan yang terjadi adalah pada penjualannya, yang mana itu menuntut perusahaan untuk lebih meningkatkan efisiensi tingkat penjualan yang terjadi dengan memanfaatkan data-data penjualan yang dimiliki.

Perusahaan yang saat ini masih belum bisa mengacu pada data untuk melihat perkembangan tren, maka akan sulit untuk bisa tetap bersaing (Sasongko & E, 2020). Berdasarkan hasil riset yang dilakukan oleh Dell Technologies bersama Forrester Consulting, mengungkapkan bahwa perusahaan-perushaan di Indonesia dinilai masih belum cakap dalam mengolah data. Riset Dell melaporkan, 69% responden di Indonesia menyatakan bahwa data sangatlah bernilai bagi bisnisnya, namun hanya 22% yang sudah memanfaatkan datanya untuk mengembangkan bisnis mereka. Hal ini membuktikan bahwa masih banyak perusahaan-perusahaan di Indonesia yang belum memanfaatkan data yang dimiliki secara maksimal (Kuncorojati, 2021).

Salah satu startup investasi dan *marketplace* ternak di Denpasar yaitu Panak.id, juga belum memanfaatkan data-data penjualan yang dimiliki secara maksimal. Data tersebut hanya dijadikan sebagai arsip, dan tidak dimanfaatkan sebagai informasi yang dapat digunakan dalam pengambilan keputusan dalam hal

strategi penjualan dan juga pengadaan stok produk berdasarkan tingkat penjualan yang ada. Produk-produk olahan ternak yang dimiliki untuk dijual, tentu tidak semuanya termasuk ke dalam kategori sangat laris, laris, maupun kurang laris. Hal ini berpengaruh terhadap sistem pengadaan yang terjadi di Panak.id, dikarenakan produksi stok yang terus dilakukan tanpa memperhatikan tingkat penjualan dari produk yang dikeluarkan. Banyak stok produk yang menumpuk dan tidak terjual, bahkan tidak bisa dijual lagi akibat kadaluwarsa, yang mana ini menyebabkan kurangnya efisiensi penjualan jika dilihat dari perbandingan atau rasio antara pengadaan dan penjualannya.

Berdasarkan masalah tersebut, diperlukan sebuah solusi yang tepat untuk mengatasi masalah yang terjadi agar bisa meningkatkan efisiensi perbandingan atau rasio pengadaan dan penjualan produk. Solusi yang akan penulis berikan untuk mengatasinya adalah dengan melakukan *data mining* pada data-data penjualan yang dimiliki oleh Panak.id. *Data mining* merupakan bidang ilmu interdisipliner yang menyatukan teknik pembelajaran dari mesin (*machine learning*), pengenalan pola (*pattern recognition*), statistik, basis data, dan visualisasi untuk mengatasi masalah ekstraksi informasi dari basis data yang besar (Sari & Hartama, 2018). Metode yang akan digunakan adalah metode klasterisasi. Klasterisasi adalah sebuah proses menggabungkan satu set objek fisik atau abstrak ke dalam kelas dari objek yang sama, sehingga objek data dengan tingkat kemiripan maksimum berada dalam klaster yang sama, sebaliknya klaster berbeda akan ditempati oleh objek data dengan tingkat kemiripan minimum (Rohmatullah, Rahmalia, & Pradana, 2019). Untuk menerapkan klasterisasi, dapat menggunakan algoritma K-Means, yang merupakan salah satu metode *cluster analysis* non hirarki yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih klaster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu klaster yang sama, dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam klaster yang lain (Hasanah, Defit, & Nurcahyo, 2021).

Pada penelitian-penelitian sebelumnya telah dilakukan teknik data mining yang serupa, namun dengan studi kasus dan tahapan implementasi yang berbeda. Di penelitian ini, penulis ingin memberikan solusi untuk masalah yang terjadi pada startup Panak.id. Metode yang dipilih bukan hanya sekadar mengganti tempat meneliti, namun juga sangat memperhatikan dan menyesuaikan kebutuhan yang ada di lapangan. Perbedaan yang paling mendasar yaitu, pada penelitian-penelitian terdahulu hanya sampai dengan melakukan analisis data, sedangkan pada penelitian ini, penulis juga akan membangun sebuah program *data mining* sehingga dapat memudahkan pihak perusahaan untuk kembali mengklasterisasi data-data yang baru pada tahun berikutnya secara mandiri, hanya dengan menginputkan data penjualan terbaru.

Dengan begitu, Panak.id dapat mengetahui jumlah stok yang harus disediakan di gudang secara cepat dan mudah setelah mengetahui pengetahuan berupa tingkat penjualan pada masing-masing produk. Hal ini tentu dapat mengurangi terjadinya kerugian yang diakibatkan tidak terjualnya produk yang dikarenakan penumpukan stok. Selain itu, pihak Panak.id juga dapat memanfaatkan pengetahuan tersebut untuk megambil keputusan mengenai strategi apa yang sebaiknya mereka ambil untuk meningkatkan penjualan produk-produk yang berada pada klaster kurang laris.

Sebelum melakukan penentuan metode yang akan digunakan, penulis telah terlebih dahulu melakukan perbandingan pada beberapa algoritma *data mining* yang sesuai dengan masalah pada objek penelitian. Terdapat banyak metode yang dapat digunakan dalam melakukan data mining. Pada analisis klasterisasi, ada beberapa jenis algoritma yang dapat diterapkan. Pertama yaitu algoritma K-Means, yang memungkinkan setiap data untuk masuk ke dalam klaster tertentu pada suatu tahapan proses klasterisasi, dan pada tahapan berikutnya dapat berpindah ke klaster lain. Lalu yang kedua yaitu Hierarchical Clustering, yang tidak memperbolehkan sebuah data untuk berpindah ke klaster lain pada tahapan proses berikutnya. Dan yang ketiga, yaitu Overlapping Clustering dengan algoritma Fuzzy C-Means, yang memungkinkan sebuah data untuk menjadi anggota dari beberapa klaster sekaligus.

Berdasarkan perbandingan tersebut, dapat disimpulkan bahwa algoritma klasterisasi yang paling cocok untuk digunakan pada penelitian ini adalah algoritma K-Means, yang bersifat lebih dinamis, sehingga memungkinkan data untuk berpindah klaster sesuai dengan jumlah data yang di-*input*, dan juga pada satu tahapan *output*, data yang sama hanya boleh menempati satu klaster saja, yang mana ini akan lebih sesuai dengan solusi yang penulis berikan yaitu mengklasterisasi data penjualan sesuai tingkat penjualannya. Apabila 1 data bisa menempati 2 klaster yang berbeda, hal itu tentu menjadi ketidaksesuaian dan juga menyebabkan pihak Panak.id mengalami kesulitan dalam pengambilan keputusan nantinya.

Dengan dilakukannya penelitian ini, diharapkan dapat membantu pihak Panak.id dalam memanfaatkan data penjualan yang dimiliki, dengan memberikan *output* atau keluaran berupa pola atau pengetahuan baru seperti klaster produk, karakteristik klaster masing-masing produk, dan visualisasi pola dalam bentuk bagan *scatter plot* yang dapat digunakan untuk menentukan atau meningkatkan strategi dalam pengambilan

keputusan, sehingga dapat meningkatkan efisiensi proses bisnis pada bagian pengadaan stok produk dan juga meningkatkan strategi pemasarannya.

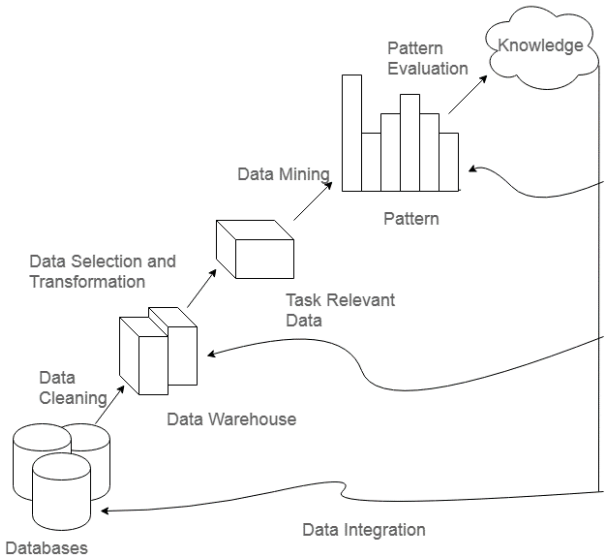
METODE

Metode Penelitian

KDD yang mewakili *Knowledge Discovery in Databases* atau Penemuan Pengetahuan dalam Basis Data, merupakan sebuah proses yang luas untuk menemukan pengetahuan dalam data dan menekankan aplikasi tingkat tinggi dari teknik penambangan data. KDD merupakan bidang yang menarik bagi para peneliti di berbagai bidang, seperti kecerdasan buatan, *machine learning*, pengenalan pola, basis data, statistik, akuisisi pengetahuan untuk sistem profesional, dan visualisasi data (Rahmah, 2020).

KDD menggunakan beberapa algoritma yang bersifat belajar mandiri (*self-learning*) untuk menyimpulkan pola yang berguna dari data yang diproses. Prosesnya adalah berupa umpan balik dari sebuah perulangan (*loop*) konstan yang tertutup, di mana banyak iterasi terjadi antara berbagai langkah sesuai dengan permintaan algoritma dan interpretasi pola (Muliono & Sembiring, 2019).

Langkah-langkah yang terlibat dalam proses KDD dapat ditunjukkan pada gambar di bawah ini (Andrean, Fendy, & Nugroho, 2019):



Gambar 1. Proses KDD

Berdasarkan gambar proses KDD di atas, dapat diuraikan prosesnya sebagai berikut (Andrean, Fendy, & Nugroho, 2019):

1. *Data Cleaning*, didefinisikan sebagai penghapusan data yang bising atau *noise* dan tidak relevan dari pengumpulan data.
2. *Data Integration*, didefinisikan sebagai data heterogen dari berbagai sumber yang digabungkan dalam satu sumber yang sama (*Data Warehouse*)
3. *Data Selection*, didefinisikan sebagai proses di mana data yang relevan dengan analisis diputuskan dan diambil dari pengumpulan data.
4. *Data Transformation*, didefinisikan sebagai proses mengubah data menjadi bentuk yang sesuai yang dibutuhkan oleh prosedur penambangan.
5. *Data Mining* atau penambangan data, didefinisikan sebagai teknik cerdas yang diterapkan untuk mengekstrak pola yang berpotensi berguna.
6. *Pattern Evaluation*, didefinisikan sebagai mengidentifikasi pola yang meningkat secara ketat yang mewakili pengetahuan berdasarkan ukuran yang diberikan.
7. *Knowledge Representation*, didefinisikan sebagai teknik yang menggunakan alat visualisasi untuk mewakili hasil *data mining*.

Jenis Data

Jenis data yang digunakan dalam penelitian data mining pada startup Panak.id ini adalah:

1. Data kuantitatif, merupakan jenis data berupa informasi numerik yang dapat diukur atau dihitung (Zein, et al., 2019). Pada penelitian ini, data kuantitatif yang digunakan adalah data penjualan untuk produk olahan ternak Panak.id pada periode Juli 2021 sampai dengan Juli 2022.
2. Data kualitatif, merupakan data yang tidak dapat diukur atau dihitung secara objektif, atau data yang mengungkapkan kualitas subjektif dan interpretatif dari suatu item atau proses (Hasanah H. , 2016). Data kualitatif yang digunakan dalam penelitian ini adalah proses bisnis perusahaan di bagian penjualan produk olahan ternak. Namun data ini hanya dijadikan sebagai data pendukung yang bisa membantu penulis dalam mengidentifikasi permasalahan yang dialami pada bagian penjualan dan pengadaan produk olahan ternak di perusahaan *startup* Panak.id.

Sumber Data

Sumber data terbagi menjadi dua, yaitu sumber data primer dan data sekunder. Data primer adalah data asli dan bersifat unik yang dikumpulkan secara langsung oleh peneliti dari suatu sumber seperti observasi, survei, angket, studi kasus dan wawancara sesuai dengan kebutuhannya. Berbeda dengan data sekunder yang lebih mudak diakses namun bersifat tidak murni lagi karena telah melalui banyak perlakuan statistik. Sumber data sekunder adalah publikasi pemerintah, website, buku, artikel jurnal, catatan internal. Sumber data yang yang dimanfaatkan pada penelitian ini adalah (Maulana, Susilo, & Riyadi, 2015):

1. Data primer, merupakan sumber data yang diperoleh atau dikumpulkan langsung dari sumber asli tanpa perantara (wawancara, kuesioner, observasi). Contoh data ini adalah proses bisnis perusahaan *startup* Panak.id.
2. Data sekunder, merupakan sumber data yang diperoleh secara tidak langsung dan melalui media perantara. Data sekunder dapat berasal dari sumber yang telah ada atau data sudah tersedia dan dikumpulkan oleh pihak lain (dokumentasi perusahaan). Dalam hal ini, peneliti memanfaatkan data sekunder yang sudah tersedia dan dikumpulkan oleh pihak perusahaan *startup* Panak.id berupa dokumentasi perusahaan pada bagian penjualan produk olahan ternak. Data tersebut nantinya akan diolah terlebih dahulu sebelum dapat digunakan secara lebih lanjut.

Tempat dan Waktu Penelitian

Penelitian ini mengambil data di perusahaan startup Panak.id yang beralamat di Jl. Nangka No. 16, Dangin Pura Kaja, Kec. Denpasar Utara, Kota Denpasar Bali 80234. Periode penelitian ini berlangsung dari bulan Januari sampai dengan November 2022, dengan mengakses data laporan penjualan dari bulan Juli 2021 sampai dengan Juli 2022.

Metode Pengumpulan Data

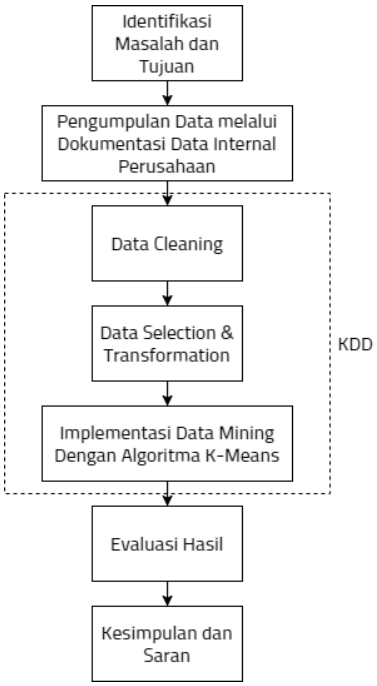
Metode pengumpulan data pada penelitian ini dilakukan dengan beberapa cara yang dijabarkan sebagai berikut:

1. *Library Research* (Penelitian Kepustakaan)
   Yaitu mengumpulkan, mempelajari, dan membandingkan data-data dari berbagai sumber berupa studi literatur dan sumber-sumber lain dengan permasalahan yang serupa atau memiliki kesamaan teknik dan metodologi penelitian (Sari & Asmendri, 2020), dimana ini bertujuan untuk membantu menambah pengetahuan serta wawasan penulis dalam memahami studi dalam bidang data science khususnya data mining, serta algoritma-algoritma yang dapat digunakan dalam pengimplementasiannya.
2. *Field Research* (Penelitian Lapangan)
   Yaitu pengumpulan data secara langsung di lapangan (Nurdiani, 2014), yang dalam hal ini adalah di lokasi kantor perusahaan startup Panak.id. Adapun beberapa cara yang dilakukan dalam penelitian ini adalah:
   a. Wawancara, yaitu cara pengumpulan data dengan melakukan tanya jawab kepada pihak Panak.id terkait informasi yang ingin dikumpulkan sebagai bahan penelitian. Wawancara dilakukan kepada Saudari Tetsuya Aisyah, selaku CEO dari startup Panak.id. Beberapa hal yang ditanyakan terkait dengan sekilas mengenai profil *startup*, proses bisnis perusahaan, dan permasalahan yang dialami di bagian produksi.
   b. Observasi, yaitu mengamati secara langsung proses bisnis yang terjadi pada bagian produksi dan penjualan produk olahan ternak di perusahaan *startup* Panak.id.

c. Dokumentasi, yaitu melakukan pengumpulan data dari dokumen-dokumen internal perusahaan berupa data penjualan pada periode tertentu. Data tersebut didapatkan dengan cara meminta secara langsung kepada pihak perusahaan yang kemudian dilakukan pendokumentasian data-data mentah penjualan pada periode yang diberikan. Data tersebut disimpan secara berurut di perangkat komputer sesuai dengan periodenya agar bisa digunakan secara baik selama penelitian ini berlangsung.

Alur Penelitian

Alur penelitian ini dapat digambarkan secara umum seperti pada gambar berikut ini:



Gambar 2. Alur Penelitian

Secara umum, bagan alur penelitian tersebut dapat dijelaskan sebagai berikut:

1. Identifikasi masalah dan tujuan, yaitu mendefinisikan masalah-masalah penelitan dan tujuannya yang akan dijadikan bahan penelitian.
2. Pengumpulan data, yaitu dilakukan dengan cara wawancara ke pihak perusahaan yang akan diteliti, dan melakukan pendokumentasian data-data internal perusahaan.
3. *Data cleaning*, yaitu pembersihan data-data mentah perusahaan sebelum dapat digunakan secara lebih lanjut.
4. *Data selection & transformation*, yaitu melakukan pemilahan lebih lanjut terkait data-data yang akan digunakan, lalu ditransformasi sesuai dengan keperluan pengolahan.
5. Implementasi *data mining*, yaitu mengimplementasikan teknik *data mining* menggunakan algoritma K-Means terhadap data-data yang telah diolah tadi untuk dianalisis dan ditemukan polanya sehingga menghasilkan informasi baru yang dapat membantu perusahaan dalam membuat keputusan.
6. Evaluasi hasil, yaitu penaksiran atau penilaian terkait hasil dari penelitian yang telah dilakukan dengan menghitung tingkat akurasi untuk memastikan bahwa hasil klasterisasi sudah valid.
7. Kesimpulan dan saran, yaitu membuat kesimpulan, serta memberikan saran untuk penelitian-penelitian kedepannya.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Proses pengumpulan data dilakukan secara langsung dengan mengunjungi kantor Panak.id. Data yang digunakan berasal dari dokumen penjualan bulanan produk olahan ternak dalam format file Excel. Pengumpulan data ini dilakukan setiap bulan sesuai dengan kebutuhan data yang akan digunakan, yaitu berkisar dari bulan Juli 2021 sampai dengan Juli 2022, dengan total data yang terkumpul sejumlah 1.728 buah

data. Beberapa variabel pada data yang sudah dikumpulkan akan dihilangkan untuk menyesuaikan kebutuhan penelitian.

Tabel 1. Sampel Data Penjualan

No	Tanggal	Produk	Terjual
1.	Kamis, 1 Juli 2021	Susu Kambing	2
2.	Kamis, 1 Juli 2021	Lele Bumbu Kuning 400gr	2
3.	Kamis, 1 Juli 2021	Lele Bumbu Kuning 400gr	2
4	Kamis, 1 Juli 2021	Lele Bumbu Kuning 400gr	5
5	Kamis, 1 Juli 2021	Lele Bumbu Kuning 400gr	3
6	Kamis, 1 Juli 2021	Lele Bumbu Kuning 400gr	3
7	Kamis, 1 Juli 2021	Lele Mati Bersih	2
8	Jumat, 2 Juli 2021	Lele Nugget	2
9	Jumat, 2 Juli 2021	Lele Nugget	4
10	Jumat, 2 Juli 2021	Lele Bumbu Kuning 400gr	2
	…	…	…
11	Sabtu, 30 Juli 2022	Ayam Bumbu Kuning 250gr	1

Sumber: Data Penjualan Panak.id Bulan Buli 2021 – Juli 2022

Dari hasil pengumpulan data, didapatkan 26 produk yang akan menjadi kategori dimana setiap data yang memiliki nama sama akan diakumulasikan untuk menghasilkan variabel total ‘Terjual’ dan dihitung jumlah munculnya untuk menghasilkan variabel total ‘Transaksi’.

Proses KDD

Proses pembersihan data dilakukan setelah data terkumpul. Tujuan dari pembersihan data adalah untuk memastikan tidak adanya *error* berupa kesalahan tipografi, duplikat, data tidak lengkap atau hal lainnya yang menyebabkan data tersebut tidak sesuai dengan salah satu dari 26 kategori yang ada pada data penjualan. Pada tahapan ini sudah cukup terbantu oleh data penjualan mentah yang bersifat ‘bersih’, dikarenakan pihak penyedia data sudah menggunakan *list* produk pada aplikasi Excel sehingga meminimalisir kemungkinan terjadinya salah ketik pada data penjualan yang di-*input*.

Namun, untuk memastikan data sudah benar-benar bersih, *testing data* baru akan dilakukan pada tahap implementasi agar lebih menghemat tenaga dan waktu. Jika ada data yang *error* berupa kesalahan tipografi dan lain sebagainya, akan langsung terdeteksi pada implementasi karena data tersebut akan membentuk kategori produk baru, yang mana itu berarti tidak sesuai dengan kategori yang sudah ditetapkan pada pengumpulan data.

Pada tahap *data selection & transformation*, variabel mana saja yang akan digunakan ketika implementasi *data mining* akan ditentukan. Sesuai dengan tujuan penelitian yaitu untuk mengetahui tingkat penjualan produk pada *startup* Panak.id, maka variabel yang akan digunakan adalah ‘Jumlah Transaksi’ dan ‘Total Terjual’ dari masing-masing produk.

Variabel ‘Jumlah Transaksi’ dapat diperoleh dengan cara mengakumulasikan jumlah munculnya suatu produk dalam data penjualan, sedangkan untuk variabel ‘Total Terjual’ diperoleh dengan cara menjumlahkan total terjualnya masing-masing produk. Setelah variabel ditentukan, dan ditentukan logika pembentukan variabelnya, maka akan terbentuk *dataset* baru yang nantinya akan langsung digunakan dalam implementasi klasterisasi menggunakan algoritma K-Means.

Penerapan klasterisasi dengan algoritma K-Means dalam penelitian ini dilakukan dengan bantuan bahasa pemrograman Python, yang ditulis menggunakan *code editor* Visual Studio Code. Untuk memudahkan penggunaan, program yang dibangun adalah dalam bentuk *Graphical User Interface* (GUI) dengan memanfaatkan Tkinter sebagai *library* dari UI-nya. Selain itu, untuk memvisualisasikan hasil klasterisasi dapat memanfaatkan *library* Matplotlib, serta Pandas dan Sklearn sebagai pemroses algoritma K-Means. Adapun langkah – langkah implementasinya, yaitu sebagai berikut:

1. Implementasi dengan Python

Program yang dibuat memungkinkan pengguna untuk mengimpor *file* penjualan yang sudah terlebih dahulu melalui proses transformasi data, sehingga memerlukan pendefinisian file Excel pada Python ke dalam bentuk *array*. Untuk melakukannya, dapat menggunakan *code* sebagai berikut:

Tabel 2. *Source Code* untuk *Import* File Excel

```
global df, df5, df4
import_file_path = filedialog.askopenfilename()
read_file = pd.read_excel (import_file_path)
df = DataFrame(read_file, columns=['Terjual', 'Produk'])
df1 = df.groupby(['Produk'])['Terjual'].count().reset_index()
df2 = df1.rename(columns = {"'Terjual":"Transaksi"})
df3 = df.groupby('Produk',
sort=False)["'Terjual"].sum().reset_index(name="'Terjual")
df4 = df2.merge(df3)[['Produk','Transaksi','Terjual']]
df5 = df4[['Transaksi','Terjual']]
```

Proses pengimporan file ini juga secara langsung mem-*filter* variabel-varibel yang diperlukan saja, yaitu variabel ‘Produk’, dan ‘Terjual’. Selanjutnya, Produk dengan nama yang sama, akan dikelompokkan, lalu dihitung jumlah munculnya sehingga membentuk variabel baru dengan nama ‘Transaksi’. Selain itu, jumlah terjual untuk masing-masing produk juga dijumlahkan untuk memperoleh total 'Terjual’, sehingga variabel akhir yang akan digunakan dalam proses K-Means adalah variabel ‘Transaksi’, dan ‘Terjual’.

Setelah file Excel berhasil dipanggil, maka proses K-Means siap untuk dilakukan. Untuk melakukannya, dapat menuliskan *code* sebagai berikut:

Tabel 3. *Source Code* K-Means

```
global df5,df4
global numberOfClusters
numberOfClusters = int(self.entryInput.get())

kmeans = KMeans(n_clusters=numberOfClusters).fit(df5)
centroids = kmeans.cluster_centers_
```

Meskipun pada penelitian ini hanya memerlukan 3 buah klaster, terdapat penambahan fitur yang memungkinkan pengguna untuk menginputkan jumlah klaster yang diinginkan, sehingga dapat menyesuaikan dengan kebutuhan mereka kedepannya.

Untuk memvisualisasikan hasil klaster, *scatter plot* digunakan, yang nantinya akan menampilkan posisi masing-masing data dan juga titik klasternya dalam bentuk petak sebar. Untuk merealisasikannya, dapat menuliskan *code* sebagai berikut:

Tabel 4. *Source Code* untuk Visualisasi *Scatter Plot*

```
figure1 = plt.Figure(figsize=(3,4), dpi=100)
figure1.patch.set_facecolor('#ffffff')
ax1 = figure1.add_subplot(111)
ax1.patch.set_facecolor('#ffffff')
ax1.scatter(df5['Transaksi'], df5['Terjual'], c= kmeans.labels_.astype(float), s=25, alpha=0.5)
ax1.scatter(centroids[:, 0], centroids[:, 1], c='red', s=25, marker="*")
scatter1 = FigureCanvasTkAgg(figure1, master=self.label_info_1)
scatter1.get_tk_widget().grid(row=0, column=0, pady=0, padx=0, stick="nwe")
```

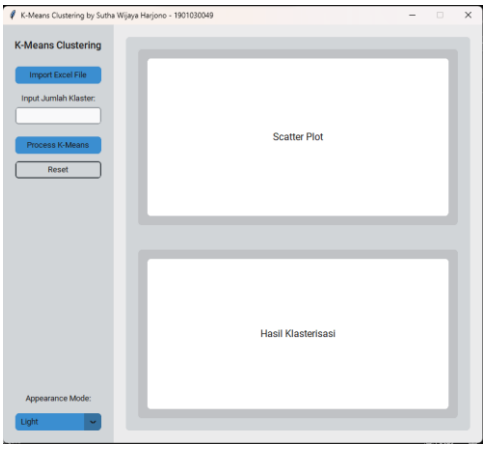
Selain visualisasi hasil klaster, hasil akhir data-data yang sudah diklasterisasi dalam bentuk teks juga ditambahkan, dan dilakukan penyortiran dari data produk yang paling laris ke yang kurang laris. Variabel akhir yang ditampilkan adalah: Produk, Transaksi, Terjual, Klaster. Untuk melakukannya, dapat menuliskan *code* sebagai berikut:

Tabel 5. *Source Code* Hasil Akhir Klaster

```
y_predicted = kmeans.fit_predict(df5[['Transaksi','Terjual']])
df5['Klaster'] = y_predicted
df4['Terjual'] = df4["Terjual"].astype(int)
df6 = df5['Klaster']
dframe = pd.DataFrame(pd.concat([df4,df6], axis =1, ignore_index=True))
dframe.columns=['Produk','Transaksi','Terjual','Klaster']
txt = Text(master=self.label_info_2, width=1)
txt.grid(row=0, padx=0, pady=0, stick="nwes")
class PrintToTXT(object):
    def write(self, s):
        txt.insert(END, s)
sys.stdout = PrintToTXT()
dframe1 = dframe.sort_values(by='Terjual', ascending=False)
dframe2 = dframe1.reset_index(drop=True)
dframe2.index += 1
print(dframe2)
```

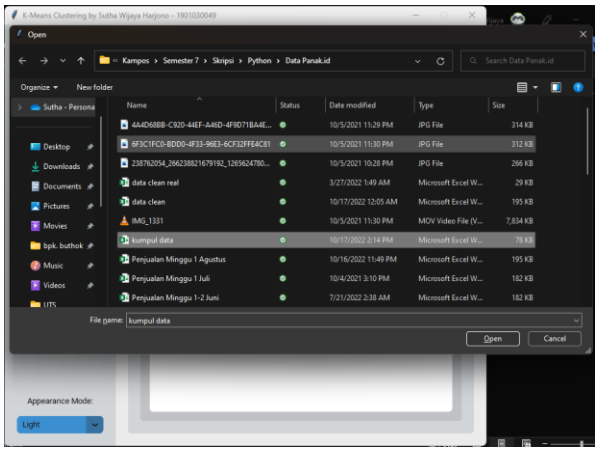
2. Hasil Implementasi GUI

Dengan memanfaatkan Python TKinter, UI program dapat dibentuk agar lebih mudah digunakan oleh pihak perusahaan nantinya. Sesuai dengan proses implementasi Python yang sudah dilakukan, berikut ini adalah hasil implementasi GUI-nya:



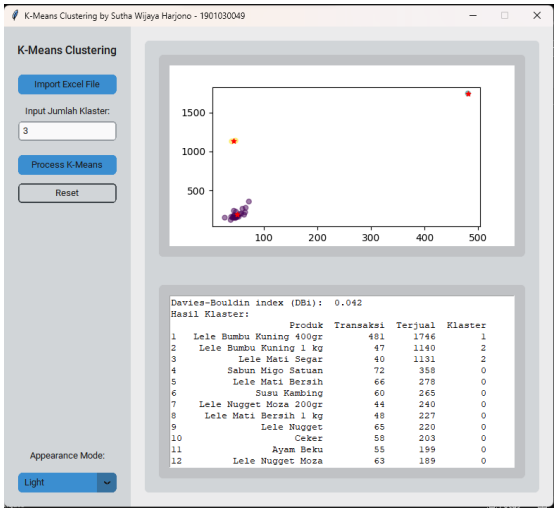
Gambar 3. Hasil Implementasi GUI

Langkah pertama yang dilakukan yaitu melakukan impor file Excel dengan mengklik tombol ‘Import Excel File’, lalu memilih file Excel yang diperlukan, dan klik ‘Open’.



Gambar 4. Proses *Import* File Excel

Pada jendela ‘Open’, *file* Excel dengan nama ‘kumpul data.xlsx’ digunakan. Setelah *file* berhasil diimpor, pengguna dapat menginputkan jumlah klaster, yang dalam penelitian ini, membutuhkan 3 buah klaster. Jika sudah, maka proses K-Means dapat langsung dilakukan dengan mengklik tombol ‘Process K-Means’ sehingga akan langsung muncul hasilnya di sebelah kanan antarmuka.



Gambar 5. Hasil Klaster

3. Penjelasan Hasil Klaster

Berdasarkan Proses K-Means yang sudah dilakukan, didapatkan hasil produk yang sudah terklaster sebagai berikut:

Tabel 6. Hasil Klaster

Produk	Transaksi	Terjual	Klaster
Lele Bumbu Kuning 400 gr	481	1746	1
Lele Bumbu Kuning 1 kg	47	1140	2
Lele Mati Segar	40	1131	2
Sabun Migo Satuan	72	358	0
Lele Mati Bersih 500 gr	66	27	0
Susu Kambing	60	265	0
Lele Nugget Moza 200 gr	44	240	0
Lele Mati Bersih 1 kg	48	227	0
Lele Nugget	65	220	0
Ceker	58	203	0
Ayam Beku	55	199	0
Lele Nugget Moza 100 gr	63	189	0
Ati Ampela	45	179	0
Keripik Kulit Lele	44	179	0
Ayam Potong	51	175	0
Lele Bumbu Kuning 2 kg	51	175	0
Sabun Migo Box	53	163	0
Ayam 1 kg	39	162	0
Keripik Tulang Lele	42	161	0
Lele Nugget Moza 400 gr	46	159	0
Ayam Bumbu Kuning 150 gr	49	456	0

Ayam Bumbu Kuning 1 kg	27	151	0
Ayam Bumbu Kuning 500 gr	45	149	0
Lele Bumbu Kuning 150gr	42	248	0
Sabun Migo 2 pcs	47	147	0
Ayam Bumbu Kuning 250 gr	38	122	0

Berdasarkan hasil klasterisasi, berikut ini adalah penjabaran produk per klasternya beserta total produk pada masing-masing klaster:

a. Klaster 0 (23 Produk): Sabun Migo Satuan, Lele Mati Bersih, Susu Kambing, Lele Nugget Moza 200 gr, Lele Mati Bersih 1 kg, Lele Nugget, Ceker, Ayam Beku, Lele Nugget Moza 100 gr, Ati Ampela, Keripik Kulit Lele, Ayam Potong, Lele Bumbu Kuning 2 kg, Sabun Migo Box, Ayam 1kg, Keripik Tulang Lele, Lele Nugget Moza 400 gr, Ayam Bumbu Kuning 150gr, Ayam Bumbu Kuning 1kg, Ayam Bumbu Kuning 500 gr, Lele Bumbu Kuning 150gr, Sabun Migo 2 pcs, dan Ayam Bumbu Kuning 250 gr.
b. Klaster 1 (1 Produk): Lele Bumbu Kuning 400 gr.
c. Klaster 2 (2 Produk): Lele Bumbu Kuning 1 kg, dan Lele Mati Segar.

4. *Centroid* Terakhir

Dalam pengaplikasian K-Means, titik pusat klaster atau *centroid* akan berakhir pada suatu titik, yang menjadi titik pusat dimana setiap data menempati klaster-klaster yang sesuai. Untuk mengetahuinya, hanya perlu memanggil kembali variabel ‘centroid’ pada *code* di Python, lalu menyusunnya dalam bentuk tabel dengan menuliskan *code* sebagai berikut:

Tabel 7. *Source Code Centroid* Terakhir

```
print("Centroid Terakhir:")
centroidAkhir = pd.DataFrame(
    {
        "Centroid" : ['C0','C1','C2'],
        “Transaksi" : [round(centroids[0,0], 2), round(centroids[1,0],2),
round(centroids[2,0],2)],
        "Terjual" : [round(centroids[0,1],2), round(centroids[1,1],2),
round(centroids[2,1],2)]
    }
)
print(centroidAkhir)
print("")
print("Jumlah Produk Dalam Klaster")
print("Klaster 0 = ", dframe1[dframe1['Klaster']==0]['Produk'].count())
print("Klaster 1 = ", dframe1[dframe1['Klaster']==1]['Produk'].count())

print("Klaster 2 = ", dframe1[dframe1['Klaster']==2]['Produk'].count())
```

Dengan hasil pada GUI sebagai berikut:

```
Centroid Terakhir:
  Centroid  Transaksi  Terjual
0       C0        50.0   191.52
1       C1       481.0  1746.00
2       C2        43.5  1135.50
```

Gambar 6. *Centroid* Terakhir di GUI

Berdasarkan *Centroid* Terakhir yang sudah dipanggil dan ditampilkan hasilnya di dalam GUI, berikut ini adalah penjelasannya:

a. C0: [50 191,52]
   Centroid 0 memiliki rata-rata total transaksi sebesar 50 kali, dengan rata-rata total terjual sebesar 191,52 buah. Ini menunjukkan bahwa produk-produk yang masuk ke dalam C0 adalah yang termasuk ke dalam kategori 'Kurang Laris'.
b. C1: [481 1746]
   Centroid 1 memiliki rata-rata total transaksi sebanyak 481 buah, dengan rata-rata total terjual sebesar 1.746 buah. Ini menunjukkan bahwa produk yang masuk ke dalam C1 merupakan produk yang termasuk ke dalam kategori 'Sangat Laris'.
c. C2: [43,5 1135,5]
   Centroid 2 memiliki rata-rata total transaksi sebanyak 43,5 buah, dengan rata-rata total terjual sebesar 1.134,5, yang menunjukkan bahwa produk-produk yang masuk ke dalam C2 adalah yang termasuk ke dalam kategori 'Laris'.

5. Tingkat Akurasi

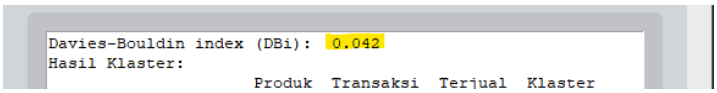
Untuk mengukur seberapa baik klasterisasi yang sudah dilakukan, maka perlu melakukan evaluasi algoritma dengan menggunakan Davies-Bouldin Index. Apabila nilai yang dihasilkan semakin mendekati 0, maka dapat dikatakan klasterisasi yang dilakukan sudah baik. Begitu pula sebaliknya, apabila nilai mendekati angka 1, maka klasterisasi yang dilakukan adalah buruk dan perlu dikaji kembali data ataupun jumlah klaster yang digunakan. Langkah ini dapat dilakukan secara langsung pada Python, menggunakan *code* sebagai berikut:

Tabel 8. *Source Code* Davies-Bouldin Index

```
dfArray = pd.DataFrame(df5, columns = ['Transaksi', 'Terjual'])
myArray = dfArray.to_numpy()
kmeans = KMeans(n_clusters=numberOfClusters, random_state=30)
labels = kmeans.fit_predict(myArray)
db_index = davies_bouldin_score(myArray, labels)

print("Davies-Bouldin index (DBi): ", round(db_index, 3))
```

Dengan hasil pada GUI sebagai berikut:



Gambar 7. Hasil Evaluasi DBi

Dari hasil evaluasi, didapatkan hasil 0,042 yang berarti klasterisasi yang dilakukan memiliki tingkat akurasi yang sangat baik, dimana jarak antar data dalam satu klaster memiliki tingkat kemiripan yang sangat rapat atau tinggi, dan juga jarak antar klaster memiliki perbedaan yang signifikan.

KESIMPULAN DAN SARAN

Berdasarkan hasil dari penerapan *data mining* pada data penjualan produk-produk olahan ternak pada *startup* Panak.id menggunakan algoritma K-Means, maka dapat disimpulkan bahwa algoritma K-Means dapat diimplementasikan pada proses klasterisasi menggunakan bahasa pemrograman Python. Pengujian klaster dikelompokkan menjadi 3 klaster yaitu: sangat laris, laris, dan kurang laris. Secara garis besar penjualan produk olahan ternak pada *startup* Panak.id masuk ke dalam kategori 'kurang laris', yaitu sebanyak 23 jenis produk, diikuti oleh produk dengan kategori 'laris' sebanyak 2 jenis, dan produk 'sangat laris' sebanyak 1 jenis produk. Sehingga dapat disimpulkan bahwa tingkat penjualan dengan jenis produk terbanyak adalah kategori kurang laris yaitu sebanyak 23 jenis produk, yang berarti ada sangat banyak jenis produk yang ditawarkan oleh Panak.id yang tidak diminati oleh masyarakat. Adapun hasil evaluasi menggunakan Davies-Bouldin Index diperoleh sebesar 0,042 sehingga tingkat akurasi dari hasil klaster termasuk sangat baik, karena nilai DBI sudah mendekati angka nol.

Beberapa saran yang dapat penulis sampaikan untuk penelitian ini dan juga perusahaan *startup* Panak.id, antara lain: menjadikan penelitian ini sebagai bahan rujukan bagi para peneliti yang akan melakukan penelitian

dengan topik sejenis dan dapat menambahkan algoritma klasterisasi partisi lain seperti K-Medoids sehingga dapat digunakan sebagai pembanding; hasil penelitian dapat dijadikan acuan oleh pihak *startup* Panak.id terutama pada produk-produk yang masuk ke dalam kategori 'Kurang Laris' agar bisa lebih meningkatkan strategi maupun inovasi produk, serta melakukan prediksi yang dapat membantu dalam hal pengontrolan stok produk di gudang; program atau aplikasi yang dibangun dapat digunakan dengan baik dan membantu pihak *startup* Panak.id dalam melakukan klasterisasi pada data penjualan mereka untuk mengetahui tingkat penjualan masing-masing produk; menggunakan format penulisan produk pada data penjualan sesuai dengan yang sudah ditetapkan dalam penelitian ini untuk menghindari *error* ketika penggunaan program atau aplikasi klasterisasi.

DAFTAR RUJUKAN

Andrean, R., Fendy, S., & Nugroho, A. (2019). Klasterisasi Pengendalian Persediaan Aki . *Journal of Information Technology and Computer Science*, 5-12.

BPS. (2016, Mei 20). *Jumlah Penduduk dan Laju Pertumbuhan Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Timur, 2010, 2014, dan 2015*. Retrieved from Badan Pusat Statistik Provinsi Jawa Timur: http://jatim.bps.go.id

Hasanah, H. (2016). TEKNIK-TEKNIK OBSERVASI. *Jurnal at-Taqaddum*, 21-46.

Hasanah, M., Defit, S., & Nurcahyo, G. W. (2021). Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade . *Jurnal Sistim Informasi dan Teknologi*, 30-35.

Kuncorojati, C. (2021). *Perusahaan Indonesia Dinilai Belum Cakap Mengolah Data*. (medcom.id) Retrieved from https://www.medcom.id/teknologi/news-teknologi/aNr91RVK-perusahaan-indonesia-dinilai-belum-cakap-mengolah-data

Maulana, S. M., Susilo, H., & Riyadi. (2015). IMPLEMENTASI E-COMMERCE SEBAGAI MEDIA PENJUALAN ONLINE (STUDI KASUS PADA TOKO PASTBRIK KOTA MALANG). *Jurnal Administrasi Bisnis*, 1-9.

Muliono, R., & Sembiring, Z. (2019). DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA. *DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA*, 272-279.

Nurdiani, N. (2014). TEKNIK SAMPLING SNOWBALL DALAM PENELITIAN LAPANGAN . *ComTech*, 1110-1118.

Rahmah, S. A. (2020). KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA). *Journal of Information Technology Research*.

Rohmatullah, A., Rahmalia, D., & Pradana, M. S. (2019). KLASTERISASI DATA PERTANIAN DI KABUPATEN . *Jurnal Ilmiah Teknosains*, 86-93.

Sari, M., & Asmendri. (2020). Penelitian Kepustakaan (Library Research) dalam Penelitian Pendidikan IPA. *Jurnal Penelitian Bidang IPA dan Pendidikan IPA*, 41-53.

Sari, R. W., & Hartama, D. (2018). Data Mining: Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi*, 322-326.

Sasongko, Y. A., & E, A. D. (2020). *Big Data Semakin Ngetren, SDM Kompetensi Data Science Dilirik Industri*. (KOMPAS.com) Retrieved from https://edukasi.kompas.com/read/2020/12/10/093500371/big-data-semakin-ngetren-sdm-kompetensi-data-science-dilirik-industri

Utami, N. W., & Paramitha, A. I. (2021). PENERAPAN DATA MINING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI DI STMIK PRIMAKARA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. *JUTIK*, 456-463.

Zein, S., Yasyifa, L., Ghozi, R., Harahap, E., Badruzzaman, F., & Darmawan, D. (2019). PENGOLAHAN DAN ANALISIS DATA KUANTITATIF . *Jurnal Teknologi Pendidikan dan Pembelajaran*, 839-845.