

Nama : Kevin Mohammad Arsy
NIM : A11.2019.12084
Kelompok : A11.4619

LAPORAN HASIL EKSPERIMEN TUGAS AKHIR DATA MINING

1. JUDUL

“Penerapan Metode *Clustering K-Means* Berdasarkan Model RFM Untuk *Customer Segmentation Online Retail* Dalam Periode 8 Bulan Pada Tahun 2010”

2. PEMBUATAN KODING PYTHON

Pada pembuatan koding untuk eksperimen data mining ini menggunakan bahasa pemrograman Python dan text editor Visual Code. Berikut adalah penjelasan mengenai langkah-langkah serta source code dalam eksperimen data mining tersebut :

- Import Library

```
#import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Library yang digunakan dalam eksperimen ini menggunakan pandas, numpy, matplotlib.

- Memanggil dataset dari file .csv

```
#Memanggil dataset file .csv
df =pd.read_csv("G:\\Kuliah\\Semester6\\Data
Mining\\TA\\OnlineRetail.csv", encoding='unicode_escape')
#Melihat semua nama kolom
df.keys()
#Menemukan no. dari baris dan kolom
df.shape
#Melihat beberapa detail dasar kumpulan data
df.describe()
```

Dalam eksperimen ini, menggunakan dataset dengan nama OnlineRetail dengan format .csv

3. IMPLEMENTASI SOURCE CODE UNTUK KLASTERISASI

Berikut adalah implementasi dari proses klasterisasi data *online retail* di UK dalam periode 8 bulan pada tahun 2010.

- Menghapus redundansi/preprocessing data

```
#Menghapus redundansi/preprocessing data
df=df[df["Quantity"]>0]
df.shape
df.describe() #melihat detail dasar tentang dataset setelah menghapus
redundansi
df.head()
df.info() #mendapatkan semua informasi tentang kumpulan data
df["InvoiceDate"] = pd.to_datetime(df["InvoiceDate"]) #mengubah menjadi
datetime
df.info() #mendapatkan semua informasi tentang kumpulan data
```

Setelah menghapus redundansi, sekarang kita dapat melihat bahwa InvoiceDate telah telah dikonversi ke format tanggal dalam hari agar kita dapat menemukan frequency.

```
#Menghitung Monetary
df["Sale"]=df.Quantity*df.UnitPrice
#membuat kolom penjualan
df.head()
#Data dikelompokkan berdasarkan ID Pelanggan dan jumlah total penjualan
memberikan uang
monetry = df.groupby("CustomerID").Sale.sum()
monetry = monetry.reset_index()
monetry.head()

#Menghitung Frequency
frequency = df.groupby("CustomerID").InvoiceNo.count() #Untuk menghitung
frekuensi disini kita juga menggunakan groupby function yang
mengelompokkan CustomerId beserta jumlah invoice no. kita bisa mendapatkan
frekuensi kita.
frequency = frequency.reset_index() #mengatur ulang indeks kita, frekuensi
kita memiliki multiindex jadi kita menghapus
frequency
frequency.head()

#Menghitung Recency
LastDate=max(df.InvoiceDate) #menghitung tanggal terakhir InvoiceDate
LastDate
```

```
LastDate = LastDate + pd.DateOffset(days = 1)
LastDate
#Diff adalah perbedaan antara Lastdate dan InvoiceData
df["Diff"] = LastDate - df.InvoiceDate

#data dikelompokkan berdasarkan ID Pelanggan dan kemudian Perbedaan
minimum menghasilkan Recency
recency = df.groupby("CustomerID").Diff.min()
recency = recency.reset_index()
recency.head()
```

RMF adalah data utama yang sedang dikerjakan sampai sekarang kita melakukan preprocessing data

```
#Menyatukan semua kerangka RFM data menjadi satu df
rmf = monetry.merge(frequency, on = "CustomerID")
rmf = rmf.merge(recency, on = "CustomerID")
rmf.columns = ["CustomerID", "Monetry", "Frequency", "Recency"]
rmf
RMF1 = rmf.drop("CustomerID", axis = 1)
RMF1.Recency = RMF1.Recency.dt.days
RMF1
```

- Melakukan scaling data

```
from sklearn.preprocessing import StandardScaler
ssd = StandardScaler()
ssd.fit(RMF1)
ssd.transform(RMF1)
```

Scaling data menggunakan StandardScaler.

```
#Konversi Data di DataFrame
RMF1 = pd.DataFrame(RMF1, columns=["Monetry", "Frequency", "Recency"])
```

- Menerapkan algoritma K-Means

```
from sklearn.cluster import KMeans
```

Import library

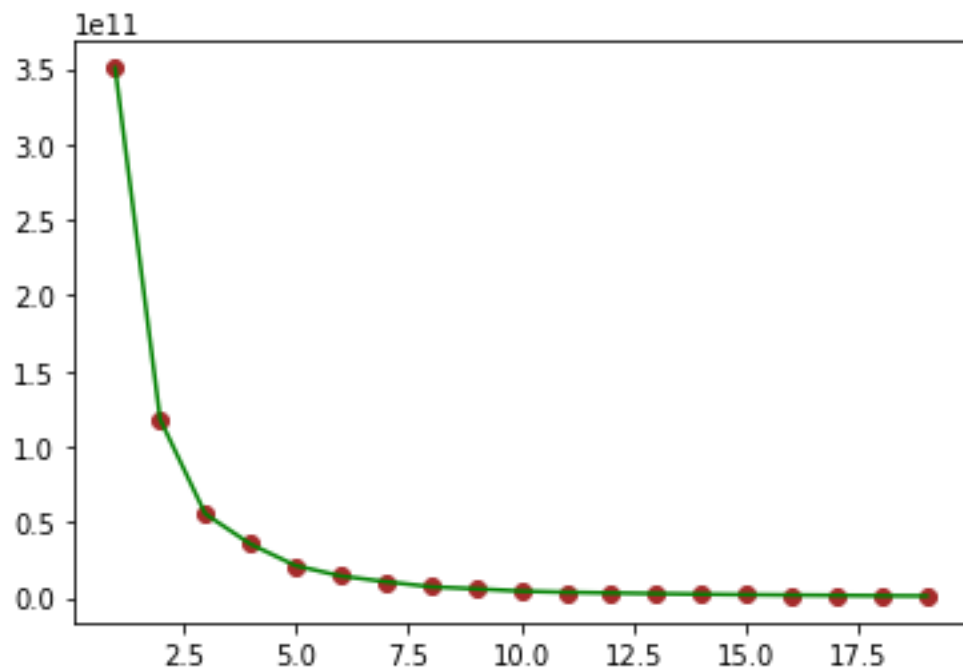
```
#Pelatihan data dan fitting
#Menemukan inersia untuk rentang nilai k
ssd = []
for k in range(1, 20):
```

```

km = KMeans(n_clusters=k)
km.fit(RMF1)
ssd.append(km.inertia_)

#Merencanakan nilai wrt 'k' inersia
plt.plot(np.arange(1,20), ssd, color = "green")
plt.scatter(np.arange(1,20), ssd, color = "brown")
plt.show()

```



```

#Melatih dengan KMeans dan memprediksi ClusterID.
model = KMeans(n_clusters=5)
ClusterID = model.fit_predict(RMF1)

RMF1["ClusterID"] = ClusterID
RMF1

km_cluster_sale = RMF1.groupby("ClusterID").Monetry.mean()
km_cluster_Recency = RMF1.groupby("ClusterID").Recency.mean()
km_cluster_Frequence = RMF1.groupby("ClusterID").Frequence.mean()

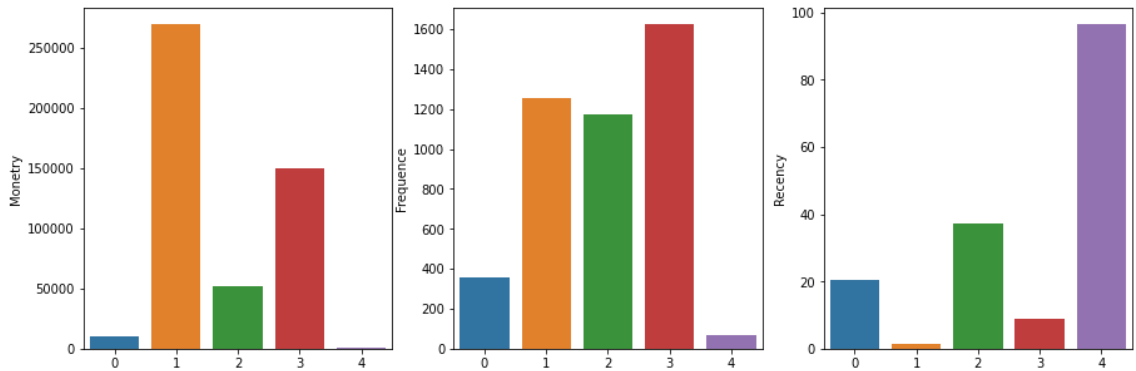
```

4. HASIL KLAUSTERISASI DATASET

Berikut adalah hasil dari proses klasterisasi data ritel online di UK dalam periode 8 bulan pada tahun 2010.

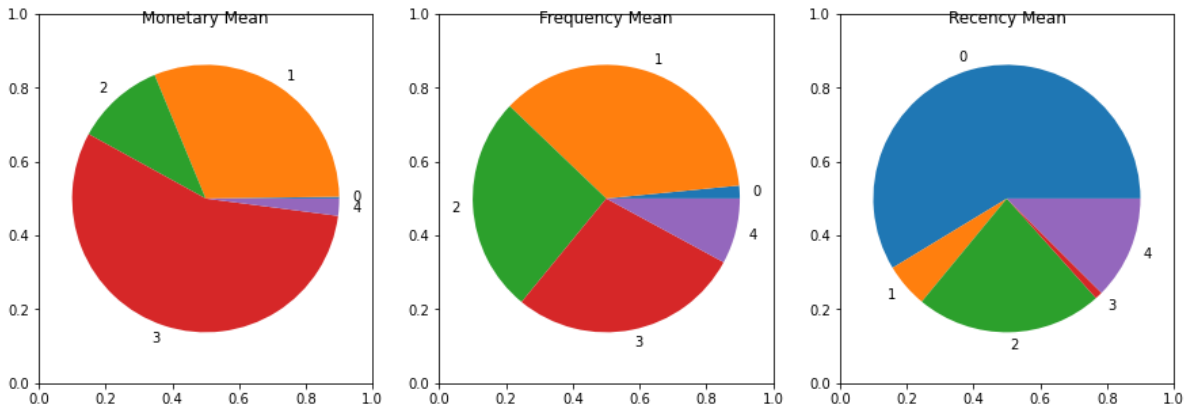
```
import seaborn as sns
```

```
fig, axs = plt.subplots(1,3, figsize = (15, 5))
sns.barplot(x = [0,1,2,3,4], y = km_cluster_sale , ax = axs[0])
sns.barplot(x = [0,1,2,3,4], y = km_cluster_Frequency , ax = axs[1])
sns.barplot(x = [0,1,2,3,4], y = km_cluster_Recency , ax = axs[2])
```



```
fig, axs = plt.subplots(1,3, figsize = (15, 5))
ax1 = fig.add_subplot(1, 3, 1)
plt.title("Monetary Mean")
ax1.pie(km_cluster_sale, labels = [0,1,2,3,4])
ax1 = fig.add_subplot(1, 3, 2)
plt.title("Frequency Mean")
ax1.pie(km_cluster_Frequency, labels = [0,1,2,3,4])
ax1 = fig.add_subplot(1, 3, 3)
plt.title("Recency Mean")
ax1.pie(km_cluster_Recency, labels = [0,1,2,3,4])

plt.axis("off")
plt.show()
```



5. HASIL/KESIMPULAN

Kesimpulan dari eksperimen ini adalah dari diagram lingkaran di atas kita dapat dengan mudah memahami 5 kelompok berdasarkan rata-rata Recency, rata-rata Frequency dan rata-rata Monetary, sebagai berikut

Kelompok 3 adalah kelompok pelanggan yang menghabiskan jumlah uang maksimum dan juga memiliki frequency yang baik dan tingkat resensi yang rendah. Kelompok 1 adalah pelanggan yang tingkat frequency maksimum, nilai monetary baik dan tingkat recency juga cukup baik, sedangkan Kelompok 0 adalah kelompok pelanggan yang memiliki tingkat recency yang maksimum artinya mereka belum pernah membeli apapun sebelumnya.

DAFTAR PUSTAKA

- AMIK BSI Yogyakarta, E. M. (2018). Komparasi Metode Clustering K-Means Dan K-Medoids Dengan Model Fuzzy Rfm Untuk Pengelompokan Pelanggan. *Evolusi : Jurnal Sains Dan Manajemen*, 6(2). <https://doi.org/10.31294/evolusi.v6i2.4600>
- Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1), 60–64. <https://doi.org/10.18196/st.211211>
- Darmi, Y. D., & Setiawan, A. (2017). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*, 12(2), 148–157. <https://doi.org/10.37676/jmi.v12i2.418>
- Grobogan, K. (2019). *Pertanian Grobogan Dalam Angka 2019*. 2019.
- Hadi, F., Octari Rahmadia, D., Hadi Nugraha, F., Putri Bulan, N., Mustakin, & Monalisa, S. (2017). Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru). *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 15(1), 69–76. <http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/4575>
- Hadi, S. W., Julianto, M. F., Rahmatullah, S., & Gata, W. (2020). Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means. *Bianglala Informatika*, 8(2), 86–90. <https://doi.org/10.31294/bi.v8i2.8191>
- Hadi Yansah, A. N. M., Puspaningrum, E. Y., & Rizky, A. M. (2021). Implementasi Kombinasi Metode Rfm Dan Algoritma Kmeans Untuk Klasterisasi Penjualan Minuman Kemasan. *Prosiding Seminar Nasional Informatika Bela Negara*, 2, 122–126. <https://doi.org/10.33005/santika.v2i0.96>
- Komputer, F. I., & Nuswantoro, U. D. (n.d.). *Jurnal_14114*. 1–5.
- Missa, D., Achmadi, S., & Mahmudi, A. (2021). Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Data Penghasilan Orang Tua Siswa. *JATI*

(*Jurnal Mahasiswa Teknik Informatika*), 5(1), 125–133.
<https://doi.org/10.36040/jati.v5i1.3275>

Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, 3(1), 10–17.

Novita Sari, V., Yupianti, Y., & Maharani, D. (2018). Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan. *Jurteksi*, 4(2), 133–140. <https://doi.org/10.33330/jurteksi.v4i2.53>

Prastyawan, Z. G., Ridho Bagaskara, M., & Fitriati, D. (2018). Segmentasi Pelanggan Restoran Menggunakan Metode Clustering Simple K-Means (Studi Kasus Xyz). *Informatika Dan Keamanan Siber (SEINASI-KESI) Jakarta-Indonesia*, 198–203.

Ruly, R. D. F., Purbandini, & Wuryanto, E. (2017). Penerapan Clustering K-Means Pada Customer Segmentation Berbasis Recency Frequency Monetary (RFM) (Studi Kasus : PT . Sinar Kencana Intermoda Surabaya). *Seminar Nasional Matematika Dan Aplikasinya*, 418–427.

Sani, A. (2018). Penerapan metode k-means clustering pada perusahaan. *Jurnal Ilmiah Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri*, May, 1–7.

Savitri, A. D., Bachtiar, F. A., & Setiawan, N. Y. (2018). Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(9), 2957–2966.

Sulistiyowati, Ketherin, B. E., Arifiyanti, A. A., & Sodik, A. (2018). Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering. *Seminar Nasional Sains Dan Teknologi Terapan VI 2018*, 51–58.

Tetap, M. K. (2020). *Program studi teknik informatika - universitas pgri madiun* / 43. 43–48.

Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 1(1), 204–215. <https://doi.org/10.24002/konstelasi.v1i1.4293>

Wijayanto, S., & Fathoni, M, Y. (2021). Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means. *Jurnal JUPITER*, 13(2), 212–219.

(- AMIK BSI Yogyakarta, 2018; 3, n.d.; Asroni et al., 2018; Darmi & Setiawan, 2017; Grobogan, 2019; F. Hadi et al., 2017; S. W. Hadi et al., 2020; Hadi Yansah et al., 2021; Komputer & Nuswantoro, n.d.; Missa et al., 2021; Muningsih & Kiswati, 2015; Novita Sari et al., 2018; Prastyawan et al., 2018; Ruly et al., 2017; Sani, 2018; Savitri et al., 2018; Sulistiyowati et al., 2018; Tetap, 2020; Widiyanto & Witanti, 2021; Wijayanto & Fathoni, M, 2021)