```python
# Latihan 1
# import pandas
import pandas as pd
# import numpy
import numpy as np
# import Library SelectKBest
from sklearn.feature_selection import SelectKBest
# import Library chi kuadrat/squared
from sklearn.feature_selection import chi2

# load dataset
data = pd.read_csv('Iris.csv')
data
```

```
      Id  SepalLengthCm  ...  PetalWidthCm         Species
0      1            5.1  ...           0.2     Iris-setosa
1      2            4.9  ...           0.2     Iris-setosa
2      3            4.7  ...           0.2     Iris-setosa
3      4            4.6  ...           0.2     Iris-setosa
4      5            5.0  ...           0.2     Iris-setosa
..   ...            ...  ...           ...             ...
145  146            6.7  ...           2.3  Iris-virginica
146  147            6.3  ...           1.9  Iris-virginica
147  148            6.5  ...           2.0  Iris-virginica
148  149            6.2  ...           2.3  Iris-virginica
149  150            5.9  ...           1.8  Iris-virginica

[150 rows x 6 columns]
```

```python
# Latihan 2
# Menghilangkan kolom Id
df1 = data.drop('Id', 1)
# lalu tampilkan
df1
```

```
      SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
Species
0               5.1           3.5            1.4           0.2
Iris-setosa
1               4.9           3.0            1.4           0.2
Iris-setosa
2               4.7           3.2            1.3           0.2
Iris-setosa
3               4.6           3.1            1.5           0.2
Iris-setosa
4               5.0           3.6            1.4           0.2
Iris-setosa
..              ...           ...            ...           ...
...
145             6.7           3.0            5.2           2.3  Iris-
virginica
```

```
146              6.3            2.5            5.0            1.9  Iris-
virginica
147              6.5            3.0            5.2            2.0  Iris-
virginica
148              6.2            3.4            5.4            2.3  Iris-
virginica
149              5.9            3.0            5.1            1.8  Iris-
virginica

[150 rows x 5 columns]
```

```python
# Latihan 3
#independent columns --> SepalLengthCm, SepalWidthCm, PetalLengthCm,
PetalWidthCm
X = df1.iloc[:,0:4]
# target columns --> species
y = df1.iloc[:,-1]

# Latihan 4
#Apply SelectKBest class to extract

bestfeature = SelectKBest(score_func=chi2, k=4)
fit = bestfeature.fit(X,y)
dfscores = pd.DataFrame(fit.scores_)
dfcolums = pd.DataFrame(X.columns)

# Latihan 5
#gabungkan 2 dataframe tersebut untuk visualisasi yang lebih bagus

featureScores = pd.concat([dfcolums, dfscores],axis=1)
featureScores.columns = ['Field', 'Score']
print(featureScores.nlargest(10,'Score'))
```

```
          Field        Score
2   PetalLengthCm   116.169847
3    PetalWidthCm    67.244828
0   SepalLengthCm    10.817821
1    SepalWidthCm     3.594499
```

```python
# Latihan 6
data = pd.read_csv('Iris.csv')

# Menghilangkan kolom Id
df2 = data.drop('Id', 1)
# lalu tampilkan
df2
```

```
     SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
Species
0              5.1           3.5            1.4           0.2
Iris-setosa
```

1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
..	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

[150 rows x 5 columns]

```python
# Latihan 7
#independent columns --> SepalLengthCm, SepalWidthCm, PetalLengthCm,
PetalWidthCm
A = df2.iloc[:,0:4]
# target columns --> species
b = df2.iloc[:,-1]

# Latihan 8
# Import library ExtraTreesClassifier
from sklearn.ensemble import ExtraTreesClassifier
# Import library matplotlib
import matplotlib.pyplot as plt

# fit model ExtraTreesClassifier
model = ExtraTreesClassifier()
model.fit(A,b)

ExtraTreesClassifier()

# Latihan 9
print(model.feature_importances_)
feat_importance = pd.Series(model.feature_importances_,
index=A.columns)
feat_importance.nlargest(10).plot(kind='barh')
plt.show()

[0.1118297  0.06335808 0.39293668 0.43187554]
```
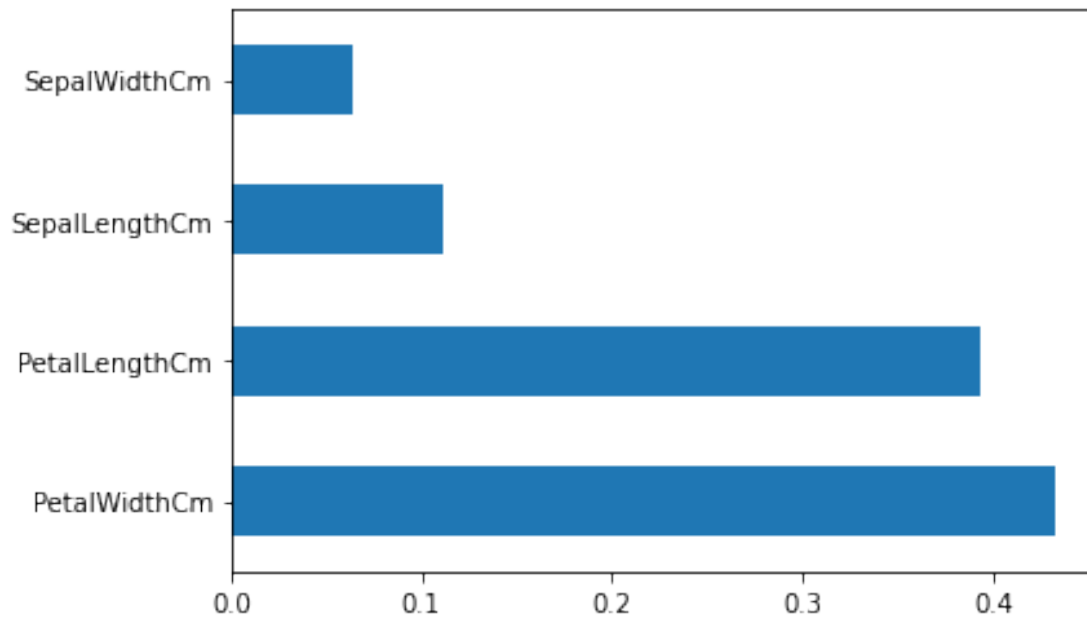
```
import pandas as pd
import numpy as np
import seaborn as sns

data = pd.read_csv('Iris.csv')
df3= data.iloc[:,1:]
df3
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
..	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-

```
virginica

[150 rows x 5 columns]
```

```python
# Latihan 10
#independent columns --> SepalLengthCm, SepalWidthCm, PetalLengthCm,
PetalWidthCm
K = df3.iloc[:,0:4]
# target columns --> species
j = df3.iloc[:,-1]

# mendapatkan korelasi di setiap fitur dalam dataset
corrmat = df3.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))

# plot heatmap
h =
sns.heatmap(df3[top_corr_features].corr(),annot=True,cmap="RdYlGn")
```
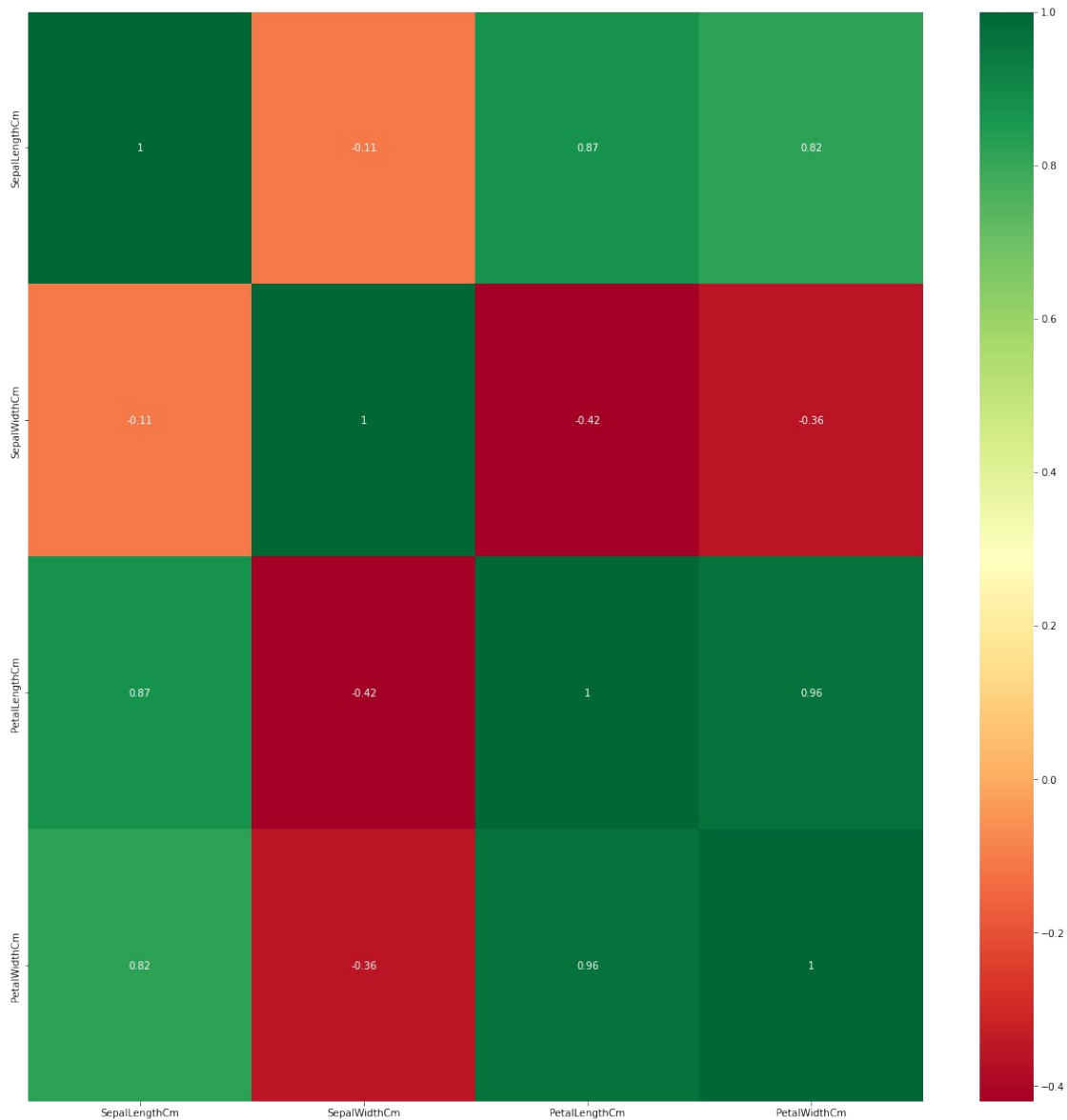
Kesimpulan :

1. lihat pada baris terakhir yaitu price range, korelasi antara PetalWidthCm dengan fitur lain dimana ada relasi kuat dengan variabel SepalLengthCm dan diikuti oleh var petalLengthCm.
2. sedangkan utk var PetalWidthCm berkorelasi lemah dengan SepalLengthCm