Nama Anggota Kelompok:

Beatrice Ruth Elizabeth Lumban Tobing / 3311901065
Juniar Indah Purnama Raya Situmorang / 3311901081
Elsa Maulidya / 3311901083

Penerapan Multiple Regression pada Dataset "Women Entrepreneurship and Labor Force"

Link GitHub: https://github.com/juniariprs/Tugas-Proyek-Data-Mining

Dataset:

Dataset yang digunakan yaitu diperoleh dari laporan Women Entrepreneurship Index dan Global Entrepreneurship Index yang diterbitkan pada tahun 2015. Riset ini terbatas pada negara-negara OECD di mana semua data untuk tahun 2015 tersedia pada waktu yang sama di database.

Proses Data Mining

Proses data mining yang dilakukan pada dataset:

1. Data Pre-processing
   Sebelum melakukan proses data mining, dataset terlebih dahulu melalui tahap pre-processing yaitu seleksi atribut secara manual. Atribut – atribut yang dieliminasi dari dataset tersebut adalah:
   Level of Development, European Union Membership, dan Currency.

2. Data Mining
   Proses data mining yang dilakukan adalah multiple linear regression untuk mengetahui nilai korelasi atribut-atribut yang ada terhadap atribut target Entrepreneurship Index dan menemukan formula regresi untuk prediksi nilai atribut Entrepreneurship Index.

Nama Anggota Kelompok:

Beatrice Ruth Elizabeth Lumban Tobing / 3311901065
Juniar Indah Purnama Raya Situmorang / 3311901081
Elsa Maulidya / 3311901083

Full Code:

```
#set and get location

setwd("G:/Tugas-Proyek_Data-Mining")

getwd()


#read dataset

dataset <- read.csv("entrepreneurship_after-preprocessing.csv", sep = ";")

head(dataset)


#see correlation

cor(dataset$Women.Entrepreneurship.Index, dataset$Entrepreneurship.Index)

cor(dataset$Inflation.Rate, dataset$Entrepreneurship.Index)

cor(dataset$Female.Labor.Force, dataset$Entrepreneurship.Index)


#make scatter plot between Women Entrepreneurship Index and Entrepreneurship Index

scatter.smooth(x=dataset$Women.Entrepreneurship.Index, y=dataset$Entrepreneurship.Index,
main="Entrepreneurship Index ~ Women Entrepreneurship Index")


#make model

linearMod <- lm(Entrepreneurship.Index ~

Women.Entrepreneurship.Index+Inflation.Rate+Female.Labor.Force, data = dataset)

summary(linearMod)

print(linearMod)
```

Nama Anggota Kelompok:

Beatrice Ruth Elizabeth Lumban Tobing / 3311901065
Juniar Indah Purnama Raya Situmorang / 3311901081
Elsa Maulidya / 3311901083

```r
#set and get location
setwd("G:/Tugas-Proyek_Data-Mining")
getwd()

#read dataset
dataset <- read.csv("entrepreneurship_after-preprocessing.csv", sep = ";")
head(dataset)

#see correlation
cor(dataset$Women.Entrepreneurship.Index, dataset$Entrepreneurship.Index)
cor(dataset$Inflation.Rate, dataset$Entrepreneurship.Index)
cor(dataset$Female.Labor.Force, dataset$Entrepreneurship.Index)

#make scatter plot between Women Entrepreneurship Index and Entrepreneurship Index
scatter.smooth(x=dataset$Women.Entrepreneurship.Index, y=dataset$Entrepreneurship.Index, main="Entrepreneurship Index ~ Women Entrepreneurship Index")

#make model
linearMod <- lm(Entrepreneurship.Index ~
Women.Entrepreneurship.Index+Inflation.Rate+Female.Labor.Force, data = dataset)
summary(linearMod)
print(linearMod)
```

Hasil Data Mining

Hasil data mining yang diperoleh nilai korelasi antar variabel sebagai berikut:

1. Nilai korelasi masing – masing variabel bebas terhadap variabel terikat Entrepreneurship Index, yaitu:

```
> #see correlation
> cor(dataset$Women.Entrepreneurship.Index, dataset$Entrepreneurship.Index)
[1] 0.9145797
> cor(dataset$Inflation.Rate, dataset$Entrepreneurship.Index)
[1] -0.3953699
> cor(dataset$Female.Labor.Force, dataset$Entrepreneurship.Index)
[1] 0.3341705
```

   - Women Entrepreneurship Index – Entrepreneurship Index = 0.9145797
   - Inflation Rate – Entrepreneurship Index = -0.3953699
   - Female Labor Force – Entrepreneurship Index = 0.3341705

2. Summary linear model

Nama Anggota Kelompok:

Beatrice Ruth Elizabeth Lumban Tobing / 3311901065
Juniar Indah Purnama Raya Situmorang / 3311901081
Elsa Maulidya / 3311901083

```
> summary(linearMod)

Call:
lm(formula = Entrepreneurship.Index ~ Women.Entrepreneurship.Index +
    Inflation.Rate + Female.Labor.Force, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-12.4450  -2.6065  -0.2787   3.1065  15.7011

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  0.43640    4.73417   0.092    0.927
Women.Entrepreneurship.Index 1.09988    0.08146  13.503   <2e-16 ***
Inflation.Rate               0.10138    0.19575   0.518    0.607
Female.Labor.Force          -0.10380    0.07537  -1.377    0.175
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.611 on 47 degrees of freedom
Multiple R-squared:  0.8433,    Adjusted R-squared:  0.8333
F-statistic: 84.34 on 3 and 47 DF,  p-value: < 2.2e-16
```

3. Print linear model

```
> print(linearMod)

Call:
lm(formula = Entrepreneurship.Index ~ Women.Entrepreneurship.Index +
    Inflation.Rate + Female.Labor.Force, data = dataset)

Coefficients:
                 (Intercept)  Women.Entrepreneurship.Index
                      0.4364                        1.0999
               Inflation.Rate           Female.Labor.Force
                      0.1014                       -0.1038
```

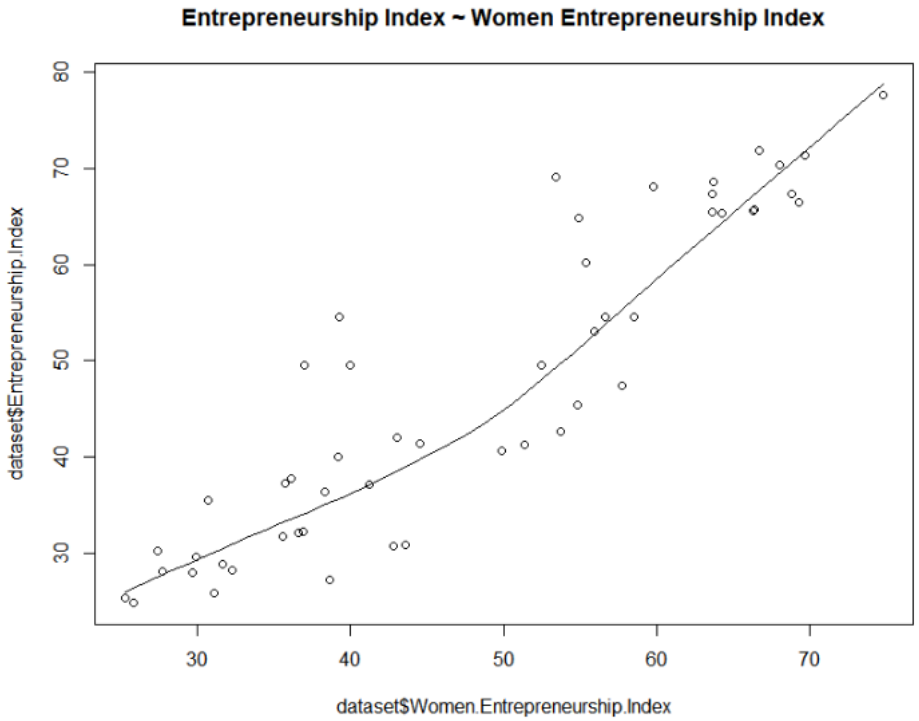
Hasil Analisa Data Mining

Berdasarkan nilai korelasi masing – masing variabel bebas terhadap variabel terikat, variabel yang memiliki nilai korelasi tertinggi adalah variabel Women Entrepreneurship Index yakni sebesar 0.9145797, di mana variabel ini mempunyai pengaruh lebih besar terhadap nilai variabel terikat Entrepreneurship Index.

Visualisasi Plot Scatter antara Women Entrepreneurship Index dan Entrepreneurship Index

Nama Anggota Kelompok:

Beatrice Ruth Elizabeth Lumban Tobing / 3311901065
Juniar Indah Purnama Raya Situmorang / 3311901081
Elsa Maulidya / 3311901083



Formula regresi sebagai model prediksi yang dihasilkan dari proses multiple linear regression ini (dengan y = Entrepreneurship Index), yaitu:

```
> print(linearMod)

Call:
lm(formula = Entrepreneurship.Index ~ Women.Entrepreneurship.Index +
    Inflation.Rate + Female.Labor.Force, data = dataset)

Coefficients:
                (Intercept)  Women.Entrepreneurship.Index
                     0.4364                        1.0999
              Inflation.Rate           Female.Labor.Force
                     0.1014                       -0.1038
```

y = 0.4364 + (1.0999 × Women Entrepreneurship Index) + (0.1014 × Inflation Rate) – (0.1038 × Female Labor Force)

Referensi

https://www.kaggle.com/babyoda/women-entrepreneurship-and-labor-force

https://www.machinelearningplus.com/machine-learning/complete-introduction-linear-regression-r