**Penerapan Multiple Linear 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.

1. 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.

**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)

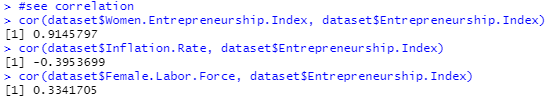
print(linearMod)

****

**Hasil Data Mining**

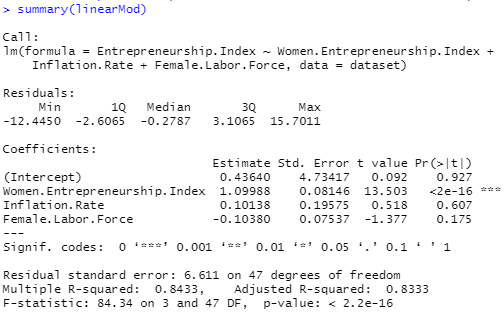
Hasil data mining yang diperoleh nilai korelasi antar variabel sebagai berikut:

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

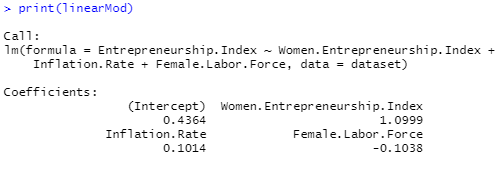


* Women Entrepreneurship Index – Entrepreneurship Index = 0.9145797
* Inflation Rate – Entrepreneurship Index = -0.3953699
* Female Labor Force – Entrepreneurship Index = 0.3341705

1. Summary linear model



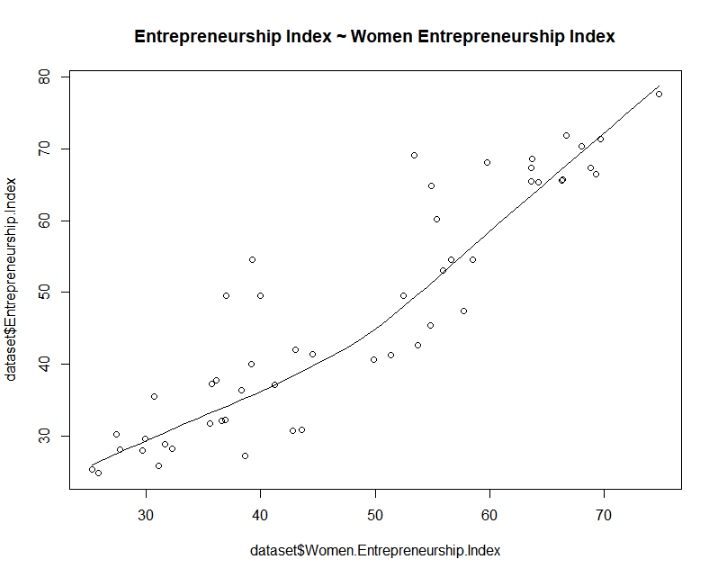
1. Print linear model



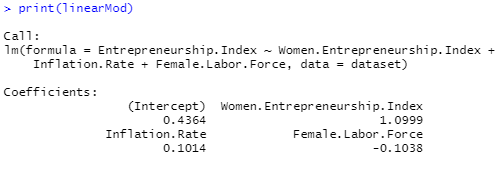
**Hasil Analisa Data Mining**

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



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



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