**Nama :Habibie Luthfi Mubarak**

**NIM :A11.2022.14104**

**Kelas : A11.4413**

**Mata Kuliah :Machine Learning**

**Judul Project : Prediksi Penyakit Diabetes Menggunakan Metode SVM**

**Ringkasan dan Permasalahan Project**

Diabetes merupakan sebuah penyakit kronis yang ditandani dengan tingginya kadar gula darah. Glukosa merupakan sumber energi utama bagi sel tubuh manusia. Akan tetapi, pada penderita diabetes, glukosa tersebut tidak dapat digunakan oleh tubuh. Berdasarkan data dari International Diabetes Federation (IDF) menunjukkan jumlah penderita diabetes di dunia pada tahun 2021 mencapai 537 juta jiwa. Menurut IDF, Indonesia menduduki peringkat kelima negara dengan jumlah diabetes terbanyak dengan 19,5 juta penderita di tahun 2021.

Berdasarkan data tersebut, diabetes merupakan salah satu penyakit yang cukup banyak diderita oleh masyarakat Indonesia. Selain memberikan sosialisasi terkait penyakit diabetes dan cara menanggulanginya, salah satu cara yang bisa dilakukan adalah melakukan prediksi dengan memanfaatkan machine learning.

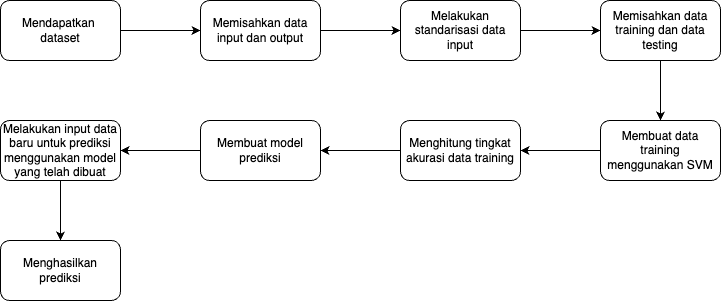
Untuk melakukan prediksi apakah seseorang terindikasi penyakit diabetes menggunakan machine learning salah satu cara yang dilakukan adalah menggunakan pendekatan supervised learning. Supervised learning merupakan sebuah metode pendekatan dalam machine learning yang dimana mesin akan dilatih untuk mempelajari sebuah data yang sudah memiliki label sebelumnya.

Dalam project ini, penulis menggunakan metode Support Vector Machine (SVM) untuk memprediksi penyakit diabetes berdasarkan data-data yang sudah memiliki label. Data-data yang akan digunakan seperti Pregnancies, Glucose, Blood Pressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, Age. Data-data tersebut akan digunakan sebagai input untuk selanjutnya menghasilkan sebuah output apakah seseorang terindikasi penyakit diabetes atau tidak.

**Tujuan Yang Akan Dicapai**

Tujuan dari project ini adalah untuk melakukan prediksi atau mengetahui apakah seseorang terindikasi penyakit diabetes atau tidak berdasarkan data-data yang dimasukkan.

**Model/Alur Penyelesaian**

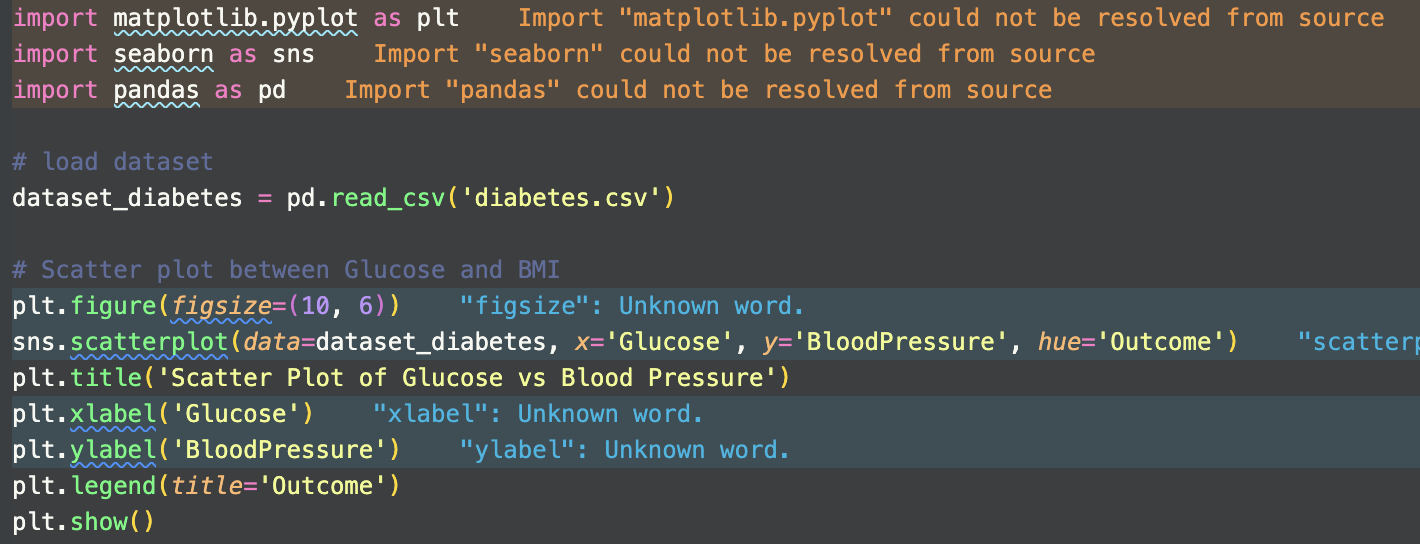


**Penjelasan Dataset, EDA dan Features Dataset**

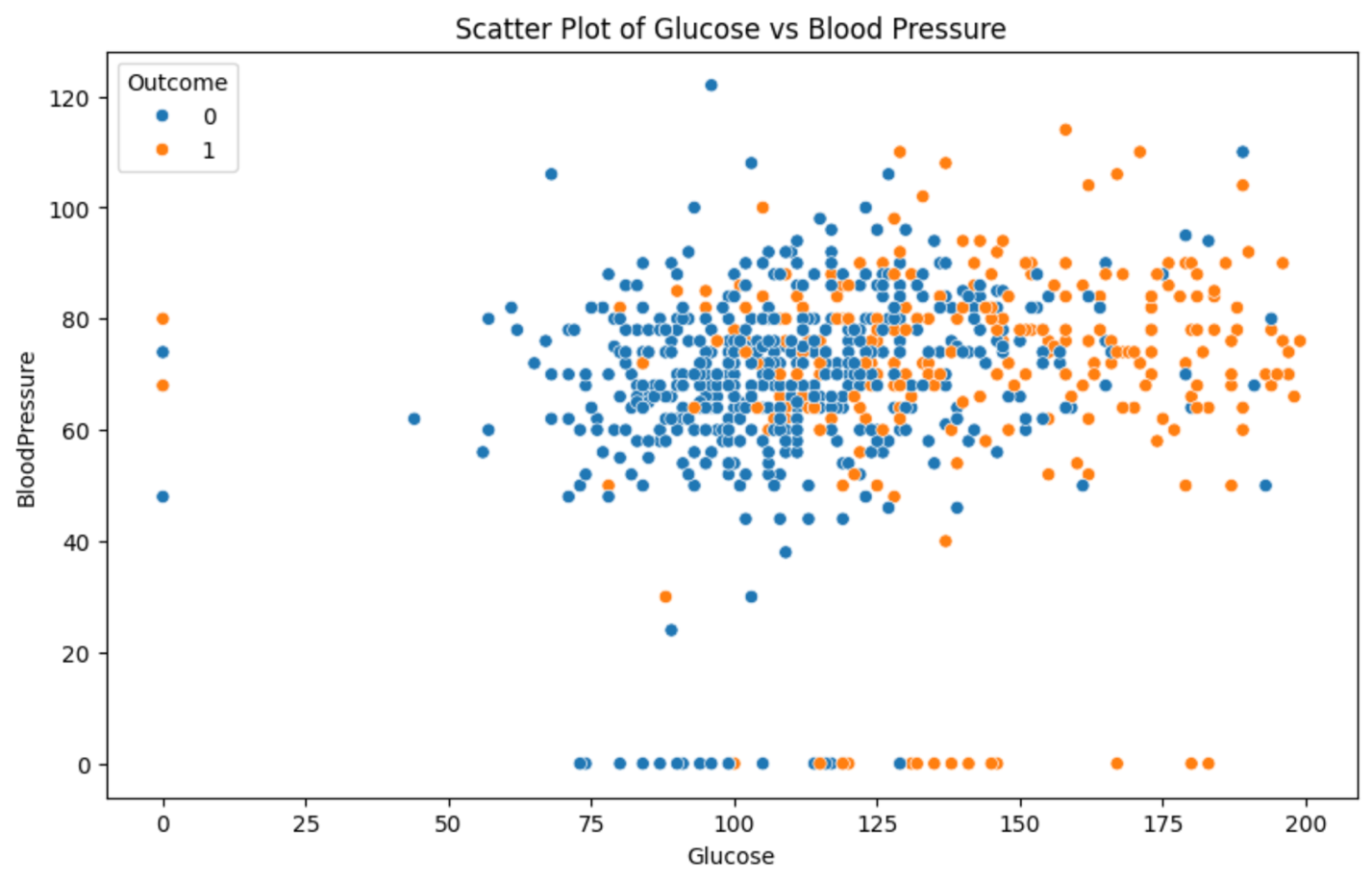
Dataset yang digunakan pada project ini didapatkan dari situs kaggle.com dengan judul “Pima Indians Diabetes Database” (<https://www.kaggle.com/datasets/mragpavank/diabetes/data>). Memiliki 9 kolom dengan 8 kolom input (Pregnancies, Glucose, Blood Pressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, Age) dan 1 kolom output (Outcome) serta memiliki 768 baris data.

Sebelum melakukan pengolahan data menggunakan SVM, dataset terlebih dahulu dilakukan analisis awal atau yang disebut dengan Exploratory Data Analysis (EDA). Disini, penulis melakukan analisis menggunakan teknik analisis bivariat untuk mengetahui hubungan dua variabel. Dua variabel yang digunakan adalah Glucose dan Blood Pressure dan divisualisasikan menggunakan scatter plot.

Source Code :



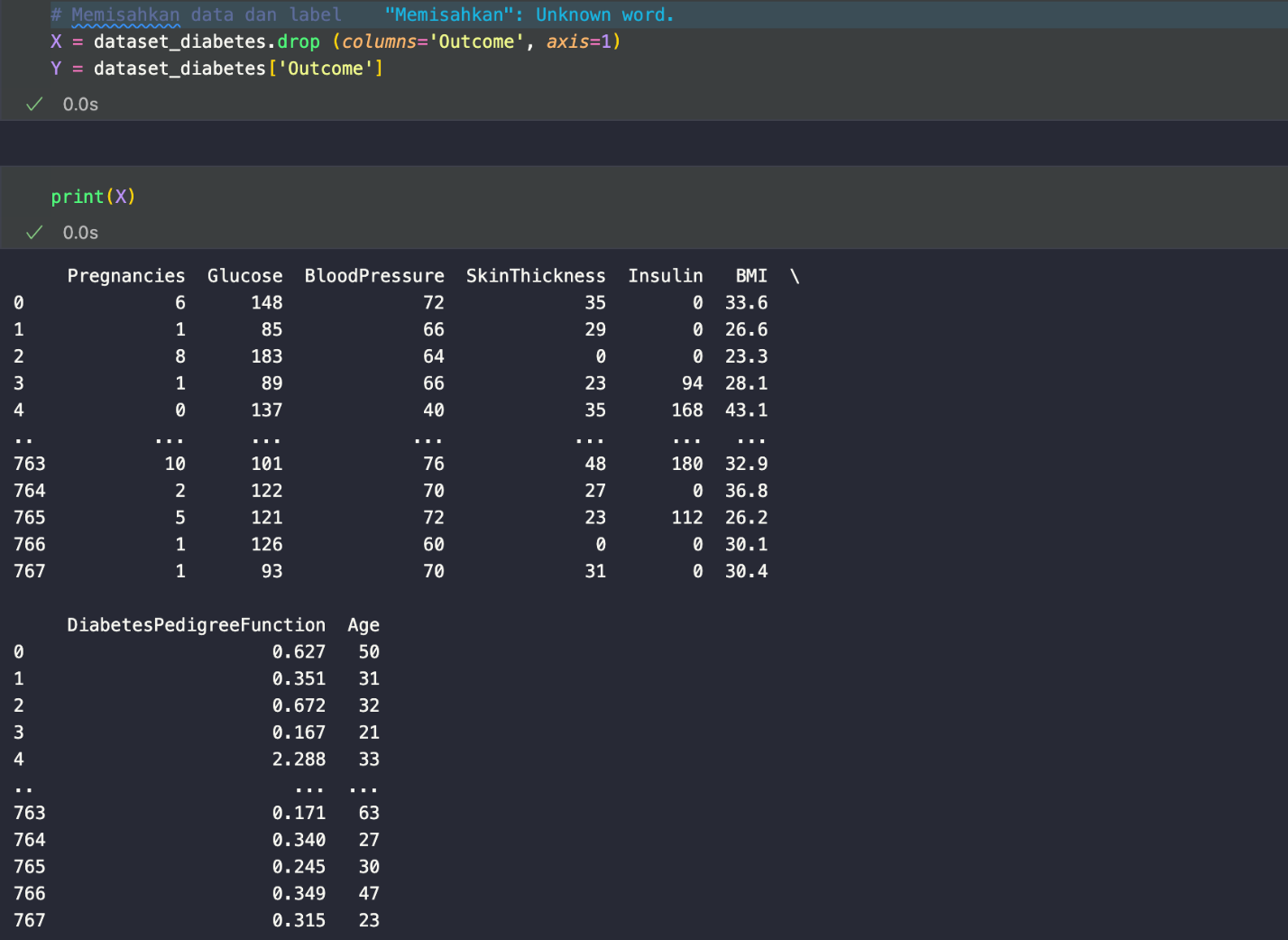
Hasil Scatter Plot



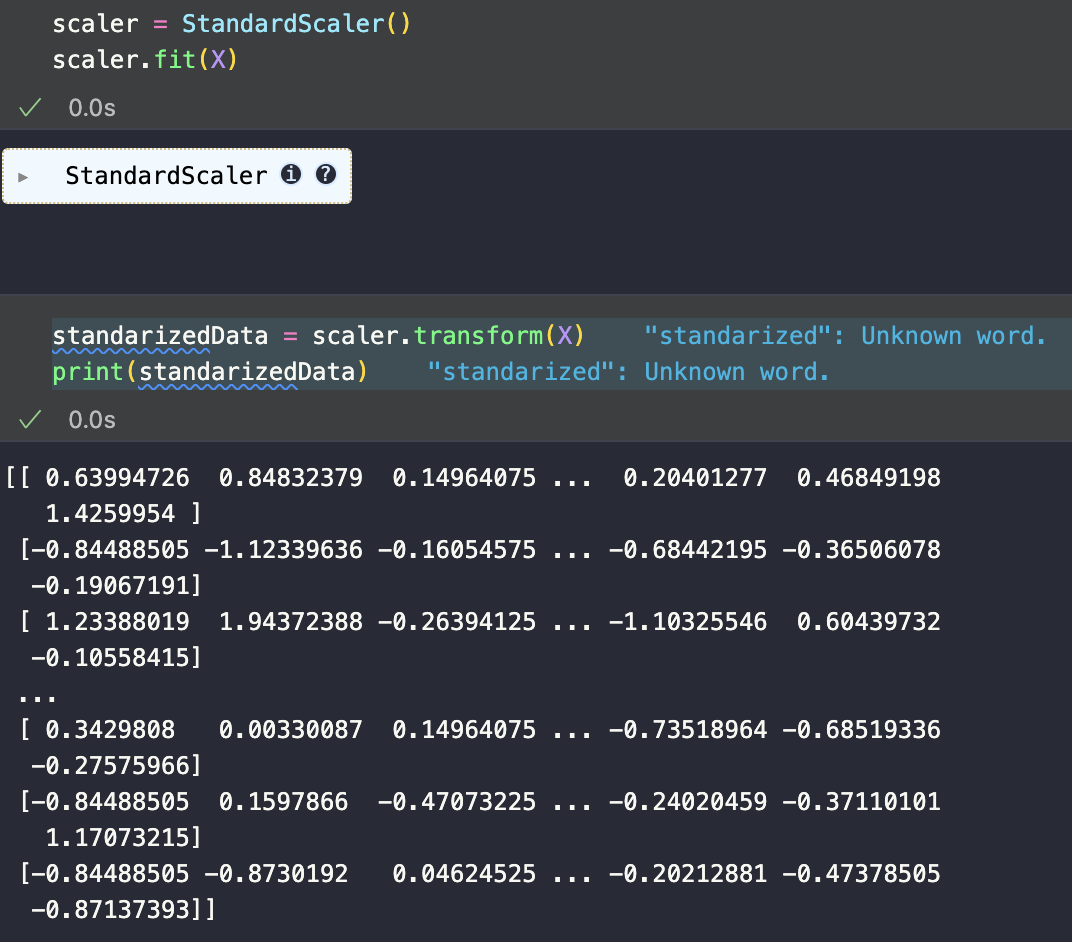
Setelah melakukan Exploratory Dataset Analysis, langkah selanjutnya adalah melakukan Features Dataset. Langkah ini merupakan sebuah langkah dalam menentukan variabel input dan output serta melakukan transformasi atau standarisasi data dengan tujuan data dapat diolah dan menghasilkan model untuk digunakan dalam melakukan prediksi.

Source code

1. Memisahkan data input (label) dengan output



1. Melakukan standarisasi data pada data input

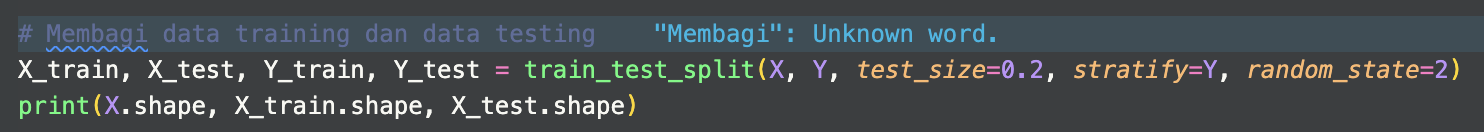


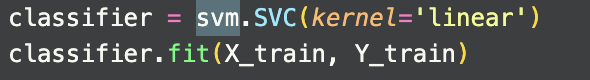
**Proses Learning/Modelling**

Setelah dataset sudah siap, langkah selanjutnya adalah proses learning dan modelling. Disini dataset akan dibagi menjadi 2 yaitu data training dan data testing. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk melakukan pengujian terhadap data training.

Dalam project ini, dataset dibagi menjadi data training dan data testing dengan perbandingan 8:2 atau 80% untuk data training dan 20% untuk data testing. Setalah itu data training dimasukkan ke dalam metode SVM.

Source code

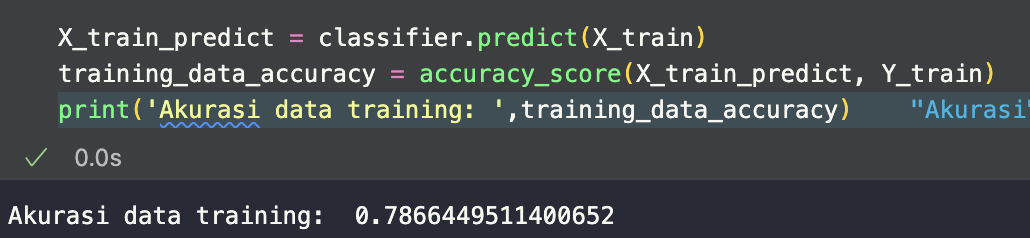




**Performa Model**

Langkah selanjutnya adalah melakukan perhitungan terhadap data training dari metode SVM. Dan membuat model dari data tersebut.

Source code





**Diskusi Hasil dan Kesimpulan**

Berdasarkan pengolahan dataset dan pengujian model, jika seseorang memiliki data sebagai berikut :

* Pregnancies = 6
* Glucose = 148
* Blood Pressure = 72
* Skin Thickness = 35
* Insulin = 0
* BMI = 33.6
* Diabetes Pedigree Function = 0.627
* Age = 50

Diprediksi positif menderita penyakit diabetes dengan tingkat akurasi sebanyak 78.67%