

Major Project on Supervised Learning using Logistic Regression, KNN, SVM & Decision Tree Classification Algorithm

About the Project

This project uses dataset, originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage. Have trained the Dataset by splitting 80% data for training and 20% for testing.

Tools

Python - Python is a general-purpose programming language, so it can be used for many things. Python is used for web development, AI, machine learning, operating systems, mobile application development, and video games.

Google Colab – Colab is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

Algorithm

Logistic Regression - In statistics, the logistic model is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc.

KNN - The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.

SVM - Support Vector Machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text

Decision Tree - Decision Trees are a type of Supervised Machine Learning where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves.

Outcome/Result

	Logistic Regression	KNN	SVM	Decision Tree
Accuracy Score	79.8%	81.8%	81.8%	70.6%

Documents

Have enclosed the program code file (Diabetes_Project.ipynb) for reference.

Report By
Prasanth N