Stroke Prediction Using Machine Learning

Implement by Artificial Intelligence and Machine Learning module

20-10-2025

In here I implement a model to predict stroke or not in human. I came up with a data set and first I cleaned that data set. After that I saw the data was unbalanced so for that I came up with another solution as smote data set. After the smote data set was balanced and then from that data set I learned my model. I chose my model as the Logistic Regression. For the reason is It's good for classification problems, Simple and easy to interpret, Efficient and fast and Supports regularization (L1, L2). I use parameters to train models. And I use three variants to tarin my model and finally I end up with best one.

Preprocessing:

I used a dataset: final_preproccessed_dataset.csv, which contained 8 Principal Components. The data was split, balanced with SMOTE (ratio 0.55), and scaled with StandardScaler.

Model Design & Tuning:

1. Variant 1 (Baseline): Default LogisticRegression (C=1.0, l2 penalty).
2. Variant 2 (Regularized): Stronger regularization (C=0.1).
3. Variant 3 (GridSearch): GridSearchCV with C [0.001-100], penalty ['l1', 'l2'], and solver ['liblinear', 'saga'].

Evaluation (F1-Score):

- Variant 1: 0.3099
- Variant 2: 0.3077
- Variant 3: 0.2644

- Best CV Params: 'C': 0.01, 'penalty': 'l2', 'solver': 'liblinear'
- Best Model: Basic Logistic Regression (F1-Score: 0.3099).