|  |  |
| --- | --- |
| **Stroke Prediction in Patients Report** |  |

*LAMYA ALARWAN*

**Introduction**

Stroke is a blood clot or bleeding in the brain which can cause long-term neurological damage and complications. stroke doesn't only cost medical financial burden and permanent disability but can eventually lead to death. Prevention and early detection of a stroke risk has many benefits on the patient and healthcare system.

This project predicts whether a person is at risk of having a stroke or not. They will get benefit from this project Patients and healthcare workers.

**Data**

The dataset source is Kaggle. It is containing 5110 records and12 features. Record contains patient information and if patient had a stroke or not. Target features (stroke)=1 if the patient had a stroke or 0 if not.

**Handling Missing Values**

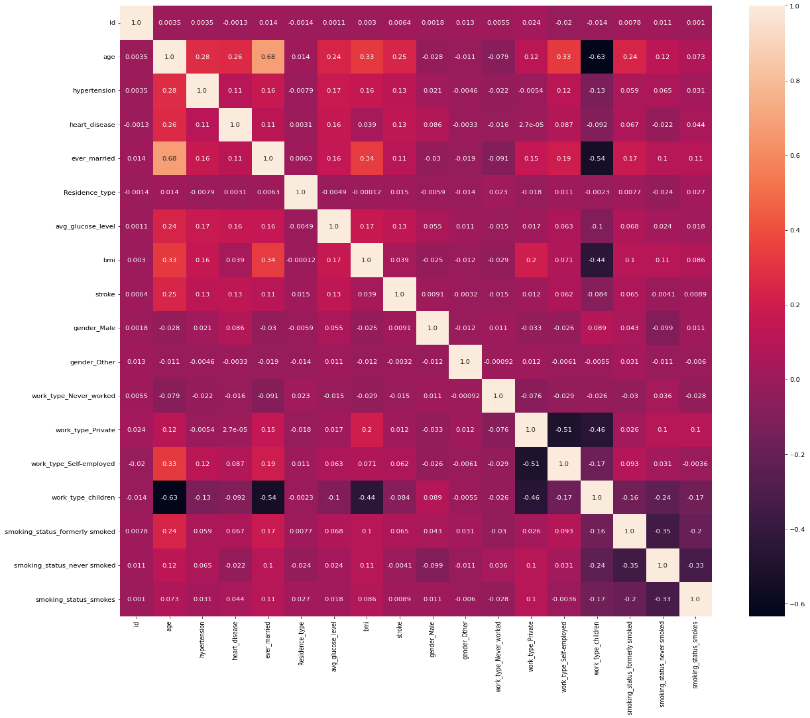
|  |  |  |
| --- | --- | --- |
| **صورة تحتوي على نص, استلام  تم إنشاء الوصف تلقائياً** |  | **صورة تحتوي على منضدة  تم إنشاء الوصف تلقائياً** |

**Handling Columns**

|  |  |  |
| --- | --- | --- |
|  | **صورة تحتوي على نص  تم إنشاء الوصف تلقائياً** | **صورة تحتوي على نص, صحف, لقطة شاشة, استلام  تم إنشاء الوصف تلقائياً** |

**Handling Class Imbalance**

|  |
| --- |
|  |
| **صورة تحتوي على نص  تم إنشاء الوصف تلقائياً** |
|  |

**Visualization & Correlation**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**Models & Performance**

After Handling Imbalance

Before Handling Imbalance

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ML Models** | **Accuracy** | **F1\_score** | **Confusion matrix** | **Accuracy** | **F1\_score** | **Confusion matrix** |
| LogisticRegression() | 94.4% | 0.0229 | [1447 3]  [82 1] | 67.97% | 0.2093 | [977 473] [18 65] |
| DecisionTreeClassifier(random\_state=1) | 91.51% | 0.1975 | [1387 63] [67 16] | 85.45% | 0.1115 | [1296 154] [69 14] |
| SVC() | 94.58% | 0.0 | [1450 0] [83 0] | 83.49% | 0.0664 | [1271 179] [74 9] |
| RandomForestClassifier() | 94.58% | 0.0 | [1449 1] [83 0] | 90.34% | 0.0864 | [1378 72] [76 7] |
| KNeighborsClassifier() | 94.58% | 0.0235 | [1449 1] [82 1] | 67.34% | 0.1363 | [1000 450] [44 39] |
| GaussianNB() | 92.82% | 0.2666 | [1403 47] [63 20] | 67.86% | 0.2314 | [960 490] [8 75] |
| LinearSVC() | 94.58% | 0.0 | [1450 0] [83 0] | 74.62% | 0.1525 | [1109 341] [48 35] |
| MLPClassifier() | 94.39% | 0.0 | [1447 3] [83 0] | 94.06% | 0.0808 | [1438 12] [79 4] |
| GradientBoostingClassifier() | 94.52% | 0.0454 | [1447 3] [81 2] | 86.30% | 0.1532 | [1304 146] [64 19] |
| CatBoostClassifier(verbose=0) | 94.25% | 0.0 | [1445 5] [83 0] | 89.49% | 0.08 | [1365 85] [76 7] |
| LGBMClassifier() | 93.93% | 0.0 | [1440 10] [83 0] | 90.15% | 0.0621 | [1395 55] [76 7] |
| XGBClassifier(eval\_metric='mlogloss') | 93.998% | 0.0612 | [1440 10] [83 0] | 90.34% | 0.0864 | [1378 72] [76 7] |