HOBBY PROJECT

① Imported csv file using pandas read-csv

② Checked shapes of both DataFrames. using shape function

③ Checked Null values using isnull function

④ Dropped some duplicated rows and columns which aren't required. using drop function

⑤ checked unique values from categorical coloums, & removed outliers values with deviate signifi. from rest. using value_count()

⇒ used Boxplot to check outliers in Numerical coloumns & removed outliers (if exits)

⑥ Visualised data using different graphs (Pie, Bar graph, Histogram, Scatter Plot) using Matplotlib & Seaborn library.

⑦ Used Label Encoder to convert categorical columns to numerical columns.

⑧ Selected Features & Targeted columns to numerical coloumns.

⑨ Created Training & testing Dataset using train_test_split.

⑩ Used classification model & checked accuracy, confusion matrix.

1. LR
2. Decision Tree classification
3. Random Forest "
4. SVM classification (support vector Machine)
5. KNN (K Near Neighbour)
6. XGBoost classification.

↳ ≈ 75% → Accuracy, precision score isn't good.

⑪ Normalized data using MinMax Scaler,

⑫ Again the data was Imbalanced, used imblearn. over_sampling to balance data

Again used all ML model, and got max accuracy 84.14% using XGBoost Model.

(21) used kfold. & stratified shuffle split for validation.

pickle →

Std. way of serializing objects in python.

↓

converting objects into stream of bytes to store the object or transmit it to memory or a file.

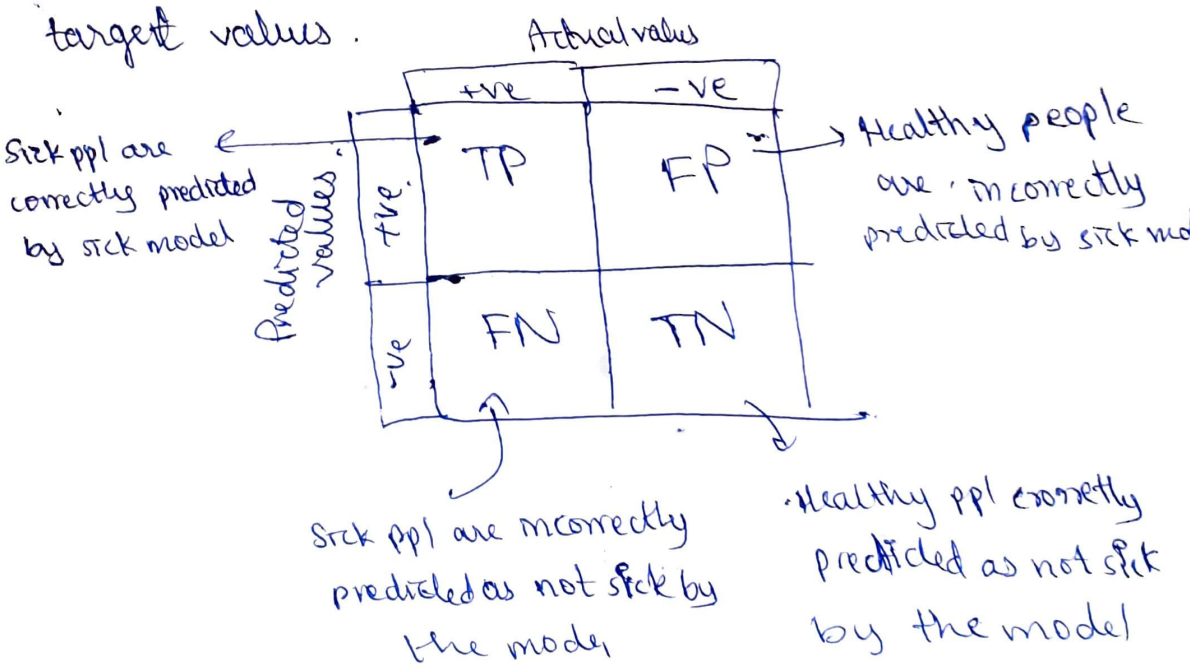
Reg.) Class

↓

· Regression algo → used to pred

Confusion Matrix

↳ It's N×N matrix used to evaluate the performance of classification model, when N is number of target values.



12 features

↓

30 features