BUILDING A STUDENT INTERVENTION SYSTEM

PRANAV N RAO
1BG16CS076
BNM INSTITUTE OF TECHNOLOGY
BANGALORE - 70

PYTHON

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

MACHINE LEARNING

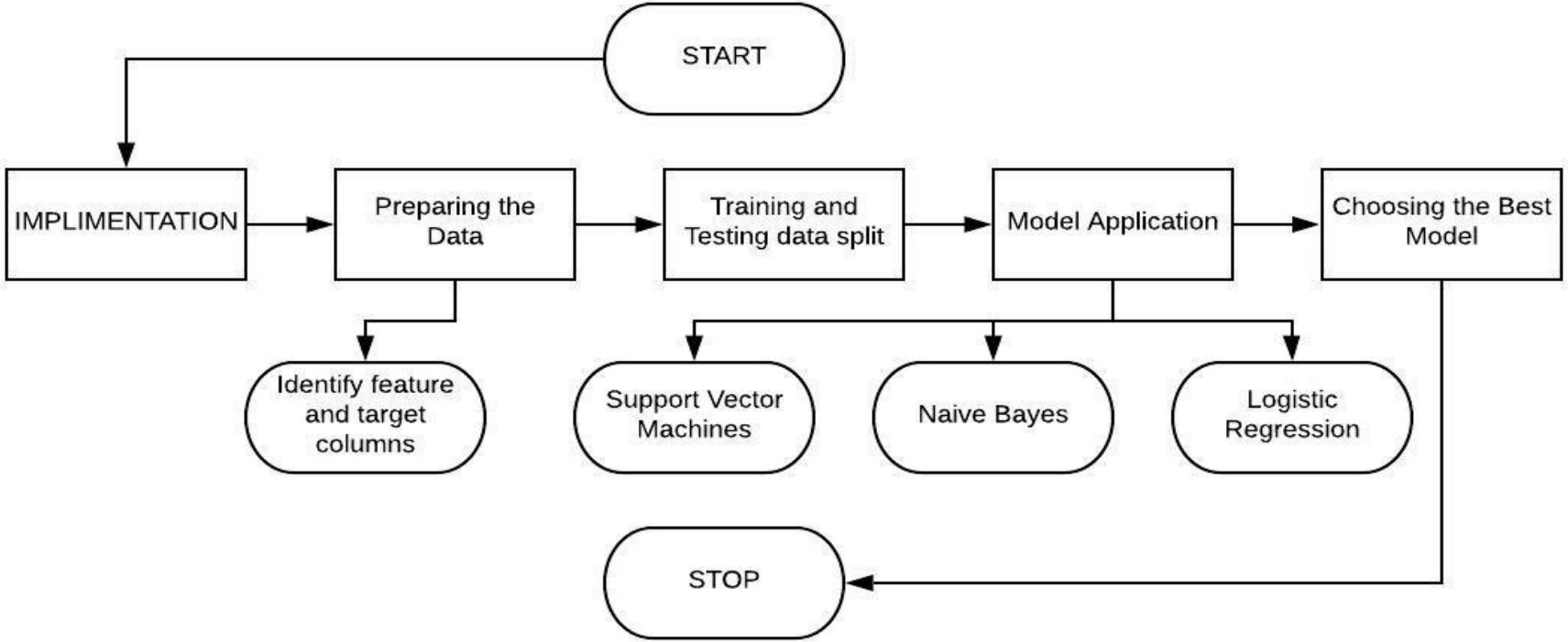
Machine learning (**ML**) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

INTRODUCTION

Here we are going to find the students who are going to graduate and who are not going to graduate by seeing their overall gradals and performances.

Our goal is to identify students who might need early intervention before they fail to graduate and to provide which type of supervised learning so that they can graduate successfully.

FLOW CHART


START
IMPLIMENTATION
Preparing the Data
Training and Testing data split
Model Application
Choosing the Best Model
Identify feature and target columns
Support Vector Machines
Naive Bayes
Logistic Regression
STOP

IMPLEMENTATION : DATA EXPLORATION

Let’s begin by investigating the dataset to determine how many students we have information on, and learn about the graduation rate among these students. You will need to compute following:

- The total number of students, n_students.
- The total number of features for each student, n_features.
- The number students who passed, n_passed.
- The number of students who failed, n_failed.
- The graduation rate of the class, grad_rate, in percent (%).

PREPARING THE DATA

Identify features and target columns

It is often the case that the data you obtain contains non-numeric features. This can be a problem, as most machine learning algorithms expect numeric data to perform computations with.

IMPLEMENTATION : TRAINING AND TESTING DATA SPLITS

So far, we have converted all categorical features into numeric values. For the next step, we split the data (both features and corresponding labels) into training and test sets. You will have to implement the following :

- Randomly shuffle and split data (X_all, y_all) into training and testing subsets.
- Use 300 training points (approx. 75%) and 95 testing points (approx. 25%).
- Set a random_state for the function(s) you use, if provide.
- Store the result in X_train, X_test, y_train, y_test

MODEL APPLICATION

There are three supervised learning models :

1. Naive Bayes
2. Logistic Regression
3. Support Vector Machines

CHOOSING THE BEST MODEL

In this final section, you will choose from the three supervised learning models the best model to use on the student data. You will then perform a grid search optimization for the model over the entire training set (X_train and y_train) by tuning at least one parameter to improve upon the untuned model's F1 score.

THANK YOU