

SportingBet

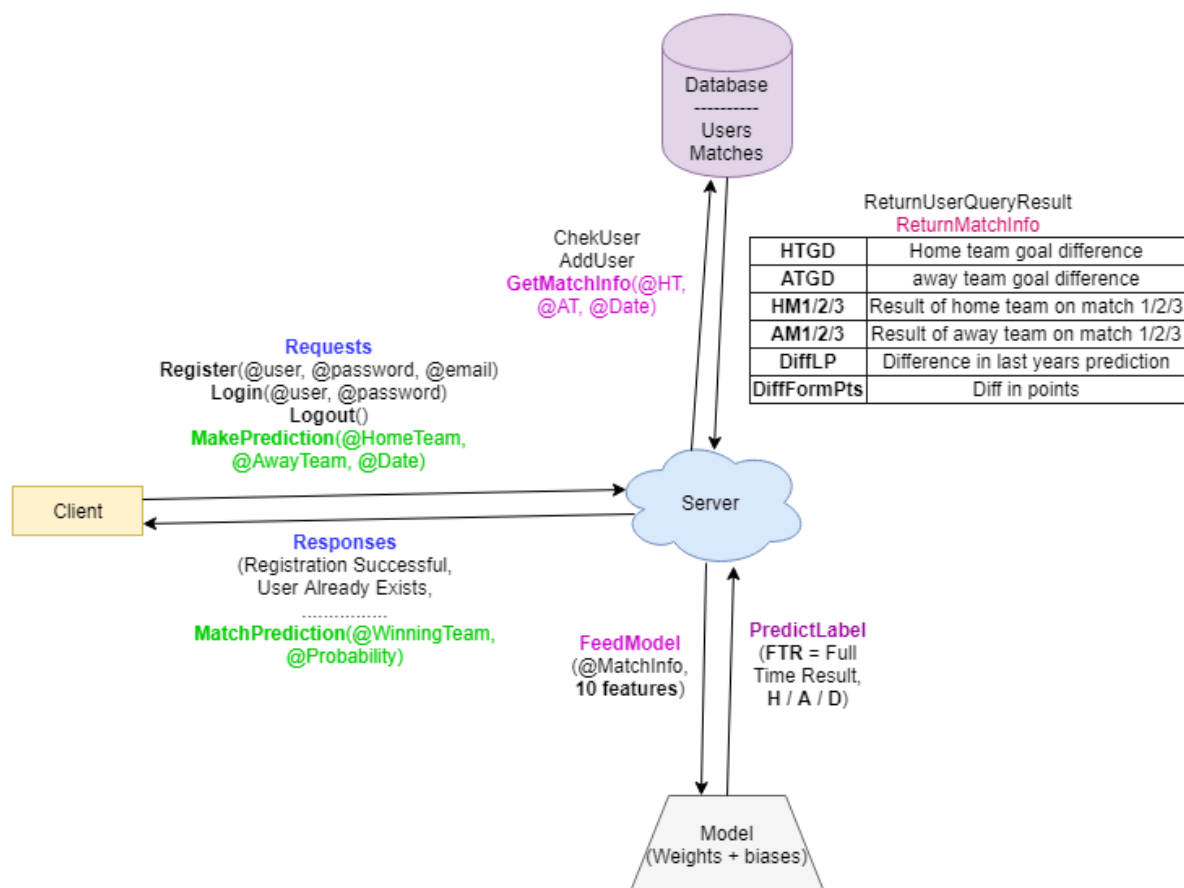
The application aims to predict the outcome of an upcoming football match using machine learning classification models based on previous matches from the same league.

Each prediction will be treated as a classification problem with three possible outcomes :

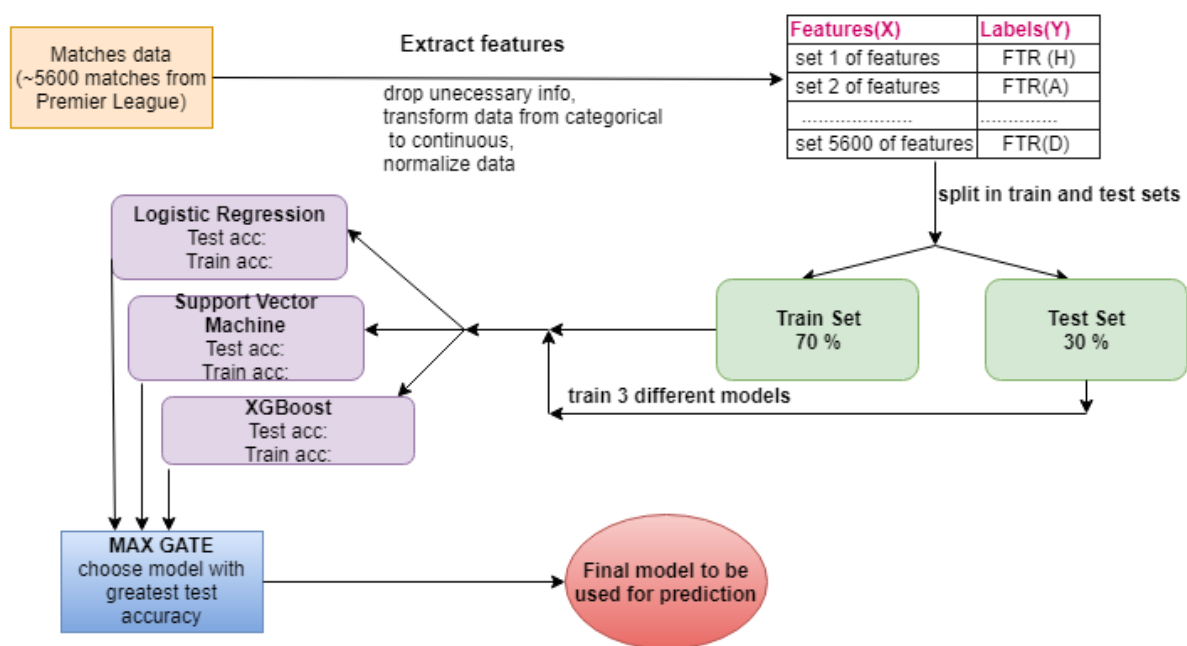
1. Home team wins
2. Away team wins
3. Draw

The user only needs to input the match for which the prediction is wanted and the system will automatically retrieve necessary information to feed the prediction model. The features used include Home/Away Team Goal Difference, Difference in Points in Current Scoreboard, results of the last matches for each team (important indicator for the current form of the team). The output of the system will be the predicted outcome of the match as well as the probability of this outcome.

Application architecture



Model choosing



Classification Models

1. Logistic Regression

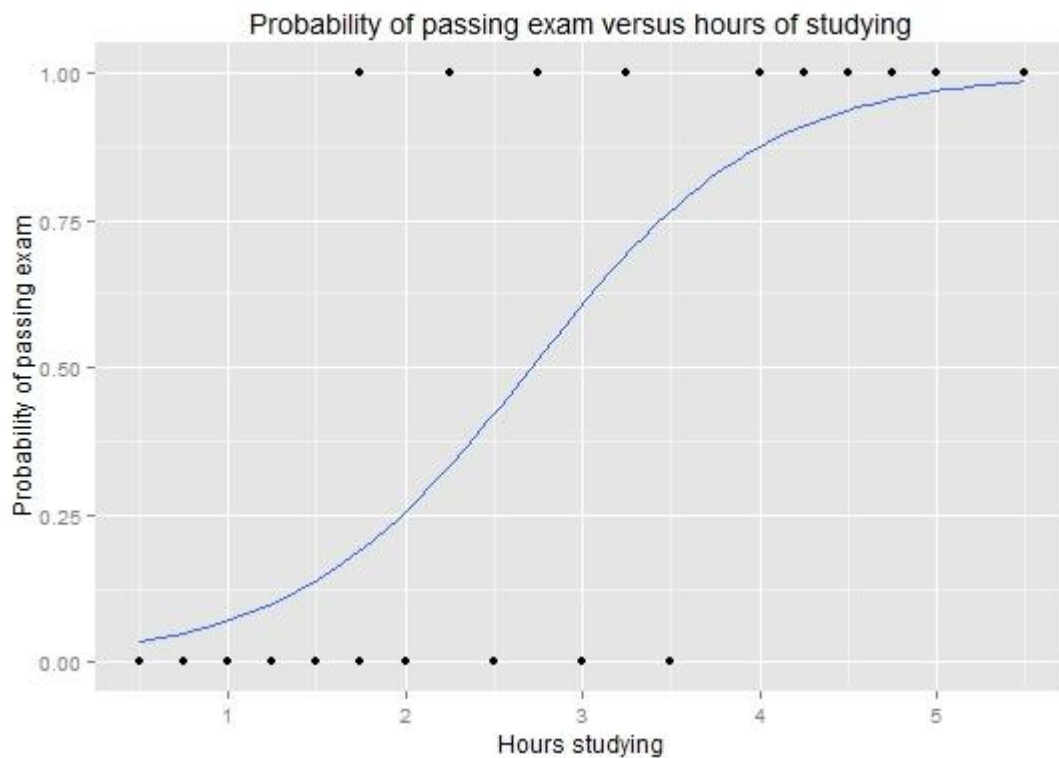
A regression model where the dependent variable (DV) is categorical.

Is used for prediction of the probability of occurrence of an event by fitting data to a logistic curve

Logistic regression makes use of several predictor variables that may be either numerical or categorical.

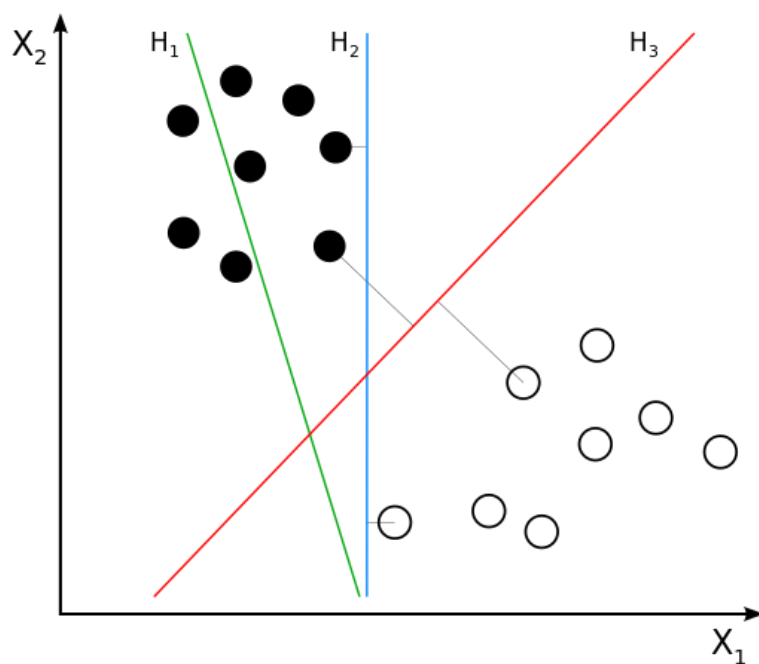
For example, the probability that a person has a heart attack within a specified time period might be predicted from knowledge of the person's age, sex and body mass index.

We can turn the logistic regression into a classification model by setting a threshold for the predicted probability : if probability < 0.5 then Class A else Class B.



2. Support Vector Machine

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

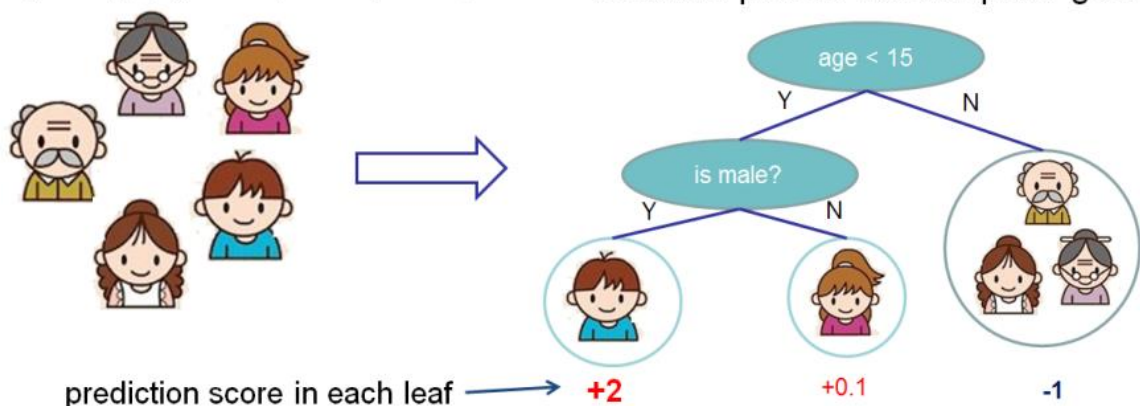


3. Tree ensembles.

The tree ensemble model is a set of classification and regression trees (CART). Here's a simple example of a CART that classifies whether someone will like computer games.

Input: age, gender, occupation, ...

Does the person like computer games



We classify the members of a family into different leaves, and assign them the score on the corresponding leaf. A CART is a bit different from decision trees, where the leaf only contains decision values. In CART, a real score is associated with each of the leaves, which gives us richer interpretations that go beyond classification. This also makes the unified optimization step easier, as we will see in a later part of this tutorial.

Usually, a single tree is not strong enough to be used in practice. What is actually used is the so-called tree ensemble model, which sums the prediction of multiple trees together.

