**Classification of unlabeled LoL match records with “Win/Loss”**

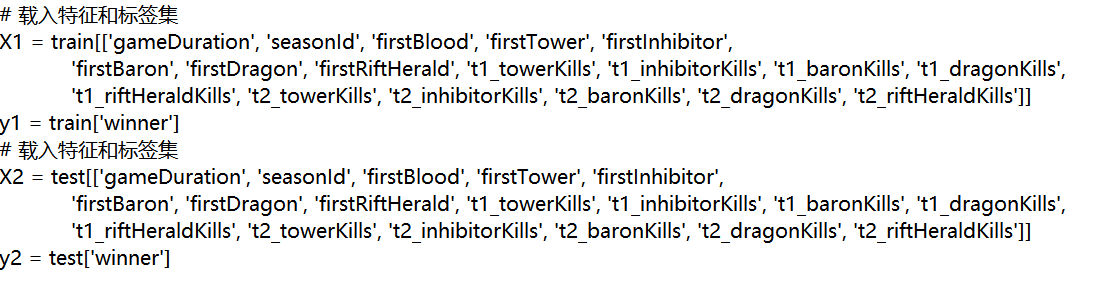
Dengyuyang, Intelligent Manufacturing

**Introduction**: The project is a binary problem. LOL is one of the most played eSports in the world nowadays. In this project, 3 Million match records are provided as training set and other 2 Million records are testing set. Each record comprises of all publicly available game statistics of a match. I will create some classifiers that take as inputs from any fields from the records and labels this records as a “1” or a “2”. Support Vector Machine, Logistic Regression, Decision Tree and K-Nearest Neighbors will be used to predict the results of matches.

**Algorithms**: I use four classifiers.

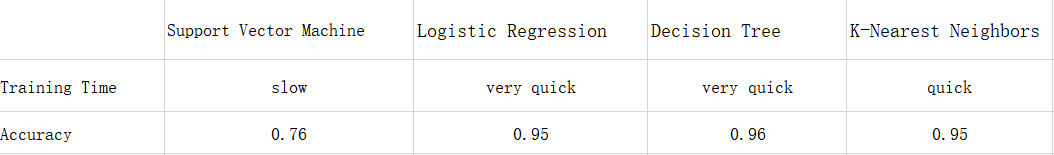
1. A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they’re able to categorize new text.
2. Logistic regression is a classification algorithm. It is used to predict a binary outcome based on a set of independent variables. A binary outcome is one where there are only two possible scenarios—either the event happens (1) or it does not happen (0). Independent variables are those variables or factors which may influence the outcome (or dependent variable). So Logistic regression is the correct type of analysis to use when you’re working with binary data. You know you’re dealing with binary data when the output or dependent variable is dichotomous or categorical in nature; in other words, if it fits into one of two categories (such as “yes” or “no”, “pass” or “fail”, and so on).
3. A decision tree is a type of flowchart that shows a clear pathway to a decision. In terms of data analysis, it is a type of algorithm that includes conditional ‘control’ statements to classify data. A decision tree starts at a single point (or ‘node’) which then branches (or ‘splits’) in two or more directions. Each branch offers different possible outcomes, incorporating a variety of decisions and chance events until a final outcome is achieved. When shown visually, their appearance is tree-like.
4. K-Nearest Neighbors is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

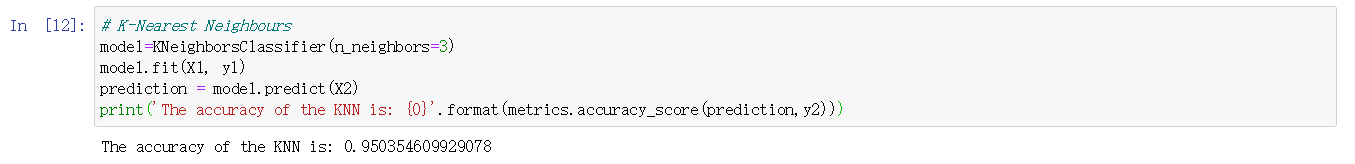
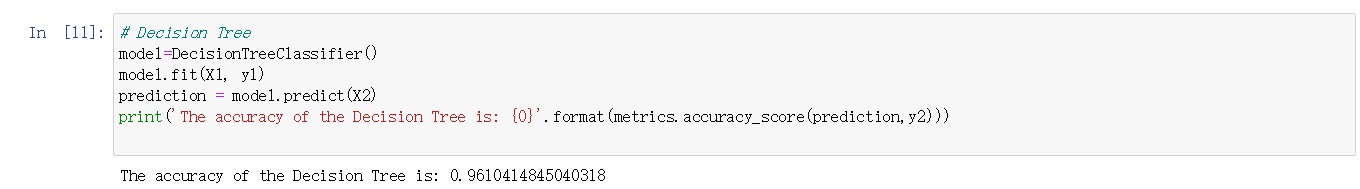
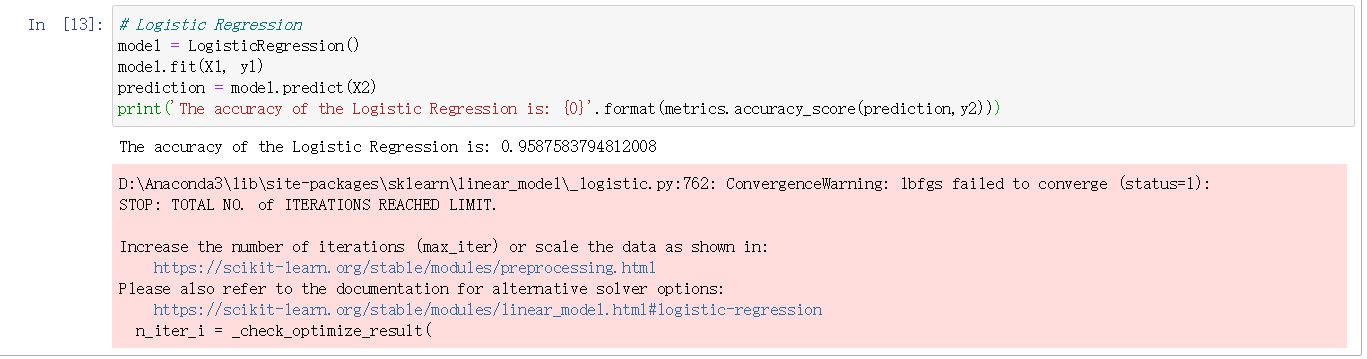
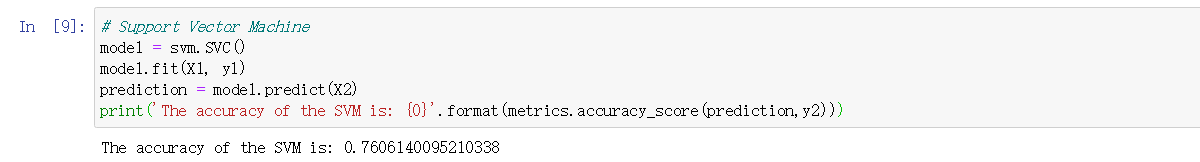
The parameters used are showed below.



**Requirements**: Have a knowledge of the meaning of attributes and search information about classifiers.

**Results**:





**Comparison and discussion**: I gained more information about classifiers and have a better grasp of classifiers. And I make a mistake because I don not understand what CreationTime means so firstly I use it as one attributes. So I think I should do more preparation before I write the code. And if more time is given, I want to analyse why other three classifiers are more accurate than SVM on this project and I want to know more about how to solve the problem if the result is a continuous variable.