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

**Chen Yunan**

**1.Introduction**

League of Legends (LoL) is one of the most played eSports in the world at the moment.In this game, there are many different factors that may change the game outcomes including the creation time, game duration and the number of towers, inhibitor each team has.

Though it has the uncertainty of the game outcomes, we can try to investigate what are the better strategy to win this game. In this project, you will have access to about 3 Million match records of solo gamers as training set. Each record comprises of all publicly available game statistics of a match played by some gamer, including an important field called "winner". Then predict the result of the other 2 million such records.

**2.Algorithm**

**A.Decision Tree(DT)**

Decision tree is based on the probability of known situations, by a decision tree to calculate the net present value of the probability of expectation is greater than or equal to zero, the project risk evaluation, and judge its feasibility decision analysis method, a graphic method is intuitive to use probability analysis because of this decision branch painted graphics is like a tree branches, therefore calls the decision tree decision tree is a tree structure, where each internal node said test on an attribute, each branch represents a test output, each leaf node represents a decision tree is a kind of very common categories of classification method.It is a kind of supervised learning, that is, given a set of samples, each sample has a set of attributes and a category, these categories are determined in advance, then a classifier can be obtained through learning, which can give the correct classification of the newly emerged objects.

**B.K Nearest Neighbor(KNN)**

Adjacent algorithm, or K nearest neighbor classification algorithm is one of the most simple method of data mining classification technology called K nearest neighbor, is the meaning of K nearest neighbor, said is each sample can use it to the closest K neighbor to represent the KNN algorithm is the core idea is that if a sample in the feature space K most of the adjacent sample belongs to a category, then the sample also belong to this category, and has the characteristics of this category sample this method in determining the classification decisions based on only the most adjacent categories to determine one or a few sample to sample points belong to the category of the KNN method is only related to a very small number of adjacent samples when making category decision. Since KNN method mainly relies on the limited adjacent samples rather than the method of discriminating class domains to determine the category belonging to it,KNN method is more suitable than other methods for the sample sets to be divided which have more overlapping or overlapping class domains.

**C.Support Vector Machine(SVM)**

SVM is referring to the support vector machine (SVM) is a common assessment method in the field of machine learning, is a supervised learning model, the SVM classification is usually used for pattern recognition and regression analysis method is through a nonlinear mapping p, the sample space is mapped to a high-dimensional as well as in the feature space of infinite dimension (the Hilbert space), made in original sample space nonlinear separable problem into linear separable problem in feature space.

**D.Multi-Layer Perceptron(MLP)**

MLP, namely the multi-layer perceptron, is a kind of tendency of the structure of the artificial neural network, the mapping of a set of input vectors to accommodate a set of output vector can be seen as a directed graph, is composed of multiple nodes layer, each layer of the entire company received the next layer in addition to the input nodes, each node is a nonlinear activation function of neurons (or processing unit) is known as the supervision of the back propagation algorithm learning method is often used to trained MLP. MLP is an extension of the perceptron overcomes the perceptron unable to realize the linear score according to identify faults.

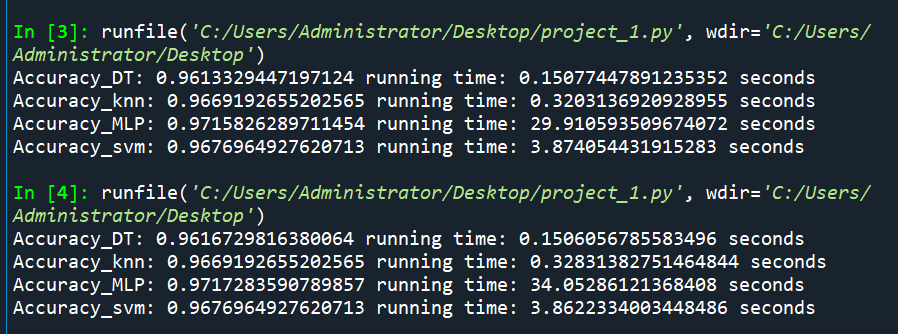
**3.Requirement**

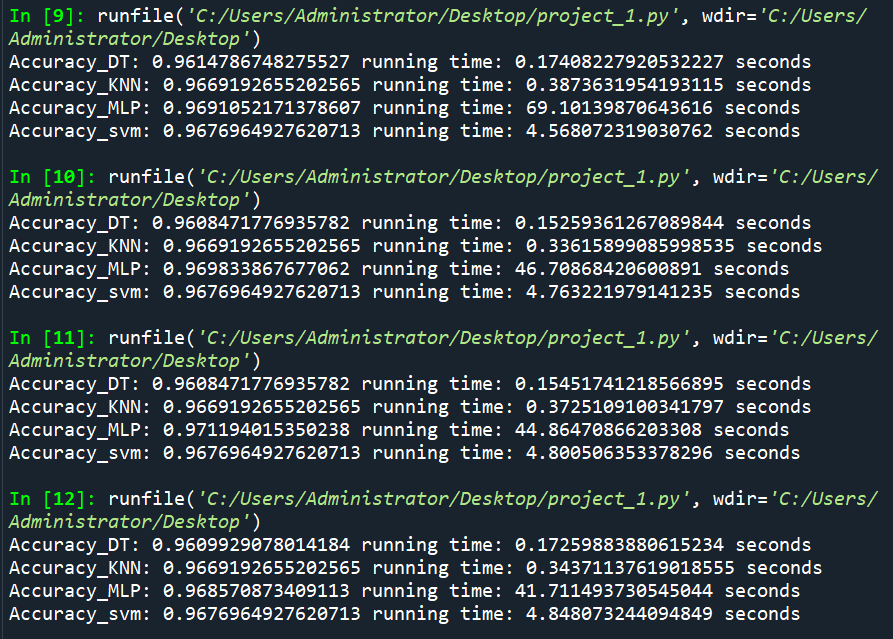
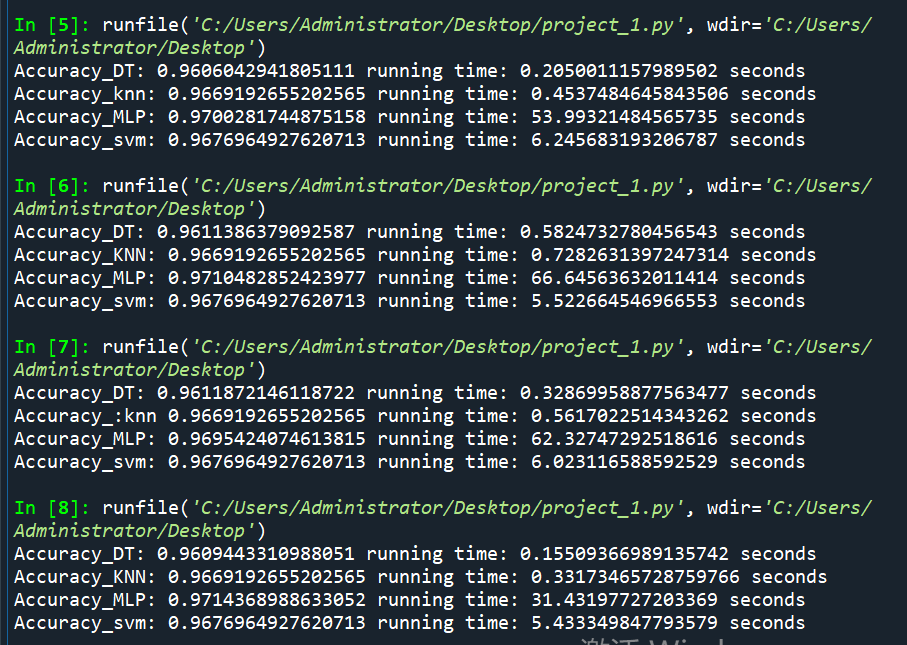
**Scikit-learn**

Scikit-learn (formerly SciKits. Learn, also known as Sklearn) is a free software machine learning library for the Python programming language. It features a variety of classification, regression, and clustering algorithms, including support vector machines, random forests, gradient lift, K-means, and DBSCAN, and is designed to work with Python numerical Science libraries NumPy and SciPy.

Scikit-learn is written primarily in Python, and numPY is widely used for high-performance linear algebra and array operations.In addition, Cython has been used to write some core algorithms to improve performance.The support vector machine is implemented by Cython wrapper around LIBSVM;Logistic regression and linear support vector machines are similarly packaged around LIBLINEAR.In this case, you may not be able to extend these methods using Python.

**4.Results**

****

****

**5.Comparison and discussion.**

The classification algorithm in supervised learning is an effective way to predict the category of unknown samples. The sample data used in this project comes from the game information of LOL(League of Legends). The purpose of the project is to predict the winner of the game based on the information of the game. Initially, the sample has 20 features and one label. In the process of data preprocessing, after correlation analysis, some of the features of the sample did not have much effect on label prediction, so these invalid features were deleted in the process of data preprocessing, and the sample left 16 attribute features and one label feature . Subsequently, four classifiers (DT, KNN,MLP,SVM) were trained using the training set, and then they become the relatively accurate classification model, and they can do a relatively accurate prediction on the test samples.

The next step in this project is to try real-time monitor the information of game, and give game suggestions to players. When a player is playing a game, the game information will be collected by this classifier in real time. This classifier predicts the game result based on the game information, and proposes game suggestions to both parties based on the prediction result and the current game information.