Using python to predict

the housing price

Team:陈文帅、于鹏、叶恺、罗绍博、王程灏、茅宇润

The important that we need to do such a predict:

Nowadays,the price of house in many cities is always high and the trendency is still going up . But many people need such a house to live but the price is increasing quickly so that the pressure for the people is going up.

In this background,we plan to do a program to predict the housing price.For this target,we try to analyse the reason of the increasing from four parts.

Some factors may influence the price:

(1)The region (e.g. the price is different between XuHui and MinHang)

(2)The distance to the nearest railway station

(3)The distance to the nearest school

(4)The distance to the nearest park(scenary)

The next step → find the solution

Two stages:

(1) stage 1: Choose the most suitable Algorithm.

(2)stage 2:Predict the housing price

The algorithm we prepared:

(1)Kneibors Regressor

(2)Linear Regressor

(3)Decision Tree Regressor

(4)Random Forest Regressor

The procedure:

(1)We collect 200 different data of housing price and input them in to a file.

↓

(2) Then, we prepared feature matrix X and label vector y, and use holdout validation to split X and y into two parts: training (80%) and testing (20%).

↓

(3)Running the program to compute the MSE for further comparing by every algorithm.

Algorithm	MSE
KNeighbors Regressor	5.891485
Linear Regressor	0.986493
DecisionTree Regressor	1.383725
RandomForest Regressor	0.655586

This chart shows what the program has got.

So we choose the last algorithm.

Stage 2

When we apply the algorithm we have chosen, as soon as the user input the three kinds of distance information into a file called “predict price.csv” .

And then users can run the program, they can learn about the prices of houses about to sell.

The problems we meet when we do the project:

(1)we can't distinguish the value of the region by the code .

(2)It is different that every railway station has it own value,and it is difficult to adjust the different data for different station in the code.

The problems we meet when we do the project:

(3)There are many kinds of school near the house(e.g the college or the primary school)

what we need to do is classify all the kinds of school and utilize it into our project .

(4)The last problem is the nearest park(scenary),some of the houses have many great parks in the vicinity but some of them don't have the park or the scenary in their vicinity, in this case, it may get a negetive figure or a impractical figure in the code

We think what we can improve our program......

(1)upgrade the process of inputting the statistics.

(2)expand the factors and try to find some other easier ways to collect statistics.

(3)......

THANKS!