IBDA Individual Project

Jack 1930026143

Abstract

Our group is tasked with helping banks market their wealth management products. In business data analysis, we first need to determine the type of our model. Since our ultimate goal is to determine who is more likely to buy our product and when the promotion is more likely to be successful, this is a classification problem. We can conduct business analysis based on known classification results, and even predict whether customers will buy or not.

Challenge and Solution

The part I am in charge of is modeling, which I think is a very important step in business analysis. We chose random forest as our classification model because it has higher error tolerance and accuracy than decision trees. However, due to the limited local CPU resources, we must reduce the dimensionality of the data in the face of a large number of features. After debugging the results, the memory overflow problem was solved. After obtaining the random forest model, we can know the importance of each feature which means that which feature can effect the reust most. We analyze according to the features with the highest importance, and we can draw some rules through probability formulas and visual analysis. But for the variables of PCA, because it is difficult to explain, we can use PCA as a reference indicator to filter out some people who are unlikely to buy and reduce the sales cost of banks. Here is the best solution for the bank:

Promotion time:

1. Contact customers every three days or twice a week
2. Peddle during the holidays

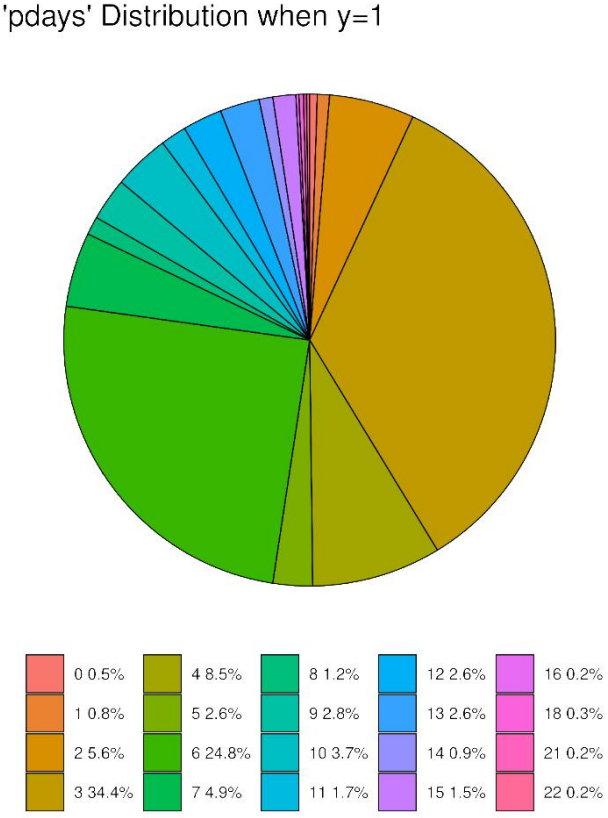
Target groups:

1. Peddle to middle-aged people
2. Administrators and entrepreneurs
3. People who have university degree
4. Married people
5. Non-explanatory factors

   The same principal component analysis is performed on the customer data, and among the obtained 6 PC features, the 3 most important ones are selected for auxiliary judgment. If customers' $PC1 > 0.4195, PC4 > 0.7921$ and $PC5 > 0.6613$, then we can filter it out.

Result interpretation Example

1. Since both month and day_of_week refer to the same information, the date information of the last contact, we consider them for completeness. Customers who purchase this product should contact them again on the same day of the following year. Agencies should pay more attention to the consumer's religion, hometown, and even country (if any). Trading is more likely when certain holidays are celebrated. For example, some undergraduate students may also get lucky money when it comes to Chinese New Year. However, some of them may just have all their money in the bank and have no idea how to properly manage their money. It's also an opportunity for bankers to provide these college students with some solid and simple financial management plans.
2. We can easily see that customers who bought a product three days ago and seven days ago play such a big role in this pie chart. In addition, for the same set of customers, it may be difficult to sell to them because they may have forgotten who you are or the product details. Therefore, people working for banks are strongly advised to contact their clients again three to seven days after the last conversation. (Visual method)



3. For the type of target group, we use the Bayesian formula to calculate the conditional probability of each type of purchasers to get the most likely type.