

# Risk Analysis in Medical Cosmetic Surgeries Loans Based on Decision Tree

Tianyu Wu

International Curriculum Center, the High School Affiliated to Renmin University, Beijing, China

**Keywords:** Medical Cosmetic Surgeries, Surgeries Loans, Risk, Decision Tree

**Abstract:** This research paper uses medical cosmetic surgery loan as an example, utilizing 1200 people's data from a typical medical cosmetic surgery loan company from May,2016 to October,2018. It contrasts current data mining model-Decision Tree with traditional qualitative methods (such as whether a person has mortgage before borrowing money, how long a person has lived in one city, and whether the person borrows money from multiple companies at the same time) which companies normally use in determining if a person has higher risk of not paying back money on time and identifies the key factors that can influence a person's punctuality in giving back money. The main conclusions of my research are that firstly, decision tree is doing a better job in judging whether people are going to pay back their money on time. Secondly, the default rate in big cities is bigger than that in small cities. What's more, there is threshold existing within the scope of people's educational level. People have good payback rate beyond that level and have low payback rate below that level.

## 1. Introduction

As the society develops, people do not only care about bread and butter, but they are also pursuing higher living qualities. Having medical cosmetic surgeries to improve their outlook has already becoming a common way for people to achieve a higher living quality. According to Townley (2019), up to 2017, there were approximately 13.6 million people, well above world average, who had already taken the medical cosmetic surgeries.(See Figure 1) He also made a prediction that people in China who had done such surgeries would surpass those in Brazil in 2018 and rank the second place in the world. (Townley,2019).

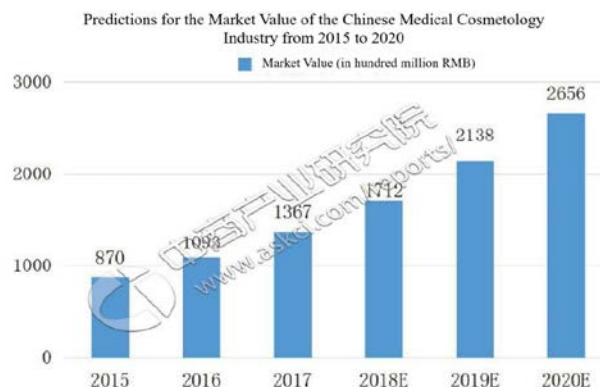


Figure 1. Medical Cosmetology Data in China

But many people, because of financial reasons, cannot take out a great amount of money to afford the high cost operation fees. According to Toronto Cosmetic Clinic (2019), the current price of doing medical cosmetic surgeries is indeed very high: programs such as Breast Augmentation, Breast Lift, and Liposuction are all above 5000 dollars. (See Table 1)

Table 1. Price of Medical Cosmetic Surgeries

Names of operation	Price
Breast Augmentation	\$8500
Breast Lift	\$10,500
Liposuction	\$5,000

To address the needs of many people, companies have medical cosmetic surgeries loan programs for those who need money for operation in a very short time. People can borrow the money to get their surgeries done first and then give back their money in a certain amount of time. However, even though it seems easy to give loans to people and wait until they pay back, reality tells these companies a different thing. There are always some people who do not pay their money back on time. Just as a survey done by Kaiser Family Foundation (2012) revealed, 7% of people who borrowed money couldn't give their money back until years later, and 48% of people would give their money back about one year later. ( See Figure 2)

**Percent of Privately Insured Adults with Difficulty Paying Medical Bills by Deductible, 2012**

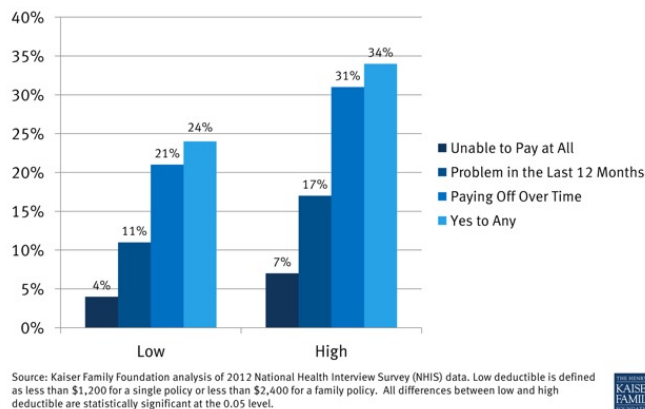


Figure2. Data Analysis

These behaviors will cause companies not able to get money on time after giving out money to other people, leading to loss of money or even bankruptcy in the end. Crowd psychology may make this situation even worse (Zhu et al., 2018). Therefore, how to keep the risk of not paying back money on time at its lowest point becomes the current problem which people are studying now. This research uses decision tree analysis in data mining model to analyze what causes people not to pay back money on time so that companies can have a more comprehensive and thorough review of the people who want to borrow money. The decision tree will help ruling out people that are unlikely to pay their money back and only borrowing money to people who pass the tree test. The tree will help to lower the risk and make companies earn economic profits in the long run, which is the meaning of this research.

## 2. Literature Review and Methodology

### 2.1. Literature Review

So far, some analysts in medical cosmetic surgeries companies have already tried to use qualitative methods to analyze what kinds of people are more likely to pay their money back on time. Their conclusions have already let the companies to gather information such as people's wages, whether they are on the black list, whether they have recently borrowed money from other companies, and whether they have lived a while in a city in order to rule out some people who are not qualified. What's more, these companies lend low interest loans as many as possible and make sure that people have mortgages before borrowing money to stifle the possibility of people not giving back money forever. According to the data from the medical cosmetic surgeries loan company where 1200 data were acquired, over 1100 out of 1200 people paid back their money in the end. Besides, some researchers have found out that a person is more likely to default if he writes a long statement when applying the loan (Herzenstein et al., 2011); some researchers have discovered that people are not paying back money on time if they use real names instead of nicknames (Guo, 2016); and other researchers pointed out that if a person looks nice, then he will tend to pay the money back. (Duarte, 2012)

## 2.2 Current Methods

But all of the aforementioned methods have flaws. Using blacklist cannot rule out new borrowers who will not pay their money on time, and asking about people's wages, whether people have borrowed money from other companies, and whether they have lived here for a relatively long time will involve in the privacy issues with people's personal communicational record and voice recordings of phone calls. What's more, people mortgage their properties to borrow money for operation and pay a little more than they should if they pay all the money once can only ensure that they will eventually pay their money back but not pay their money back on time. According to the data from the research, only 694 people never default from the beginning to the end. In addition, the criteria which the researchers have proposed such as "more statements", "use real names", and "look nice" are vague and varies with people. It is hard for companies to use them to differentiate each person. So, the methods which companies use now are not enough to solve the default problems. It is better to use decision tree in data mining to analyze the influence that different factors have towards people to refine the current qualitative methods. After looking at the research paper, people can know about the meaning of using decision tree, the current status of studying medical cosmetic surgeries and the methods' problems, the definition and principles of the decision tree, the advantages of using the decision tree method, and how to use decision tree to process data and get the conclusion.

## 3. The principles and advantages of using decision tree

### 3.1 The principle of decision tree

Decision tree is a tree-like structure (See Figure 3) which lists all the factors that are related with the final results in different layers according to their importance. In each layer, conditions are used to separate each factor into multiple branches, and each person with different traits can find himself on a particular branch. If most of the people on a branch pays money back on time or not on time, then all the people on this branch will be concluded as paying back on time or not paying back on time.

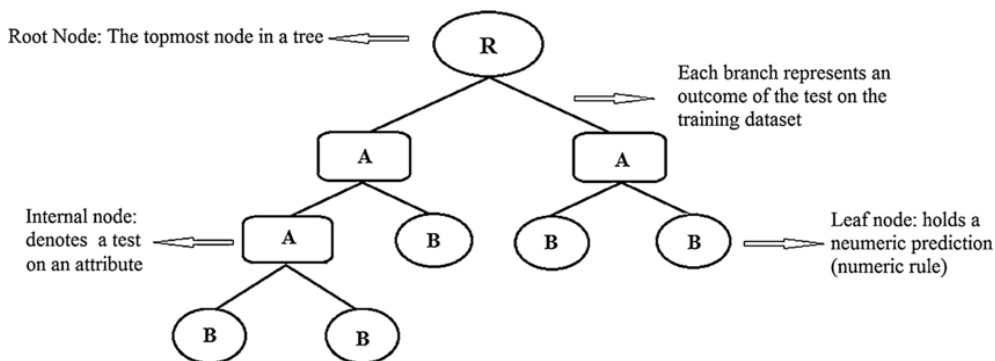


Figure 3. Decision Tree Model

### 3.2. The advantages of drawing the decision tree

Users have no troubles understanding the decision tree, for it doesn't require much knowledge. As long as the test data can be written as "factors conclusions", then decision tree can be applied to the data. Decision tree is efficient and suitable for large-sized data, and they can categorize different data really well. The sampling method in each decision tree is a tree-like structure, so it makes tree clear and simple and can turn every path to the internal nodes to the "if then" form. This alteration makes decision tree very easy to understand. Rules and procedures have already been proposed to help people successfully drawing a decision tree. So far there are three methods to draw the tree: ID3, C45, and CART. In this research paper, the ID3 method will be used to process the data. ID3 was proposed by Quinlan in 1986, and it is a well-known method in data mining modeling. The proposal of ID3 opens up a new era of decision tree making. ID3 is one of the most decision tree making methods in the world. According to the ID3 method, the concept of information entropy is

introduced to measure the importance of each factor to the results. The formula of calculating information entropy is  $E = -N \log_2 N$  (Shannon Formula). Since the factor with the smallest value of information entropy is more important in deciding whether a person will eventually pay back their money on time, it will be put at the top layer.

#### 4. The realization of decision tree and data processing

##### 4.1 Variables declaration

The data source in this research paper is 1200 borrowers' micro-data (see Table 2) retrieved from an internet medical surgery loan platform from May 2016 to October 2018. There are primarily eighteen variables that are involved in this data, and they are divided into five different categories according to their types:

The first type of variables is personal information data (9 variables): The work order ID of the borrowers, their phone numbers, the date when their contract started, and their occupations, educational levels, salaries, marriages, hospitals where the operations will be held, as well as the banks they are using. These are the fundamental data of these borrowers.

The second type of variables is loan data (5 variables): The detailed information of the current loan that people borrow from this platform such as the amount of money that people have borrowed, the date when they started to borrow money, whether they are defaulting on loans, and whether they have experiences of borrowing money before.

The third type of variables is loan products' data (1 variable): whether people choose to buy loan products on the platform.

The fourth type of variables is credit report from the bank (1 variable): the credit report given by a bank to evaluate a person's credibility.

The fifth type of variables is risk evaluation report (2 variables): borrowers' Tencent anti-cheating score and Share-shield anti-cheating score.

Table 2. Data Statistics

Money	Date	Occupation	Periods	Place	Hospital	Bank	Educational level	Marriage	Default
1	2016/11/4	Service	12	Shandong	HHH	CCB	5	1	1
4	2016/11/24	Industry	6	Anhui	FOH	ICBC	4	2	1
1	2016/11/21	Service	12	Shandong	JDH	CCB	3	1	1
4	2016/11/24	Service	12	Hunan	CHH	BOC	4	0	1
4	2016/11/27	Industry	12	Henan	HYH	ABC	3	0	1
2	2016/11/30	Service	12	Shandong	TMH	ABC	4	0	1
1	2016/12/12	Retail	12	Anhui	HAH	ICBC	5	1	1

##### 4.2 Data processing

In the original form, there are many word descriptions and vacant data which cannot be identified by the computers, so data cleaning is done to rule out all the empty data and other data such as the time when they borrowed money, the IP number of people's computers, and the time they borrowed money which are unrelated to the results. What's more, words' descriptions are changed to numbers to represent the status of these borrowers. (if these borrowers do not default on loans, then their results are written as "0", but if these borrowers default on loans, then their results are written as "1". Their educational levels range from 1-8, with one represents primary school degree, and eight represents PHD degree. The places where they come from are divided into 27 different groups according to the division of Chinese provinces. Moreover, people are divided into four groups (1-4) due to the amount of money they borrowed from the medical cosmetic surgery loan company.) For people who are defaulting on loans, those who gave their money to the companies in two days after the deadline are not considered as "default" and are put into the "0" group. They may just simply forget the date to give back money, or there are some delays in the money transferring process of banks. Companies are not likely to suffer any losses just because

people return money in two days after the deadline, and these people are not defaulting on loans because of financial reasons. However, for people who return their money back more than two days after the deadline, they are intentionally not giving their money back on time. Among the 1200 people who are not in the blacklist, 905 of them didn't give their money back on time, which consists of 76.5% of people in the whole group. The remaining 295 people do not pay their money back on time. This fact reveals that these borrowers are overall trustworthy and credible.(See table 3)

Table 3. Data Classification

Status	back on time	Default( $\leq 2$ )	Default( $> 2$ )
numbers	694	295	211
Ratio	57.8%	24.6%	17.6%

Then, 60% of the overall 1200 data were used to construct the data mining model –decision tree. The other 40% of my data are used to test my data. By using the formula:  $E = -N \cdot \log_2 N$  to calculate the information entropy of multiple factors, three factors, the places they are living in, they amount of money they borrowed, and their educational levels, are found out to be appealing-their information entropies are relatively low, indicating that these factors are very important in determining whether people are going to default on loans or not. Through further calculations, the amount of money is proved to have the lowest value of information entropy, so it is put in the top layer. The place where people live in at the second layer, and the educational level is the third important factor in determining whether people will give back their money on time. After these calculations, the amount of people who give their money on time is compared with the amount of people who don't on each branch of the decision tree. If one branch has more people who give back their money on time than people who do not give their money back on time (or the ratio is above 1:1), then the whole branch is concluded as giving their money back on time, vice versa. Finally, the decision tree is tested in the remaining 40% of the 1200 data, and the success rate is 75.7%.

## 5. Conclusion

Overall, the decision tree reveals that big cities such as Beijing, Shanghai, and Guangzhou don't have great results of paying back money on time (See part of my decision tree in Figure 4). The ratio of paying back money on time to not paying back money on time is below average of the overall data (905:295). But in other provinces such as Jiangsu, Zhejiang, Hunan, and Hubei have perfect record of people giving back money on time, which is above the average of the overall data. The reason attributes to the characters of local people or the domestic culture in the area, so people in that area will give back money on time after they borrow money. What's more, people who have borrowed more money are thought to be more likely to default on loans, whereas people who have borrowed less money are thought to be less likely to default on loans. Besides, there's an obvious threshold in educational levels. People who have educational levels above and equal to bachelor's degree tend to give their money back on time compared to people who are below bachelor's degree. This phenomenon is conspicuous in big cities (Beijing, Shanghai, Guangdong) of China. Therefore, decision tree is proved to be more scientific and useful, as well as easier to judge than other qualitative methods which companies are using now.

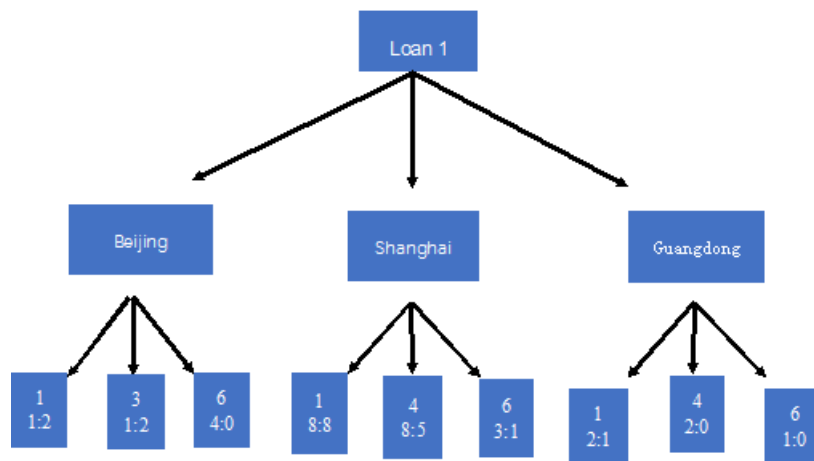


Figure 4. Decision Tree

## References

- [1] Herzentein, M., S. Sonnenstein and U. M. Dholakia (2011): "Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions", *Journal of Marketing Research*, 48(SPL), S138.
- [2] Duarte, J., S. Seigel and L. Young (2012): "Trust and credit: the role of appearance in peer-to-peer lending", *Review of Financial Studies*, 25(8), 2455-2484
- [3] Feng, G. (2016). The Usage of Internet Nickname with the Success of Paying back P2P Loan on Time. *Economic Science*, 2016(6):102-116.
- [4] Zhu, J., Zhu, J., Ghosh, S., Wu, W., & Yuan, J. (2018). Social influence maximization in hypergraph in social networks. *IEEE Transactions on Network Science and Engineering*.
- [5] Jingsong, L. (2016). Research on Whether People will pay back money on time in a commercial bank based on data mining. *Economics Studies*, 2016(7): 150-159
- [6] Townley, C.(2019). Cosmetic Surgery is on The Rise, New Data Reveal. Retrieved June 17, 2019, from <https://www.medicalnewstoday.com/articles/324693.php>.
- [7] Franklin, J.(2012). What Happens When Debt is Unpaid? Retrieved June 16, 2019, from <https://smallbusiness.chron.com/happens-debt-unpaid-43266.html>.
- [8] Bird, R.(2018). Decision Trees- A Simple Way to Visualize A Decision. Retrieved June 15, 2019, from <https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>.
- [9] Jazuli, H.(2018). An Introduction to Decision Tree Learning: ID3 Algorithm. Retrieved June 16, 2019, from <https://medium.com/machine-learning-guy/an-introduction-to-decision-tree-learning-id3-algorithm-54c74eb2ad55>.
- [10] Birken, E. (2016). How Do Loan Companies Make Money? Retrieved June 17, 2019, from <https://studentloanhero.com/featured/loan-companies-make-money/>.
- [11] Toronto Cosmetic Clinic (2019). Plastic Surgery Costs. Retrieved June 17, 2019, from <https://www.tcclinic.com/>
- [12] Kaiser Family Foundation (2012). Analysis of 2012 National Health Interview Survey Data. Retrived June 16, 2019, from <http://ww25.kaiserfamilyfoundation.com/>.