

# 基于 Logistic 回归的汽车经销商客户流失预警与回厂预测

宋丽娜 张文杰 王金良

(指导教师: 周少玲)

河北工程大学

## 一、案例背景及问题描述

在经济社会快速发展的今天,人们对物质生活的要求不断提高。而汽车作为出行的重要交通工具,近年来,人们对此的需求也不断增多,这就助推了汽车市场的迅速发展。与此同时,4S店也快速兴起,其主要经营汽车销售、零配件、售后服务及信息反馈等。因为其通过与汽车生产厂家进行产销合作,因此,4S店在品牌知名度、售后服务及购物环境等方面都存在明显的优势。但汽车产业快速发展的同时,汽车品牌及其数量也迅速增多,从而使得行业竞争激烈,然而4S店售后服务却没有得到强有力的发展,合资品牌车和国产车客户出现较为严重的流失<sup>[1-4]</sup>。近两年来,随着汽车行业几乎达到饱和状态,新车销量增长也有所减缓,作为汽车服务行业拥有大量存量客户的汽车4S店,也开始越来越重视售后服务。因此,通过对汽车4S店存量客户流失预警与客户回厂预测的研究,可以为汽车4S店对流失客户制定针对性的挽回策略提供依据,对提高汽车4S店的管理效率,促进汽车服务行业中客户管理的发展具有重要意义。

本文将首先通过汽车4S店的客户数据进行分析建模,对客户流失建立分级标准,并根据该标准和所建立的模型对相应的客户给出流失得分及其比例;其次,再针对客户以往的回厂情况,对客户下次回厂的时间进行预测;最后,再对已建立的模型进行评判。

## 二、基于 Logistic 回归的客户流失分级预警

### (一) 问题描述

第一题主要是建立汽车4S店客户的流失预警分级标准,将测试集中的客户数据划分为五个等级,并给出相应的流失得分(1-5)及其比例。此题主要是建立合适的模型,分析现有公司的客户流失比例,维护现有客户,提高客户满意度。

## （二）数据预处理

### 1. 数据来源及指标选择

本次原始数据来源于大赛提供的数据，参考其他相关文献，以及对数据进行模型验证，存在一些指标未通过模型检验，遂选取以下七个指标较为合适，分别为 IS\_LOST，CUST\_AGE，CAR\_AGE，CAR\_PRICE，ALL\_BUYINS\_N，SALES\_TOTAL，MEAN\_MAINTENANCE\_TIME 这七个指标。其中 IS\_LOST 为分类变量，其余指标都为连续变量。提供了选取指标含义及符号说明。

表 1 问题一指标选取及符号说明

指标名	含义
CUST_ID	客户 ID
IS_LOST	客户是否流失：1=流失，0=未流失 $Y$
CUST_AGE	客户年龄 $X_1$
CAR_AGE	车龄：单位：天 $X_2$
CAR_PRICE	车辆销售价格 $X_3$
ALL_BUYINS_N	在 4S 店购买保险总次数 $X_4$
SALES_TOTAL	总费用 $X_5$
MEAN_MAINTENANCE_TIME	平均保养维修时间 $X_6$
ORDER_DATE	订单日期
CROSSTOWN_DATE	本次订单完成交车时间

### 2. 数据处理

#### 1) 数据缺失值处理

为了保证数据的完整性，首先对数据重复值进行了查找，结果如表 2 所示，其中总的的数据共有 51075 条，“在 4S 店购买的总次数”和“客户年龄”分别有 4631 和 475 个缺失数据。首先考虑缺失量相对较多的变量——“在 4S 店购买的总次数”，相对于总的的数据量而言，其缺失的数据量较少，且该数据为离散值，客户与客户之间的购买情况也相互独立，互不影响，如果利用未缺失的数据对其进行填补的话可能难以反映该数据的真实情况，因此对于“在 4S 店购买的总次数”的缺失值进行剔除。

表 2 缺失值情况 1

	客户年龄	车龄	车辆销售价格	总费用	平均保养修理维修时间	在 4S 店购买保险的总次数	是否流失
缺失数量	475	0	0	0	0	4631	0

剔除后总的数据为 46444 条，再次对数据的重复值进行查看，从表 2 可以看出“客户年龄”的变量仍然存在 409 个缺失数据，缺失数据较少，且从图 1 可知，“客户年龄”的分布具有集中趋势，因此针对该变量的缺失数据，就利用众数对其进行了填充。

表 3 缺失值情况 2

	客户年龄	车龄	车辆销售价格	总费用	平均保养修理维修时间	在 4S 店购买保险的总次数	是否流失
缺失数量	409	0	0	0	0	0	0

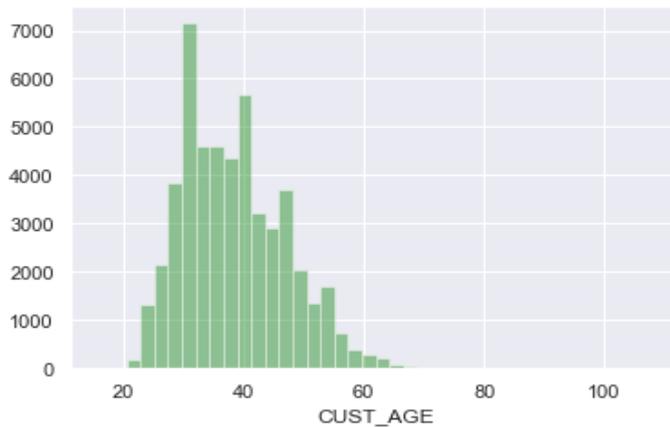


图 1 客户年龄分布情况

## 2) 数据计算

所选指标中是否流失、客户年龄、车龄、车辆销售价格和在 4S 店购买保险总次数由原始数据“car\_info\_train”/“car\_info\_test”提供。“car\_srv\_train”/“car\_srv\_test”中为客户具体保养维修信息，“car\_info\_train”/“car\_info\_test”中每个用户的总费用为“car\_srv\_train”/“car\_srv\_test”对应客户对应的每次维修保养费用之和，“car\_info\_train”/“car\_info\_test”中每个用户平均保养维修时间计算公式为：

$$\text{每个用户平均保养维修时间} = \frac{\sum_{i=1}^n (\text{订单完成时间} - \text{订单发生时间})}{\text{保养维修次数}}$$

其中， $i$  为每个用户第几次保养维修， $n$  为每个用户维修总次数。数据结果见文件“问题一结果数据”。

### 3) 多重共线性检验

本文在对已选指标做 Logistic 回归分析前，需要对数据进行多重共线性检验。以  $X_j$  为解释变量作对其他解释变量辅助线性回归的可决系数用  $R_j^2$  表示，则可以说明解释变量  $X_j$  参数估计值  $\hat{\beta}_j$  的方差可以表示为

$$\text{Var}\left(\hat{\beta}_j\right)=\frac{\sigma^2}{\sum x_j^2} * \frac{1}{1-R_j^2} = \frac{\sigma^2}{\sum x_j^2} * \text{VIF}_j$$

式中  $\text{VIF}_j$  是变量  $X_j$  的方差膨胀因子，即

$$\text{VIF}_j = \frac{1}{1-R_j^2}$$

由于  $R_j^2$  度量了  $X_j$  与其他解释变量的线性相关程度，这种程度越强，说明变量间的多重共线性越强， $\text{VIF}_j$  也就越大。反之， $X_j$  与其他解释变量的线性相关程度越弱，说明变量间的共线性越弱， $\text{VIF}_j$  也就越接近 1。经验表明， $\text{VIF}_j \geq 10$  时，说明解释变量与其余解释变量之间有严重的多重共线性。

接下来对已选指标进行多重共线性检验，结果见下表 4:

**表 4 问题二模型多重共线性检验**

	特征值	条件指标	常量	客户年龄	车龄	车辆销售价格	在 4S 店购买保险总次数	总费用	平均保养修理时间
1	3.949	1.000	0.00	0.00	0.01	0.01	0.01	0.01	0.01
2	0.955	2.034	0.00	0.00	0.07	0.00	0.00	0.68	0.18
3	0.867	2.134	0.00	0.00	0.87	0.00	0.00	0.10	0.01
4	0.814	2.203	0.00	0.00	0.05	0.00	0.00	0.21	0.80
5	0.253	3.955	0.00	0.00	0.00	0.40	0.59	0.00	0.00
6	0.138	5.345	0.06	0.09	0.00	0.58	0.37	0.00	0.00
7	0.024	12.96	0.94	0.91	0.00	0.00	0.02	0.00	0.00

由表 4 可知，各变量之间的 VIF 值都小于 10，由此认为此模型不存在多重

共线性，可以用于后续的 Logistic 回归分析。

### (三) Logistic 回归模型

汽车 4S 店要想实现长期稳定的盈利，必须重视售后客户关系维护，重视客户的流失率管控，对汽车 4S 店存量客户进行流失预警有着十分重要的意义<sup>[4-6]</sup>，为对流失客户制定针对性的挽回策略提供依据，促进汽车 4S 店的管理效率。对客户流失水平进行分级，有利于提供有适用性与针对性的策略。所选因变量为训练数据集“car\_info\_train”中客户是否流失  $Y$  是分类变量，数据的类别只有两种：0 和 1，即数据是 0-1 型因变量问题。根据需要最终将客户划分为五个等级，故考虑使用逻辑回归对问题进行建模求解。

逻辑回归模型是一种经典的分类算法，该模型可以根据一组自变量输入计算出归属每种类别的概率，一般常用的是二项逻辑回归。

其条件概率分布为：

$$P(Y = 1|x, \theta) = \frac{1}{1 + e^{-\theta^T x}}$$
$$P(Y = 0|x, \theta) = \frac{1}{1 + e^{\theta^T x}}$$

Logistic 回归方程为：

$$f(x) = \frac{1}{1 + e^{-x}}$$

其中， $x \in R^n$  是输入，称为实例的特征； $Y \in (0, 1)$  是输出，两个分类变量对应是客户未流失和客户流失； $\theta \in R^n$  是参数，称为权值向量，且

$$\theta^T x = \sum_{i=1}^n \theta_i x_i, i = 0, \dots, n.$$

权值向量反应了输入中每一项对分类结果的影响力，某输入项对应的权重为正数表示此项增大了结果的概率，为负表示减小了结果的概率；权重绝对值很大表示该项对结果的影响力很强，反之权重绝对值接近于表示该项对结果几乎没有影响<sup>[7]</sup>。

#### (四) 模型建立与求解

##### 1. 模型建立

对于给定的训练数据集“car\_info\_train”与“car\_srv\_train”，选取合适的指标对建立模型有着非常重要的意义。初步选取的指标变量为客户年龄、车龄、车辆销售价格、在 4S 店购买保险总次数，以及通过数据集“car\_srv\_train”计算出的客户在店内消费总费用与平均保养修理时间。本案例利用 SPSS 软件运用 Logistic 回归方法对客户是否流失进行统计分析。

对该数据利用 SPSS 软件进行 Logistic 回归分析，得到表：

表 5 原始数据的模型参数估计

	B	S.E.	Wald	df	Sig.	Exp(B)
客户年龄 $X_1$	0.000112	0.001	0.006	1	0.938	1.000
车龄 $X_2$	0.001	0.000	2924.275	1	0.000	1.001
车辆销售价格 $X_3$	0.000002	0.000	109.83	1	0.000	1.000
在 4S 店购买保险总次数 $X_4$	-0.296	0.013	552.718	1	0.000	0.743
总费用 $X_5$	0.000037	0.000	415.811	1	0.000	1.000
平均保养修理时间 $X_6$	0.007	0.001	36.265	1	0.000	1.007
常数	-1.851	0.060	944.034	1	0.000	0.157

表中，Wald 是回归系数检验的统计量值

$$\text{Wald} = \left( \frac{B}{\text{S.E.}} \right)^2$$

Sig. 是 Wald 检验的显著性水平。可以看到，客户年龄  $X_1$  的统计量  $P$  值远大于 0.05，即回归系数不显著，基于此，决定将其剔除，按照上述步骤重新进行 Logistic 回归分析。

为得到估计模型，需对模型中参数进行估计，估计结果如下表：

表 6 剔除客户年龄  $X_1$  后的模型参数估计

	B	S.E.	Wald	df	Sig.	Exp(B)
车龄 $X_2$	0.001	0.000	3081.907	1	0.000	1.001
车辆销售价格 $X_3$	0.000002	0.000	119.718	1	0.000	1.000
在 4S 店购买保险总 次数 $X_4$	-0.297	0.013	562.506	1	0.000	0.743
总费用 $X_5$	0.000037	0.000	420.213	1	0.000	1.000
平均保养修理时间 $X_6$	0.007	0.001	36.274	1	0.000	1.007
常数	-1.854	0.037	2496.23	1	0.000	0.157

从表中可以看出，在显著性水平为 0.05 时，车龄  $X_2$ 、车辆销售价格  $X_3$ 、在 4S 店购买保险总次数  $X_4$ 、总费用  $X_5$ 、平均保养修理时间  $X_6$  都是显著的，所以保留在模型中。正相关系数表明，自变量值的增加会增加客户流失的概率，负相关系数表明，自变量值的增加会减少客户流失的概率。结合现实意义，观察各变量可以发现，随着车龄、车辆价格、在 4S 店的总消费和平均保养修理时间的增加，客户流失的概率也会随之增加，即客户回厂进行保养与维修的几率减小。随着客户在 4S 店购买保险总次数增加，客户流失的几率就会减少，即回厂的几率就会增加。这与实际意义相符，说明模型效果较好。

## 2. 模型求解

对于给定的训练数据集，SPSS 运用极大似然估计法估计的模型参数  $\theta$ ，根据表估计的参数，得到估计的逻辑函数为：

$$\hat{f}(X) = -1.854 + 0.001X_2 + 0.000002X_3 - 0.297X_4 + 0.000037X_5 + 0.007X_6$$

根据逻辑函数可以写出 Logistic 回归方程：

$$\ln \frac{p}{1-p} = \hat{f}(X)$$

将式子转化为概率形式：

$$p = \frac{\exp[\hat{f}(X)]}{1 + \exp[\hat{f}(X)]}$$

## （五）模型评价

建立的模型得到下表所示的拟合效果。

表 7 模型预测的正确率

观察值 \ 预测值	预测值		正确百分比
	0	1	
0	35728	736	98.0
1	9111	869	8.7
整体百分比			78.8

表反映了模型拟合预测的正确率,从表中可以看出,观测值  $Y=0$  有 36464 个,预测 35728 个为 0,预测正确率为 98%,观测值  $Y=1$  有 9980 个,预测 869 个为 1,预测正确率为 8.7%,其中总预测率达 78.8%,高预测率说明模型拟合效果较好。

表 8 曲线下面积

Area	标准错误	渐进显著性	95%信赖区间的下限	95%信赖区间的上限
0.755	0.003	0	0.75	0.76

表为 ROC 曲线下面积 AUC, AUC 的值就是处于曲线下方的那部分面积的大小,它可以衡量一个分类器的性能好坏。一般 AUC 的值介于 0 和 1 之间,较大的代表了较好的表现。如果模型做的好,则其曲线是凸形的,大于否则为凹形的,说明模型分类结果是反的,把正例判为负例,负例判成正例了,此时只需要将判别结果取反即可<sup>[7]</sup>。可见 AUC 的值为 0.755 大于 0.75,显示该模型预测模型的区分能力较好。

对测试数据集“car\_info\_test”与“car\_srv\_test”进行处理,计算出总费用与平均修理时间两个变量,并将其与其余变量代入模型中进行测试,预测值与原始因变量的值进行比较。结果发现,预测值的准确性达 83.73%。预测结果见文件“问题一结果数据”。

## （六）分级标准

对于一个实例输入  $X$ ，按照该模型可以计算出  $P(Y = 1|X, \theta)$ ，这是一个介于 0 和 1 之间的数，利用该模型对测试集中数据进行预测，并对预测所得概率设定阈值，以达到客户流失预警分级的效果。阈值设定，即分级标准如下表：

表 9 分级标准

阈值	分级	得分
[0,0.2)	1 级	1 分
[0.2,0.4)	2 级	2 分
[0.4,0.6)	3 级	3 分
[0.6,0.8)	4 级	4 分
[0.8,1]	5 级	5 分

概率越趋近于 1，说明客户流失的概率越大，越趋近于 0，说明客户流失的概率越小。

根据所得拟合模型对训练集“car\_info\_test”中客户是否流失进行预测，预测结果与客户相应的流失得分在文件“问题一结果数据”中。其中，客户流失得分数量及比例如下表：

表 10 客户流失得分及比例

得分	数量	比例
1 分	5729	56.50%
2 分	3462	34.20%
3 分	795	7.80%
4 分	134	1.30%
5 分	2	0.20%
总计	10122	100%

### 三、基于 Logistic 回归的客户回厂时间预测

#### (一) 问题描述

问题二在于建立汽车客户回厂预测模型,预测客户下次可能回 4S 店的时间。利用现有数据,建立合适的模型,对现有客户预测下次回厂以日为单位的准确时间,分析客户回厂时间间隔问题。

#### (二) 数据预处理

##### 1. 数据来源及指标选取

本文在查阅相关文献<sup>[8-14]</sup>,在进行模型检验后,剔除不显著变量,选取合适的指标作为汽车客户回厂时间预测的因素,各 2925 条数据,分别为 IS\_LOST, AVG\_ORDER\_GAP, AVG\_DAYS\_IN\_DEALERSHIP, ORDER\_CNT, AVG\_MILEAGE\_GAP, TOTAL\_DISCOUNT 这六个指标,其中 IS\_LOST 为分类变量,其余指标都为连续变量,

表 11 提供了选取指标含义及符号说明。

表 11 问题二指标选取及符号说明

指标名	含义
IS_LOST	客户是否回厂: 1=不回厂, 0=回厂. $F$
AVG_ORDER_GAP	时间周期: 平均每隔多少天数去保养或者维修 $Z_1$
AVG_DAYS_IN_DEALERSHIP	在 4S 维修平均天数 $Z_2$
ORDER_CNT	做过几次维修保养 $Z_3$
AVG_MILEAGE_GAP	里程周期: 平均每隔多少 <i>mile</i> 去保养或者维修 $Z_4$
TOTAL_DISCOUNT	维修保养总折扣 $Z_5$
LAST_MAINTENANCE_TIME	最近维修时间

##### 2. 数据处理

###### 1) 数据计算

“car\_info\_train”提供汽车客户的原始信息,“data\_info”为客户保养维修的

具体信息，提取“car\_info\_train”和“data\_info”中相同的客户的信息，包括“data\_info”中客户是否流失，时间周期，在4s维修平均天数，做过几次维修保养，里程周期，维修保养总折扣等六个指标，其中将客户是否流失解释为客户是否回厂，1表示不回厂，0表示回厂，且数据不存在缺失。最晚维修时间是将“car\_info\_train”和“data\_info”中相同的客户的信息提取出，计算在“data\_info”中每个客户最晚维修保养的时间。对客户下次回厂时间进行预测，数据结果见文件“问题二结果”。

## 2) 多重共线性检验

在进行 Logistic 回归分析前，需要对已选指标进行多重共线性检验。

表 12 问题二模型多重共线性检验

特征值	条件指标	常量	时间周期	在 4s 维修平均天数	做过几次维修保养	里程周期	维修保养总折扣
3.383	1.000	0.01	0.02	0.0	0.02	0.02	0.01
0.978	1.860	0.00	0.01	0.00	0.01	0.01	0.85
0.806	2.049	0.00	0.02	0.90	0.00	0.02	0.00
0.502	2.597	0.02	0.05	0.08	0.35	0.12	0.14
0.223	3.896	0.11	0.30	0.00	0.12	0.76	0.00
0.108	5.584	0.86	0.61	0.00	0.50	0.06	0.00

由表 12 可以看出，VIF 值都远低于 10，说明各变量之间无明显相关性，无多重共线性，可进行 Logistic 回归分析。

## (三) 模型建立与求解

### 1. 模型建立

对客户回厂时间进行预测，不仅可以提前准备为客户提供更好的服务，而且可以适当的鼓励和引导客户回厂以增加回厂率。因此，对客户的回厂时间进行准确预测有着十分重要的意义。选取客户是否流失作为回厂依据，即因变量为客户是否回厂 F，是分类变量，并对数据集“data\_info”、“car\_info\_train”与“car\_srv\_test”进行处理，选取的自变量为时间周期  $Z_1$ 、在 4S 维修平均天数  $Z_2$ 、做过几次保养维修  $Z_3$ 、里程周期  $Z_4$  以及维修保养总折扣  $Z_5$ 。对数据利用 SPSS 软件运用 Logistic 回归方法对客户是否回厂进行统计分析，并对模型进行求解转换，从而实现回

厂时间的预测。

对处理后的数据利用 SPSS 软件进行 Logistic 回归分析，得到表 13：

表 13 模型参数估计

	B	S.E.	Wald	df	Sig.	Exp(B)
时间周期 $Z_1$	-0.003	0.001	20.35	1	0.000	0.997
在 4s 维修平均天数 $Z_2$	0.015	0.004	13.71	1	0.000	1.015
做过几次维修保养 $Z_3$	-0.105	0.013	70.414	1	0.000	0.9
里程周期 $Z_4$	0.000019	0	3.919	1	0.048	1.000
维修保养总折扣 $Z_5$	0.00028	0	7.123	1	0.008	1.000
常数	-0.415	0.123	11.327	1	0.001	0.66

从表 13 中可以看出，Wald 检验统计量的显著性水平 Sig. 皆小于 0.05，即回归系数显著性检验皆通过。

## 2. 模型求解

运用极大似然估计的方法对模型参数进行估计，且估计结果如下所示，得到估计的逻辑函数为：

$$\hat{g}(Z) = -0.415 - 0.003Z_1 + 0.015Z_2 - 0.105Z_3 + 0.000019Z_4 + 0.00028Z_5$$

根据逻辑函数可以写出 Logistic 回归方程：

$$\ln \frac{p'}{1-p'} = \hat{g}(Z)$$

转化为概率形式：

$$p' = \frac{\exp[\hat{g}(Z)]}{1 + \exp[\hat{g}(Z)]}$$

由上式可以得到客户回厂的概率值，但是我们要对客户回厂日期进行预测，又已知自变量中有关于客户回厂的平均时间间隔，因此对模型进行转换，将时间周期转化为因变量，其余变量为自变量，并为分类变量的概率设置  $P$  值，预测在此概率下，客户下次的回厂时间。

转换后的模型为：

$$Z_1 = -333 \ln \frac{p'}{1-p'} - 138 + 5Z_2 - 35Z_3 + 0.0063Z_4 + 0.0933Z_5$$

所用数据是否回厂为 0-1 分类变量，其中 0 是回厂，1 是不回厂。当概率越小时，回厂几率就越大，反之亦然。为使预测的时间得到充分应用，确定回厂概率阈值为 0.2，回厂几率大，且得分为 1 分的客户，即预测客户回厂几率较大时，客户的回厂时间。当  $p'$  值为 0.2 时，代入模型，预测时间间隔的方程为：

$$Z_1 = 323.538 + 5Z_2 - 35Z_3 + 0.0063Z_4 + 0.0933Z_5$$

由该模型可以计算出各客户平均回厂时间间隔。

接下来对数据集“*car\_srv\_train*”中的数据进行处理，提取出各客户最近一次订单日期，用于客户下次回厂日期预测。预测模型为：

$$T_{next} = T_{last} + Z_1$$

其中， $T_{next}$  为客户下次回厂时间， $T_{last}$  为客户最近一次回厂时间， $Z_1$  为客户平均回厂时间间隔。客户下次回厂时间预测数据见文件“问题二结果”。

#### （四）模型评价

所建立的 Logistic 回归模型的拟合效果如表 14 所示：

表 14 模型预测的正确率

观察值 \ 预测值	预测值		正确百分比
	0	1	
0	2359	2	99.9
1	561	3	0.5
整体百分比			80.8

表 14 反应了模型预测的正确率，从表中可以看出，观测值  $Y=0$  有 2361 个，预测 2359 个为 0，预测正确率为 99.9%，观测值  $Y=1$  有 564 个，预测 3 个为 1，预测正确率为 0.5%，其中总预测率达 80.8%。说明模型拟合效果较好。

## 四、总结与建议

### （一）总结

针对客户的流失状况。首先利用了“*car\_info\_train*”和“*car\_info\_test*”两个

文件的数据，对其进行了相应的计算及数据的预处理，再通过 SPSS 软件对变量的显著性进行了分析，最终筛选出了车龄、车辆销售价格、在 4S 店购买保险总次数、总费用及平均保养修理时间这五个变量作为自变量，其因变量为客户的流失情况，为分类变量。从而，针对该数据集的特征建立了 Logistic 回归模型，通过模型的预测率及 ROC 曲线下面积 AUC 值可知所建模型较好，符合现实意义，能够较好的反映客户流失的真实情况，说明该模型能够很好的对客户流失状况进行预测。最后，利用所建立的 Logistic 回归模型对测试集进行了预测，并通过设定阈值将客户分为了五个等级，其分别表示客户的不同流失程度。同时，对各等级所占比例进行了统计，能够很好的起到预警的作用。

针对客户下次回厂的时间预测。提取“car\_info\_train”和“data\_info”中相同的客户的信息，选取了时间周期、在 4S 维修平均天数、做过几次保养维修、里程周期以及维修保养总折扣五个变量为自变量，其中将客户是否流失解释为客户是否回厂（1 表示不会厂，0 表示回厂），接着对已选指标进行多重共线性检验，结果显示变量间无多重共线性，可进行 Logistic 回归分析。最后，建立 Logistic 回归模型对客户是否回厂进行统计分析，并对模型进行求解转换，从而实现了对回厂时间的预测，通过对模型预测正确率达到 99.9%，说明该模型拟合效果较好。

## （二）建议

（1）注重对现有客户的维护。从流失得分比例来看，流失概率为 0~0.4 的客户占总客户的 90.7%，其中流失概率在 0.2~0.4 之间的客户比例为 34.2%。也就表明绝大部分的客户的流失概率均处于较低的水平，因此 4S 店应该更加注重对现有客户的维护，定期对老客户进行回访，并采集相关建议。

（2）根据客户建议不断提高服务水平。4S 店作为服务行业的一员，其发展方向与服务方式都应从客户的需求出发，只有充分考虑了客户的体验，才有利于实现其长远的发展目标。因此，4S 店应及时采取新老客户的建议，通过对其建议的充分考虑，进而有针对性的对其综合服务模式、水平做出改进，不断提高其服务水平。

（3）充分利用客户的历史数据。客户是 4S 店盈利和发展的基础，是极为重要的无形资产。因此，应充分利用客户的历史数据，定期对客户流失状况和回厂

时间进行分析与预测，并与实际的情况进行反复对比，进而不断改进对客户  
的分析能力，不断从历史数据中挖掘出其中蕴含的规律，更好掌握客户的情况，从而  
不断改变 4S 店的经营策略，助推 4S 店的发展。

## 参考文献

- [1] 王兵.一种基于逻辑回归模型的搜索广告点击率预估方法的研究[D].浙江大学,2013.
- [2] 钟超.扬州大长江 4S 店客户流失影响因素及应对研究[D].广西师范大学,2019.
- [3] 刘科娟.浅议汽车 4S 店客户流失原因及解决方案[J].才智, 32:327+329,2014.
- [4] 张宽.基于 TFM 模型的汽车 4S 店客户流失预警技术应用研究[D].兰州理工大学,2019.
- [5] 黄展辉.基于数据挖掘的电信客户流失分析[D].广东工业大学,2011.
- [6] 孙海.奥迪 4S 店客户流失原因分析及服务改进对策研究[D].吉林大学,2014.
- [7] 高昉.加油卡客户流失分析和预测——基于“跨行业数据挖掘标准流程”[J].国际石油经济, 27(10):99-105, 2019.
- [8] 罗静.汽车维修企业客户管理现状及改善建议[J].汽车维修与保养,(01):102-103, 2013.
- [9] 季广征.斯柯达 4S 店售后服务客户流失研究[D].山东大学,2016.
- [10] 崔秀敏.HG 一汽大众 4S 店的客户价值和客户细分研究[D].东北大学,2015.
- [11] 杨霄霄.逻辑回归和判别分析在财务危机预警模型中的应用[D].北京理工大学,2016.
- [12] 毛躍霖.汽车服务企业客户流失预测模型的分析与应用[D].东华大学,2015.
- [13] 孙琳.基于逻辑回归算法,搭建客户流失预测模型[J].信息系统工程, (06): 134, 2017.
- [14] 关欣,王征.基于 Logistic 回归和 BP 神经网络的财务预警模型比较[J].统计与决策, (17):179-181,2016.