

# 什么样的餐厅更受顾客欢迎

张晓宇      刘妍      陈锟      高伦      刘懋霖

## 目录

1 文件读取	1
2 数据探索	1
2.1 逐步回归法	1
2.2 回归树	6
2.3 随机森林回归	7
3 模型变量重要程度	8
4 最终变量选择	16

## 1 文件读取

```
data1<-read.csv("cating.csv")
```

该数据集是大众点评 2017 年广州餐厅的数据集，包含了 3124 条个案，18 个变量（如环境评分、服务评分等），这些变量是由客户的评价统计而得，我们将寻求其中好评率与哪些因素密切相关。

## 2 数据探索

### 2.1 逐步回归法

逐步回归就是从自变量  $x$  中挑选出对  $y$  有显著影响的变量，已达到最优。逐步回归分析是以 AIC 信息统计量为准则，通过选择最小的 AIC 信息

统计量，来达到删除或增加变量的目的。

```
a<-step(lm(ApplauseRate~.,data1))
```

```
## Start: AIC=13286.93
```

```
## ApplauseRate ~ ReviewNum + Level + FlavorScore + EnvironmentScore +
```

```
## ServiceScore + X5StarReviewNum + High.qualityMerchant + PopularArea +
```

```
## PictureNum + ParkingNum + ParkingInfo + GroupPurchase + Promotion +
```

```
## AdvanceReservation + TakeOut + PerConsumption + BusinessDay
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## - Promotion	1	7	217196	13285
## - AdvanceReservation	1	56	217245	13286
## - Level	1	70	217259	13286
## <none>			217189	13287
## - TakeOut	1	454	217643	13292
## - ParkingInfo	1	492	217681	13292
## - PopularArea	1	644	217833	13294
## - PictureNum	1	1219	218408	13302
## - EnvironmentScore	1	1360	218549	13304
## - FlavorScore	1	1451	218640	13306
## - PerConsumption	1	1696	218885	13309
## - ParkingNum	1	2718	219907	13324
## - BusinessDay	1	5062	222251	13357
## - ServiceScore	1	5314	222503	13360
## - GroupPurchase	1	5983	223172	13370
## - ReviewNum	1	8358	225547	13403
## - X5StarReviewNum	1	12202	229391	13456
## - High.qualityMerchant	1	96786	313975	14436

```
##
```

```
## Step: AIC=13285.02
```

```
## ApplauseRate ~ ReviewNum + Level + FlavorScore + EnvironmentScore +
```

```
## ServiceScore + X5StarReviewNum + High.qualityMerchant + PopularArea +
```

```
## PictureNum + ParkingNum + ParkingInfo + GroupPurchase + AdvanceReservation +
```

```
##      TakeOut + PerConsumption + BusinessDay
##
##              Df Sum of Sq    RSS    AIC
## - AdvanceReservation    1      53 217249 13284
## - Level                  1      70 217265 13284
## <none>                    217196 13285
## - TakeOut                1     452 217648 13290
## - ParkingInfo            1     493 217688 13290
## - PopularArea            1     647 217843 13292
## - PictureNum             1    1255 218450 13301
## - EnvironmentScore       1    1353 218549 13302
## - FlavorScore            1    1454 218650 13304
## - PerConsumption        1    1711 218907 13308
## - ParkingNum             1    2723 219919 13322
## - BusinessDay            1    5067 222263 13355
## - ServiceScore           1    5308 222504 13358
## - GroupPurchase          1    5980 223176 13368
## - ReviewNum              1    8366 225562 13401
## - X5StarReviewNum        1   12268 229464 13455
## - High.qualityMerchant   1   96846 314041 14435
##
## Step:  AIC=13283.79
## ApplauseRate ~ ReviewNum + Level + FlavorScore + EnvironmentScore +
##      ServiceScore + X5StarReviewNum + High.qualityMerchant + PopularArea +
##      PictureNum + ParkingNum + ParkingInfo + GroupPurchase + TakeOut +
##      PerConsumption + BusinessDay
##
##              Df Sum of Sq    RSS    AIC
## - Level                  1      67 217316 13283
## <none>                    217249 13284
## - TakeOut                1     455 217704 13288
## - ParkingInfo            1     486 217735 13289
## - PopularArea            1     637 217887 13291
```

```

## - PictureNum          1      1240 218489 13300
## - EnvironmentScore    1      1334 218583 13301
## - FlavorScore         1      1461 218710 13303
## - PerConsumption      1      1775 219024 13307
## - ParkingNum          1      2735 219984 13321
## - BusinessDay         1      5056 222305 13354
## - ServiceScore        1      5383 222632 13358
## - GroupPurchase       1      6257 223506 13370
## - ReviewNum           1      8410 225659 13400
## - X5StarReviewNum     1     12278 229527 13454
## - High.qualityMerchant 1     96794 314043 14433
##
## Step:  AIC=13282.76
## ApplauseRate ~ ReviewNum + FlavorScore + EnvironmentScore + ServiceScore +
##      X5StarReviewNum + High.qualityMerchant + PopularArea + PictureNum +
##      ParkingNum + ParkingInfo + GroupPurchase + TakeOut + PerConsumption +
##      BusinessDay
##
##              Df Sum of Sq    RSS    AIC
## <none>                217316 13283
## - TakeOut           1      450 217767 13287
## - ParkingInfo       1      495 217811 13288
## - PopularArea       1      640 217957 13290
## - PictureNum        1     1234 218551 13298
## - EnvironmentScore  1     1417 218734 13301
## - PerConsumption   1     1762 219078 13306
## - ParkingNum        1     2830 220147 13321
## - FlavorScore       1     3327 220644 13328
## - BusinessDay       1     4989 222305 13352
## - GroupPurchase     1     6259 223575 13370
## - ServiceScore      1     6320 223636 13370
## - ReviewNum         1     8409 225725 13399
## - X5StarReviewNum   1    12278 229595 13452

```

```
## - High.qualityMerchant 1 96998 314314 14434
```

根据逐步回归的分析结果，从中挑选出最佳拟合的变量，用其建立多元线性回归模型。

```
mylm<-lm(formula = ApplauseRate ~ ReviewNum + FlavorScore + EnvironmentScore + ServiceScore +
  X5StarReviewNum + High.qualityMerchant + PopularArea + PictureNum +
  ParkingNum + ParkingInfo + GroupPurchase + TakeOut + PerConsumption +
  BusinessDay, data = data1)
summary(mylm)
```

```
##
## Call:
## lm(formula = ApplauseRate ~ ReviewNum + FlavorScore + EnvironmentScore +
##     ServiceScore + X5StarReviewNum + High.qualityMerchant + PopularArea +
##     PictureNum + ParkingNum + ParkingInfo + GroupPurchase + TakeOut +
##     PerConsumption + BusinessDay, data = data1)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-25.447	-5.937	-0.002	5.255	34.482

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.667e+01	2.479e+00	-14.790	< 2e-16 ***
ReviewNum	-4.762e-03	4.341e-04	-10.968	< 2e-16 ***
FlavorScore	3.109e+00	4.507e-01	6.899	6.31e-12 ***
EnvironmentScore	1.793e+00	3.982e-01	4.503	6.95e-06 ***
ServiceScore	4.972e+00	5.229e-01	9.508	< 2e-16 ***
X5StarReviewNum	1.185e-02	8.938e-04	13.254	< 2e-16 ***
High.qualityMerchant	1.407e+01	3.777e-01	37.252	< 2e-16 ***
PopularArea	-9.507e-01	3.141e-01	-3.027	0.00249 **
PictureNum	-2.663e-03	6.336e-04	-4.202	2.72e-05 ***
ParkingNum	2.760e-02	4.337e-03	6.363	2.26e-10 ***
ParkingInfo	-1.040e+00	3.908e-01	-2.660	0.00784 **

```
## GroupPurchase      3.158e+00  3.337e-01   9.463 < 2e-16 ***
## TakeOut            8.474e-01  3.339e-01   2.538 0.01119 *
## PerConsumption     1.392e-02  2.772e-03   5.020 5.44e-07 ***
## BusinessDay        -1.876e-03  2.221e-04  -8.448 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.361 on 3109 degrees of freedom
## Multiple R-squared:  0.7402, Adjusted R-squared:  0.739
## F-statistic: 632.6 on 14 and 3109 DF,  p-value: < 2.2e-16
```

建立线性模型后，使用 `summary()` 函数查看模型，观察到所使用的变量都通过了显著性检验，且回归模型的 R square 值为 0.739，拟合效果较好。

## 2.2 回归树

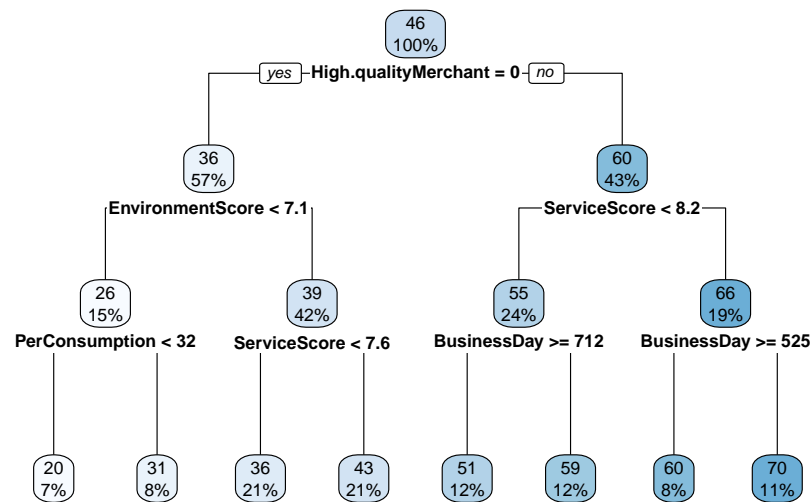
决策树 (Decision Tree) 是在已知各种情况发生概率的基础上，通过构成决策树来求取净现值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析方法，是直观运用概率分析的一种图解法。由于这种决策分支画成图形很像一棵树的枝干，故称决策树。在机器学习中，决策树是一个预测模型，他代表的是对象属性与对象值之间的一种映射关系。Entropy = 系统的凌乱程度，使用算法 ID3, C4.5 和 C5.0 生成树算法使用熵。决策树是一种树形结构，其中每个内部节点表示一个属性上的测试，每个分支代表一个测试输出，每个叶节点代表一种类别。

回归树是可以用于回归的决策树模型，一个回归树对应着输入空间（即特征空间）的一个划分以及在划分单元上的输出值。

```
library(rpart.plot)
```

```
## Loading required package: rpart
```

```
a2<-rpart(ApplauseRate~.,data1)
rpart.plot(a2)
```



```

SST<-sum((data1[,1]-mean(data1[,1]))^2)
resa<-data1[,1]-predict(a2,data1[,,-1])
SSEa<-sum(resa^2)
R2a<-1-SSEa/SST
R2a

```

```
## [1] 0.7167177
```

建立回归树，可以从树中看出被选作分类变量的变量。计算模型拟合的 R square 值，为 0.716，较多元回归模型效果较差。

## 2.3 随机森林回归

随机森林实际上是一种特殊的 bagging 方法，它将决策树用作 bagging 中的模型。首先，用 bootstrap 方法生成  $m$  个训练集，然后，对于每个训练集，构造一颗决策树，在节点找特征进行分裂的时候，并不是对所有特征找到能使得指标（如信息增益）最大的，而是在特征中随机抽取一部分特征，在抽到的特征中间找到最优解，应用于节点，进行分裂。随机森林的方法由于有了 bagging，也就是集成的思想在，实际上相当于对于样本和特征都进

行了采样（如果把训练数据看成矩阵，就像实际中常见的那样，那么就是一个行和列都进行采样的过程），所以可以避免过拟合。

```
library(randomForest)

## Warning: package 'randomForest' was built under R version 3.5.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

a3<-randomForest(ApplauseRate~.,data1,importance=T,localImp=T,proximity=T)
resa3<-data1[,1]-predict(a3,data1[, -1])
SSEa3<-sum(resa3^2)
R2a3<-1-SSEa3/SST
R2a3

## [1] 0.9646181
```

建立模型后通过计算随机森林预测的综合 R square 值，得到模型的 R square 值为 0.964，拟合效果很好。经过上述 3 次探索，最终选择随机森林回归作为最终的回归模型。

### 3 模型变量重要程度

```
a3$importance
```

##	%IncMSE	IncNodePurity
## ReviewNum	18.2523091	26916.433
## Level	13.6250544	52255.296
## FlavorScore	13.4845302	50047.640
## EnvironmentScore	16.5480091	82790.449
## ServiceScore	38.4674369	123444.313
## X5StarReviewNum	41.3442479	43792.783
## High.qualityMerchant	142.6642919	273413.334



## PopularArea	0.8623767	4311.790
## PictureNum	13.6294824	26633.992
## ParkingNum	6.3900445	12261.009
## ParkingInfo	2.5541670	4522.214
## GroupPurchase	4.1441949	10979.544
## Promotion	0.2596281	1637.185
## AdvanceReservation	0.1627183	1599.006
## TakeOut	0.8124798	3393.124
## PerConsumption	12.6577796	47081.831
## BusinessDay	17.4650825	56195.133

a3\$rsq

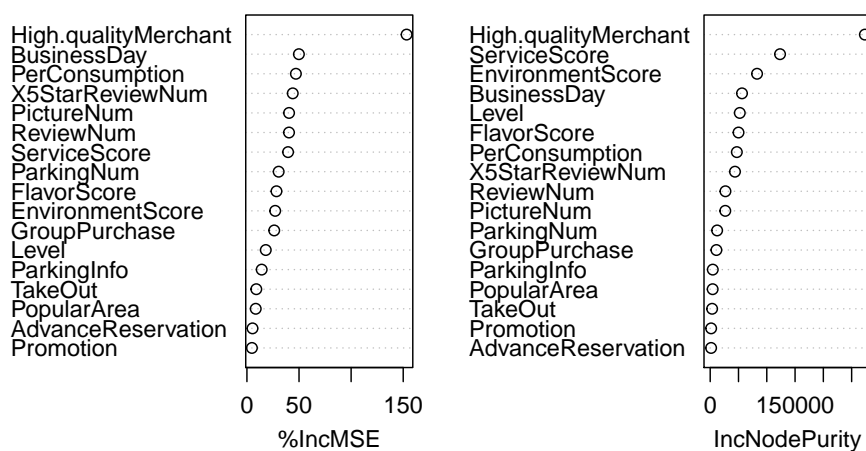
```
## [1] 0.6217150 0.6387206 0.6516186 0.6698428 0.6960535 0.7169715 0.7270773
## [8] 0.7403223 0.7520462 0.7576240 0.7642096 0.7707436 0.7736647 0.7766596
## [15] 0.7812001 0.7839411 0.7856935 0.7880215 0.7911953 0.7919577 0.7935831
## [22] 0.7955317 0.7985251 0.8005896 0.8027463 0.8044810 0.8049130 0.8050456
## [29] 0.8056953 0.8063089 0.8069559 0.8078719 0.8092532 0.8092128 0.8102809
## [36] 0.8101728 0.8104804 0.8105394 0.8109397 0.8111702 0.8115451 0.8116394
## [43] 0.8121922 0.8128238 0.8129670 0.8132020 0.8134697 0.8138380 0.8144684
## [50] 0.8139552 0.8146803 0.8150593 0.8149385 0.8149209 0.8154443 0.8151818
## [57] 0.8153266 0.8155131 0.8156644 0.8161085 0.8162454 0.8164654 0.8166257
## [64] 0.8164988 0.8167121 0.8166023 0.8167087 0.8163150 0.8162629 0.8163671
## [71] 0.8163306 0.8164440 0.8163696 0.8164987 0.8168522 0.8166435 0.8166556
## [78] 0.8170138 0.8174259 0.8173977 0.8171352 0.8174096 0.8174656 0.8176777
## [85] 0.8175979 0.8175644 0.8173557 0.8173180 0.8172158 0.8171811 0.8173322
## [92] 0.8173708 0.8175834 0.8176419 0.8176698 0.8176066 0.8178053 0.8181073
## [99] 0.8183182 0.8184585 0.8185675 0.8183426 0.8182700 0.8182854 0.8183013
## [106] 0.8183369 0.8182295 0.8182600 0.8182627 0.8184307 0.8187542 0.8187130
## [113] 0.8186267 0.8185737 0.8186331 0.8185041 0.8184784 0.8182468 0.8183044
## [120] 0.8184185 0.8185650 0.8186760 0.8187310 0.8189314 0.8187711 0.8186516
## [127] 0.8187667 0.8188246 0.8188700 0.8188571 0.8189480 0.8190529 0.8191585
## [134] 0.8191878 0.8194139 0.8194422 0.8194171 0.8192686 0.8193316 0.8193218
## [141] 0.8194073 0.8195375 0.8193951 0.8193524 0.8195279 0.8195735 0.8196342
```

```
## [148] 0.8198181 0.8198766 0.8199333 0.8199330 0.8200164 0.8199497 0.8201729
## [155] 0.8199180 0.8198799 0.8198645 0.8199836 0.8201350 0.8200923 0.8202322
## [162] 0.8203332 0.8203237 0.8202893 0.8203437 0.8204243 0.8205455 0.8205604
## [169] 0.8205767 0.8205830 0.8204631 0.8204895 0.8202981 0.8202119 0.8201244
## [176] 0.8200928 0.8201491 0.8203066 0.8204729 0.8206205 0.8206116 0.8206135
## [183] 0.8205538 0.8206831 0.8207219 0.8206394 0.8206719 0.8205972 0.8206534
## [190] 0.8208450 0.8209768 0.8209579 0.8208889 0.8209479 0.8209612 0.8210592
## [197] 0.8209476 0.8210476 0.8210671 0.8210482 0.8210874 0.8211266 0.8210806
## [204] 0.8209805 0.8209912 0.8210565 0.8211185 0.8210001 0.8210492 0.8211179
## [211] 0.8211398 0.8211814 0.8212240 0.8213229 0.8213542 0.8212678 0.8212656
## [218] 0.8213579 0.8213693 0.8213073 0.8212271 0.8212177 0.8211959 0.8211492
## [225] 0.8210268 0.8210675 0.8212547 0.8211867 0.8212590 0.8212873 0.8212897
## [232] 0.8213128 0.8213756 0.8213982 0.8213997 0.8214218 0.8214008 0.8214108
## [239] 0.8214118 0.8214021 0.8213821 0.8213009 0.8212811 0.8212649 0.8213257
## [246] 0.8213604 0.8213320 0.8211903 0.8211921 0.8211338 0.8211595 0.8211758
## [253] 0.8210957 0.8211092 0.8210838 0.8209215 0.8209005 0.8208683 0.8209429
## [260] 0.8208731 0.8207876 0.8208407 0.8207671 0.8207326 0.8206988 0.8207427
## [267] 0.8206330 0.8206389 0.8206383 0.8206399 0.8205399 0.8206293 0.8206438
## [274] 0.8205464 0.8204792 0.8205436 0.8205070 0.8205160 0.8204932 0.8205901
## [281] 0.8207275 0.8207667 0.8207868 0.8207307 0.8206598 0.8206981 0.8207143
## [288] 0.8208197 0.8208843 0.8208788 0.8208355 0.8209566 0.8208849 0.8208403
## [295] 0.8208880 0.8208543 0.8208433 0.8208702 0.8208253 0.8209123 0.8209085
## [302] 0.8209170 0.8209117 0.8208889 0.8209841 0.8210152 0.8209923 0.8209981
## [309] 0.8209872 0.8210622 0.8210859 0.8210120 0.8210498 0.8210956 0.8210825
## [316] 0.8211179 0.8210957 0.8211568 0.8211720 0.8211982 0.8212366 0.8212214
## [323] 0.8212437 0.8212278 0.8212861 0.8212847 0.8213807 0.8213906 0.8214150
## [330] 0.8214364 0.8215105 0.8214972 0.8215455 0.8215065 0.8215166 0.8214624
## [337] 0.8214178 0.8214191 0.8213534 0.8213437 0.8214310 0.8214649 0.8214753
## [344] 0.8214993 0.8214772 0.8214357 0.8215124 0.8214944 0.8215138 0.8215452
## [351] 0.8214622 0.8214415 0.8214864 0.8214081 0.8214147 0.8214492 0.8214163
## [358] 0.8214594 0.8215441 0.8215195 0.8215846 0.8216615 0.8216542 0.8216555
## [365] 0.8217341 0.8216873 0.8216418 0.8216073 0.8215830 0.8215225 0.8214585
## [372] 0.8215231 0.8215368 0.8215552 0.8216358 0.8216830 0.8217137 0.8217610
```

```
## [379] 0.8217524 0.8217294 0.8216874 0.8216950 0.8216856 0.8216795 0.8216467
## [386] 0.8216620 0.8216847 0.8216376 0.8216920 0.8217215 0.8217205 0.8216809
## [393] 0.8217177 0.8217970 0.8217642 0.8217976 0.8217960 0.8217802 0.8217667
## [400] 0.8217902 0.8217944 0.8218491 0.8218674 0.8219655 0.8219170 0.8219152
## [407] 0.8220163 0.8219594 0.8219963 0.8219962 0.8219904 0.8219989 0.8219308
## [414] 0.8218974 0.8219164 0.8218987 0.8219616 0.8219036 0.8220109 0.8220669
## [421] 0.8220972 0.8221033 0.8221510 0.8221693 0.8221796 0.8220919 0.8221090
## [428] 0.8221203 0.8221188 0.8220643 0.8220768 0.8220705 0.8220962 0.8221262
## [435] 0.8221339 0.8220667 0.8220641 0.8220488 0.8220175 0.8220197 0.8220052
## [442] 0.8219792 0.8219914 0.8219743 0.8219815 0.8220084 0.8220081 0.8220017
## [449] 0.8219960 0.8219386 0.8219495 0.8218869 0.8218560 0.8218433 0.8218267
## [456] 0.8218379 0.8218212 0.8217625 0.8216727 0.8216533 0.8216511 0.8216859
## [463] 0.8216773 0.8216552 0.8216093 0.8215850 0.8215555 0.8215365 0.8215440
## [470] 0.8215568 0.8215901 0.8215859 0.8215415 0.8215537 0.8215371 0.8215056
## [477] 0.8215280 0.8215641 0.8215345 0.8215853 0.8215750 0.8215852 0.8216545
## [484] 0.8216725 0.8216554 0.8216341 0.8216089 0.8216261 0.8216125 0.8216298
## [491] 0.8216460 0.8216574 0.8216357 0.8216880 0.8217036 0.8216570 0.8216375
## [498] 0.8215691 0.8216225 0.8216272
```

```
varImpPlot(a3)
```

a3



使用代码查看变量的重要程度，发现其中影响最大的变量其影响力远高于影响第二的变量，判断其与因变量之间存在某种决定性关系，需将其剔除后再次建模。

```
data2<-data1[,-8]
a4<-randomForest(ApplauseRate~.,data2,importance=T,localImp=T,proximity=T)
resa4<-data2[,1]-predict(a4,data2[, -1])
SSEa4<-sum(resa4^2)
R2a4<-1-SSEa4/SST
a4$importance
```

##	%IncMSE	IncNodePurity
## ReviewNum	45.9883332	48036.963
## Level	26.4735133	77684.509
## FlavorScore	28.6597210	77837.875
## EnvironmentScore	26.1623585	104943.334
## ServiceScore	89.4515246	187674.106
## X5StarReviewNum	106.8735921	79144.259
## PopularArea	4.7366018	10370.629

## PictureNum	35.1722878	51107.988
## ParkingNum	12.2605524	17571.396
## ParkingInfo	4.9968456	6565.920
## GroupPurchase	6.0140582	16032.189
## Promotion	0.9087477	2860.533
## AdvanceReservation	0.1881044	2321.392
## TakeOut	1.7670755	5665.737
## PerConsumption	8.4639955	43605.618
## BusinessDay	32.6146192	79996.803

```
a4$rsq
```

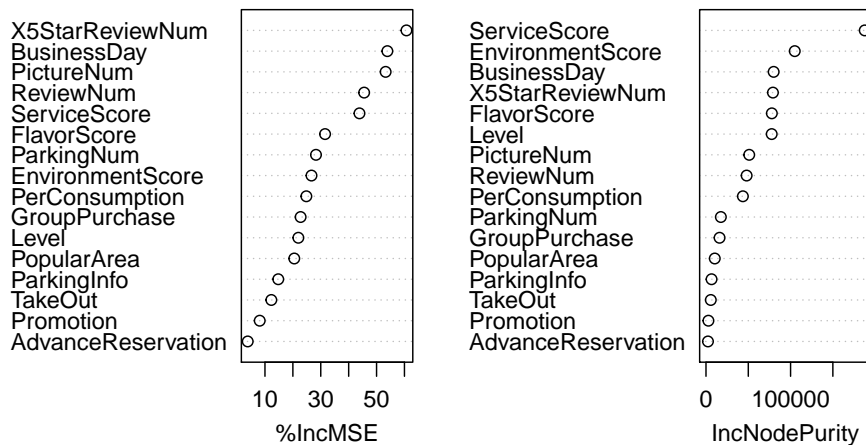
```
## [1] 0.2418995 0.3418204 0.3841936 0.4467788 0.4947428 0.5210016 0.5477059
## [8] 0.5671990 0.5860291 0.5992010 0.6061610 0.6184885 0.6283420 0.6326904
## [15] 0.6391743 0.6484135 0.6536247 0.6582109 0.6593403 0.6617233 0.6648890
## [22] 0.6678617 0.6708006 0.6736352 0.6750470 0.6776268 0.6802868 0.6816714
## [29] 0.6823434 0.6833071 0.6848207 0.6867729 0.6874787 0.6890139 0.6891001
## [36] 0.6909011 0.6905263 0.6912277 0.6946902 0.6948374 0.6955722 0.6976505
## [43] 0.6992322 0.6989672 0.6999041 0.6994257 0.7009864 0.7015027 0.7007374
## [50] 0.7011781 0.7008038 0.7013724 0.7024083 0.7028510 0.7031540 0.7035008
## [57] 0.7043632 0.7053829 0.7052446 0.7052867 0.7058931 0.7058918 0.7065185
## [64] 0.7072384 0.7082954 0.7085288 0.7086878 0.7091605 0.7091251 0.7094111
## [71] 0.7091570 0.7087782 0.7091372 0.7090483 0.7089034 0.7086091 0.7087762
## [78] 0.7084443 0.7091189 0.7093791 0.7091954 0.7094992 0.7091686 0.7092628
## [85] 0.7085248 0.7090621 0.7096863 0.7094571 0.7097548 0.7101141 0.7096834
## [92] 0.7097797 0.7103163 0.7103985 0.7105485 0.7103501 0.7105988 0.7111352
## [99] 0.7112959 0.7111081 0.7109539 0.7106732 0.7109115 0.7106738 0.7109263
## [106] 0.7113944 0.7114976 0.7115374 0.7115393 0.7113928 0.7111144 0.7114799
## [113] 0.7113954 0.7111587 0.7112987 0.7113300 0.7114055 0.7120235 0.7121216
## [120] 0.7123463 0.7122112 0.7116807 0.7122907 0.7128987 0.7126280 0.7126298
## [127] 0.7129700 0.7125431 0.7126566 0.7124556 0.7126497 0.7129261 0.7131531
## [134] 0.7130192 0.7129320 0.7132605 0.7132739 0.7131360 0.7132979 0.7134326
## [141] 0.7133365 0.7131974 0.7129297 0.7127778 0.7126900 0.7130562 0.7131714
## [148] 0.7127724 0.7127040 0.7126395 0.7125225 0.7122479 0.7122389 0.7122624
```

```
## [155] 0.7119901 0.7123214 0.7121217 0.7119935 0.7123426 0.7124702 0.7129152
## [162] 0.7130849 0.7133255 0.7134617 0.7133477 0.7136310 0.7136068 0.7139806
## [169] 0.7140186 0.7142501 0.7142320 0.7141801 0.7141386 0.7141157 0.7140906
## [176] 0.7140128 0.7142760 0.7143044 0.7144083 0.7142823 0.7143350 0.7142203
## [183] 0.7147646 0.7147320 0.7148321 0.7146673 0.7144531 0.7144879 0.7145706
## [190] 0.7147232 0.7148159 0.7148987 0.7149764 0.7150257 0.7151987 0.7155159
## [197] 0.7154560 0.7159827 0.7160308 0.7158923 0.7158668 0.7160500 0.7161723
## [204] 0.7161859 0.7160963 0.7162610 0.7163097 0.7163542 0.7166984 0.7167864
## [211] 0.7168964 0.7169004 0.7170972 0.7170683 0.7171382 0.7170351 0.7172506
## [218] 0.7173341 0.7171916 0.7170859 0.7171373 0.7171523 0.7173345 0.7172423
## [225] 0.7175043 0.7175791 0.7176243 0.7175685 0.7176281 0.7178472 0.7179249
## [232] 0.7179473 0.7179407 0.7179051 0.7181601 0.7182540 0.7180201 0.7180341
## [239] 0.7180457 0.7180962 0.7179881 0.7180692 0.7181740 0.7179165 0.7178634
## [246] 0.7178664 0.7178728 0.7177957 0.7180335 0.7181474 0.7181730 0.7182594
## [253] 0.7184400 0.7184303 0.7184629 0.7185505 0.7185404 0.7183859 0.7184941
## [260] 0.7183062 0.7181818 0.7182629 0.7184249 0.7187572 0.7188711 0.7190182
## [267] 0.7189879 0.7189149 0.7186867 0.7185200 0.7183863 0.7184523 0.7182344
## [274] 0.7181524 0.7181801 0.7180926 0.7181273 0.7182917 0.7181865 0.7180878
## [281] 0.7181761 0.7182034 0.7182525 0.7183181 0.7182797 0.7181528 0.7182177
## [288] 0.7183424 0.7183615 0.7185860 0.7185947 0.7185077 0.7182583 0.7184924
## [295] 0.7185106 0.7184903 0.7183614 0.7183427 0.7183458 0.7184346 0.7183408
## [302] 0.7182684 0.7183900 0.7182666 0.7182189 0.7182497 0.7182247 0.7182570
## [309] 0.7182508 0.7183422 0.7183927 0.7184245 0.7184692 0.7184838 0.7183914
## [316] 0.7183399 0.7184170 0.7183774 0.7185125 0.7185894 0.7186922 0.7186158
## [323] 0.7186991 0.7187042 0.7188567 0.7188295 0.7187481 0.7186663 0.7186772
## [330] 0.7186613 0.7184973 0.7184802 0.7185806 0.7186998 0.7186848 0.7188265
## [337] 0.7188921 0.7188698 0.7188938 0.7188662 0.7189605 0.7189821 0.7189129
## [344] 0.7189147 0.7188943 0.7188245 0.7188434 0.7188537 0.7189182 0.7188779
## [351] 0.7187980 0.7188158 0.7186372 0.7185561 0.7184921 0.7184583 0.7184103
## [358] 0.7183421 0.7182970 0.7182174 0.7181671 0.7181592 0.7181776 0.7181794
## [365] 0.7181569 0.7181328 0.7180008 0.7179924 0.7180582 0.7181142 0.7180270
## [372] 0.7180004 0.7179698 0.7178832 0.7177720 0.7177770 0.7176838 0.7176449
## [379] 0.7175656 0.7174989 0.7174891 0.7175619 0.7175434 0.7174405 0.7174507
```

```
## [386] 0.7174682 0.7176078 0.7177438 0.7177446 0.7176939 0.7177370 0.7178063
## [393] 0.7177640 0.7177757 0.7177330 0.7177374 0.7177732 0.7177752 0.7181398
## [400] 0.7180507 0.7181754 0.7181719 0.7182086 0.7182728 0.7182344 0.7182877
## [407] 0.7182253 0.7181735 0.7181211 0.7180753 0.7181806 0.7182028 0.7182553
## [414] 0.7181972 0.7181616 0.7180950 0.7182074 0.7181223 0.7182118 0.7182195
## [421] 0.7181324 0.7181335 0.7180557 0.7180196 0.7181326 0.7181918 0.7181609
## [428] 0.7180683 0.7181574 0.7181030 0.7179881 0.7180399 0.7181144 0.7179545
## [435] 0.7181392 0.7180082 0.7181777 0.7181328 0.7180820 0.7180772 0.7180856
## [442] 0.7180993 0.7180857 0.7179795 0.7179267 0.7179507 0.7179440 0.7179840
## [449] 0.7179554 0.7179312 0.7178890 0.7179027 0.7178438 0.7178804 0.7179045
## [456] 0.7178772 0.7179270 0.7180538 0.7181015 0.7181438 0.7182862 0.7181886
## [463] 0.7181747 0.7182344 0.7183326 0.7182259 0.7181440 0.7181145 0.7180906
## [470] 0.7180696 0.7181504 0.7181368 0.7181578 0.7182227 0.7182505 0.7182423
## [477] 0.7182105 0.7182106 0.7181589 0.7181524 0.7181739 0.7181031 0.7179650
## [484] 0.7178621 0.7178252 0.7179471 0.7178764 0.7178989 0.7178015 0.7178376
## [491] 0.7178711 0.7178305 0.7177901 0.7177316 0.7177710 0.7177244 0.7177646
## [498] 0.7178743 0.7178015 0.7177694
```

```
varImpPlot(a4)
```

a4



再次建模，通过代码以及图片输出查看各变量对模型的重要程度。其中两个重要指标解释：

1. 置换精度: 将某个变量去除以后，如果模型精确度大大下降，就说明该变量对于模型的影响较大

2. 结点纯度: 通过基尼指数算的结点不纯度，某个变量对于拆分结点不纯度降低的重要性

3. 节点不纯度: 集合中某个结果随机应用于某一数据项的期望误差率绘制局部重要图，反映了每个变量在每颗树中的影响力大小。

之后绘制模型对于各变量的边缘依赖图，表示将其他变量的影响求和加总后单个变量对模型的影响程度，反映了变量在怎样一个取值范围内对模型作出影响。

```
layout(matrix(c(1,2,3,3),nrow = 2,b=T))
for(i in 1:2){
  title(colnames(a4$importance)[i])
}
matplot(2:17,a4$local,type = "l",xlab = "Variable",ylab = "Local importance",main="Local importance")

par(mfrow=c(4,4))
partialPlot(a4,pred.data = data2,ReviewNum);partialPlot(a4,pred.data = data2,Level)
partialPlot(a4,pred.data = data2,FlavorScore);partialPlot(a4,pred.data = data2,Environment)
partialPlot(a4,pred.data = data2,ServiceScore);partialPlot(a4,pred.data = data2,X5StarRating)
partialPlot(a4,pred.data = data2,PopularArea);partialPlot(a4,pred.data = data2,PictureNumber)
partialPlot(a4,pred.data = data2,ParkingNum);partialPlot(a4,pred.data = data2,ParkingInconvenience)
partialPlot(a4,pred.data = data2,GroupPurchase);partialPlot(a4,pred.data = data2,Promotion)
partialPlot(a4,pred.data = data2,AdvanceReservation);partialPlot(a4,pred.data = data2,TravelTime)
partialPlot(a4,pred.data = data2,PerConsumption);partialPlot(a4,pred.data = data2,BusinessType)
```

## 4 最终变量选择

通过上述判断变量重要程度以及影响范围的指标，对变量进行重要程度排序，选取排名前四的变量作为最能影响餐厅受欢迎程度的变量。选出的影响因素最大的四个变量分别是：



ServiceScore;  
flavourScore;  
businessday;  
X5starReviewnum.