PUBLISHED RALATIVE PERFORMANCE PREDICTION

1) Initial Variables

Vendor Name

Model Name

MCYT

MMIN

MMAX

CACH

CHMIN

CHMAX

PRP

ERP

2) Pre-processing

1) Named the columns of given data

2) Nullified the unrequired feature variables like ‘Vendor Name’ and ‘Model Name’ and ‘ERP’

3) Checked for the NULL values

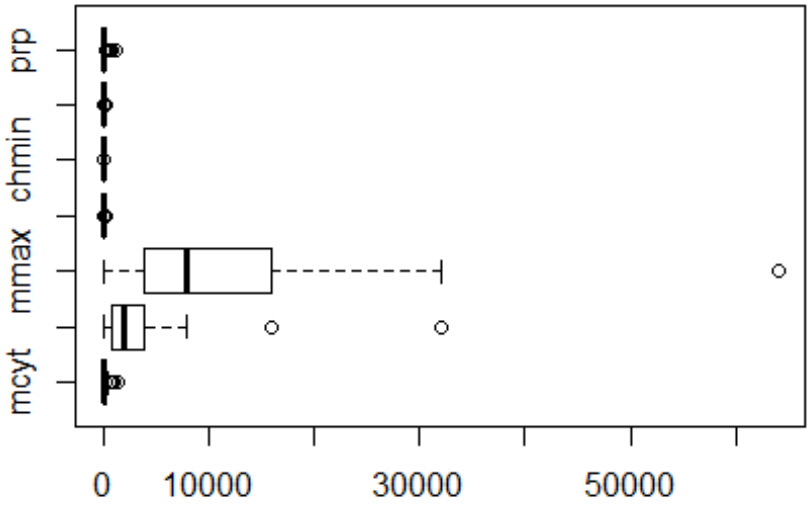
4) Checked the correlation between the feature variables

```
##            mcyt       mmin       mmax       cach      chmin      chmax
## mcyt   1.0000000 -0.3370714 -0.3795919 -0.3404142 -0.3007337 -0.2556289
## mmin  -0.3370714  1.0000000  0.7578268  0.6027875  0.5266649  0.2938772
## mmax  -0.3795919  0.7578268  1.0000000  0.6006801  0.5685938  0.5623875
## cach  -0.3404142  0.6027875  0.6006801  1.0000000  0.5881277  0.4235497
## chmin -0.3007337  0.5266649  0.5685938  0.5881277  1.0000000  0.5417623
## chmax -0.2556289  0.2938772  0.5623875  0.4235497  0.5417623  1.0000000
## prp   -0.3065714  0.7983105  0.8655762  0.7046424  0.6088411  0.6213091
##            prp
## mcyt  -0.3065714
## mmin   0.7983105
## mmax   0.8655762
## cach   0.7046424
## chmin  0.6088411
```

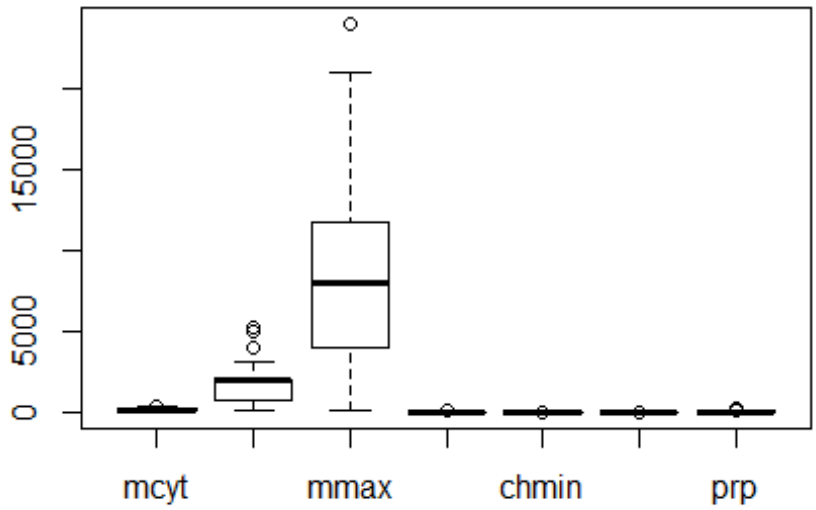
```
## chmax  0.6213091
## prp    1.0000000
```

5) Checked the outliers of the data and replaced it with mean of that respective column

-Before outlier replacement



-After outlier replacement



3) Feture Selection

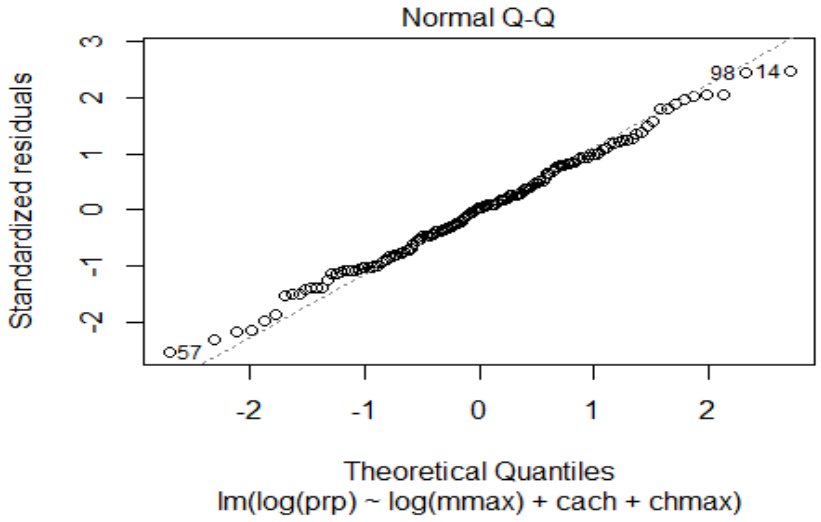
1) First of all feature variables are taken into consideration
2) And then through analysis and relevance, scaling is done and feature variables are selected

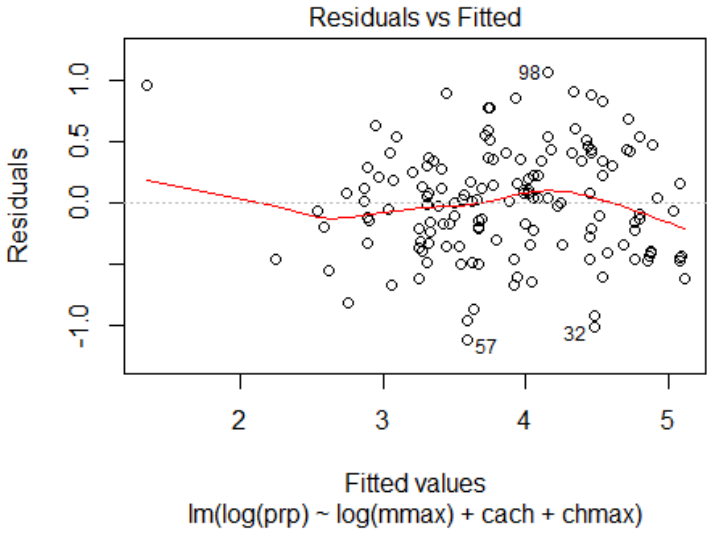
4) Model Development

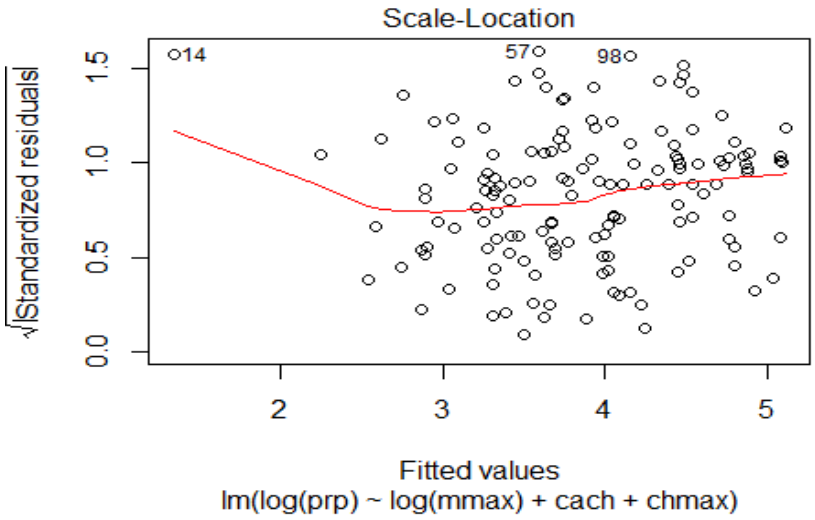
1) Linear algorithm is used for model development

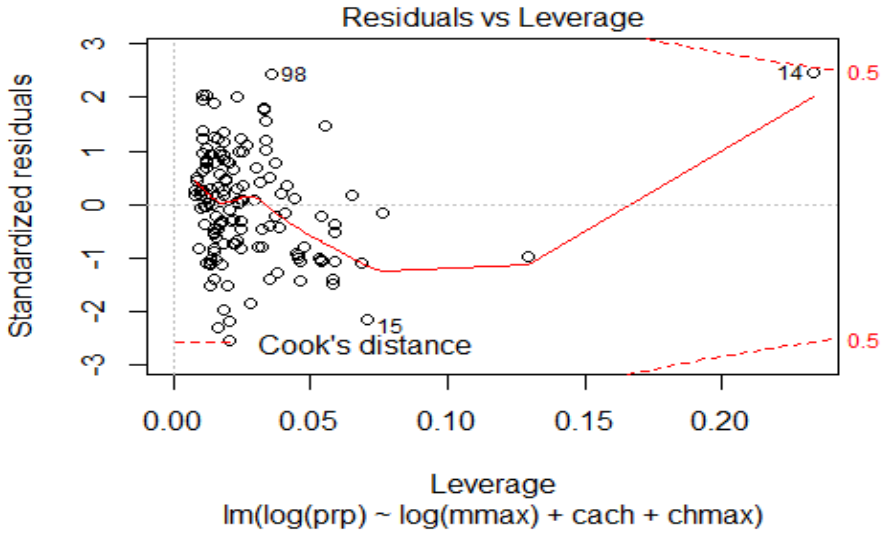
5) Model Testing and Accuracy Checking

1) Linearity between data and model is checked









2) Analysis of variance table is also done

```
## Response: log(prp)
##             Df Sum Sq Mean Sq F value    Pr(>F)
## log(mmax)   1 54.100  54.100 274.853 < 2.2e-16 ***
## cach        1 10.290  10.290  52.280 2.731e-11 ***
## chmax       1  4.448   4.448  22.599 4.842e-06 ***
## Residuals 142 27.950   0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3) RMSE is checked for developed model

```
## [1] 0.4407441
```

6) Codes

```
setwd("E:/Aegis/Machine Learning/Data")

cpu11=read.csv("data set.csv",stringsAsFactors = TRUE)


colnames(cpu11)=c("vendor","model
name","mcyt","mmin","mmax","cach","chmin","chmax","prp","erp")


cpu11$vendor=NULL

cpu11$`model name`=NULL

cpu11$erp=NULL


anyNA(cpu11)


cor(cpu11)


boxplot(cpu11,horizontal = TRUE)


outlier.removal.with.mean=function(a){
 for(j in 1:ncol(a)){
   v=summary(a[,j])
   rightlimit=v[5]+1.5*IQR(v)
   leftlimit=v[2]-1.5*IQR(v)


   for(i in 1:nrow(a)){
     if(a[i,j]>rightlimit || a[i,j]<leftlimit){
       a[i,j]=mean(a[,j])
     }
```

```
   }
 }
 return(a)
}


cpu11=outlier.removal.with.mean(cpu11)
boxplot(cpu11)



indi=sample(1:nrow(cpu11),round(0.70*(nrow(cpu11))))
cpu11train=cpu11[indi,]
cpu11test=cpu11[-indi,]


#fitt14=lm(formula = prp ~ mcyt + mmin + mmax + cach + chmin + chmax ,
data = cpu11train)
#summary(fitt11)
#plot(fitt11)


#fitt14=lm(log(prp) ~ mcyt + mmin + mmax + cach + chmin +
chmax,cpu11train)#0.4865599
#summary(fitt12)
#plot(fitt12)



#fitt14=lm(log(prp)~ log(mcyt) + log(mmin) + log(mmax) + cach + chmin
,cpu11train)
#summary(fitt13)
#plot(fitt13)
```

```
fitt14=lm(formula = log(prp) ~   log(mmax) + cach + chmax, data =
cpu11train)

summary(fitt14)

plot(fitt14)

anova(fitt14)


out14=predict(fitt14,cpu11test)


error=out14-log(cpu11test[7])


RMSE14=sqrt(mean((error)^2))

RMSE14
```