Logistic Regression Model

The data from the MySql database was made longitudinal again and was exported as an CSV file. This file as sourced to the R.

A logistic Regression model was built over this dataset to predict whether the thr A Route would be On-Time, or Delayed or Early.

As described earlier, if the actual time is a minute or earlier than the schedule time, it is marked Early. If the bus arrives a minute before or after the scheduled time, then it is termed as On-Time. If the bus is delayed by more than a minute, it is termed as late.

In essense, On time : one min before and after the scheduled time Early: Earlier than a minute to the scheduled time Late: Later than a minute to the scheduled time

Thus, the logistic model would have 3 levels, On-Time, Early and Delayed.

Model 1 The model was built in such a way that the prediction in first stop would dependant on the tie of the hour and the trip of the day. **Model 2** The second model was built in such a way that this would consider the status of arrival in the first stop, hour and the day of travel. **Iterative Models** The consecutive models for stops iteratively built on this process, all constant factors and the previous stop status

```
setwd("C:\\Users\\Ganesh\\Google Drive\\Courses\\CSCI B 565\\Bus Project\\nefarious-octo-rutabaga\\R Mo
library(nnet)
routea<-read.csv("DATA.csv")
# Stop 1 Model
stop1.fit<-multinom(Stop.1.Status~Day+Hr.Day,data=routea)
```

```
## # weights:  24 (15 variable)
## initial  value 2614.551165
## iter  10 value 1136.297534
## iter  20 value 1087.220212
## final  value 1087.212606
## converged
```

```
# Predict Stop 1 Status
stop1.prob<-data.frame(Day=c("M"),Hr.Day=10)
predict(stop1.fit,newdata = stop1.prob,"probs")
```

```
## Data Missing      Delayed        Early      On Time
##   0.03082123   0.77910313   0.07317609   0.11689955
```

```
#Stop 2 status
stop2.fit<-multinom(Stop.2.Status~Day+Hr.Day+Stop.1.Status,data=routea)
```

```
## # weights:  36 (24 variable)
## initial  value 2614.551165
## iter  10 value 2117.757741
## iter  20 value 2011.097163
## iter  30 value 1998.078206
## iter  40 value 1997.885031
## final  value 1997.884320
## converged
```

```r
#Prediction for Stop 2 status
stop2.prob<-data.frame(Day=c("M"),Hr.Day=10,Stop.1.Status=c("Delayed"))
predict(stop2.fit,newdata = stop2.prob,"probs")
```

```
## Data Missing      Delayed        Early      On Time
##   0.01664346   0.14224930   0.36711064   0.47399660
```

```r
#Stop 3 Status
stop3.fit<-multinom(Stop.3.Status~Day+Hr.Day+Stop.1.Status+Stop.2.Status,data=routea)
```

```
## # weights:  48 (33 variable)
## initial  value 2614.551165
## iter  10 value 861.842048
## iter  20 value 708.720592
## iter  30 value 685.807742
## iter  40 value 684.351992
## iter  50 value 684.315511
## final  value 684.313809
## converged
```

```r
#Prediction for Stop 3
stop3.prob<-data.frame(Day=c("M"),Hr.Day=10,Stop.1.Status=c("Delayed"),Stop.2.Status=c("On Time"))
predict(stop3.fit,newdata = stop3.prob,"probs")
```

```
## Data Missing      Delayed        Early      On Time
## 3.381334e-03 7.365777e-11 7.682267e-01 2.283920e-01
```

```r
# Stop 4 Model
stop4.fit<-multinom(Stop.4.Status~Day+Hr.Day+Stop.1.Status+Stop.2.Status+Stop.3.Status,data=routea)
```

```
## # weights:  60 (42 variable)
## initial  value 2614.551165
## iter  10 value 1261.966796
## iter  20 value 1136.788612
## iter  30 value 1099.663056
## iter  40 value 1096.691497
## iter  50 value 1096.544668
## iter  60 value 1096.531568
## final  value 1096.531328
## converged
```

```r
#Prediction
stop4.prob<-data.frame(Day=c("M"),Hr.Day=10,Stop.1.Status=c("Delayed"),Stop.2.Status=c("On Time"),Stop.3
predict(stop4.fit,newdata = stop4.prob,"probs")
```

```
## Data Missing      Delayed        Early      On Time
## 2.966553e-18 9.588480e-01 8.424239e-12 4.115202e-02
```

A keen observer can see that this prediction is as same as the results off Tableau Graphs. Even there, the pattern that was observed is that if the bus starts delayed,