**ind <-sample(2,nrow(train\_transaction\_v2), replace = T, prob = c(0.001, 0.999))**

> tab3 <- table(predicted = p3, Actual = train$is\_churn)

> tab3

Actual

predicted 0 1

0 708 48

1 11 29

> (1-sum(diag(tab3))/sum(tab3)) \* 100

[1] 7.41206

> (sum(diag(tab3))/sum(tab3)) \* 100

[1] 92.58794

> p4 <- predict(lrmodel, test, type = "response")

> p4 <- ifelse(p4>0.5,1,0)

> tab4 <- table(predicted = p4, Actual = test$is\_churn)

> tab4

Actual

predicted 0 1

0 317 23

1 7 9

> (1-sum(diag(tab4))/sum(tab4)) \* 100

[1] 8.426966

> (sum(diag(tab4))/sum(tab4)) \* 100

[1] 91.57303

**ind <-sample(2,nrow(train\_transaction\_v2), replace = T, prob = c(0.005, 0.995))**

> tab3 <- table(predicted = p3, Actual = train$is\_churn)

> tab3

Actual

predicted 0 1

0 3502 232

1 42 137

> (1-sum(diag(tab3))/sum(tab3)) \* 100

[1] 7.0023

> (sum(diag(tab3))/sum(tab3)) \* 100

[1] 92.9977

> p4 <- predict(lrmodel, test, type = "response")

> p4 <- ifelse(p4>0.5,1,0)

> tab4 <- table(predicted = p4, Actual = test$is\_churn)

> tab4

Actual

predicted 0 1

0 1538 102

1 20 56

> (1-sum(diag(tab4))/sum(tab4)) \* 100

[1] 7.109557

> (sum(diag(tab4))/sum(tab4)) \* 100

[1] 92.89044

# Presentation:

## How many people churn and not churn

0 1

1018463 113573

## Logistic Regression

Data Split: 70-30

> p3 <- predict(lrmodel, train, type = "response")

> p3 <- ifelse(p3>0.5,1,0)

c

**[6:02 PM, 12/6/2017] Abhishek UMASS: decision tree**

[6:02 PM, 12/6/2017] Abhishek UMASS: (tab1 <- table(predicted = p1, Actual = train$is\_churn))

Actual

predicted 0 1

0 705756 30622

1 6745 48901

> 1 - sum(diag(tab1))/sum(tab1)

[1] 0.04717912589

> sum(diag(tab1))/sum(tab1)

[1] 0.9528208741

> p2 <- predict(tree, test, type = 'response')

> (tab2 <- table(predicted = p2, Actual = test$is\_churn))

Actual

predicted 0 1

0 302914 13259

1 3048 20791

> 1-sum(diag(tab2))/sum(tab2)

[1] 0.047960072

> sum(diag(tab2))/sum(tab2)

[1] 0.952039928

**Neural Network**

> tab5

Actual

predicted 0.00000102990854412128 0.00000205981708824256

0 713228 79398

> table(train\_transaction\_v2$is\_cancel)

0 1

1099153 32883

> table(train\_transaction\_v2$is\_churn)

0 1

1018463 113573

**People cancel =1**

> table(df$is\_churn)

0 1

13293 19590

Total: 32883 == > 59%

**Cancel = 0**

> table(df1$is\_churn)

0 1

1005170 93983

Total: 1099153 🡺 8.5%