DeSimone\_MIS64060\_Final

Heather DeSimone

4/30/2022

##Loading my train and test data sets.

DF.Test=read.csv("C:/Users/hdesi/Desktop/MBA/Machine Learning/Back\_order\_test.csv")  
DF.Train=read.csv("C:/Users/hdesi/Desktop/MBA/Machine Learning/Back\_order\_train.csv")  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(caret)

## Warning: package 'caret' was built under R version 4.1.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.1.2

## Loading required package: lattice

library(class)  
library(ISLR)

## Warning: package 'ISLR' was built under R version 4.1.1

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.3

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble 3.1.2 v purrr 0.3.4  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'readr' was built under R version 4.1.3

## Warning: package 'stringr' was built under R version 4.1.2

## Warning: package 'forcats' was built under R version 4.1.3

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x purrr::lift() masks caret::lift()

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.1.3

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(stats)  
DF.Test <- na.omit(DF.Test) ##Remove attributes missing data  
DF.Train <- na.omit(DF.Train) ##Remove attributes missing data

##Making sure variables are numerical

sapply(DF.Train, class) ##Making sure variables are numerical

## sku national\_inv lead\_time in\_transit\_qty   
## "numeric" "numeric" "numeric" "numeric"   
## forecast\_3\_month forecast\_6\_month forecast\_9\_month sales\_1\_month   
## "numeric" "numeric" "numeric" "numeric"   
## sales\_3\_month sales\_6\_month sales\_9\_month min\_bank   
## "numeric" "numeric" "numeric" "numeric"   
## potential\_issue pieces\_past\_due perf\_6\_month\_avg perf\_12\_month\_avg   
## "character" "numeric" "numeric" "numeric"   
## local\_bo\_qty deck\_risk oe\_constraint ppap\_risk   
## "numeric" "character" "character" "character"   
## stop\_auto\_buy rev\_stop went\_on\_backorder   
## "character" "character" "character"

##Removing uneeded attributes and changing yes/no attributes to factors

DF.Train$sku<-NULL ##I do not want sku to affect prediction of model  
DF.Train$potential\_issue=as.factor(DF.Train$potential\_issue)  
DF.Train$deck\_risk=as.factor(DF.Train$deck\_risk)  
DF.Train$oe\_constraint=as.factor(DF.Train$oe\_constraint)  
DF.Train$ppap\_risk=as.factor(DF.Train$ppap\_risk)  
DF.Train$stop\_auto\_buy=as.factor(DF.Train$stop\_auto\_buy)  
DF.Train$rev\_stop=as.factor(DF.Train$rev\_stop)  
DF.Train$went\_on\_backorder=as.factor(DF.Train$went\_on\_backorder)  
  
DF.Test$sku<-NULL ##I do not want sku to affect prediction of model  
DF.Test$potential\_issue=as.factor(DF.Test$potential\_issue)  
DF.Test$deck\_risk=as.factor(DF.Test$deck\_risk)  
DF.Test$oe\_constraint=as.factor(DF.Test$oe\_constraint)  
DF.Test$ppap\_risk=as.factor(DF.Test$ppap\_risk)  
DF.Test$stop\_auto\_buy=as.factor(DF.Test$stop\_auto\_buy)  
DF.Test$rev\_stop=as.factor(DF.Test$rev\_stop)  
DF.Test$went\_on\_backorder=as.factor(DF.Test$went\_on\_backorder)  
summary(DF.Train)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. : -499.00 Min. : 0.000 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 4.00 1st Qu.: 4.000 1st Qu.: 0.00 1st Qu.: 0.0   
## Median : 14.00 Median : 8.000 Median : 0.00 Median : 0.0   
## Mean : 315.18 Mean : 7.652 Mean : 33.58 Mean : 107.2   
## 3rd Qu.: 72.25 3rd Qu.: 8.000 3rd Qu.: 0.00 3rd Qu.: 2.0   
## Max. :44286.00 Max. :52.000 Max. :27430.00 Max. :20000.0   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. : 0.0 Min. : 0.0 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.: 0.0   
## Median : 0.0 Median : 0.0 Median : 0.00 Median : 1.0   
## Mean : 190.9 Mean : 273.1 Mean : 29.69 Mean : 100.1   
## 3rd Qu.: 8.0 3rd Qu.: 14.0 3rd Qu.: 4.00 3rd Qu.: 12.0   
## Max. :33480.0 Max. :47880.0 Max. :5047.00 Max. :16625.0   
## sales\_6\_month sales\_9\_month min\_bank potential\_issue  
## Min. : 0.0 Min. : 0.0 Min. : 0.00 No :1821   
## 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.: 0.00 Yes: 179   
## Median : 2.0 Median : 3.0 Median : 0.00   
## Mean : 202.6 Mean : 329.2 Mean : 32.76   
## 3rd Qu.: 24.0 3rd Qu.: 38.0 3rd Qu.: 2.00   
## Max. :36510.0 Max. :54616.0 Max. :7458.00   
## pieces\_past\_due perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## Min. : 0.000 Min. :-99.000 Min. :-99.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.620 1st Qu.: 0.630 1st Qu.: 0.000   
## Median : 0.000 Median : 0.820 Median : 0.800 Median : 0.000   
## Mean : 0.683 Mean : -9.792 Mean : -9.447 Mean : 0.451   
## 3rd Qu.: 0.000 3rd Qu.: 0.960 3rd Qu.: 0.940 3rd Qu.: 0.000   
## Max. :600.000 Max. : 1.000 Max. : 1.000 Max. :525.000   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy rev\_stop went\_on\_backorder  
## No :1513 No :1618 No :1755 No : 77 No :1971 No :1981   
## Yes: 487 Yes: 382 Yes: 245 Yes:1923 Yes: 29 Yes: 19   
##   
##   
##   
##

##Double checking variables are numeric

sapply(DF.Train, class) ##Double checking variables are factored

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## "numeric" "numeric" "numeric" "numeric"   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## "numeric" "numeric" "numeric" "numeric"   
## sales\_6\_month sales\_9\_month min\_bank potential\_issue   
## "numeric" "numeric" "numeric" "factor"   
## pieces\_past\_due perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## "numeric" "numeric" "numeric" "numeric"   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy   
## "factor" "factor" "factor" "factor"   
## rev\_stop went\_on\_backorder   
## "factor" "factor"

##Data is ready ##Creating a new data set for the normalization process - I have removed the target variable: Went on Backorder. ##I have also removed the attributes that were factored

Normalization\_DF <- data.frame(DF.Train)  
Normalization\_DF$potential\_issue<-NULL  
Normalization\_DF$deck\_risk<-NULL  
Normalization\_DF$oe\_constraint<-NULL  
Normalization\_DF$ppap\_risk<-NULL  
Normalization\_DF$stop\_auto\_buy<-NULL  
Normalization\_DF$rev\_stop<-NULL  
Normalization\_DF$went\_on\_backorder<-NULL  
summary(Normalization\_DF)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. : -499.00 Min. : 0.000 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 4.00 1st Qu.: 4.000 1st Qu.: 0.00 1st Qu.: 0.0   
## Median : 14.00 Median : 8.000 Median : 0.00 Median : 0.0   
## Mean : 315.18 Mean : 7.652 Mean : 33.58 Mean : 107.2   
## 3rd Qu.: 72.25 3rd Qu.: 8.000 3rd Qu.: 0.00 3rd Qu.: 2.0   
## Max. :44286.00 Max. :52.000 Max. :27430.00 Max. :20000.0   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. : 0.0 Min. : 0.0 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.: 0.0   
## Median : 0.0 Median : 0.0 Median : 0.00 Median : 1.0   
## Mean : 190.9 Mean : 273.1 Mean : 29.69 Mean : 100.1   
## 3rd Qu.: 8.0 3rd Qu.: 14.0 3rd Qu.: 4.00 3rd Qu.: 12.0   
## Max. :33480.0 Max. :47880.0 Max. :5047.00 Max. :16625.0   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## Min. : 0.0 Min. : 0.0 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 2.0 Median : 3.0 Median : 0.00 Median : 0.000   
## Mean : 202.6 Mean : 329.2 Mean : 32.76 Mean : 0.683   
## 3rd Qu.: 24.0 3rd Qu.: 38.0 3rd Qu.: 2.00 3rd Qu.: 0.000   
## Max. :36510.0 Max. :54616.0 Max. :7458.00 Max. :600.000   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## Min. :-99.000 Min. :-99.000 Min. : 0.000   
## 1st Qu.: 0.620 1st Qu.: 0.630 1st Qu.: 0.000   
## Median : 0.820 Median : 0.800 Median : 0.000   
## Mean : -9.792 Mean : -9.447 Mean : 0.451   
## 3rd Qu.: 0.960 3rd Qu.: 0.940 3rd Qu.: 0.000   
## Max. : 1.000 Max. : 1.000 Max. :525.000

##Normalizing the data

Norm\_model <- preProcess(Normalization\_DF,   
 method = c("center", "scale"))  
backorder\_norm=predict(Norm\_model,Normalization\_DF)  
summary(backorder\_norm)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. :-0.3966 Min. :-1.08628 Min. :-0.05227 Min. :-0.1140   
## 1st Qu.:-0.1516 1st Qu.:-0.51848 1st Qu.:-0.05227 1st Qu.:-0.1140   
## Median :-0.1467 Median : 0.04933 Median :-0.05227 Median :-0.1140   
## Mean : 0.0000 Mean : 0.00000 Mean : 0.00000 Mean : 0.0000   
## 3rd Qu.:-0.1183 3rd Qu.: 0.04933 3rd Qu.:-0.05227 3rd Qu.:-0.1119   
## Max. :21.4174 Max. : 6.29517 Max. :42.64156 Max. :21.1660   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. :-0.1197 Min. :-0.1218 Min. :-0.1228 Min. :-0.1192   
## 1st Qu.:-0.1197 1st Qu.:-0.1218 1st Qu.:-0.1228 1st Qu.:-0.1192   
## Median :-0.1197 Median :-0.1218 Median :-0.1228 Median :-0.1180   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.:-0.1146 3rd Qu.:-0.1155 3rd Qu.:-0.1062 3rd Qu.:-0.1049   
## Max. :20.8677 Max. :21.2238 Max. :20.7460 Max. :19.6717   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## Min. :-0.1204 Min. :-0.1201 Min. :-0.1250 Min. :-0.03737   
## 1st Qu.:-0.1204 1st Qu.:-0.1201 1st Qu.:-0.1250 1st Qu.:-0.03737   
## Median :-0.1192 Median :-0.1190 Median :-0.1250 Median :-0.03737   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.00000   
## 3rd Qu.:-0.1062 3rd Qu.:-0.1062 3rd Qu.:-0.1174 3rd Qu.:-0.03737   
## Max. :21.5780 Max. :19.7989 Max. :28.3316 Max. :32.79221   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## Min. :-2.9033 Min. :-2.9583 Min. :-0.03705   
## 1st Qu.: 0.3389 1st Qu.: 0.3329 1st Qu.:-0.03705   
## Median : 0.3454 Median : 0.3385 Median :-0.03705   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.00000   
## 3rd Qu.: 0.3499 3rd Qu.: 0.3431 3rd Qu.:-0.03705   
## Max. : 0.3512 Max. : 0.3451 Max. :43.09675

##Adding back in the attributes that were removed for normalization

backorder\_norm$potential\_issue=DF.Train$potential\_issue  
backorder\_norm$deck\_risk=DF.Train$deck\_risk  
backorder\_norm$oe\_constraint=DF.Train$oe\_constraint  
backorder\_norm$ppap\_risk=DF.Train$ppap\_risk  
backorder\_norm$stop\_auto\_buy=DF.Train$stop\_auto\_buy  
backorder\_norm$rev\_stop=DF.Train$rev\_stop  
backorder\_norm$went\_on\_backorder=DF.Train$went\_on\_backorder  
summary(backorder\_norm)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. :-0.3966 Min. :-1.08628 Min. :-0.05227 Min. :-0.1140   
## 1st Qu.:-0.1516 1st Qu.:-0.51848 1st Qu.:-0.05227 1st Qu.:-0.1140   
## Median :-0.1467 Median : 0.04933 Median :-0.05227 Median :-0.1140   
## Mean : 0.0000 Mean : 0.00000 Mean : 0.00000 Mean : 0.0000   
## 3rd Qu.:-0.1183 3rd Qu.: 0.04933 3rd Qu.:-0.05227 3rd Qu.:-0.1119   
## Max. :21.4174 Max. : 6.29517 Max. :42.64156 Max. :21.1660   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. :-0.1197 Min. :-0.1218 Min. :-0.1228 Min. :-0.1192   
## 1st Qu.:-0.1197 1st Qu.:-0.1218 1st Qu.:-0.1228 1st Qu.:-0.1192   
## Median :-0.1197 Median :-0.1218 Median :-0.1228 Median :-0.1180   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.:-0.1146 3rd Qu.:-0.1155 3rd Qu.:-0.1062 3rd Qu.:-0.1049   
## Max. :20.8677 Max. :21.2238 Max. :20.7460 Max. :19.6717   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## Min. :-0.1204 Min. :-0.1201 Min. :-0.1250 Min. :-0.03737   
## 1st Qu.:-0.1204 1st Qu.:-0.1201 1st Qu.:-0.1250 1st Qu.:-0.03737   
## Median :-0.1192 Median :-0.1190 Median :-0.1250 Median :-0.03737   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.00000   
## 3rd Qu.:-0.1062 3rd Qu.:-0.1062 3rd Qu.:-0.1174 3rd Qu.:-0.03737   
## Max. :21.5780 Max. :19.7989 Max. :28.3316 Max. :32.79221   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty potential\_issue  
## Min. :-2.9033 Min. :-2.9583 Min. :-0.03705 No :1821   
## 1st Qu.: 0.3389 1st Qu.: 0.3329 1st Qu.:-0.03705 Yes: 179   
## Median : 0.3454 Median : 0.3385 Median :-0.03705   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.00000   
## 3rd Qu.: 0.3499 3rd Qu.: 0.3431 3rd Qu.:-0.03705   
## Max. : 0.3512 Max. : 0.3451 Max. :43.09675   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy rev\_stop went\_on\_backorder  
## No :1513 No :1618 No :1755 No : 77 No :1971 No :1981   
## Yes: 487 Yes: 382 Yes: 245 Yes:1923 Yes: 29 Yes: 19   
##   
##   
##   
##

##Separating my training data set into a training and validation set

Train\_Index = createDataPartition(DF.Train$went\_on\_backorder,p=0.6, list=FALSE)  
Train.df=backorder\_norm[Train\_Index,]  
Validation.df=backorder\_norm[-Train\_Index,]   
  
sapply(lapply(Train.df, unique), length)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## 293 19 76 135   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## 180 206 104 170   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## 214 259 114 7   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty potential\_issue   
## 81 84 10 2   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy   
## 2 2 2 2   
## rev\_stop went\_on\_backorder   
## 2 2

sapply(lapply(Validation.df, unique), length)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## 244 21 71 114   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## 134 155 93 138   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## 170 202 94 7   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty potential\_issue   
## 76 81 8 2   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy   
## 2 2 2 2   
## rev\_stop went\_on\_backorder   
## 2 2

##Inputing 1st set of attributes for prediction

To\_Predict=data.frame(national\_inv=490, lead\_time=7.872, in\_transit\_qty=45.5, forecast\_3\_month=188.7, forecast\_6\_month=365.3, forecast\_9\_month=536, sales\_1\_month=56.9, sales\_3\_month=178.5, sales\_6\_month=352.2, sales\_9\_month=544, min\_bank=53.2, potential\_issue=0, pieces\_past\_due=2.17, perf\_6\_month\_avg=-1.015, perf\_12\_month\_avg=-0.5532, local\_bo\_qty=0.633, deck\_risk=0, oe\_constraint=0, ppap\_risk=0, stop\_auto\_buy=1, rev\_stop=0)  
  
print(To\_Predict)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month forecast\_6\_month  
## 1 490 7.872 45.5 188.7 365.3  
## forecast\_9\_month sales\_1\_month sales\_3\_month sales\_6\_month sales\_9\_month  
## 1 536 56.9 178.5 352.2 544  
## min\_bank potential\_issue pieces\_past\_due perf\_6\_month\_avg perf\_12\_month\_avg  
## 1 53.2 0 2.17 -1.015 -0.5532  
## local\_bo\_qty deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy rev\_stop  
## 1 0.633 0 0 0 1 0

##Normalizing prediction data

To\_Predict\_norm=To\_Predict  
To\_Predict\_norm$potential\_issue<-NULL  
To\_Predict\_norm$deck\_risk<-NULL  
To\_Predict\_norm$oe\_constraint<-NULL  
To\_Predict\_norm$ppap\_risk<-NULL  
To\_Predict\_norm$stop\_auto\_buy<-NULL  
To\_Predict\_norm$rev\_stop<-NULL  
To\_Predict\_norm$went\_on\_backorder<-NULL  
To\_Predict\_norm=predict(Norm\_model,To\_Predict\_norm)

##Adding attributes back in

To\_Predict\_norm$potential\_issue=To\_Predict$potential\_issue  
To\_Predict\_norm$deck\_risk=To\_Predict$deck\_risk  
To\_Predict\_norm$oe\_constraint=To\_Predict$oe\_constraint  
To\_Predict\_norm$ppap\_risk=To\_Predict$ppap\_risk  
To\_Predict\_norm$stop\_auto\_buy=To\_Predict$stop\_auto\_buy  
To\_Predict\_norm$rev\_stop=To\_Predict$rev\_stop  
To\_Predict\_norm$went\_on\_backorder=To\_Predict$went\_on\_backorder  
summary(To\_Predict\_norm)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. :0.08515 Min. :0.03116 Min. :0.01855 Min. :0.08674   
## 1st Qu.:0.08515 1st Qu.:0.03116 1st Qu.:0.01855 1st Qu.:0.08674   
## Median :0.08515 Median :0.03116 Median :0.01855 Median :0.08674   
## Mean :0.08515 Mean :0.03116 Mean :0.01855 Mean :0.08674   
## 3rd Qu.:0.08515 3rd Qu.:0.03116 3rd Qu.:0.01855 3rd Qu.:0.08674   
## Max. :0.08515 Max. :0.03116 Max. :0.01855 Max. :0.08674   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. :0.1093 Min. :0.1172 Min. :0.1125 Min. :0.09329   
## 1st Qu.:0.1093 1st Qu.:0.1172 1st Qu.:0.1125 1st Qu.:0.09329   
## Median :0.1093 Median :0.1172 Median :0.1125 Median :0.09329   
## Mean :0.1093 Mean :0.1172 Mean :0.1125 Mean :0.09329   
## 3rd Qu.:0.1093 3rd Qu.:0.1172 3rd Qu.:0.1125 3rd Qu.:0.09329   
## Max. :0.1093 Max. :0.1172 Max. :0.1125 Max. :0.09329   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## Min. :0.0889 Min. :0.07833 Min. :0.07798 Min. :0.08136   
## 1st Qu.:0.0889 1st Qu.:0.07833 1st Qu.:0.07798 1st Qu.:0.08136   
## Median :0.0889 Median :0.07833 Median :0.07798 Median :0.08136   
## Mean :0.0889 Mean :0.07833 Mean :0.07798 Mean :0.08136   
## 3rd Qu.:0.0889 3rd Qu.:0.07833 3rd Qu.:0.07798 3rd Qu.:0.08136   
## Max. :0.0889 Max. :0.07833 Max. :0.07798 Max. :0.08136   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty potential\_issue  
## Min. :0.2857 Min. :0.2938 Min. :0.01495 Min. :0   
## 1st Qu.:0.2857 1st Qu.:0.2938 1st Qu.:0.01495 1st Qu.:0   
## Median :0.2857 Median :0.2938 Median :0.01495 Median :0   
## Mean :0.2857 Mean :0.2938 Mean :0.01495 Mean :0   
## 3rd Qu.:0.2857 3rd Qu.:0.2938 3rd Qu.:0.01495 3rd Qu.:0   
## Max. :0.2857 Max. :0.2938 Max. :0.01495 Max. :0   
## deck\_risk oe\_constraint ppap\_risk stop\_auto\_buy rev\_stop  
## Min. :0 Min. :0 Min. :0 Min. :1 Min. :0   
## 1st Qu.:0 1st Qu.:0 1st Qu.:0 1st Qu.:1 1st Qu.:0   
## Median :0 Median :0 Median :0 Median :1 Median :0   
## Mean :0 Mean :0 Mean :0 Mean :1 Mean :0   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0 3rd Qu.:1 3rd Qu.:0   
## Max. :0 Max. :0 Max. :0 Max. :1 Max. :0

Train.df$potential\_issue=as.numeric(Train.df$potential\_issue)  
Train.df$deck\_risk=as.numeric(Train.df$deck\_risk)  
Train.df$oe\_constraint=as.numeric(Train.df$oe\_constraint)  
Train.df$ppap\_risk=as.numeric(Train.df$ppap\_risk)  
Train.df$stop\_auto\_buy=as.numeric(Train.df$stop\_auto\_buy)  
Train.df$rev\_stop=as.numeric(Train.df$rev\_stop)  
  
Validation.df$potential\_issue=as.numeric(Validation.df$potential\_issue)  
Validation.df$deck\_risk=as.numeric(Validation.df$deck\_risk)  
Validation.df$oe\_constraint=as.numeric(Validation.df$oe\_constraint)  
Validation.df$ppap\_risk=as.numeric(Validation.df$ppap\_risk)  
Validation.df$stop\_auto\_buy=as.numeric(Validation.df$stop\_auto\_buy)  
Validation.df$rev\_stop=as.numeric(Validation.df$rev\_stop)

##Finding best k-value

set.seed(123)  
  
fitControl <- trainControl(method = "repeatedcv",  
 number = 3,  
 repeats = 2)  
  
searchGrid=expand.grid(k = 1:10)  
  
Knn.model=train(went\_on\_backorder~.,   
 data=Train.df,  
 method='knn',  
 tuneGrid=searchGrid,  
 trControl = fitControl,)  
  
Knn.model

## k-Nearest Neighbors   
##   
## 1201 samples  
## 21 predictor  
## 2 classes: 'No', 'Yes'   
##   
## No pre-processing  
## Resampling: Cross-Validated (3 fold, repeated 2 times)   
## Summary of sample sizes: 801, 801, 800, 801, 800, 801, ...   
## Resampling results across tuning parameters:  
##   
## k Accuracy Kappa   
## 1 0.9808510 0.055616770  
## 2 0.9808489 0.062858617  
## 3 0.9883416 -0.002237136  
## 4 0.9887583 -0.001787912  
## 5 0.9895916 -0.000669344  
## 6 0.9900083 0.000000000  
## 7 0.9900083 0.000000000  
## 8 0.9900083 0.000000000  
## 9 0.9900083 0.000000000  
## 10 0.9900083 0.000000000  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was k = 10.

##Best K-Value is 10 ##Prediction on my training data

Train\_Prediction <-knn(train=Train.df[1:21],   
 test=To\_Predict\_norm,  
 cl=Train.df$went\_on\_backorder,  
 k=10)  
print(Train\_Prediction)

## [1] No  
## Levels: No Yes

##Validating my model with validation data

Validation\_prediction <-predict(Knn.model,Validation.df)  
confusionMatrix(Validation\_prediction,Validation.df$went\_on\_backorder)##Creating my confusion matrix

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 792 7  
## Yes 0 0  
##   
## Accuracy : 0.9912   
## 95% CI : (0.982, 0.9965)  
## No Information Rate : 0.9912   
## P-Value [Acc > NIR] : 0.59871   
##   
## Kappa : 0   
##   
## Mcnemar's Test P-Value : 0.02334   
##   
## Sensitivity : 1.0000   
## Specificity : 0.0000   
## Pos Pred Value : 0.9912   
## Neg Pred Value : NaN   
## Prevalence : 0.9912   
## Detection Rate : 0.9912   
## Detection Prevalence : 1.0000   
## Balanced Accuracy : 0.5000   
##   
## 'Positive' Class : No   
##

##99% Accuracy points to a strong predictive model for my training/validation set ##Normalization of test data set

Normalization\_DF\_Test <- data.frame(DF.Test)  
Normalization\_DF\_Test$potential\_issue<-NULL  
Normalization\_DF\_Test$deck\_risk<-NULL  
Normalization\_DF\_Test$oe\_constraint<-NULL  
Normalization\_DF\_Test$ppap\_risk<-NULL  
Normalization\_DF\_Test$stop\_auto\_buy<-NULL  
Normalization\_DF\_Test$rev\_stop<-NULL  
Normalization\_DF\_Test$went\_on\_backorder<-NULL  
  
sapply(Normalization\_DF\_Test, class)##Checking that format is correct

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## "numeric" "numeric" "numeric" "numeric"   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## "numeric" "numeric" "numeric" "numeric"   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## "numeric" "numeric" "numeric" "numeric"   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## "numeric" "numeric" "numeric"

##Normalization

Norm\_model\_Test <- preProcess(Normalization\_DF,   
 method = c("center", "scale"))  
backorder\_norm\_test=predict(Norm\_model\_Test,Normalization\_DF\_Test)  
summary(backorder\_norm\_test)

## national\_inv lead\_time in\_transit\_qty forecast\_3\_month   
## Min. :-12.5322 Min. :-0.94433 Min. : -0.05227 Min. : -0.114   
## 1st Qu.: 0.7674 1st Qu.:-0.80238 1st Qu.: 0.20805 1st Qu.: 3.505   
## Median : 1.7120 Median : 0.04933 Median : 1.11586 Median : 5.576   
## Mean : 5.2739 Mean :-0.23074 Mean : 3.65153 Mean : 16.635   
## 3rd Qu.: 4.1737 3rd Qu.: 0.04933 3rd Qu.: 3.17662 3rd Qu.: 11.553   
## Max. :383.0325 Max. : 2.60445 Max. :290.42136 Max. :1607.159   
## forecast\_6\_month forecast\_9\_month sales\_1\_month sales\_3\_month   
## Min. : -0.1197 Min. : 3.360 Min. : -0.1228 Min. : -0.1192   
## 1st Qu.: 4.2676 1st Qu.: 4.515 1st Qu.: 3.0528 1st Qu.: 3.2155   
## Median : 6.5000 Median : 6.984 Median : 5.0148 Median : 5.1491   
## Mean : 18.8901 Mean : 19.471 Mean : 15.0796 Mean : 14.2024   
## 3rd Qu.: 13.4977 3rd Qu.: 14.144 3rd Qu.: 10.3933 3rd Qu.: 10.1712   
## Max. :1352.0371 Max. :1409.658 Max. :1236.9929 Max. :1293.3498   
## sales\_6\_month sales\_9\_month min\_bank pieces\_past\_due   
## Min. : -0.1204 Min. : -0.1201 Min. : -0.125 Min. : -0.037   
## 1st Qu.: 3.1994 1st Qu.: 2.9076 1st Qu.: 2.667 1st Qu.: -0.037   
## Median : 5.1734 Median : 4.6697 Median : 4.551 Median : -0.037   
## Mean : 14.2493 Mean : 13.1247 Mean : 12.650 Mean : 8.630   
## 3rd Qu.: 10.3771 3rd Qu.: 9.5328 3rd Qu.: 9.668 3rd Qu.: -0.037   
## Max. :1249.9531 Max. :1165.2036 Max. :751.045 Max. :4375.270   
## perf\_6\_month\_avg perf\_12\_month\_avg local\_bo\_qty   
## Min. :-2.9033 Min. :-2.9583 Min. : -0.0371   
## 1st Qu.: 0.3441 1st Qu.: 0.3365 1st Qu.: -0.0371   
## Median : 0.3490 Median : 0.3425 Median : -0.0371   
## Mean : 0.2797 Mean : 0.2969 Mean : 5.3015   
## 3rd Qu.: 0.3506 3rd Qu.: 0.3441 3rd Qu.: -0.0371   
## Max. : 0.3512 Max. : 0.3451 Max. :511.9817

##Adding factors back in

backorder\_norm\_test$potential\_issue=DF.Test$potential\_issue  
backorder\_norm\_test$deck\_risk=DF.Test$deck\_risk  
backorder\_norm\_test$oe\_constraint=DF.Test$oe\_constraint  
backorder\_norm\_test$ppap\_risk=DF.Test$ppap\_risk  
backorder\_norm\_test$stop\_auto\_buy=DF.Test$stop\_auto\_buy  
backorder\_norm\_test$rev\_stop=DF.Test$rev\_stop  
backorder\_norm\_test$went\_on\_backorder=DF.Test$went\_on\_backorder

##Changing factors to numeric values to match model training set parameters

backorder\_norm\_test$potential\_issue=as.numeric(backorder\_norm\_test$potential\_issue)  
backorder\_norm\_test$deck\_risk=as.numeric(backorder\_norm\_test$deck\_risk)  
backorder\_norm\_test$oe\_constraint=as.numeric(backorder\_norm\_test$oe\_constraint)  
backorder\_norm\_test$ppap\_risk=as.numeric(backorder\_norm\_test$ppap\_risk)  
backorder\_norm\_test$stop\_auto\_buy=as.numeric(backorder\_norm\_test$stop\_auto\_buy)  
backorder\_norm\_test$rev\_stop=as.numeric(backorder\_norm\_test$rev\_stop)  
  
##Using my test set to check for overfitting  
  
Test\_Prediction <-knn(train=Train.df[1:21],   
 test=backorder\_norm\_test[1:21],  
 cl=Train.df$went\_on\_backorder,  
 k=10)

##Check for overfitting

Test\_Prediction<-predict(Knn.model,backorder\_norm\_test)  
confusionMatrix(Test\_Prediction,backorder\_norm\_test$went\_on\_backorder)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 1965 35  
## Yes 0 0  
##   
## Accuracy : 0.9825   
## 95% CI : (0.9757, 0.9878)  
## No Information Rate : 0.9825   
## P-Value [Acc > NIR] : 0.5448   
##   
## Kappa : 0   
##   
## Mcnemar's Test P-Value : 9.081e-09   
##   
## Sensitivity : 1.0000   
## Specificity : 0.0000   
## Pos Pred Value : 0.9825   
## Neg Pred Value : NaN   
## Prevalence : 0.9825   
## Detection Rate : 0.9825   
## Detection Prevalence : 1.0000   
## Balanced Accuracy : 0.5000   
##   
## 'Positive' Class : No   
##

##Test set accuracy decreased slightly, but not enough for concern

##Backorder prediction on 1st sku

Predict\_Sku\_3449673=data.frame(national\_inv=389335, lead\_time=8, in\_transit\_qty=25884, forecast\_3\_month=1021940, forecast\_6\_month=2094452, forecast\_9\_month=3162260, sales\_1\_month=299190, sales\_3\_month=1086554, sales\_6\_month=2103389, sales\_9\_month=3195211, min\_bank=168852, potential\_issue=0, pieces\_past\_due=79964, perf\_6\_month\_avg=0.82, perf\_12\_month\_avg=0.54, local\_bo\_qty=0, deck\_risk=0, oe\_constraint=0, ppap\_risk=0, stop\_auto\_buy=1, rev\_stop=0)  
  
Predict\_Sku\_3449673\_norm=Predict\_Sku\_3449673  
Predict\_Sku\_3449673\_norm$potential\_issue<-NULL  
Predict\_Sku\_3449673\_norm$deck\_risk<-NULL  
Predict\_Sku\_3449673\_norm$oe\_constraint<-NULL  
Predict\_Sku\_3449673\_norm$ppap\_risk<-NULL  
Predict\_Sku\_3449673\_norm$stop\_auto\_buy<-NULL  
Predict\_Sku\_3449673\_norm$rev\_stop<-NULL  
Predict\_Sku\_3449673\_norm$went\_on\_backorder<-NULL  
Predict\_Sku\_3449673\_norm=predict(Norm\_model,Predict\_Sku\_3449673\_norm)  
  
Predict\_Sku\_3449673\_norm$potential\_issue=Predict\_Sku\_3449673$potential\_issue  
Predict\_Sku\_3449673\_norm$deck\_risk=Predict\_Sku\_3449673$deck\_risk  
Predict\_Sku\_3449673\_norm$oe\_constraint=Predict\_Sku\_3449673$oe\_constraint  
Predict\_Sku\_3449673\_norm$ppap\_risk=Predict\_Sku\_3449673$ppap\_risk  
Predict\_Sku\_3449673\_norm$stop\_auto\_buy=Predict\_Sku\_3449673$stop\_auto\_buy  
Predict\_Sku\_3449673\_norm$rev\_stop=Predict\_Sku\_3449673$rev\_stop  
Predict\_Sku\_3449673\_norm$went\_on\_backorder=Predict\_Sku\_3449673$went\_on\_backorder  
  
Predict\_Sku\_3449673\_norm$potential\_issue=as.numeric(Predict\_Sku\_3449673\_norm$potential\_issue)  
Predict\_Sku\_3449673\_norm$deck\_risk=as.numeric(Predict\_Sku\_3449673\_norm$deck\_risk)  
Predict\_Sku\_3449673\_norm$oe\_constraint=as.numeric(Predict\_Sku\_3449673\_norm$oe\_constraint)  
Predict\_Sku\_3449673\_norm$ppap\_risk=as.numeric(Predict\_Sku\_3449673\_norm$ppap\_risk)  
Predict\_Sku\_3449673\_norm$stop\_auto\_buy=as.numeric(Predict\_Sku\_3449673\_norm$stop\_auto\_buy)  
Predict\_Sku\_3449673\_norm$rev\_stop=as.numeric(Predict\_Sku\_3449673\_norm$rev\_stop)  
  
Prediction\_Sku\_3449673 <-knn(train=Train.df[1:21],   
 test=Predict\_Sku\_3449673\_norm,  
 cl=Train.df$went\_on\_backorder,  
 k=10)##best k value  
print(Prediction\_Sku\_3449673)

## [1] No  
## Levels: No Yes

##Prediction of no backorder

##Backorder prediction on 2nd sku

Predict\_Sku\_3301406=data.frame(national\_inv=627349, lead\_time=8, in\_transit\_qty=123648, forecast\_3\_month=1092576, forecast\_6\_month=2157024, forecast\_9\_month=3124704, sales\_1\_month=290705, sales\_3\_month=912833, sales\_6\_month=1762825, sales\_9\_month=2616764, min\_bank=191682, potential\_issue=0, pieces\_past\_due=0, perf\_6\_month\_avg=0.55, perf\_12\_month\_avg=0.74, local\_bo\_qty=0, deck\_risk=0, oe\_constraint=0, ppap\_risk=0, stop\_auto\_buy=1, rev\_stop=0)  
  
Predict\_Sku\_3301406\_norm=Predict\_Sku\_3301406  
Predict\_Sku\_3301406\_norm$potential\_issue<-NULL  
Predict\_Sku\_3301406\_norm$deck\_risk<-NULL  
Predict\_Sku\_3301406\_norm$oe\_constraint<-NULL  
Predict\_Sku\_3301406\_norm$ppap\_risk<-NULL  
Predict\_Sku\_3301406\_norm$stop\_auto\_buy<-NULL  
Predict\_Sku\_3301406\_norm$rev\_stop<-NULL  
Predict\_Sku\_3301406\_norm$went\_on\_backorder<-NULL  
Predict\_Sku\_3301406\_norm=predict(Norm\_model,Predict\_Sku\_3301406\_norm)  
  
Predict\_Sku\_3301406\_norm$potential\_issue=Predict\_Sku\_3301406$potential\_issue  
Predict\_Sku\_3301406\_norm$deck\_risk=Predict\_Sku\_3301406$deck\_risk  
Predict\_Sku\_3301406\_norm$oe\_constraint=Predict\_Sku\_3301406$oe\_constraint  
Predict\_Sku\_3301406\_norm$ppap\_risk=Predict\_Sku\_3301406$ppap\_risk  
Predict\_Sku\_3301406\_norm$stop\_auto\_buy=Predict\_Sku\_3301406$stop\_auto\_buy  
Predict\_Sku\_3301406\_norm$rev\_stop=Predict\_Sku\_3301406$rev\_stop  
Predict\_Sku\_3301406\_norm$went\_on\_backorder=Predict\_Sku\_3301406$went\_on\_backorder  
  
Predict\_Sku\_3301406\_norm$potential\_issue=as.numeric(Predict\_Sku\_3301406\_norm$potential\_issue)  
Predict\_Sku\_3301406\_norm$deck\_risk=as.numeric(Predict\_Sku\_3301406\_norm$deck\_risk)  
Predict\_Sku\_3301406\_norm$oe\_constraint=as.numeric(Predict\_Sku\_3301406\_norm$oe\_constraint)  
Predict\_Sku\_3301406\_norm$ppap\_risk=as.numeric(Predict\_Sku\_3301406\_norm$ppap\_risk)  
Predict\_Sku\_3301406\_norm$stop\_auto\_buy=as.numeric(Predict\_Sku\_3301406\_norm$stop\_auto\_buy)  
Predict\_Sku\_3301406\_norm$rev\_stop=as.numeric(Predict\_Sku\_3301406\_norm$rev\_stop)  
  
Prediction\_Sku\_3301406 <-knn(train=Train.df[1:21],   
 test=Predict\_Sku\_3301406\_norm,  
 cl=Train.df$went\_on\_backorder,  
 k=10)  
print(Prediction\_Sku\_3301406)

## [1] No  
## Levels: No Yes

##Prediction of no backorder

##Backorder prediction on 3rd sku

Predict\_Sku\_3337435=data.frame(national\_inv=786698, lead\_time=8, in\_transit\_qty=37248, forecast\_3\_month=1510592, forecast\_6\_month=1858864, forecast\_9\_month=1858864, sales\_1\_month=187474, sales\_3\_month=197602, sales\_6\_month=197602, sales\_9\_month=202734, min\_bank=131888, potential\_issue=0, pieces\_past\_due=0, perf\_6\_month\_avg=0.89, perf\_12\_month\_avg=0.9, local\_bo\_qty=0, deck\_risk=0, oe\_constraint=0, ppap\_risk=0, stop\_auto\_buy=1, rev\_stop=0)  
  
Predict\_Sku\_3337435\_norm=Predict\_Sku\_3337435  
Predict\_Sku\_3337435\_norm$potential\_issue<-NULL  
Predict\_Sku\_3337435\_norm$deck\_risk<-NULL  
Predict\_Sku\_3337435\_norm$oe\_constraint<-NULL  
Predict\_Sku\_3337435\_norm$ppap\_risk<-NULL  
Predict\_Sku\_3337435\_norm$stop\_auto\_buy<-NULL  
Predict\_Sku\_3337435\_norm$rev\_stop<-NULL  
Predict\_Sku\_3337435\_norm$went\_on\_backorder<-NULL  
Predict\_Sku\_3337435\_norm=predict(Norm\_model,Predict\_Sku\_3337435\_norm)  
  
Predict\_Sku\_3337435\_norm$potential\_issue=Predict\_Sku\_3337435$potential\_issue  
Predict\_Sku\_3337435\_norm$deck\_risk=Predict\_Sku\_3337435$deck\_risk  
Predict\_Sku\_3337435\_norm$oe\_constraint=Predict\_Sku\_3337435$oe\_constraint  
Predict\_Sku\_3337435\_norm$ppap\_risk=Predict\_Sku\_3337435$ppap\_risk  
Predict\_Sku\_3337435\_norm$stop\_auto\_buy=Predict\_Sku\_3337435$stop\_auto\_buy  
Predict\_Sku\_3337435\_norm$rev\_stop=Predict\_Sku\_3337435$rev\_stop  
Predict\_Sku\_3337435\_norm$went\_on\_backorder=Predict\_Sku\_3337435$went\_on\_backorder  
  
Predict\_Sku\_3337435\_norm$potential\_issue=as.numeric(Predict\_Sku\_3337435\_norm$potential\_issue)  
Predict\_Sku\_3337435\_norm$deck\_risk=as.numeric(Predict\_Sku\_3337435\_norm$deck\_risk)  
Predict\_Sku\_3337435\_norm$oe\_constraint=as.numeric(Predict\_Sku\_3337435\_norm$oe\_constraint)  
Predict\_Sku\_3337435\_norm$ppap\_risk=as.numeric(Predict\_Sku\_3337435\_norm$ppap\_risk)  
Predict\_Sku\_3337435\_norm$stop\_auto\_buy=as.numeric(Predict\_Sku\_3337435\_norm$stop\_auto\_buy)  
Predict\_Sku\_3337435\_norm$rev\_stop=as.numeric(Predict\_Sku\_3337435\_norm$rev\_stop)  
  
Prediction\_Sku\_3337435 <-knn(train=Train.df[1:21],   
 test=Predict\_Sku\_3337435\_norm,  
 cl=Train.df$went\_on\_backorder,  
 k=10)  
print(Prediction\_Sku\_3337435)

## [1] No  
## Levels: No Yes

##Prediction of no backorder