

AI in School Safety

RAzad

9/27/2019

Introduction

There is nothing more essential than safeguarding the well-being of our children and youth. Creating safe and accommodating schools is central to this resolution and must be a national priority. Moreover, school safety cannot be achieved with a single program or piece of security equipment. Rather, effective school safety starts with prevention. There is no clear research evidence that the use of metal detectors, security cameras, or heavily armed guards in schools is effective in preventing school violence[1]. In fact, research has shown that their presence negatively impacts students' perceptions of safety and even increases fear among some students. In addition, studies suggest that restrictive school security measures have the potential to harm school learning environments. Therefore, the school authorities should come up with more intuitive approaches for preventative measures. In this regard, machine learning can be a very useful mechanism.

This paper is a manuscript of 'AI in School Safety' which illustrates the possibility of school safety measures using machine learning algorithm applied to previous school safety statistics. These school safety statistics involve the number of violent crimes, non-violent crimes, and property damages to the schools and so on. This machine learning model can predict the possibility of violence in any New York City borough schools within the same school year using their prior history of violence. In this project, NYC school safety statistics were used for the school years 2013-2017. In order to train the model, school years 2013-2016 were utilized and for validation purpose the 2016-2017 school year data was used. This school safety data set is open source and acquired from "<https://data.cityofnewyork.us/browse?q=school+safety>".

The objective of this model is to apply this model to any NYC school and determine whether the school has any possibility of violence for the current school year. If so, then the authorities can decide on what kind of preventative measures they need to implement. However, an optimized threshold value plays a critical role in deciding whether the school is in danger or not. therefore, our ultimate goal for this project will involve in determining an optimized threshold value.

Key factors Responsible for School Violence

According to the report "Causes of School Safety", published by Constitutional Rights foundation, school violence does not limit itself to the student population. Eight percent of teachers say that they are threatened with violence on school grounds at least once a month. Two percent of them reported being physically attacked each year[2]. Besides the problem of school violence has roots in a broader culture beyond the school, therefore the solution will also not be found in a single school climate.

As Conoley and Goldstein hypothesized that a culture of violence exists because of the following: skill deficits, abuse, poverty, racism, unemployment, poor classroom climate, the easy accessibility of weapons, drugs, and alcohol, an estrangement from one's culture, the lack of supervised constructive activities for youth, a reduction in the influence of social institutions, and media that glorify violence[3]. Nonetheless, the following factors have the most influence on school violence amongst all.

- 1) Access to Weapons
- 2) Cyber Abuse
- 3) Environmental Impact i.e. School Environments, Gangs at schools, School size
- 4) Community Environments i.e. economic well-being, and racial and ethnic mix
- 5) Family Environments i.e. parental alcohol abuse, domestic violence, the presence of guns in the home may encourage a child to follow in his/her parents' footsteps.

Analysis

Using the NYC Open Data source for school safety reports of 2013-2016 years, the proposed machine learning algorithm predicts the probability of school violence risk score in the current school year. Amongst various deciding factors which influence school violence that are already mentioned above, the following were used as the explanatory variables or input predictor variables while the school violence risk score is used as the output or the dependent variable. The Conoley and Goldstein hypothesis[3] have major inspiration behind these variable selection. Out of 23 variables, only 6 were used to generate the model.

- * Registered Student percentile
- * School Population rate compared to the school capacity
- * High School Graduation Rate in each school year
- * High School Dropout Rate in each school year
- * Borough Poverty Rate
- * Borough Unemployment Rate

After a massive data cleaning, the school safety data sets are ready to use for analysis. A few general properties of this data set are displayed below-

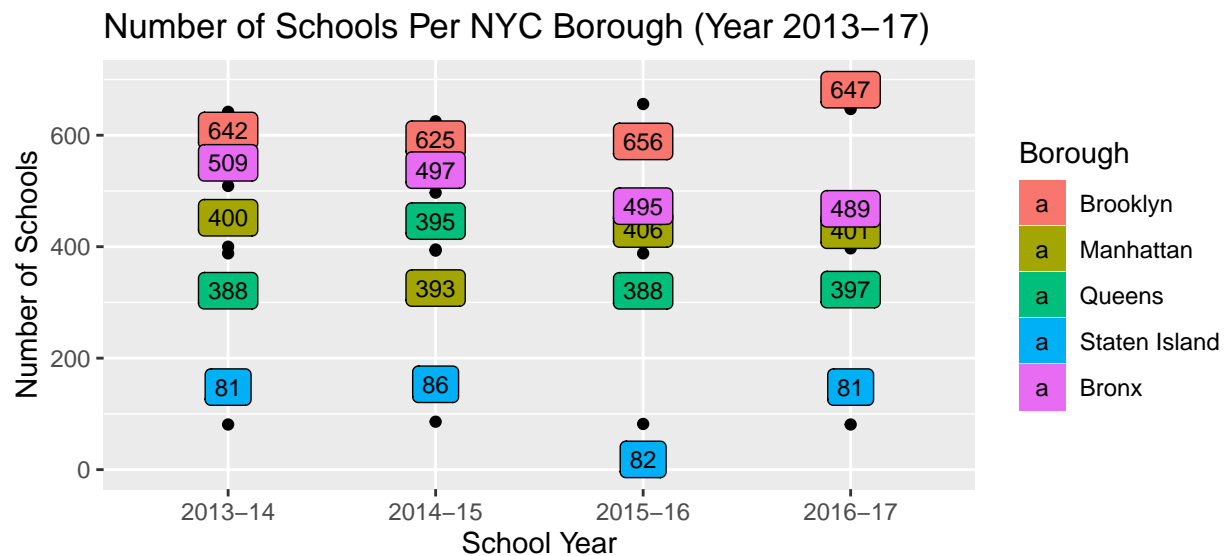
```
str(school)
```

```
## 'data.frame': 8064 obs. of 24 variables:
## $ School.Year : Factor w/ 4 levels "2013-14","2014-15",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Location.Code : Factor w/ 2004 levels "", " ", "K001",...: 3 4 121 1 5 7 8 9 10 11 ...
## $ Borough : chr "K" "K" "K" "K" ...
## $ Register : Factor w/ 1938 levels "", "#N/A", "1,002",...: 113 1198 1102 1629 1236 98 ...
## $ Total.Population : chr "1251-1500" "251-500" "251-500" "751-1000" ...
## $ Criminal.Offense : num 0 0 0 1 2 1 0 0 0 0 ...
## $ Other.Crime : num 2 0 0 5 0 1 1 1 0 0 ...
## $ NonCriminal : num 1 0 0 2 0 0 2 0 0 0 ...
## $ Property.Crime : num 1 0 0 2 2 2 0 0 0 0 ...
## $ Violent.Crime : num 0 0 0 4 0 0 0 0 0 0 ...
## $ AvgOfCriminal : num 0.86 0 0 0.52 0.33 0.35 0.33 0.56 0.52 0 ...
## $ AvgOfOther : num 3.26 0 0 1.71 1.32 1.06 1.32 2.4 1.71 0 ...
## $ AvgOfNoCrim : num 5.55 0 0 2.49 1.76 1.09 1.76 3.56 2.49 0 ...
## $ AvgOfProp : num 2.17 0 0 1.16 0.83 0.73 0.83 1.36 1.16 0 ...
## $ AvgOfVio : num 1.29 0 0 0.75 0.59 0.5 0.59 1.05 0.75 0 ...
## $ Registered.Students : num 1277 479 397 876 513 ...
## $ School.Population : num 1500 500 500 1000 750 500 750 1250 1000 750 ...
## $ BachelorDegree.Rate : num 31 31 31 31 31 31 31 31 31 31 ...
## $ Poverty.Rate : num 23 23 23 23 23 23 23 23 23 23 ...
## $ Unemployment.Rate : num 7.6 7.6 7.6 7.6 7.6 7.6 7.6 7.6 7.6 7.6 ...
## $ HighSchoolGraduation.Rate : num 68.1 68.1 68.1 68.1 68.1 68.1 68.1 68.1 68.1 68.1 ...
## $ DropOut.Rate : num 11.3 11.3 11.3 11.3 11.3 11.3 11.3 11.3 11.3 11.3 ...
## $ SchoolPopulation.Rate : num 85.1 95.8 79.4 87.6 68.4 ...
## $ FLAG_VIOL_INC : chr "yes" "no" "no" "yes" ...
```

The following graph displays the breakdown of total schools in each borough in each school year since 2013.

```
##### Number of Schools Per NYC Borough in Each School Year from 2013 to 2017
schoolsPerBorough %>%
  ggplot(aes(x = School.Year, y = Total.Schools, fill = Borough)) +
  geom_point() +
```

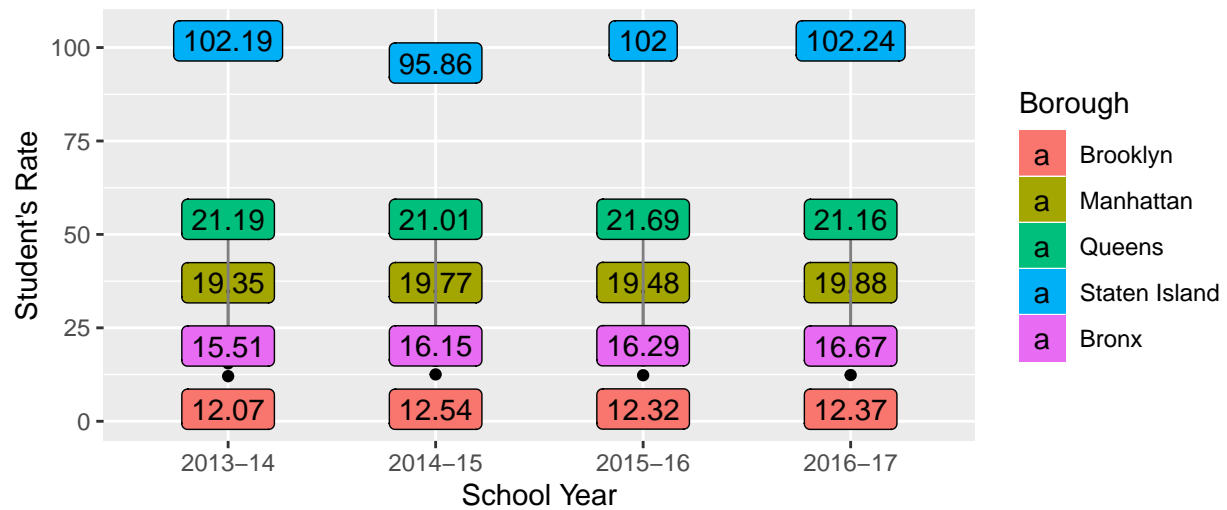
```
geom_label_repel(aes(label = Total.Schools), size = 3,
  box.padding = 0.3,
  point.padding = 0.2,
  segment.color = 'grey50',
  direction = 'y') +
ylab("Number of Schools") +
xlab("School Year") +
ylim(0,700) +
ggtitle("Number of Schools Per NYC Borough (Year 2013-17)") +
scale_fill_discrete(breaks=c("K", "M", "Q", "R", "X"),
  labels=c("Brooklyn", "Manhattan", "Queens",
    "Staten Island", "Bronx"))
```



The following graph displays the breakdown of average registered students compared to school capacity in each borough in each school year since 2013.

```
##### Rate of Registered Students Per NYC Borough in Each School Year from 2013 to 2017
schoolStat %>%
  ggplot(aes(x = School.Year, y = Students.Rate, fill = Borough)) +
  geom_point() +
  geom_label_repel(aes(label = Students.Rate), size = 4,
    box.padding = 0.3,
    point.padding = 0,
    segment.color = 'grey50',
    direction = 'y') +
  ylab("Student's Rate") +
  xlab("School Year") +
  ylim(0,105) +
  ggtitle("Registered Students' Rate Compared to School Capacity Per NYC Borough") +
  theme(plot.title = element_text(size = 12, face = "bold")) +
  scale_fill_discrete(breaks=c("K", "M", "Q", "R", "X"),
    labels=c("Brooklyn", "Manhattan", "Queens",
      "Staten Island", "Bronx"))
```

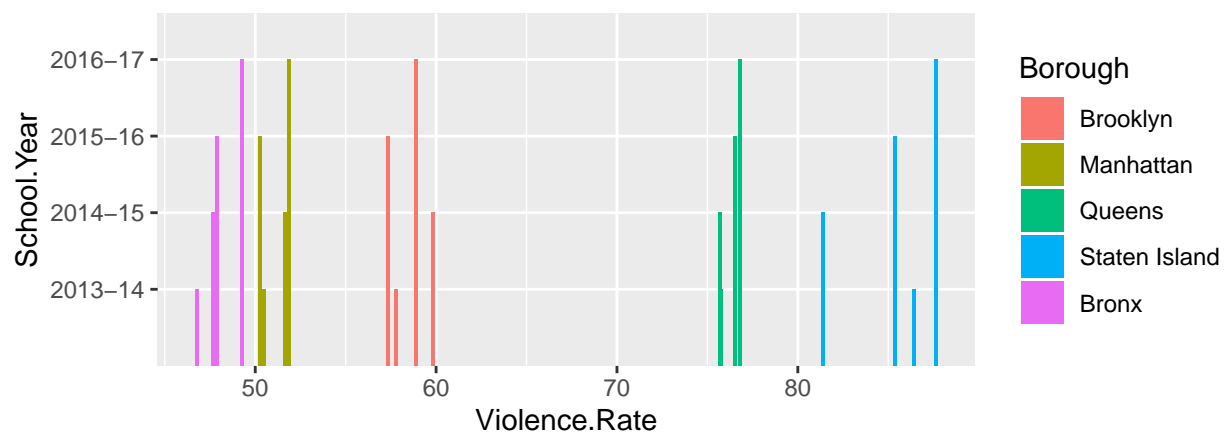
Registered Students' Rate Compared to School Capacity Per NYC Borough



If there was any Criminal activity or Violent activity occurred for a specific school within a school year, a new column has been generated as FLAG_VIOL_INC with a value of 'yes' otherwise 'no'. The following graph displays the violence rate of each borough using the total number of schools per borough.

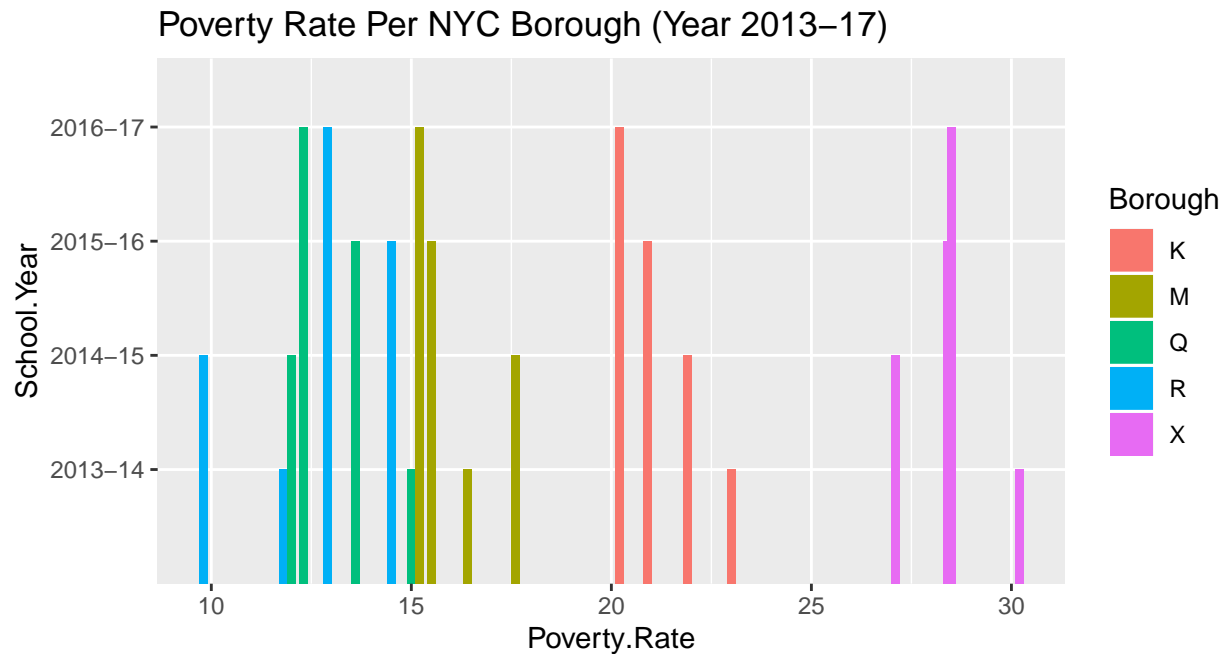
```
##### Rate of Violence Per NYC Borough in Each School Year from 2013 to 2017
df <- school %>%
  filter(FLAG_VIOL_INC == "yes") %>%
  group_by(School.Year, Borough) %>%
  dplyr::summarise(violentSchools = n())
schoolStat <- schoolsPerBorough %>%
  inner_join(df, by = c("School.Year", "Borough")) %>%
  mutate(Violence.Rate = round((violentSchools/Total.Schools)*100, digit = 2))
schoolStat %>%
  ggplot(aes(x = Violence.Rate, y = School.Year, fill = Borough)) +
  geom_col(width = 0.2) +
  ggtitle("Rate of Violence in Schools Per NYC Borough (Year 2013-17)") +
  scale_fill_discrete(breaks=c("K", "M", "Q", "R", "X"),
    labels=c("Brooklyn", "Manhattan", "Queens",
      "Staten Island", "Bronx"))
```

Rate of Violence in Schools Per NYC Borough (Year 2013-17)

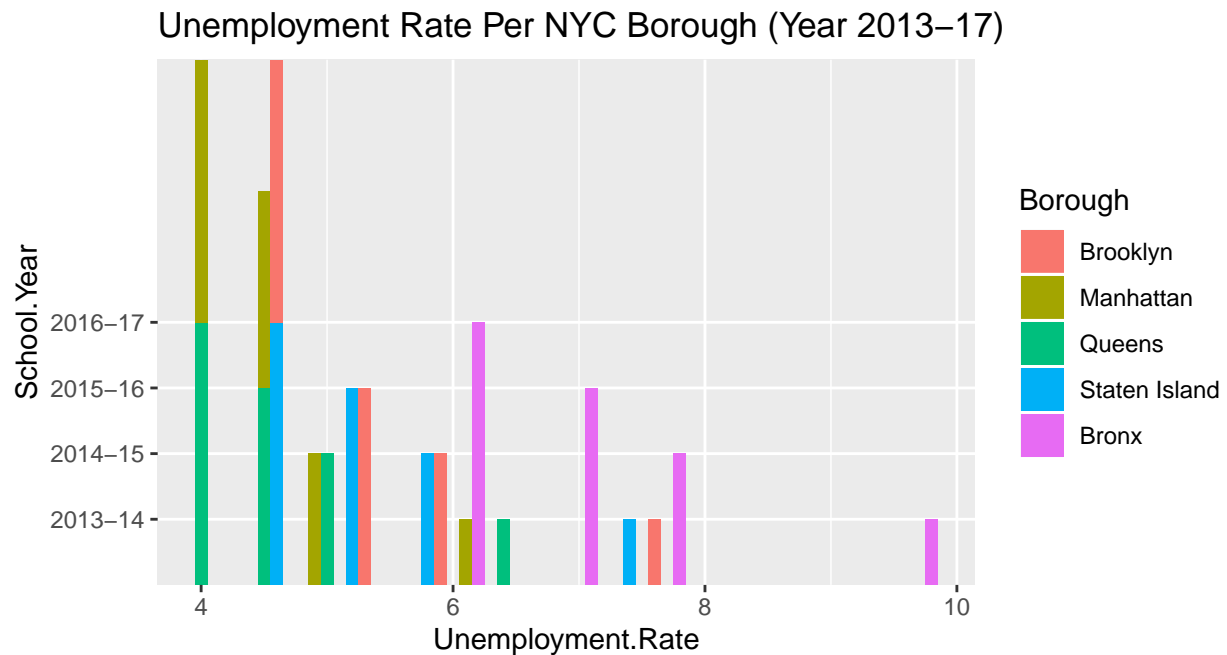


How the poverty rate in each borough has changed overtime in each borough, is displayed as follows-

```
##### NYC Borough demographics from 2013 to 2017
## Poverty rate
borough_data %>%
  filter(Borough != '0') %>%
  ggplot(aes( x = Poverty.Rate, y = School.Year, fill = Borough))+
  geom_col(width = 0.22, stat = "identity") +
  ggtitle("Poverty Rate Per NYC Borough (Year 2013-17)")
```



How the unemployment rate in each borough has changed overtime in each borough, is displayed as follows-



When data quality varies, knowledge about the potential performance of alternate predictive models can help a decision maker to design an appropriate information system in terms of predictive accuracy [7]. Data quality can be described in terms of a number of dimensions, including frequency, accuracy, and response time. In general, a model performs better the more frequent and the more accurate the data, and the shorter the time to obtain it. However, due to the wide variety of data collected for school safety measures, caret ensemble model was used to create the proposed school safety model. Ensemble models are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging) and bias (boosting), also improve predictions (stacking)[4]. In general, an ensemble model combines multiple ‘individual’ (diverse) models together and delivers superior prediction power.

Therefore, for the proposed school safety model the methods were used-

- a) Random Forest,
- b) Support Vector Machine (SVM) and
- c) Neural network models were used as a combination of one.

```
clean_school_safety_data <- function(school_data, highSchool = FALSE){

  if(!exists("school_data", mode = "function") && !highSchool){
    ## school safety model data preparation
    school_data <- school
    school_data$FLAG_VIOL_INC <- relevel(as.factor(school_data$FLAG_VIOL_INC), "yes")

    school_data <- school_data %>%
      select(School.Year, Registered.Students,
             Poverty.Rate, Unemployment.Rate,
             HighSchoolGraduation.Rate, DropOut.Rate,
             SchoolPopulation.Rate, FLAG_VIOL_INC)
  }else{
    school_data <- school_data %>%
      select(Registered.Students, Poverty.Rate,
             Unemployment.Rate, HighSchoolGraduation.Rate,
             DropOut.Rate, SchoolPopulation.Rate)
  }
}

#school safety model
set.seed(2019, sample.kind="Rounding")

school_data <- clean_school_safety_data()
# generate training set using the school years 2013-2016
data_train <- school_data %>%
  filter(School.Year != '2016-17') %>%
  select(-School.Year)
#down sample the data for training
data_train <- downSample(data_train, data_train$FLAG_VIOL_INC, list = FALSE)
data_train <- data_train %>% select(-c(Class))

mycontrol <- trainControl(method = "cv",
                           number = 5,
                           classProbs = TRUE,
                           savePredictions = "final",
                           allowParallel = T )
```

```

registerDoParallel(cores = 2)
model_list <- caretList(FLAG_VIOL_INC ~ .,
  data = as.data.frame(data_train),
  trControl = mycontrol,
  metric = "ROC",
  tuneList = list(
    rf = caretModelSpec(method = "rf", nodesize = 10,
      tuneLength = 15),
    svmPoly = caretModelSpec(method = "svmPoly",
      tuneLength=10,
      tuneGrid = data.frame(degree = 1,
        scale = 1,
        C = 1),
      preProcess = c("pca", "scale", "center")),
    nnet = caretModelSpec( method = "nnet", tuneLength=6,
      trace=FALSE,
      preProcess = c('center', 'scale'))))

ensem_model <- caretEnsemble(model_list)
## save the model as an RDS file
fname <- "./Data/SchoolSafety/Model/School_ensembleModel.rds"
saveRDS(ensem_model, file = fname)

```

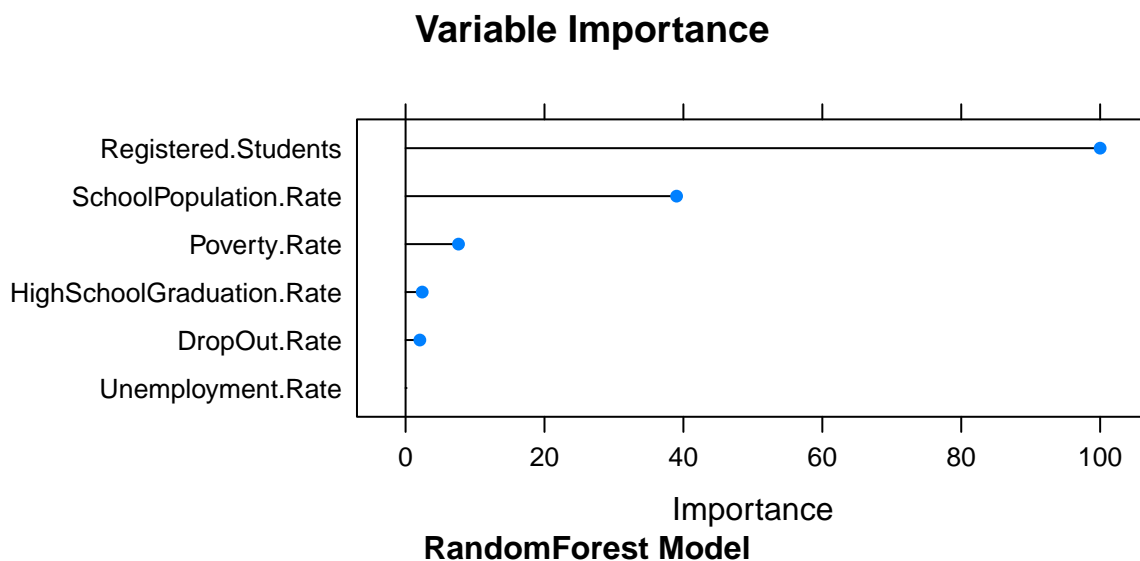
Random Forest

Random Forest is a supervised learning algorithm. To simplify, Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction [5]. One huge advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems. In our case, the school safety risk score prediction is a classification problem. Therefore, random forest is an appropriate selection of model. Using the 2013-2016 school safety statistics, the model ranks the variable importance as follows-

```

plot(varImp(ensem_model$model$rf), main = " Variable Importance", sub = "RandomForest Model")

```

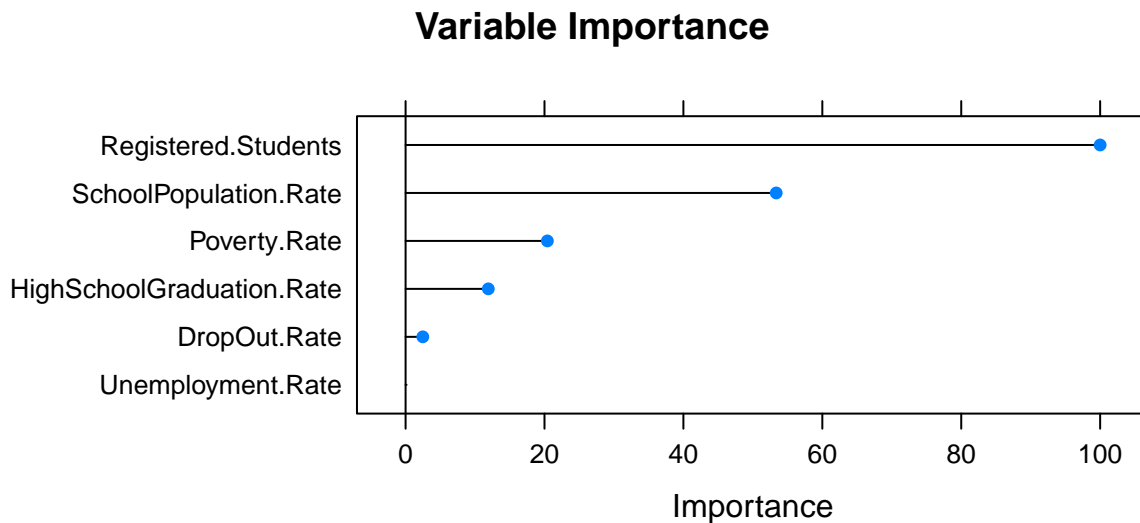


Therefore, it is safe to assume that the borough unemployment rate has no effect on school violence while random forest model is being used.

Support Vector Machines

The Support Vector Machines algorithm is implemented in practice using a kernel. The learning of the hyperplane in linear SVM is done by transforming the problem using linear algebra. “Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. A powerful insight is that the linear SVM can be rephrased using the inner product of any two given observations, rather than the observations themselves. The inner product between two vectors is the sum of the multiplication of each pair of input values[6]. Using the 2013-2016 school safety statistics, the model ranks the variable importance as follows-

```
plot(varImp(ensem_model$model$svmPoly), main = " Variable Importance",  
     sub = "Support Vector Machines with Polynomial Kernel")
```



Support Vector Machines with Polynomial Kernel

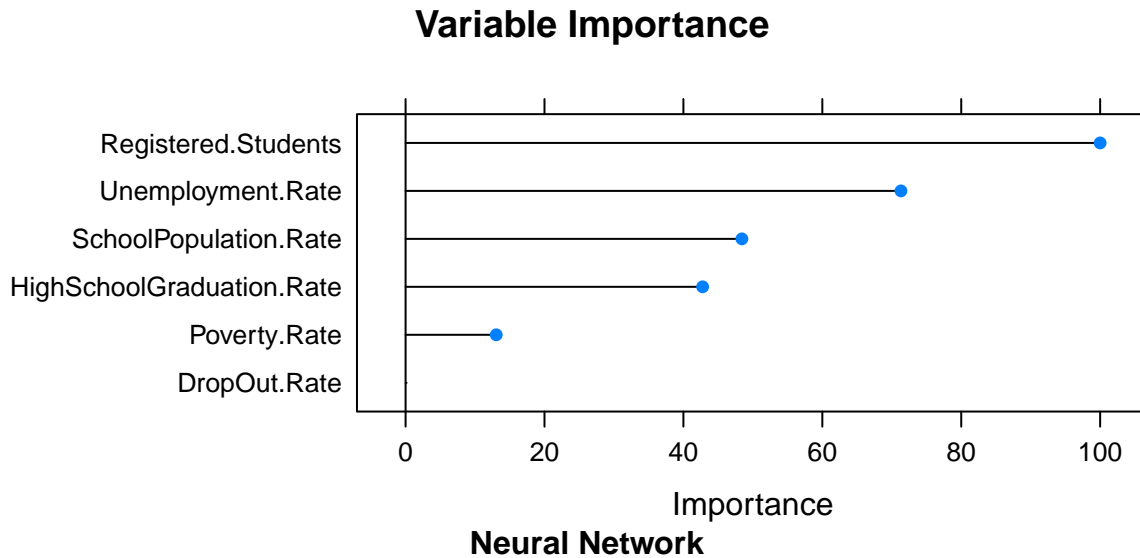
Therefore, it is safe to assume that the borough unemployment rate has no effect on school violence while support vector machines model is being used. However, all the other variables have similar importance levels when comparing random forest and svm model together.

Neural Network

Artificial Neural Networks are relatively crude electronic models based on the neural structure of human brain. Our brains fundamentally learns from experience. It is a natural proof that some problems that are beyond the scope of current computers are indeed solvable by small energy efficient packages. This brain modeling also promises a less technical way to develop machine solutions[8]. In supervised training as the school safety model, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the system, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually tweaked. The set of data which enables the training is called the “training set.”

In this case, it is the 2013-2016 school years’ dataset. During a network training, the same set of data is processed many times as the connection weights are refined. Hence, the model ranks the variable importance as follows-

```
plot(varImp(ensem_model$model$nnet), main = " Variable Importance", sub = "Neural Network")
```

Therefore, it is safe to assume that the dropout rate has no effect on school violence while using neural network model. However, all the variables have very different importance levels i.e. correlation to the data when all three models are compared against each other.

Results & Findings

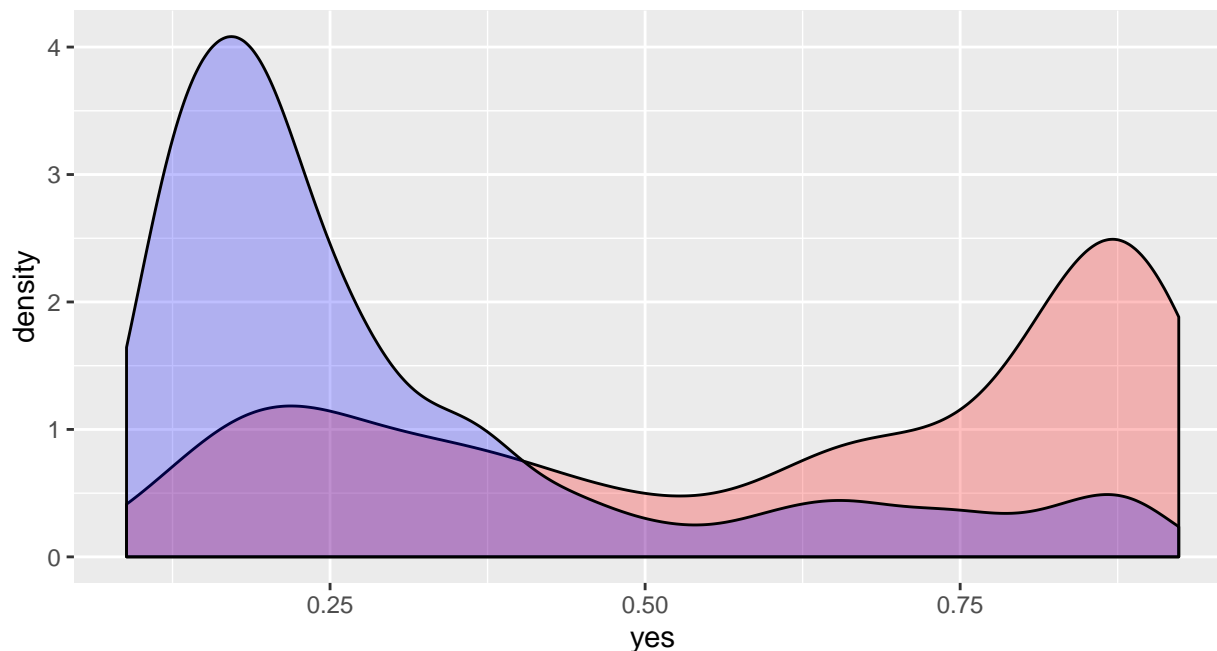
In order to measure the model performance, the school safety model was tested against a brand new data set which was never part of the initial model training set. For testing the model, the 2016-2017 school year statistics were used. This dataset includes all the NYC borough school information along with crime statistics for the school year 2016-2017. There were around 2000 observations in this test dataset. There are several functions that can be used to describe the performance of any classification models. The following is the model prediction for the test dataset as “yes” meaning the schools will encounter violent crimes in near future which is in this case within the same school year.

```
#school safety model
set.seed(2019, sample.kind="Rounding")
SCHOOL_MODEL <- readRDS("C:/Users/razad/workspace-Shiny/EdxCapStone/Data/SchoolSafety/Model/School_en

# generate test set using the school years 2016-2017
SCHOOL_DATA_TEST <- school_data %>%
  filter(School.Year == '2016-17') %>%
  select(-School.Year)

ensem_pred_yes <- predict(SCHOOL_MODEL, newdata = SCHOOL_DATA_TEST, type="prob")
ensem_probs <- data.frame(1-ensem_pred_yes, ensem_pred_yes)
colnames(ensem_probs) <- c("yes", "no")

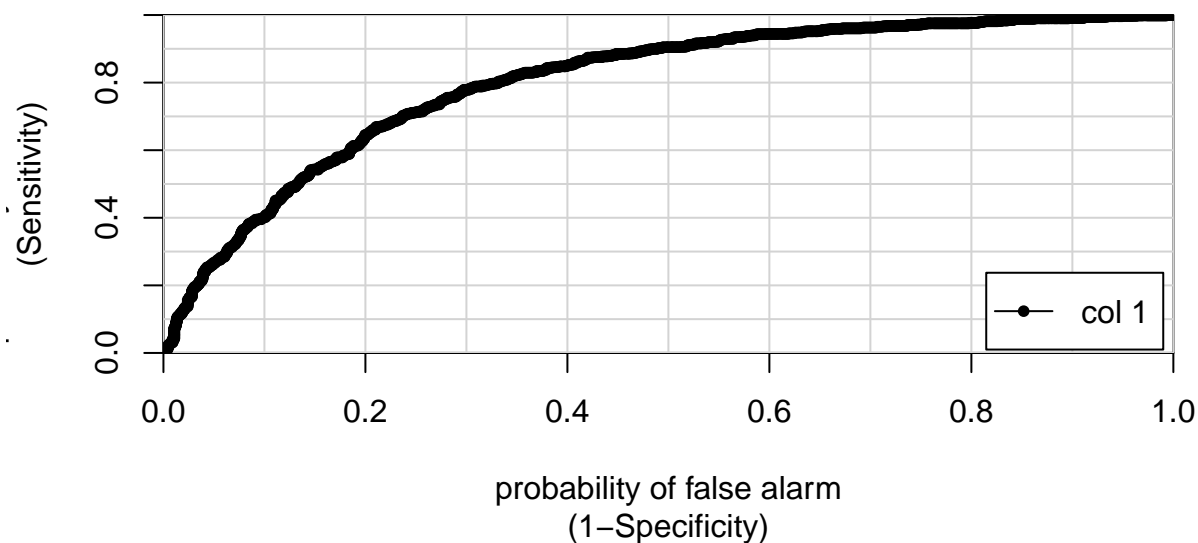
ggplot(ensem_probs, aes(x=yes)) +
  geom_density(data = ensem_probs[SCHOOL_DATA_TEST$FLAG_VIOL_INC == "yes",],
    alpha=0.25, fill = rgb(1,0,0)) +
  geom_density(data = ensem_probs[SCHOOL_DATA_TEST$FLAG_VIOL_INC == "no",],
    alpha=0.25, fill = rgb(0,0,1))
```



The very straightforward measure of model performance is Receiver Operating Characteristic (ROC). The area under the ROC curve can be used as an evaluation metric to compare the efficiency of the models. With the test data set, the ROC turned out to be 80.18%. The following is the ROC curve of the school safety model.

```
### AUC value on the test data set
colAUC(ensem_probs$yes, SCHOOL_DATA_TEST$FLAG_VIOL_INC, plotROC = TRUE)
```

ROC Curves



```
## [1]
## yes vs. no 0.8017733
```

The confusionMatrix is used to compute various summaries for this classification model. However, the threshold value is a major factor in this regard. A comparison for three different threshold values are discussed below. The first confusion matrix is calculated having the threshold value set to 0.325 meaning if the probability score is greater than 0.325, the model will consider that to be a “yes” predictor. Therefore, model accuracy as 73.41% and balanced accuracy is 73.07%.

```
pred_class <- ifelse(ensem_probs$yes > 0.325, "yes", "no")
caret::confusionMatrix(as.factor(pred_class), SCHOOL_DATA_TEST$FLAG_VIOL_INC, positive = "yes")

## Confusion Matrix and Statistics
##
##           Reference
## Prediction yes  no
##           yes 903 232
##           no  304 577
##
##           Accuracy : 0.7341
##           95% CI : (0.7143, 0.7533)
##           No Information Rate : 0.5987
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.4547
##
##           Mcnemar's Test P-Value : 0.002164
##
##           Sensitivity : 0.7481
##           Specificity : 0.7132
##           Pos Pred Value : 0.7956
##           Neg Pred Value : 0.6549
##           Prevalence : 0.5987
##           Detection Rate : 0.4479
##           Detection Prevalence : 0.5630
##           Balanced Accuracy : 0.7307
##
##           'Positive' Class : yes
##
```

The following confusion matrix is calculated by setting the threshold to 0.375. Hence, the model accuracy has lowered to 73.21%. However, the balanced accuracy has increased to 73.88%. Moreover, the number of schools that were predicted as not having any threats to violence which actually encountered violence (false negative) has decreased from 232 to 184. Conversely, the number of schools that were predicted as a threat to violence which in reality did not experience violence (false positive) has increased from 304 to 356.

```
#Calculate confusion matrix using threshold = 0.375
pred_class <- ifelse(ensem_probs$yes > 0.375, "yes", "no")
caret::confusionMatrix(as.factor(pred_class), SCHOOL_DATA_TEST$FLAG_VIOL_INC, positive = "yes")

## Confusion Matrix and Statistics
##
##           Reference
## Prediction yes  no
##           yes 851 184
##           no  356 625
```

```
##
##          Accuracy : 0.7321
##          95% CI : (0.7122, 0.7514)
##    No Information Rate : 0.5987
##    P-Value [Acc > NIR] : < 2.2e-16
##
##          Kappa : 0.4614
##
##    McNemar's Test P-Value : 1.858e-13
##
##          Sensitivity : 0.7051
##          Specificity : 0.7726
##    Pos Pred Value : 0.8222
##    Neg Pred Value : 0.6371
##    Prevalence : 0.5987
##    Detection Rate : 0.4221
##    Detection Prevalence : 0.5134
##    Balanced Accuracy : 0.7388
##
##    'Positive' Class : yes
##
```

The following confusion matrix is calculated by setting the threshold to 0.45. Hence, the model accuracy has lowered to 71.92% and the balanced accuracy has also lowered to 73.60%. On the other hand, the number of schools that were predicted as not having any threats to violence which actually encountered violence (false negative) has decreased from 184 to 145. Conversely, the number of schools that were predicted as a threat to violence which in reality did not experience violence (false positive) has increased from 356 to 421.

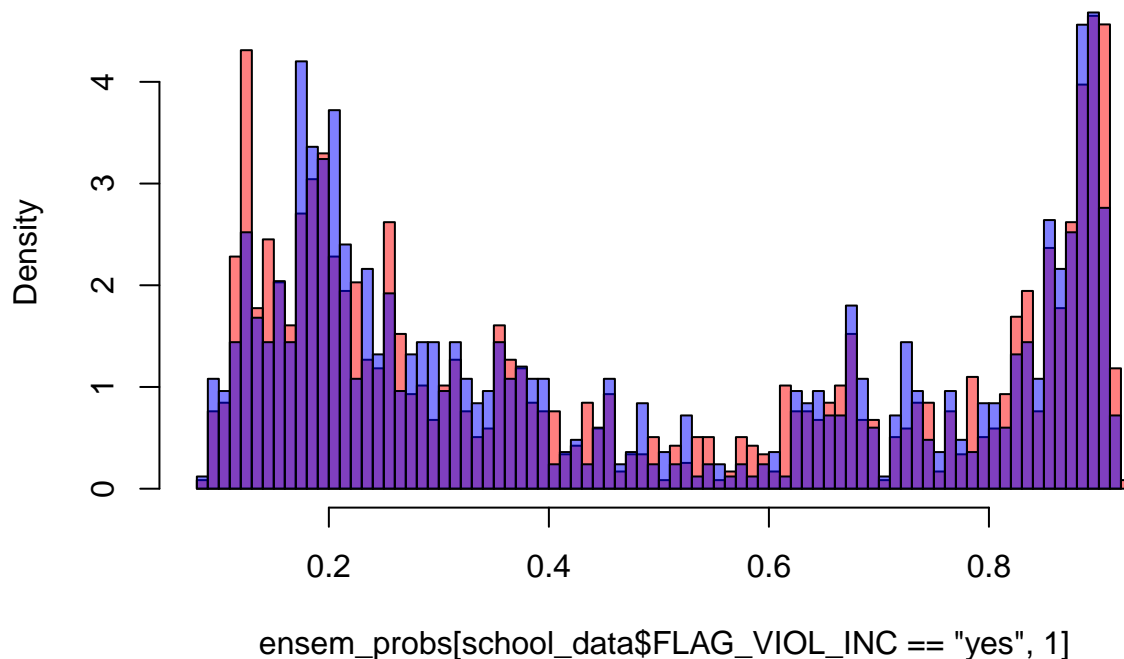
```
#Calculate confusion matrix using threshold = 0.45
pred_class <- ifelse(ensem_probs$yes > 0.45, "yes", "no")
caret::confusionMatrix(as.factor(pred_class), SCHOOL_DATA_TEST$FLAG_VIOL_INC, positive = "yes")
```

```
## Confusion Matrix and Statistics
##
##          Reference
## Prediction yes  no
##          yes 786 145
##          no  421 664
##
##          Accuracy : 0.7192
##          95% CI : (0.6991, 0.7388)
##    No Information Rate : 0.5987
##    P-Value [Acc > NIR] : < 2.2e-16
##
##          Kappa : 0.4468
##
##    McNemar's Test P-Value : < 2.2e-16
##
##          Sensitivity : 0.6512
##          Specificity : 0.8208
##    Pos Pred Value : 0.8443
##    Neg Pred Value : 0.6120
##    Prevalence : 0.5987
```

```
##          Detection Rate : 0.3899
##    Detection Prevalence : 0.4618
##          Balanced Accuracy : 0.7360
##
##          'Positive' Class : yes
##
```

Therefore, it is a matter of user priority and resource allocation, which would determine whether we want threshold value to go higher or lower. Clearly the authorities would not want to over allocate their resources for the schools having false positive predictions whether as they should not let their guards down for the schools with false negative predictions.

Histogram of `ensem_probs[school_data$FLAG_VIOL_INC == "yes",`



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated distribution of prediction on the 2016-17 school year data.

For the time being, we assume that the best threshold value for this model is 0.315 with accuracy rate of 73.81% and a balanced accuracy rate of 73.24%. The number of schools that were predicted as not having any threats to violence which actually encountered violence (false negative) is 240. Conversely, the number of schools that were predicted as a threat to violence which in reality did not experience violence (false positive) is 288.

```
## Best threshold value is determined based on Balanced Accuracy
## meaning if the probability of violence is greater than 0.315,
## only then the school will be listed as a potential threat in violence in this school year

threshold <- 0.315
pred_class <- ifelse(ensem_probs$yes > threshold, "yes", "no")
caret::confusionMatrix(as.factor(pred_class), SCHOOL_DATA_TEST$FLAG_VIOL_INC, positive = "yes")
```

```

## Confusion Matrix and Statistics
##
##           Reference
## Prediction yes  no
##           yes 919 240
##           no  288 569
##
##           Accuracy : 0.7381
##           95% CI : (0.7183, 0.7572)
##           No Information Rate : 0.5987
##           P-Value [Acc > NIR] : < 2e-16
##
##           Kappa : 0.4602
##
## Mcnemar's Test P-Value : 0.04081
##
##           Sensitivity : 0.7614
##           Specificity : 0.7033
##           Pos Pred Value : 0.7929
##           Neg Pred Value : 0.6639
##           Prevalence : 0.5987
##           Detection Rate : 0.4559
##           Detection Prevalence : 0.5749
##           Balanced Accuracy : 0.7324
##
##           'Positive' Class : yes
##

```

For any classification models, the functions `sensitivity`, `specificity`, `posPredValue` and `negPredValue` can be used to characterize model performance where there are only two classes in terms of 1 (true) and 0 (false). Moreover, the `confusionMatrix` frames the errors in terms of sensitivity and specificity. In this model, the true value represents that the schools will experience any violent crimes within the school year and false represents that they will not.

Sensitivity is the percentage of actual 1's that were correctly predicted. It shows what percentage of 1's were covered by the model. In this case, the 1's represent that the schools will experience violence within the same school year and the value is 76.14%. Sensitivity matters more when classifying the 1's correctly is more important than classifying the 0's. Just like what we need here in school safety case, where you do not want to miss out any possible violence at school to be classified as no or low risk.

Likewise, Specificity is the proportion of actual 0's that were correctly predicted. In this case, the 0's represent that the schools will not experience any violence and the value is 70.33%. Maximizing specificity is more relevant in cases like email spam detection, where you strictly do not want genuine messages (0's) to end up in spam (1's). However, in school safety, identifying a low risk school as a high risk for potential violence could be waste of resources yet not life threatening.

Challenges and limitations

Currently the school safety model is trained using three school years' (2013-2016) worth of data for all schools in NYC boroughs. The dataset includes around 8500 records. For each school year, there was only a single record available per school, meaning the crime statistics data was available as yearly basis. If the school crime statistics were accessible more frequently i.e. monthly basis, the model predictions would have been more accurate. Better yet, instead of predicting the risk of violence within the same school year, the model could have predicted for the next 30 or 60 days.

Moreover, as it is mentioned earlier in this paper, that school quality is a factor which influences the school violence; could be an explanatory variable to the model. However, school quality report was not available for all the school years in question. For some school years, it was only available for high schools. Therefore, it could not be used as a predictor for school safety risk.

Conclusion

Even when a machine learning system is no more, or less, prone to error than a human performing the same task, relying on the machine can feel uncertain because it raises new concerns [10]. When any errors are made or at the time of system failure, who do we assign accountability? If any support is provided by the system designers for adjustments, upgrades, and maintenance. These concerns are especially critical for the fields such as medicine, spacecraft, finance, and real-time systems, or safety and security. However, it is always advisable to implement machine learning models to solve real life problems if the data is already been collected. In this case, the Department of Education has already collected hundreds of thousands of statistics on school safety. These models can be a useful tool for the authority to assist in difficult decision making and ensure school safety.

Reference

- 1) Tillery, A. D., Varjas, K., Roach, A. T., Kuperminc, G. P., & Meyers, J. The importance of adult connections in adolescents' sense of school belonging: Implications for schools and practitioners. *Journal of School Violence*, 2013
- 2) <http://www.crf-usa.org/school-violence/causes-of-school-violence.html>
- 3) Conoley, J. C., & Goldstein, A. P. The known, unknown, and future of violence reduction. In A. P. Goldstein & J. C. Conoley (Eds.), *School violence intervention: A practical handbook*. New York, NY: Guilford Press, 1997
- 4) Zhi-Hua Zhou, "Ensemble Methods: Foundations and Algorithms", CRC Press, 2012
- 5) Niklas Donges, "The Random Forest Algorithm", SAP Machine Learning Foundation, 2015
- 6) Jason Brownlee, "Support Vector Machines for Machine Learning", *Machine Learning Mystery*, 2016
- 7) Pauziah Mohd Arsad, Norlida Buniyamin, Jamalul-lail Ab Manan, *Neural Network and Linear Regression methods for prediction of students' academic achievement*, IEEE, 2014
- 8) Sonali. B. Maind, Priyanka Wankar, *Research Paper on Basic of Artificial Neural Network*, *International Journal on Recent and Innovation Trends in Computing and Communication*, ISSN: 2321-8169
- 9) Avneet Pannu, *Artificial Intelligence and its Application in Different Areas*, *International Journal of Engineering and Innovative Technology*, 2015
- 10) Kiri L. Wagstaf, *Machine Learning that Matters*, California Institute of Technology, 2010