Machine-in-the-Loop: a Machine Learning Exploration

Ping Ng Chong, Haiyao Ni, Ethan Paulsen

April 16, 2019

Motivation

The motivation behind this project is being able to parse through data generated data from four satellites around the globe and determine which data should be downloaded back down to Earth for further study. Since the orbits of the satellites only allow for 2% to 4% of the data to be transmitted back to Earth, a machine learning method needs to be implemented in order for the data to be parsed in real-time, and eliminate the need for a scientist-in-the-loop, or SITL.

Related Works

Evaluation Criteria

Methods Evaluated

```
  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
```

```
  data <- read.csv("spaceDataset.csv")
  test <- read.csv("testset1.csv")
  data.frame <- data.frame(data)
  data.frame$X1 <- 1:nrow(data.frame)
  DES.N <- data.frame$DES.N
  FGM.Bt <- data.frame$FGM.Bt
  DIS.N <- data.frame$DIS.N
  DES.T_para <- data.frame$DES.T_para
  DES.T_perp <- data.frame$DES.T_perp
  Selected <- data.frame$Selected
```

```
  fit <- lm(Selected ~ DES.N + FGM.Bt + DES.T_para + DES.T_perp, data = data.frame)
  predict_lm <- predict(fit, test, interval = "predict")
```

```
  glm.fit <- glm(Selected ~ DES.N + FGM.Bt + DES.T_para + DES.T_perp, family = binomial(), data = data.
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

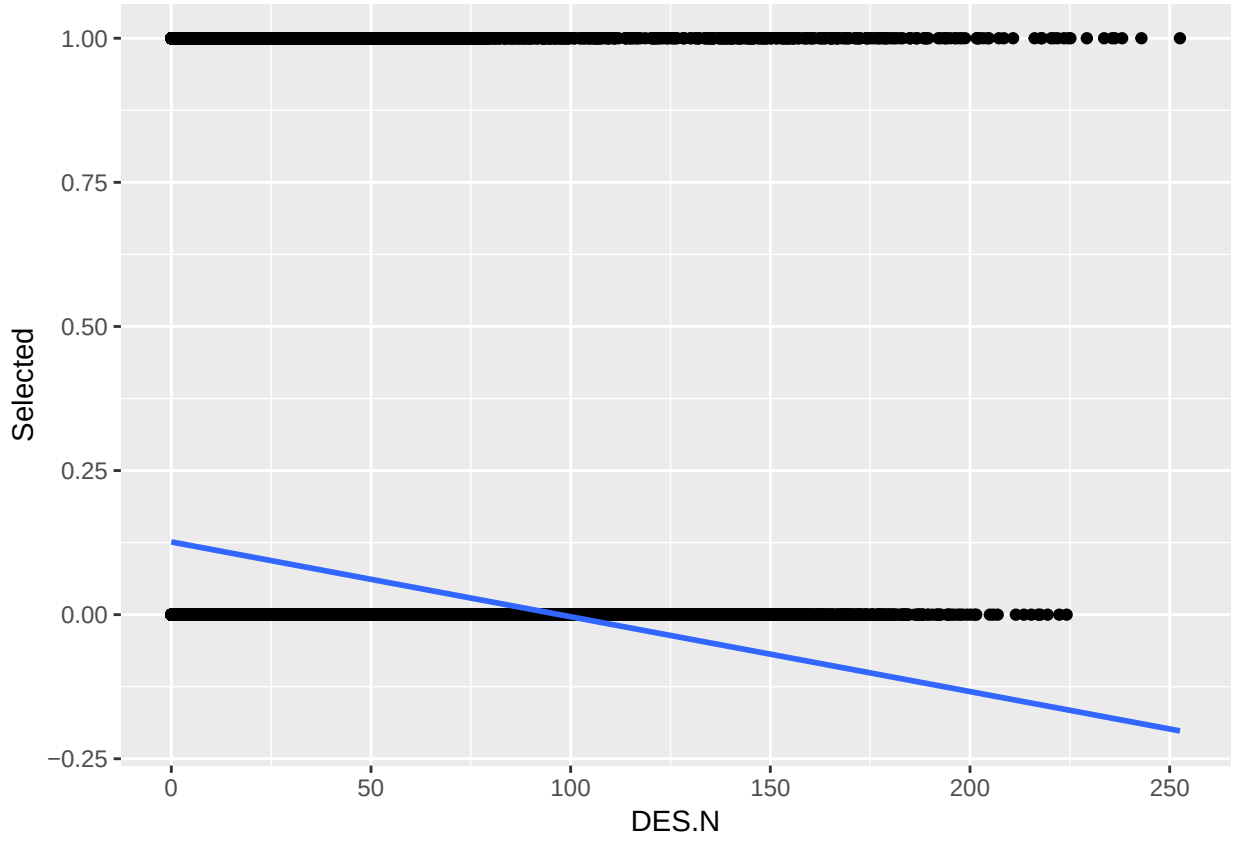
```
  summary(glm.fit)
```

```
##
## Call:
## glm(formula = Selected ~ DES.N + FGM.Bt + DES.T_para + DES.T_perp,
##     family = binomial(), data = data.frame)
```

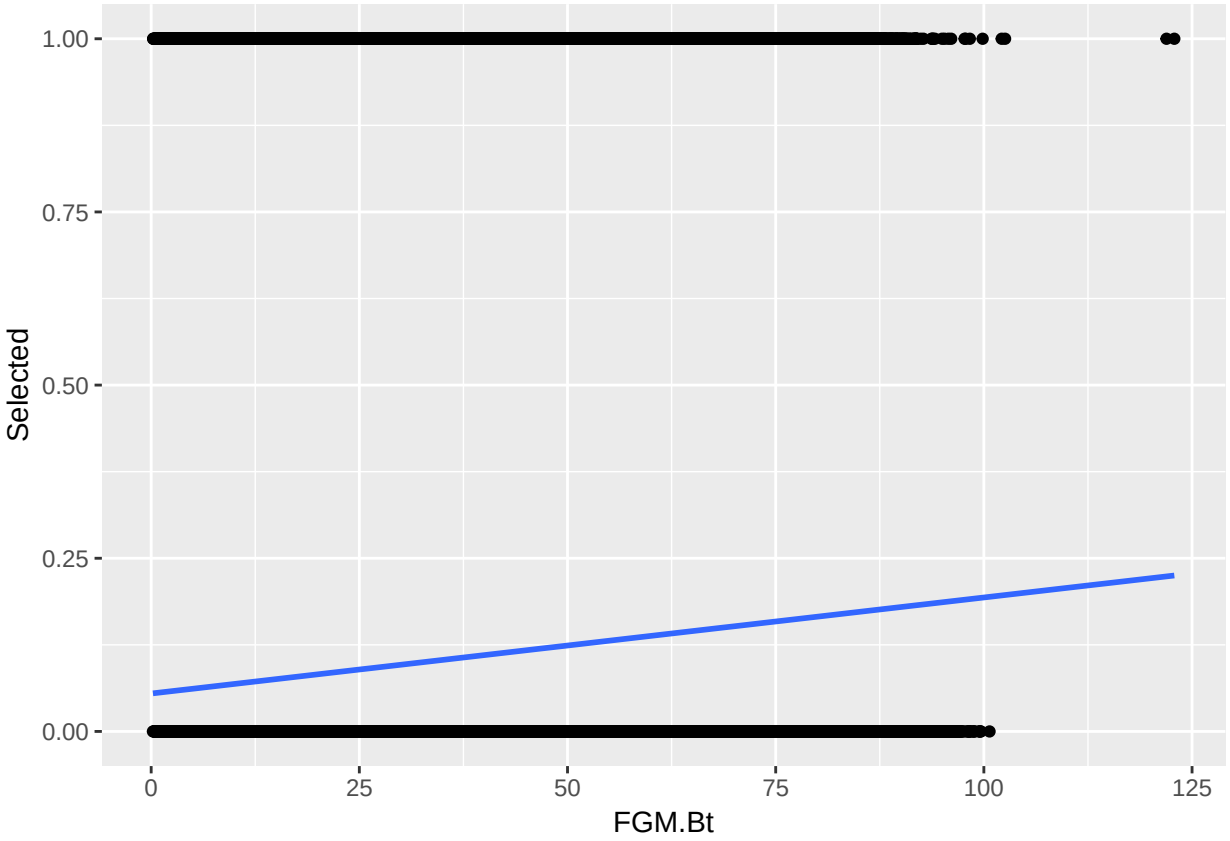
```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5683  -0.4784  -0.4100  -0.2898   5.8152
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2321658  0.0169617 -131.60   <2e-16 ***
## DES.N       -0.0249200  0.0004432  -56.23   <2e-16 ***
## FGM.Bt       0.0237496  0.0003363   70.62   <2e-16 ***
## DES.T_para   0.0136927  0.0002253   60.78   <2e-16 ***
## DES.T_perp  -0.0150197  0.0002079  -72.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 256842  on 395457  degrees of freedom
## Residual deviance: 236447  on 395453  degrees of freedom
## AIC: 236457
##
## Number of Fisher Scoring iterations: 6
```

```
predict_glm <- predict(glm.fit, test, interval = "predict")
```

```
library(ggplot2)
ggplot(data.frame, mapping = aes(x = DES.N, y = Selected)) + geom_point() + geom_smooth(method = "lm"
```



```
ggplot(data.frame, mapping = aes(x = FGM.Bt, y = Selected)) + geom_point() + geom_smooth(method = "gl
```



Recommended Methods

Analysis of Results

NOTES

Try using these functions (using DES.N and FGM.Bt): - PCA - K-Means - SVM