Name: Md Imanur Rahman Emon ID: 1813181642

Course: CSE445 Section: 03

Name: Md. Imanur Rahman Emon
Id: 1813181642
Sec: 03
Course: CSE445
Email: manur.emon@northsouth.edu
Semester: spring 21

Ans: to the qu:no (1)

a) Batch gradient Descent: Batch gradient Descent deals with whole numbers of sample or training data. To find out slope we need to measure all Θ's value.

Pros: It produce a stable gradient descent where we find a stable error. It is useful for all resources where it deals with not with a single sample rather all the training dataset.

Cons: It may take a local minimum as a global minimum. If the training set is too large it may take a huge time. also it has a low learning rate.

Stochastic Gradient Descent: Here we take the Θ's value (1) randomly from the training data. We randomly chose one iteration and find the slope.

Pros: For choosing the Θ's value randomly it take a small time for that the learning rate is very fast.

Name: Md. Imanur Rahman Emon
Id: 1813181642
Sec: 03
course: CSE445

Cons: It will take a random value if the value is not right than the variance will high. ~~will~~ May take a wrong θ's value which can perform worst.

Mini-Batch Gradient Descent: Mini-Batch Gradient Descent take a small sub set of training sample to measure the θ's value. ultimately MSE.

Pros: It is faster than BGD. It takes average data from training sample because of that this will produce stable error. Easily fits in the memory.

Cons: It may take some data from the training data but it can be measure a local minimam as a global minimum. For this it rejects noice to other gradient descent.

Name: Md. Imanur Rahman Emon

Id: 1813181642

1) B) Momemtum is a term which used in gradeent descent algorithms. Gradient descent is a optimization algorithm which works to find the perfect slope on that point. Towards that direction in each steps, the value of function to be minimized. The problem in this direction may change gradually in some points of the function where there best path to go usually does not contain a lot of turns. So, it use so called momentum to improve convergence. Which already going along for sometime before it changes its direction.

Name: Md. Imanur Rahman Emon . Course: Cse 445

Id: 1813181642 Semester: Spring 2021

Section: 3 NSU Email: Imanur.emon@northsouth.edu

Ans: to the qu: no (2)

Here, the system is failing is A fails & atleast 2 of B, C, D fails.

So, the probability is $\frac{1}{3}$ (A fails) + 3 $(1/3)^2 \times (2/3)$,

(BD or BC or DC fail) = $\frac{1}{3} + \frac{2}{9} = \frac{5}{9} = 0.555$

Probability that only A fails $= \frac{1}{3} \times (2/3)^3 = 0.0987$

Probability that only A fails given that when system is failing $= \frac{0.0987}{0.555}$

$= 0.1778$

a) Probability that each system works is 2/3, and each fails is $\frac{1}{3}$.

So, the probability that all system work is

$= 2/3 \times 2/3 \times 2/3 \times 2/3$

Name: Md. Imanur Rahman Emon
Id: 1813181642
Course: CSE 445
Sec: 3

$= \frac{16}{81}$

$= 0.1975$

Ans:

(b) Probability that system does not work properly is $= \left(1-(2/3)^4\right)$

$= 1-(0.66)^4$

$= 1-0.197$

$= 0.803$

Probability that only A system fails is $= \frac{1}{3} \times (2/3)^3$

$= \frac{1}{3} \times \frac{8}{27}$

$= \frac{8}{81}$

$= 0.09876$

Probability that only A ~~fails given~~ fails also system does not work properly is $= \frac{0.09876}{0.803}$

$= 0.1229$

$= 0.123$

Ans:

Name: Md. Imanur Rahman Emon

Id: 1813181642

Course: CSE445

Sec: 3

(d) The same is brought out in empirical probability. A, B, C, D failures or functioning are simulated using randombetween function in the range 1 to 9. If 1 to 3 comes it's failure else success. System A only failed 98 times out of 810 improper working of system. The resulting probability is 98/810

= 0.120

Ans:

Name: Md. Imanur Rahman Emon
Id: 1813181642
Course: CSE 445
Sec: 3

Am: to the qu no (3)

Bias: In our Model when we measure the performance from the known data we trained the model, by doing this with bias. Bias is a source of error in our model it's the reason of underfitting.

~~Variance: While we train our model we took~~

Variance: The variance is an error from sensitivity with small function in the training dataset. When we testing our model from unknown data then variance comes.

Bias vs variance tradeoff deals with the prediction error via bias and varience. When testing accuracy is too bad like we find 80% accuracy while we training on the other hand when in comes of testing the model with an unknown data and shows 90% accuracy that means

Name: Md. Imarun Rahman Emon
Id : 1813181642
Course: CSE 445
Sec : 3

there's variance is very high. For measuring the model while we took a data from training data and it shows that the training accuracy is 90% and testing accuracy is 60%. it means there's data is over-fitting, In our machine learning model we need a performance where there is low bias and low variance means there is no overfitting or underfitting data. So we need to train our model with a good amount of data, and try to split the Data by 80/20 and the same time try to preprocess the data correctly. Then we will find a good model where a low bias and a low variance which will helps to predict the data/system perfectly. If not have to change the model.

Name: Md. Imanur Rahman Emon

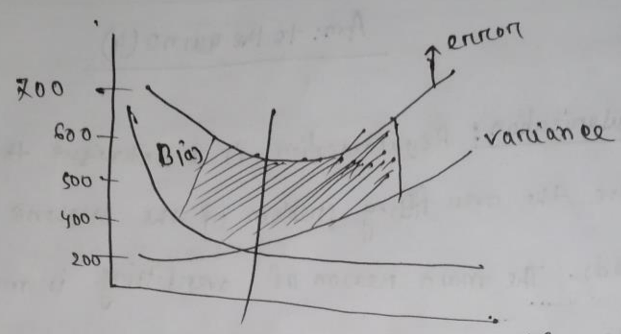
Id: 1813181642

Email: Imanur.emon@northsouth.edu

Course: CSE445

Semester: Spring 2021

Sec: 3



Bias - variance tradeoff.

Higher variance and lower bias lead to overfitting. training accuracy is good when the bias. While testing accuracy is too bad then high varience.

Name: Md. Imanur Rahman Emon
Id: 1813181642
sec: 03
course: CSE445

Ans: to the ques no (4)

Regularization: Regularization is a technique that is used to solve the over-fitting problem of the Machine learning Methods. The main reason of overfitting is making a model more complex than necessary. If we find a way to reduce the complexity, then overfitting issue is solved.

Two common methods of regularization are L1 and L2 regularization.

L1 regularization: L1 Regularization is the the penalty for lasso regression, it's cost function is defined as,

$$\min_{B} (y - XB)^T (y - XB) + \lambda \| B \|$$

This additional cost will cause the resulting weights to do this additional cost. Unfortunately L1 regularization does not have a closed form solution because it is not differentiable when a weight B falls to 0. The another

Name: Md. Imanur Rahman Emon
Id: 1813181442
sec: 03
course: CSE445

of the constraint. The contraint have very sharp edges which lie on each dimensional axis at a distance e from the origin.

L1 Regulization penalizes |weight|

Lasso regression modifies the over fitted or underfitted by adding the penalty equivalent to the sum of the absolute values of coefficients.

cost function = Loss + $\delta \times \Sigma |w|$

L2 Regularization: $\text{Min}_B \ (y-yB)^T (y-XB) + \delta \, ||B||^2$

The only difference is the added regularization term δ with in the threrse.

Ridge Regression (L2) Example:

cost function = Loss + $\delta \times \Sigma ||w||^2$

This is why regularizers are often called a penalty term.

Name: Md. Imanur Rahman Emon

Id: 1813181642

Course: CSE 445

Sec: 03

Semester: Spring 21

Ans: to the ques no ~~(4)~~ (6)

Here,

	S_1	S_2	S_3	S_4
P_i	0.03	0.62	0.26	0.09

Have to compute the entropy of the system.

We know,

$$\text{Entropy} = H(x) = \sum_{i=1}^{n} P(x_i) \log_2 \left(\frac{1}{P(x_i)} \right)$$

$$= (0.03)\log_2(1/0.03) + (0.62)\log_2(1/0.62)$$

$$+ (0.26)\log_2(0.26) + (0.09)\log_2(0.09)$$

$$= (0.03)(5.05) + (0.62)(0.689) + (0.26)(1.94)$$

$$+ (0.09)(3.47)$$

$$= 0.15 + 0.427 + 0.5 + 0.31$$

$$= 1.387$$

~~Ans~~

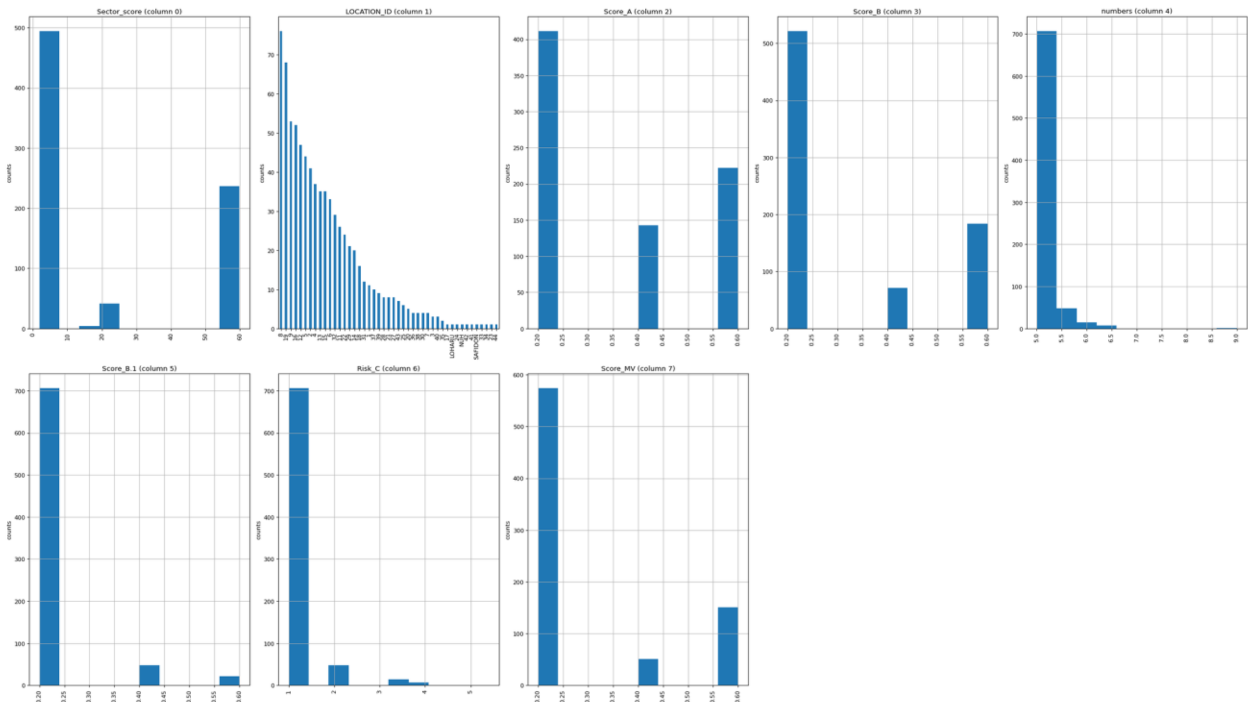
The entrophy of the system is 1.387

Ans

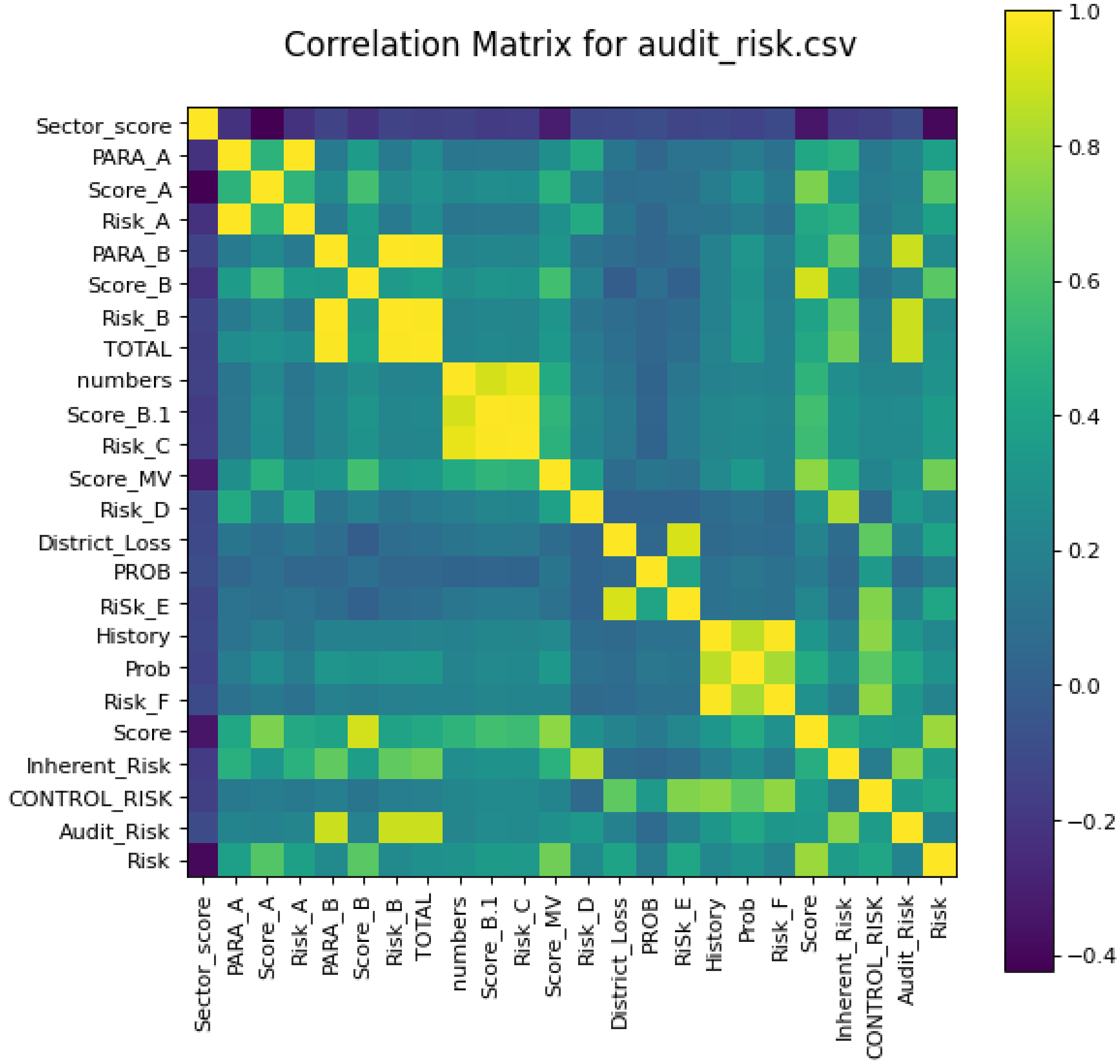
Report:

Problem 1

EDA

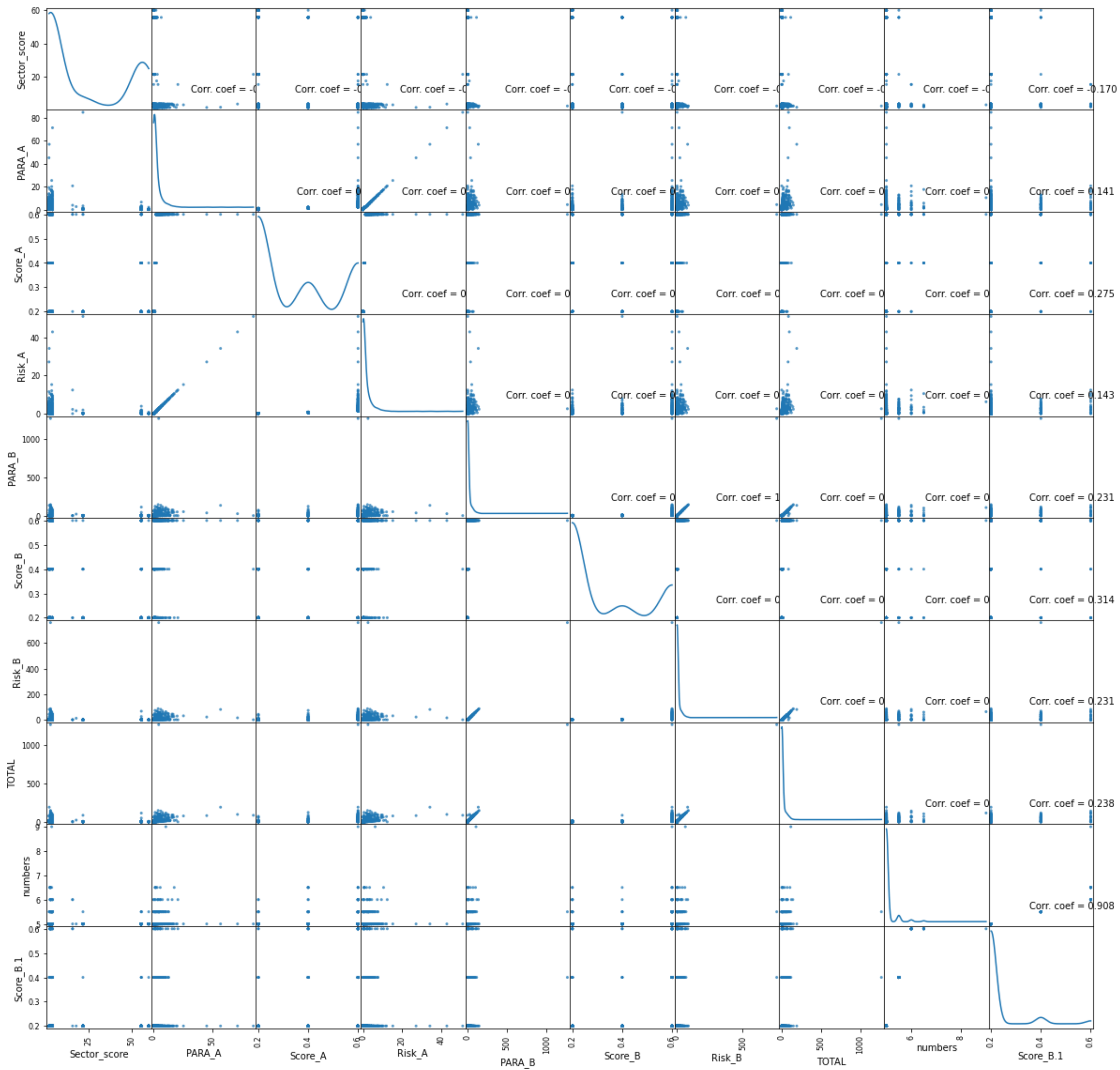


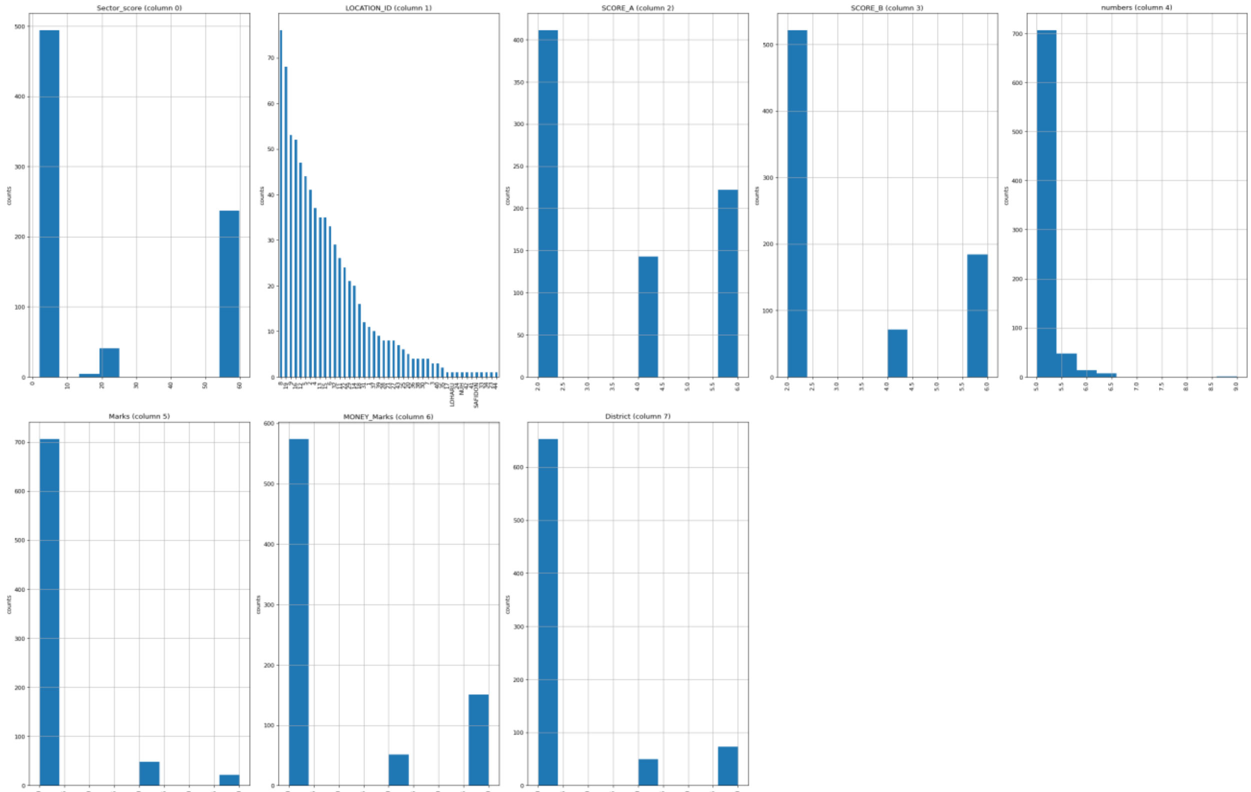
Per column Data



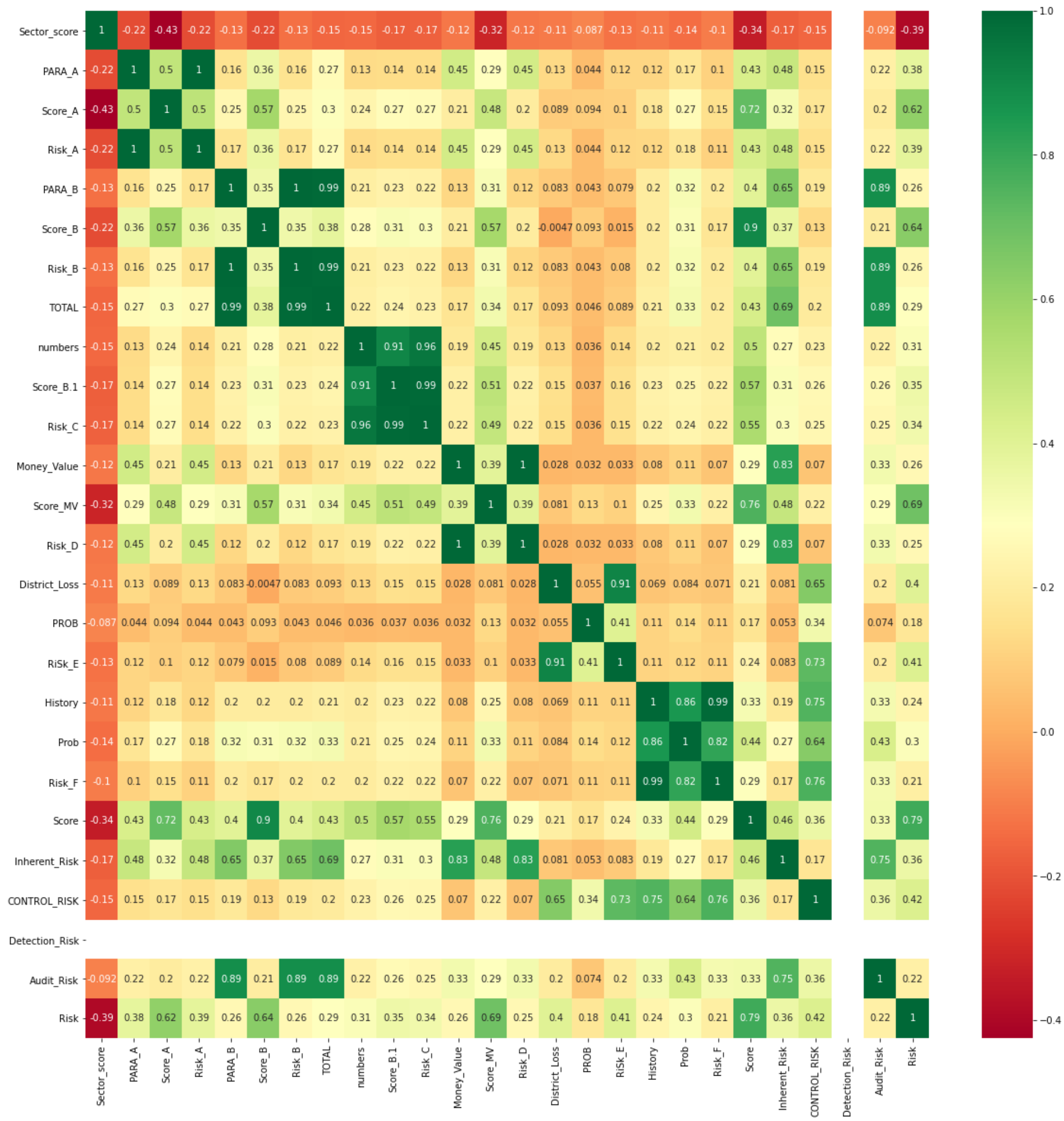
Confusion Matrix

Scatter and Density Plot





Per Column Data



Hit Map

Classifiers:

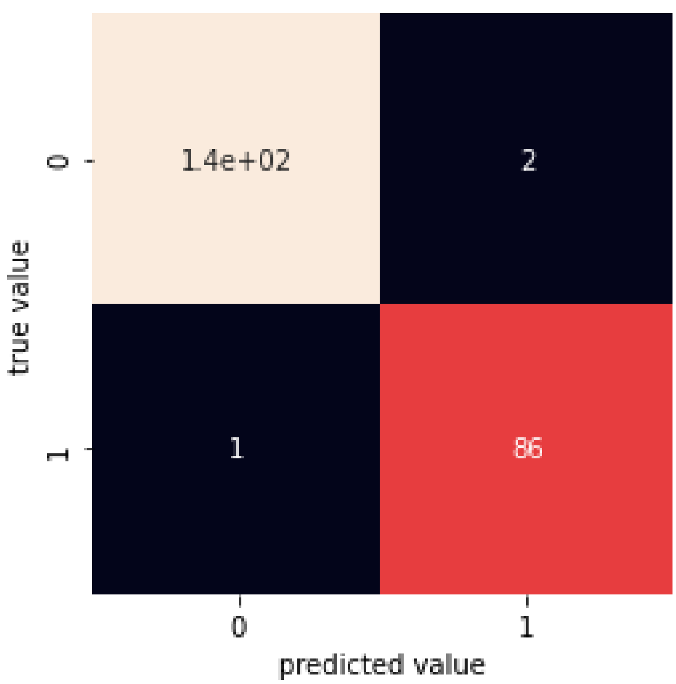
Logistic Regression

Decision Tree

Random Forrest

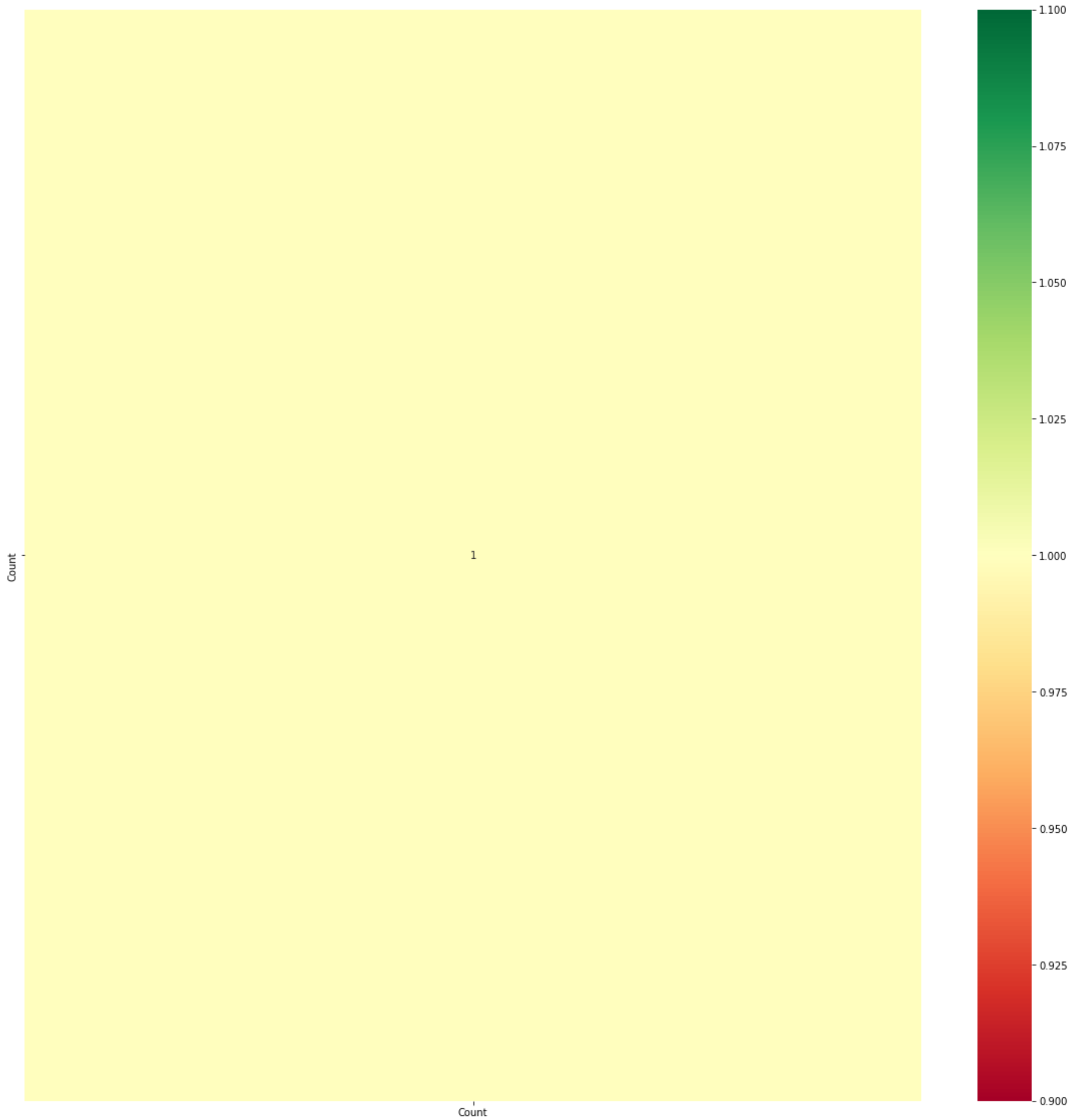
Accuracy

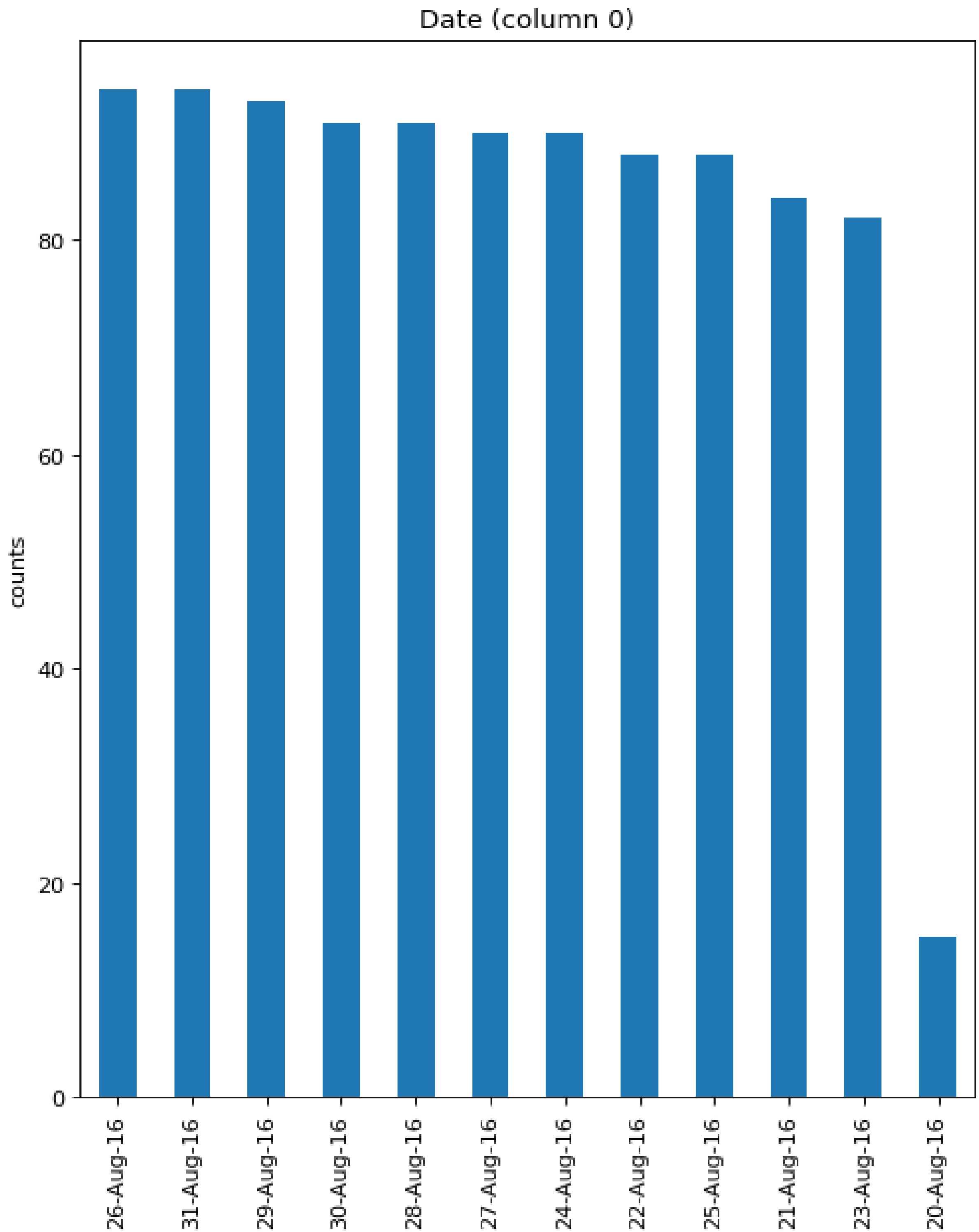
```
                 Logistic Regression: 98.71%
                 K-Nearest Neighbors: 90.13%
                       Decision Tree: 100.00%
 Support Vector Machine (Linear Kernel): 99.14%
    Support Vector Machine (RBF Kernel): 97.00%
                      Neural Network: 99.14%
                       Random Forest: 100.00%
                   Gradient Boosting: 100.00%
```



Confusion Matrix

Assignment 2:




Date (column 0)
counts
0
20
40
60
80
26-Aug-16
31-Aug-16
29-Aug-16
30-Aug-16
28-Aug-16
27-Aug-16
24-Aug-16
22-Aug-16
25-Aug-16
21-Aug-16
23-Aug-16
20-Aug-16