ANOMALY DETECTION

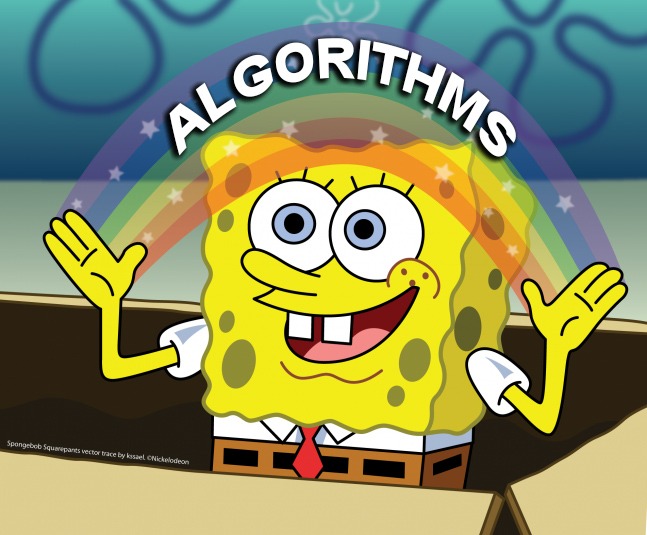
Group name: Algorithms and roses

Batıkan Özkul - 117200087

Adilcan Cebeci – 117200086

Yunus Emre Sancak – 117200067

Hüseyin Ezen-117200123



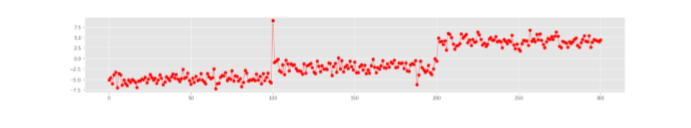
Outline:

Main idea in this Project is that we have a data structure and we need to write 2 different codes for anomally detection and use both of them on the same data and compare the results.

Introduction:

The codes we should use were determined by our teacher Uzay Çetin and he wanted us to change them from python to java.

Visual presentation of the data structure that we will be using anomally detection codes on is:



And basicly what the codes we have does is:

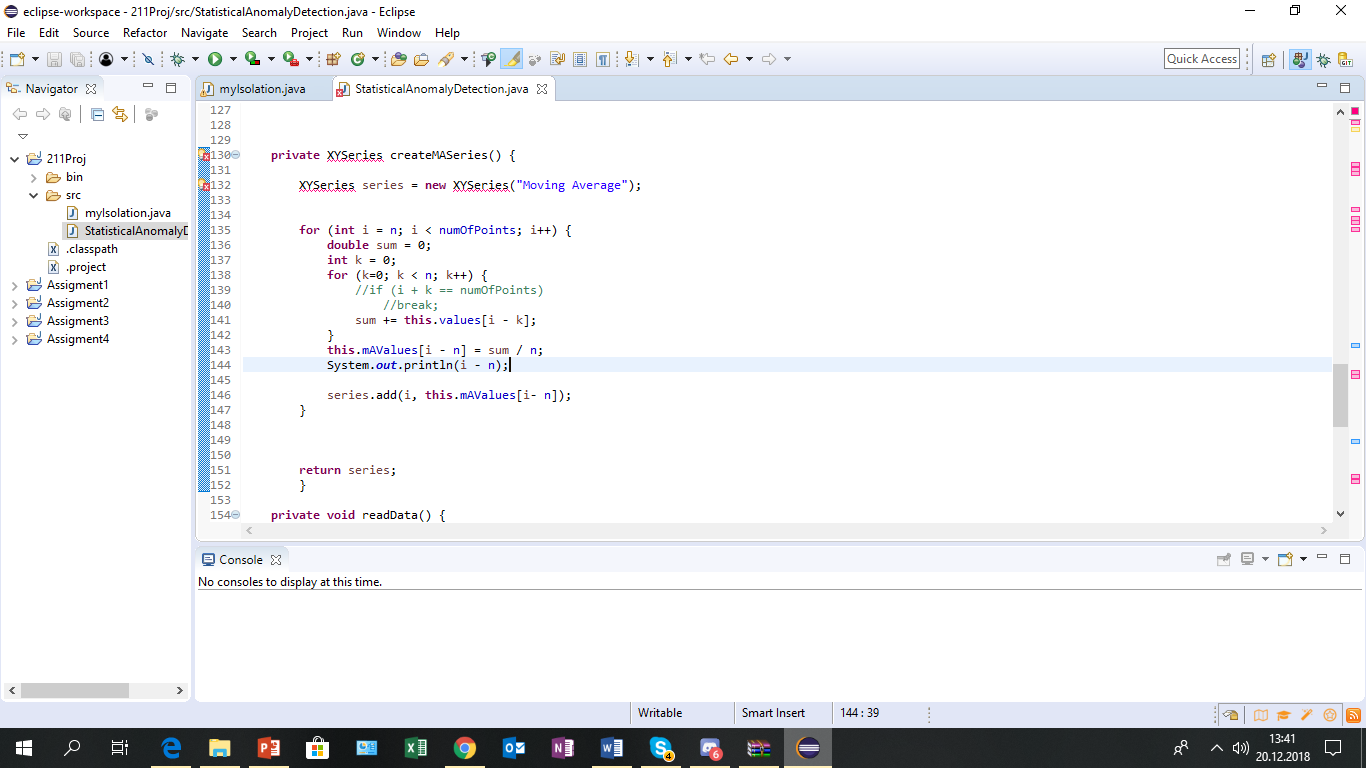
Statistical Anomally Detection:.It finds the standart deviation,mean and the moving avarage of the variables the data has and uses Mean and standart deviation to find the anomally by applying 3σ rule.

Isolation Forest: Select the point to isolate.For each feature, set the range to isolate between the minimum and the maximum.Choose a feature randomly.Pick a value that’s in the range, again randomly: If the chosen value keeps the point *above*, switch the *minimum* of the range of the feature to the value.If the chosen value keeps the point *below*, switch the *maximum* of the range of the feature to the value.Repeat these steps until the point is isolated. That is, until the point is the only one which is inside the range for all features.Count how many times you’ve had to repeat steps 3 & 4. We call this quantity the **isolation number.**

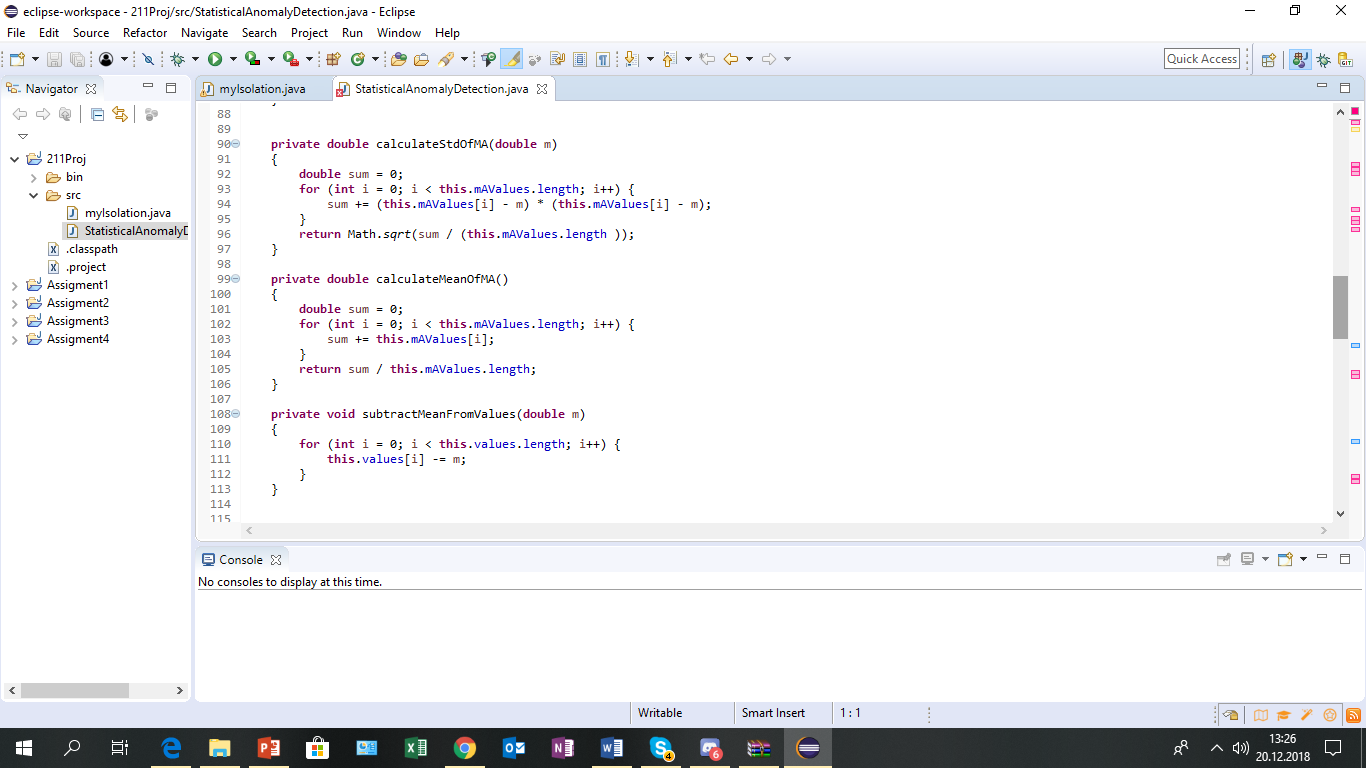
The variable with the lowest isolation number is the most easily isolated one which is the normally.

STATİSCAL ANOMALLY DETECTİON:

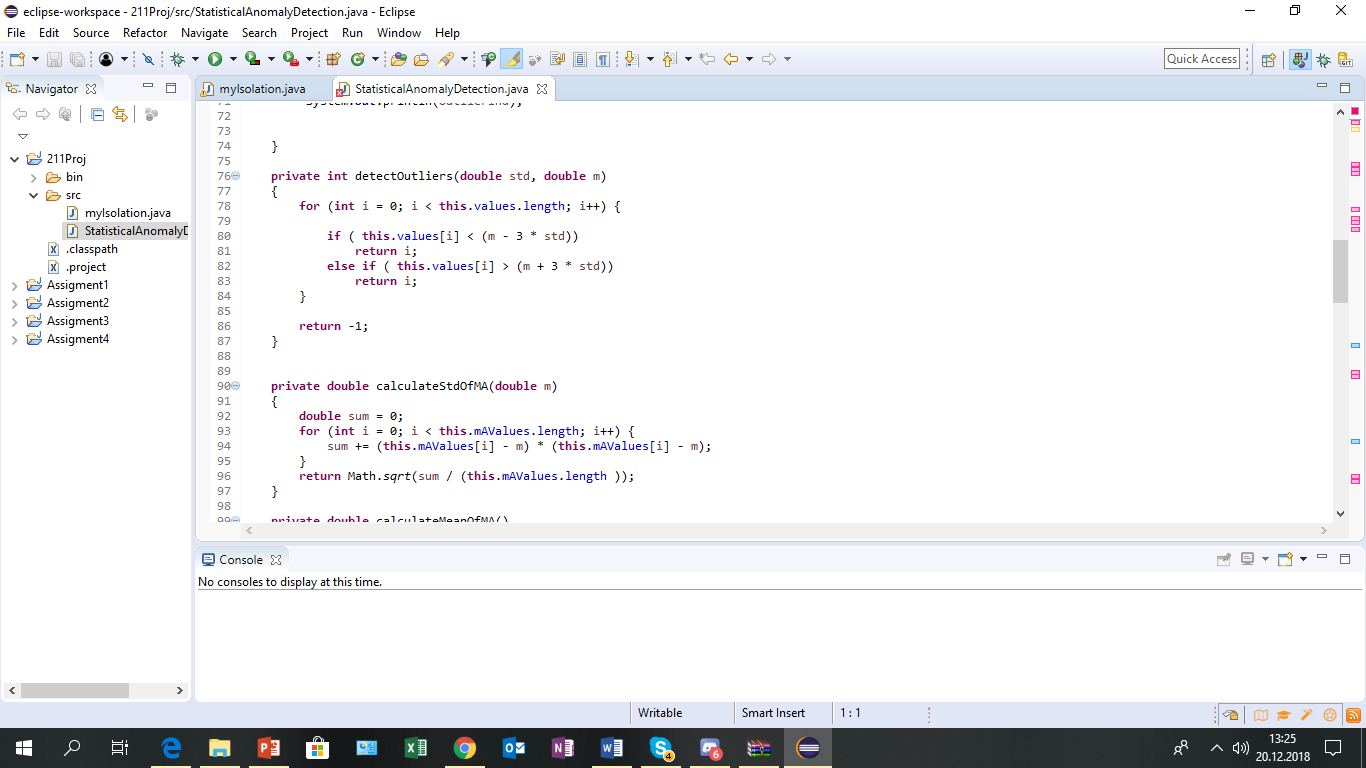
First we used File reading to take the variables from the messages.txt and turned them into a array.after turning the data into a array we wrote methods which finds the Moving avarage .



This method creates a new array which has the moving avarage of the numbers ing the message.txt file.After creating the method which takes the moving avarage we wrote methods which finds the Mean ,Standart Deviation of moving avarage array And the method which subtracts the means from the values array.



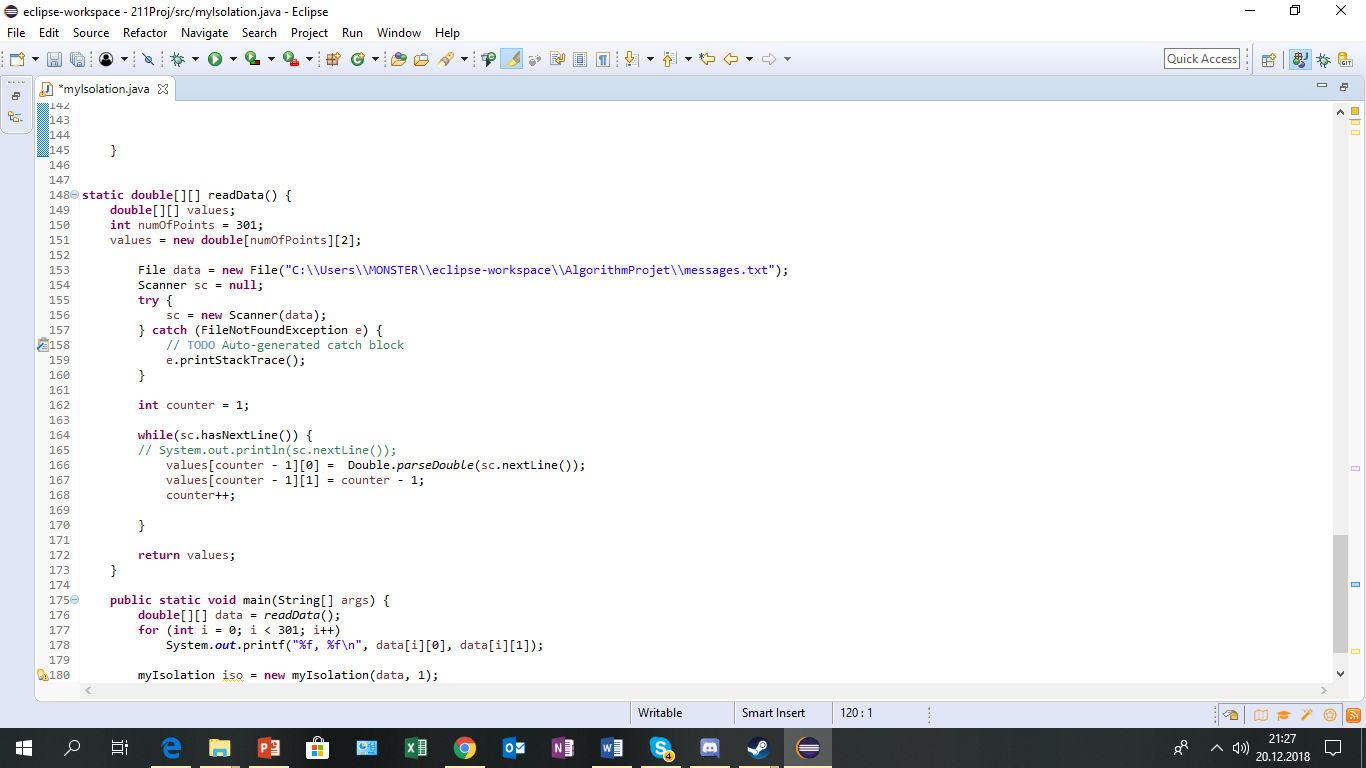
After creating these methods we created a method which applies the 3σ rule to find the anomallies.



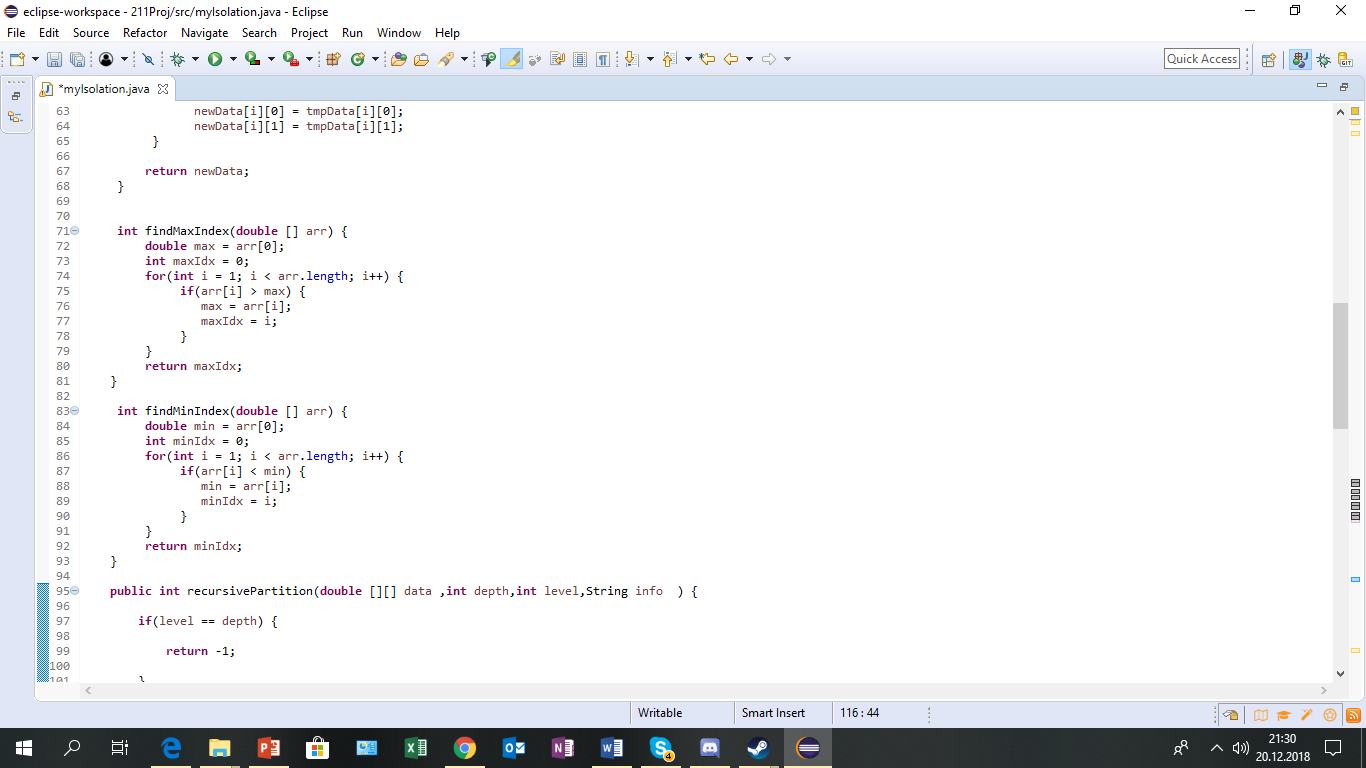
After writing all these methods we downloaded and used used javafreechart library to draw a chart.

Isolation Forest:

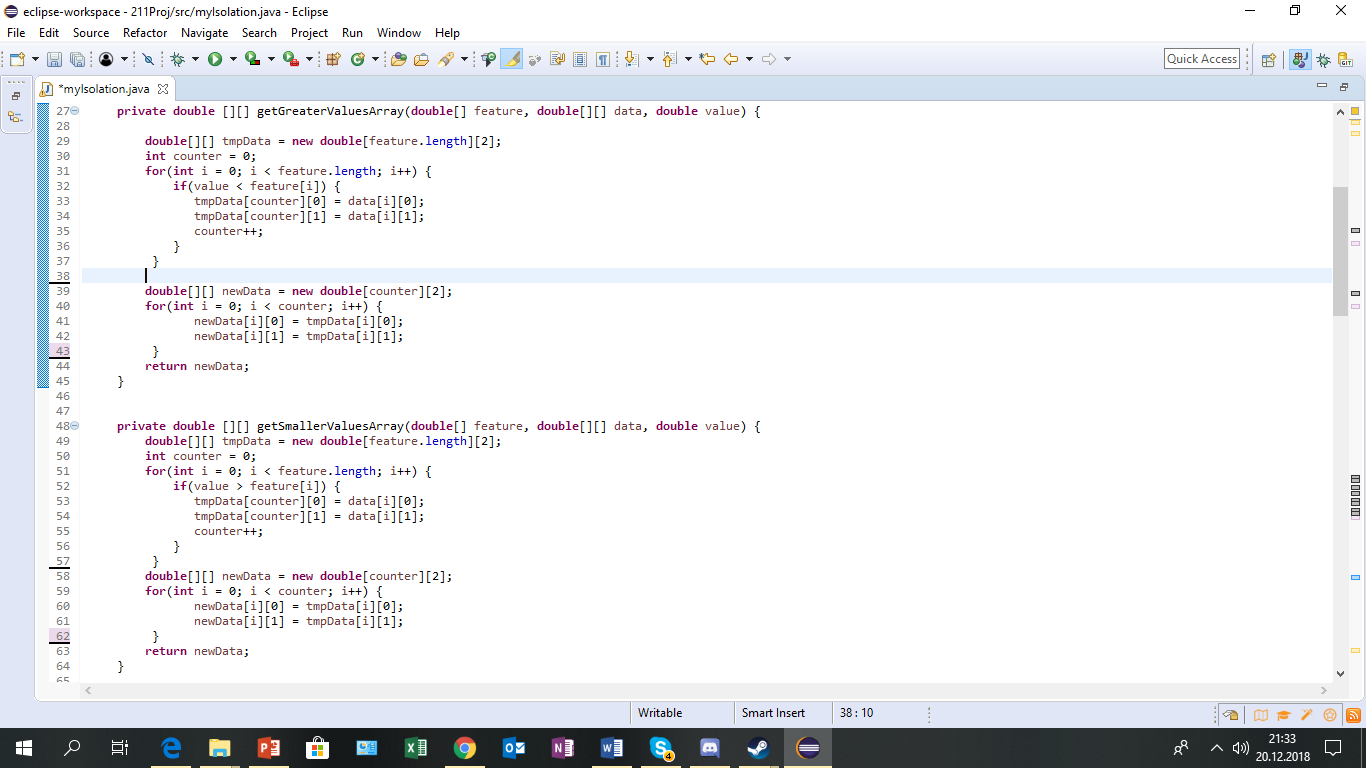
First we read the data by using file reading commands and for loops note that we are creating a double array



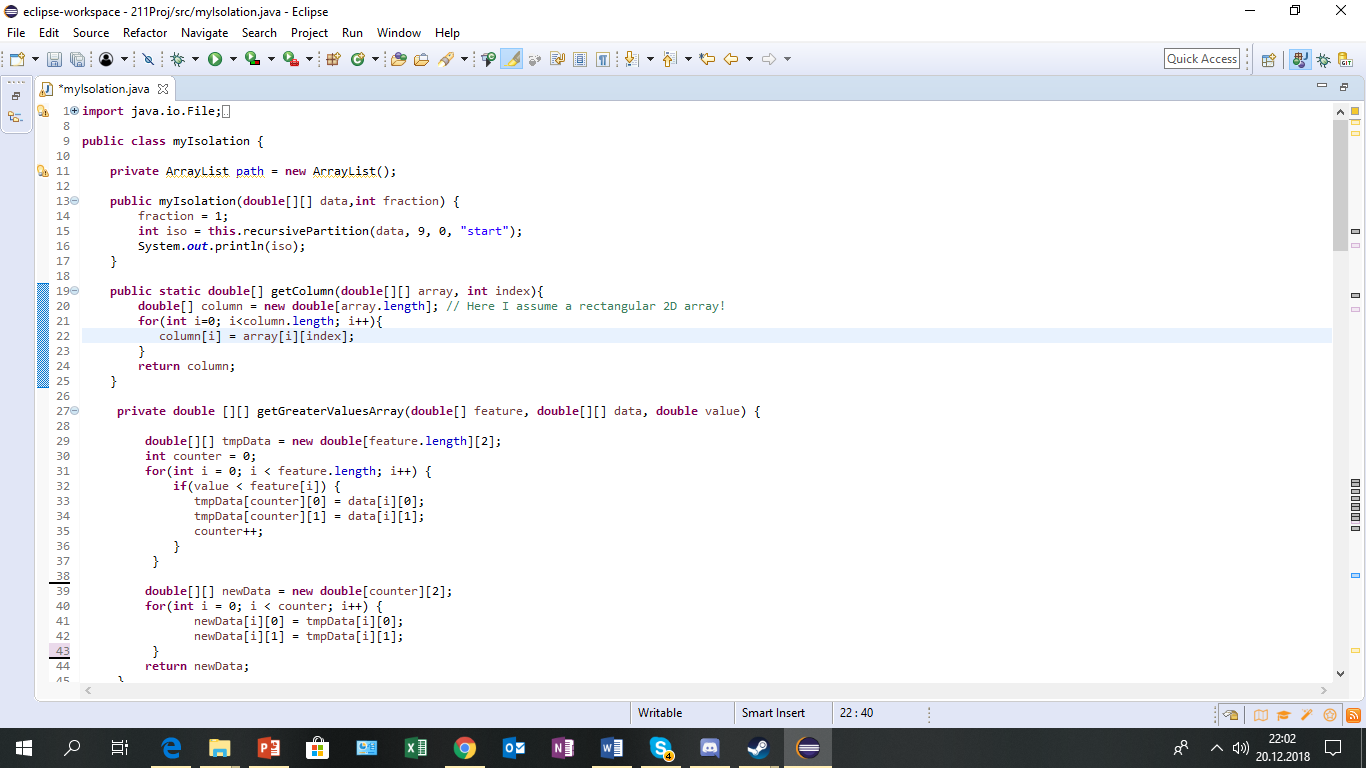
Then we wrote 2 different methods which find the max and min value in the array we got from previous method



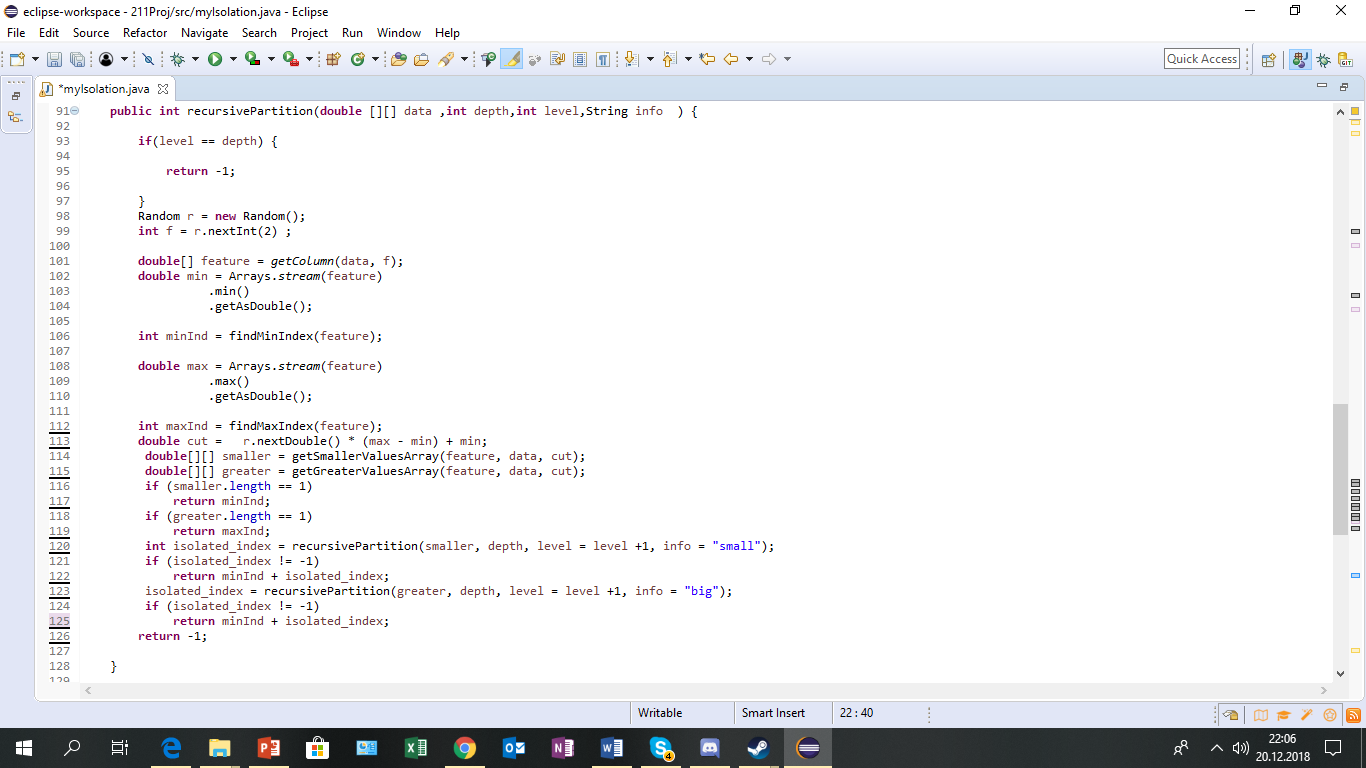
After that we created 2 more method 1 of them takes a array and a random double value and than creates a new array which contains the variables that are bigger or smaller than that random double value.



And this is the method which takes the column of the data we have we wrote and used it’s output on other methods



And in this step what we do is basicly count how many times we use getSmallerVal. And GetGreaterVal. methods and compare it to depth note that we will be taking as “9”.Other than that level will start from 0 and data will be our data taken from “messages.txt”.



Methodology:

In first part we managed to completely finish Statistical Anomally Detection without any problems and drawed the chart.

And in second part of the Project we managed to come until to the part where we count how many times we are splitting the array in the data.but then we stuck in the part where we must print which variable was taken how many times and isolation took how many steps such as :

0: [5, 3, 7, 8, 4, 6, 3, 4, 3, 2]

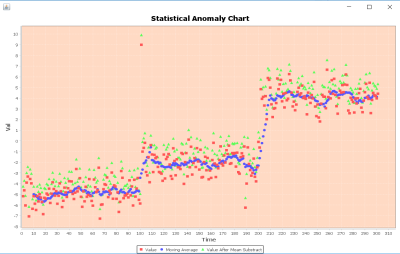
1: [2, 14]

2: [7, 7, 6, 8]

And the part after that where we are using storage table to store average path length of isolations to find the anomally

OUTPUTS:

Statistical anomally detection’s output:



And we weren’t able to create a output for isolationforest.

DISCUSSION:

We weren’t sure how to print the table.this problem was caused by multiple things starting from no one knows python incredibly well.in the last days of the Project we tried to create a different class for recursive partition method so “level” would be a private variable for a object but we werent able to complete it due to time.other then that creating the recursivepartition method was a problem on its own which took lot of time which was more then we guessed and it caused us some time problems like i said.

In order to avoid the same mistakes.In the future i would try to start coding a lot earlier and with a plan/outline so in the last days i wouldn’t have to code all of it in a hurry and if i manage to do it and since i would be working with a outline the wouldn’t be any needs for rewriting/reworking the code due to imcompability between the methods.