**Implementation Document**

**Business Understanding**

**Business Objective**

The objective of the entire project is to aid in the reduction of crime within Baltimore. This objective is to be carried out by making use of data science and machine learning techniques. One area that has been marked out for focus is “Crime and Youth”. The objective for this focus is to help reduce crime rates by the youth of Baltimore.

**Assess Situation**

Currently Baltimore has a record high rate of homicide. It has a higher rate of homicide than any other city in America. Baltimore’s rise in homicide and other crimes is being credited to a “Drug Turf-War”.

**Data Mining Goals**

1. Predict the likelihood of a crime being a felony crime based off the offenders age.
2. Predict the likelihood of a crime being a felony crime based off the offenders age and the poverty rating of the neighbourhood the offence was committed in.
3. Predict the most dangerous areas in Baltimore
4. Predict the most dangerous times in Baltimore’s highly populated areas.

**Project Plan**

Using Rapid Miner the datasets that have been gleaned and prepared from various places will be used to make the predictions previously mentioned.

The results of these predictions/ models will be assessed and analysed before being presented in a manner that conveys the finding clearly and correctly.

**Data Understanding**

**Collect Data**

Two datasets (Crime, Incidents) were collected from the BaltimorePD website.

A dataset containing information on the Baltimore neighbourhoods was gleaned from, by using a scraper made from Python using beautiful soup 4.

A dataset was gleaned from Maryland State Commission on Criminal Sentencing Policy by downloading the pdf and using PDFElement to convert the table inside the PDF into a csv file.

**Describe Data**

Explain each dataset

**Explore Data**

ASK GREG??

**Verify Data Quality**

Both the Arrests and Incidents datasets had missing values in a number of their rows. Apart from these missing values the data was relatively acceptable with only a small number of exceptions.

The dataset that was gleaned from , was constructed by the authors and as such was tailor made to suit the project without any faults.

The dataset from MSCCSP outlining charges was constructed by PDFElement and as such was imperfect and had various issues with it.

**Data Preparation**

**Select Data**

State the data to be used for each goal

**Clean Data**

Arrests:

Neighbourhoods:

ChargeDesription: This dataset had to be cleaned substantially. There was several problems and these included merging of two columns that needed to be separated. Extra characters appended to the end of the charge strings. There was also some missing values that had o be removed.

**Construct Data**

The project requires the Arrests dataset to have a ‘Felony’ Boolean column added to it. This is done by comparing the code in the arrests data to the code in the OffenceTable dataset and using the Type column in the OffenceTable a true or false value is determined and added to the Arrests dataset.

The Neighbourhoods dataset had two columns added to it. These were PopulationRank and IncomeRank. These were added through Python by calculating appropriate labels for each neighbourhood based off their density and household income in comparison to the other neighbourhoods.

**Integrate Data**

Not Applicable

**Format Data**

Not Applicable

**Analysis of Crime and Youth**

**Objective:**

The objective of this section is to analyse and asses the situation in Baltimore with regards to the youth of Baltimore and committing felony crime. The final objective of this section is to provide information gleaned from the work undertaken that will aid in preventing young people from following a life of crime.

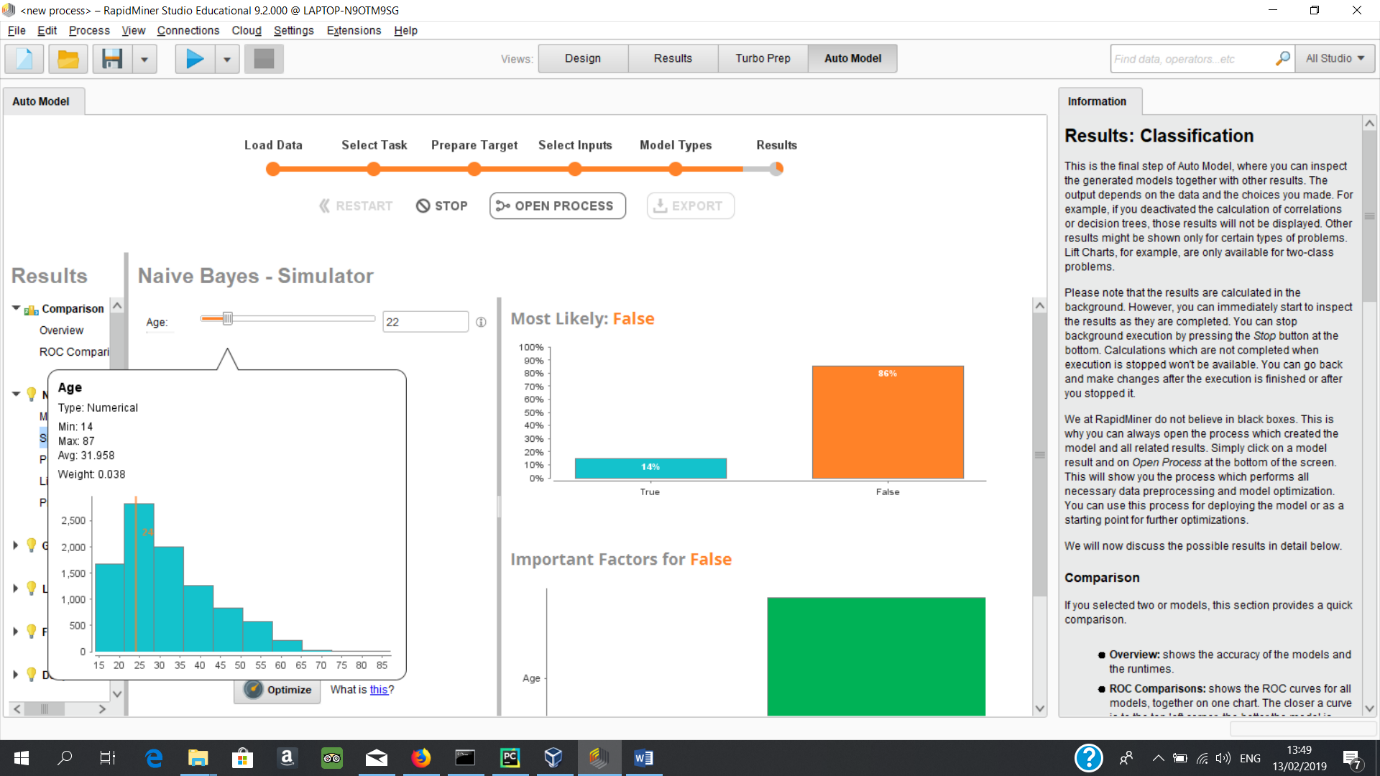
**Data:**

The data used for this section is the Arrests data set that has been completed by the addition of the felony Boolean column.

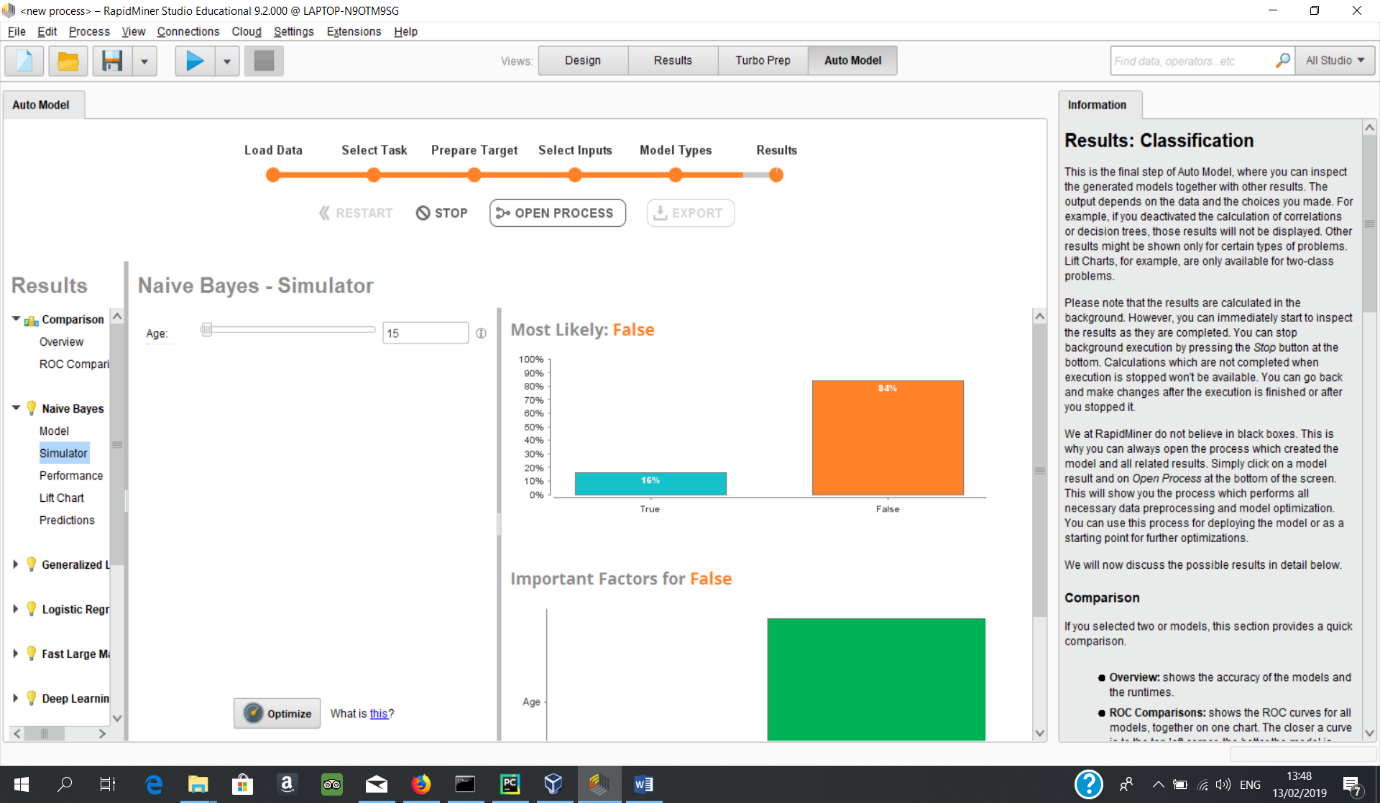
**Modelling Technique:**

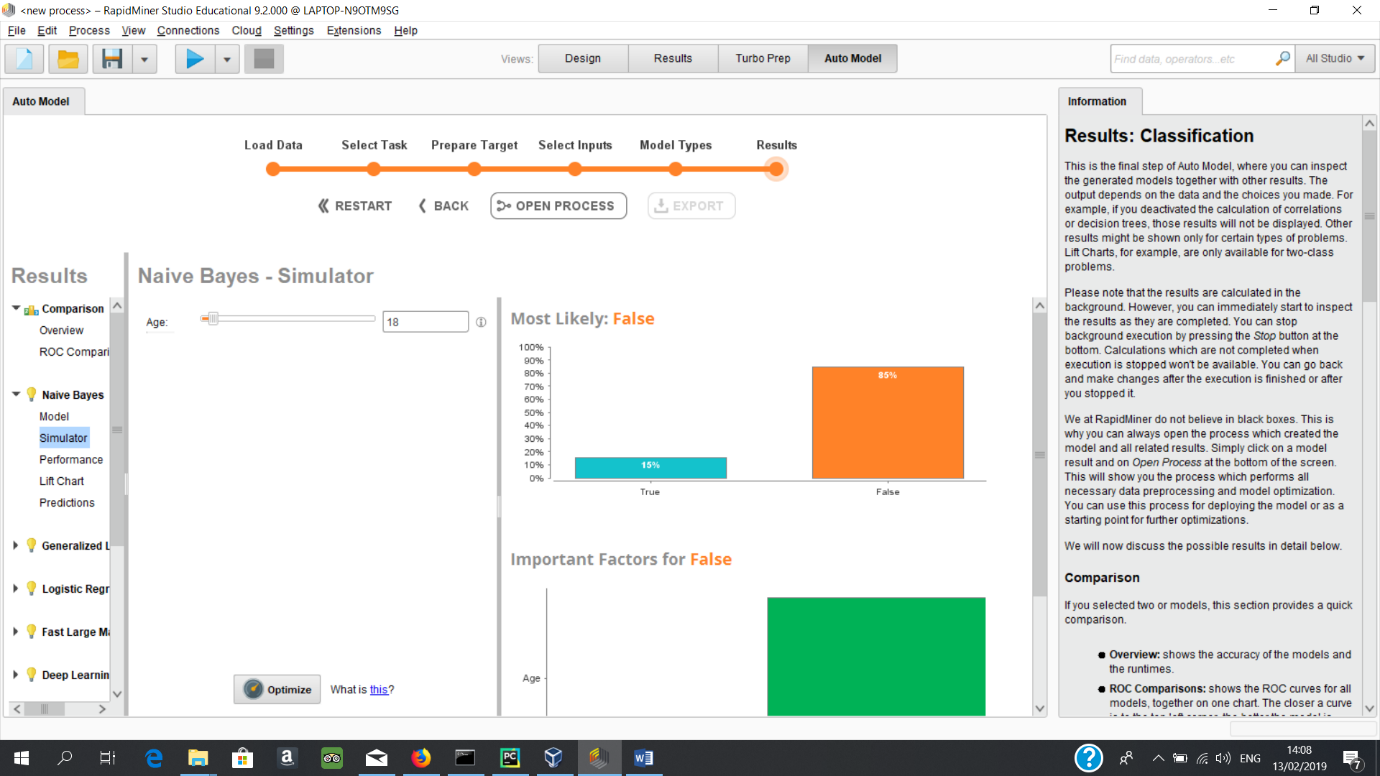
The modelling technique chosen here is Naïve Bayes. This was chosen as only one column is being used to predict the outcome of felony or misdemeanour. Naïve Bayes will be able to do this task with extreme ease.

**Results:**

****

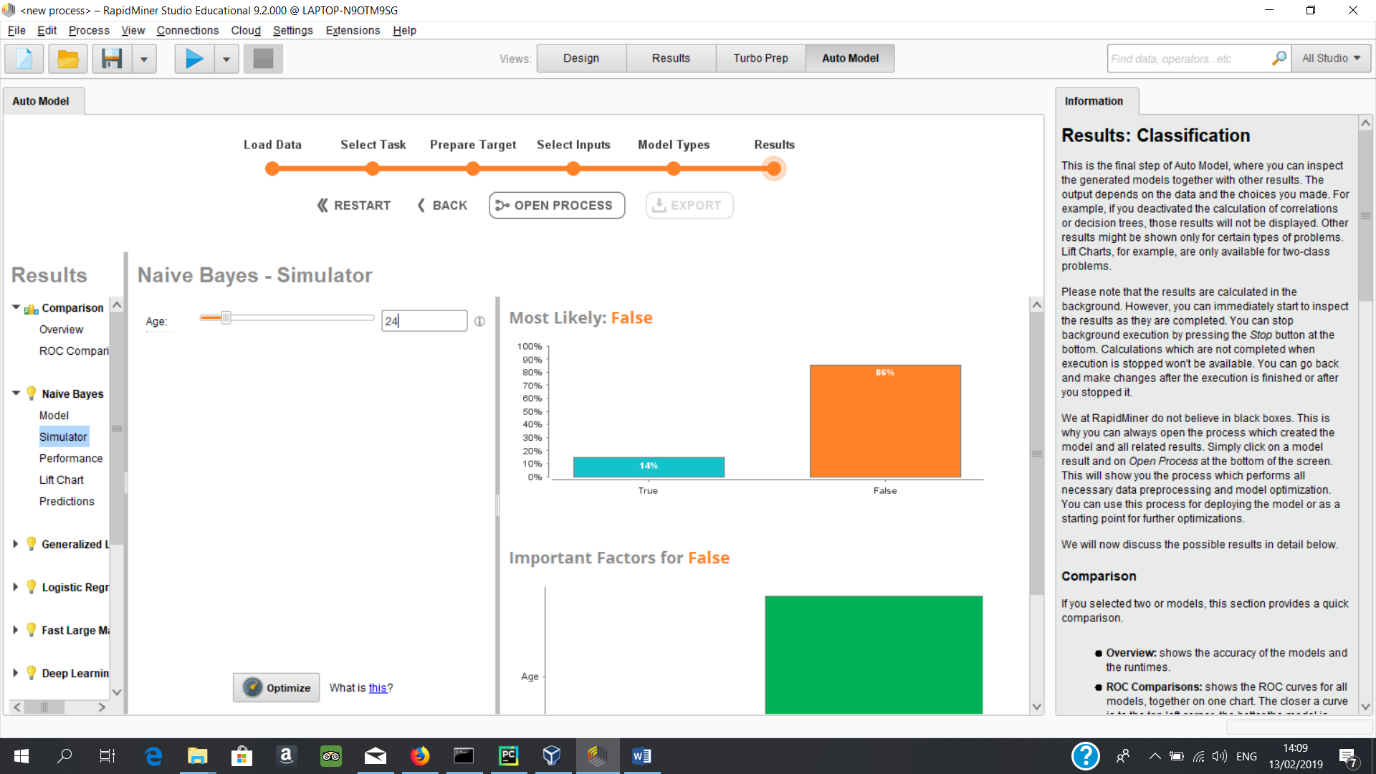
This Screenshot demonstrate a histogram that shows the frequency of crimes based in the ages of the offender.

****This screenshot shows the results of Naïve Bayes prediction of the probability of a crime being a felony crime if the age of the offender is 15. The result is 16% chance of the crime being a felony crime.

****

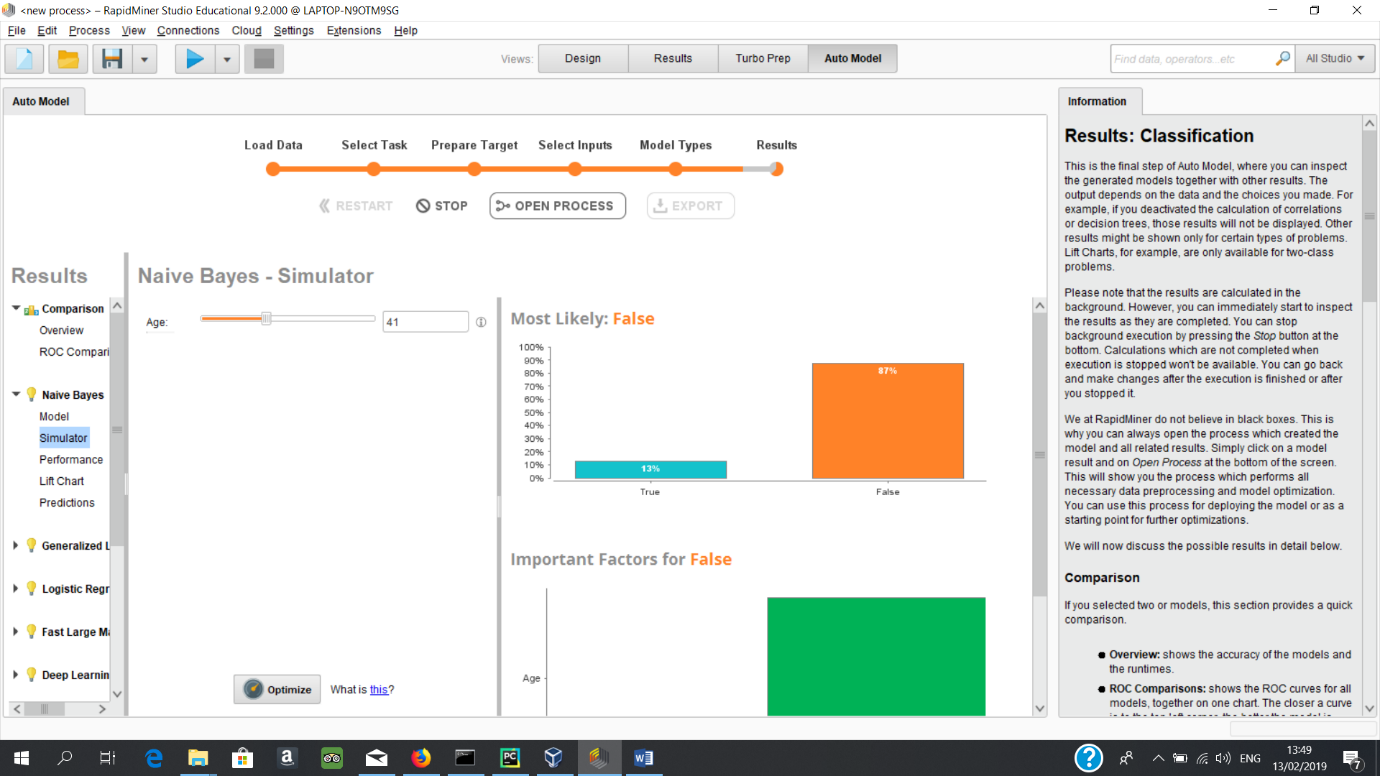
This screenshot shows the results of Naïve Bayes prediction of the probability of a crime being a felony crime if the age of the offender is 18. The result is 15% chance of the crime

being a felony crime.

****

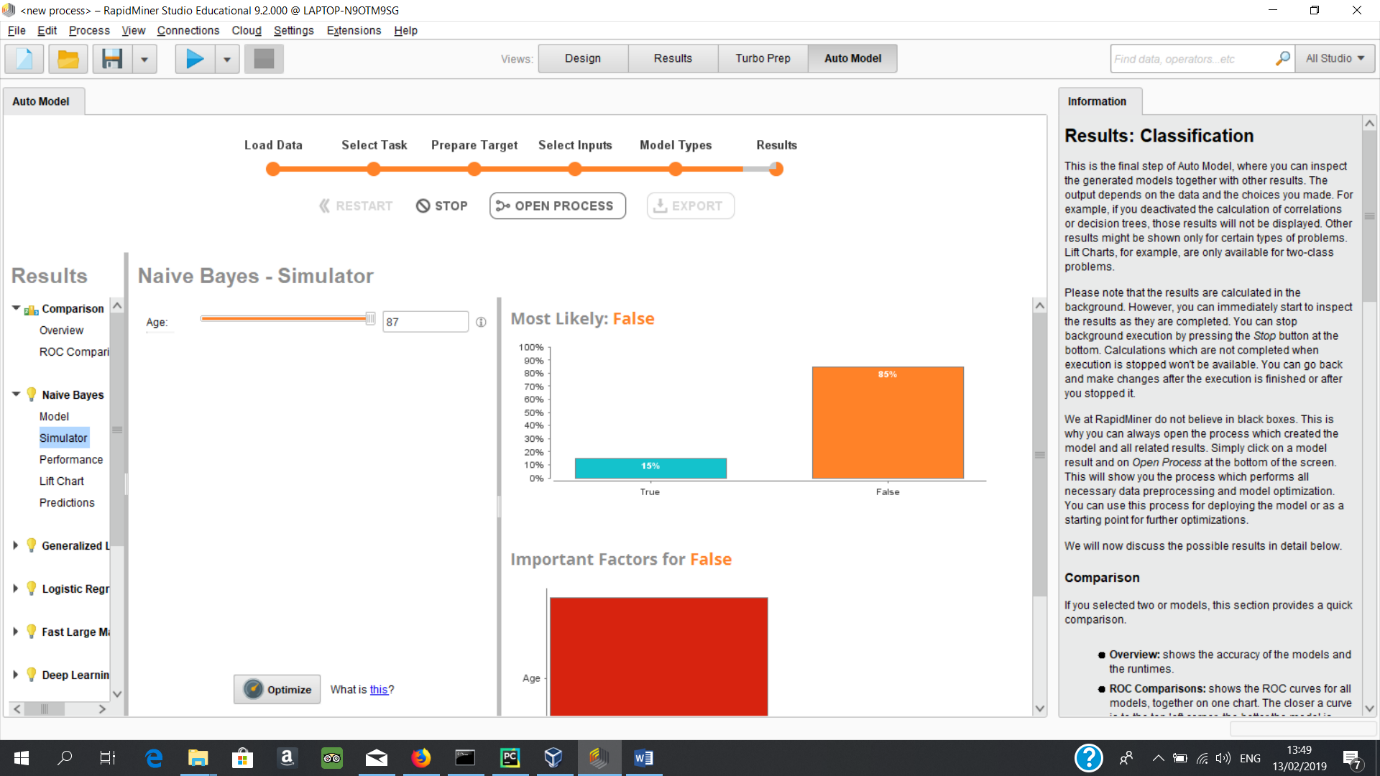
This screenshot shows the results of Naïve Bayes prediction of the probability of a crime being a felony crime if the age of the offender is 24. The result is 14% chance of the crime

being a felony crime.

****

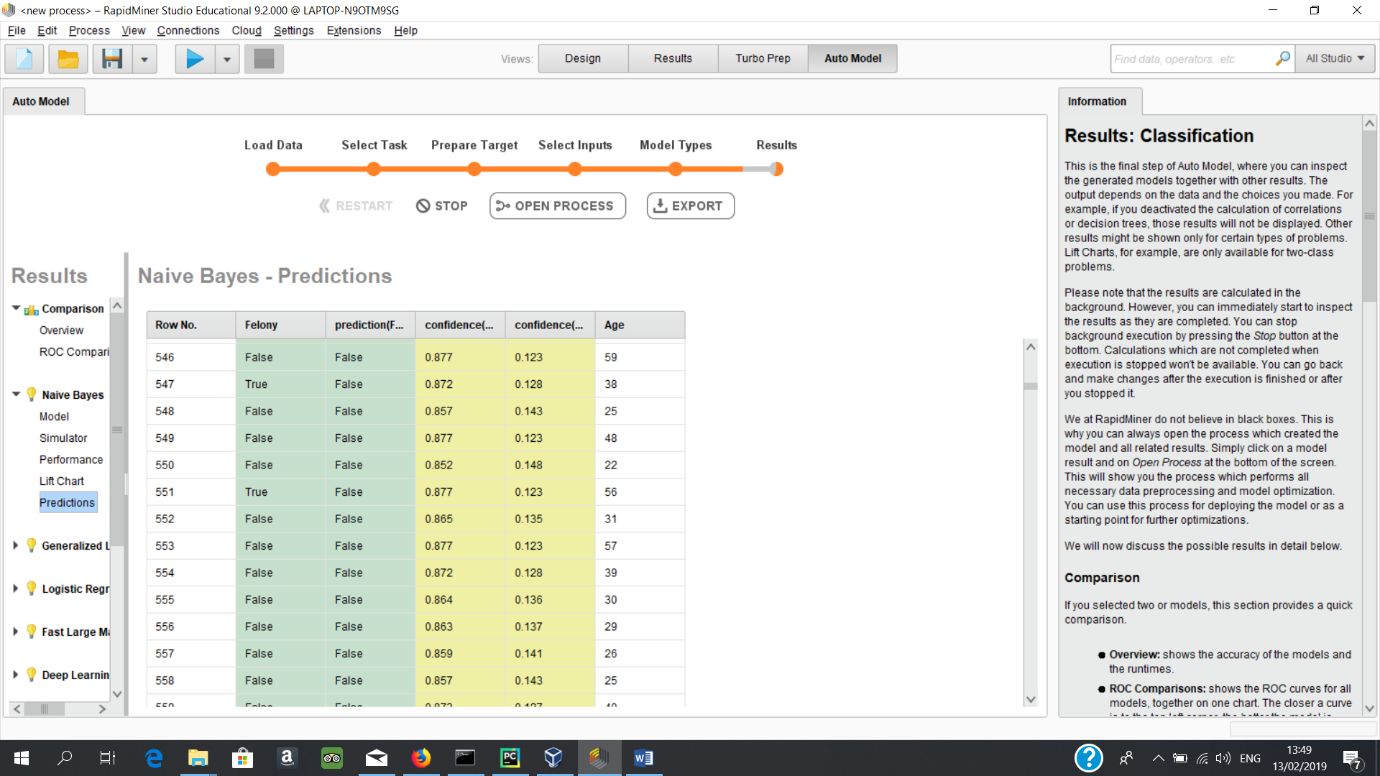
This screenshot shows the results of Naïve Bayes prediction of the probability of a crime being a felony crime if the age of the offender is 41. The result is 15% chance of the crime

being a felony crime.

****

This screenshot shows the results of Naïve Bayes prediction of the probability of a crime being a felony crime if the age of the offender is 87. The result is 16% chance of the crime

being a felony crime.

****

This is a screenshot of a snippet of predictions made.

**Assessment:**

The results for the likelihood of offences being a felony based off age begin at the age of 15. Here the result is a 16% chance of a felony crime. This statistic stays static until 18 when it becomes 15% and the next change is at 24 where the result becomes 14%. 14% is the lowest the result gets to and actually begins to rise again at 41 to 15% and rises again to 16% at 78.

The histogram of crimes does not support these findings as the histogram suggests 20 to 28 year old are committing far more crime and even 28 to 35 year olds are committing more crime than people under 20.

Combining both results suggests that the reason for the high rate of felony crimes in youth is due to a lack of arrests for misdemeanour crimes when they are committed by people under 20. This also rings through to the elderly as people over 78 have the same statistic.

**Recommendations:**

A possible solution to preventing youth offenders from following a life of crime or to prevent them from getting a long term jail sentence in their early 20’s could be to crack down on misdemeanour offences by youth’s. What this means is instead of the perceived light treatment of youths who commit misdemeanour as suggested by the datasets, the police should begin to arrest youths committing misdemeanours more vigorously. This would increase the number of arrests of youths but could instil in these youths that any crime they commit will be punished and would in turn deter these very youths from committing more serious crimes. Not only could this protect the public but could also protect the youths themselves.

**Analyse the Neighbourhoods in Baltimore**

Analyse the neighbourhoods in Baltimore and the different levels of crime in each neighbourhood.

Good

Bad

In-Between

**Analyse Times in Baltimore**

Analyse times.

Best and worst etc.

**Analyse Densely Populated Neighbourhoods**

**Referencing**

<https://nationalinterest.org/feature/why-police-cant-stop-crime-baltimore-34397>

<https://eu.usatoday.com/story/news/2018/09/25/baltimore-homicide-murder-rate-fbi-statistics-death-crime-killings/1426739002/>

https://baltimore.cbslocal.com/2018/10/17/baltimore-drug-turf-war-causing-crime-spike/