**Date:** 28/09/2018

**Present:** Alex Freitas, Adam Preston, Marcus Robinson-Bailey, Euel Domingo

**Not In Attendance:** Kester Gideon

**What has been done:**

*This is the first group project meeting.*

**Goal for next week:**

The group must read through the Privacy Policy of mobile telecommunication companies and highlight statements or sentences that will be considered as risky or where users would have a second look. Each member of the group should be able to have a copy of the privacy policy of the mobile company with the highlighted text/sentences.

**Who is doing what:**

**Adam Preston:** Privacy Policy for O2 and Virgin

**Marcus Robinson-Bailey:** Privacy Policy for Three and EE

**Euel Domingo:** Privacy Policy for Giffgaff and Lycamobile

**Kester Gideon:** Privacy Policy for Tesco and Asda mobile

**Issues/Notes/Further Discussion:**

**Implementation**

Alex advised the group to first identify the data that the group will be working on before planning the implementation.

**Identifying the data**

Everyone agreed to look at privacy policy for mobile telecommunication companies

**Recommended reading material for project**

The group was advised to have a read regarding data mining. The book suggested was: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems)

**Date:** 05/10/2018

**Not In Attendance:** Kester

**What’s Been Done since the Previous Meeting:**

**Adam:** Highlighted 'risky' statements from Virgin mobile contract. Added to google drive. Started reading chapter 1 from Data mining book book.

**Marcus:** Highlighted 'risky' statements from EE contract. Added to google drive

**EJ:** Highlighted 'risky' statements from Virgin mobile contract. Added to google drive

**Kester:** Highlighted 'risky' statements from Tesco mobile contract. Added to google drive

**What’s Being Done:**

**Adam:** Reviewed EE contract. Voting 'up' or 'down' on statements in contract. Finished chapter 1 of data mining book. Reading up on chapter 2.

**Marcus:** Reviewed virgin mobile contract. Voting 'up' or 'down' on statements in contract.

**EJ:** Highlighted 'risky' statements from Lycamobile contract. Added to google drive

**Kester:** Highlighted ‘risky’ statements from Tesco and Asda mobile contracts

**Further Discussion:**

We discussed about the preparation of our data. Moreover detailing that once the contracts had been analysed, we must extract the statements from our contracts into an excel format. This is so it can be easily processed by the neural network.

Alex also advised that during the later stages of our project we should start reading the 'Text Mining' book.

**Date:** 12/10/2018

**Not In Attendance:** None

**What’s Been Done since the Previous Meeting:**

**Adam:** Used the voting system across contracts “EE”

**Marcus:** Used the voting system across contracts “Giff Gaff, Lycamobile, Virgin Mobile”

**EJ:** Used the voting system across contracts “EE, Three and Virgin Mobile”

**Kester:** Used the voting system across contracts “”

Info: Voting system is reviewing and voting up (+2) or down (0) on statements that the person agreed or disagreed with.

**What’s Being Done:**

**Adam:** Looking for a program to create and format the collected data.Which in turn will create our dataset. Reviewing Asda, Lycamobile and O2 contracts.

**Marcus:** Looking for a program to create and format the collected data. Which in turn will create our dataset. Reviewing O2 contract.

**EJ:** Looking for a program to create and format the collected data.Which in turn will create our dataset. Reviewing Asda contract.

**Kester:** Looking for a program to create and format the collected data.Which in turn will create our dataset. Reviewing Three, Giff Gaff and O2 contracts.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).

**Further Discussion:**

Alex advised us to find a program that could easily separate and format our sentences into variable columns. A way to process the data easily was vital for us to progress.

We looked at both Microsoft Excel and python implementations, and will come to a decision of what will use during next week.

**Date:** 17/10/2018  
**Not In Attendance:** Kester  
  
**What has been done**  
**Adam:** Reviewed and used the voting system across contracts for Asda, Lycamobile, Tesco and O2.  
**Marcus:** Reviewed and used the voting system across contracts for Asda, Lycamobile, Tesco and O2.  
**EJ:** Reviewed and used the voting system across contracts for Asda, Lycamobile, Tesco and O2.  
**Kester:** Reviewed and used the voting system across contracts for Asda, Lycamobile, Tesco and O2.   
  
**What’s Being Done:**  
**Adam:** Reviewing software for text mining.  
**Marcus:** Reviewing software for text mining.  
**EJ:** Reviewing software for text mining.  
**Kester:** Reviewing software for text mining.  
  
(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**  
  
We looked at a few software in the internet. Group has agreed to review one or two software and decide which one will be used for text mining.  
  
Group has agreed to install and review software for text mining on their own pc/laptop.

**Date:** 19/10/2018

**Not In Attendance:** None

**What’s been done**

**Adam:** Reviewed the software that should be used for text mining.

**Marcus:** Reviewed the software that should be used for text mining

**EJ:** Reviewed the software that should be used for text mining

**Kester:** Reviewed the software that should be used for text mining

**What’s being done**

**Adam:** Working on the script in order to tokenize sentences in the contract.

**Marcus:** Due to inadequate and inaccurate data from the policies, they had to be reviewed again and words from sentences are being placed into an excel document.

**EJ:** Produce a script that reads sentences and putting words into columns.

**Kester:** Reviewed the data policies along with Marcus and extracted the words form the sentences and placed them into excel. Also made the documentation for the system requirement of the analyser.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).

**Further Discussion:**

We are currently looking through the excel document that broke down all the sentences into their own cell and also removed punctuations. The excel document was turned into a csv file in order to allow us to use the script which would take all the words in the excel document and turn them into headers. The script also removes any duplicate words.

More tasks are being delegated within the group so as to ensure that we all know what we are doing. More documentation documents are also being produced in order to ensure that we are up to date and can also refer to what had been done and be able to track whether we have made a mistake anywhere and rectify them

**Date:** 25/10/2018  
**Not In Attendance:** None  
  
**What’s been done**

**Adam:** Developed a script using python that extracts sentences from each contract, classifies them and puts them into a dataset

**Marcus:** Created a python script to calculate the word frequency within CSV document. This will allow the group to see any words that could be considered as a header word in our final dataset document.

**EJ:** Created a python script that separated dataset sentences into individual words and then put those words into individual columns. (Our dataset headers)   
**Kester:** Cleaning up Dataset document, by reviewing all sentences in policy contracts

**What’s being done**  
**Adam:** Merging EJ’s script into mine to create one final pre processing script. The script examines our contracts, imports the processed sentences into a table and collates each word in the sentence against a feature.

**Marcus:** Documentation and help with user interface

**EJ:** Continuing python script but including stemming of words, removal of stop words and looking at any adverbs, verbs and adjectives within the contracts.

**Kester:** Creating the design for our user Interface and making a start on the implementation of the interface.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**

Current roles within group:

* **Dataset coding:** EJ & Adam
* **User interface coding:** Kester
* **Documentation:** Marcus
* **User Interface designs:** Marcus & Kester

**Date:** 2/11/2018  
**Not In Attendance:** None  
  
**What’s been done**

**Adam:** Created pre processing script

**Marcus:** Documentation and designs for user interface

**EJ:** Created script that focuses on stemming of words, removal of stop words, adverbs, verbs and adjectives.  
**Kester:** Created User Interface

**What’s being done**  
**Adam:** Continue focusing on current role (Python script & Dataset document)

**Marcus:** Continue focusing on current role (Documentation & User Interface)

**EJ:** Continue focusing on current role (Python script & Dataset document)

**Kester:** Continue focusing on current role (User Interface), Creating gantt chart that documents progress of our project

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**

Focus on Dataset document and make that the document is 100% accurate and correct

**Date:** 15/11/2018  
**Not In Attendance:** None  
  
**What’s been done**

**Adam:** Continued on pre-processing script and finishing the dataset

**Marcus:** Continued on user interface designs and documentation

**EJ:** Continued on pre-processing script and finishing the dataset  
**Kester:** Continued with completing the implementation for user interface

**What’s being done**  
**Adam:** Consult and discuss implementation on dataset

**Marcus:** Focusing on pre-processing script, looking at removing punctuation, stop words and anything other features that stop the dataset from being finalised

**EJ:** Consult and discuss implementation on dataset

**Kester:** Focusing on completing all implementation for user interface

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**

Focus on Dataset document and make that the document is 100% accurate and correct

Focus on User Interface implementation

**Date:** 26/11/2018  
**Not In Attendance:** None  
  
**What’s been done**

**Adam:** Discussed improvements for dataset

**Marcus:** Completed implementation to pre-processing script

**EJ:** Discussed improvements for dataset  
**Kester:** Completed implementation to pre-processing script

**What’s being done**  
**Adam:** Discuss and help with implementation of pre-processing script

**Marcus:** Discuss and help with implementation of pre-processing script

**EJ:** Discuss and help with implementation of pre-processing script

**Kester:** Complete final implementation on pre-processing script by removing any anomalies out of the dataset i.e. ca, bb etc

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**

Focus on Dataset document and make that the document is 100% accurate and correct

See what Alex says in our next meeting

**Date:** 06/12/2018  
**Not In Attendance:** Marcus, Kester  
  
**What’s been done**   
  
Adam: Updated pre-processing script by removing anomalies and finalising dataset   
Marcus: Updated pre-processing script by removing anomalies and finalising dataset  
EJ: Updated pre-processing script by removing anomalies and finalising dataset  
Kester: Updated pre-processing script by removing anomalies and finalising dataset  
  
**What’s being done**  
  
**All:** Submit a project documentation\* to Alex before end of term. Further updates required on the pre-processing script for the final dataset.   
• Irrelevant words to be removed on the dataset. (i.e. pronouns, nouns, prepositions, conjunctions, ‘andor’, etc ) – implement a part of speech tag  
• Classification missing on dataset – to be added  
• Skipping lines – to be removed  
  
(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**  
  
\*The project documentation must contain the description of the applications/software used, decisions made during the process and other significant information.  
  
No meeting next week (13 December 2018). Next meeting to be confirmed by Alex.

**Date**: 17/01/2019

**Not In Attendance:** None

**What’s been done**

\* Dataset completed and sent to Alex to check.

\* All project files have been transferred to git

\* All group members started machine learning course just before Christmas break. (https://www.coursera.org/learn/machine-learning/)

**What’s being done**

**All:**

\* Submit a project documentation\* to Alex before end of term.

\* Add word count/frequency functionality to remove word that appear less than 3.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book). Further Discussion: \*The project documentation must contain the description of the applications/software used, decisions made during the process and other significant information.

**Date**: 25/01/2019

**Not In Attendance:** Kester

**What’s been done**

\* Frequency functionality within Pre-Processing script has been completed

**What’s being done**

**All:**

\* Learning how to use the WEKA Machine Learning software

\* Starting Technical Report

\* Amendments to the Pre-Processing script i.e. for each feature (word), the number of instances where the word has a frequency greater than 0.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).

Further Discussion: \*The project documentation must contain the description of the applications/software used , decisions made during the process and other significant information. (Relevant to dataset)

**Date:** 01/02/2019  
**Not In Attendance:** None  
  
**What’s been done**   
  
Adam: On the technical report, will focus on writing the quality assurance and the development done in the project.   
Marcus: Will be focusing on the general background of the project and the challenges we faced as a group during the project.  
EJ: Will be addressing the aims of the project, what we wanted to achieve and the requirements of the project. (data mining and automated machine learning)  
Kester: Will write about the abstract and the introduction of the technical report.   
  
**What’s being done**  
  
**All:** Allocated tasks to each other for the technical report. Ensured that each member of the group knows what they must do and will make a start on the technical report. Ensure that the script is changed to show the tally of occurrences of risky words in a sentence. Ranked the words from the highest to the lowest. Review the requirements of the technical report and ensure the group is working in line with them.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book).  
  
**Further Discussion:**  
  
\*The project documentation must contain the description of the applications/software used, decisions made during the process and other significant information.

**Date**: 15/02/2019

**Not In Attendance**: Kester

**What’s been done**

Adam: Further changes made to script and dataset. Applied Oversampling to the dataset and used WEKA to check for data precision and recall.

Marcus: Will be focusing on the general background of the project and the challenges we faced as a group during the project. – in progress

EJ: Addressing the aims of the project, what we wanted to achieve and the requirements of the project. (data mining and automated machine learning) – in progress

Kester: Will write about the abstract and the introduction of the technical report. – in progress

**What’s being done**

Adam: On the technical report, will focus on writing the quality assurance and the development done in the project. Looking for solutions regarding better data precision and recall.

Marcus: Will be focusing on the general background of the project and the challenges we faced as a group during the project.

EJ: Will be addressing the aims of the project, what we wanted to achieve and the requirements of the project. Send list of stop words and other words that are being removed by the algorithm for analysis.

Kester: Will write about the abstract and the introduction of the technical report.

(All still reading Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems) book and Text Mining book). Further Discussion: \*The project documentation must contain the description of the applications/software used, decisions made during the process and other significant information.