

2017

Financial Complaints: Sentiment Analysis

SOFTWARE PROJECT
ARLON JUNIOR MOLEKA

NATIONAL COLLEGE OF IRELAND | Mayor Street, IFSC, Dublin 1, D01 Y300

Software Project

Final Technical Report

FINANCIAL COMPLAINTS: SENTIMENT ANALYSIS

First Name: Arlon Junior

Last Name: Moleka

Student Number: x13562617

Email address: x13562617@student.ncirl.ie

BSc (Hons) in Business Information Systems

10 May 2017

Declaration Cover Sheet for Project Submission

SECTION 1 *Student to complete*

Name: Arlon Junior Moleka
Student ID: x13562617
Supervisor: Lisa Murphy

SECTION 2 Confirmation of Authorship

The acceptance of your work is subject to your signature on the following declaration:

I confirm that I have read the College statement on plagiarism (summarised overleaf and printed in full in the Student Handbook) and that the work I have submitted for assessment is entirely my own work.

Signature: _____ Arlon Junior Moleka _____ Date: 10/05/2017

NB. If it is suspected that your assignment contains the work of others falsely represented as your own, it will be referred to the College's Disciplinary Committee. Should the Committee be satisfied that plagiarism has occurred this is likely to lead to your failing the module and possibly to your being suspended or expelled from college.

Executive Summary

In this project, we are going to be analysing a large file of dataset (consumers' complaints) to make sure that consumers are treated fairly by financial companies such as banks, lenders and so on. This data is made available by the Consumer Financial Protection Bureau, this dataset provides consumers' experience in their own words explaining what happened in different sectors of finance like in mortgage services, prepaid card services, student loan and so on.

To do so, we analyse consumers' complaints on specific services (Prepaid Cards and Student Loan) so by adding their consents, consumers help to improve the financial marketplace for better experience in the future and companies have a clear understanding in what must be done to improve their services or products to meet consumers' need.

Measuring sentiment will help us to understand the overall feeling of consumers on a specific subject, in this scenario student loan and prepaid card services.

This helps in creating a complete image of the consumers' feedback on the service.

Sentiment is a point of view of an individual towards a specific subject, that point of view can be either positive, negative or neutral.

Two algorithms (Score sentiment algorithm and Naïve Bayes algorithm) are used in order to produce sentiment from the consumers' complaints dataset (Prepaid Cards and Student Loan) and to then classify the data of each service into "positive", "Neutral" and "negative" sentiment. For this analysis 9,403 consumers' complaints have been collected on student loan service and 1,271 consumers' complaints on prepaid cards.

Both of datasets have been collected on the 28th April 2017, they both successfully classified using the score sentiment approach and the machine learning approach.

Table of Contents

1	Introduction.....	6
1.1	Background.....	6
1.2	Aims.....	6
1.3	Technologies.....	7
2	System.....	7
2.1	Requirements.....	7
2.1.1	Functional requirements.....	7
2.1.2	User requirements.....	8
2.1.3	Datasets.....	8
2.2	Design and Architecture.....	9
2.2.1	Score Sentiment Approach.....	9
2.2.2	Machine Learning Approach.....	10
2.3	Implementation.....	11
2.3.1	Score Sentiment Approach.....	12
2.3.2	Machine Learning Approach.....	13
2.4	Testing.....	16
2.4.1	Description.....	16
2.4.2	Methods.....	16
2.5	Results.....	17
2.5.1	Prepaid cards.....	17
2.5.2	Student loan.....	22
3	Conclusions.....	26
4	Further development or research.....	27
5	References.....	27
6	Appendix.....	28
6.1	Project Materials.....	28
6.2	Initial Project Proposal.....	28
6.3	Initial Project Requirements Specification.....	32
6.4	Monthly Journals.....	40
6.5	Sentiment Coding.....	44
6.5.1	On Prepaid Cards.....	44
6.5.2	On Student Loan.....	54
6.5.3	Comparing the 2 datasets.....	64
6.6	Complaints' Score and complaints' Polarity & emotion (Tables).....	65
6.6.1	Database 1: Complaints on Prepaid Cards.....	65

6.6.2 Database 2: Complaints on Student Loan..... 70

1 Introduction

1.1 Background

Sentiment analysis is the process of determining the opinion or feeling of a piece of text. It seems to be easy for human being to detect if a sentence is either positive or negative.

For instance, in this sentence “I love my mother”; it is easy for anyone to tell that this sentence is expressing a positive sentiment. When performing sentiment analysis, we seek for opinions in that context where we pick sentiment out of the opinion.

According to the above example, it is my feeling towards my mother, in this opinion “I love my mother”, we note the verb love that express a sentiment of positivity; note that usually, verbs describe sentiments.

Companies across the world have implemented machines learning how to automatically detect sentiments on their customers. It is super useful for gaining insight into customers’ opinions.



This approach is very important because, once you understand how a customer feel about your product/service after analysing his comment/feedback, you can find out what are their expectations, what change will be relevant to meet their needs.

1.2 Aims

This project would attempt to extract data from the unstructured elements by means of text processing techniques and use the extracted data to identify the nature of complaints which are likely to be upheld and/or rejected.

For this project, the text processing technique adopted will be sentiment analysis. Surely, we would like to know what a sentiment analysis is, in brief what we know about sentiment analysis is detecting and understanding how the audience is reacting to a product/service, either positively or negatively. In this case, how consumers are reacting on a service offered by a financial company. The customer feedback on the service is extremely important for the organisation.

The purpose of this project is to create a system that will automatically detect the customer feedback which can be positive, neutral or negative. The performance of this system will show how well the service is appreciated by the clients, otherwise the organisation can easily identify complains about their services from the consumers and respond to their queries quickly with efficiency.

Here, the financial organisation focus on the extracted data to identify the nature of complaints which are likely to be upheld and/or rejected. The priority will be the complaints

highly qualified negative as the organisation will look to satisfy them first, this category is upheld. The organisation operates in descendant manner which means from negative through neutral to positive. Among these 3 categories, positive is where the organisation will be rejecting consumers feedback to resolve where they went wrong in the negative group. The neutral category may require a deep analysis to see where the organisation can bring change to completely satisfy consumers in this category.

1.3 Technologies



This is an open source programming language generally used between statisticians and data miners who look for analysing data or developing statistical software. R is a language and environment for statistical computing. Also, it provides graphics supported by the R foundation for statistical computing.

R is a great environment for text mining, packages such sentiment R and text mining have been well developed by Timothy P. Jurka when using the machine learning approach. In the machine learning approach, sentiment analysis is more fitting.

2 System

2.1 Requirements

2.1.1 Functional requirements

The Functional requirements describes what the system will achieve in ranked order. Here is the list of steps involved in the construction of this system:

- Collection of data from the Consumer Financial Protection Bureau website on prepaid card and student loan consumers' complaints
- The procedure of cleansing data from the previous stage for a cleaner version of it that will be readable with no error.
- Data obtained in the previous stage will then be stored in a database.
- Classification: here the data will be classified in 3 categories which are Positive, Neutral and Negative.
- Analysis: we will finally perform our analysis on the data classified in the previous stage. Here with the result obtained, we can clearly see what are the complaints more likely to be upheld or rejected.

More details on each stage can be found on the appendix 3: Initial Project Requirements Specification

2.1.2 User requirements

This system is intended to perform a sentiment analysis on consumers' feedback on financial services stored at CFPB, the system will then classify the results into mainly 3 categories. (Positive, Neutral or Negative). Depending on the score attributed to each customer feedback, the ones with the highest negative values are the most likely to be upheld and the ones with the highest positive values are the most likely to be rejected.

The output of this system will be nothing else that a report on how the financial services (prepaid card and student loan) are seen by customers, what are they thinking about them? What are consumers' feelings towards them?

Clients of this system (financial institutions), they want an efficient system that will perform the sentiment analysis accurately on a target categories of the feedback of their customers. For instance, in this scenario we have focused on prepaid card and student loan. After the consumers' voices have been analysed by mean of text processing, this helps the financial institution (Bank or Lenders) to measure their consumers' satisfaction on their financial services they provide and gives them a large understanding in which sector they need to improve their services.

2.1.3 Datasets

The Consumer Complaint Database is a group of complaints on a variety of consumer financial products and services. This data is sent to the companies as feedback on where they went wrong but that can help them to improve as well. This database usually updates on a daily basis.

Each complaint of this database contains certain information on source of the complaint in the US, the date of that complaint's submission, the company the complaint was addressed for attention and many more.

The CFPB make this data available for more exploration in the sector, before this data is published; all personal information is removed. This data is available in 2 different format (CSV or JSON), also it can be downloaded according to the product variable.

Two different files have been downloaded in the csv format for the purpose of this project, one containing consumers' complaints on the prepaid card and the other complaints on student loan.

- The Consumer_Complaint_Narratives.csv file contains 1,271 observations of 18 variables, file size is 1.38MB.
- The Student_Loan_Complaints_with_Consumer_Complaint_Narratives.csv file contains 9,403 observations of 18 variables, file size is 12.5MB.

The 18 variables are listed as follow:

Date received	The date the CFPB received the complain
Product	The type of product/service the consumer identified in the complaint
Sub-product	The type of sub-product the consumer identified in the complaint
Issue	The issue the consumer identified in the complaint
Sub-issue	The sub-issue the consumer identified in the complaint
Consumer complaint narrative	Actual feedback/complaint
Company public response	The company's optional, public response to a consumer's complaint
Company	The complaint is about this company
State	The consumer's reported mailing state for the complaint
Zip code	Mailing ZIP code provided by the consumer
Submitted via	How the complaint was submitted to CFPB
Tags	Data that supports easier searching by or on behalf of consumers
Consumer consent provided	Finds if the consumer agreed to publish their complaint narrative
Date sent to company	The date the CFPB sent the complaint to the company
Company response to consumer	This is how the company responded
Timely response	Whether the company gave a timely response
Consumer disputed	Whether the consumer disputed the company's response
Complaint ID	The unique identification number for a complaint

2.2 Design and Architecture

This project will involve 2 approaches on performing the sentiment analysis, the purpose of this is to compare results of both methods to each other to measure the degree of accuracy.

2.2.1 Score Sentiment Approach

The `score.sentiment()` function is simply an algorithm used in R to perform sentiment analysis based on Breen's approach named after the seminal elucidating slides on twitter sentiment analysis conducted by Jeffrey Breen. For this project, this approach will calculate the score of each complaint and classify it as either positive, neutral or negative according to the score obtained.

The calculation of the overall score of a complaint is defined by the following formula:

$$\text{Score} = \text{Number of positive words} - \text{Number of negative words}$$

The `score.sentiment()` function allocates a score of 1 to each positive word in the consumer complaint narrative of an individual and add them up to have the 'number of positive words' and allocates a score of -1 to each negative word in the consumer complaint narrative of an individual and add them up to have the 'number of negative words'.

The overall score of the above formula determines the category of the complaints, this consists of comparing the final score to the value 0, if:

- Score is greater than 0 (score is > 0), the complaint will have an overall 'positive opinion'
- Score is less than 0 (Score is < 0), the complaint will have an overall 'negative opinion'
- Score is equal to 0 (Score is = 0), the complaint will be a 'neutral opinion'

The sentiment lexicon is the last factor required to count the number of positive or negative opinion words. In this project, we use the opinion lexicon by Hu and Liu, It is basically a list of English positive and negative words determining an opinion. It contains in total about 6,800 words (2,006 positive words and 4,783 negative words), this list of opinion words can be modified by adding or deducting opinion words from it.

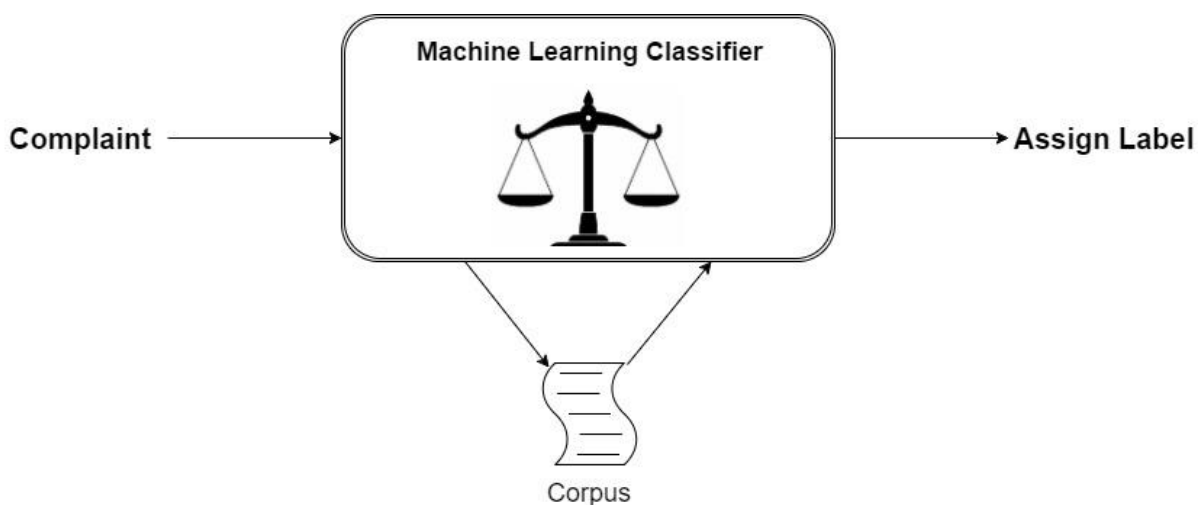
2.2.2 Machine Learning Approach

Machine learning approach is a type of artificial intelligence providing computers with the abilities to learn instead of being programmed. This computer program is more complex as it changes when new data comes in. In other words, this computer program is set up as a human being's brain works.

The defining characteristic of the machine learning algorithm is that it requires some data which to start with and this is where the corpus get involved. briefly, what is known by the corpus is that it is a body of text already classified which will serve as input in the machine learning algorithm.

The machine learning approach has this obvious stage known as the training step in which the machine learning is trained on the corpus. The result obtained through this training stage is then the machine learning classifier.

The machine learning classifier is used to classify the new problem instance that the computer program has never seen. It is important to know that this classifier is the result of training the machine learning algorithm.



There are some types of machine learning such as Naïve Bayes or Support Vector Machines. In this paper, we will focus on the Naïve Bayes algorithm.

Naïve Bayes Algorithm

Naïve base rapidly generates mining models that can actually be used for classification or prediction of complaints. The Naïve Bayes makes the model a good choice for the exploration of the data.

It calculates probabilities for each possible state of input attribute, giving each state of predictable attribute. This means that it calculates the occurrence distribution of input in the classes of the expected attribute. So, these probabilities will then be used in later stage to

envisage the output of the predicted attribute that is founded on the known input attributes. Also, all the input attributes are treated uniquely, this means that the algorithm takes one input after another.

Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In R, the function `classify_emotion()` will be used to assign label according to emotion each complaint expressing. The sentiment package used in this project uses the emotion in-built dataset that categorises words into 7 categories which are known as fear, surprise, disgust, joy, sadness, anger and unknown. This dataset contains approximately 1,500 words.

And the `classify_polarity()` function is then used for classifying words into polarity categories know as positive, neutral or negative. This is also a portion of the R sentiment package.

2.3 Implementation

The prepaid card and student loan complaints have been implemented separately following the exact same procedure. Only the prepaid card complaints implementation is explained in stages here, the R code for both analysis are attached in the appendix.

Getting the data

Before starting the proper analysis, we must first grab the data from the Consumer Financial Protection Bureau. The CFPB gives access to its API for more option on the data or alternatively it gives access to download the entire dataset in 2 format which are CSV and JSON. The dataset can be filtered depending on the business product or service such as student loan, prepaid card, debt collection, mortgage and so on. Also, it gives option to download the dataset including blanks in the consumer narrative column or filtering the data with only consumer narratives.

I have filtered and downloaded individually the prepaid card and student loan complaints for this analysis as it can be viewed on the above image.

2.3.1 Score Sentiment Approach

2.3.1.1 *Sentiment function in R*

This function consists of getting the sum of positive words in each complaint and subtracting from that the sum of negative words in the same complaint and obtain overall score of that specific complaint.

```

5
6 score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
7 - {
8   require(plyr)
9   require(stringr)
10
11 # we got a vector of sentences.
12 #plyr will handle a list or a vector as an "l" for us
13 # we want a simple array ("a") of scores back, so we use
14 # "l" + "a" + "ply" = "laply":
15
16 scores = laply(sentences, function(sentence, pos.words, neg.words) {
17   # clean up sentences with R's regex-driven global substitute, gsub():
18   sentence = gsub('[[[:punct:]]]', '', sentence) #Remove punctuation
19
20   sentence = gsub('[[[:cntrl:]]]', '', sentence) #Remove Control characters
21
22   sentence = gsub('\\d+', '', sentence) #Remove digits
23
24 # and convert to lower case:
25
26 sentence = tolower(sentence)
27
28 # split into words. str_split is in the stringr package
29
30 word.list = str_split(sentence, '\\s+')
31
32 # sometimes a list() is one level of hierarchy too much
33
34 words = unlist(word.list)
35
36 # compare our words to the dictionaries of positive & negative terms
37
38
39
40 pos.matches = match(words, pos.words)
41 neg.matches = match(words, neg.words)
42
43 # match() returns the position of the matched term or NA
44 # we just want a TRUE/FALSE:
45
46 pos.matches = !is.na(pos.matches)
47
48 neg.matches = !is.na(neg.matches)
49
50 # and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():
51
52 score = sum(pos.matches) - sum(neg.matches)
53
54 return(score)
55
56 }, pos.words, neg.words, .progress=.progress )
57 scores.df = data.frame(score=scores, text=sentences)
58 return(scores.df)
59 }

```

To do this in R, we require 2 packages which are stringr and plyr. These packages are therefore installed and loaded in R using the library function.

Firstly, the sentiment function's parameters include 3 vectors (text to score (sentence), word of positive sentiment, word of negative sentiment).

Secondly, the array is created using the laply function. The sentence is cleaned by removing punctuation, digit and other specific characters. The sentence is then converted to lower case and split up in words using the stringr package. These words are compared to the dictionary of positive and negative words.

by getting the position of the match term, we get the true or false value, the overall score of each complaint is now calculated and that put in a data frame, the object is called "score.df".

2.3.1.2 *Input corpus in R*

The sentiment lexicon made available by Hu and Liu was used as the input corpus for this project. These positive and negative words predefined can be downloaded at

<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

It contains around 6800 words in total subdivided in 2 categories (positive words and negative words). Also, it can be modified by adding or subtracting words.

```
63 #Load sentiment word lists
64 hu.liu.pos = scan("C:/Users/molsj/documents/4th Year/Semester 2/Software Project II/R Codes/positive-words.txt", what='character', comment.char=';')
65 hu.liu.neg = scan("C:/Users/molsj/documents/4th Year/Semester 2/Software Project II/R Codes/negative-words.txt", what='character', comment.char=';')
66
67 #Add words to list
68 pos.words = c(hu.liu.pos, 'upgrade')
69 neg.words = c(hu.liu.neg, 'wtf', 'fuck', 'shit', 'wait','waiting', 'epicfail', 'mechanical')
70
71
72 #Import csv file (Dataset)
73 DatasetCustomers_Narratives <- read.csv("C:/Users/molsj/documents/4th Year/Semester 2/Software Project II/R Codes/Consumer_Complaint_Narratives.csv")
74 DatasetCustomers_Narratives$Consumer.complaint.narrative<-as.factor(DatasetCustomers_Narratives$Consumer.complaint.narrative)
75
```

In the above image, the opinion lexicon is loaded into R, new words have been added to each category (1 in the positive category and 7 in the negative category).

Finally, read the prepaid card dataset into R and force the argument to a factor.

2.3.1.3 *Scoring complaints and adding column*

After completing the previous stages, all the complaints can be finally be given a numerical value that will determine its category. The complaint is said to be positive if its score is over than value 0, neutral if the its score is equal to the value 0 and negative if its score is less than value 0.

```
77 #score all complaints
78 Consumer_Complaint_Narratives.scores = score.sentiment(DatasetCustomers_Narratives$Consumer.complaint.narrative, pos.words,neg.words,
79 .progress='Consumer.complaint.narrative')
80
81
82
83 path<-"C:/Users/molsj/documents/4th Year/Semester 2/Software Project II/R Codes"
84 write.csv(Consumer_Complaint_Narratives.scores,file=paste(path,"Consumer_Complaint_Narratives_Scores.csv",sep=""),row.names=TRUE)
85
86
87 Consumer_Complaint_Narratives.scores$Feedback = 'Consumer_Complaint_Narratives'
88
```

The above image illustrate how each complaint get a numerical score using the score.sentiment() function ran earlier, those scores for each complaints will be stored as a csv file in the working folder.

The last step here is the creation of a new column called feedback beside the score and complaints narrative columns. This column differentiates the 2 datasets as one is analysing prepaid cards complaints and the other analysis the student loan complaints. The feedback column added is then used to plot an unique graph with the sentiment score when comparing the distribution of the 2 sectors (prepaid card and student loan).

2.3.2 *Machine Learning Approach*

2.3.2.1 *Data preparation*

Firstly, the text (complaint) must be grabbed from the dataset downloaded earlier for a cleansing process before starting the analysis.

```

119 #Get the text
120 Consumer_Complaint_Narratives_txt = DatasetCustomers_Narratives$Consumer.complaint.narrative
121
122 #Data cleansing to prepare text for the sentiment analysis
123 Consumer_Complaint_Narratives_txt = gsub("[[:punct:]]", "", Consumer_Complaint_Narratives_txt) #remove punctuation
124
125 Consumer_Complaint_Narratives_txt = gsub("(RT|via)((?:\b\\w*\b)+)", "", Consumer_Complaint_Narratives_txt) #remove retweet entities
126
127 Consumer_Complaint_Narratives_txt = gsub("@\\w+", "", Consumer_Complaint_Narratives_txt) #remove @
128
129 Consumer_Complaint_Narratives_txt = gsub("[[:digit:]]", "", Consumer_Complaint_Narratives_txt) #remove digits
130
131 Consumer_Complaint_Narratives_txt = gsub("^\\s+|\\s+$", "", Consumer_Complaint_Narratives_txt) #remove tabs
132
133 Consumer_Complaint_Narratives_txt = gsub("[ \\t]{2,}", "", Consumer_Complaint_Narratives_txt) #remove spaces
134
135 Consumer_Complaint_Narratives_txt = gsub("http\\w+", "", Consumer_Complaint_Narratives_txt) #remove link
136
137 Consumer_Complaint_Narratives_txt = gsub("xx", "", Consumer_Complaint_Narratives_txt) #remove xx
138
139 Consumer_Complaint_Narratives_txt = gsub("xxx", "", Consumer_Complaint_Narratives_txt) #remove xxx
140
141 Consumer_Complaint_Narratives_txt = gsub("xxxx", "", Consumer_Complaint_Narratives_txt) #remove xxxx]
142
143 Consumer_Complaint_Narratives_txt = gsub("xxxxxxx", "", Consumer_Complaint_Narratives_txt) #remove xxxxxxxx
144

```

In the above image, text is grabbed from the prepaid card database in the consumer complaint narrative column (DatasetCustomers_Naratives\$Complaint.narrative.complaint). This allows R to get individual complaints and proceed in data cleansing by removing, digit, punctuation, tabs, specific terms and so on.

2.3.2.2 Transform to lower case

The complaint seized from the previous stage (Consumer_Complaint_Narratives_txt) is then transformed into lower case.

```

145 #Defining a function which can handle "tolower error handling",
146 #in case arises any while converting all the words into lower case.
147 try.error = function(x)
148 {
149   #create missing value
150   y = NA
151
152   #trycatch error
153   try_error = trycatch(tolower(x), error = function(e) e)
154
155   #if not an error
156   if(!inherits(try_error, "error"))
157     y = tolower(x)
158
159   #result
160   return(y)
161 }
162
163 #Transforming all the words into lower case using the
164 #try.error function created above with the supply function
165 Consumer_Complaint_Narratives_txt = sapply(Consumer_Complaint_Narratives_txt, try.error)
166
167 #Remove NAs, if any exists, from Consumer_Complaint_Narratives_txt
168 Consumer_Complaint_Narratives_txt = Consumer_Complaint_Narratives_txt[!is.na(Consumer_Complaint_Narratives_txt)]
169
170 #Also remove names (column headings) from the text,
171 #as we do not want them in the sentiment analysis
172 names(Consumer_Complaint_Narratives_txt) = NULL
173

```

Before the transformation of the complaint in lower case, we define the try.error function that can handle the “tolower error handling” in case it occurs while changing the complaint to lower case, the transformation to lower case is executed using the supply function along with the try.error function just created.

It checks for any NA existing in Consumer_Complaint_Narratives_txt to remove it and also, it removes column headings as they are not required in the sentiments analysis.

2.3.2.3 Classification

To perform the classification, R requires some packages such as sentiment, Rstem, tm and NLP. These packages can be found in R packages directory to download, one of the these packages, “sentiment” package is not available and requires to be downloaded in the cran archive in this link <https://cran.r-project.org/src/contrib/Archive/sentiment/>

After the installation of all the packages, they are loaded in R using the library function.

```

174 #installing sentiment 0.2
175 install.packages("C:/Users/moljsj/Documents/4th Year/Semester 2/Software Project II/R Codes/sentiment_0.2.tar.gz", repos = NULL, type="s
176
177 library(RcolorBrewer)
178 library(wordcloud)
179 library(NLP)
180 library(tm)
181 library(Rstem)
182 library(sentiment)
183
184
185 #We are using Bayes' algorithm to:
186
187 #classify emotion
188 #This function returns an object of class data.frame with seven columns
189 #(anger, disgust, fear, joy, sadness, surprise, best_fit) and one row for each document:
190 class_emo = classify_emotion(Consumer_Complaint_Narratives_txt, algorithm="bayes", prior=1.0)
191 #get emotion best fit
192 emotion = class_emo[,7]
193 # Replacing NA's (if any, found while processing classification)
194 #by the word "unknown"
195 emotion[is.na(emotion)] = "unknown"
196
197
198 #classify polarity
199 #This process will classify the text data into four categories
200 #(pos, neg, pos/neg, best_fit)
201 class_pol = classify_polarity(Consumer_Complaint_Narratives_txt, algorithm = "bayes")
202 #get polarity best fit
203 polarity = class_pol[,4]
204

```

We are using the most popular algorithm known as the Naïve Bayes Algorithm. This algorithm classifies the complaints based on emotion. The `classify_emotion()` function classifies the complaints into 7 categories (fear, joy, disgust, sadness, surprise, anger, unknown).

Also, it classifies the complaints based on polarity. The `classify_polarity()` function groups complaints into 3 categories (positive, neutral and negative).

2.3.2.4 Data frame

This is the creation of the data frame that helps in the visualisation to fill in polarity or emotion in the relevant graphs.

```

205 #Creating data frame and rearrange data for plotting purposes
206
207 #Creating data frame
208 sent_categ = data.frame(text=Consumer_Complaint_Narratives_txt,
209                         polarity=polarity, stringsAsFactors=FALSE)
210
211 #sort data frame
212 sent_categ = within(sent_categ,
213                     emotion <- factor(emotion, levels=names(sort(table(emotion), decreasing=TRUE))))
214
215

```

These lines of code create a data frame called `sent_categ`, where it groups the prepaid cards complaints in polarity or emotion.

2.4 Testing

This system consists of performing a sentiment analysis, the computer reads the data and assign a polarity to the data. The system analysis 2 datasets (prepaid card and student loan) both with 18 variables and 1,241 observations for prepaid card and 9,403 observations for student loan. As we are analysing 2 datasets, thus the second one (student loan) can be considered as a test data too.

2.4.1 Description

The second data to be tested is the student loan dataset, downloaded from the CFPB websites in a csv file.

Firstly, we run the sentiment analysis with the score sentiment algorithm.

We write the `score.sentiment()` function and cleans the data the load the positive and negative dictionary into R. Scoring the complaints then plot graphs and store the results in a csv file.

Secondly, we run the sentiment analysis with the Naïve Bayes algorithm using the R sentiment package. Here we clean the data first to prepare the analysis itself, this procedure consists of removing links, specific characters, digits and many more. Then the data is transformed to lower case before classifying the data into emotion using the `classify_emotion()` function or into polarity using the `classify_polarity()` function. Results can then be displayed and saved in csv files.

2.4.2 Methods

The first algorithm chosen to conduct a sentiment analysis is the score sentiment algorithm, this consists of counting the number of positive and negative words in each complaint in the database. The overall score is then the subtraction of all the negative words from the positive words.

This method is not possible until we add a dictionary of words into R.

For this, we download the opinion lexicon of positive and negative words in order to count the number of positive and negative words in a complaint. This is made available online by Hu and Liu, it contains about 6,800 words grouped in 2 categories (positive and negative). Once obtained it is loaded into R, then we can run the analysis.

The second algorithm chosen for this project is the Naïve Bayes algorithm, this algorithm uses the Naïve Bayes classifier and will do a further investigation on the data.

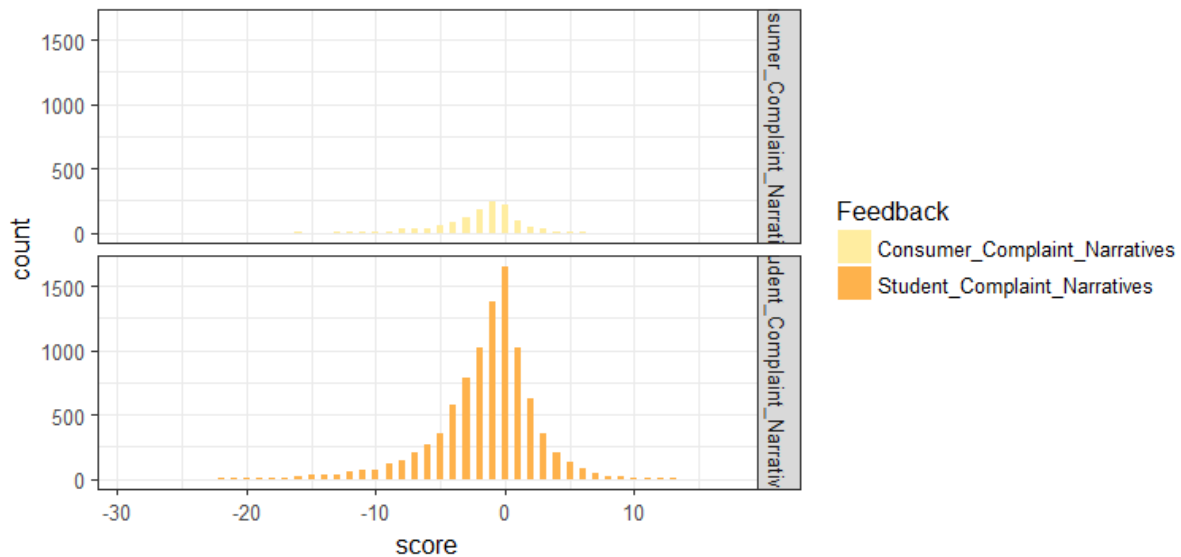
The R sentiment package and other relevant packages are loaded in R; therefore, we can use function like `classify_emotion` to classify complaints by emotion and `classify_polarity` to classify complaints by polarity.

The polarity consists of classifying data into 3 groups known as positive, neutral or negative and the emotion consists of classifying data into 7 groups known as anger, disgust, fear, joy, sadness, surprise and unknown.

2.5 Results

Comparing 2 datasets (Prepaid Card vs Student Loan)

The objective of this analysis is to automate a system that will retain the complaints with the highest degree of negativity in order for a company to respond first to situations that matter then moving on to the one that follows the previous. We understand that not every consumer of the service leaves bad feedback on the company, for those people, their feedbacks are not taken into account in this procedure as they all fall on the positive category, these complaints will then be rejected.



The above diagram compares 2 financials’ sectors, it helps the organisation at identifying which sector has the highest negative score and helps to estimate the number of complaints having a particular score. Alternatively, I have gotten the frequency of each score occurring in R in each financial sector and used excel to generate a table. This diagram is generated using the score.sentiment() in R, it assigns score to each complaint in the database. Also, this graph requires the ggplot2 package in R.

2.5.1 Prepaid cards

2.5.1.1 Summary results

2.5.1.1.1 [Sentiment function polarity summary](#)

ID	Minimum	Maximum	Polarity	Count	Percentage
1	-26	-1	Negative	846	67
2	0	0	Neutral	216	17
3	1	11	Positive	209	16

The summary result of the sentiment function on the prepaid card database indicates that the polarity distribution is as follow:

Out of 1,271 complaints provided by the database, 846 (67%) complaints are found to be negative, 216 (17%) complaints are neutral and 209 (16%) complaints are positive.

Knowing the range of the score distribution is between -26 to 11, I defined the range of each category of polarity to generate this table by using the MATCH formula in excel.

2.5.1.1.2 [Naïve Bayes polarity summary](#)

ID	Polarity	Count	Percentage
1	Negative	684	54
2	Neutral	410	32
3	Positive	177	14

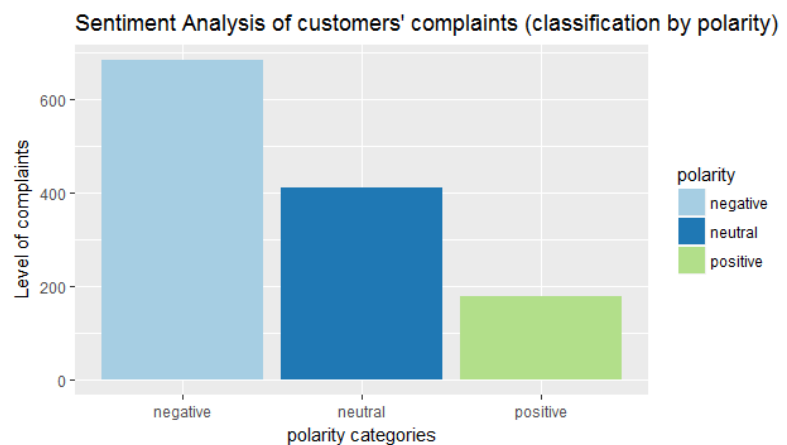
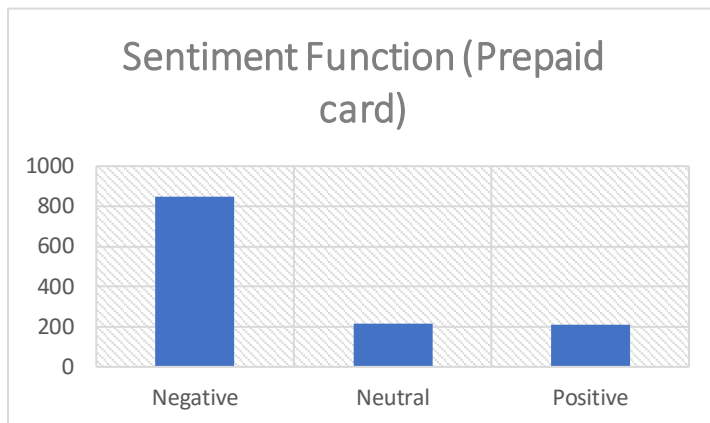
In the other hand, the summary result of the Naïve Bayes on prepaid card database indicates the polarity is as follow:

Out of 1,271 complaints provided by the database, 684 (54%) complaints are found to be negative, 410 (32%) complaints are neutral and 177 (14%) complaints are positive.

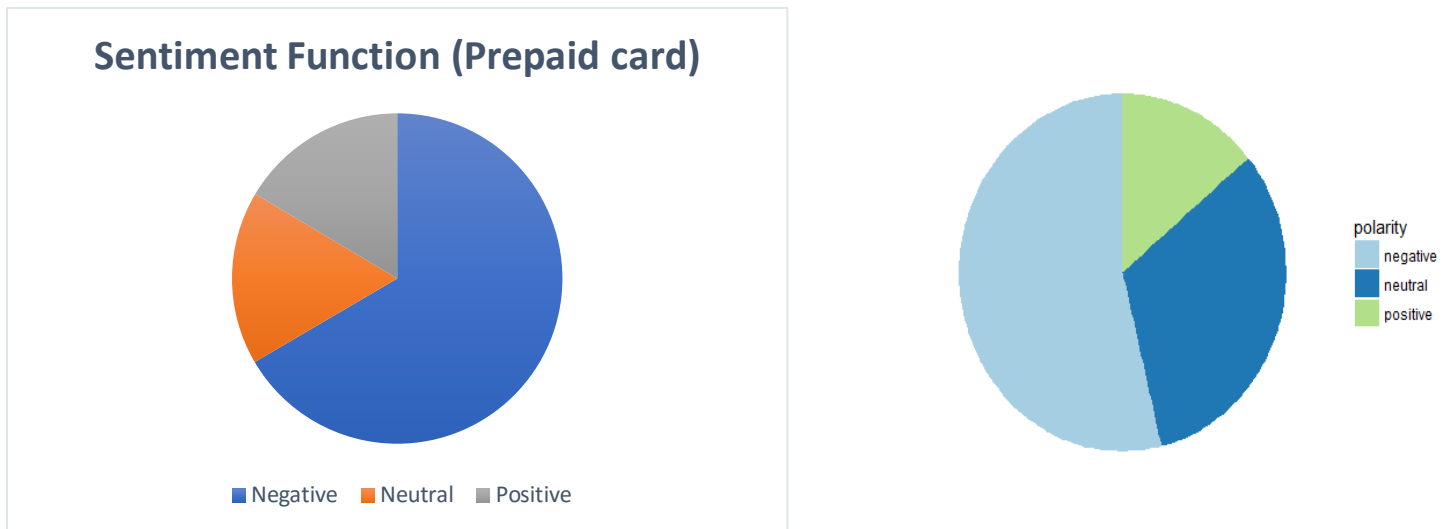
I have used the count() function in R the get the exact figures then I used excel to generate the above table.

2.5.1.2 [Comparing approaches](#)

Sentiment function histogram VS Naïve Bayes histogram



Both approaches have classified more complaints in the negative category, in the second position comes the neutral category and finally the positive category. We note that in the sentiment function it is very tight between the neutral category (17%) and the positive neutral (16%).

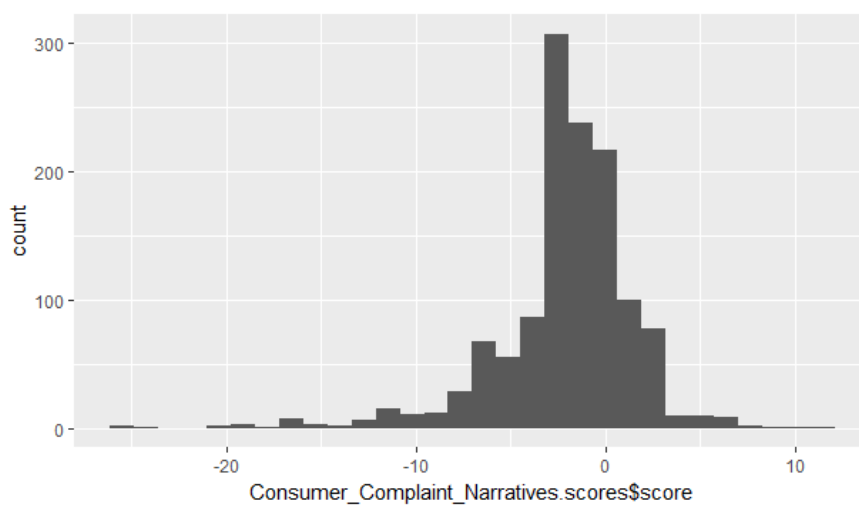
Sentiment function pie chart VS Naïve Bayes pie chart

The above pie charts diagrams confirm the result on the histograms generated previously.

2.5.1.3 Additional graphs

2.5.1.3.1 Sentiment function graphs

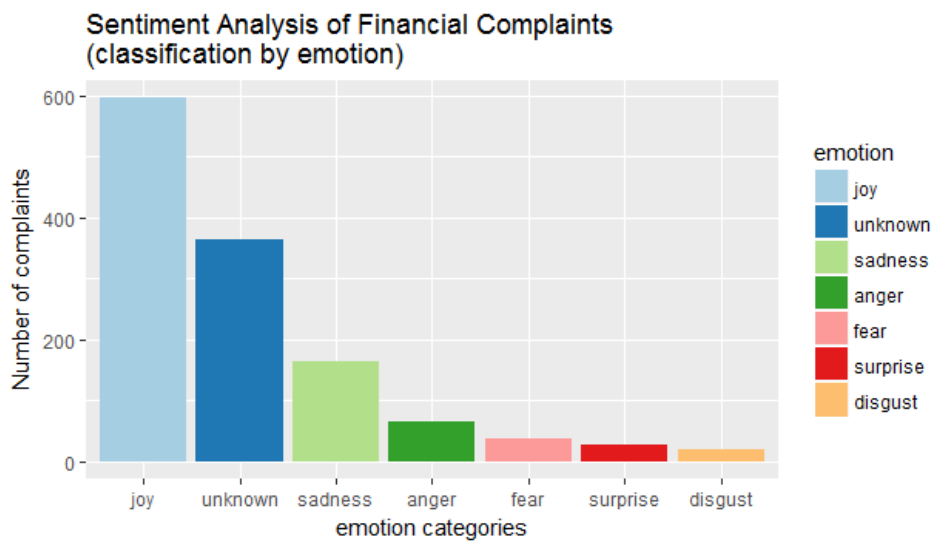
Histogram on sentiment score



This is a histogram generated using the `qplot()` in the sentiment function, it represents the distribution of the score attributed to each complaint of the prepaid card database. The x axis represent the range of score and the y axis represents the number of complaints. The graph shows that is is a normal distribution.

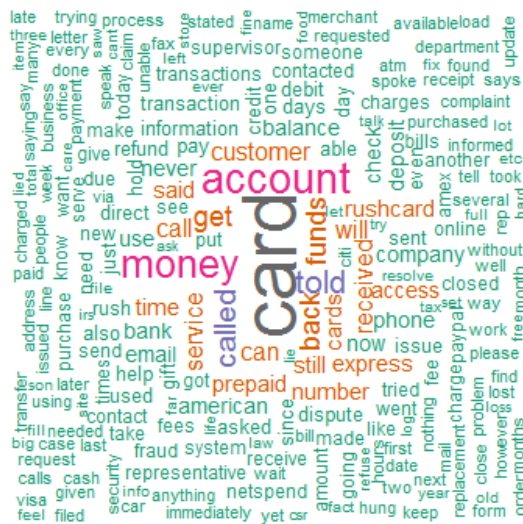
2.5.1.3.2 [Machine learning \(Naïve Bayes\) graphs](#)

Sentiment Analysis of consumers complaints (classification by emotion)



This is the classification by emotion, each complaint is classified by emotion and this is the histogram representing each group of emotion. Generally, joy is the emotion with most complaints followed by unknown, sadness and so on.

Word cloud of the prepaid card database



This word cloud represents all the words available in the prepaid card database, the size of each word tells its frequency in the database, the bigger the word is, the more frequent that word appears in the database. "Card" is the word with the highest frequency in the prepaid card database.

2.5.2 Student loan

2.5.2.1 Summary results

2.5.2.1.1 [Sentiment function polarity summary](#)

ID	Minimum	Maximum	Polarity	Count	Percentage
1	-29	-1	Negative	5209	55
2	0	0	Neutral	1648	18
3	1	17	Positive	2546	27

The summary result of the sentiment function on the student loan database indicates that the polarity distribution is as follow:

Out of 9,403 complaints provided by the database, 5,209 (55%) complaints are found to be negative, 1,648 (18%) complaints are neutral and 2,546 (27%) complaints are positive.

Knowing the range of the score distribution is between -29 to 17, I defined the range of each category of polarity to generate this table by using the MATCH formula in excel.

2.5.2.1.2 [Naïve Bayes polarity summary](#)

ID	Polarity	Count	Percentage
1	Negative	3851	41
2	Neutral	3504	37
3	Positive	2048	22

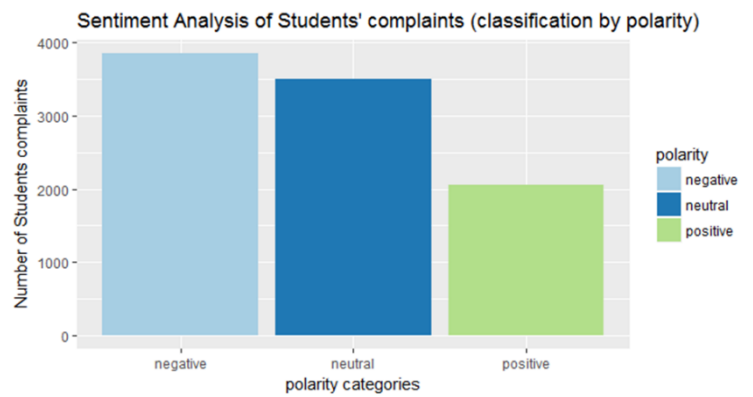
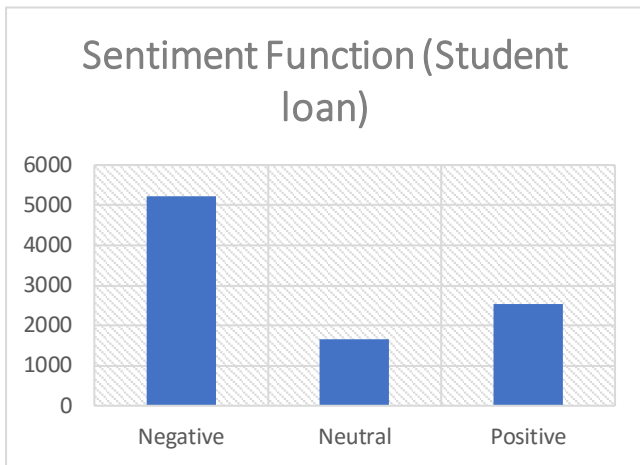
In the other hand, the summary result of the Naïve Bayes on student loan database indicates the polarity is as follow:

Out of 9,403 complaints provided by the database; 3,851 (41%) complaints are found to be negative; 3,504 (37%) complaints are neutral and 2,048 (22%) complaints are positive.

I have used the count() function in R the get the exact figures then I used excel to generate the above table.

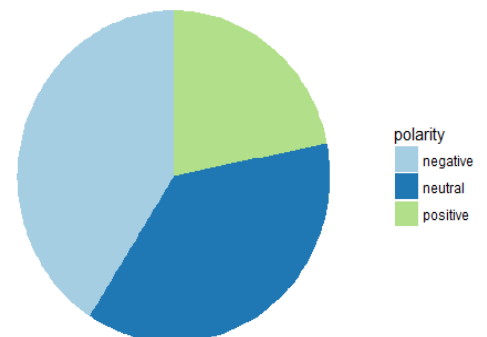
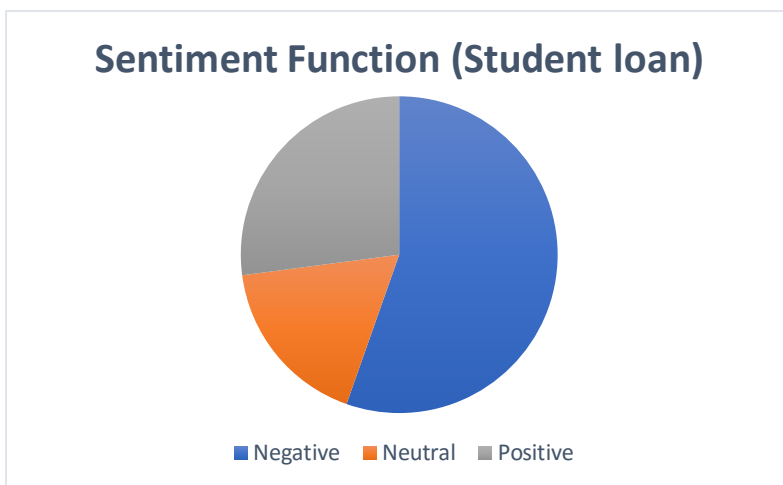
2.5.2.2 Comparing approaches

Sentiment function histogram VS Naïve Bayes histogram



Both approaches have classified more complaints in the negative category. In the second position the sentiment function approach has positive category and the Naïve Bayes approach has neutral category and finally in the third position the sentiment function approach has neutral category and the Naïve Bayes approach has positive category

Sentiment function pie chart VS Naïve Bayes pie chart

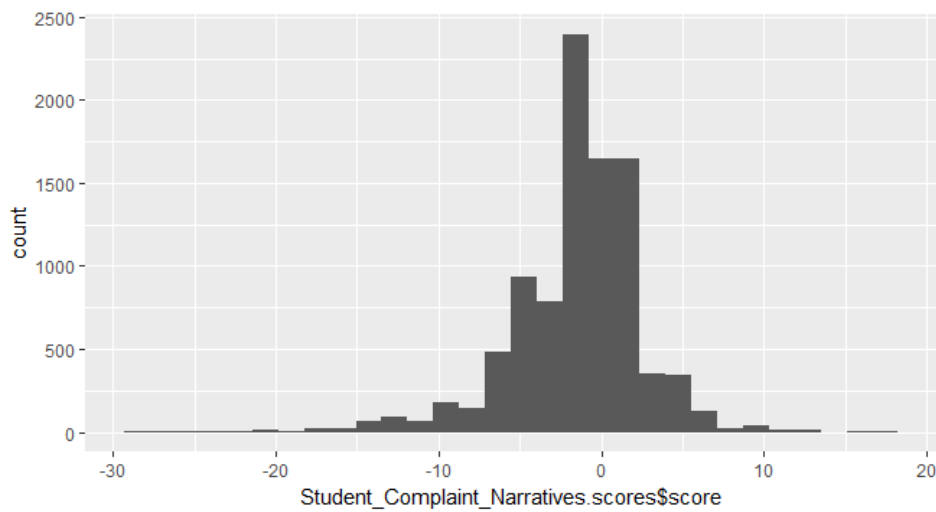


The above pie charts diagrams confirm the result on the histograms generated previously.

2.5.2.3 Additional graphs

2.5.2.3.1 Sentiment function graphs

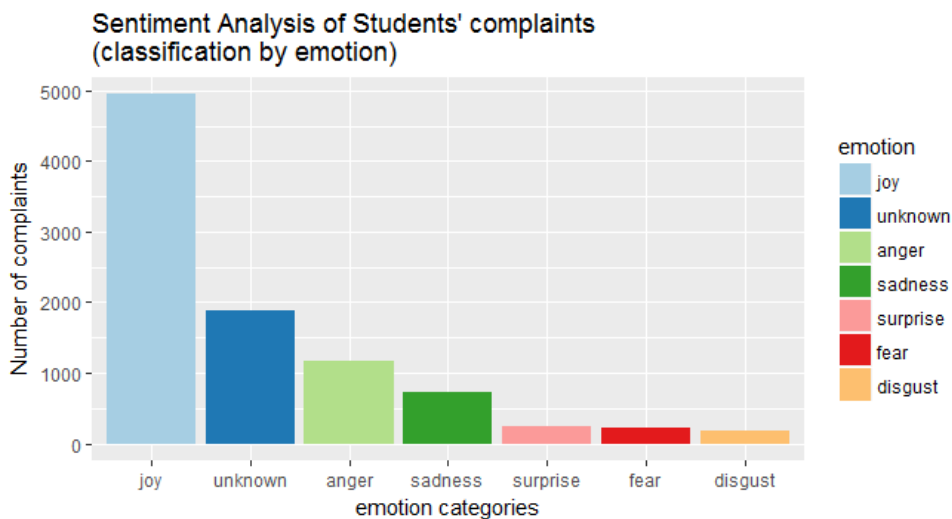
Histogram on sentiment score



This is a histogram generated using the `qplot()` in the sentiment function, it represents the distribution of the score attributed to each complaint of the student loan database. The x axis represent the range of score and the y axis represents the number of complaints. The graph shows that is is a normal distribution.

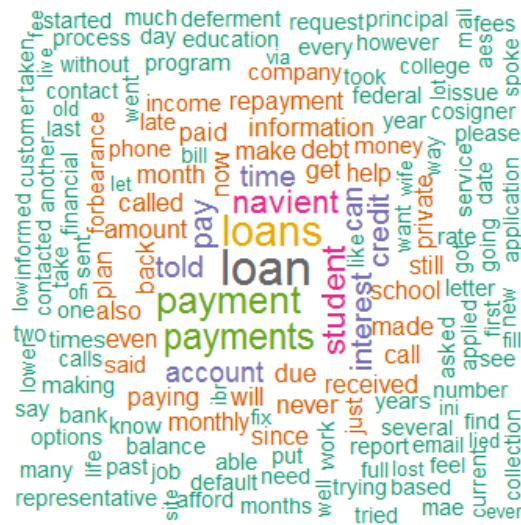
2.5.2.3.2 Machine learning (Naïve Bayes) graphs

Sentiment Analysis of Students' complaints (classification by emotion)



This is the classification by emotion, each complaint is classified by emotion and this is the histogram representing each group of emotion. Generally, joy is the emotion with most complaints followed by unknow, anger and so on.

Word cloud of the student loan database



This word cloud represents all the words available in the student loan database, the size of each word tells its frequency in the database, the bigger the word is, the more frequent that word appears in the database. “loan” is the word with the highest frequency in the student loan database.

Word Cloud comparison based on emotion



The above word cloud represents words found in the student loan database, those words are grouped by emotion. Words of each emotion category are represented with a unique color in the word cloud.

It is concluded that the score sentiment algorithm is the best approach for this system as it attributes an individual score to each complaint that also gives its degree of negativity. This approach in both datasets, has lower percentage on neutral complaints compared to the Naïve Bayes where the neutral percentage are nearly close to the negative percentage. This can also mean that the scores are more accurate than in the Naïve Bayes and we can assume that the Naïve Bayes are classifying most of positive and negative complaints as neutral.

4 Further development or research

The CFPB offers multiple products/services other than the one used for this project (prepaid card and student loan), a deeper research in the sector and comparison of products/services can be done with a powerful user interface to filter the results efficiently.

I have started working on this project with Python until I moved to R and I have faced so many difficulties that I failed to overcome with the Python software. A further analysis with different algorithm such as decision tree algorithm or support vector algorithm may be applied to compare the results with the one obtained in this project.

5 References

- Angnakoon, Paeng. "Mining Twitter With R - Tutorial 1: Building A Corpus From Twitter Data". YouTube. N.p., 2013. Web. 2 May 2017.
- Angnakoon, Paeng. "Mining Twitter With R - Tutorial 2: Scoring Tweets". YouTube. N.p., 2013. Web. 2 May 2017.
- Angnakoon, Paeng. "Mining Twitter With R - Tutorial 3: Histogram ()". YouTube. N.p., 2013. Web. 2 May 2017.
- Angnakoon, Paeng. "Mining Twitter With R - Tutorial 4: Polarity & Emotion Classification". YouTube. N.p., 2013. Web. 2 May 2017.
- Breen, Jeffrey. "Twitter Text Mining R Slides". jeffreymbreen.wordpress. N.p., 2011. Web. 11 Apr. 2017.
- "Consumer Financial Protection Bureau". Consumer Financial Protection Bureau. N.p., 2017. Web. 13 Apr. 2017.
- Liu, Bing. Sentiment Analysis And Opinion Mining. 1st ed. [S.l.]: Morgan & Claypool, 2012. Print.
- "Polarity | R Handbook". Rhandbook.wordpress.com. N.p., 2017. Web. 15 Apr. 2017.
- "Sentiment Analysis On Twitter Data : Text Analytics Tutorial". DataScienceKumar. N.p., 2017. Web. 15 Apr. 2017.
- Thakkar, Harsh, and Dhiren Patel. "Approaches For Sentiment Analysis On Twitter: A State-Of-Art Study". <https://arxiv.org/>. Web. 28 Apr. 2017.
- Uncategorized, DataScientist. "Sentiment Analysis On Twitter Data : Text Analytics Tutorial". mkmanu.wordpress. N.p., 2014. Web. 2 May 2017.
- Wang, Cheng-Jun. "Sentiment Analysis With Machine Learning In R". R-bloggers. N.p., 2016. Web. 11 Apr. 2017.
- "Solutions - Sentiment Analysis". *Xiphostec.com*. N.p., 2017. Web. 2 May 2017.

Srinivasan, Vitthal. "Building Sentiment Analysis Systems In Python". *App.pluralsight.com*. N.p., 2017. Web. 3 May 2017.

6 Appendix

6.1 Project Materials

All the coding used and output obtained in this project is available via this link below:

Note that, no username and/or password is required.

<https://github.com/arlonjunior/Software-Project.git>

6.2 Initial Project Proposal

1. Objectives & Background

Introduction

In this project, we are going to be analysing a large file of data to make sure that customers are treated fairly by financial companies such as banks, lenders and so on. To do so, we analyse customers' complaints on a particular service so by adding their consents, consumers help improve the financial marketplace.

Measuring sentiment will help us to understand the overall feeling of customers surrounding a particular subject, in this scenario prepaid card services.

This helps in creating a complete image of the customers' feedback on the service.

Sentiment is seen as a point of view of an individual towards a specific subject, that point of view can be positive, negative or impartial.

CFPB:

Consumer Financial Protection Bureau is a US Government company which makes available all the Consumer Complaints Database in many sectors (Prepaid Card, Student Loan, Money transfer and so on) available. The database is composed a different type of data, from structured, semi structured to unstructured data of customers' complaints in the banking sector the United states.

Objective

This project would attempt to extract data from the unstructured elements by means of text processing techniques and use the extracted data to identify the nature of complaints which are likely to be upheld and/or rejected.

For this project, the text processing technique adopted will be sentiment analysis.

Surely we would like to know what a sentiment analysis is, in brief what we know about sentiment analysis is detecting and understanding how the audience is reacting to a product/service, either positively or negatively. In this case, how customers are reacting on a service offered by a financial company. The customer feedback on the service is extremely important for the organisation.

The purpose of this project is to create a system that will automatically detect the customer feedback either is positive or negative. This performance of this system will show how well the

service is appreciated by the clients, otherwise the organisation can easily identify complains about their customers and respond to their queries quickly with efficiency.

Background

Sentiment analysis is the process of determining the opinion or feeling of a piece of text. It seems to be easy for human being to detect if a sentence is positive or negative. For example, in this sentence "I love the new iPhone7", it is easy for anyone to tell that this sentence is expressing a positive sentiment. When performing sentiment analysis, we seek for opinions in that context where we pick sentiment out of the opinion.

According to the above example, it is an opinion on the new iPhone7, in this opinion "I love the new iPhone7", we note the verb love that express a sentiment of positivity; note that usually, verbs describe sentiments.

Companies across the world have implemented machines learning how to automatically detect sentiments on their customers. It is super useful for gaining insight into customers' opinions.

This approach is very important because, once you understand how a customer feel about your product/service after analysing his comment/feedback, you can find out what are their expectations, what change will be relevant to meet their needs.

Contribution to Knowledge

Sentiment analysis provides a means of tracking opinions and attitudes on the web and determines if they are positively or negatively received by the public. As we combine the knowledge based techniques and machine learning techniques, this system will give a full data analysis. It is clearly understood that technique is complementary to the analysis.

Note that as the company has learned sentiment on customers towards their services, this is an advantage for them to identify where they went wrong. This analysis let the company to anticipate for a better future. Issues can identify in advance before competitor, also the communication is improved with its target audience and improve the marketing team for further services/products.

2. Technical Approach

Sentiment Analysis has been practiced in a pile of topics, we note that this analysis is performed using different approaches in order to conduct to the same result.

Here are listed 3 types of approaches that can be used to perform our sentiment analysis:

- Lexical Analysis
- Machine Learning
- Combined Analysis

For this project, we will only be focusing on machine learning as it is the approach selected for the construction of our system.

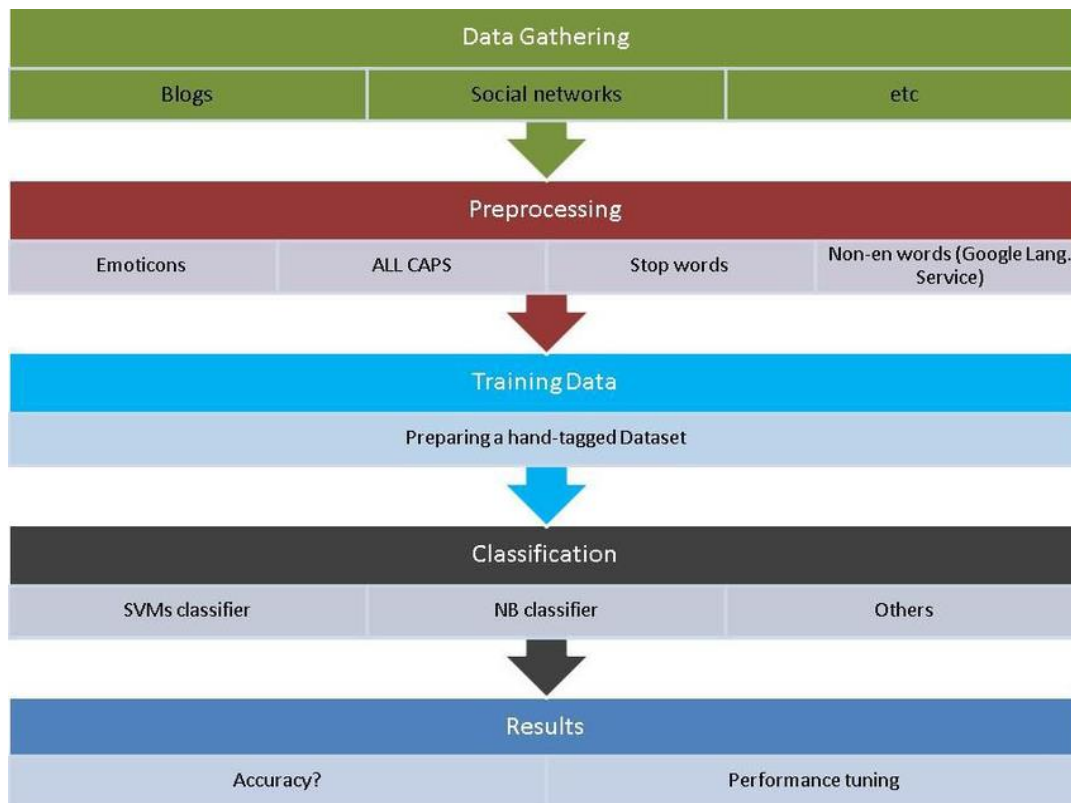


Figure 1: Machine Learning based process

Machine learning is seen like a type of artificial intelligence that gives computers the ability of learning instead of it being programmed. The computer understands how to categorise unknown words and classifying them in an appropriate manner.

Machine learning is the most preferred due to its adaptability and accuracy.

It consists of 5 steps:

- Data Gathering

The Consumer Financial Protection Bureau made available dataset of the US bank sector in different topic, the data used for this project is the customer prepaid card complaints.

- Pre-processing

Before feeding the data into the classifier, it is necessary to proceed in the cleansing of data. In case customers may have included emoticons to express the facial emotions such as happiness “😊”, sadness “☹️” and so on.

This process unifies the lower case and upper case to a common case that the system will understand, also the removal of all the unnecessary white spaces.

- Training Data

The dataset is trained to create a model that is employed to unknow text for the classification reasons.

- Classification

The classifier is finally ready to be used after cleansing. The cleaned version of data from the previous stage can now be deployed for sentiment extraction reason.

- Results

This is the visualisation of the result that can be represented by graphs, charts, etc.

3. Special resources required

Book:

Sentiment Analysis and Opinion Mining by Bing Liu

Dataset:

The Consumer Financial Protection Bureau (Prepaid Card Complaints)

Web:

Public Sentiment Lexicon

Hardware:

- Python

As we are going to process text in this project, Python is well reputed because of the large libraries it offers for sentimental analysis.

- Natural Language Toolkit (NLKT)

NLTK is a leading platform for building Python programs to work with human language data.

A pile of easy-to-use interfaces are provided such as WordNet. The NLTK do an incredible job for Python, it can process a huge amount of text processing and analysis.

4. Project Plan

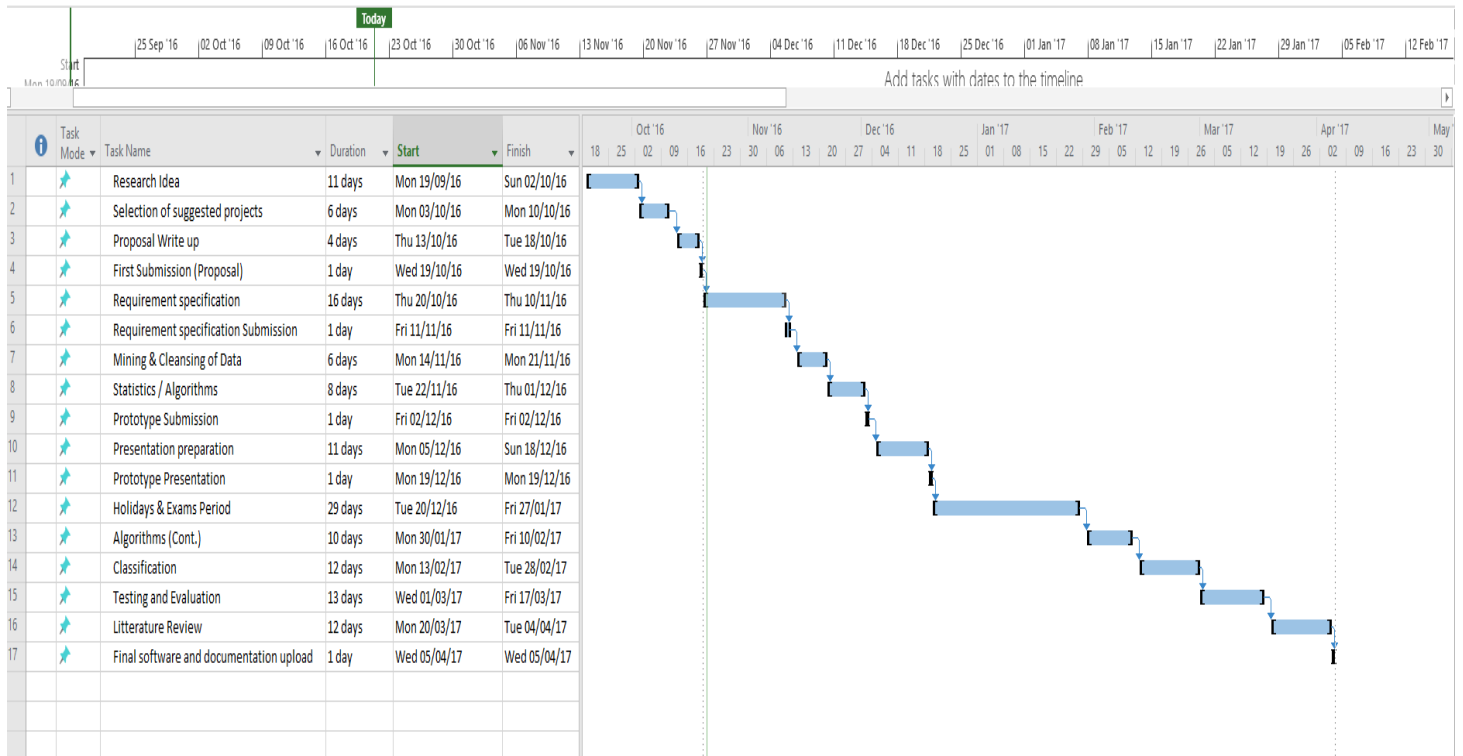


Figure 2: Gantt Chart

5. Evaluation

At the end of this project, the system will be able to compute a customer’s complaint to tell if it is positive or negative, then the organisation can extract information needed quickly for other purposes such as change to make or responding to a customer.

6.3 Initial Project Requirements Specification

1. Introduction

Purpose

The purpose of this document is to make discover what the requirements of a sentiment analysis will be on performing a text analyse on customers’ complaint about their prepaid cards.

The purpose of this analysis is to measure customers’ complaints on prepaid cards provided by The Consumer Financial Protection Bureau (CFPB). The system developed should assess on what end the complaint falls. e.g. Upheld, Indifferent or Rejected.

Project Scope

The scope of the project is to develop a system that allow us to perform a sentiment analysis on a service. After performing this analysis, this system will then classify the sentiment either like upheld, indifferent or rejected depending on what the complaint is about.

These 3 classification will then be retained as the main classification, there will not be in-between classes as we might have 5 classes instead of 3 classes. For example, a rejected sentiment will not depend on the level of rejection, all type of rejection sentiment will fall in the same zone of "Rejected".

This system will be able to automatically determine if the customer complaint will be upheld, indifferent or rejected after performing a sentiment analysis on a customer complaint narrative.

An appropriate system must be constructed to allow this performance to take place. Therefore, data involving complaints of customers on prepaid cards will be collected from CFBB, proceed to the cleansing process for a clean version of data that can be readable on the system and finally classify that data into the system. The analysis will then be performed on the final classified data.

The final system built will not require high level of technical knowledge to be used, a client with low level of technical knowledge should be able to understand and use it.

Definitions, Acronyms, and Abbreviations

CFPB "Consumer Financial Protection Bureau"

Consumer Financial Protection Bureau is a US Government company which makes available all the Consumer Complaints Database in many sectors (Prepaid Card, Student Loan, Money transfer and so on) available. The database is composed a different type of data, from structured, semi structured to unstructured data of customers' complaints in the banking sector the United states.

API "Application Programming Interface"

application programming interface is used to help a programmer to build software and/or applications by providing a set of subroutines definitions, protocols and tools.

NPL "Natural Language Processing"

Is an artificial intelligence tool that analyse, understand and generate languages used by humans so it can interact with computers in either written or spoken.

GUI "Graphical User Interface"

2. User Requirements Definition

This system is intended to perform a sentiment analysis on customers' complaints issue, the system will then classify the result in 3 categories (Upheld, Indifferent or Rejected). The output of this system will be nothing else that a report on how the prepaid card service is seen by customers, what are they thinking about? What is their feeling towards it? Clients of this system they want an efficient system that will perform the sentiment analysis accurately on a target category of complaints which are complaints on prepaid cards. Feedback received from this system will then help the client to measure its customers' satisfaction on their prepaid card service and gives them a large understanding in which sector they need to improve their service.

3. Requirements Specification

In this project, the analysis performed must be easily understood by a client who does not require a high technical level to use it. Results on performed analysis will then be display on board. Therefore, we must create a graphical form to show the result to the client on a screen.

The system will make sure that the accuracy has a certain level in the system. For this project, the system accuracy with 55% or more are going to be acceptable. Since the system is dealing the dataset from CFPB website on customers' complaints, data may not be cleansed as it may contain abbreviations and so on.

Functional requirements

The Functional requirements describes what the system will achieve in ranked order. Here is the list of steps involved in the construction of this system:

- Collection of data from the Consumer Financial Protection Bureau website on prepaid card customers' complaints
- The procedure of cleansing data for a cleaner version of it that will be readable with no error.
- Data obtained in the previous stage will then be stored in a database.
- Classification: here the data will be classified in 3 categories which are Upheld, Indifferent and Rejected.
- Analysis: we will finally perform our analysis on the data classified in the previous stage.

i. Use Case Diagram

This diagram below provides an overview of the entire functional requirement for this system, starting from the data collection to the analysis.

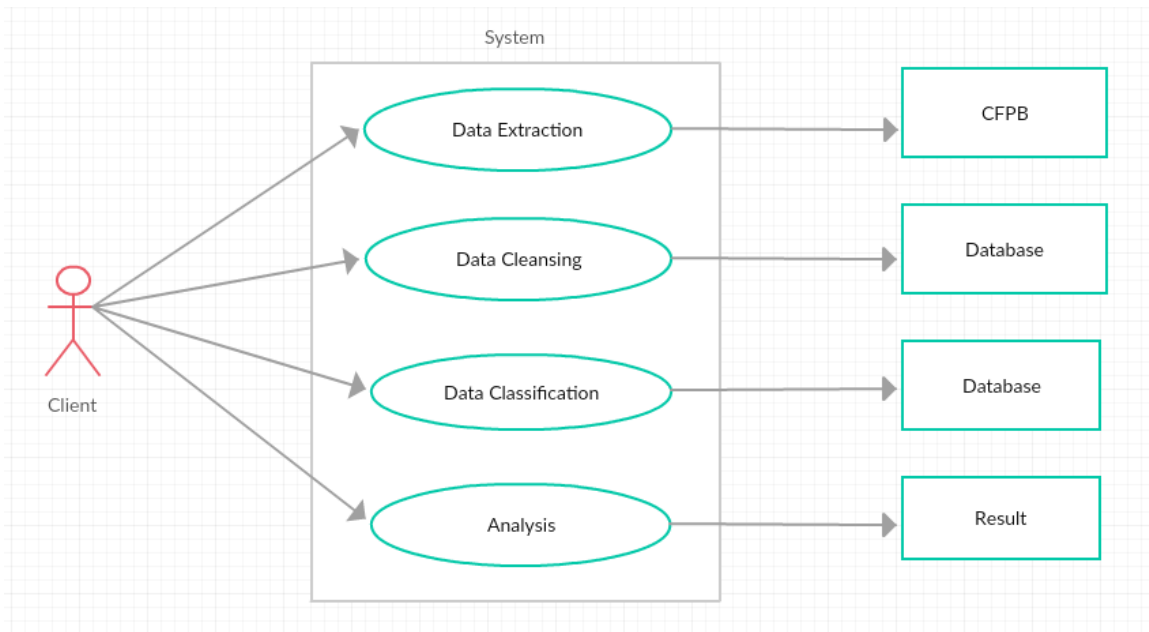


Figure 1: Use Case Diagram

ii. Requirement 1: Data Extraction

1. Description & Priority

For the first requirement “Data Extraction”, we will be mainly focused on data collection available on an open data API open to the public on the Consumer Financial Protection Bureau website on a specific service.

This is an essential requirement as in order to process on the analysis we must have some data which is extracted on this stage and stored in a database for further processing.

2. Use Case

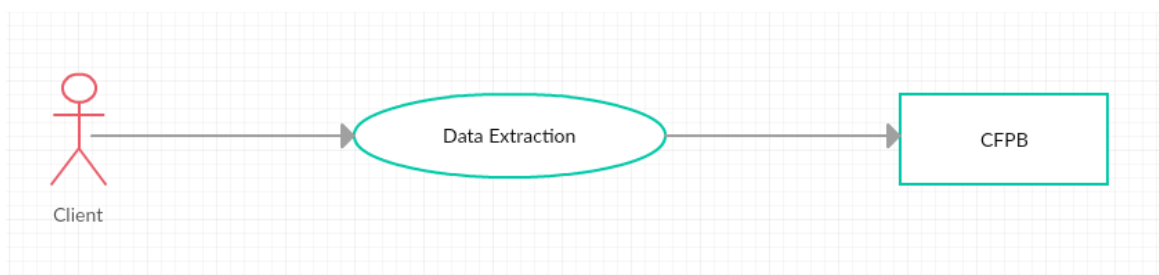


Figure 2: Data Extraction (Use Case Diagram)

- Scope

The scope of this use case is the extraction of prepaid card complaints data from the CFPB API then store it in the database for further processing.

- Description

This use case describes the data collection stage of the construction of this system. The user access the CFPB API. Once the connection of the API is made, the data can be store into a database from the stream.

- Precondition

There no precondition as this is the first action we take on the construction of this system.

- Activation

This use case starts when the client runs scripts on Python.

- Main flow

- Activation of the process by the client
- Database connection with the system
- CFPB API connection to the system
- The connection is established between the system and API
- Data streaming has started
- Data is transferred to the SQL database
- Client finishes the system.

- Exceptional flow

- Action of the process by the client
- Database connection with the system
- CFPB API connection to the system
- The connection is established between the system and API
- Data streaming has started
- Database or API becomes inaccessible
- System waits few seconds
- Loop back to the database connection

- Termination

The system will terminate when the user ends the process or when data has been streamed through the SQL database.

- Post condition

The system goes into a wait state when the API cannot be connected or there is a problem with the database on the system. At this stage, the system allows few seconds before retrying to process again.

iii. Requirement 2: Data Cleansing

1. Description & Priority

In this stage, we will concentrate on cleansing data. Some data may have contents and/or characters that will seem to be difficult for the system to read. Many of customers' complaints contain hidden content such as date content hidden (xx/xx/xxxx) and so on.

This stage is where the streamed data must be transformed to a clean version before passing it to the SQL database to proceed to the next stage of classification

2. Use Case



Figure 3: Data Cleansing (Use Case Diagram)

- Scope

The scope of this use case is to clean the extracted data and have it ready on SQL database for the next stage of data classification.

- Description

This use case describes the data cleansing process as the stream data may contain unwanted texts and/or characters.

- Precondition

The system at this stage, requires the previous stage (data extraction) to be successfully completed before the cleansing process begins.

- Activation

This use case starts when the client runs scripts on Python.

- Main flow

- Activation of the process by the client
- Database connection with the system
- The system reads transferred data in the SQL database
- The system cleans out unwanted texts and/or characters
- The system stores the new cleansed data back into the SQL database
- Client finishes the system.

- Exceptional flow

- Action of the process by the client

- Database connection with the system
 - The system reads transferred data in the SQL database
 - The system cleans out unwanted texts and/or characters
 - Database becomes inaccessible
 - System waits few seconds
 - Loop back to the database connection
 - The system stores the new cleansed data back into the SQL database
 - Client finishes the system.
- Termination

The system ends immediately when the cleansing process is completed.

- Post condition

The system goes into a wait state when the database is inaccessible. The system allows few seconds the reconnect to the database and try to process the command one more time again.

iv. Requirement 3: Data Classification

1. Description & Priority

This is the last step the system will go through before performing the analysis on data stored on the SQL database. In this last requirement, the cleansed data will be divided into different categories to allow the classification of data to happen. Therefore, we will classify data into 3 categories which are upheld, indifferent and rejected.

This requirement is the most important cause the system needs this classified data to perform analysis on.

2. Use Case



Figure 4: Data Classification (Use Case Diagram)

- Scope

The scope of this use case will be the classification of cleansed data, at the end of this process the classified data will be used to perform analysis on it.

- Description

This use case describes the process of classifying the cleansed data into their respective categories.

- Precondition

The system at this stage, requires the previous stage (data cleansing) to be successfully completed before the classification process begins.

- Activation

This use case starts when the client runs scripts on Python.

- Main flow

- Activation of the process by the client
- Database connection with the system
- The system reads transferred data in the SQL database
- The system stats the classification of data into their respective categories
- The system saves the classified data into the database
- Client finishes the system.

- Termination

The system ends immediately when the classification process is completed.

4. Interface requirements

This section describes how the software interfaces with other software products or users for input or output. Examples of such interfaces include APIs, web services, shared memory, data streams, and so forth. Most systems would have a GUI. Add more subsections for other interfaces as required.

GUI

For this project, there is no Graphical User Interface required since it will not be passed to a third party.

Application Programming Interfaces (API)

This system is not expected to offer any interface but it uses the CFPB API to access the stream data on customers' complaints on a specified service that will be analyzed.

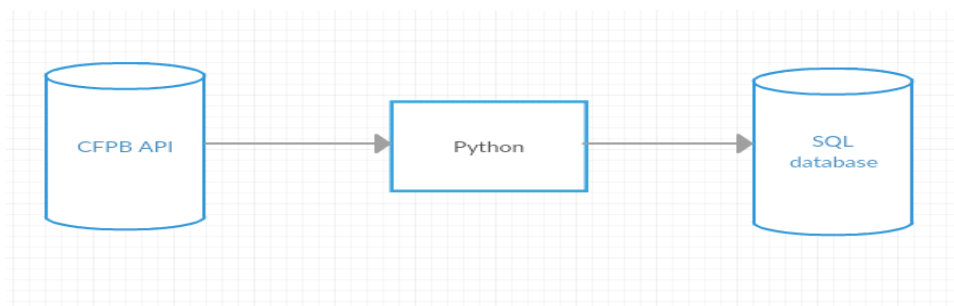


Figure 5: CFPB API extraction map

5. System Architecture

In the diagram below, we have the system architecture that demonstrate how the system will be performing from its initial stage till the end.

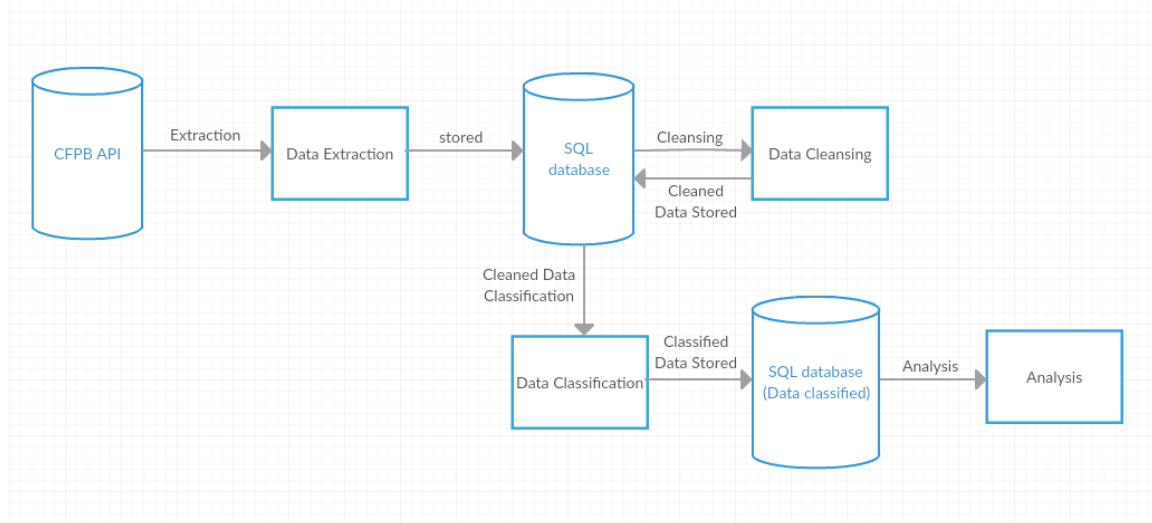


Figure 6: System Architecture map

6. System Evolution

This section describes how the system could evolve over time.

Due to the time limited on the construction of this system, we have delimited the amount of work that could be added over time. In the future, this system may add sub-categories on categories (classes) created on this project.

Depending on business needs, we may increase our area of expertise from customers' complaints on prepaid card to all types of complaints such as Student Loan, Money transfer and so on.

6.4 Monthly Journals

Month: September

My Achievements

During this period, I had to come up with a unique idea for my final year project, something more complex and challenging to implement. My original idea based on a web application development using php have been rejected. The back-up plan was to choose a project idea from a list of projects suggested by the college and ask for the permission to the author of the idea before start working on it. Out of all the projects proposed, I have picked Oisin Creaner's idea for the interpretation of Banking Complaints data by mean of text processing such as sentiment analysis.

My Reflection

Before start working on this project, I have contacted its author to ask for permission which has been granted. The next step was to start the research on the field and understand the project requirements for a better understanding.

Supervisor Meetings

Date of Meeting: None

Items discussed: None

Action Items: None

Month: October

My Achievements

After the decision of choosing what project I will be working on this year, I have decided to start the learning process. What is the project about, how can it be achieved, how long does it take, what software is needed for this project to be completed and so on. These questions can properly be answer through the project proposal submitted this month. I was able to define what is a sentiment analysis and its importance in an organisation.

My Reflection

I did a research on how to approach a construction of this type of software, I discovered a large amount of help online with different approaches. I do believe that this project is feasible.

However, let's move to the project requirement specification and start the actual coding for the following month.

Supervisor Meetings

Date of Meeting: None

Items discussed: None

Action Items: None

Month: November

My Achievements

This month I was able to provide the project requirement specification after submitting the proposal idea last month. At this stage, the project is becoming clearer to understand and the project requirement specification serves as a good guide for the implementation part intended to begin by next month.

My Reflection

I was still a bit confused on the approach to adopted for this project. After researching in different sources, I required an advice from another person on the topic, then I have contacted Lisa Murphy who is my supervisor for the first time.

Supervisor Meetings

Date of Meeting: 06/12/2016

Items discussed: Useful resources for coding in python for my project.

Action Items: She sent me useful links to use Python

Month: December

My Achievements

I have started coding in python for the first time, I have managed to read the data into Python and I could tokenise by sentence or by word with the data uploaded. I prepared my technical report with all the work done so far to upload (this technical report will be updated as we go along). Additionally, I had to prepare my mid-point presentation in the same month as well.

My Reflection

I became more confident when I have seen the data collected from the Consumer Financial Protection Bureau to be read in Python.

However, I was not successful to classify this data as expected for the mid-point presentation. Also, I have reduced the amount of data as Python was not able to read large file of data.

Supervisor Meetings

Date of Meeting: None

Items discussed: None

Action Items: None

Month: January

My Achievements

After completing the first semester, I have tried to resolve issues I had with Python in performing the sentiment analysis, and be able to use large file of data without the software to crash but I still was not successful. Thus, I went back from the beginning to learn more about R instead, I found it more understandable using R over Python. I already had a bit of background knowledge in R with my other module advanced business data analysis.

My Reflection

I have contacted my supervisor to help me fixing these issues in Python and if she knew more about R and why would be her advice.

However, I was pretty sure that I would be able to work in R than Python based on series of tutorials I learn on performing a sentiment analysis.

Supervisor Meetings

Date of Meeting: 12/02/2017

Items discussed: Trouble in using Python and idea of changing the software from Python to R.

Action Items: She gave me more notes on the issue in Python that could help provided by her colleague Sarah.

Month: February

My Achievements

I had finally decided to change the software that I was using for this project to R. By changing this, I was able to perform my first sentiment analysis using Breen's approach that attributes scores to each complaint. I still had trouble on the classifying by polarity based on this approach and other techniques to display the result other than the histogram. This required a deep research on the field and I decided to use the Naïve Bayes algorithm instead to fix this problem.

My Reflection

Learned how the Naïve Bayes algorithm works, I watched online tutorial explaining this method and apply this in R. Minor errors were still occurring but I was able to fix them, it was just a matter of time.

I have informed my supervisor Lisa of my new approach about the project and I also asked for advice on areas I could not fix on my own with the new software R. Also, I asked for advice from my Eugene who is my Advanced Business data analysis in a few issues I encountered in R.

Supervisor Meetings

Date of Meeting: 21/03/2017

Items discussed: New software used and advice on the area.

Action Items: Suggested to meet lectures with R background for further discussions.

Month: March

My Achievements

I have succeeded to implement a sentiment analysis on the consumers' financial complaints using 2 different algorithms and interpreted the results obtained. Graphically I could display the polarity of the complaints through the histogram based on the Naïve Bayes approach and the scores on complaints on the histogram based on the Breen approach. The success of this implementation led me to add a second dataset (student loan complaints) to compare to the one we had before (prepaid card complaints).

My Reflection

I felt it will be more interesting comparing the results obtained with the Breen's approach to the one obtained with the Machine Learning's approach. Also, providing more results graphically other than histograms.

However, I failed generating the word cloud in both datasets and required more tutorials in graphics using R studio.

Supervisor Meetings

Date of Meeting: 29/03/2017

Items discussed: Graphics in R

Action Items: Contacted her colleague Catherine for me for more tutorials in this topic.

6.5 Sentiment Coding

6.5.1 On Prepaid Cards

```
library (plyr) #Tools for Splitting, Applying and Combining Data
library (stringr) #Make it easier working with string

score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  require(plyr)
  require(stringr)

  # we got a vector of sentences.
  #plyr will handle a list or a vector as an "l" for us
  # we want a simple array ("a") of scores back, so we use
  # "l" + "a" + "ply" = "lapply":

  scores = lapply(sentences, function(sentence, pos.words, neg.words) {

    # clean up sentences with R's regex-driven global substitute, gsub():

    sentence = gsub('[:punct:]', '', sentence) #Remove punctuation

    sentence = gsub('[:cntrl:]', '', sentence) #Remove Control characters

    sentence = gsub('\\d+', '', sentence) #Remove digits

    # and convert to lower case:

    sentence = tolower(sentence)

    # split into words. str_split is in the stringr package

    word.list = str_split(sentence, '\\s+')

    # sometimes a list() is one level of hierarchy too much

    words = unlist(word.list)

    # compare our words to the dictionaries of positive & negative terms

    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)
```

```
# match() returns the position of the matched term or NA
# we just want a TRUE/FALSE:

pos.matches = !is.na(pos.matches)

neg.matches = !is.na(neg.matches)

# and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():

score = sum(pos.matches) - sum(neg.matches)

return(score)

}, pos.words, neg.words, .progress=.progress )
scores.df = data.frame(score=scores, text=sentences)
return(scores.df)
}

# Scoring Customer feedbacks (complaints) & Adding a column

#Load sentiment word lists
hu.liu.pos = scan('C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/positive-words.txt', what='character', comment.char=';')
hu.liu.neg = scan('C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/negative-words.txt', what='character', comment.char=';')

#Add words to list
pos.words = c(hu.liu.pos, 'upgrade')
neg.words = c(hu.liu.neg, 'wtf', 'fuck', 'shit', 'wait','waiting', 'epicfail',
'mechanical')

#Import csv file (Dataset)
DatasetCustomers_Narratives <- read.csv("C:/Users/molsj/Documents/4th
Year/Semester 2/Software Project II/R Codes/Consumer_Complaint_Narratives.csv")
DatasetCustomers_Narratives$Consumer.complaint.narrative<-
as.factor(DatasetCustomers_Narratives$Consumer.complaint.narrative)
```

```
#Score all complaints
Consumer_Complaint_Narratives.scores =
score.sentiment(DatasetCustomers_Narratives$Consumer.complaint.narrative,
pos.words,neg.words, .progress='Consumer.complaint.narrative')

path<-"C:/Users/molsj/Documents/4th Year/Semester 2/Software Project II/R Codes"
write.csv(Consumer_Complaint_Narratives.scores,file=paste(path,"Consumer_Complain
t_Narratives_Scores.csv",sep=""),row.names=TRUE)

Consumer_Complaint_Narratives.scores$Feedback = 'Consumer_Complaint_Narratives'

# Visualizing
library(ggplot2)
library(QPot)

#The positive values stand for positive feedback
#and the negative values for negative feedback.
hist(Consumer_Complaint_Narratives.scores$score)
qplot(Consumer_Complaint_Narratives.scores$score)

#Mean of the score distribution
mean(Consumer_Complaint_Narratives.scores$score)

#Range of the score distribution
range(Consumer_Complaint_Narratives.scores$score)

#Median of the score distribution
median(Consumer_Complaint_Narratives.scores$score)

#Frequency of each score
count(Consumer_Complaint_Narratives.scores$score)
```

```
#Histogram with ggplot2
all.scores = rbind(Consumer_Complaint_Narratives.scores)
ggplot(data=all.scores) + # ggplot works on data.frames, always
  geom_histogram(mapping=aes(x=score, fill=Feedback), binwidth=0.5) +
  facet_grid(Feedback~.) + # make a separate plot for each hashtag
  theme_bw(base_size = 12) + scale_fill_brewer(palette = 18) # plain display,
nicer colors

# Classification

#Get the text
Consumer_Complaint_Narratives_txt =
DatasetCustomers_Narratives$Consumer.complaint.narrative

#Data cleansing to prepare text for the sentiment analysis
Consumer_Complaint_Narratives_txt = gsub("[[:punct:]]", "",
Consumer_Complaint_Narratives_txt) #remove punctuation

Consumer_Complaint_Narratives_txt = gsub("(RT|via)((?:\\b\\w*@\\w+)+)", "",
Consumer_Complaint_Narratives_txt) #remove retweet entities

Consumer_Complaint_Narratives_txt = gsub("@\\w+", "",
Consumer_Complaint_Narratives_txt) #remove @

Consumer_Complaint_Narratives_txt = gsub("[[:digit:]]", "",
Consumer_Complaint_Narratives_txt) #remove digits

Consumer_Complaint_Narratives_txt = gsub("^\\s+|\\s+$", "",
Consumer_Complaint_Narratives_txt) #remove tabs

Consumer_Complaint_Narratives_txt = gsub("[ \\t]{2,}", "",
Consumer_Complaint_Narratives_txt) #remove spaces

Consumer_Complaint_Narratives_txt = gsub("http\\w+", "",
Consumer_Complaint_Narratives_txt) #remove link
```



```
Consumer_Complaint_Narratives_txt = gsub("xx", "",
Consumer_Complaint_Narratives_txt) #remove xx

Consumer_Complaint_Narratives_txt = gsub("xxx", "",
Consumer_Complaint_Narratives_txt) #remove xxx

Consumer_Complaint_Narratives_txt = gsub("xxxx", "",
Consumer_Complaint_Narratives_txt) #remove xxxx

Consumer_Complaint_Narratives_txt = gsub("xxxxxxx", "",
Consumer_Complaint_Narratives_txt) #remove xxxxxxxx

#Defining a function which can handle "tolower error handling",
#in case arises any while converting all the words into lower case.
try.error = function(x)
{
  #create missing value
  y = NA

  #trycatch error
  try_error = tryCatch(tolower(x), error = function(e) e)

  #if not an error
  if(!inherits(try_error, "error"))
    y = tolower(x)

  #result
  return(y)
}

#Transforming all the words into lower case using the
#try.error function created above with the sapply function
Consumer_Complaint_Narratives_txt = sapply(Consumer_Complaint_Narratives_txt,
try.error)

#Remove NAs, if any exists, from Consumer_Complaint_Narratives_txt
Consumer_Complaint_Narratives_txt =
Consumer_Complaint_Narratives_txt[!is.na(Consumer_Complaint_Narratives_txt)]
```

```
#Also remove names (column headings) from the text,
#as we do not want them in the sentiment analysis
names(Consumer_Complaint_Narratives_txt) = NULL

#installing sentiment 0.2
install.packages("C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/sentiment_0.2.tar.gz", repos = NULL, type="source")

library(RColorBrewer)
library(wordcloud)
library(NLP)
library(tm)
library(Rstem)
library(sentiment)

#We are using Bayes' algorithm to:

#classify emotion
#This function returns an object of class data.frame with seven columns
#(anger, disgust, fear, joy, sadness, surprise, best_fit) and one row for each
document:
class_emo = classify_emotion(Consumer_Complaint_Narratives_txt,
algorithm="bayes", prior=1.0)
#get emotion best fit
emotion = class_emo[,7]
# Replacing NA's (if any, found while processing classification)
#by the word "unknown"
emotion[is.na(emotion)] = "unknown"

#classify polarity
#This process will classify the text data into four categories
#(pos, neg, pos/neg, best_fit)
class_pol = classify_polarity(Consumer_Complaint_Narratives_txt, algorithm =
"bayes")
#get polarity best fit
polarity = class_pol[,4]
```

```
#Creating data frame and rearrange data for plotting purposes

#Creating data frame
sent_categ = data.frame(text=Consumer_Complaint_Narratives_txt,
                        polarity=polarity, stringsAsFactors=FALSE)

#sort data frame
sent_categ = within(sent_categ,
                    emotion <- factor(emotion, levels=names(sort(table(emotion),
decreasing=TRUE))))

#Results in numbers

#Frequency of eah observation of polarity
count(sent_categ, "polarity")
str(count(sent_categ, "polarity"))

#Frequency of eah observation of emotion
count(sent_categ, "emotion")
str(count(sent_categ, "emotion"))

#Visualization

#Plot distribution of polarity
ggplot(sent_categ, aes(x=polarity)) +
  geom_bar(aes(y=..count.., fill=polarity)) +
  scale_fill_brewer(palette="Paired") +
  labs(x="polarity categories", y="Number of complaints",
       title = "Sentiment Analysis of customers' complaints (classification by
polarity)",
       plot.title = element_text(size=12))
```

```
#Plot distribution of emotions
ggplot(sent_categ, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Paired") +
  labs(x="emotion categories", y="Number of complaints",
       title = "Sentiment Analysis of Financial Complaints\n(classification by
emotion)",
       plot.title = element_text(size=12))

#Pie Chart with size polarity
ggplot(data = sent_categ) +
  geom_bar(mapping = aes(x = polarity, fill = polarity), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_linedraw() +
  coord_polar()

#Pie Chart with size emotion
ggplot(data = sent_categ) +
  geom_bar(mapping = aes(x = emotion, fill = emotion), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_linedraw() +
  coord_polar()

#Pie Chart Same Size polarity
ggplot(data = sent_categ) +
  geom_bar(mapping = aes(x = factor(1), fill = polarity), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_void() +
  coord_polar(theta = "y")

#Pie Chart Same Size emotion
ggplot(data = sent_categ) +
  geom_bar(mapping = aes(x = factor(1), fill = emotion), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_void() +
  coord_polar(theta = "y")

#Word Cloud
library(tm)
library(wordcloud)
```

```
#Word cloud of the full dataset (non-adjusted)
wordcloud(Consumer_Complaint_Narratives_txt)

#Word cloud adjusted
wordcloud(Consumer_Complaint_Narratives_txt, scale=c(3,0.6), max.words=1000,
          random.order=FALSE, rot.per=0.35, colors=brewer.pal(8,"Dark2"))

#Comparing word cloud

#Compare words by emotion
#separating text by emotion
emos = levels(factor(sent_cat$emotion))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = Consumer_Complaint_Narratives_txt[emotion == emos[i]]
  emo.docs[i] = paste(tmp, collapse="")
}

# remove stopwords
emo.docs = removeWords(emo.docs, stopwords("english"))
# create corpus
corpus = Corpus(VectorSource(emo.docs))
tdm = TermDocumentMatrix(
  corpus,
  control = list(
    wordLengths=c(0,Inf),
    removePunctuation = TRUE,
    stopwords = c("prayformh370", "prayformh", stopwords("english")),
    removeNumbers = TRUE, tolower = TRUE) )
tdm = as.matrix(tdm)
colnames(tdm) = emos

library(RColorBrewer)

# comparison word cloud by emotion
```

```
comparison.cloud(tdm, max.words=1000, colors = brewer.pal(nemo, "Dark2"),
                 scale = c(3,.5), random.order = FALSE, title.size = 1.5)

#Compare words by polarity
#separating text by polarity
emos = levels(factor(sent_categ$polarity))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = Consumer_Complaint_Narratives_txt[polarity == emos[i]]
  emo.docs[i] = paste(tmp, collapse="")
}

# remove stopwords
emo.docs = removeWords(emo.docs, stopwords("english"))
# create corpus
corpus = Corpus(VectorSource(emo.docs))
tdm = TermDocumentMatrix(
  corpus,
  control = list(
    wordLengths=c(0,Inf),
    removePunctuation = TRUE,
    stopwords = c("prayformh370", "prayformh", stopwords("english")),
    removeNumbers = TRUE, tolower = TRUE) )
tdm = as.matrix(tdm)
colnames(tdm) = emos

library(RColorBrewer)

# comparison word cloud by polarity
comparison.cloud(tdm, max.words=3000, colors = brewer.pal(nemo, "Dark2"),
                 scale = c(3,.5), random.order = FALSE, title.size = 1.5)
```

6.5.2 On Student Loan

```
library (plyr) #Tools for Splitting, Applying and Combining Data
library (stringr) #Make it easier working with string

score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  require(plyr)
  require(stringr)

  # we got a vector of sentences.
  #plyr will handle a list or a vector as an "l" for us
  # we want a simple array ("a") of scores back, so we use
  # "l" + "a" + "ply" = "laply":

  scores = laply(sentences, function(sentence, pos.words, neg.words) {

    # clean up sentences with R's regex-driven global substitute, gsub():

    sentence = gsub('[:punct:]', '', sentence) #Remove punctuation

    sentence = gsub('[:cntrl:]', '', sentence) #Remove Control characters

    sentence = gsub('\\d+', '', sentence) #Remove digits

    # and convert to lower case:

    sentence = tolower(sentence)

    # split into words. str_split is in the stringr package

    word.list = str_split(sentence, '\\s+')

    # sometimes a list() is one level of hierarchy too much

    words = unlist(word.list)

    # compare our words to the dictionaries of positive & negative terms

    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)

    # match() returns the position of the matched term or NA
    # we just want a TRUE/FALSE:

    pos.matches = !is.na(pos.matches)
```

```
neg.matches = !is.na(neg.matches)

# and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():

score = sum(pos.matches) - sum(neg.matches)

return(score)

}, pos.words, neg.words, .progress=.progress )
scores.df = data.frame(score=scores, text=sentences)
return(scores.df)
}

# Scoring students feedbacks (complaints) & Adding a column

#Load sentiment word lists
hu.liu.pos = scan('C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/positive-words.txt', what='character', comment.char='')
hu.liu.neg = scan('C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/negative-words.txt', what='character', comment.char='')

#Add words to list
pos.words = c(hu.liu.pos, 'upgrade')
neg.words = c(hu.liu.neg, 'wtf', 'fuck', 'shit', 'wait', 'waiting', 'epicfail',
'mechanical')

#Import csv file (Dataset)
DatasetStudent_Narratives <- read.csv("C:/Users/molsj/Documents/4th Year/Semester
2/Software Project II/R
Codes/Student_Loan_Complaints_with_Consumer_Complaint_Narratives.csv")
DatasetStudent_Narratives$Consumer.complaint.narrative<-
as.factor(DatasetStudent_Narratives$Consumer.complaint.narrative)

#Score all complaints
```



```
Student_Complaint_Narratives.scores =  
score.sentiment(DatasetStudent_Narratives$Consumer.complaint.narrative,  
pos.words,neg.words, .progress='Consumer.complaint.narrative')  
  
path<-"C:/Users/molsj/Documents/4th Year/Semester 2/Software Project II/R Codes"  
write.csv(Student_Complaint_Narratives.scores,file=paste(path,"Student_Complaint_N  
arratives_scores.csv",sep=""),row.names=TRUE)  
  
Student_Complaint_Narratives.scores$Feedback = 'Student_Complaint_Narratives'  
  
#Visualizing  
library(ggplot2)  
library(QPot)  
  
#The positive values stand for positive feedback  
#and the negative values for negative feedback.  
hist(Student_Complaint_Narratives.scores$score)  
qplot(Student_Complaint_Narratives.scores$score)  
  
#Mean of the score distribution  
mean(Student_Complaint_Narratives.scores$score)  
  
#Range of the score distribution  
range(Student_Complaint_Narratives.scores$score)  
  
#Median of the score distribution  
median(Student_Complaint_Narratives.scores$score)  
  
#Frequency of each score  
count(Student_Complaint_Narratives.scores$score)
```

```
#Histogram with ggplot2
all.scores = rbind(Student_Complaint_Narratives.scores)
ggplot(data=all.scores) + # ggplot works on data.frames, always
  geom_histogram(mapping=aes(x=score, fill=Feedback), binwidth=0.5) +
  facet_grid(Feedback~.) + # make a separate plot for each hashtag
  theme_bw(base_size = 12) + scale_fill_brewer(palette = 18) # plain display,
nicer colors

# Classification

#Get the text
Student_Complaint_Narratives_txt =
DatasetStudent_Narratives$Consumer.complaint.narrative

#Data cleansing to prepare text for the sentiment analysis
Student_Complaint_Narratives_txt = gsub("[[:punct:]]", "",
Student_Complaint_Narratives_txt)

Student_Complaint_Narratives_txt = gsub("(RT|via)((?:\b\\w*@\w+)+)", "",
Student_Complaint_Narratives_txt)

Student_Complaint_Narratives_txt = gsub("@\w+", "",
Student_Complaint_Narratives_txt) #remove @

Student_Complaint_Narratives_txt = gsub("[[:digit:]]", "",
Student_Complaint_Narratives_txt) #remove digits

Student_Complaint_Narratives_txt = gsub("^\s+|\s+$", "",
Student_Complaint_Narratives_txt) #remove tabs

Student_Complaint_Narratives_txt = gsub("[ \t]{2,}", "",
Student_Complaint_Narratives_txt) #remove spaces

Student_Complaint_Narratives_txt = gsub("http\w+", "",
Student_Complaint_Narratives_txt) #remove link
```

```
Student_Complaint_Narratives_txt = gsub("xx", "",
Student_Complaint_Narratives_txt) #remove xx

Student_Complaint_Narratives_txt = gsub("xxx", "",
Student_Complaint_Narratives_txt) #remove xxx

Student_Complaint_Narratives_txt = gsub("xxxx", "",
Student_Complaint_Narratives_txt) #remove xxxx

Student_Complaint_Narratives_txt = gsub("xxxxxxx", "",
Student_Complaint_Narratives_txt) #remove xxxxxxxx

#Defining a function which can handle "tolower error handling",
#in case arises any while converting all the words into lower case.
try.error = function(x)
{
  #create missing value
  y = NA

  #trycatch error
  try_error = tryCatch(tolower(x), error = function(e) e)

  #if not an error
  if(!inherits(try_error, "error"))
    y = tolower(x)

  #result
  return(y)
}

#Transforming all the words into lower case using the
#try.error function created above with the sapply function
Student_Complaint_Narratives_txt = sapply(Student_Complaint_Narratives_txt,
try.error)

#Remove NAs, if any exists, from Student_Complaint_Narratives_txt
Student_Complaint_Narratives_txt =
Student_Complaint_Narratives_txt[!is.na(Student_Complaint_Narratives_txt)]
```

```
#Also remove names (column headings) from the text,
#as we do not want them in the sentiment analysis
names(Student_Complaint_Narratives_txt) = NULL

#installing sentiment 0.2
install.packages("C:/Users/molsj/Documents/4th Year/Semester 2/Software Project
II/R Codes/sentiment_0.2.tar.gz", repos = NULL, type="source")

library(RColorBrewer)
library(wordcloud)
library(NLP)
library(tm)
library(Rstem)
library(sentiment)

#We are using Bayes' algorithm to:

#classify emotion
#This function returns an object of class data.frame with seven columns
#(anger, disgust, fear, joy, sadness, surprise, best_fit) and one row for each
document:
class_emo = classify_emotion(Student_Complaint_Narratives_txt, algorithm="bayes",
prior=1.0)
#get emotion best fit
emotion = class_emo[,7]
# Replacing NA's (if any, found while processing classification)
#by the word "unknown"
emotion[is.na(emotion)] = "unknown"

#classify polarity
#This process will classify the text data into four categories
#(pos, neg, pos/neg, best_fit)
class_pol = classify_polarity(Student_Complaint_Narratives_txt, algorithm =
"bayes")
#get polarity best fit
```

```
polarity = class_pol[,4]

#Creating data frame and rearrange data for plotting purposes

#Creating data frame
sent_categ2 =
data.frame(Consumer.complaint.narrative=Student_Complaint_Narratives_txt,
           polarity=polarity, stringsAsFactors=FALSE)

# sort data frame
sent_categ2 = within(sent_categ2,
                    emotion <- factor(emotion, levels=names(sort(table(emotion),
decreasing=TRUE))))

#Results in numbers

#Frequency of eah observation of polarity
count(sent_categ2, "polarity")
str(count(sent_categ2, "polarity"))

#Frequency of eah observation of emotion
count(sent_categ2, "emotion")
str(count(sent_categ2, "emotion"))

#Visualization

#Plot distribution of polarity
ggplot(sent_categ2, aes(x=polarity)) +
  geom_bar(aes(y=..count.., fill=polarity)) +
  scale_fill_brewer(palette="Paired") +
  labs(x="polarity categories", y="Number of Students complaints",
       title = "Sentiment Analysis of Students' complaints (classification by
polarity)",
       plot.title = element_text(size=12))
```

```
#Plot distribution of emotions
ggplot(sent_categ2, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Paired") +
  labs(x="emotion categories", y="Number of complaints",
       title = "Sentiment Analysis of Students' complaints\n(classification by
emotion)",
       plot.title = element_text(size=12))

#Pie Chart with size of polarity
ggplot(data = sent_categ2) +
  geom_bar(mapping = aes(x = polarity, fill = polarity), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_linedraw() +
  coord_polar()

#Pie Chart with size of emotion
ggplot(data = sent_categ2) +
  geom_bar(mapping = aes(x = emotion, fill = emotion), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_linedraw() +
  coord_polar()

#Pie Chart of polarity
ggplot(data = sent_categ2) +
  geom_bar(mapping = aes(x = factor(1), fill = polarity), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_void() +
  coord_polar(theta = "y")

#Pie Chart Same Size emotion
ggplot(data = sent_categ2) +
  geom_bar(mapping = aes(x = factor(1), fill = emotion), width = 1) +
  scale_fill_brewer(palette="Paired") +
  theme_void() +
  coord_polar(theta = "y")

#Word Cloud
library(tm)
library(wordcloud)
```

```
#Word cloud of the full dataset (non-ajusted)
wordcloud(Student_Complaint_Narratives_txt)

#Word cloud ajusted
wordcloud(Student_Complaint_Narratives_txt, scale=c(2,0.8), max.words=1000,
          random.order=FALSE, rot.per=0.35, colors=brewer.pal(8, "Dark2"))

#Comparing word cloud

#Compare words by emotion
# separating text by emotion
emos = levels(factor(sent_categ2$emotion))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = Student_Complaint_Narratives_txt[emotion == emos[i]]
  emo.docs[i] = paste(tmp, collapse="")
}

# remove stopwords
emo.docs = removeWords(emo.docs, stopwords("english"))
# create corpus
corpus = Corpus(VectorSource(emo.docs))
tdm = TermDocumentMatrix(
  corpus,
  control = list(
    wordLengths=c(0,Inf),
    removePunctuation = TRUE,
    stopwords = c("prayformh370", "prayformh", stopwords("english")),
    removeNumbers = TRUE, tolower = TRUE) )
tdm = as.matrix(tdm)
colnames(tdm) = emos

library(RColorBrewer)

# comparison word cloud by emotion
comparison.cloud(tdm, max.words=1000, colors = brewer.pal(nemo, "Dark2"),
                scale = c(3,.5), random.order = FALSE, title.size = 1.5)
```

```
#Compare words by polarity
#separating text by polarity
emos = levels(factor(sent_categ2$polarity))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = Student_Complaint_Narratives_txt[polarity == emos[i]]
  emo.docs[i] = paste(tmp, collapse="")
}

# remove stopwords
emo.docs = removeWords(emo.docs, stopwords("english"))
# create corpus
corpus = Corpus(VectorSource(emo.docs))
tdm = TermDocumentMatrix(
  corpus,
  control = list(
    wordLengths=c(0,Inf),
    removePunctuation = TRUE,
    stopwords = c("prayformh370", "prayformh", stopwords("english")),
    removeNumbers = TRUE, tolower = TRUE) )
tdm = as.matrix(tdm)
colnames(tdm) = emos

library(RColorBrewer)

# comparison word cloud by polarity
comparison.cloud(tdm, max.words=3000, colors = brewer.pal(nemo, "Dark2"),
  scale = c(3,.5), random.order = FALSE, title.size = 1.5)
```


6.5.3 Comparing the 2 datasets

```
#Frequency of each score in score in prepaid card score distribution

#Frequency of each score in Student Loan score distribution
count(Student_Complaint_Narratives.scores$score)

#Compare prepaid card complaints to student loan complaints datasets
all.scores = rbind(Consumer_Complaint_Narratives.scores,
Student_Complaint_Narratives.scores)
ggplot(data=all.scores) + # ggplot works on data.frames, always
  geom_histogram(mapping=aes(x=score, fill=Feedback), binwidth=0.5) +
  facet_grid(Feedback~.) + # make a separate plot for each hashtag
  theme_bw(base_size = 12) + scale_fill_brewer(palette = 18) # plain
display, nicer colors
```

6.6 Complaints' Score and complaints' Polarity & emotion (Tables)

6.6.1 Database 1: Complaints on Prepaid Cards

The tables below represent the first 10 rows of the prepaid card complaints

6.6.1.1 Scores

	score	text	Feedback
1	2	i loaded my card and i made an online purchase in XXXX, in XXXX of this year i loaded more funds on the card and made a few purchases, then when i checked my card it said i had XXXX dollar left available to me, upon calling them i was told that i had a pending transaction for XXXX dollars back in XXXX, and that when i used my card and loaded it they took the XXXX dollars, i do not understand how it is that u get a prepaid card so i can not spend what i do not have available to me. i have spoken to them and i was told that it was a glitch on there end, but then i am not responsible for there mistakes, i would like my eighty dollars back, furthermore i am thinking my card had money on it because i loaded 150 dollars on it and i only made XXXX transactions after loading the card with XXXX dollars, XXXX transactions in the amount of four dollars and twenty five cents, and XXXX for forty-one dollars and another for seventeen dollars then when i go out for dinner thinking i have enough fund to pay for my meal i find out i have no funds available to me, as i understand it a prepaid card can not allow things to go through if u do not have the available funds to pay for it, how can the card company allow a transaction to go through when u do not have the funds on the card, the purpose of having a prepaid card is so u do not over spend.	Consumer_Complaint_Narratives
2	2	I tried to get a refund on {\$4500.00} worth of cards. I called reloadit and they said that if I would send them a copy of the receipts I would receive my money in 20 days. This was on the XXXX of XXXX. Then they asked for a copy of the reloadit cards with the numbers on them and they said it was approved.=, but they wanted me to send them a copy of my drivers license and my utility bill. Than they said they sent the check out on XXXX XXXX. I still have n't received nothing so far and i told them if I did n't received my check by the XXXX XXXX I was going to call the FBI and the FBI said that I should contact you. I	Consumer_Complaint_Narratives
3	-2	My paypal account was permanently limited by Paypal on XXXX. I have contacted Paypal several times, but customer service department could n't explain to me why it was limited since it was done by another department. I have over {\$3000.00} in the account and could n't be transferred to my bank account due to limitation. Paypal said that it will be held for 180 days. I dont think that it is legal to hold the money for that long time without proper explanation.	Consumer_Complaint_Narratives
4	0	XXXX has frozen my safe with {\$100.00} in it and is not allowing me to access my funds. I have been trying for 2 months to access my funds, but the account has been frozen. I have been trying to contact customer service, but have been unable to reach them. They are holding my {\$100.00}, and I want to either load my card or have them mail it to me. My username on their site to access my safe is XXXXXXXXXXXX 'm attaching a picuture of what happens when I try to log in. There is {\$100.00} in the safe that is mine and they are holding.	Consumer_Complaint_Narratives
5	2	XXXX/XXXX/15 I was a victim of a IRS scam, some XXXX called me and I fall for it and purchased the prepaid cards XXXX it and them the numbers of the XXXX prepaid cards, an around two hours later I realized it was a scam so I called the number in the back of the cards, a CSR inform me the the money was on the cards and she (XXXX) was putting a stop on the cards. I was happy as thinking did n't loose my money and everything was fine just like she said, also she give me a ITS number of the claim and explain the it will take XXXX to XXXX business day to send me a check, I explain the situation and ask her if she was sure the sto payment was done because this was a time sensitive situation and she said yes. Week latter I called the company to seek the status of my claim and the inform me they only put a stop on XXXX cards and the XXXX of them were redeem, they only can return the money on those cards	Consumer_Complaint_Narratives

		and the rest was lost because the cards area like cash, I did understand that but when I called them the money was there and they supposedly put a stop, if they would it told that the stop would take time, I could redeem the cards my self and get my money back but the make me believe that my money was save and I believe them because after searched their name it seem like a legitimate company. The are regulated by the federal government, and license from XXXX XXXX XXXX XXXX XXXX XXXX FDIC. I feel the this company is as guilty as the people how called me, or more because they deliberated did n't inform me of the time a stop would take, on the other hand the make me believe the money was save and everything was fine that I should just wait for my money to come in the mail.	
6	8	I opened an account for a prepaid Visa card with First Progress in XX/XX/XXXX to help build up my credit score and everything was fine made all my payments on time and never had any issues until I decided to close this account out in XX/XX/XXXX for I was no longer using this account in I received my statement for the months of XX/XX/XXXX thru XX/XX/XXXX which showed they had charged me the annual fee of {\$39.00} when I called in to close the account in XX/XX/XXXX the representative for First Progress informed me that as long as there were no charges made to my account that I would receive the refund of the {\$39.00} annual fee along with the balance of my prepaid Visa Card which totaled {\$150.00} but there would be a 60 day waiting period to make sure that there were no charges made to my account finally I received my refund in XX/XX/XXXX but it was on for the amount of {\$150.00} and did not include the refund for the annual fee of {\$39.00} in which the representative for First Progress indicated I was entitled to receive so I contacted customer service with First Progress and was told I would have to write a letter to the First Progress Correspondence department to request my refund which I did on XX/XX/XXXX, I received a response dated XX/XX/XXXX indicating that I would have to contact customer service to resolve this issue so I contacted customer service and they said that I would have to contact the Progress Correspondence department what they are doing is giving me the run around and violating their own policies for which I have been told XXXX different times by a Customer service representative that I am entitled to the refund of the annual fee of {\$39.00} as long as there were no charges mad to my account in which I have the statements to show that no charges were made.	Consumer_Complaint_Narratives
7	-1	AMEX Prepaid card purchased by my company in XX/XX/XXXX and handed out to employees at a sales training event. Card has been in my possession since I received it. I did not attempt to the use card until XX/XX/XXXX. Merchant declined the card due to {\$0.00} balance. Called AMEX and found out the card was used at a gas station in XXXX XX/XX/XXXX. I live in XXXX, XXXX and have never been to XXXX. AMEX tells me they can not do anything. They tried to give me the phone # of the merchant in XXXX which was unavailable.	Consumer_Complaint_Narratives
8	-3	I have purchased {\$3000.00} in reload packs from the company, XXXX, a subsidiary of XXXX Holdings. I have been only able to load {\$1500.00} and this only after hundreds of attempts to use the online service. XXXX of attempts result in the same error message, 'We are experiencing technical difficulties '. I have made multiple calls to their customer service department and was told, after HOURS on hold, that a customer service supervisor would get back to me. That never happened. I need the money I have paid to use this companies service for a medical emergency. Please help.	Consumer_Complaint_Narratives
9	5	I deposited a check using their American Express Serve mobile app on XXXX XXXX, 2015 and it said that the check will clear in XXXX business days. When I logged into my account on the XXXX business day the money was not available, I called American Express Serve to let them know the money was not available and the transferred me to a XXXX party company called XXXX. XXXX told me that the check was approved and it was up to American Express Serve to make the funds available. I called back American Express Serve to tell them what XXXX said to me and I got the run around, rude and sarcastic answers from the rep. I do n't understand why my money is not available to me since the check was cleared. I am counting on that money to put gas in my car to go to work and buy groceries now I am without money today.	Consumer_Complaint_Narratives

10	0	<p>On XXXX XXXX, 2015 I checked my American Express Serve Card balance and had approximately {\$380.00} dollars available. This was to be used to pay a bill so on Monday, XXXX XXXX, 2015 I checked my balance again so that I could withdraw the money to pay my bill and my available balance was {\$9.00}. I immediately called American Express Serve and explained to them that for some reason I only had {\$9.00} available when I should have had {\$380.00}. I saw that there were XXXX transactions for {\$120.00}, {\$120.00} and a pending charge for {\$120.00}. They all were from XXXX. I requested a fraud dispute document. I explained on the document that I did not make the charges and did not authorize anyone else to make them on my behalf. However, a couple of days later I received an email from American Express Serve that they had done the following: 1. Closed my account. 2. Determined that the charges were valid. I again called American Express Serve and asked them how they came to that determination especially since I did not charge anything from XXXX, and did not authorize anyone to do so. Afterward I was told to read paragraph XXXX of the User Agreement. Upon doing so, it gave me no information regarding fraudulent transactions, or how I would get the money back that was taken from my account. Furthermore, the email that was sent to me said I could request what information led them to determine that there was no fraud. However, when I called Amex Serve they kept referring me to paragraph XXXX and no help regarding what they found or how I will get the money that was taken from my account.</p>	Consumer_Complaint_Narratives
----	---	---	-------------------------------

6.6.1.2 Polarity & emotion

	Text	polarity	emotion
1	<p>i loaded my card and i made an online purchase in in of this year i loaded more funds on the card and made a few purchases then when i checked my card it said i had dollar left available to me upon calling them i was told that i had a pending transaction for dollars back in and that when i used my card and loaded it they took the dollars i do not understand how it is that u get a prepaid card so i can not spend what i do not have available to me i have spoken to them and i was told that it was a glitch on there end but then i am not responsible for there mistakes i would like my eighty dollars back furthermore i am thinking my card had money on it because i loaded dollars on it and i only made transactions after loading the card with dollars transactions in the amount of four dollars and twenty five cents and for fortyone dollars and another for seventeen dollars then when i go out for dinner thinking i have enough fund to pay for my meal i find out i have no funds available to me as i understand it a prepaid card can not allow things to go through if u do not have the available funds to pay for it how can the card company allow a transaction to go through when u do not have the funds on the card the purpose of having a prepaid card is so u do not over spend</p>	neutral	disgust
2	<p>i tried to get a refund on worth of cards i called reloadit and they said that if i would send them a copy of the receipts i would receive my money in days this was on the of then they asked for a copy of the reloadit cards with the numbers on them and they said it was approved but they wanted me to send them a copy of my drivers license and my utility bill than they said they sent the check out on i still have nt received nothing so far and i told them if i did nt received my check by the i was going to call the fbi and the fbi said that i should contact you i</p>	negative	sadness
3	<p>my paypal account was permanently limited by paypal on i have contacted paypal several times but customer service department could nt explain to me why it was limited since it was done by another department i have overin the account and could nt be transferred to my bank account due to limitation paypal said that it will be held for days i dont think that it is</p>	negative	joy

	legal to hold the money for that long time without proper explanation		
4	has frozen my safe within it and is not allowing me to access my funds i have been trying formonths to access my funds but the account has been frozen i have been trying to contact customer service but have been unable to reach them they are holding myand i want to either load my card or have them mail it to me my username on their site to access my safe is m attaching a picuture of what happens when i try to log in there isin the safe that is mine and they are holding	negative	sadness
5	i was a victim of a irs scam some called me and i fall for it and purchased the prepaid cards it and them the numbers of the prepaid cards an around two hours later i realized it was a scam so i called the number in the back of the cards a csr inform me the the money was on the cards and shewas putting a stop on the cards i was happy as thinking did nt loose my money and everything was fine just like she said also she give me a its number of the claim and explain the it will take to business day to send me a check i explain the situation and ask her if she was sure the sto payment was done because this was a time sensitive situation and she said yes week latter i called the company to seek the status of my claim and the inform me they only put a stop on cards and the of them were redeem they only can return the money on those cards and the rest was lost because the cards area like cash i did understand that but when i called them the money was there and they supposedly put a stop if they would it told that the stop would take time i could redeem the cards my self and get my money back but the make me believe that my money was save and i believe them because after searched their name it seem like a legitimate company the are regulated by the federal government and license from fdic i feel the this company is as guilty as the people how called me or more because they deliberated did nt inform me of the time a stop would take on the other hand the make me believe the money was save and everything was fine that i should just wait for my money to come in the mail	neutral	sadness
6	i opened an account for a prepaid visa card with first progress in to help build up my credit score and everything was fine made all my payments on time and never had any issues until i decided to close this account out in for i was no longer using this account in i received my statement for the months of thru which showed they had charged me the annual fee ofwhen i called in to close the account in the representative for first progress informed me that as long as there were no charges made to my account that i would receive the refund of theannual fee along with the balance of my prepaid visa card which totaledbut there would be aday waiting period to make sure that there were no charges made to my account finally i received my refund in but it was on for the amount ofand did not include the refund for the annual fee ofin which the representative for first progress indicated i was entitled to receive so i contacted customer service with first progress and was told i would have to write a letter to the first progress correspondence department to request my refund which i did on i received a response dated indicating that i would have to contact customer service to resolve this issue so i contacted customer service and they said that i would have to contact the progress correspondence department what they are doing is giving me the run around and violating their own policies for which i have been told different times by a customer service representative that i am entitled to the refund of the annual fee ofas long as there were no charges mad to my account in which i have the statements to show that no charges were made	neutral	joy
7	amex prepaid card purchased by my company in and handed out to employees at a sales training event card has been in my possession since i received it i did not attempt to the use card until merchant declined the card due tobalance called amex and found out the card was used at a gas station in i live in and have never been to amex tells me they can not do anything they tried to give me the phoneof the merchant in which was unavailable	negative	joy

8	<p>i have purchased in reload packs from the company a subsidiary of holdings i have been only able to load and this only after hundreds of attempts to use the online service of attempts result in the same error message we are experiencing technical difficulties i have made multiple calls to their customer service department and was told after hours on hold that a customer service supervisor would get back to me that never happened i need the money i have paid to use this companies service for a medical emergency please help</p>	negative	joy
9	<p>i deposited a check using their american express serve mobile app on and it said that the check will clear in business days when i logged into my account on the business day the money was not available i called american express serve to let them know the money was not available and the transferred me to a party company called told me that the check was approved and it was up to american express serve to make the funds available i called back american express serve to tell them what said to me and i got the run around rude and sarcastic answers from the rep i do not understand why my money is not available to me since the check was cleared i am counting on that money to put gas in my car to go to work and buy groceries now i am without money today</p>	neutral	unknown
10	<p>on i checked my american express serve card balance and had approximately dollars available this was to be used to pay a bill so on monday i checked my balance again so that i could withdraw the money to pay my bill and my available balance was immediately called american express serve and explained to them that for some reason i only had available when i should have had i saw that there were transactions for and a pending charge for they all were from i requested a fraud dispute document i explained on the document that i did not make the charges and did not authorize anyone else to make them on my behalf however a couple of days later i received an email from american express serve that they had done the following closed my account determined that the charges were valid i again called american express serve and asked them how they came to that determination especially since i did not charge anything from and did not authorize anyone to do so afterward i was told to read paragraph of the user agreement upon doing so it gave me no information regarding fraudulent transactions or how i would get the money back that was taken from my account furthermore the email that was sent to me said i could request what information lead them to determine that there was no fraud however when i called amex serve they kept referring me to paragraph and no help regarding what they found or how i will get the money that was taken from my account</p>	neutral	joy

6.6.2 Database 2: Complaints on Student Loan

The tables below represent the first 10 rows of the student loan complaints

6.6.2.1 Scores

	score	text	Feedback
1	0	I received a forberance on my loans last spring. XXXX 2016. When i saw i was recently reported the last few mths for non payment i contacted the Fed Loan Servicing center and my info was not entered correctly on the fed loan side. The women i spoke with stated shefixed the error and suggested i file a dispute. They will correct the statemenr of multiple late payments.	Student_Complaint_Narratives
2	-3	Was not contacted 4 years later about some private loan I supposedly took out and they are trying to collect it. I keep trying to call XXXX college but they keep giving me bogus numbers where nobody answers. I called the collection agency to get some information and they will not give me any info at all. They said call XXXX college but again this gets me nowhere with bogus numbers where noone answers or hangs up on me.	Student_Complaint_Narratives
3	-1	This is a continuation of a previous issue with Citibank that has reached another XXXX. Citibank increased my monthly payments XXXX and I have been unable to reach an agreement with them about my monthly payment. I am requesting that Citibank reconsider my request for graduated repayment, so that I am not forced into Collections. I would like to pursue a payment plan that works with my current financial situation and allows me be responsible for my debt.	Student_Complaint_Narratives
4	1	In 2008 I attended XXXX University in XXXX off XXXX and XXXX while attending XXXX University I was only known of the XXXX XXXX loans I had. While going to the school they kept telling me I did not have enough money to continue that I had to keep taking out loans just to continue. Right before I graduated they told me I did n't have enough money to XXXX that I had to take out another loan before I can XXXX an at that time is when I found out I had a XXXX lending loan. Then after XXXX i ask for my XXXX to find out they took more then XXXX loan out on the same day at the same time. I also find out that they XXXX about how much money my XXXX will cost it was suppose to cost about XXXX an when i XXXX it was almost XXXX. I was also suppose to XXXX after XXXX years and i did n't XXXX till about XXXX years. Then after XXXX and continuing XXXX i was told i can continue with my XXXX even if i had a XXXX record and it withhold of adjudication. well I just found out that i wasted my time at XXXX and XXXX college In XXXX to get my XXXX and almost my XXXX and XXXX wont even let me withdraw when i told them my situation. Even though they said i am not even suppose to be certified but they still wont withdraw me. I told genesis lending of my situation but they wont drop the loan. I do n't feel I need to pay any loans because it was more then what i was told and I was lied to just so they can get money I wasted my time when I could of went to XXXX for something else.	Student_Complaint_Narratives
5	-3	I have a private loan with Sallie Mae. I called them to try to get them to lower my payment or interest rate due to my income. I have XXXX children and only made a small on my taxes last year I need help with my payment or it will ruin my credit. They REFUSED to help me at all. I am VERY unhappy.	Student_Complaint_Narratives
6	-4	My husband lost his job in XX/XX/XXXX and was unemployed for almost XXXX months. He has recently started a new job with XXXX and we are getting our finances back on track. However, by the end of XXXX, I struggled to keep up with the payments and my student loans became dangerously past due. On or around XX/XX/XXXX, I arranged a {\$250.00} student loan payment on line. I did not realize until it was too late that I had forgotten to change my banking information and the payment was rejected by my bank. (I have since cleared this up.) I called AES to discuss further options. On XX/XX/XXXX, I was told by a representative that I qualified for XXXX-day forbearance. When we discussed the unfortunate payment error, she said, " do n't worry about it ; the forbearance period	Student_Complaint_Narratives

		<p>would cover the non-payment. On XX/XX/XXXX, my credit bureau monitoring service notified me that my credit score had dropped by over XXXX points due to a XXXX-day late payment reported by AES. This was AFTER I was assured that I qualified for the forbearance. When I called AES to inquire about this, I was told that credit reporting operated on a separate system and that credit reporting was " automatic, " but that he would note the situation in my file. All of the representatives have been very pleasant and helpful. However, as I explained to XXXX of the representatives, I am a consumer. When a representative tells me " not to worry about a payment issue, " I listen. My husband and I are XXXX veterans and have to commute 2 hours to work every day. We are responsible and have not missed a student loan payment in several years. We need to buy a home closer to our work. The problem is this. The loans with AES are a consolidated debt of XXXX different loans. When the loans were reported to credit bureau as XXXX days late, my credit scores dropped almost XXXX points for EACH LOAN. That 's approximately XXXX points!!! This is neither fair nor accurate. When payments are made, amounts are applied to each individual loan. The way this is set up, even if XXXX of the XXXX loans is late, ALL XXXX report late. No credit report can sustain and instant XXXX point drop in credit score over something this confusing. Like I stated earlier, I am a veteran and a responsible citizen. I struggled to pay the loans on time for the entire XXXX months my husband was out of work. It is not fair that I should have to suffer the insult of a XXXX day late payment due to misinformation that I was given. If I am late making payments due to neglect, then I would feel like I have earned this discredit. However, I was acting on information supplied to me by a company representative. I followed their rules. Therefore, I respectfully ask that the XXXX day late be removed from all XXXX credit bureaus so that we can move on with our lives.</p>	
<p>7</p>	<p>-12</p>	<p>Mohela called me on XXXX/XXXX/15, I told the rep I was at work he then advised me it was an urgent matter. I was concerned so I asked my team lead to give me about XXXX minutes because I had an urgent phone call. I then verified my information and he then informed me I was past due on my loan (only XXXX days) but the payment was pending in their system because I always make my payment through their system. I had to tell that rep the payment was already made and that he should have looked at the payment screen before he asked me to make a payment. This incident has happened before when a payment was made but not yet posted. I was only XXXX days late, I started a new job and my paydays are different from my previous employer, I have never been XXXX days late with them and this matter was not an urgent matter because the payment was already and I felt like they wasted my time, the co-signer and was upset because she states no one told her a payment was pending in the system. It seems to me that Mohela wants to misled and employ scare tactics, the payment was made for the XXXX and the money came out of my account on the XXXX (it was showing as pending in my checking acct on the XXXX). Mohela has did this to me before, I had to tell the rep to look into the system because I made a payment, I also heard a call my mom recorded of Mohela and not one did he tell her a payment was pending in the system, he asked her to make a payment. She was upset with me because they implied she needed to make a payment when I already her one was made but not yet posted, nowhere on that call was that information given. I actually looked online on mohela 's online page and the payment was posted on the day I said it would. I also spoke with a consumer rights attorney and he advised me that Mohela should make sure they are looking to see if a payment is scheduled online before they ask someone to make another payment and the rep misrepresented the call by saying it was an urgent matter ; it was not an important issue because the payment was made and it was only XXXX days past due. I was nowhere near the XXXX day mark so I am not sure they discussed anything about credit reporting for this payment because it sounded stupid to say such a thing. A supervisor states they have to call when a payment is late (this is their job) but make sure you are</p>	<p>Student_Complaint_Narratives</p>

		<p>actually looking to see if a payment was made before you stay anything stupid like asking to make another payment. I was also advised by an attorney that Mohela should not be telling people it is an urgent matter because it was not in this case and nothing about a past due bill is an emergency. It was utterly deceptive and this is not the first time this has happened, I would prefer to get manual calls not computer calls on my cell phone but they told me I had to endure these stupid calls. The rep asked me if I had another phone call number to use, I said no then I was told I would have to endure these useless calls especially since a payment was already made but not posted, it was pending in the system and I wish people at Mohela would take the time to look at the online screen because they wasted my time because I was at work and I had to stop what I was doing to deal with this matter.</p>	
8	2	<p>I ca n't pay for my private student loans. They have ballooned to more than double what I borrowed. I feel like I have a " payday " student loan. They want me to pay over {\$220000.00}, who can afford to pay that? They were marketed to me as a way to pay for the things you want or need while in college. To buy books, a nice apartment or even a new car. I got emails frequently saying it 's time to sign up again for your Chase Education One loan. I thought the Chase loans were different than the XXXX XXXX loans because XXXX XXXX mentions additional fees in their advertisements were as Chase did not. Well to make a long story short, they were sold to a different bank while in forbearance. I received no notification that they were sold and no notification that the sale took them out of forbearance. I also received no notification they went into default, I only learned they were in default by a debt collector call months after they were sold and place into default. I thought I was being scammed because I had paperwork saying they were in forbearance. I 've have searched for advice the last 6 years on how to settle and handle the loans only to be misinformed, sold some unneeded service debt service and flat out told you have no hope to get the loans out of default. Now that information is becoming more available I have learned in order to get them out of default I have find find a competed private student loan attorney, pay \$ 2-4000 in lawyer fees and another \$ 4-8000 as a good faith deposit just to get them out of default. Likely many, I could n't come up with all those fees and now I am being summons to court. How can I while making \$ XXXXk a year afford to get caught up on these loans, let alone pay off the loans. These loans have left me helpless and hopeless!</p>	Student_Complaint_Narratives
9	-2	<p>My XXXX payment is almost half of my monthly income. I have contacted Aes several times about lowering my payment but every time they refuse to help. They say I am not eligible for any of the other options. I am a XXXX and will be forced to default if American Education Services refuses to help with a lower payment.</p>	Student_Complaint_Narratives
10	-6	<p>Discover Student Loans, the current servicer of my private student loan recently sent me a letter indicating that during a " recent review, " they discovered that unpaid interest from a previous deferment period was not added to my principal loan balance. Because of this, my minimum monthly payment was lower than what it should have been. To resolve this, they added the unpaid interest to my principal balance ; however, the exact amount added is not disclosed in the letter. Along with this, they have indicated that my monthly payment will also increase, in which that amount is also not disclosed. They have indicated that the adjustments will be reflected on my monthly statement within the next 60 days, but no explanation on the amount added back in and how they arrived at the dollar figure. The letter also indicated that a credit of {\$210.00} has been applied to the interest balance of the loan, which is not known how they arrived at that exact figure either. The letter sent is extremely vague and does not discuss how they arrived at the decision or calculation used to add back in the amount to the principal balance or what the new monthly payment would be. The letter also states that this was done to ensure the loan is paid in full at the end of its repayment period. Additionally, the letter states that any adjustments made will not increase the total cost of the loan ;</p>	Student_Complaint_Narratives

	<p>however, my balance increased along with my monthly payment. Upon logging in to my account, it was discovered that Discover Student Loans added back in over {\$2500.00} to my principal balance. To find an answer to my questions, I contacted Discover Student Loans where I was unable to obtain any answers, but they did mention that my payment has increased by approximately {\$20.00} per month. Out of the XXXX representatives I have spoken to in the last two days, I have not been given an answer as to how they arrived at the calculation used to determine the exact amount added back to my principal balance and they have failed to follow up with me regarding my concerns. I have now been in repayment for close to five years, with 100 % on-time payments. I feel that I am being mislead and I am having to pay for a mistake that I did not make. As you will see in the recent statement attached, I have paid {\$20000.00} through XXXX XX/XX/2015 and of that amount, {\$20000.00} has gone to interest alone. It also indicates {\$0.00} in late fees to date. School : XXXX XXXX ; XXXX, OH</p>	
--	--	--

6.6.2.2 Polarity & emotion

	Text	polarity	emotion
1	<p>i received a forberance on my loans last spring xxxwhen i saw i was recently reported the last few mths for non payment i contacted the fed loan servicing center and my info was not entered correctly on the fed loan side the women i spoke with stated shefixed the error and suggested i file a dispute they will correct the statemenr of multiple late payments</p>	neutral	disgust
2	<p>was not contactedyears later about some private loan i supposedly took out and they are trying to collect it i keep trying to call xxxx college but they keep giving me bogus numbers where nobody answers i called the collection agency to get some information and they will not give me any info at all they said call xxxx college but again this gets me nowhere with bogus numbers where noone answers or hangs up on me</p>	negative	sadness
3	<p>this is a continuation of a previous issue with citibank that has reached another xxxx citibank increased my monthly payments xxxx and i have been unable to reach an agreement with them about my monthly payment i am requesting that citibank reconsider my request for graduated repayment so that i am not forced into collections i would like to pursue a payment plan that works with my current financial situation and allows me be responsible for my debt</p>	positive	joy
4	<p>ini attended xxxx university in xxxx off xxxx and xxxx while attending xxxx university i was only known of the xxxx xxxx loans i had while going to the school they kept telling me i did not have enough money to continue that i had to keep taking out loans just to continue right before i graduated they told me i did nt have enough money to xxxx that i had to take out another loan before i can xxxx an at that time is when i found out i had a xxxx lending loan then after xxxx i ask for my xxxx to find out they took more then xxxx loan out on the same day at the same time i also find out that they xxxx about how much money my xxxx will cost it was suppose to cost about xxxx an when i xxxx it was almost xxxx i was also suppose to xxxx after xxxx years and i did nt xxxx till about xxxx years then after xxxx and continuing xxxx i was told i can continue with my xxxx even if i had a xxxx record and it withhold of adjudication well i just found out that i wasted my time at xxxx and xxxx college in xxxx to get my xxxx and almost my xxxx and xxxx wont even let me withdraw when i told them my situation even though they said i am not even suppose to be certified but they still wont withdraw me i told genesis lending of my situation but they wont drop the loan i do nt feel i need to pay any loans because it was more then what i was told and i was lied to just so they can get money i</p>	negative	sadness

	wasted my time when i could of went to xxxx for something else		
5	i have a private loan with sallie mae i called them to try to get them to lower my payment or interest rate due to my income i have xxxx children and only made a small on my taxes last year i need help with my payment or it will ruin my credit they refused to help me at all i am very unhappy	negative	sadness
6	my husband lost his job in xxxxxxxx and was unemployed for almost xxxx months he has recently started a new job with xxxx and we are getting our finances back on track however by the end of xxxx i struggled to keep up with the payments and my student loans became dangerously past due on or around xxxxxxxx i arranged astudent loan payment on line i did not realize until it was too late that i had forgotten to change my banking information and the payment was rejected by my banki have since cleared this upi called aes to discuss further options on xxxxxxxx i was told by a representative that i qualified for xxxxday forbearance when we discussed the unfortunate payment error she saiddo nt worry about itthe forbearance period would cover the nonpayment on xxxxxxxx my credit bureau monitoring service notified me that my credit score had dropped by over xxxx points due to a xxxxday late payment reported by aes this was after i was assured that i qualified for the forbearance when i called aes to inquire about this i was told that credit reporting operated on a separate system and that credit reporting wasautomaticbut that he would note the situation in my file all of the representatives have been very pleasant and helpful however as i explained to xxxx of the representatives i am a consumer when a representative tells menot to worry about a payment issuei listen my husband and i are xxxx veterans and have to commutehours to work every day we are responsible and have not missed a student loan payment in several years we need to buy a home closer to our work the problem is this the loans with aes are a consolidated debt of xxxx different loans when the loans were reported to credit bureau as xxxx days late my credit scores dropped almost xxxx points for each loan that s approximately xxxx points this is neither fair nor accurate when payments are made amounts are applied to each individual loan the way this is set up even if xxxx of the xxxx loans is late all xxxx report late no credit report can sustain and instant xxxx point drop in credit score over something this confusing like i stated earlier i am a veteran and a responsible citizen i struggled to pay the loans on time for the entire xxxx months my husband was out of work it is not fair that i should have to suffer the insult of a xxxx day late payment due to misinformation that i was given if i am late making payments due to neglect then i would feel like i have earned this discredit however i was acting on information supplied to me by a company representative i followed their rules therefore i respectfully ask that the xxxx day late be removed from all xxxx credit bureaus so that we can move on with our lives	negative	joy
7	mohela called me on xxxxxxxx i told the rep i was at work he then advised me it was an urgent matter i was concerned so i asked my team lead to give me about xxxx minutes because i had an urgent phone call i then verified my information and he then informed me i was past due on my loanonly xxxx daysbut the payment was pending in their system because i always make my payment through their system i had to tell that rep the payment was already made and that he should have looked at the payment screen before he asked me to make a payment this incident has happened before when a payment was made but not yet posted i was only xxxx days late i started a new job and my paydays are different from my previous employer i have never been xxxx days late with them and this matter was not an urgent matter because the payment was already and i felt like they wasted my time the cosigner and was upset because she states no one told her a payment was pending in the system it seems to me that mohela wants to misled and employ scare tactics the payment was made for the xxxx and the money came out of my account on the xxxxit was showing as pending in my checking acct on the xxxxmohela has did this to me before i had to tell the rep to look into the system	negative	joy

	<p>because i made a payment i also heard a call my mom recorded of mohela and not one did he tell her a payment was pending in the system he asked her to make a payment she was upset with me because they implied she needed to make a payment when i already her one was made but not yet posted nowhere on that call was that information given i actually looked online on mohela s online page and the payment was posted on the day i said it would i also spoke with a consumer rights attorney and he advised me that mohela should make sure they are looking to see if a payment is scheduled online before they ask someone to make another payment and the rep misrepresented the call by saying it was an urgent matter it was not an important issue because the payment was made and it was only xxxx days past due i was nowhere near the xxxx day mark so i am not sure they discussed anything about credit reporting for this payment because it sounded stupid to say such a thing a supervisor states they have to call when a payment is late this is their job but make sure you are actually looking to see if a payment was made before you say anything stupid like asking to make another payment i was also advised by an attorney that mohela should not be telling people it is an urgent matter because it was not in this case and nothing about a past due bill is an emergency it was utterly deceptive and this is not the first time this has happened i would prefer to get manual calls not computer calls on my cell phone but they told me i had to endure these stupid calls the rep asked me if i had another phone call number to use i said no then i was told i would have to endure these useless calls especially since a payment was already made but not posted it was pending in the system and i wish people at mohela would take the time to look at the online screen because they wasted my time because i was at work and i had to stop what i was doing to deal with this matter</p>		
<p>8</p>	<p>i ca nt pay for my private student loans they have ballooned to more than double what i borrowed i feel like i have apayday student loan they want me to pay over who can afford to pay that they were marketed to me as a way to pay for the things you want or need while in college to buy books a nice apartment or even a new car i got emails frequently saying it s time to sign up again for your chase education one loan i thought the chase loans were different than the xxxx xxxx loans because xxxx xxxx mentions additional fees in their advertisements were as chase did not well to make a long story short they were sold to a different bank while in forbearance i received no notification that they were sold and no notification that the sale took them out of forbearance i also received no notification they went into default i only learned they were in default by a debt collector call months after they were sold and placed into default i thought i was being scammed because i had paperwork saying they were in forbearance i ve have searched for advice the last years on how to settle and handle the loans only to be misinformed sold some unneeded service debt service and flat out told you have no hope to get the loans out of default now that information is becoming more available i have learned in order to get them out of default i have find find a competed private student loan attorney pay in lawyer fees and another as a good faith deposit just to get them out of default likely many i could nt come up with all those fees and now i am being summons to court how can i while making xxxxx a year afford to get caught up on these loans let alone pay off the loans these loans have left me helpless and hopeless</p>	<p>neutral</p>	<p>joy</p>
<p>9</p>	<p>my xxxx payment is almost half of my monthly income i have contacted aes several times about lowering my payment but every time they refuse to help they say i am not eligible for any of the other options i am a xxxx and will be forced to default if american education services refuses to help with a lower payment</p>	<p>positive</p>	<p>unknown</p>
<p>10</p>	<p>discover student loans the current servicer of my private student loan recently sent me a letter indicating that during a recent review they discovered that unpaid interest from a previous deferment period was not added to my principal loan balance because of this my minimum monthly payment was lower than what it should have been to resolve this they added</p>	<p>negative</p>	<p>joy</p>

the unpaid interest to my principal balance however the exact amount added is not disclosed in the letter along with this they have indicated that my monthly payment will also increase in which that amount is also not disclosed they have indicated that the adjustments will be reflected on my monthly statement within the next days but no explanation on the amount added back in and how they arrived at the dollar figure the letter also indicated that a credit of has been applied to the interest balance of the loan which is not known how they arrived at that exact figure either the letter sent is extremely vague and does not discuss how they arrived at the decision or calculation used to add back in the amount to the principal balance or what the new monthly payment would be the letter also states that this was done to ensure the loan is paid in full at the end of its repayment period additionally the letter states that any adjustments made will not increase the total cost of the loan however my balance increased along with my monthly payment upon logging in to my account it was discovered that discover student loans added back in over to my principal balance to find an answer to my questions i contacted discover student loans where i was unable to obtain any answers but they did mention that my payment has increased by approximately per month out of the xxxx representatives i have spoken to in the last two days i have not been given an answer as to how they arrived at the calculation used to determine the exact amount added back to my principal balance and they have failed to follow up with me regarding my concerns i have now been in repayment for close to five years without time payments i feel that i am being misled and i am having to pay for a mistake that i did not make as you will see in the recent statement attached i have paid through xxxx xxxx and of that amount has gone to interest alone it also indicates in late fees to date school xxxx xxxxxxxx oh