

What Members Really Feel About Your Credit Card

Anirban Chakraborty, Oklahoma State University; Surya Bhaskar, Oklahoma State University;
Dr. Goutam Chakraborty, Oklahoma State University

ABSTRACT

Immediately after a new credit card product is launched and in the wallets of cardholders, sentiment begins to build. Positive and negative experiences of current customers posted online generate impressions among prospective cardholders in the form of technological word of mouth. Credit card issuing companies can use sentiment analysis to understand how its product is being received by the consumers and by taking suitable measure can propel the card's market success. With the help of text mining and sentiment analysis using SAS Enterprise Miner and SAS Sentiment Analysis Studio we are trying to answer: which aspects of a credit card garnered the most favor, and conversely, which generated negative impressions among the consumers.

Credit Karma is a free credit and financial management platform for US consumers available on the web and major mobile platforms, it provides free weekly updated credit scores and credit reports from national credit bureaus TransUnion and Equifax.

The implications of this project are as follows:

- All credit card issuing companies can use this technique to know how their product is fairing in the market and based on public opinion can make business decisions to improve the flaws.
- Sentiment analysis can simulate word of mouth of millions of existing users about a credit card and know about the salient features.

INTRODUCTION

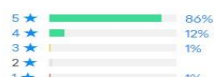
Credit Karma is a free credit and financial management platform for US consumers available on the web and major mobile platforms. Founded in 2007, it provides free weekly updated credit scores and credit reports from national credit bureaus TransUnion and Equifax, alongside daily credit monitoring from TransUnion.

Credit Karma also provides credit tools, such as a Credit Score Simulator, which simulates the effect of potential financial actions on a user's credit score; and tailored financial recommendations based on each individual user's credit profile.

In addition to its free credit reports and tools, Credit Karma offers a My Spending Tool through account aggregation service Yodlee, which allows users to track their credit card, loan transactions and balances in Credit Karma's interface. Credit Karma also hosts user forums for financial product reviews and credit advice, and provides calculator tools for debt repayment, amortization, home affordability and simple loans. In our project we have leveraged the use of the customer reviews for all the credit cards listed in the website in order to get a sense of the products amongst the customers. This would help the existing and prospective users to get the facts and interpret the opinions. Opinions are so important that whenever we need to make a decision, we want to hear others' opinions. This would further help the participating organizations to understand how their product is fairing in the market, the pros and cons, the customer segment for targeted marketing etc.

Chase Sapphire Preferred® Card Cardholders' Choice¹

4.8 out of 5 stars

★★★★★ 146 Reviews
FROM CARDHOLDERS IN THE LAST YEAR

Apply Now

Add Review

Approval Odds²

Sign up or log in to see your Approval Odds

☐ Compare This Card

ADDITIONAL RATINGS YOU MIGHT FIND HELPFUL

APPLICATION PROCESS
4.5 out of 5 starsCUSTOMER SERVICE
4.5 out of 5 stars

Card Details

Balance Transfer Intro APR

Not Applicable

Balance Transfer Regular APR

16.49% - 23.49% Variable

Annual Fee \$0 intro, \$95 after first year

Purchase Intro APR

Not Applicable

Purchase Regular APR

16.49% - 23.49% Variable

[See Details, Rates, & Fees](#)

Our Take

The Chase Sapphire Preferred® Card is a travel rewards card aimed at excellent credit consumers.

What to Like

Cardholders get **2 points for every dollar spent on travel and restaurants**, and 1 point for every other dollar spent on purchases. The Chase Ultimate Rewards program offers a **20% discount on travel purchases made using points**. The card also has **no foreign transaction fee** and offers a **50,000 point bonus** if you spend \$4,000 in the first 3 months.

Look Out For

After the first year, there's a **\$95 annual fee**. There's **no intro APR for purchases or balance transfers**.BMACK750 CARDHOLDER

I currently have the Chase Freedom, Chase Freedom Unlimited and Chase Slate cards with credit lines ranging from \$3200-\$5500. I opened a Chase checking and savings last year when I received my 1st chase credit card. Since then I've opened up 2 highschool checking accounts for my kids. My Fico scores are 696 Equifax, 681 TransUnion & 682 Experian. My Credit Karma scores are 744 Equifax & 738 TransUnion. I was notified by Chase that they had selected me for the Sapphire card offer a few months ago. [more](#)

Was this review helpful?



★★★★★ Mar 02, 2017

"Went into bank,"

Anonymous CARDHOLDER

Went into the bank. 9k later thanks Chase

Was this review helpful?



★★★★★ Feb 28, 2017

"Just approved"

themdash1 CARDHOLDER

I was just approved for this card @ \$5k with an approx credit score of 670. Approval odds = FAIR. I received a PENDING APPROVAL and waited 4 days then decided to call their reconsideration dept. After being transferred to verification dept they approved me.

Fig. 1. Sample Reviews on the credit karma website

DESCRIPTION & METADATA

The reviews data has 21 different credit cards with around 5000 user reviews. Embedded is the data dictionary:



VARIABLE	LEVEL	DESCRIPTION
ID	ID	This field represents the unique review number
CARD NAME	Text	This variable describes the name of the credit card
USER REVIEW	Text	This variable represents the actual credit card review posted by the user
USER RATING	Integer	This field has the user rating

Table 1: Data Dictionary

METHODOLOGY

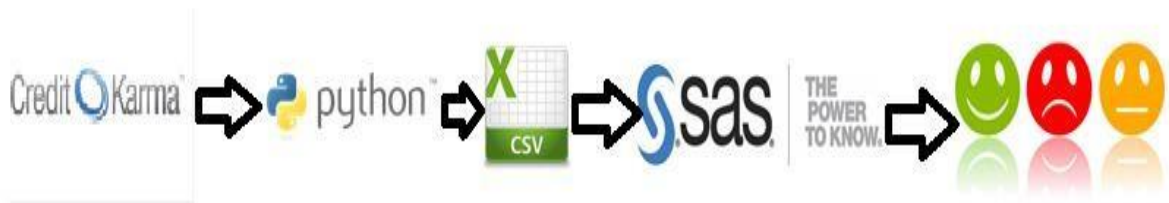


Fig. 2. Diagrammatic approach of the project

- The website www.creditkarma.com was used as the data source. 72 distinct credit cards having 10500 user reviews were identified.
- Using Python code, the web-scraping was done in order to collect these user reviews.
- The scraped user comments were collected in the csv format where the data cleansing part was done.
- Finally, the cleaned data (user reviews) were imported to SAS where different text mining techniques like text parsing, text filtering, text profiling, Rule builder etc. were used in order to perform sentiment analysis.

CODES

Python Code to scrape the user reviews from the website

```

"""
Spyder Editor
This is a temporary script file.
"""

```

```
from lxml import html
```

```

import requests
import math
from bs4 import BeautifulSoup
inF=open("D:\\2FALL2016\\SASSYMPORIUM\\input.txt", "r")
outStar = open("D:\\2FALL2016\\SASSYMPORIUM\\myOutFileStar.txt", "w")

for line in inF:
    print line
    nameNo = line.split(',')
    Pages=int(math.ceil(int(nameNo[1])/10))+1
    print nameNo
    print str(Pages)
    for pageNo in range(1,Pages+1):
        #print nameNo[0]
        finalResult=[]
        finalResult.append([])
        finalResult.append([])
        finalResult.append([])

        page = requests.get('https://www.creditkarma.com/reviews/credit-
card/single/id/'+nameNo[0]+'?pg='+str(pageNo)+'&pgsz=10')
        pageContent=page.content.replace('<p>','').replace('</p>','')
        pageContent=pageContent.replace('ck-stars','ckstars')
        tree = html.fromstring(pageContent)
        #This will create a list of buyers:
        buyers = tree.xpath('//div[@class="readmoreInner"]/text()')
        #This will create a list of prices
        soup = BeautifulSoup(pageContent)
        prices=soup.find_all(ckstars=["5 blue","4 blue","3 blue","2 blue","1 blue"])
        #print 'Buyers: ', buyers
        #print 'Prices: ', prices
        i=0
        for line in buyers:
            i=i+1
            if i> 2:
                Review=str(BeautifulSoup(line,'xml'))
                finalResult[0].append(nameNo[0])
                finalResult[2].append(Review.replace('<html><body><p>','').replace('</p></body></html>',''))
        i=0
        for line in prices:
            i=i+1
            if i> 2:
                finalResult[1].append(str(line).replace('<span ckstars=""','').replace('blue"></span>',''))

        for res in range(0,min(10,i-4)):
            print res
            print >>outStar,str((pageNo-1)*10+res)+'|'+finalResult[0][res]+'|'+finalResult[1][res]+'|'+finalResult[2][res]
            fileName=finalResult[0][res]+str((pageNo-1)*10+res)+".txt"
            if finalResult[1][res]>'3':
                fileName="PVE\\"+fileName
            else:
                if finalResult[1][res]<'3':
                    fileName="NVE\\"+fileName
                else:
                    fileName="NUT\\"+fileName
            outRev = open("D:\\2FALL2016\\SASSYMPORIUM\\"+fileName, "w")
            print >>outRev,finalResult[2][res]
            outRev.close()
print 'Program Quitting Normally'
outStar.close()

```

SAS Code to separate the positive and negative reviews

```
LIBNAME MYLIB 'D:\2FALL2016\SASSYMPORIUM';

DATA MYLIB.CARD_POS MYLIB.CARD_NEG;
  SET MYLIB.REVIEWS;
  NewId=card || PUT(ID,4.) ;
  IF Rating >3 then output MYLIB.CARD_POS;
  IF Rating <3 then output MYLIB.CARD_NEG;
  DROP id CARD RATING F5 F6 F7 F8 F9 F10 F11;

RUN;

DATA MYLIB.REVIEWSRULE;
  SET MYLIB.REVIEWS;
  IF Rating >3 then Target='P';
  IF Rating <3 then Target='N';
  IF Rating <> 3 then output;
  DROP id CARD RATING F5 F6 F7 F8 F9 F10 F11;

RUN;
```

DATA PREPARATION & EXPLANATORY ANALYSIS

We wrote a Python script to scrape the website in order to download the credit card user reviews. In the Python code we gave a txt file as an input which consisted of the name of the credit card (as mapped in the website) and the number of reviews (to ensure required number of iterations run to retrieve all the reviews). Once the process completed all the reviews were collected for every credit card along with the rating received by the users. There are 72 different credit cards in the website but for our analysis we considered only those credit cards which had more than at least 50 reviews, thus we got 21 distinct credit cards. The downloaded datasets were in the form of CSV file, which were converted into SAS Dataset.

Once we had the SAS data set with all the review we separated these reviews in to negative and positive based on the user ratings, here our assumption was to consider reviews with corresponding user rating 4 or more as a positive review and those reviews with user rating 2 or less as a negative review.

Credit Cards with Most User Reviews

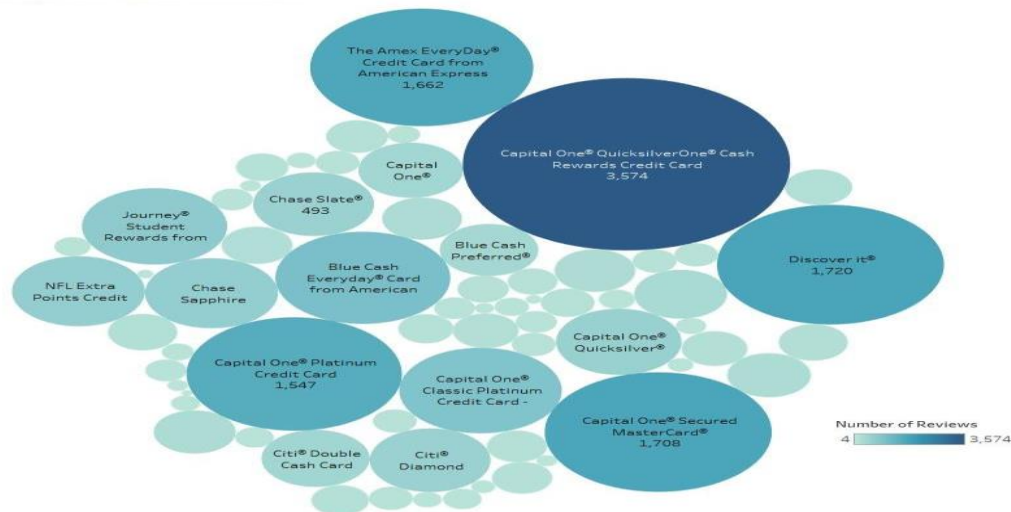


Fig. 3. Credit Cards with Most User Reviews

Some of the most talked credit cards are Capital One cash rewards card, Capital One Master card, Discover IT, AMEX Everyday card and Capital One Platinum card.

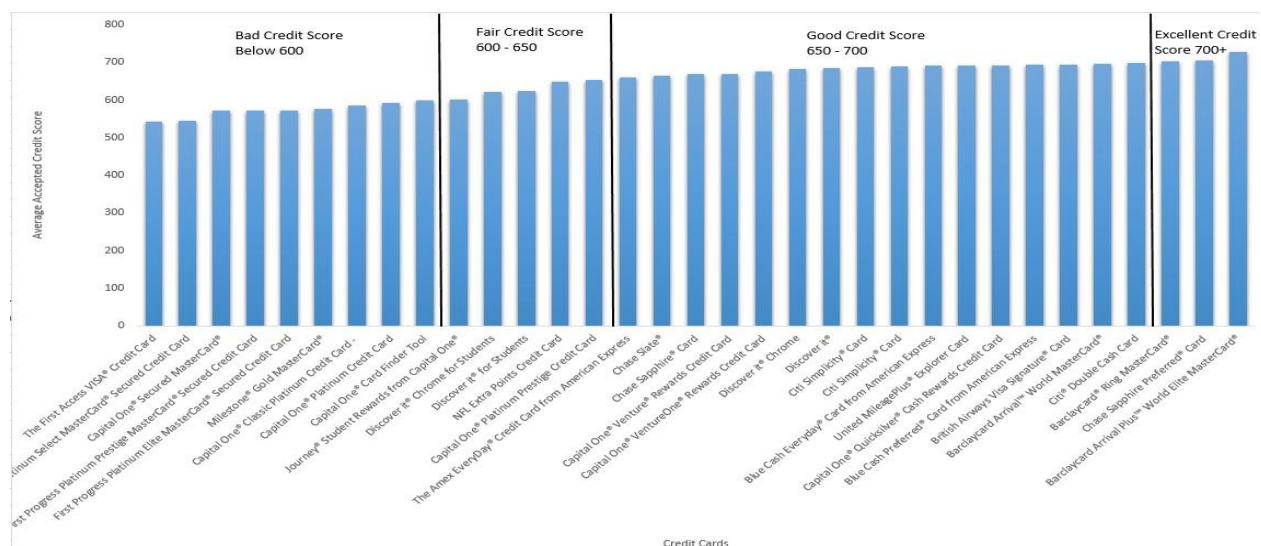


Fig. 4. Credit Cards with Average Accepted Credit Score

The above image shows the various cards as per their average accepted credit. We can see that there are a lot of options in different categories of credit score.

Different credit cards with their average household income

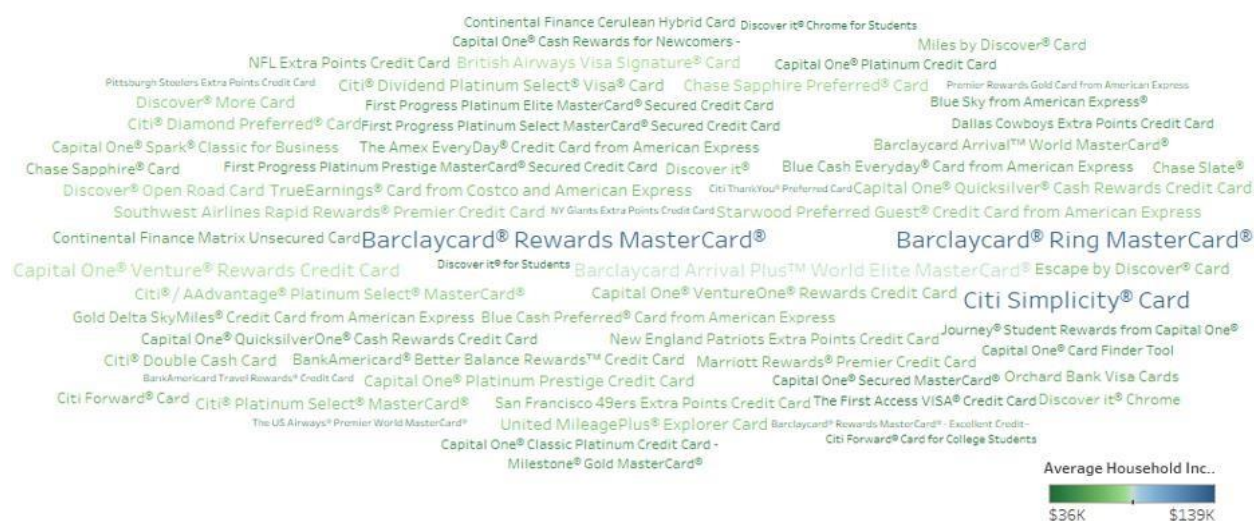


Fig. 5. Distribution of Credit Cards by Average Household Income

From the above image we can see that people with high household income mostly own Barclaycard Rewards MasterCard, Barclaycard Ring MasterCard, Citi Simplicity Card etc.

Credit Cards with their Average Credit Limit

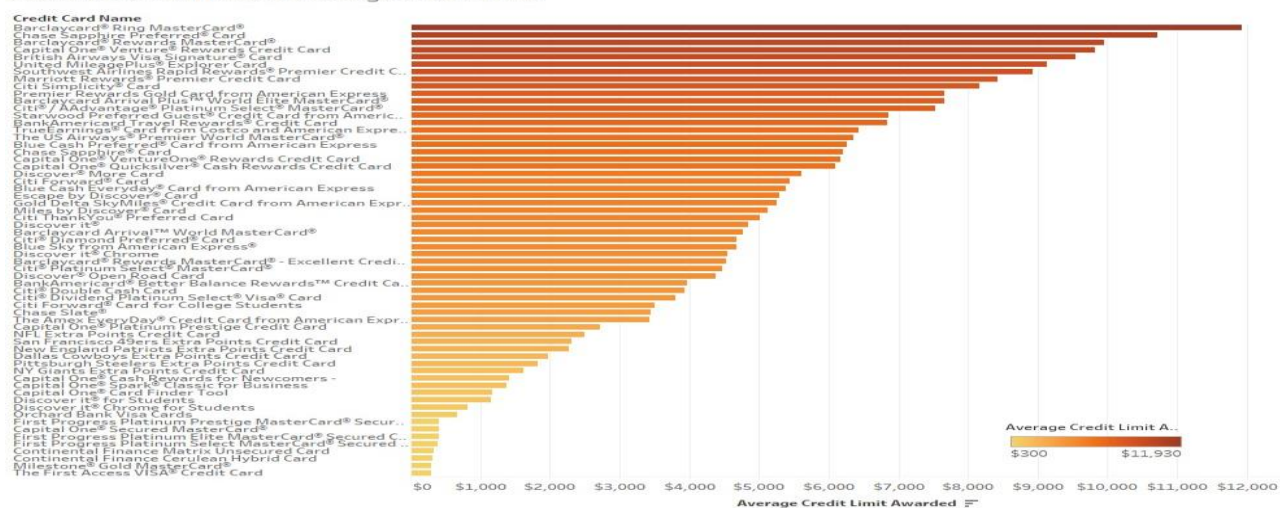


Fig. 5. Distribution of Credit Cards by Average Credit Limit

Some of the credit cards with highest credit limits are Barclaycard Ring MasterCard, Chase Sapphire with limit above \$10000 whereas the cards with least credit limit are The First Access Visa Card, cards by Continental Finance, First Progress, Capital One Secured Master Card etc.

TEXT ANALYSIS

We performed the project as 3 steps.

1. Analysis of all the reviews to understand what users are talking about.
2. Separating the reviews as positive (User ratings 4 & 5) and negative (User ratings 1 & 2) and then analyzing the reviews to get a better picture of the pros and cons of the credit cards.
3. Sentiment Analysis of individual major credit card companies.

1. ANALYSIS OF ALL THE REVIEWS

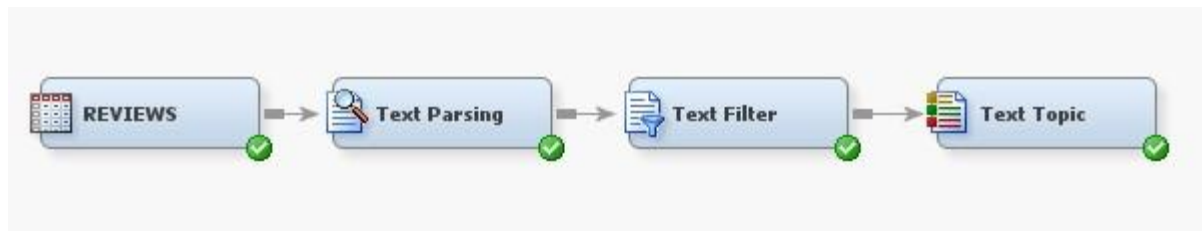


Fig. 6. Text Mining Process

Text Import

The downloaded data, in csv format was converted in to SAS dataset where each record was identified as a user review. This dataset was then imported in SAS® Enterprise Miner™ for further text analysis.

Text Parsing

After importing the text, the text parsing node is attached to it and a few modifications are made to clean up the unstructured text data. Using the properties panel:

- The 'find entities' option is set to standard.
- The 'detect different parts of speech' option is set to no to be able to represent one word or term as a whole and not have repetitive terms with different parts of speech.
- abbr, prop and num parts of speech have been ignored apart from the default option

With the help of this node the term by frequency document matrix is created, which is used to understand the most frequently occurring term and the number of documents it has occurred in. It is also used to analyze the terms that are rarely used. From research it is seen that typically the terms that are used moderately are the ones that are the most helpful in exploration and modeling.

Terms								
Term	Role	Attribute	Freq ▼	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable numdocs
+ be	... Verb	Alpha	9041	3217N		+	20140	2
+ card	... Noun	Alpha	7785	3243Y		+	3243	1
+ have	... Verb	Alpha	6455	2815N		+	20021	3
+ credit	... Noun	Alpha	3889	2188Y		+	3335	5
+ get	... Verb	Alpha	3848	2183N		+	19965	6
+ not	... Adv	Alpha	3695	1931N		+	20040	7
+ approve	... Verb	Alpha	3315	2401Y		+	10751	4
+ month	... Noun	Alpha	3151	1900Y		+	3551	9
+ apply	... Verb	Alpha	2767	1922Y		+	12478	8
+ do	... Verb	Alpha	2477	1429N		+	20198	10
+ pay	... Verb	Alpha	2075	1335Y		+	11419	11
+ score	... Noun	Alpha	1811	1294Y		+	3107	12
+ year	... Noun	Alpha	1589	1098Y		+	9489	13
+ use	... Verb	Alpha	1528	1086N		+	20028	14
+ payment	... Noun	Alpha	1506	1035Y		+	2066	16
+ capital	... Noun	Alpha	1488	1001Y		+	2321	17
+ s	... Noun	Alpha	1445	978N		+	19963	18
+ limit	... Noun	Alpha	1444	1073Y		+	10376	15
+ go	... Verb	Alpha	1309	947N		+	19989	21
i	... Noun	Alpha	1281	508N			20268	53
no	... Adv	Alpha	1246	940N			20229	22
+ now	... Adv	Alpha	1241	974N		+	20252	19
one	... Prop	Alpha	1199	744N			19899	29
+ make	... Verb	Alpha	1190	874N		+	20087	25
+ good	... Adj	Alpha	1165	968Y		+	1049	20

Fig. 7. Text Parsing Output

The most frequently used terms are card, credit, approve, month and apply which is expected since the reviews are for a credit card. Terms like 'be', 'have', 'not' etc. which is eliminated later using the text filter node.

Text Filter

The text filter node is added which is used to eliminate the terms that occur minimal number of times in all the documents by manually entering the desired required number of documents the words should be present. We can also perform spell check by enabling the option present in the properties panel. Spell check option also handles the terms that could be potential synonyms. The terms 'apporval', 'approval' were changed to approval.

	numdocs	term	childndocs	parent	termrole	parentrole	minsped	dict	te
1	5	managable	1	manageable	Noun	Adj	5	N	
2	52	aplication proc...	1	application pro...	NOUN_GROUP	NOUN_GROUP	2	N	
3	14	americanexpress	1	american express	Prop	NOUN_GROUP	6	N	
4	8	wouldnt	1	wouldnt	Prop	Noun	0	N	
5	8	wouldn	1	wouldnt	Noun	Noun	5	N	
6	41	qualifty	1	qualify	Noun	Verb	7	N	
7	41	prevoius	1	previous	Noun	Adj	6	N	
8	6	quarterly	1	quarterly	Noun	Adj	0	Y	
9	23	prequalified	1	prequalified	Noun	Adj	4	N	
10	23	prequalified	3	prequalified	Prop	Adj	0	N	
11	23	prequalifier	2	prequalified	Noun	Adj	8	N	
12	20	mention	3	mention	Noun	Verb	0	Y	
13	20	meantion	2	mention	Noun	Verb	7	N	
14	48	raiting	1	rating	Prop	Noun	8	N	
15	5	communcation	1	communication	Noun	Noun	4	N	
16	7	milestone	3	milestone	Prop	Noun	0	Y	

Fig. 8. Text Filter Spell Check

Terms											
Term	Role	Attribute	Status ▲	Weight	Imported Frequency	Freq	Number of Imported Documents	# Docs	Rank	Parent/Child Status	Parent ID
+ be	Verb	Alpha	Drop	0.000	9041	9041	3217	3217	2+		20140
+ have	Verb	Alpha	Drop	0.000	6455	6455	2815	2815	3+		20021
+ get	Verb	Alpha	Drop	0.000	3848	3848	2183	2183	6+		19965
+ not	Adv	Alpha	Drop	0.000	3695	3695	1931	1931	7+		20040
+ do	Verb	Alpha	Drop	0.000	2477	2477	1429	1429	10+		20198
+ use	Verb	Alpha	Drop	0.000	1528	1528	1086	1086	14+		20028
+ s	Noun	Alpha	Drop	0.000	1445	1445	978	978	18+		19963
+ now	Adv	Alpha	Drop	0.000	1241	1241	974	974	19+		20252
+ go	Verb	Alpha	Drop	0.000	1309	1309	947	947	21+		19989
no	Adv	Alpha	Drop	0.000	1246	1246	940	940	22		20229
+ just	Adv	Alpha	Drop	0.000	1125	1125	884	884	23+		19996
+ give	Verb	Alpha	Drop	0.000	1101	1101	876	876	24+		20168
+ make	Verb	Alpha	Drop	0.000	1190	1190	874	874	25+		20087
+ very	Adv	Alpha	Drop	0.000	977	977	801	801	27+		19998
+ say	Verb	Alpha	Drop	0.000	1022	1022	763	763	28+		20199
one	Prop	Alpha	Drop	0.000	1199	1199	744	744	29		19899
one	Num	Alpha	Drop	0.000	867	867	681	681	31		20069
+ so	Adv	Alpha	Drop	0.000	771	771	620	620	34+		20072
k	Noun	Alpha	Drop	0.000	1002	1002	613	613	35		20127
+ also	Adv	Alpha	Drop	0.000	671	671	560	560	45+		19922
+ then	Adv	Alpha	Drop	0.000	661	661	524	524	49+		19954
only	Adv	Alpha	Drop	0.000	575	575	517	517	50		20266
+ take	Verb	Alpha	Drop	0.000	599	599	516	516	51+		20142
+ ago	Adv	Alpha	Drop	0.000	577	577	509	509	52+		20148
i	Noun	Alpha	Drop	0.000	1281	1281	508	508	53		20268
+ other	Adj	Alpha	Drop	0.000	557	557	473	473	57+		20223

Fig. 9. Text Filter Output

After running the text filter node, we can see that terms such as be, not, have, do, get etc., are dropped from the text since they do not contribute towards any meaning in the review. Only words that are related to a credit card in some way are kept.

Text filter is also used to group synonyms together. It can be done by importing a file with all the synonyms or manually by dragging and dropping the terms into each other.

18	83	aprove	2	approve	Noun	Verb	4	N
19	83	approv	1	approve	Noun	Verb	5	N
20	16	good approval ...	1	good approval ...	NOUN_GROUP	NOUN_GROUP	2	N
21	574	aproval	4	approval	Noun	Noun	3	N
22	574	apporval	1	approval	Noun	Noun	6	N
23	574	approaval	1	approval	Noun	Noun	6	N
24	28	pre-approved	2	pre-approved	Prop	Verb	0	N
25	28	pre-apprived	1	pre-approved	Noun	Verb	4	N
26	2347	approved	1	approved	Noun	Verb	6	N
27	2347	approved	7	approved	Noun	Verb	7	N
28	2347	approved	1	approved	Noun	Verb	3	N
29	2347	apprived	1	approved	Prop	Verb	12	N
30	2347	approed	1	approved	Noun	Verb	7	N
31	2347	approved	5	approved	Prop	Verb	0	Y
32	12	aprovals	1	approvals	Noun	Noun	3	N

Fig. 10. Synonyms Grouping

Terms							
	TERM ▲	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
+	approval verse	1	1	☐	0.0	Noun Group	Alpha
	approval website	1	1	☐	0.0	Noun Group	Alpha
-	approve	3315	2401	☒	0.096	Verb	Alpha
	approve	87	83	☐		Verb	Alpha
	apprived	1	1	☐		Prop	Alpha
	aprove	2	2	☐		Noun	Alpha
	approv	1	1	☐		Noun	Alpha
	approves	3	3	☐		Verb	Alpha
	approved	1	1	☐		Noun	Alpha
	approed	1	1	☐		Noun	Alpha
	approved	5	5	☐		Prop	Alpha
	approving	7	7	☐		Verb	Alpha
	approved	1	1	☐		Noun	Alpha
	approved	3206	2347	☐		Verb	Alpha

Fig. 11. Synonyms Grouping

The above screenshot shows the synonyms for the term 'approve'. Terms such as 'approve', 'approved', 'approving' are grouped together as 'approve' using the interactive filter viewer.

Concept Links

Concept links can be viewed in the interactive filter viewer, in the text filter node. This enables us to understand the type of association between the terms used. Concept links can be produced for all the terms present in the documents, however it makes more sense to produce the concept links for the important terms. The term to be analyzed is in the center and the terms that it is highly used with it comes at the end of the links. The width of the link between the terms depicts the strength of association between them. The thicker

the link the stronger is the association and the more important it is and vice versa. Some of the concept links for important terms are shown below:

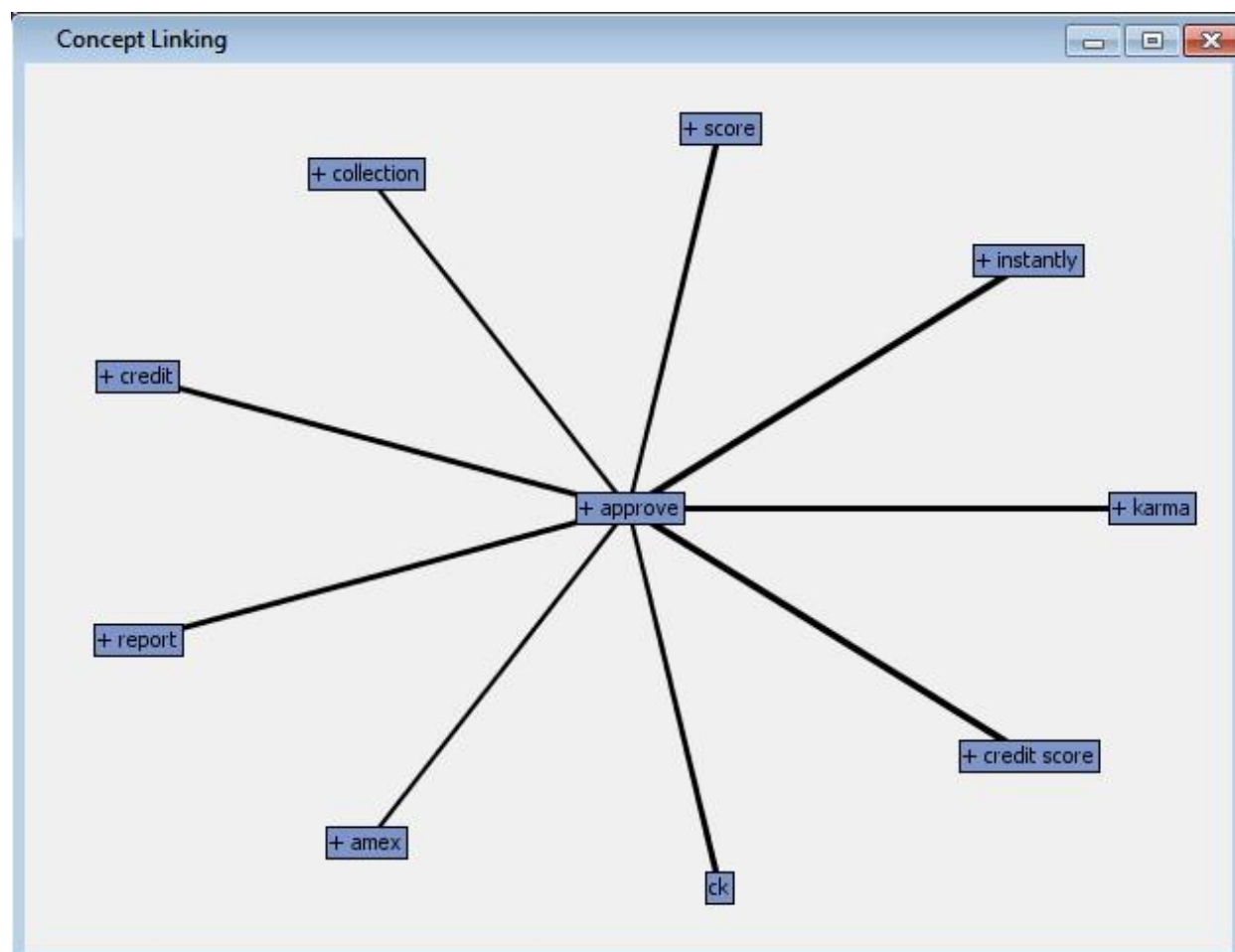


Fig. 12. Concept Link for 'Approve'

We viewed the concept link for the word "Approve" and found that the words "Instantly" and "Credit" have the strongest link with "Approve" and the second strongest links were with terms "score" and "report" as there were many reviews that text like "Approved with high credit score", "Applied and approved instantly", "Approved with high credit limit", "Approved with a limit of \$5k" etc.

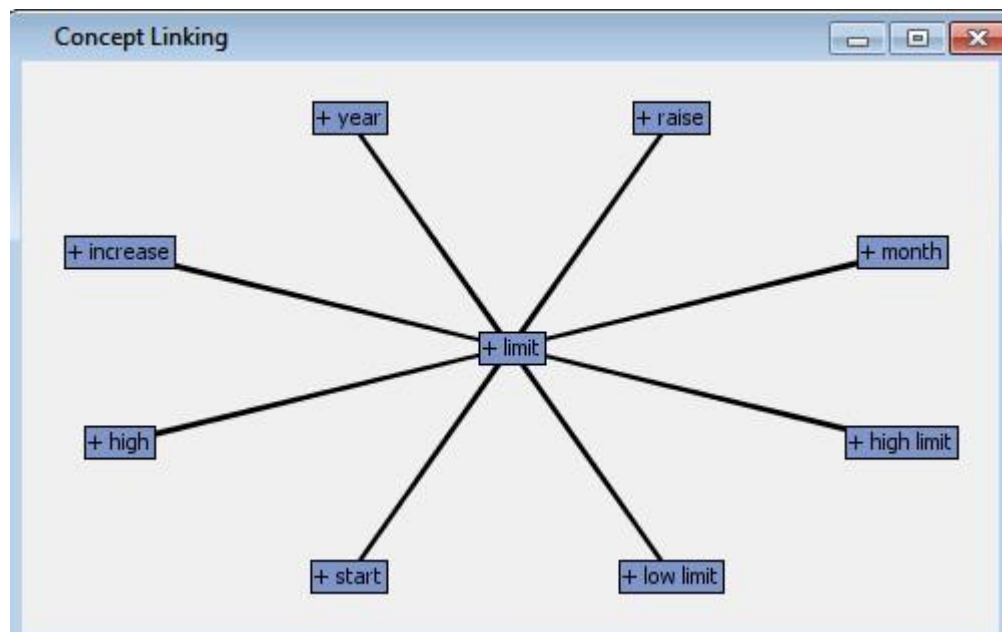


Fig. 13. Concept Link for 'Limit'

The term 'Limit' has strongest link with 'Increase', 'high limit' followed by 'start', 'month' since there were reviews like "approved with high limit", "the limit was increased after a year" etc.

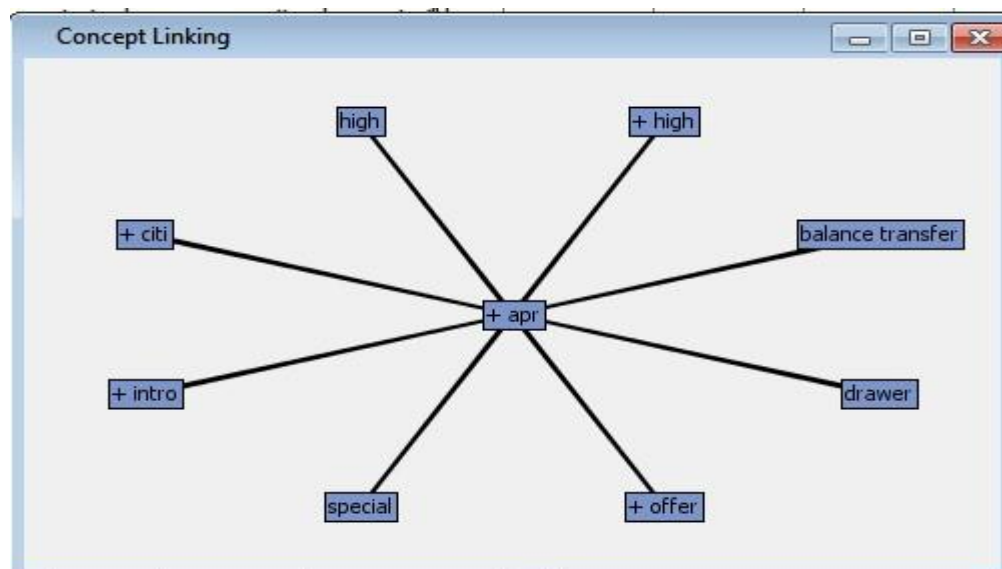


Fig. 14. Concept Link for 'Apr.'

The term 'Apr.' has strongest link with 'offer', 'intro' followed by 'high', 'special' since there were reviews like "offered with high Apr.", "special Apr. offers" etc.

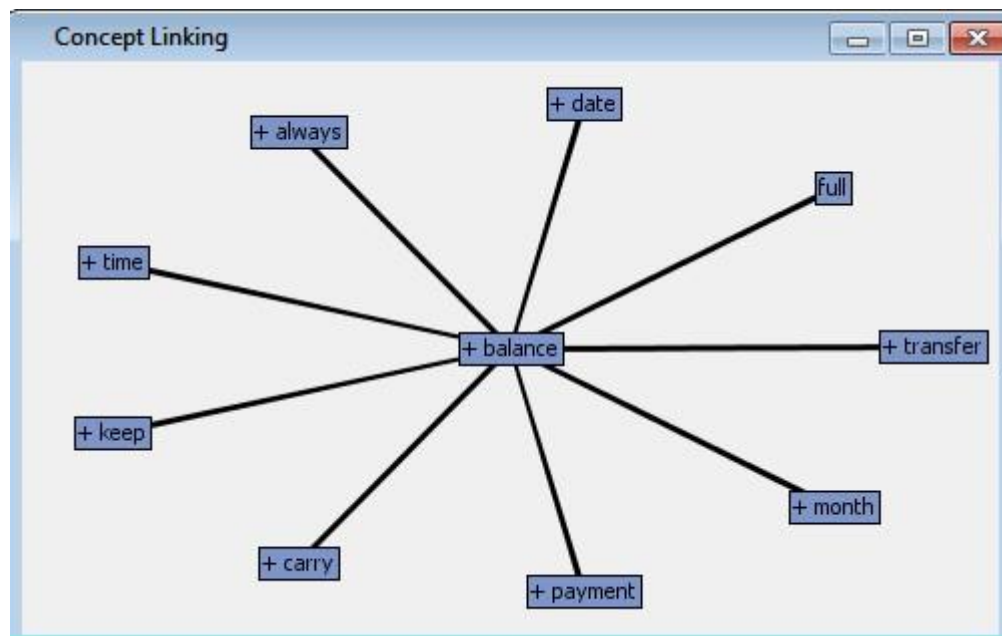


Fig. 15. Concept Link for 'Balance'

The term 'Balance' has strongest link with 'transfer', 'month' followed by 'payment', 'full' since there were reviews like "balance transfer every month", "full balance transfers" etc.

2. ANALYSIS OF POSITIVE & NEGATIVE REVIEWS SEPARATELY

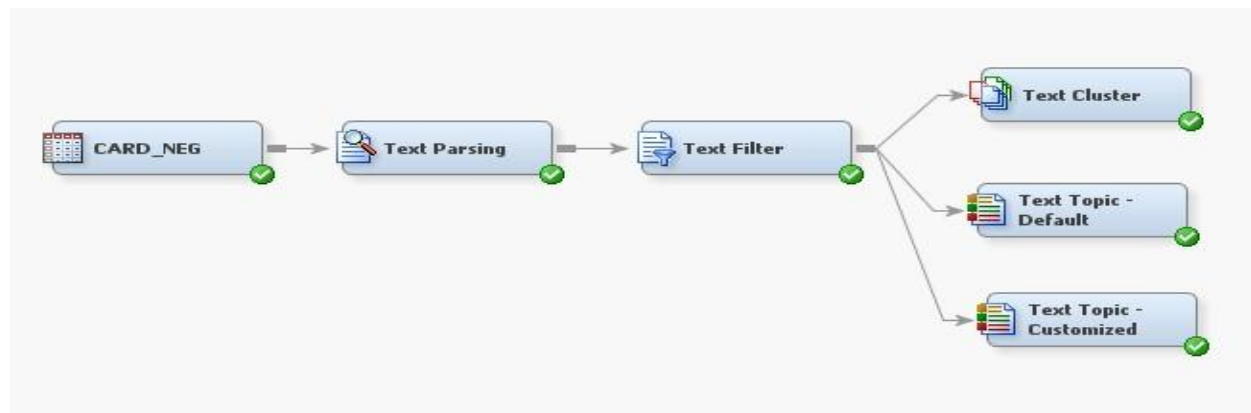


Fig. 16. Text Mining Process for Negative Reviews

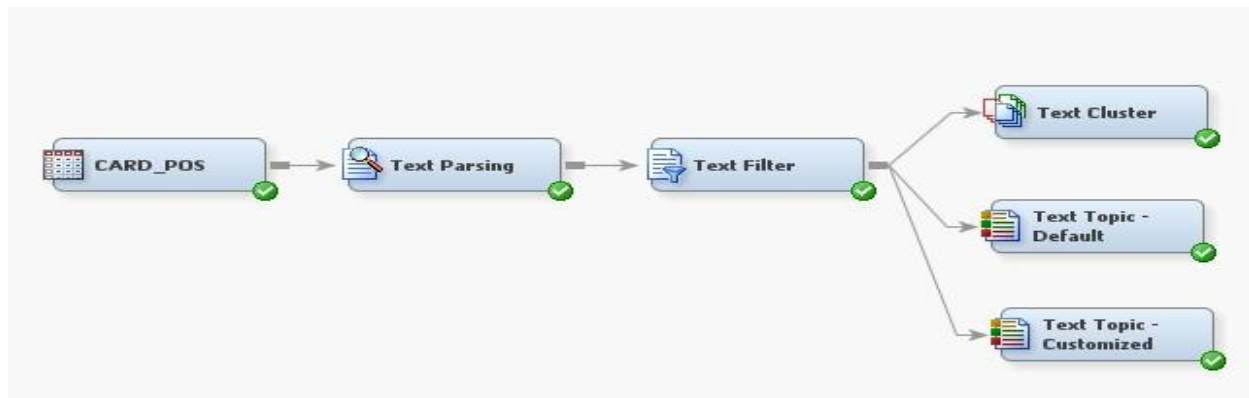


Fig. 17. Text Mining Process for Positive Reviews

In the above analysis, rest all steps i.e. from Text Import through Text Filter remain the same the interesting part comes when we do Text Clustering and Text Topic

Text Clustering

Once the text has been filtered using the Text Filter node we group similar terms in the dataset together. With the help of SAS® Enterprise Miner™ we can group terms closely related to each other into separate clusters of related terms. By default, 10 clusters are created by using Expectation-Maximization Cluster Algorithm and 8 descriptive terms that describe the cluster.

Negative Reviews:

For Negative Reviews, we can see that the 9 clusters generated are well separated from each other. The pie chart shows the distribution of the cluster frequencies. Apart from the cluster number 6,7, 4 and 3 the frequencies are well distributed among all the clusters as can be seen below.



Fig. 18. Distance Between the Clusters for Negative Reviews

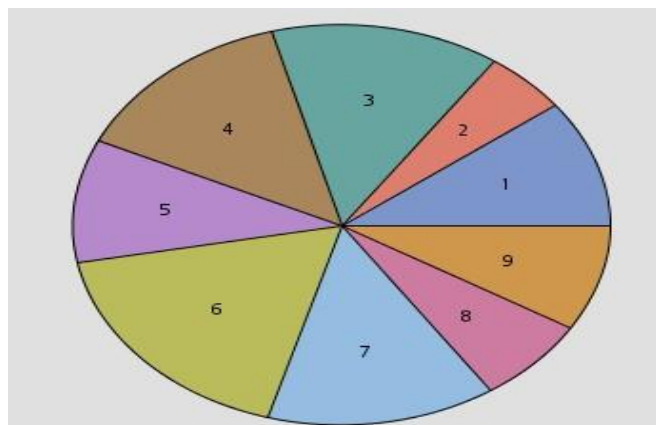


Fig. 19. Distribution of frequencies between the clusters

ClusterID	Descriptive Terms	Percentage	Explanation
1	+cash +reward +purchase +spend	10%	Talks about the rewards on spending and purchase
2	+owe +amount +limit +bank +account +payment +report	5%	This cluster is a group of terms for the balance owed to the bank
3	+rate interest +credit score +point +high +credit +good	14%	This cluster talks about the Interest rates depending on the credit score
4	+customer service +dispute +charge +time	14%	Describes the customer service complaints due to frauds and disputes
5	+phone +cancel +close +day +receive +look	10%	Talks about the cancellation or closing the credit card on
6	+payment money +bill +pay capital +report	17%	This cluster describes about the Monthly bill amount
7	+increase +inquiry +approve +line +limit	14%	Describes about the monthly Talks about the limit
8	+annual fee +high +amount +capital	7%	This cluster talks about the high annual fee
9	+balance transfer +balance +approve +apply +offer +look	9%	This cluster talks about the balance transfer feature applicable in the credit card

Table 2: Text Clusters Generated for negative reviews

Positive Reviews:

For Positive Reviews, we can see that the 11 clusters generated are well separated from each other. The pie chart shows the distribution of the cluster frequencies. Apart from the cluster number 1, 5, 9 and 10 the frequencies are well distributed among all the clusters as can be seen below.



Fig. 20. Distance Between the Clusters for Positive Reviews

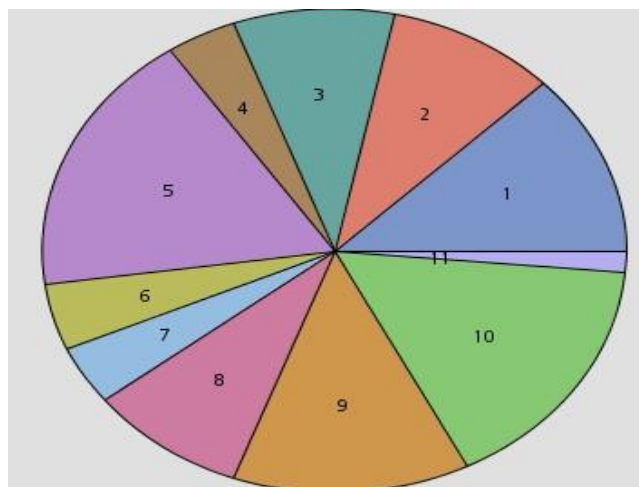


Fig. 21. Distribution of frequencies between the clusters

ClusterID	Descriptive Terms	Percentage	Explanation
1	+approve +score +apply +good odds	12%	This cluster describes about good approval rate of the credit card
2	+reward cash +back +purchase +love	10%	Talks about the good cashback and reward features
3	+balance interest +pay +rate +high +purchase	9%	This cluster groups all the reviews that talk about high interest rate on balance
4	+capital +quicksilver +start +deposit	4%	Describes the Good account to start for credit cards
5	+credit +increase +month +limit +score	18%	This cluster talks about Increasing the limit of
6	+date +report due +deposit +pay +balance +point +increase	4%	Talks about the increase of the credit score if the balance is deposited by due date
7	+fee annual +back +rate cash	4%	This cluster speaks about the annual fee of the credit cards
8	+service +customer service +problem +love +great	9%	Talks about good customer service and how the problems were resolved
9	+day +application +hard +pull +fraud +receive	13%	This explains that if there is any fraud then the credit line is hard stopped
10	+year +loan +secure +work +report	16%	This cluster talks about the Loan eligibility based on the credit report
11	+alert +problem +active +opportunity +hard	1%	Explains about the alerts when there is any issue and also gives

Table 3: Text Clusters Generated for positive review

Text Topic

After connecting the Text Filter node in SAS® Enterprise Miner™ we join the Text Topic node which will enable us to combine the term into topics so that we can analyze further. The image below.

Topic
+point,+score,+credit score,+report,+drop
+transfer,+balance,+transfer,+balance transfer,+approve
+fee,annual,+annual fee,+charge,first
+payment,+day,on time,+miss,+pay
interest,+rate,+high,high interest,+charge
service,+customer,customer service,+day,terrible
cash,back,+cash,+reward,back
+inquiry,hard,+hard inquiry,+report,+credit report
security,fraud,+activate,social,+verify
+close,+account,capital,+deposit,+secure
+line,+credit line,+credit,+approve,+limit
+credit limit,+limit,+amount,+wait,+release
+date,due,+deposit,+bill,due date
+reward,+purchase,+low,+point,great
+increase,+time,credit increase,+deny,waste
+approve,karma,+apply,approval,fico
capital,+reward,boa,+discover,+spend
+credit card,+card,+credit,+spend,+credit
+company,worst,late,+late fee,+bad card
+review,+increase,information,+month,+rep
+charge,+cancel,+dispute,+dispute,+pay
late,+late fee,+report,+open,+rate
+bank,money,chase,+amount,+post
+report,+bureau,+credit bureau,+bad card,worst
+credit score,+year,+limit,+score,joke

Fig. 22. Grouping of Negative text topics

Topic
+credit limit,+limit,+increase,+credit,+start
+fee,annual,+annual fee,+upgrade,foreign
+activate,+state,+inquiry,watchful,watchful eye
+approval,odds,approval odds,instant,instant approval
+great,+great card,+card,great customer service,+rebuild
+customer,+service,+customer service,great,great customer service
+loan,+year,+student,+student loan,+car
interest,+rate,+high,+balance,+interest rate
+line,+credit line,+credit,+increase,+credit line increase
+increase,+month,+request,+credit increase,+late
cash,+back,back,+cash,+cash back
+rebuild,+help,+start,+recommend,+report
+day,+receive,+application,mail,+apply
capital,+secure,unsecured,quicksilver,platinum
+easy,+process,+application,application process,+quick
cl,+apply,+approve,+month,cs
history,credit history,+build,+recommend,+credit
+date,due,+due date,+balance,+pay
+credit score,+score,+point,+credit,+jump
+payment,on time,+time payment,+increase,+love
+karma,+credit karma,+chance,poor,odds
+reward,+point,+earn,+love,+bonus
great,fico,utilization,+hard,+inquiry
+love,+credit card,capital,+card,first
+good,good card,+chance,+good chance,+credit card

Fig. 23. Grouping of Positive text topics

SENTIMENT ANALYSIS

SAS® Sentiment Analysis Studio is used to build a statistical model on the text data and to be able to classify the reviews as good or bad using the terms contained in it. The statistical model is run on the test data to check if the model is predicted accurately. 80% of the reviews are used for training the statistical model.

Training corpus	Card_Review
Set percentage for training	80%
Solution	Bayes Method
Probability threshold	0.50
Text normalization model	Smoothed Relative Frequency
Contextual extraction (optional)	
Runtime stop words (optional)	
Text Result Graphical Result	
With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [No Feature Ranking]: Overall precision: 90.54% Positive precision: 94.00% Negative precision: 56.92%	
With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Risk Ratio]: Overall precision: 90.83% Positive precision: 95.58% Negative precision: 44.62%	
With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Chi Square]: Overall precision: 70.49% Positive precision: 68.72% Negative precision: 87.69%	
With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Information Gain]: Overall precision: 73.64% Positive precision: 72.51% Negative precision: 84.62%	
BEST MODEL is Smoothed Relative Frequency and Risk Ratio	

Fig. 24. Statistical Model Output



Fig. 25. Best Model

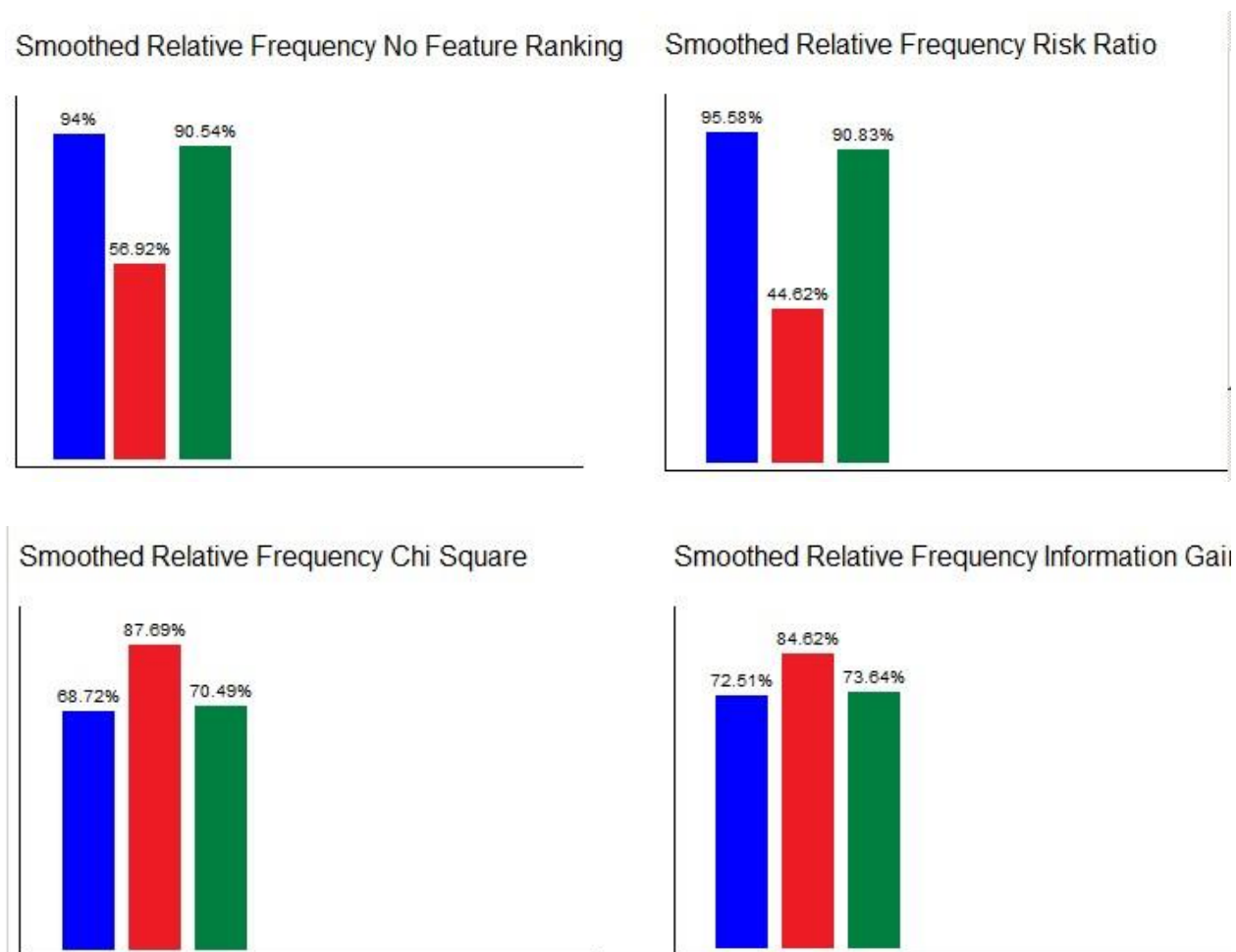


Fig. 25. Statistical Model Graphical Results

By running the statistical models on the training data we find that the Smoothed Relative Frequency and Risk Ratio is the best model chosen with its overall precision at 95.58%. The smoothed relative frequency algorithm is a text normalization method that corrects for the length of the document and the number of feature words per document to maintain consistency since some of the documents may be small while others are large. Now, the model is tested using the test data.

Model Testing

The statistical model built is used to test the model accuracy on the test dataset for both the positive and negative reviews. We have used a total of 1421 reviews for the testing the accuracy of the statistical model.

Test for Positive Reviews:

Text Result	Graphical Result
Results for selected folder: This directory is Positive Number of articles:1274 Number of positive articles:1233 Number of negative articles:34 Number of neutral articles:7 Positive percent:96.78%.	

Sentiment Distribution

■ Positive ■ Negative ■ Neutral



Fig. 26. Positive Reviews testing Output

It can be seen that the precision for identifying positive reviews is 96.78%. It is higher than the model accuracy obtained by using the training data. It correctly classifies 1233 out of 1274 articles as positive.

Test for Negative Reviews:

Text Result	Graphical Result
Results for selected folder: This directory is Positive Number of articles:147 Number of positive articles:84 Number of negative articles:62 Number of neutral articles:1 Positive percent:57.14%.	

Sentiment Distribution

■ Positive ■ Negative ■ Neutral



Fig. 27. Negative Reviews testing Output

By observing the above table and graph, we can see that the model classifies 42.86% of the documents correctly using the statistical model built which is higher than the percent correctly classified in the training data. 62 out of 147 documents are correctly classified as negative.

Though the statistical model shows good precision in predicting the positive and negative reviews it cannot be explained easily since it is like a black box. We do not know what terms were considered by the model as good or bad.

3 TEXT ANALYSIS OF MAJOR CREDIT CARDS

We wanted to figure out how the major credit cards are faring in the market. Thus it was important to realize what opinion consumers have on relevant topics.



Fig. 28. Text Mining Process for Major Credit Cards

TOPICS	CREDIT CARDS				
	Capital One	American Express	Chase	CitiBank	Discover
Customer Service	Great		Great		Great
Application Process	Easy	Easy	Easy	Easy	
Annual Fee			-		
Balance Transfer			Yes	Yes	
Approval			Good	Poor	High & Instantly
Interest Rate	High			Low	
Features	Build Credit	Great Benefits	Easy Foreign Usa	Low credit Limit	First year Double Bonus, Student Loan

Fig. 28. Discussion Topics for Major Credit Car

From the above table we can see that:

- Customer Service: Capital One, Chase and Discover have very nice Customer Service.
- Application Process: The Application process for a Discover card seems not so easy.
- Balance Transfer: Chase and Citibank offers Balance Transfer.

- Approval: Discover applicants have a very high and instant approval rate on the other hand Citibank seems to have a low approval rate.
- Interest Rate: Capital One credit card has a high APR and Citibank card has low APR rate.

CONCLUSION

Around 80 percent of the data that is available in the real world is unstructured, of which text data is a major portion. Customer reviews play an important role in determining the true perception of any product among the audience and to specifically understand what they liked or disliked in the product, which in our case is a credit card. It also gives insights into what the people expected from the credit card and what they actually want to have as features from the card. This information can be leveraged by the credit card companies, banks and financial institutions to make better quality credit cards add/modify in the features to cater to the needs and expectations of the users and prospects. Concept links can be used to understand the association of one term with others depending on how often they are used together. For example, the term 'Approve' is strongly associated with 'Credit', 'Limit', etc. indicating that the audience pays more attention to the credit limit provided by the card when the application gets approved, which rests on the shoulder of the issuers. The raw data needs to be parsed and filtered before being analyzed to correct for spelling mistakes, to group synonyms together and to drop the terms that do not contribute in making sense of the data. In general, credit card user reviews will be made available immediately after a user posts about the product after few months of usage of the product or by immediate perception of the product and there will be a section of the audience who would just like to know if the credit card is worth the application or compare and assess which card to apply for. Using the statistical model and text rule based model built we can identify terms and rules that help us classify reviews into good or bad. This can help people who want value and their expectations met to have any particular credit card or not.

REFERENCES

- *Journal article* Learning Word Vectors for Sentiment Analysis by Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).
- *Book* Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® by Goutam Chakraborty, Murali Pagolu, Satish Garla.
- *Journal article* Sentiment Analysis and Opinion Mining by Bing Liu (May 2012).
- *Book* SAS Institute Inc 2014. Getting Started with SAS® Text Miner 13.2. Cary, NC: SAS Institute Inc.
- *Journal article* Huayi Li, Arjun Mukherjee, Jianfeng Si and Bing Liu. Extracting Verb Expressions Implying Negative Opinions. Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15). 2015.
- *Journal article* Zhiyuan Chen, Nianzu Ma and Bing Liu. "Lifelong Learning for Sentiment Classification" to appear in Proceedings of the 53st Annual Meeting of the Association for Computational Linguistics (ACL-2013, short paper), 26-31, July 2015, Beijing, China.

ACKNOWLEDGMENTS

- We heartily thank Dr. Goutam Chakraborty, Professor, Department of Marketing and founder of SAS and OSU Data Mining Certificate program, Director MS in Business Analytics – Oklahoma State University for his constant inputs and guidance without which we couldn't have done this research project.
- We thank SAS Global Forum 2016 conference committee for giving us an opportunity to present our work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Anirban Chakraborty
Organization: Oklahoma State University
Phone: (918)813-4998
Email: anirban.chakraborty@okstate.edu

Surya Bhaskar Ayyalasomayajula
Organization: Oklahoma State University
Phone: (918)853-8472
Email: suryaba@okstate.edu