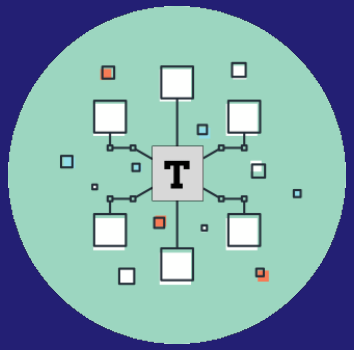
TEXT ANALYTICS

Gramener
A Data Science Company

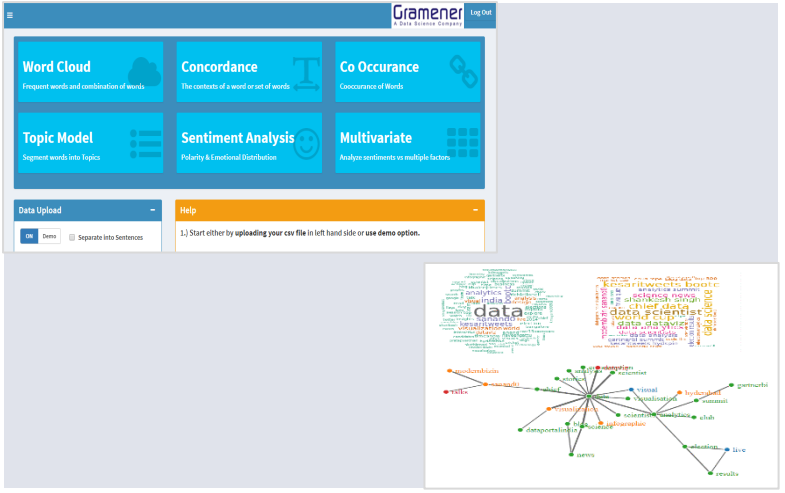
TEXT ANALYSIS – ACCELERATOR



- An engine to generate insights from mining and analyze any text data.
- This tool can help to find out impact of date (week/month) and any numerical columns like rating, no of likes on textual feedbacks, scripts etc..
- Engine and UI developed on R Studio using the Shiny interface.

FEATURES

- Uses basic analysis like word clouds (unigram, bigram and so on) , concordance and co-occurrence graph to find out basic view on textual data.
- Uses Topic modelling (LDA) to differentiate blog of text into different topics.
- Uses Sentiment analytics to identify overall emotions from the text.



APPLICATION AREAS

Social Media
Analyze social buzzes about any recent topics which can help in campaigning.

Entertainment
Analyze movie/ TV serial scripts to find out impact of particular actors, genre to get better impressions.

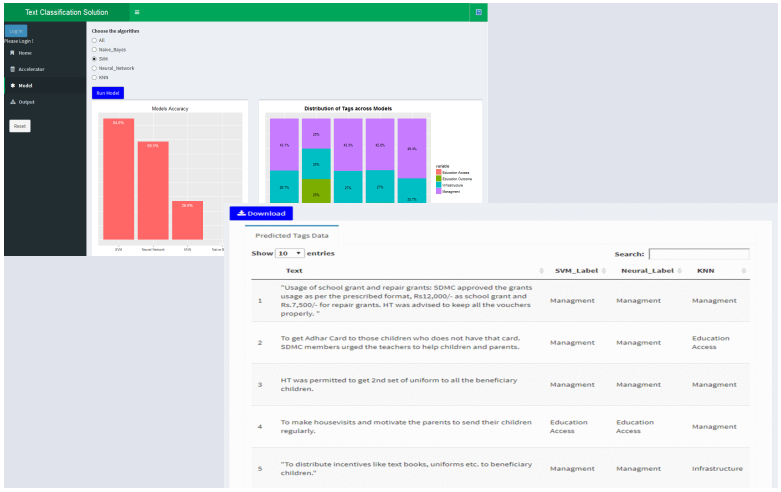
Feedback of any Services
Analyze users feedback for any kind of services like Hotel, Restaurants, Transports etc.. to serve your customer better.

TEXT CLASSIFICATION – ACCELERATOR

- An engine developed on R Studio and Shiny package
- Tool generates Tags for untagged text data by building classification models on tagged data.
- This tool can help in reducing the process of laborious and expensive ways to read each piece of feedback/response and classify them into different themes/topics

FEATURES

- Generate tags for untagged data along with the accuracy of each model and confusion matrix
- Choice of 4 different classification models – Naïve Bayes, SVM, KNN and Neural Network.
- Business has quick turnaround time and scale infinitely for text classification to perform instant decision making



APPLICATION AREAS

Brand Association Mapping
Classify and Analyze social buzzes of a product or service into different themes to understand brand perception

Infrastructure & Service Management
Classify generated IT tickets to assign a right resolver group thus reduce resolution time significantly

Feedback Analysis
Classify feedback provided both by internal & external stakeholders to understand areas with high satisfaction / scope for improvement

Improving customer satisfaction using unstructured data

Problem

A global IT hardware vendor wanted to improve their key business metric: the Net Promoter Score (NPS) for customer satisfaction.

The customer wanted to analyze the vast amount of unstructured textual and verbal feedback they receive from customers

Approach

Each piece of raw customer feedback text data was auto-tagged to a business theme. We identified its sentiment and identified prominent topics & their performance.

We then applied Impact Analysis to identify which variables / factors were most important in predicting NPS.

Outcome

On a user base of 10 million, this exercise found process improvements that could improve the NPS by 5% points.

Further, it showed that the factors affecting the NPS the most were account relationship & tech support, not installation and website

Analytics | IT & SM

VISUALIZING NET PROMOTER SCORE USING SENTIMENT & TOPIC

Each of the bubbles below is a comment. The *x-axis* signifies sentiment & the comments are segmented on the *y-axis* based on *NPS*. Click **here** to know more about the key factors that influence *NPS score* (This is at an overall level).

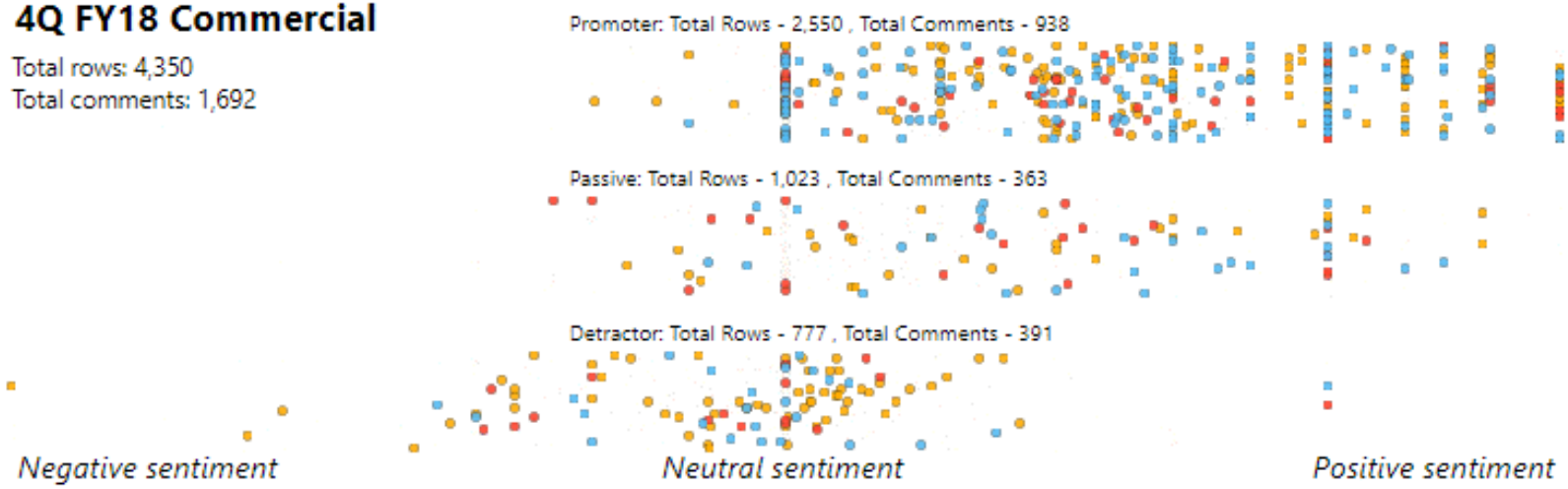
The color signifies Region - EMEA, AMERICAS & APJ.

Search comments | Reset

Represent customer sentiment along with count of comments by timeline

4Q FY18 Commercial

Total rows: 4,350
Total comments: 1,692



To know the key topics discussed, click on these 'Theme' buttons.

Technical Support | Account Relationship | CSG products | ISG products | Online | Order Experience | Compared to Competition

Text analytics for a leading technology solutions provider

Problem

A large IT hardware vendor wanted to improve the customer satisfaction score among its customers

The customer wanted to analyze the vast amount of unstructured feedback he gets from customers

Approach

Analyzed customer feedback (text data)

Keywords were tagged to business themes with a frequency count

For each set of feedback, customer sentiment was identified and represented

Impact analysis was done to identify variable importance

Outcome

Identified process improvements for 5 %age points NPS improvement on a 10mn user base

Identified the most important parameters affecting the NPS as account relationship & tech support as opposed to installation and website

Analytics | IT & SM

REPRESENT FREQUENCY COUNT (BY BUSINESS THEME) AS A WORD CLOUD

To know the key topics discussed, click on these 'Theme' buttons.

Technical Support | Account Relationship | CSG products | ISG products | Online | Order Experience | Compared to Competition

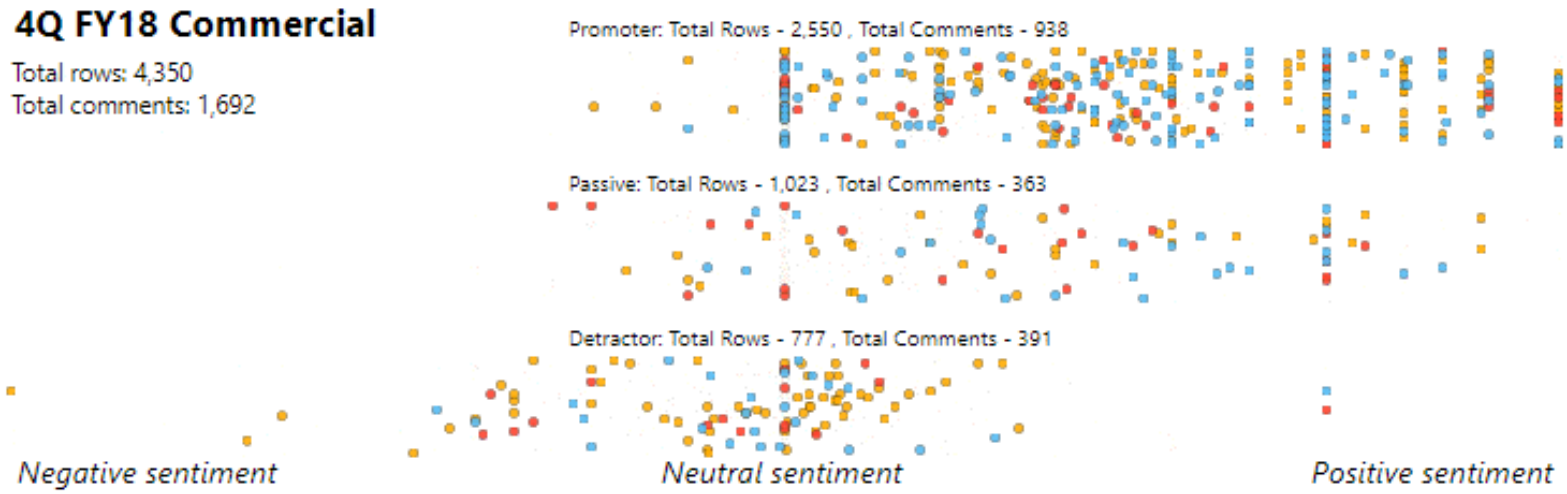
Approximately 33% of the comments talk about 'Technical Support'

REPRESENT CUSTOMER SENTIMENT ALONG WITH COUNT OF COMMENTS BY TIMELINE

Each of the bubbles below is a comment. The *x-axis* signifies sentiment & the comments are segmented on the *y-axis* based on *NPS*. Click **here** to know more about the key factors that influence *NPS score* (This is at an overall level).

The color signifies Region - EMEA, AMERICAS & APJ.

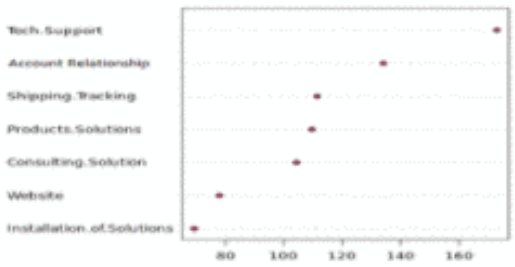
Search comments | Reset



REPRESENT VARIABLE IMPORTANCE / IMPACT ANALYSIS AS A STATIC VIEW

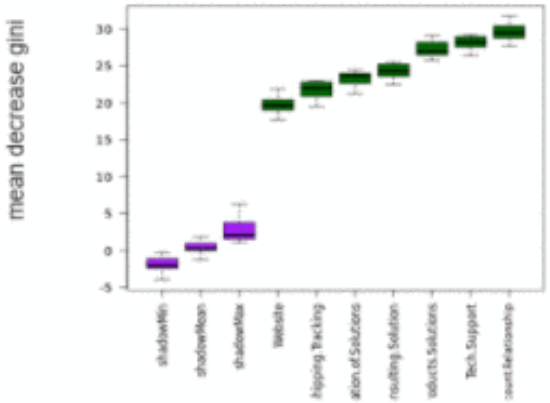
IMPACT ANALYSIS & INFERENCE – 4QFY18

Random Forest variable importance plot



Higher the value, higher is the importance

Boruta variable importance plot



Higher the value, higher is the importance

Final Rank Table

Importance	Variables	Rank - Commercial
Top	Account Relationship	1
	Tech Support	2
Middle	Products Solutions	3
	Shipping Tracking	4
	Consulting Solution	5
Bottom	Installation of Solutions	6
	Website	7

Importance	Variables	Rank - Enterprise
Top	Account Relationship	1
	Tech Support	2
Middle	Products Solutions	3
	Shipping Tracking	4
	Consulting Solution	5
Bottom	Installation of Solutions	6
	Website	7

Insights:

- For both Commercial and Enterprise, both technique 1 (Random Forest) and technique 2 (Boruta) have highlighted ***"Account Relationship" and "Tech Support",*** as the two most important features

- *The variation in the values are very marginal; This also highlights that impact on NPS also varies marginally*
- *These text are renamed for convenience: 'Sales Account relationship' as 'Account Relationship, 'Shipping, tracking delivery of Orders' as 'Shipping Tracking', 'Using products & solutions' as 'Products Solutions', 'Technical Support for Hardware, Software, Services, Solutions' as 'Tech Support'*

** The representative plots highlight the value for Commercial data*

SHOW CONTENT ANALYSIS FOR A TV CHANNEL

Problem

A leading Hindi General entertainment channel wanted to improve its TRPs for 2 shows by altering its script.

The content team wanted to analyze impact of characters and emotions on the show performance

Approach

Gramener analyzed performance of shows(TVR, Reach and TSV) at a minute level and its attributes across markets, age groups, genres, competition

Episodic text was mined and converted into structured formats analyzing key characters, linkages and sentiments experienced

Outcome

Client was able to tap in key market preferences, leverage increase/decrease in character connections and interactions, redefine roles and control plots and themes from competition/genre to improve TRPs

How did we process the Text?

Here is how we are transforming the data for the text analytics dashboard. This is the simplest example, yet indicate the level of detail involved in the computations.

Data Extraction

Scrape & combine data

We combine internal script data, with viewership ratings data and other external data sources, including Social media

Text Analytics Engine

Token-ization

Tokenization breaks a stream of text into words, phrases or other meaningful parts

POS tagging

POS marks up a word in text as corresponding to a particular part of speech

Entity detection

Then, each sentence is searched for the interesting entities in this context.

Advanced Analytics Engine

Insight mining

We then apply statistical techniques to identify strong insights and key influencers..

Chars	Attrib	Ratings
A	abc	2.3
B	def	2.7
C	ghi	3.6
D	jkl	1.4
...	...	...
...	...	...
...	...	...
...	...	...

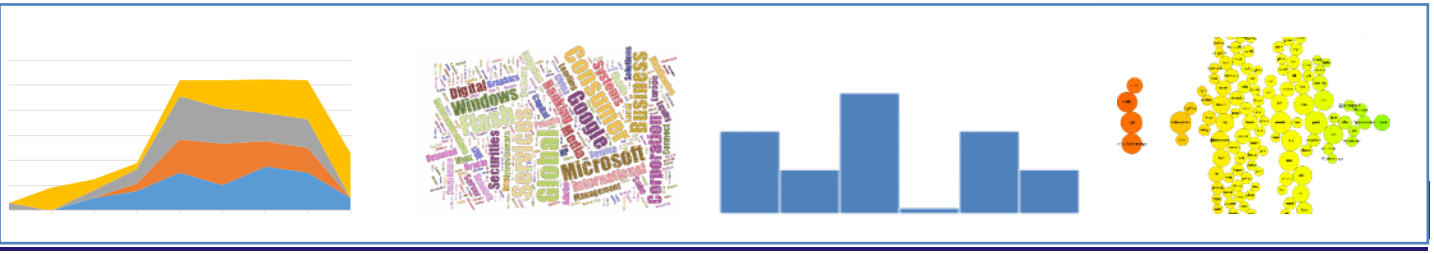
Modeling

...and statistically combine, analyze and model the extracted entitles into meaningful relationships...

Chars	Emot.	%Gr
A	53%	23%
B	51%	-35%
C	52%	95%
D	53%	101%
...	...	...
...	...	...
...	...	...
...	...	...

This transformed data serves as the input to the Visualization engine

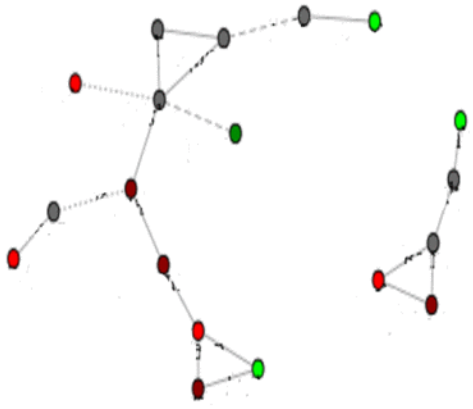
Visualization Engine



Text processing: Outcomes

- The major characters of the key Romance show either had neutral or negative impact on ratings.
- Negative external (non-family) characters were found to have positive impacts on ratings.
- Commonly repeated tropes such as reincarnation or the ‘suffering mother’ were met with lukewarm or negative responses from most markets.
- Similarly, common plots such as kidnappings or affairs have mixed reception from the audience.
- Comparatively, business and finance themed plots tend to perform well.
- Several high priority markets displayed strong preference for scenes with negative emotional content, such as fear, anger or disgust.
- Character interactions between the female lead of the key Romance show and antagonistic female characters were well received by most markets.

Character Networks, Character Types, Verb, Word Cloud



Keyword								
Love	Fear	Surprise	Trust	Joy	-	Joy	-	Joy
Tells	Anger	Disgust	Disgust/ Anger	Joy	-	Trust	Anger/ Disgust	-
House	Disgust/Anger	Trust	Anger	-	-	Anticipation	-	Anticipation
Truth	Anticipation/ Trust	Fear/Joy	Joy	-	-	Trust	Trust	-
Mother	Disgust/Anger	Surprise	Anticipation/ Trust	Trust/ Sadness	-	Sadness	-	-
Together	Fear	Fear/Joy	Joy	Anticipation	-	Anticipation	-	-
Time	Surprise/Fear	Anticipation/ Trust	Joy/Trust	Anticipation	-	Anticipation	-	-
Fight	Surprise	Fear/Anger	-	Anger/Fear	Joy/ Anticipation	Anger/Fear	-	-
Close	Neutral	Neutral	Joy/Sadness	Joy/Trust	-	-	Joy/Trust	-

Earnings Transcript Analysis: Popular Themes

Problem

A global shareholder services organization wanted to analyze earnings calls data and help understand trends in earning calls discussions.

Are there some phrases that trending up / down? Can we visually identify them?

Approach

Gramener extracted text from earning calls and auto-categorized them into themes. These were displayed them as a trending word cloud.

This allowed users to understand the popularity of each topic over the years.

Outcome

Our clients were able to quickly spot the prominent and trending topics on behalf of their clients, and coached analysts on the right topics and to improve their effectiveness in earnings call. This launched a new product line for our client.

Platform | Analytics | Text analysis | Financial Services

Earnings Transcript Analysis: Popular Themes

Popular themes in your Earning Calls - over the years *(Hover on the chart to see individual trends)*

Company: **Goldman Sachs**

Insight: Discussion over BASEL increased post 2008 Crisis *(Mouse-over)*

QUESTIONS ASKED