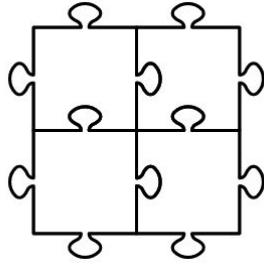


Case Study: Social Media Sentiment Analysis

Profisor Services

Sampurn Rattan Jain



Case Study Elaboration

Case Study Elaboration

5,334
customer interactions
in validation dataset

Domain: HealthCare

Business Goals: Analyze social media data to aggregate overall patient concern. Thus, allowing client company to efficiently address them in a structured, organised manner.

Sentiment analysis can influence marketing campaigns, generate leads, plan product development or improve customer service.

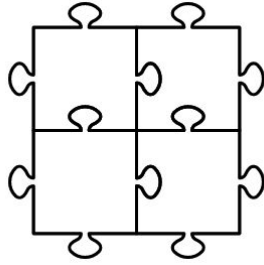
AI/ML Goals: Assign emotion labels, and corresponding score of predicted label to each customer social media interaction.

Case Study Elaboration

comment_id	comment_message
1538183672909402_1541895645871538	b'Thanks big fan here'
1540621789332257_1540637829330653	b'Reminded of trika butterfly'
1540621789332257_1540640095997093	b'BEST WISHES FOR CELEBRATING... 34 YEARS..OF TOUCHING LIVESThankyou Apollo Hospitals..'
1540621789332257_1540672129327223	b'Congrats'
1540621789332257_1540753779319058	b'Congrats'

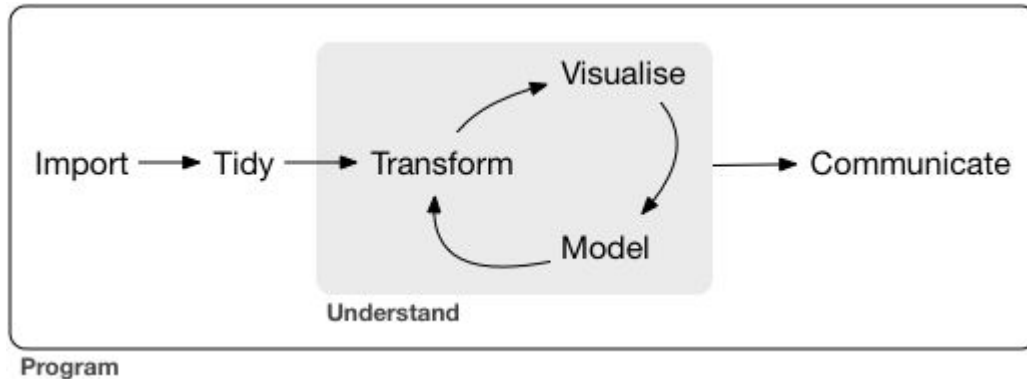
Validation Dataset Sample

- Headers
 - comment_id - unique identifier
 - comment_message - content body
- Non-parsable personalisations in text
 - Emojis
 - Irregular punctuation
 - Hinglish
 - Abbreviations



Solution Approach

Solution Approach



The ideal data science program flow.

The goal of the Transform -> Visualise -> Model iterations to understand and get the feel of the data.

However, in keeping with the goals of the case study, the number of iterations is set to one.

Solution Approach

class	name	description
2	Great CD	My lovely Pat has one of the GREAT voices of her generation...
2	One of the best game music soundtracks - for a game I didn't really play	Despite the fact that I have only played a small portion of the game...
1	Batteries died within a year ...	I bought this charger in Jul 2003 and it worked OK for a while...
2	works fine, but Maha Energy is better	Check out Maha Energy's website. Their Powerex MH-C204F...
2	Great for the non-audiophile	Reviewed quite a bit of the combo players and was hesitant due to unfavorable reviews and size of machines...

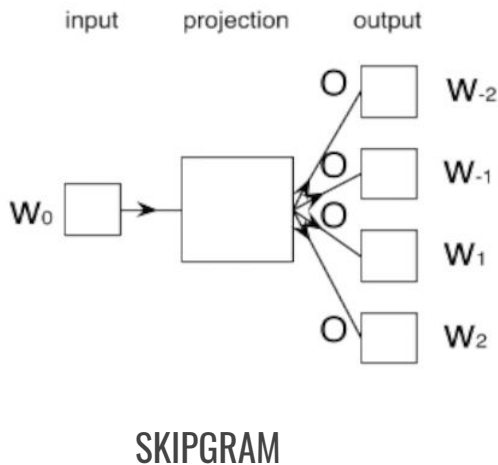
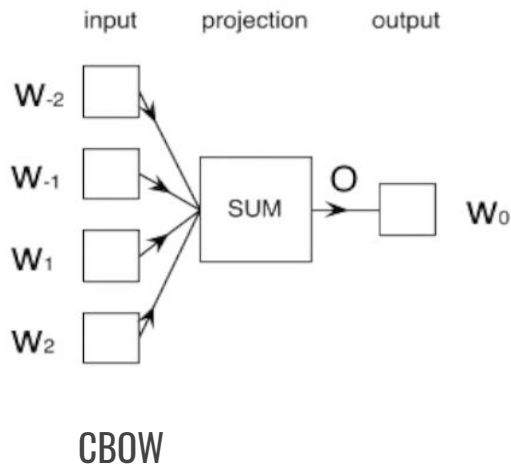
The Amazon Polarity Review dataset.

Contains 3.6 million reviews in the train set and 400,000 reviews in the test set. The reviews are positive (two) or negative (one). The dataset can be found [here](#).

Headers:

- class - 1 (negative), or 2 (positive)
- name - title of review
- description - review content body

Solution Approach



As directed, I used Facebook's powerful NLP fasttext library.

Inspired by word2vec's CBOW approach.

Most useful for polyglot models.

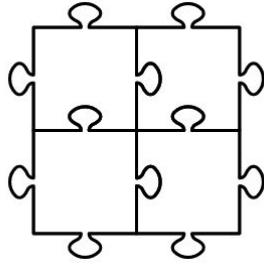
Efficient word representation learning - CBOW appended with a continuous skipgram.

Solution Approach

comment_id	comment_message	label	certainty
1538183672909402_1541895645871538	Thanks big fan here'	Positive	0.880859
1540621789332257_1540637829330653	Reminded of trika butterfly'	Positive	0.880859
1540621789332257_1540640095997093	BEST WISHES FOR CELEBRATING... 34 YEARS..OF TOUCHING LIVESThankyou Apollo Hospitals.'	Positive	0.880859
1540621789332257_1540672129327223	Congrats'	Positive	0.880859
1540621789332257_1540753779319058	Congrats'	Positive	0.880859

After training on the Amazon polarity review dataset, the model was run on the validation dataset. A label and certainty were attached to the final output.

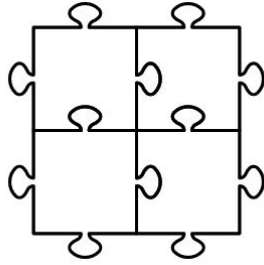
The precision of the model came out to be 0.941815. Similarly, the recall is 0.941815.



Conclusions

Conclusions

- Much was left to be desired in this initial model that has been created.
- Many improvements can be introduced to the above created model. Some include
 - Use of a more diversified dataset, with an emotional spectrum greater than two
 - Such datasets are rare in the “wild”, and all those that are available in the public domain would probably be of poor quality.
 - This ideal dataset would have to be created in-house, which is a time-consuming and painstaking exercise.
 - Model can also be trained to parse Hinglish
 - Excessively useful for an Indian healthcare company.
 - Again, the creation of such a training dataset is tedious.
 - Context-aware model
 - Many social media interactions are responses to previous posts. Model accuracy would greatly benefit if context is added to the input.



Thanks!