Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews
Natallia Casey
Bellevue University

Executive Summary

In this project, a classification analysis and Latent Dirichlet Allocation were used to perform sentiment mining with topic modeling of TuneIn Radio reviews scraped from Google Play. Exploratory data analysis, as well as data visualization, were used to derive valuable insights about the overall sentiment, the sentiment trends over time as well as the sentiment of product features.

The project draws attention to the fact that the overall sentiment of the application TuneIn Radio is positive. Yet, the positivity of the sentiment has been decreasing since 2016.

Further visualization analysis of the sentiment of the LDA topic modeling results showed that among all of the aspects of the application the users are happy with the content aspect the most. The premium subscription service, on the other hand, has the least positive sentiment.

Various recommendations based on the insights about the content of the reviews and the responses of the customer support to the users are made to help prevent customer churn and address the problems discovered.

The evaluation of the tuned multiclass classification models (logistic regression, random forest classifier, convolutional neural network, bidirectional LSTM, XGBoost, and Support Vector Machines) using the undersampled data sample showed that the Support Vector Machines was the best performing model. The metrics of F1, accuracy, recall, and precision are all the highest for this model (at about 70%). All of the models tended to struggle with the neutral class identification the most, while the Support Vector Machines performed the best for the neutral and negative classes. This is important because these two classes of reviews are prioritized for the replies by customer support specialists.

Background of the problem

Sentiment analysis of product reviews is an effective business method of gaining insights about customers’ opinions of the product and its features. Reading reviews could become unmanageable if the volume is large. Opinion mining creates an opportunity for a fast analysis of such text data.

A lot of the classification analyses exclude the neutral class. Yet, the sentiment analysis including the neutral class accounts for the neutrality of the sentiment. This is important because if only two sentiment categories are used the assumption that the text could only be either positive or negative is created. This is not always the case because there is a lot of text data that doesn't convey any sentiment. For example, the word "OK" could mean neither good nor bad.

There is text in the neutral class that conveys a mixed sentiment that is especially hard to comprehend for algorithms. For example, "The app is good, but I don't like the absence of my favorite radio station". 75% of the neutral class reviews in this dataset contain the mixed sentiment. These conclusion was made based on the presence of the conjunction "but" in these observations.

Also, the reviews with neutral sentiment tend to contain questions about the product. For example, "When is the next release?". The questions represent useful information as they are often indicators of a potential need for dialogue between a user and the company. As could be seen, there are several semantic representations of the neutral class which makes it complex and hard to predict.

Identifying the aspects of a product and their sentiment is essential for taking the necessary actions to fix the problems or understand what users like.

Tracking the responses to the most relevant reviews is a good strategy to create a strong trustful partnership with the users and prevent customer churn. It can also serve as a metric for customer support effectiveness.

TuneIn Radio provides user-centered content and could retain, and attract more users by monitoring the sentiment trends and addressing the customer concerns by replying to the reviews.

Methods

for reproducibility

OS	Languages	Libraries	**Data**	Seed
Windows Home 10,64-bit operating system	Python(3.7.2.)	pandas, matplotlib, numpy, imblearn, scikit-learn, spacy, tensorflow, keras, nltk	Scraped from Google Play TuneIn Radio on March 16th 2021. 8 variables (Appendix Part5). Imbalanced: **positive** class – 54,103, **negative** – 29,936, **neutral** – 10,036	777

Data understanding with EDA and visualizations

EDA (Appendix. Part 4)

- The information derived from identifying the percentage of missing values for the variable ReplyContent allowed to highlight the trend of the customer support assistance over time. Based on this information and the data from the variable ThubsUpCount a list of review IDs has been created to contact the unhappy users and possibly prevent customer churn.
- Another list with review IDs was created containing the recent questions discovered in the neutral class reviews in order to address any concerns or provide the necessary information to the users.

Visualizing the data.

The following insights were derived based on the visualizations (Appendix Part 1).

- The overall sentiment of the product based on the rating is mostly positive.
- The best overall opinion of the product was expressed in 2016 with a slow decline of the positive sentiment afterwards.
- The sentiment slightly increased for March in Q1, 2021.
- All of the aspects of TuneIn Radio (Content, Subscription, Technical features, and Streaming) have an overall mostly positive sentiment.
- Content is the aspect of TuneIn Radio that users are most happy with.
- Subscription is the aspect of TuneIn Radio that has the least positive sentiment.
- In comparison to 2018, there has been progress in addressing the concerns of users by replying to their reviews.
- UK is the most frequent word in the negative reviews for Q1, 2021 which indicates potential problems with the product offered for the UK-based users.
- MSNBC and BBC are also frequently mentioned in the negative reviews which could indicate that checking the streaming quality for these particular channels might be needed.
- The frequent use of “30 seconds” in negative reviews could potentially indicate a problem 30 seconds into the start of streaming.

- The frequently used phrase "7 days" might indicate that users are unhappy with the 7-day trial.

Data Preparation:

Missing values	removed for the "content" variable when working on the topic modeling and a classification model creation.
Engineered time variables	Variables "year", "month" and "day" were engineered from the variable "at".
New categorical (text) variables	"lemmatize" was created after preprocessing the variable "content" for the topic modeling with LDA. "clean_text" was created after preprocessing the variable "content" for the classification modeling.

Imbalanced data handling techniques (tested with Logistic Regression).

RandonOverSampler	SMOTE	Undersampled	Imbalanced
accuracy: 84.09% (+/- 0.78%) precision: 62.34% (+/- 0.47%) recall: 68.57% (+/- 0.51%) f1 score: 64.14% (+/- 0.59%)	accuracy: 82.48% (+/- 0.91%) precision: 60.71% (+/- 0.46%) recall: 67.12% (+/- 0.78%) f1 score: 62.52% (+/- 0.77%)	accuracy: 70.04% (+/- 0.47%) precision: 69.58% (+/- 0.46%) recall: 70.04% (+/- 0.47%) f1 score: 69.71% (+/- 0.47%)	accuracy: 89.73% (+/- 0.09%) precision: 69.19% (+/- 0.62%) recall: 60.55% (+/- 0.28%) f1 score: 60.92% (+/- 0.33%)

Even though the accuracies are higher for the oversampled data samples and the imbalanced data, their macro averaged F1 scores are smaller than that of the undersampled data due to the fact that they struggle with the neutral class identification.

Another important metric is the F1 score of the neutral class for different types of data samples:

Imbalanced	Oversampled with SMOTE	Oversampled with RandomeOverSampler	Undersampled
F1 for the neutral is ~10%	*F1 for the neutral is ~23%*	*F1 for the neutral is ~25%*	*F1 for the neutral is ~57%*

More information in Appendix Part3

Note: Neutral class results were extracted from the cross-validated logistic regression model. The final choice for the data used for the classification problem was the undersampled data due to the decision to use the data where the identification of the neutral class is also possible to learn at an acceptable level.

Vectorizing data

CountVectorizer	a choice for the topic modeling with LDA.
TfidfVectorizer	Logistic Regression, Support Vector Machine, XGBoost, Random Forest

Global Vectors embeddings (GloVe)	used for the CNN and bidirectional calssifiersLSTM neural network classifier with Keras.

Modeling and Evaluation

Topic modeling with LDA:

```
from sklearn.decomposition import LatentDirichletAllocation
import collections
lda_model_2 = LatentDirichletAllocation(n_components = 4, # number of topics
                                        random_state = 10,
                                        evaluate_every = -1,
                                        max_iter=60,# iterations
                                        n_jobs = -1,
                                       )

lda_output_2 = lda_model_2.fit_transform(vectorised)
```

The model with the clearly identified topics.

Number of topics = 4

Identified topics mapped to the product aspects:

- **Streaming** (Topic 0). The helpful keywords: *stream, listen, buffer, connection, play, stop*.
- **Subscription** (Topic 1). The useful keywords: *service, subscription, trial, year, month, money, sport, premium, account, charge, refund*
- **Content** (Topic 2). The informative keywords: *music, world, channel, country, variety, news, selection, podcast, choice, song, talk, access*.
- **Technical aspects** (Topic 3). The helpful keywords: *features, screen, battery, button, notification, alarm, background*

Classification models (all of the models can be found in Appendix Part 2)

```
              negative    neutral     positive
precision: [0.69274538 0.61061453 0.78090692]
recall:    [0.73123123 0.54704705 0.81881882]
f1 score:  [0.71146822 0.57708553 0.79941363]
------------------------------------------------
              negative    neutral     positive
precision: [0.67945076 0.60652174 0.81341822]
recall:    [0.71821822 0.55855856 0.83133133]
f1 score:  [0.69829684 0.58155289 0.82227723]
------------------------------------------------
              negative    neutral     positive
precision: [0.70458984 0.62042175 0.79104478]
recall:    [0.72222222 0.55955956 0.84884885]
f1 score:  [0.71329708 0.58842105 0.81892805]
------------------------------------------------
              negative    neutral     positive
precision: [0.6979217  0.60964912 0.78772013]
recall:    [0.72272272 0.55655656 0.82832833]
f1 score:  [0.71010573 0.5818943  0.80751403]
------------------------------------------------
              negative    neutral     positive
precision: [0.67655109 0.61007357 0.79312039]
recall:    [0.74224224 0.53953954 0.80780781]
f1 score:  [0.70787589 0.57264276 0.80039673]
------------------------------------------------
accuracy: 70.22% (+/- 0.46%)
precision: 69.83% (+/- 0.42%)
recall: 70.22% (+/- 0.46%)
f1 score: 69.94% (+/- 0.45%)
```

The best performing model (*Support Vector Machines with the decision function shape set to "ovo"*).

The evaluation of several tuned multiclass classification models (logistic regression, CNN, bidirectional LSTM, Random Forest, XGBoost, Support Vector Machines) showed that the Support Vector Machines was the best performing model. The metrics of F1, accuracy, recall, and precision are all the highest for this model (at about 70%). All of the models tended to struggle with the neutral class identification while the Support Vector Machines identified the negative and neutral reviews the most correctly. This is important because these two classes of reviews are prioritized for the replies by customer support specialists.

	algorithm	percentage_correct
0	svm	80
1	human	80
2	logistic regression	67
3	random forest	67
4	xgboost	60
5	convnet	60
6	LSTM	53
7	Vader	33

I decided to test the performance and see how the predictions made by all of the algorithms differ from the ones that I make. I tested my own sentiment identification (column human), the built models, and the Vader model on the small data set of 15 balanced observations. It looks like our best model classified the sentiment of these observations as well as I did. Note: The sentiment analysis is a complex problem even for humans. It could be very subjective. Humans tend to agree overall on about 80% of the sentiment (Barba, P.).

Questions that have been answered with the help of the sentiment analysis with topic modeling:

Questions	**Answers**
1.What is the overall sentiment analysis behind the reviews of TuneIn Radio?	The users' overall opinion of the brand is positive.
2. Has the users' opinion of the product improved over time?	The users' opinion of the product has varied over time. The most positive sentiment was expressed in 2016 with a slow but not drastic decrease afterwards.
3. What is the opinion of the product based on the recent data: Q1 2021?	The sentiment is mostly positive with March 2021 having the highest percentage of positive reviews.
4. What aspects of the product are seen as most negative?	The subscription service is the least liked aspect of TuneIn Radio.
5. What aspects of the product are liked by the users the most?	The content aspect of TuneIn Radio is the aspect that meets the listeners' expectation the most.
6. Can we predict the product sentiment from different types of	The best performing model (Support Vector Machines) with the both the F1 score and the accuracy of about 70% was able to classify the text data into the three classes at an acceptable level since on

sources with opinions about TuneIn Radio?	average even humans tend to agree on the sentiment classification only about 80% of the time (Barba, P.).
7. Is the company addressing the concerns of the users?	The overall responses to the negative reviews have increased compared to the lower number in 2018
8. Are there particular opinions that many of the customers can relate to? Have the concerns been addressed?	Most of the reviews that are considered relevant by other customers have been addressed. A list of 5 relevant unanswered review IDs has been created to help prevent customer churn.
9. Has the data proved to be useful for the proposed analysis?	Even though the data set is imbalanced, I was able to answer all of the questions originally asked and create a classifier with an F1, accuracy, precision, and recall scores of about 70%.
10. What are the main challenges of the project?	The main challenges of the project were the imbalanced data set and creating a model that would predict the complex neutral class well.
11. Did the methods prove to be effective?	The EDA, visualizations, data preprocessing, topic modeling with LDA and a classification model contributed to the successful completion of the project.
12. Are there any **recommendations** that could be made?	The recommendations for the support team: address all of the problems and concerns identified with the help of EDA and visualization analysis. Respond and answer the questions of the users whose reviews were selected for the customer support response strategy. Suggestion for the improvement of the project: -Explore the concept of mixed sentiment in the neutral reviews (sentences with the conjunction “but”) to possibly extract separate kinds of sentiment contained in one review. This should allow to keep the neutral reviews with true neutral sentiment and avoid a mix

	of positive and negative sentiment in the neutral class. The accuracy for the neutral class could be improved by using this strategy.

Challenges and limitations of the project

One of the main challenges of this project was handling the imbalanced dataset which resulted into a smaller size sample for the final classification modeling stage. Another challenge was getting an acceptable classification results for the neutral class due to the complexity of this class.

The main limitations of the project included the complexity of the language such as interpretation and handling of the following linguistic phenomena: sarcasm, irony, hyperbole, and ambiguity that can decrease the accuracy and performance for any text data analysis.

Conclusion

Sentiment analysis is a complex problem. Partially due to the complexity of the language structure and semantics, and also due to the fact that opinion mining is very subjective. Humans tend to agree on the sentiment only about 80% of the time (Barba.P.). Taking this into consideration, I can say that the best performing model built in this project with the metrics of accuracy, F1, recall and precision at about 70% is a satisfactory outcome. This result implies that 70% of the time the expected predictions of the sentiment would be correct.

Topic modeling with Latent Dirichlet Allocation used in this analysis resulted in creating 4 clear aspects of TuneIn Radio application: content, subscription, technical features, and streaming which helped expand the analysis by deriving valuable insights about each of these topics.

The project allowed to create an overall picture of the sentiment of TuneIn Radio (mostly positive), as well as the sentiment of its main features and aspects, analyze the performance of the customer support (mostly high), and identify the review IDs and questions that still need to be addressed to prevent customer churn. These findings are useful and could serve as the base of a new business or customer support strategy for product improvement.

References

1. TuneIn Radio. Retrieved from https://play.google.com/store/apps/details?hl=en&id=tunein.player

2. Google-Play-Scraper. Retrieved from https://pypi.org/project/google-play-scraper/

3. Alotaibi, Sh. 2018. Customer Satisfaction Measurement Using Sentiment Analysis. Retrieved from https://www.researchgate.net/publication/323536432_Customer_Satisfaction_Measurement_using_Sentiment_Analysis

4. Oheix, J. 2018. Detecting Bad Customer Reviews with NLP. Retrieved from https://towardsdatascience.com/detecting-bad-customer-reviews-with-nlp-d8b36134dc7e

5. Galeshchuk, S. 2019. Working with Twitter Data in Python. Retrieved from https://medium.com/analytics-vidhya/working-with-twitter-data-b0aa5419532

6. Arshad, S. 2019. Sentiment Analysis/ Text Classification Using CNN (Convolutional Neural Network). Retrieved from https://towardsdatascience.com/cnn-sentiment-analysis-1d16b7c5a0e7

7. Timaraju, A. and Khanna, V. Sentiment Analysis on Movie Reviews using Recursive and Recurrent Neural NetworkArchitectures. Retrieved from https://cs224d.stanford.edu/reports/TimmarajuAditya.pdf

8. V.S. Anoop and S. Asharaf. 2020. Aspect-Oriented Sentiment Analysis: A Topic Modeling-Powered Approach. Extracted from https://www.degruyter.com/document/doi/10.1515/jisys-2018-0299/html

9. Rana, T.A. and Cheah, Y. Aspect extraction in sentiment analysis: comparative analysis and survey. Retrieved from https://www.researchgate.net/publication/296567757_Aspect_extraction_in_sentiment_analysis_comparative_analysis_and_survey

10. Zhang, M. Employer Reviews using Topic Modeling. Retrieved from https://mickzhang.com/employer-reviews-using-topic-modeling/

11. Williams, L. 2020. Sentiment Analysis: Aspect-based opinion mining. Retrieved from https://towardsdatascience.com/%EF%B8%8F-sentiment-analysis-aspect-based-opinion-mining-72a75e8c8a6d

12. Pennington, J., Socher, R., Manning, Ch. 2014. GloVe: Global Vectors for Word Representation. Retrieved from https://nlp.stanford.edu/pubs/glove.pdf

13. Bhattacharyya, S. 2020. Classification Using Long-Short Memory and GloVe. Retrieved from https://medium.com/analytics-vidhya/classification-using-long-short-term-memory-glove-global-vectors-for-word-representation-254d02d5e158

14. Kim, R. 2018. Yet Another Twitter Sentiment Analysis Part 1 – Tackling Class Imbalance. Retrieved from https://towardsdatascience.com/yet-another-twitter-sentiment-analysis-part-1-tackling-class-imbalance-4d7a7f717d44

15. Barba, P. Sentiment Accuracy: Explaining the baseline. How to test it. https://www.lexalytics.com/lexablog/sentiment-accuracy-baseline-testing

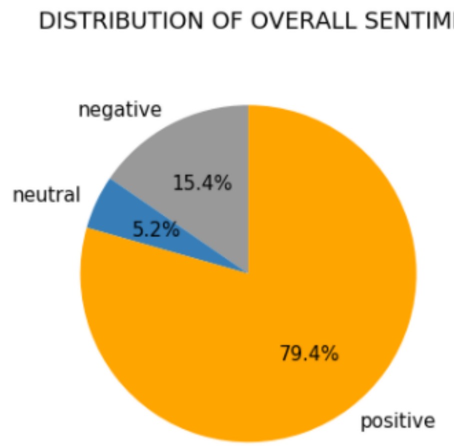
Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews

Natallia Casey
Bellevue University

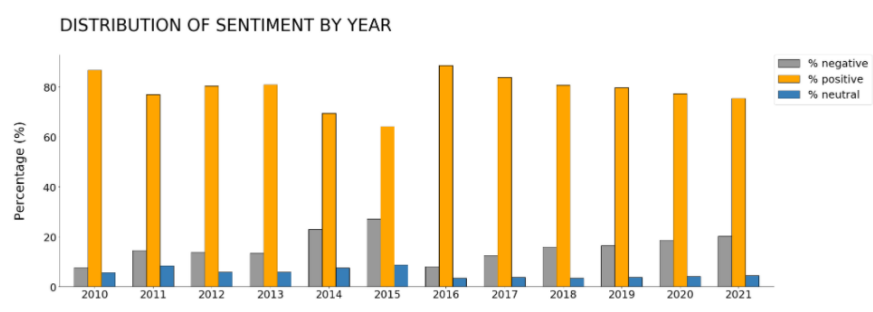
Appendix. Part 1

Visualization	Image

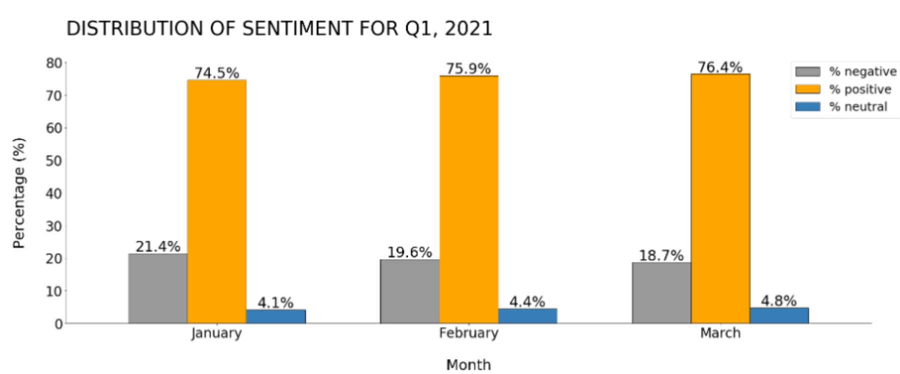
Pie chart



Grouped bar graph



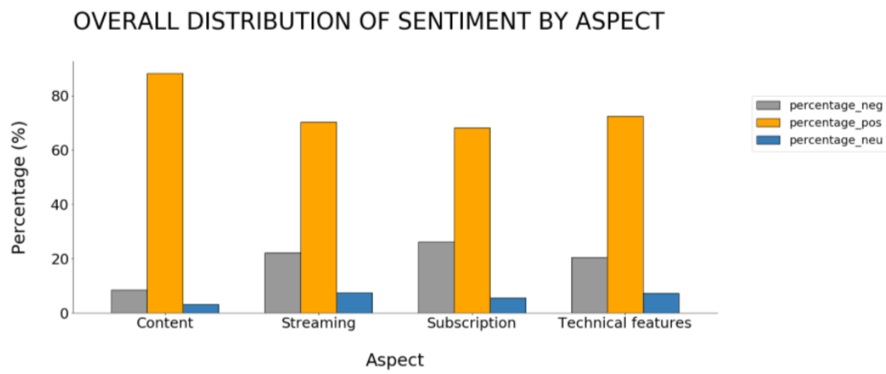
Grouped bar graph



Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews
Natallia Casey
Bellevue University

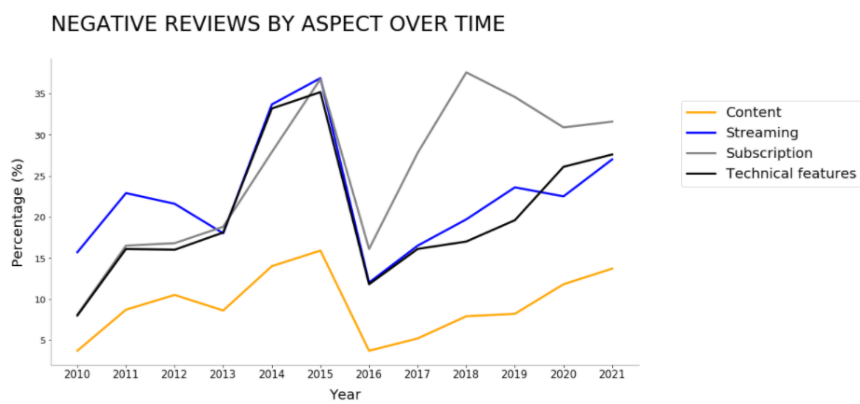
Grouped bar graph

aspects



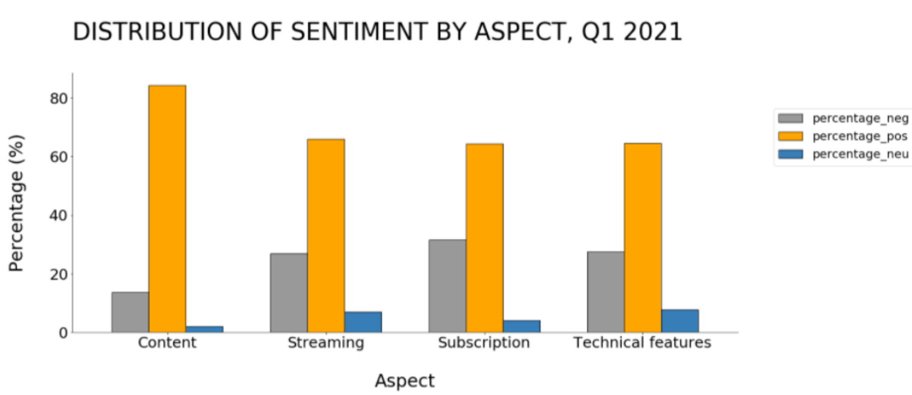
Multiple lines graph

Aspects



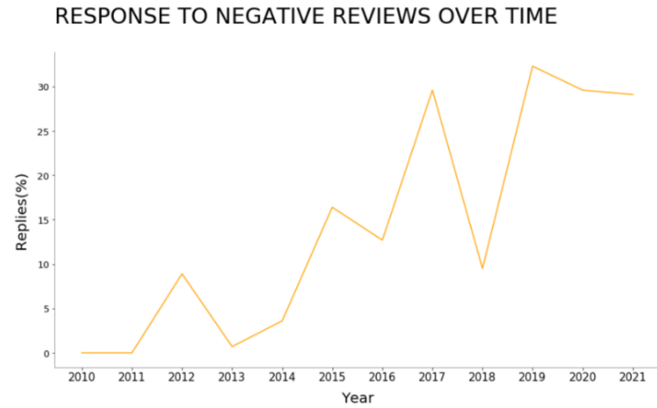
Grouped bar graph

aspects

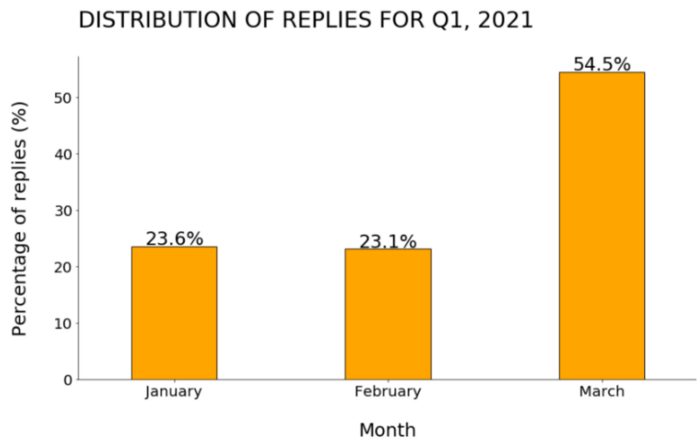


Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews
Natallia Casey
Bellevue University

Line graph



Bar graph



Word cloud

Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews
Natallia Casey
Bellevue University

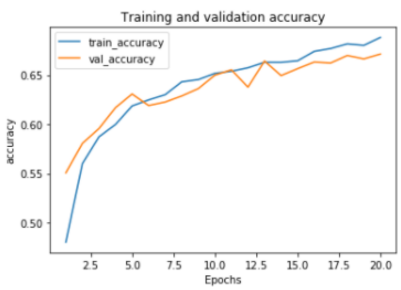
Appendix. Part2. Results of the models with undersampled data

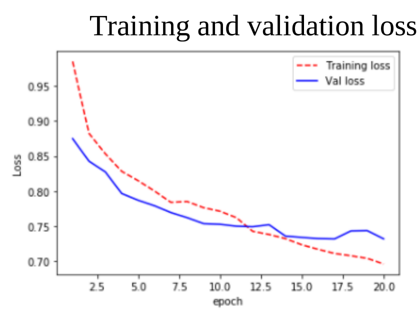
Bidirectional LSTM

Test set accuracy: 0.672

Validation set accuracy: 0.675

	precision	recall	f1-score	support
0	0.74	0.58	0.65	2500
1	0.60	0.44	0.51	2443
2	0.76	0.82	0.79	2550
micro avg	0.71	0.62	0.66	7493
macro avg	0.70	0.61	0.65	7493
weighted avg	0.70	0.62	0.65	7493
samples avg	0.62	0.62	0.62	7493



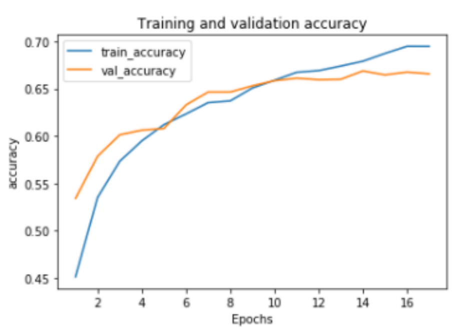


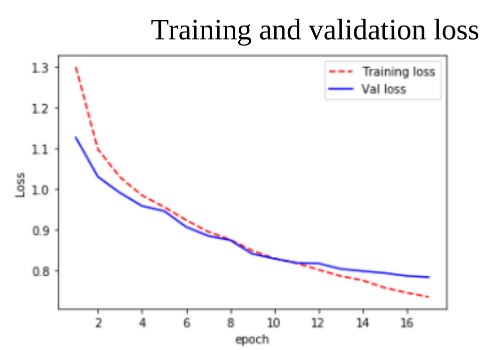
Convolutional neural network

Test set accuracy: 0.665

Validation set accuracy: 0.670

	precision	recall	f1-score	support
0	0.73	0.56	0.64	2500
1	0.58	0.51	0.55	2443
2	0.81	0.74	0.77	2550
micro avg	0.71	0.61	0.65	7493
macro avg	0.71	0.60	0.65	7493
weighted avg	0.71	0.61	0.65	7493
samples avg	0.61	0.61	0.61	7493





Logistic Regression Classifier

```
             negative    neutral     positive
precision: [0.69506941 0.6059584  0.77281279]
recall:    [0.72672673 0.53953954 0.82232232]
f1 score:  [0.71054563 0.5708234  0.79679922]
------------------------------------------------
             negative    neutral     positive
precision: [0.68218915 0.60796074 0.80163698]
recall:    [0.71121121 0.55805806 0.83333333]
f1 score:  [0.69639794 0.58194154 0.81717791]
------------------------------------------------
             negative    neutral     positive
precision: [0.70686275 0.62359551 0.78196872]
recall:    [0.72172172 0.55555556 0.85085085]
f1 score:  [0.71421496 0.58761249 0.81495686]
------------------------------------------------
             negative    neutral     positive
precision: [0.69990272 0.61033908 0.77559607]
recall:    [0.72022022 0.54954955 0.83033033]
f1 score:  [0.70991613 0.57835133 0.80203046]
------------------------------------------------
             negative    neutral     positive
precision: [0.67877095 0.60906516 0.78375781]
recall:    [0.72972973 0.53803804 0.81631632]
f1 score:  [0.70332851 0.57135264 0.79970581]
------------------------------------------------
accuracy: 70.02% (+/- 0.51%)
precision: 69.57% (+/- 0.49%)
recall: 70.02% (+/- 0.51%)
f1 score: 69.70% (+/- 0.50%)
```

Random Forest

```
             negative    neutral     positive
precision: [0.64451115 0.6146978  0.74478694]
recall:    [0.75225225 0.44794795 0.82232232]
f1 score:  [0.69422633 0.51823972 0.78163654]
------------------------------------------------
             negative    neutral     positive
precision: [0.63169257 0.59708416 0.76625173]
recall:    [0.73223223 0.45095095 0.83183183]
f1 score:  [0.67825684 0.51382948 0.79769618]
------------------------------------------------
             negative    neutral     positive
precision: [0.66104784 0.61357702 0.73665637]
recall:    [0.72622623 0.47047047 0.83583584]
f1 score:  [0.69210589 0.5325779  0.78311841]
------------------------------------------------
             negative    neutral     positive
precision: [0.64064489 0.61821975 0.74670305]
recall:    [0.75575576 0.44494494 0.82182182]
f1 score:  [0.6934558  0.51746217 0.78246366]
------------------------------------------------
             negative    neutral     positive
precision: [0.63540346 0.6        0.74687065]
recall:    [0.75275275 0.44144144 0.80630631]
f1 score:  [0.68911798 0.50865052 0.77545126]
------------------------------------------------
accuracy: 67.29% (+/- 0.35%)
precision: 66.65% (+/- 0.34%)
recall: 67.29% (+/- 0.35%)
f1 score: 66.39% (+/- 0.37%)
```

XGBoost

```
             negative    neutral     positive
precision: [0.69832402 0.61599031 0.72030329]
recall:    [0.68818819 0.50900901 0.85585586]
f1 score:  [0.69321906 0.55741299 0.78225069]
------------------------------------------------
             negative    neutral     positive
precision: [0.67103984 0.61907692 0.7425357 ]
recall:    [0.69119119 0.5035035  0.85885886]
f1 score:  [0.68096647 0.55534088 0.7964725 ]
------------------------------------------------
             negative    neutral     positive
precision: [0.69210664 0.60493827 0.72394958]
recall:    [0.66266266 0.51501502 0.86236236]
f1 score:  [0.67706469 0.55636659 0.78711741]
------------------------------------------------
             negative    neutral     positive
precision: [0.68392505 0.60733614 0.73122015]
recall:    [0.69419419 0.50550551 0.84284284]
f1 score:  [0.68902136 0.55176181 0.7830737 ]
------------------------------------------------
             negative    neutral     positive
precision: [0.68117359 0.61236987 0.72711572]
recall:    [0.6971972  0.5005005  0.84284284]
f1 score:  [0.68909226 0.55081245 0.78071395]
------------------------------------------------
accuracy: 68.20% (+/- 0.20%)
precision: 67.54% (+/- 0.20%)
recall: 68.20% (+/- 0.20%)
f1 score: 67.54% (+/- 0.19%)
```

<table>
<tr>
<td>Support Vector Machines</td>
<td><pre>
                  negative   neutral   positive
precision: [0.69274538 0.61061453 0.78090692]
recall: [0.73123123 0.54704705 0.81881882]
f1 score: [0.71146822 0.57708553 0.79941363]
----
                  negative   neutral   positive
precision: [0.67945076 0.60652174 0.81341822]
recall: [0.71821822 0.55855856 0.83133133]
f1 score: [0.69829684 0.58155289 0.82227723]
----
                  negative   neutral   positive
precision: [0.70458984 0.62042175 0.79104478]
recall: [0.72222222 0.55955956 0.84884885]
f1 score: [0.71329708 0.58842105 0.81892805]
----
                  negative   neutral   positive
precision: [0.6979217 0.60964912 0.78772013]
recall: [0.72272272 0.55655656 0.82832833]
f1 score: [0.71010573 0.5818943 0.80751403]
----
                  negative   neutral   positive
precision: [0.67655109 0.61007357 0.79312039]
recall: [0.74224224 0.53953954 0.80780781]
f1 score: [0.70787589 0.57264276 0.80039673]
----
accuracy: 70.22% (+/- 0.46%)
precision: 69.83% (+/- 0.42%)
recall: 70.22% (+/- 0.46%)
f1 score: 69.94% (+/- 0.45%)
</pre></td>
</tr>
</table>

Part 3. Sample performance tested with the Logistic Regression.

Imbalanced

```
                  negative    neutral     positive
precision: [0.76249592 0.36094675 0.92730552]
recall:    [0.7811245  0.06106106 0.97311407]
f1 score:  [0.7716978  0.10445205 0.9496577 ]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.76741536 0.40895522 0.92922254]
recall:    [0.78898929 0.06856857 0.97454973]
f1 score:  [0.77805281 0.11744535 0.95134653]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.76490996 0.37142857 0.92707881]
recall:    [0.77476573 0.05855856 0.9756591 ]
f1 score:  [0.7698063  0.10116732 0.95074878]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.76269531 0.39057239 0.92806314]
recall:    [0.78413655 0.05805806 0.97448447]
f1 score:  [0.77326733 0.10108932 0.95070748]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.76187357 0.38888889 0.92680051]
recall:    [0.77857741 0.05605606 0.97454973]
f1 score:  [0.77013492 0.09798775 0.95007555]
-------------------------------------------------
accuracy: 89.73% (+/- 0.09%)
precision: 69.19% (+/- 0.62%)
recall: 60.55% (+/- 0.28%)
f1 score: 60.92% (+/- 0.33%)
```

Undersampled

```
                  negative    neutral     positive
precision: [0.6938873  0.60690045 0.77251407]
recall:    [0.72722723 0.53703704 0.82432432]
f1 score:  [0.71016618 0.56983537 0.79757869]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.68413462 0.60926431 0.8003848 ]
recall:    [0.71221221 0.55955956 0.83283283]
f1 score:  [0.69789112 0.58335507 0.81628649]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.70562347 0.62247191 0.78192716]
recall:    [0.72222222 0.55455455 0.84884885]
f1 score:  [0.71382637 0.58655373 0.81401488]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.70048544 0.61080178 0.77595884]
recall:    [0.72222222 0.54904905 0.83033033]
f1 score:  [0.71118778 0.5782815  0.80222437]
-------------------------------------------------
                  negative    neutral     positive
precision: [0.67840593 0.61082621 0.78375781]
recall:    [0.73273273 0.53653654 0.81631632]
f1 score:  [0.70452358 0.57127631 0.79970581]
-------------------------------------------------
accuracy: 70.04% (+/- 0.47%)
precision: 69.58% (+/- 0.46%)
recall: 70.04% (+/- 0.47%)
f1 score: 69.71% (+/- 0.47%)
```

Oversampled (SMOTE)

```
                negative     neutral      positive
precision: [0.66961231 0.18380213 0.96324867]
recall:    [0.78614458 0.37937938 0.86374315]
f1 score:  [0.72321429 0.24763149 0.91078617]
--------------------------------------------------
                negative     neutral      positive
precision: [0.69327302 0.1828167  0.96303122]
recall:    [0.79501339 0.37487487 0.8695184 ]
f1 score:  [0.74066568 0.24577523 0.91388889]
--------------------------------------------------
                negative     neutral      positive
precision: [0.69544693 0.14447925 0.95984711]
recall:    [0.76422356 0.36936937 0.84393761]
f1 score:  [0.72821494 0.20771179 0.89816824]
--------------------------------------------------
                negative     neutral      positive
precision: [0.68854645 0.18387175 0.96105347]
recall:    [0.77761044 0.37887888 0.87036674]
f1 score:  [0.73037328 0.2475879  0.91346483]
--------------------------------------------------
                negative     neutral      positive
precision: [0.69420633 0.16434071 0.959096   ]
recall:    [0.78610879 0.34184184 0.86681023]
f1 score:  [0.73730476 0.22196945 0.91062094]
--------------------------------------------------
accuracy: 82.48% (+/- 0.91%)
precision: 60.71% (+/- 0.46%)
recall: 67.12% (+/- 0.78%)
f1 score: 62.52% (+/- 0.77%)
```

Oversampled (RandomOverSapmler)

```
                negative     neutral      positive
precision: [0.70787377 0.21088435 0.96952889]
recall:    [0.77325971 0.41891892 0.88971548]
f1 score:  [0.73912348 0.28054299 0.92790907]
--------------------------------------------------
                negative     neutral      positive
precision: [0.71790152 0.19441748 0.96883822]
recall:    [0.78313253 0.4009009  0.8845928 ]
f1 score:  [0.74909964 0.26185028 0.92480087]
--------------------------------------------------
                negative     neutral      positive
precision: [0.70498026 0.19777306 0.96709595]
recall:    [0.7769411  0.37337337 0.89186896]
f1 score:  [0.7392135  0.25857886 0.92796035]
--------------------------------------------------
                negative     neutral      positive
precision: [0.70511841 0.17459643 0.96733531]
recall:    [0.77225569 0.41141141 0.86384103]
f1 score:  [0.73716157 0.24515359 0.91266353]
--------------------------------------------------
                negative     neutral      positive
precision: [0.71555284 0.18277512 0.96706329]
recall:    [0.78309623 0.38238238 0.88041634]
f1 score:  [0.74780246 0.24732923 0.92170794]
--------------------------------------------------
accuracy: 84.09% (+/- 0.78%)
precision: 62.34% (+/- 0.47%)
recall: 68.57% (+/- 0.51%)
f1 score: 64.14% (+/- 0.59%)
```

Part 4. Lists of review IDs identified using EDA for customer churn prevention strategy:

Recent most relevant negative reviews that haven't been addressed:

	reviewId	thumbsUpCount
0	gp:AOqpTOHswsCtrKwtl_RIJ3C02hMKPUcFfS1K3lypoJ0VBwnJRmvkx0GGNsJkL7EsclHgpwqJyyiLD5a9k0SATg	15
1	gp:AOqpTOH70sN3rOvn7Ise6C0aHR15hFWLcUQ0GOAWPK60UllnG8mHiMbFgGbI2762Ytatu31i_nA5oC2nArwZ6g	6
2	gp:AOqpTOF5iCJA0oRV4FRQdm4EmglEaoCiiCAd69bapqg2YSS2f-ZmfAO5OhkNyUWbcqBGWWz4NuUcJCR673wvyQ	6
3	gp:AOqpTOHkWdjZ4FF371_x1BHo1uAOgZjefWAl7utWG1W5iatjt8djy8kWHopQK0_LKcK9-9WkXyf35bVUcCsYiw	4
4	gp:AOqpTOEePgZwriSFnu07bhMso94ME-AU9nAj7lypZ4RWJnh7cwa0vc12CysNAYA1embg_aieUcql-a6GrSDToQ	2
5	gp:AOqpTOHoXfUcB9js_T2BmknCNimudg2Tz6qS1gEESZGZgfrrlufpd4cqnGHx63YvN4u2NAnn26BpDILTCMrT6Q	2
6	gp:AOqpTOHWjsHZUn2q3sfaSM19WinuWMqw9BZ1Oy9mW70GCV6KKH_5-eODMjlTQwqPhUZ-Pelt4cYc1HtJolHH7Q	2

Recent reviews that need answers:

reviewId	content
gp:AOqpTOECtNu9NieZFwsucspCWYtZv0Ffh_mayv8SK81...	I JUST, JUST THAT IS, UPDATED THIS APP, MORE T...
gp:AOqpTOE-aJEAzDnZXszKtMZrutZkDZKQp_P24BjkMdg...	What happened to the fast forward & rewind but...
gp:AOqpTOGAyuZJ6hoznn845ERWSz_aFnz3J1M6sY6FJ3a...	Too expensive for just radio and boardcasting ...
gp:AOqpTOG9WA78HpgjBdz-HrKrdACkEnVc4gZRIBjotay...	Turns itself on in the car even when I don't w...
gp:AOqpTOEZNxpSaolxKGE0LG7qAYtz10w1SMxiluhF6kB...	It's OK but why the random streaming? - I don'...
gp:AOqpTOH1xZGfwgGHYYFkW-3iyXdYa_u74RQCrpC106c...	This app is very useful for podcasts and live ...
gp:AOqpTOGZSozRUDZOk8QYfL-NBYBdXDQF4FCFVQXOr42...	Why does this app auto start when Bluetooth is...
gp:AOqpTOG4qr8pWg5-wPZw8dxpVtq4c6CGZO4In5OCKCf...	2 major annoyances: - For some reason it promp...
gp:AOqpTOHPZTJUaR2W5rdcJWHkpWUNB4VZ-HxskxQke80...	What happened? Bookmarks disappeared and now I...
gp:AOqpTOHiatLYnRvpdxwp4tOsi-ffm9HQDyBXJJlOAo3...	I'm in UK and when I try to listen to my Local...

Sentiment Analysis with Topic Modeling of TuneIn Radio Reviews
Natallia Casey
Bellevue University

Codebook

reviewid – unique identifier of a review. Data type: object

content – the text of the review. Data type: object

score – a rating of a review from 1 to 5. Data type: int64

thembsUpCount – the number of thumbs-up that have been given to the review. Data type: int64

reviewCreatedVersion – the reviewed version of the application. Data type: object

at – the date of the review. Data type: datetime64

replyContent – the text of the reply. Data type: object

repliedAt – the date of the reply. Data type: object