Health Insurance

Robert Hatcher | Mohit Vaghela | Shefali Vinchurkar | Prashant Kumar | Akhila Puthran

Sentiment Analysis **12/12/19** **MSA 8040**

Background

Choosing health insurance is one of the most important investments you'll make for your family. It provides financial protection in case you have a serious accident or illness. According to the Urban Institute uninsured people on average receive less medical care and have worse health conditions compared to those who are insured. Choosing which company to provide service for you and your family can be overwhelming.

We can make this decision process easier by examining thousands or reviews from existing customers. Their qualitative data can give us insight about how well these companies provide coverage and service to their customers. For this study we compared two of the most popular insurance companies, Kaiser Permanente and United HealthCare.

Objective

By analyzing and categorizing opinions from thousands of reviews, we can determine which company is providing the better service. Our analysis in this paper concluded two main objectives:

1. Comparative analysis of the sentiments of the customers who have purchased Kaiser Permanente or United Healthcare insurance.
2. Recommendation for the Insurance based on following categories:
    a. Services (Refunds, Customer Services)
    b. Coverage plan (Cost, Deductibles, Family Plans)

Challenges:

- **Inconsistency with supervised learning -**

  Supervised learning such as rating (scale 1-5-star rating) for example wasn't a true measure for analysis because the meaning differed between users and wasn't consistency with the sentiment for a user review. Example, one user gave United Healthcare a 3 out of 5-star rating which indicated that the service was average. However, in their review the user said, "Customer service is horrible" and used other adjectives such as "lied" and "deceived." Their review read more like a 1 out of 5 instead of a 3.

['Customer Service is horrible. I have had 5 different answers on the same issue from 5 different service people. I have had misleading/wrong/inaccurate information. Train your staff! Bad experiences with coverage - have been lied to and deceived.']	['3.0']

- **Precise measurement of sentiment analysis -**

  The initial library we used to find sentiment analysis was TextBlob. However, when using TextBlob it gave a general polarity score to measure the sentiment of a review without any detail or evidence. For this reason, we were able to find a more precise library to conclude our sentiment analysis with VADER.

  Example: the phrase "I don't like food"

  TextBlob – produced a polarity and subjectivity score of 0.0 meaning this phrase is completely neutral.

  VADER – broke down the polarity score into a 51.3% negative and 48.7% neutral. Giving us a more precise measurement of sentiment, we chose to use the VADER library for the rest of the project.

```
analyzer = SentimentIntensityAnalyzer()
vs = analyzer.polarity_scores('I dont like food')
vs
```

```
{'neg': 0.513, 'neu': 0.487, 'pos': 0.0, 'compound': -0.2755}
```

```
blob = TextBlob("I dont like food")
blob.sentiment
```

```
Sentiment(polarity=0.0, subjectivity=0.0)
```

- **Customized Stop Words –**

  The English stop words provided by the nltk package also includes some words like doesn't, couldn't, won't, etc. Since these words are important for analyzing the sentiment of the sentence, the absence of these words would change the polarity of the sentence thereby giving us a false result. Hence, we created a list of all the negative words and passed this list through the original stop words list thereby giving us the customized list. (p as seen below)

```
k = stopwords.words('english')
Stop = ["no", "nor", "not", "don", "don't", "couldn't", "couldn", "didn't", "didn", "doesn't", "doesn", "hadn't", "hadn",
p = []
p = [x for x in k if x not in Stop]
```

Data Scraping:

Using Selenium package from Python we were able to scrape 1,139 reviews for Kaiser Permanente and 1,671 reviews for United HealthCare from consumeraffaris.com & bestcompany.com. The attributes we collected included:

1. Names
2. Reviews
3. Ratings
4. Location

Cleaning Data:

United HealthCare was the only dataset that contained null values for customer reviews. We performed following operations:

- In total there were 39 null values which were removed.
- Converting the words to lower case
- Removal of redundant words, punctuations and stop words
- Tokenization – This step was needed to help us understand the keywords in the data which would be helpful in conducting the sentiment analysis.
- Removal of stop words

Sentiment Analysis:

1. The reviews that were obtained after scrapping needs to be cleaned.
2. All the text is converted to lowercase using string functions
3. All the words were tokenized.
4. Create a list of stop words that needs to be eliminated and keep words such as "no", "nor", "not", "don".
5. Eliminate the stop words from the review list.
6. The reviews were further lemmatized to obtain all the words in their present tense.
7. We further cleaned the reviews by keeping only alphanumeric characters in the review and removing the rest of the characters.
8. Next step in to create a bag of words, wherein, the frequency count of all the reviews were obtained.
9. The frequency count of words was further rearranged in descending order.
10. Then using the bag of words, as well as, words from the corpus we developed the words for the parameters such as coverage and services.
11. The parameters were then used to find the related nouns in the reviews, which were used to compare the reviews between Kaiser Permanente and United healthcare.

Findings

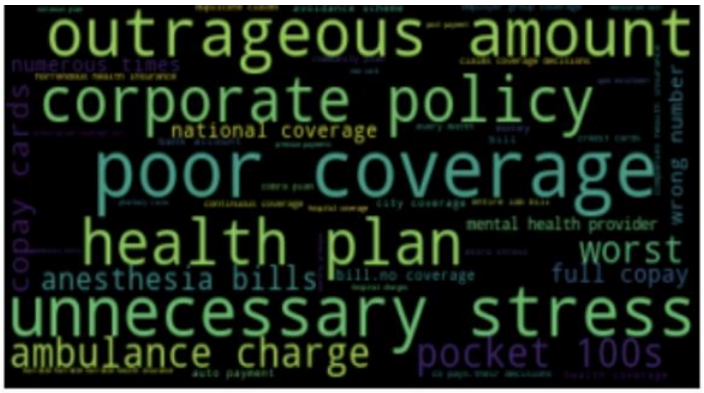
Insurance	Service	Coverage Plans
Kaiser Permanente	Good	Average
United HealthCare	Bad	Good

Comparison based on Service

Kaiser Permanente

United Healthcare

Comparison based on Coverage



Kaiser Permanente

United Healthcare

The WordCloud above highlights the strengthens between the two insurance companies. Our recommendations are based on what the customer values the most out of a insurance company. Judging from the reviews if the customer value service as their top priority Kaiser Permanente would be the better option. If they're looking for the company with the better coverage plan United Healthcare would be the better option.

Conclusion:

The unsupervised sentimental analysis of the reviews reflects that UnitedHealthcare Insurance and Kaiser Permanente Insurance are very close in quality and have the same overall sentiments.

UnitedHealthcare scores better than Kaiser Permanente across: Cost, and Coverage. Whereas, Kaiser Permanente scores better on Customer Experience and Services. Our recommendations are based on what the customer values the most out of a insurance company. Judging from the reviews, if the customer value service as their top priority Kaiser Permanente would be the better option. If they're looking for the company with the better coverage plan, United Healthcare would be the better option.

These results are based on only certain parameters which are Service, Family plans, Cost Incurred, Deductible etc...(as this was an unsupervised machine learning analysis, word-bags of these parameters were only taken)so the results may vary if other aspects like location, time period of the insurance etc... are also considered.

Future Improvements and Implementations:
As a part of future improvement, one can consider machine learning model with additional parameters. Also, if there are more reviews available, the word-bags in the model can be changed accordingly.

Currently, we have prioritized only the good parts of the reviews but as a part of future implementation, we can consider the bad sentiments also from the reviews. This part of the implementation would be done from the Insurance firms' perspective. Currently, we are giving recommendations to the users who have opted for any of the two insurance but, considering negative sentiments, we would like to suggest improvement factors to the Insurance companies.