Analysis of Health Insurance Reviews Using LDA and NER

Objective: Understand the main topics and key entities in the dataset of health insurance reviews.

1. Data Exploration:

The dataset consists of reviews related to health insurance, which have been broken down into sentences for better granularity.

2. Topic Modeling using LDA:

2.1 Negative Reviews:

Using LDA, the top topics identified for negative reviews are:

Topic #0: Administrative issues: 'told', 'called', 'said', 'phone', 'time', 'number', 'asked', 'company', 'insurance', 'day'

Topic #1: Insurance coverage and providers: 'insurance', 'company', 'health', 'care', 'policy', 'customer', 'coverage', 'service', 'plan', 'premium'

Topic #2: Medical conditions and treatments: 'doctor', 'surgery', 'pain', 'hospital', 'medication', 'drug', 'treatment', 'months', 'weeks', 'time'

Topic #3: Personal experiences and feelings: 'like', 'just', 'don', 'know', 'time', 'people', 'feel', 'want', 'years', 'way'

Topic #4: Financial concerns: 'premium', 'month', 'plan', 'year', 'coverage', 'pay', 'deductible', 'cost', 'insurance', 'company'

2.2 Positive Reviews:

For the positive reviews, the main topics are:

Topic #0: Communication and services: 'told', 'called', 'right', 'said', 'time', 'phone', 'doctor', 'information', 'days', 'new'

Topic #1: Insurance providers and care: 'insurance', 'company', 'good', 'health', 'service', 'care', 'customer', 'united', 'kaiser', 'healthcare'

Topic #2: Personal health journey: 'doctor', 'pay', 'year', 'month', 'years', 'pain', 'old', 'medication', 'went', 'high'

Topic #3: General feelings and lifestyle: 'make', 'sure', 'know', 'better', 'like', 'don', 'just', 'people', 'way', 'live'

Topic #4: Plans and medical coverage: 'plan', 'new', 'insurance', 'network', 'coverage', 'medical', 'blue', 'medicare', 'primary', 'care'

3. Named Entity Recognition (NER) on Topic 2 of Positive Reviews:

Entities were extracted from the reviews contributing to the topic related to personal health journeys. After extracting entities, they were categorized and organized:

- ORGs: Major health insurance or healthcare-related organizations.
- GPEs: Geopolitical entities, including misclassified entities.
- ORDINALs & CARDINALs: Sequence or frequency of experiences.
- LANGUAGEs: Language preferences or barriers.
- PERSONs: A mix of personal names and misclassified entities.
- TIMEs & DATEs: Duration or timing of experiences.
- WORK_OF_ARTs: Unique phrases or titles.
- NORPs: Nationalities or religious/political groups, with some misclassifications.
- PRODUCTs: Names of specific plans or services.
- PERCENTs, MONEYs: Indicate coverage amounts or cost issues.
- LAWs: Legal acts or classifications.
- FACs, LOCs: Places or establishments, with some misclassifications.
- QUANTITYs: Specific quantities related to distance or duration.
- EVENTs: Contextual times or events influencing the experience.

4. Insights and Conclusions:

The analysis effectively categorized topics in both negative and positive reviews, shedding light on areas of concern and satisfaction among reviewers.

The topics in negative reviews primarily revolved around administrative hurdles, insurance coverages, medical treatments, personal feelings, and financial issues.

Positive reviews centered on communication, satisfaction with insurance providers, personal health stories, general sentiments, and the specifics of insurance plans.

NER, while providing structured entity extraction, exhibited some misclassifications. Fine-tuning the NER model on domain-specific data might enhance accuracy.

Entities extracted paint a picture of which organizations are frequently mentioned, the geographical preferences of the reviewers, the timelines concerning their experiences, and more.

Overall, the combined use of LDA and NER provided comprehensive insights into the dataset, offering avenues for further exploration and potential actionable items for health insurance providers.

Recommendations:

Health insurance providers can utilize these insights to enhance customer experiences, rectify administrative hurdles, and communicate more effectively.

Fine-tuning the NER model can further refine entity classifications, providing more accurate results for future analyses.

End of Report.

Note: This report provides a summarized version of our analysis. Detailed insights and recommendations would require a deeper dive into individual topics and entities.