NLP SUMMIT HEALTHCARE

Sentiment Analysis of Insurance Claims using Spark NLP

Lovedeep Saini, Ph.D.
Manager, Data Science
Arthur J. Gallagher

About me

- PhD Physics, 2013
- Joined Arthur J. Gallagher in 2018
- Leading Gallagher's Natural Language Processing and Machine Learning Initiatives, where we do a variety of language analytics in the healthcare space.

Disclaimer: I work for Gallagher, a publicly traded company. Although I work for Gallagher, any opinions I express today are my own and do not reflect those of Gallagher.

NLP SUMMIT

Agenda

- Overview
  - Insurance Industry and Natural Language Processing (NLP)
  - Sentiment Analysis
  - Benefits and Suggested Applications
- Technical Implementation
  - Spark NLP and BERT
  - Process Flow Steps
  - Existing Industrial Applications
- Challenges and Mitigation Strategies
- Accomplishments and Successes
- Summary

NLP SUMMIT

Insurance Industry and NLP

Data Sources: Text messages, Social media, Chat-bots, e-mails, Phone calls, Notes, Surveys, Contracts and other digital platforms

Insurance Carriers, Brokers & Agents, 3rd-party Admins:
to communicate with customers, send reminders, provide updates etc.

Customers:
to file a claim, ask questions, or seek information about insurance products and services

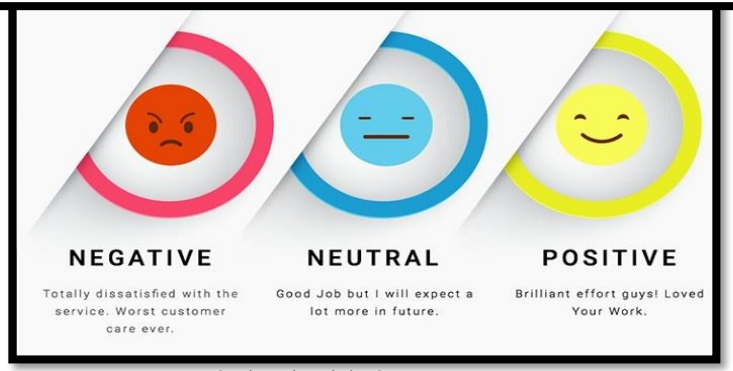
Healthcare Providers:
to communicate with customers and carriers to process medical benefits and services etc.

Natural Language Processing (NLP):
Helps by quickly analyzing large volumes of data and automating text analytics methods and ultimately resulting in improved business outcomes

NLP SUMMIT

Sentiment Analysis

- Process of determining the opinion, attitude or emotion of the writer towards a particular topic or product
- Utilize various techniques like **Natural Language Processing** (NLP), Machine Learning (ML), and Deep Learning (DL) to obtain insights from text data
- Most commonly text is classified into predefined categories such as positive, negative, or neutral



Sentiment by polarity. Source: KDnuggets

Benefits of Sentiment Analysis

Monitor Business Reputation

Improved Products

Better Customer Service

Segmentation and Target Marketing

Categorize customers based on their requirements and preferences by finding patterns in their feedback, and create focused marketing tactics to better engage each category

Generate Customized Follow-ups

By identifying the specific issues that are causing dissatisfaction, firms can develop personalized solutions to address their concerns

NLP SUMMIT

Technical Implementation

Spark NLP

- Open-source NLP library built on top of Apache Spark
- Offers wide range of NLP techniques along with easy-to-use interface
- Offers several pre-trained models
- Highly scalable, allowing users to process large datasets quickly and efficiently on distributed computing clusters
- Provides access to several state-of-art large language models like GPT2, BERT, T5 etc.

BERT

- Stands for “Bidirectional Encoder Representations from Transformers”
- Trained on a large corpus of text data using unsupervised learning techniques and can be fine-tuned for various NLP tasks
- Effectively captures the context and nuances of language, and provide highly accurate predictions of sentiment
- How do we put the idea into practice – next slide

Process Flow Steps

Data collection

Ingest data from various sources

⇩

Preprocessing and Filtering

Clean the data by removing any irrelevant information. Tokenize the text data into individual words or phrases

⇩

Embedding Generation

Use BERT/GPT/T5 to generate word

⇩

Model Training

Train a classification model OR use a pre-trained classifier instead

⇩

Model Evaluation

A challenge for a model trained on unlabeled data; Try human evaluation; label data then check accuracy, precision, recall, and F1-score etc.

⇩

Model Deployment

Deploy the trained model to a production environment to classify new text data and generate sentiment scores

Service Center Application

Voice Data: Claimant calls service center → Pre-processing: Data Converted from Voice to Text

Sentiment Classification → **Actionable Insights**

- *Thanks for your help;* **Sentiment: Positive**
- *This is the third time I am calling for my claim;* **Sentiment: Negative**

- Sentiment for every topic
- Detailed call sentiment
- Complaints can be classified according to the products, services, or operations
- Top 10 issues that members are dissatisfied with

NLP SUMMIT

Existing Industrial Applications

- Humana automated its processes for working with unstructured data using natural language processing and sentiment analysis. In order to enhance customer engagement and experience, Humana uses speech and sentiment analysis in its call centers.
- Nationwide uses sentiment analysis to detect the member's mood – showing, for instance, that people's moods worsen as the number of emails rises.
- RepuGen uses sentiment analysis to improve patient satisfaction by understanding patients emotional states.
- Allstate's Intelligent Agent Reduces Call Center Traffic and Provides Help During Quoting Process. The system was able to understand customer intent and provide personalized responses to their queries, resulting in increased customer satisfaction and a significant reduction in call center volume.
- BCBS NC's use of AI and sentiment analysis has set a new standard for customer service in the healthcare industry. The AI system implemented by BCBS NC was able to understand member intent and provide personalized responses, resulting in a significant reduction in administrative costs and an increase in member satisfaction.
- Some of the use cases mentioned in the HBR-article include tracking customer emotions during customer service calls, analyzing customer feedback on social media platforms, and using facial recognition technology to analyze customer emotions in retail stores.

Challenges and Mitigation Strategies

	Challenges	Mitigation
Imbalanced Data	○ Large number of positive claims compared to negative claims ○ Lead to biased models	Oversampling, under sampling, or use weighted loss functions etc. ○ Suggested python library imblearn
Incomplete Data	○ Missing relevant information	Data augmentation or feature engineering etc. ○ Suggested python library imblearn
Noisy Data	○ Misspellings, grammatical errors etc.	Text preprocessing, cleaning, and filtering etc. ○ Suggested python library spacy, nltk, spark-nlp
Data Storage	○ Data stored in different formats and	Store and manage data effectively ○ Data warehousing, data lake architectures, or cloud storage etc.
Ethics	○ Establishing and maintaining ethical practices	Ensure member opinions are obtained from all possible legal sources with year around supply of member sentiment data

NLP SUMMIT

Challenges and Mitigation Strategies …continued

	Challenges	Mitigation
Negations and Sarcasm	Sentiment the text conveys may not match the sentiment of the words employed ○ For example: ○ "I don't like this insurance policy". ○ "Great, now my premiums have gone up again.“	○ **Negation Handling**: ○ Modify text by adding word "not" or other negation words ○ Suggested negation handler function ○ **Sarcasm Detection**: ○ Use ML techniques to detect sarcasm in text. ○ Suggested SparkNLP model classifierdl_use_sarcasm ○ **Emotion Detection**: ○ Identify emotions *such as anger, frustration, sadness etc.* ○ Suggested SparkNLP model bert_sequence_classifier_emotion
Context Specific Words and Expressions	Insurance industry specific vocabulary ○ Certain words may have different meanings	Use domain-specific knowledge and expertise to extract relevant information ○ Fine-tune BERT or GPT-3 on domain-specific data ○ Techniques such as topic modeling or entity recognition can be used to identify relevant keywords or phrases in the data

Accomplishments and Successes

Across Industries	With Spark NLP & BERT
○ Improved Customer Retention	○ Efficient Processing
○ More Efficient Claims Handling	○ Customization
○ Better Risk Management	○ Scalability
○ Enhanced Marketing Strategies	○ Improved Accuracy and Fewer Errors

NLP SUMMIT

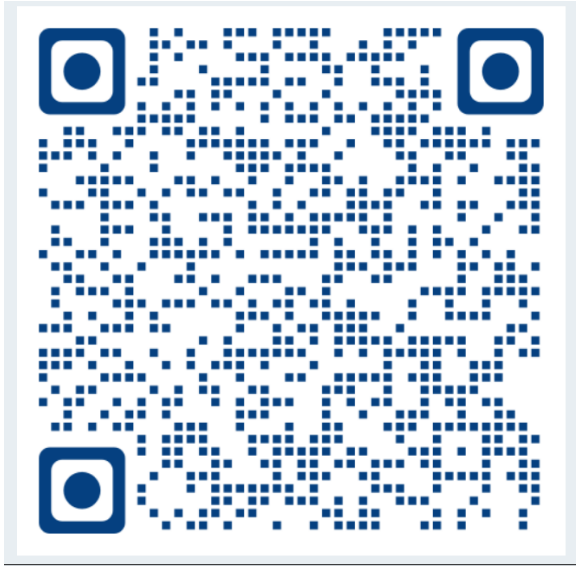
Summary

- The recent advancements in language models and contextual embeddings have enabled natural language processing (NLP) to become highly efficient, potent, and scalable.
- Businesses can readily implement NLP solutions through the Spark NLP framework, which offers pre-trained models and tools that can be deployed without significant technical expertise.
- In today's business landscape, companies are striving to become data-driven decision-makers and are gathering vast amounts of unstructured data that can be converted into valuable insights using NLP techniques.
- By following these measures or utilizing sentiment analysis and other NLP approaches, organizations can gain a competitive edge and improve customer service while enhancing their own operational efficiency.

NLP SUMMIT

Thank you for your time and attention!!!

- Questions?
- Please feel free to contact me at LinkedIn https://www.linkedin.com/in/lovedeep-kaur-saini/



NLP SUMMIT

NLP SUMMIT HEALTHCARE

Presented by John Snow LABS

www.nlpsummit.org