Risk Prediction Sentiment

Jhanavi Ghadia, Sai Gurrapu, Tarun Singh, Sai Krishna Yeshala

Instructor: Edward A. Fox
Client: Christian Johnson
Class: CS 4624 - Multimedia/Hypertext and Information Access
Presentation Outline

- Project Overview
- Solution Approach
- Work Survey Dataset
- Data Cleaning
- What is Sentiment Analysis
- Sentiment Analysis structure
- Sentence Tagging
- Speech To Text
- Django Web App Walkthrough
- Future Plans
- Acknowledgements
- References
Project Overview

● **Primary goal:** Evaluate the safety levels and risks for the workers at any workplace

● **Solution approach:** Perform sentiment analysis on the narrative data provided by the workers from their experience and provide a clear idea about the safety levels

● **Client:** Christian Johnson, Founder at www.riskinsights.ai
Solution Approach

I. Research on sentiment analysis tools
II. A relevant web interface model
III. Speech-to-text conversion tool
IV. Effective integration of all the parts

Employee Safety and Health Insights - Future
# Worker Survey Dataset

<table>
<thead>
<tr>
<th>Organization</th>
<th>Sub-Organization</th>
<th>Site</th>
<th>Dept</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site D</td>
<td>Laboratorium</td>
<td>15-Dec-20</td>
<td></td>
</tr>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site A</td>
<td>Contractors</td>
<td>15-Dec-20</td>
<td></td>
</tr>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site D</td>
<td>Energie</td>
<td>14-Dec-20</td>
<td></td>
</tr>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site A</td>
<td>Dextrose</td>
<td>12-Dec-20</td>
<td></td>
</tr>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site D</td>
<td>Energie</td>
<td>11-Dec-20</td>
<td></td>
</tr>
<tr>
<td>Global Operations</td>
<td>GO Corn Wet Milling</td>
<td>Site D</td>
<td>Drogerij 1924/1986</td>
<td>9-Dec-20</td>
<td></td>
</tr>
</tbody>
</table>

Description column blurred to preserve privacy.
Data Cleaning

- The given data set consists of various null entries, feedback in a foreign language and some irrelevant information.
- We cleaned out the dataset by removing all of these entries and passed the most accurate and relevant data to our sentiment analysis tool.
What is Sentiment Analysis?

- Use of natural language processing to assess and quantify sentences
- Determine whether a piece of writing has positive, negative or neutral sentiment
Sentiment Analysis Structure

- Each text is a list of sentences
- Each sentence is a list of tokens
- Each token is a tuple of 3 elements:
  - A word form, a word lemma, list of associated tags
Tagging Sentences

```python
expression_form = ''.join([word[0] for word in sentence[i:j]]).lower()
expression_lemma = ''.join([word[1] for word in sentence[i:j]]).lower()
if tag_with_lemas:
    literal = expression_lemma
else:
    literal = expression_form
if literal in self.dictionary:
    is_single_token = j - i == 1
    original_position = i
    i = j
    taggings = [tag for tag in self.dictionary[literal]]
    tagged_expression = (expression_form, expression_lemma, taggings)
    if is_single_token:  # if the tagged literal is a single token, conserve its previous taggings:
        original_token_tagging = sentence[original_position][2]
        tagged_expression[2].extend(original_token_tagging)
    tag_sentence.append(tagged_expression)
tagged = True
```

lack of: [inv]
not: [inv]
too: [inc] bare: [dec]
very: [inc] little: [dec]
sorely: [inc]
Speech to Text

- **Goal:** Take in detailed narrative information and descriptions of the employee’s feedback through voice
- **Input:** Voice data through computer mic
- **Output:** Text transcription generated for data cleaning/sentiment analysis

[Diagram showing the process of automatic speech recognition, natural language processing, and transcription]
Django Web App

Risk Prediction Sentiment Analysis

Upload Report
Please upload the worker feedback report below:

Choose File  No file chosen

Upload Report

Web App displaying the Upload Report page(left) and the Results page graphs(right).
Future Plans

● Work on discrete industrial data with more precise AI trained model
● Make the tool more flexible with several additional features
● Perform a data science project to prove whether safety 'sentiment' is correlated with other company performance measure
Lessons Learned

- Iteration and consistent communication
- Planning
- Cleaning the Dataset (includes foreign languages)
- Speech to Text Accuracy
- UI principles
- Add time for debugging
- Leverage existing work in the field
Acknowledgements

- Client: Christian Johnson,
  - LinkedIn: [www.linkedin.com/in/cjohnson-ehs](http://www.linkedin.com/in/cjohnson-ehs)
  - Founder: [www.riskinsights.ai](http://www.riskinsights.ai)
  - Present: Director at AstraZeneca for Safety, Health, and Environment

- Professor: Edward Fox
  - Email: fox@vt.edu
References


