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

Virginia Tech, Blacksburg, VA 24061, May 14, 2021.

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

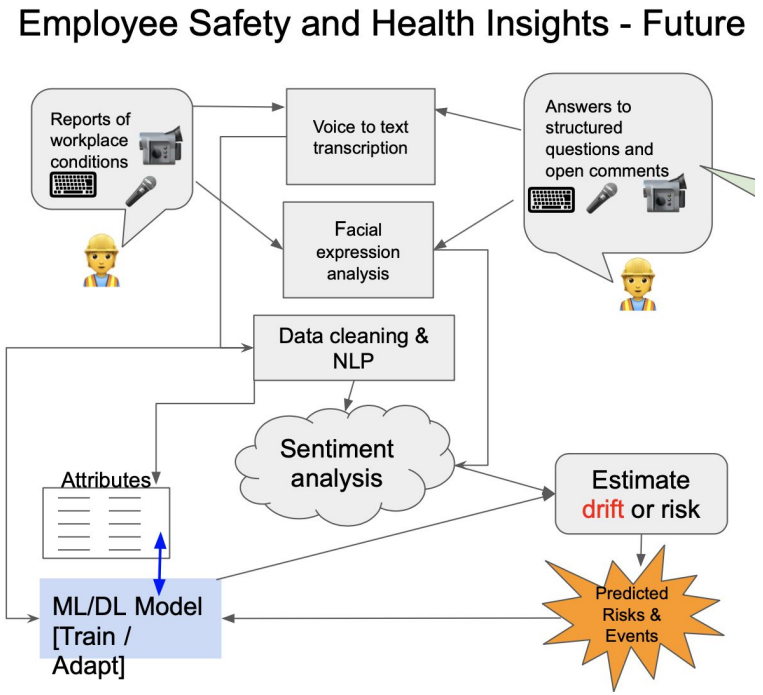
Solution Approach

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



Worker Survey Dataset

A	B	C	D	E	F
Organization	**Sub-Organization**	**Site**	**Dept**	**Date**	**Description**
Global Operations	GO Corn Wet Milling	Site D	Laboratorium	15-Dec-20	
Global Operations	GO Corn Wet Milling	Site A	Contractors	15-Dec-20	
Global Operations	GO Corn Wet Milling	Site D	Energie	14-Dec-20	
Global Operations	GO Corn Wet Milling	Site A	Dextrose	12-Dec-20	
Global Operations	GO Corn Wet Milling	Site D	Energie	11-Dec-20	
Global Operations	GO Corn Wet Milling	Site D	Drogerij 1924/1986	9-Dec-20	

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

```
[[('Manual', 'Manual', ['JJ']),
  ('die', 'die', ['NN']),
  ('cutter', 'cutter', ['NN']),
  ('had', 'had', ['VBD']),
  ('been', 'been', ['VBN']),
  ('moved', 'moved', ['VBN']),
  ('during', 'during', ['IN']),
  ('renovation', 'renovation', ['NN']),
  (';', ';', [':']),
  ('EE', 'EE', ['NNP']),
  ('was', 'was', ['VBD']),
  ('moving', 'moving', ['VBG']),
  ('mackine', 'mackine', ['NN']),
  ('back', 'back', ['RB']),
  ('and', 'and', ['CC']),
  ('had', 'had', ['VBD']),
  ('placed', 'placed', ['VBN']),
  ('it', 'it', ['PRP']),
  ('to', 'to', ['TO']),
  ('close', 'close', ['VB']),
  ('to', 'to', ['TO']),
  ('wall-grabbed', 'wall-grabbed', ['JJ']),
  ('lever/handle', 'lever/handle', ['NN']),
  ('to', 'to', ['TO']),
  ('pull', 'pull', ['VB']),
  ('it', 'it', ['PRP']),
  ('forward', 'forward', ['RB']),
  ('.', '.', ['.'])],
 [('The', 'The', ['DT']),
  ('handle', 'handle', ['NN']),
  ('came', 'came', ['VBD']),
  ('loose/I', 'loose/I', ['RB']),
  ('fell', 'fell', ['negative', 'VBD']),
  ('backwards', 'backwards', ['NNS']),
  ('on', 'on', ['IN']),
  ('floor', 'floor', ['NN']),
  ('.', '.', ['.'])],
```

Tagging Sentences

```
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_lemmas:
    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
```

```
[('The', 'The', ['DT']),
 ('handle', 'handle', ['NN']),
 ('came', 'came', ['VBD']),
 ('loose/I', 'loose/I', ['RB']),
 ('fell', 'fell', ['negative', 'VBD']),
 ('backwards', 'backwards', ['NNS']),
 ('on', 'on', ['IN']),
 ('floor', 'floor', ['NN']),
```

```
lack of: [inv]
not: [inv]
too: [inc]
very: [inc]
sorely: [inc]
barely: [dec]
little: [dec]
```

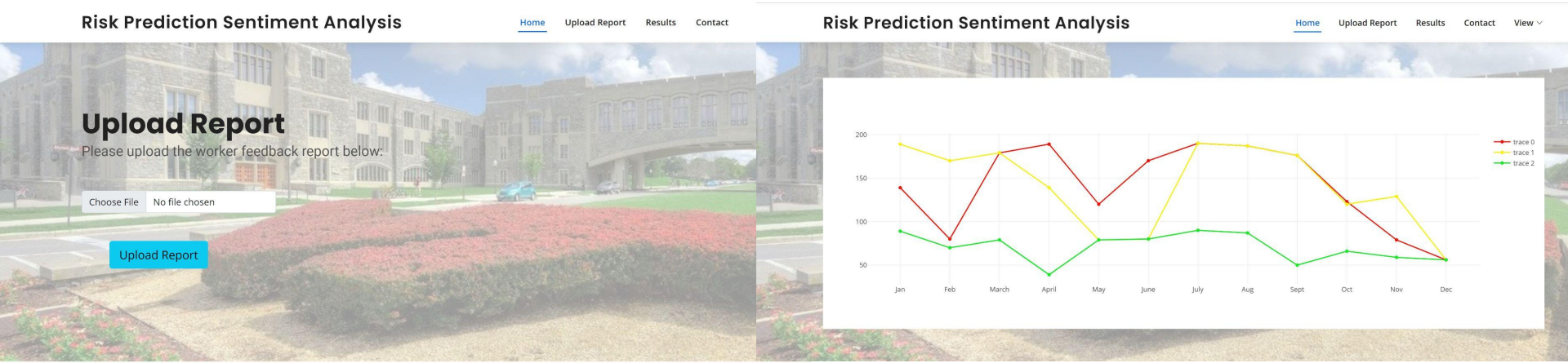
Speech to Text

- **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

python™

Google Web Speech API

Django Web App



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

django

python™

Bootstrap 4

PostgreSQL

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
  - Founder: www.riskinsights.ai
  - Present: Director at AstraZeneca for Safety, Health, and Environment

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

References

[1] Nvidia Corporation. (2020, October 13). How to build domain specific automatic speech recognition models on GPUs. Retrieved April 09, 2021, from https://developer.nvidia.com/blog/how-to-build-domain-specific-automatic-speech-recognition-models-on-GPUS/

[2] J. Wiebe and E. Riloff, "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction," in IEEE Transactions on Affective Computing, vol. 2, no. 4, pp. 175-191, Oct.-Dec. 2011, doi: 10.1109/T-AFFC.2011.19.

[3] SketchBubble. (2021). SketchBubble Data cleansing. Retrieved April 09, 2021, from https://www.sketchbubble.com/en/presentation-data-cleansing.html

[4] Google Cloud. (2021). Speech-to-Text: Automatic speech Recognition | Google Cloud. Retrieved April 09, 2021, from https://cloud.google.com/speech-to-text,

[5] Beri, A. (2020, May 27). Sentimental analysis using VADER. Retrieved April 09, 2021, from https://towardsdatascience.com/sentimental-analysis-using-VADER-a3415fef7664

[6] Clarabridge. (2019, March 05). Clarabridge Insight. Retrieved April 09, 2021, from https://www.ccma.org.uk/insight/clarabridge/

[7] Django Software Foundation. (2021). Django Web Framework. Retrieved April 09, 2021, from https://www.djangoproject.com/

[8] JavaTpoint. (2021). Django MVT - JavaTpoint. Retrieved April 9, 2021, from https://www.javatpoint.com/django-mvt

[9] Baer, T. (2018, April 20). Has the time finally come for PostgreSQL? Retrieved April 09, 2021, from https://www.zdnet.com/article/has-the-time-finally-come-for-PostgreSQL

[10] RealPython. (2021, April 03). The ultimate guide to speech recognition with Python. Retrieved April 30, 2021, from https://realpython.com/python-speech-recognition/