Risk Prediction Sentiment Analysis 3

Kaustubh Kale, Jaswant Kasinatedi, Sagar Mehta, Parvesh Samayamanthula, Arjun Vashistha
Outline

- Project Overview
- Timeline
- Frontend
- Backend
- Lessons Learned
- Future Work
- Acknowledgments
- References
Project Overview

- Hypothesis: Worker sentiment in the workplace indicates the overall safety of the workplace
- Intuitive web app for users to upload worker responses and get sentiment analysis results
Timeline

**Jan 27 - Feb 15**
- Set up AWS Permissions
- Review/Test previous base code

**Feb 15 - Feb 28**
- Test/Compare AWS and Cloud ML models
- Research transcription and login authentication API's

**Feb 28 - Mar 15**
- Implement login authentication (prioritized over Audio API)
- Finalize/Implement chosen ML model
- Manually label ~1500 worker narrative responses

**Mar 15 - Apr 15**
- Finalize UI Improvements
- Deploy Google Cloud backend function
- Start bug testing on front and backend

**April 15 - May 4**
- Connect backend to frontend
- Implemented CI/CD
- Fix all bugs
Frontend
Reset Password

Email

Reset Password

Login

Need an account? Register
Sign Up

Email

Password

Password Confirmation

Sign Up

Already have an account? Log In
**Instructions**

- **Dashboard Panel**
  - Select the Upload Tab on the top menu
  - Upload the data in CSV form with the correct header format (Date | Organization | Site | Sub-organization | Description)
  - After upload is complete, click Submit data to model button and proceed to Dashboard Panel to view outputs

- **Dashboard Panel**
  - Select the Dashboard Tab on the top menu
  - Monitor Risk and Sentiment chart to view % of positive/negative/neutral/mixed sentiments

- **Review Panel**
  - Select the Review Tab on the top menu
  - Review each sentiment output to verify correct sentiment was specified
  - If the sentiment is incorrect, manually change the clause to the specified sentiment to improve accuracy in future

- **Video Walkthrough**
# Users

- **Name**: Christian Johnson
- **Registered**: 2018/01/01
- **Role**: Owner
- **Status**: Active
Backend
Sentiment Analysis

- Services tested: Microsoft Azure, Google Cloud Platform
- Chose GCP for its custom AutoML Service
- Accepts range of sentiment scores and labeled data (Ex: 0-4)
Backend Infrastructure
AutoML Sentiment Analysis

- Labeled 1544 worker responses for supervised learning
- Initially tried a sentiment range of 0-10
- Settled on a range of 0-4
Revised Accuracy Evaluation

- Accuracy of predicting whether worker responses are positive or negative

```python
accurate = 0
[accurate := accurate + 1 for m, r in zip(model_ratings, test_ratings) if m == 2 and r == 2 or (
    (m == 3 or m == 4) and (r == 3 or r == 4)) or ((m == 0 or m == 1) and (r == 0 or r == 1))]

print(accurate / len(test_ratings))

0.8459387755102041
```
<table>
<thead>
<tr>
<th>Date</th>
<th>Organization</th>
<th>Site</th>
<th>Sub-organization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-09-18</td>
<td>Global Operations</td>
<td>Site B</td>
<td>Wetmill</td>
<td>Asked Wetmill/Feedhouse team if machine guarding audits have been completed. Not aware of a recent audit. Removable gate in question should have a lock on it. Notified Area Technologist and Area Engineer and asked that the chain be replaced at the DSL dryer.</td>
</tr>
<tr>
<td>2018-10-04</td>
<td>Global Operations</td>
<td>Site B</td>
<td>Office area</td>
<td>Talked to employee about their immediate work space and the ergonomic impact of monitor position.</td>
</tr>
<tr>
<td>2018-10-20</td>
<td>Global Operations</td>
<td>Site C</td>
<td>Operations</td>
<td>Great job and clean up</td>
</tr>
</tbody>
</table>
Upload Page Workflow

- User uploads input CSV to upload page
- CSV gets sent to the backend, is parsed, and worker descriptions are sent to the Google Cloud model
- Sentiment results are sent back from the model and appended to input CSV
- New output CSV is sent back to frontend, where it’s automatically downloaded
<table>
<thead>
<tr>
<th>Date</th>
<th>Organization</th>
<th>Site</th>
<th>Sub-organization</th>
<th>Description</th>
<th>Sentiment Score</th>
<th>Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-09-18</td>
<td>Global Operations</td>
<td>Site B</td>
<td>Wetmill</td>
<td>Asked Wetmill/Feedhouse team if machine guarding audits have been completed. Not aware of a recent audit. Removable gate in question should have a lock on it. Notified Area Technologist and Area Engineer and asked that the chain be replaced at the DSL dryer.</td>
<td>2</td>
<td>Neutral</td>
</tr>
<tr>
<td>2018-10-04</td>
<td>Global Operations</td>
<td>Site B</td>
<td>Office area</td>
<td>Talked to employee about their immediate work space and the ergonomic impact of monitor position.</td>
<td>2</td>
<td>Neutral</td>
</tr>
<tr>
<td>2018-10-20</td>
<td>Global Operations</td>
<td>Site C</td>
<td>Operations</td>
<td>Great job and clean up</td>
<td>4</td>
<td>Positive</td>
</tr>
</tbody>
</table>
CI/CD
Lessons Learned

- Time management - gauging which features to prioritize and the time to implement them
- Quality of data - skewed data results
- Building on previous team’s work and ideas
Future Work

- Audio transcription
- Continuous learning/improvement of ML model
- Testing and feedback tool
- Better visualizations UI for sentiment data
Acknowledgements

- Dr. Fox
  - Mentor
- Christian Johnson
  - Client
References


