DocFin: Multimodal Financial Prediction and Bias Mitigation using Semi-structured Documents

Puneet Mathur
University of Maryland
puneetm@umd.edu

Mihir Goyal
IIIT-Delhi
mihir17166@iiitd.ac.in

Ramit Sawhney
Georgia Institute of Technology
rsawhney31@gatech.edu

Ritik Mathur
IIT-Roorkee
rmathur@me.iitr.ac.in

Jochen L. Leidner
University of Sheffield
leidner@acm.org

Franck Dernoncourt
Adobe Research
dernonco@adobe.com

Dinesh Manocha
University of Maryland
dmanocha@umd.edu

Abstract

Financial prediction is complex due to the stochastic nature of the stock market. Semi-structured financial documents present comprehensive financial data in tabular formats, such as earnings, profit-loss statements, and balance sheets, and can often contain more than 100’s tables worth of technical analysis along with a textual discussion of corporate history, and management analysis, compliance, and risks. Existing research focuses on the textual and audio modalities of financial disclosures from company conference calls to forecast stock volatility and price movement, but ignores the rich tabular data available in financial reports. Moreover, the economic realm is still plagued with a severe under-representation of various communities spanning diverse demographics, gender, and native speakers. In this work, we show that combining tabular data from financial semi-structured documents with text transcripts and audio recordings not only improves stock volatility and price movement prediction by 5-12% but also reduces gender bias caused due to audio-based neural networks by over 30%.

1 Introduction

Financial risk modeling is of great interest to capital market participants for making sound investment decisions. Earnings calls are quarterly audio conference calls wherein company executives discuss their companies’ performance and future prospects with outside analysts and investors (Qin and Yang, 2019). Mergers and Acquisitions (M&As) conference calls are held preceding financial transactions involving two or more entities such that either one of the participant companies takes over the other(s) (“acquisition”) or combines with another to become a joint entity (“merger”) (Sawhney et al., 2021b). Both kinds of events consist of a prepared speech delivery by company executives on analysis and future expectations followed by a spontaneous analyst-driven question-answer session to seek additional information (Ye et al., 2020). Several past works have utilized the text transcripts and audio recordings from these calls to improve the stock forecasting (Mathur et al., 2022d; Yang et al., 2020; Zhou et al., 2020; Chen et al., 2020a; Ye et al., 2020; Sawhney et al., 2021b,a, 2020a). However, most prior works exclusively focus on vocal verbal information, often ignoring information from official financial documents. Financial semi-structured documents such as 10-K and 10-Q reports are publicly available, recurrent mandatory filings made by public companies to disclose their financial performance. These semi-structured financial documents present comprehensive financial data in tabular format, such as earnings, profit and loss statements and balance sheets, and can often contain more than 100’s of tables worth technical analysis. Information contained in such financial documents also includes a textual discussion of corporate history, management analysis, compliance, risks, and future plans about new projects relevant for investment decision-making (Kogan et al., 2009).

Recent studies such as Sawhney et al. (2021a) have highlighted the downside of utilizing audio-based multimodal approaches for financial risk prediction due to the inherent gender bias induced in learning models due to the imbalance of speaker demographics in call recordings. Audio features such as speakers’ pitch and intensity can vary greatly across genders. Under-representation of female executives in conference calls is amplified by deep learning models, leading to high error disparity between stock predictions across sensitive attributes.

We combine tabular from financial semi-structured documents input with existing vocal-verbal information from audio call recordings to improve stock price movement and volatility prediction. We demonstrate that supplementing existing conference calls transcripts with the tabular...
We combine publicly available earnings calls (MAEC (Li et al., 2020)) and M&A calls (Sawhney et al., 2021b) datasets with tabular data extracted from SEC-EDGAR 10-Q and 10-K company-filing documents.

We utilize tabular information from financial semi-structured documents with existing textual and audio modalities to show 8-12% relative improvement in stock volatility and price movement prediction tasks across several baseline and state-of-the-art models.

We empirically show the extent of induced gender bias due to audio modality in the financial prediction models and demonstrate the usefulness of tabular data extracted from semi-structured financial documents as an alternative to audio modality for reducing gender bias by 30% in audio-based neural networks, without significant performance degradation.

2 Methodology

Problem Formulation: We consider an input conference call recording $\chi = [t; a; tab]$, such that each call comprises of multimodal components: $N$ textual utterances $t = [t_1, t_2...t_N]$ aligned with their corresponding audio slices $a = [a_1, a_2...a_N]$, and $tab = [tab_1, tab_2...tab_M]$ corresponding to the $M$ tables extracted from the company filings relevant to the call. Each conference call is associated with speaker information denoted by $s = [s_1, s_2...s_N]$, representing the sequence of speakers for the utterances. We formulate volatility as a regression task (Kogan et al., 2009) and price movement prediction as a binary classification task (Xu and Cohen, 2018).

Measuring stock volatility: Following (Kogan et al., 2009), we formulate volatility as a regression task. For a given stock with a close price of $p_k$ on the trading day $k$, we calculate the average log volatility as the natural log of the standard deviation of return prices $r$ in a window of $\tau$ days as.

$$v_{[0,\tau]} = \ln \left( \sqrt{\frac{\sum_{k=1}^{\tau} (r_k - \bar{r})^2}{\tau}} \right)$$  

where $r_k = \frac{p_k - p_{k-1}}{p_{k-1}}$ is the return price on day $k$ for a given stock, and $\bar{r}$ is the average return price over a period of $\tau$ days.

Price movement prediction: Following (Xu and Cohen, 2018), we define price movement $y_{d-\tau,d}$ over a period of $\tau$ days as a binary classification task. For a given stock, we employ its close price, which can either rise or fall on a day $d$ compared to a previous day $d - \tau$, to formulate the classification task as:

$$y_{[d-\tau,d]} = \begin{cases} 1, & p_{d+\tau} > p_d, \\ 0, & p_{d+\tau} \leq p_d \end{cases}$$  

Given a conference call $\chi$, we experiment with several baseline and state-of-the-art multimodal financial prediction models (example M3A (Sawhney et al., 2021b)) for volatility and stock price prediction on M&A calls using dot product attention to extract condensed table representations relevant to text from transcripts, weight averages each modality through softmax layer to obtain a fused embedding, combines the fused embeddings with speaker and positional embedding, and finally passes through the Transformer model for stock price movement and volatility prediction tasks.
Table 1: Dataset statistics for the M&A dataset

<table>
<thead>
<tr>
<th>Year</th>
<th># of MA Calls</th>
<th>Mean # of Utterances</th>
<th>Mean # of Speakers</th>
<th>Mean # of Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>192</td>
<td>117.4</td>
<td>11.26</td>
<td>217.2</td>
</tr>
<tr>
<td>2017</td>
<td>206</td>
<td>99.62</td>
<td>10.14</td>
<td>216.83</td>
</tr>
<tr>
<td>2018</td>
<td>232</td>
<td>90.51</td>
<td>10.07</td>
<td>231.03</td>
</tr>
<tr>
<td>2019</td>
<td>153</td>
<td>97.41</td>
<td>10.39</td>
<td>228.64</td>
</tr>
<tr>
<td>2020</td>
<td>49</td>
<td>104.07</td>
<td>10.32</td>
<td>216.97</td>
</tr>
</tbody>
</table>

Table 2: Dataset statistics for the Earnings Call dataset

<table>
<thead>
<tr>
<th>Year</th>
<th># of Calls</th>
<th>Mean # of Utterances</th>
<th>Mean # of Speakers</th>
<th>Mean # of Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>532</td>
<td>87.35</td>
<td>1.74</td>
<td>194.74</td>
</tr>
<tr>
<td>2016</td>
<td>1125</td>
<td>87.29</td>
<td>1.74</td>
<td>211.84</td>
</tr>
<tr>
<td>2017</td>
<td>469</td>
<td>109.38</td>
<td>1.86</td>
<td>211.23</td>
</tr>
<tr>
<td>2018</td>
<td>160</td>
<td>134.14</td>
<td>2.01</td>
<td>205.52</td>
</tr>
</tbody>
</table>

et al., 2021b) in Fig. 1). We predict the average negative log volatility \( v_{[0,\tau]} \) and price movement direction \( y_{[0,\tau]} \) using the multimodal call data \( \chi = [t;a;tab] \) for \( \tau = 3, 7 \) and 15-day interval.

Encoding Text Transcript, Audio Call and Speakers: We process text and audio data following earlier works on Earnings Calls (Li et al., 2020) and M&A calls (Sawhney et al., 2021b). Each text utterance \( t_i \) is represented as a 768-dimensional encoding \( g_i \) using BERT. Each audio utterance \( a_i \) is encoded into its embedding \( h_i \) corresponding to the type of conference call. For M&A calls, we extract \( h_i \) as a 62-dimensional encoding described in (Eyben et al., 2016) using OpenSMILE\(^1\) and for Earnings Calls as a 29-dimensional low-level audio features encoding using Praat (Boersma and Van Heuven, 2001). We extract the list of speakers from the transcripts and assign each speaker \( s_j \) a sequential ID in the order of listing and represent the speaker embedding as one-hot encoding.

Encoding Tables from Company Filings: Taking inspiration from past literature (Chen et al., 2020b), we linearize each table \( tab \) into a sentence representation. For a row \( i \) with column names \( c_j \) and values \( v_{ij} \), the row is represented as ‘row \( i \)’s \( c_1 \) is \( v_{i1}; \) the \( c_2 \) is \( v_{i2}; \)…’. Each row’s representation is concatenated using punctuation to obtain a table representation which is encoded to its 768-dimensional table encoding \( k_i \) using BERT.

Combining tabular data with text-audio time series: We provide a generalized method to process, fuse and utilize the tabular data with text-audio modality such that it is extensible across different neural architectures. To this end, we use dot-product attention to allow each text utterance \( g_i \) to extract a condensed table representation \( l_i \) from the table encoding \( k_i \), such that \( l_i = DotProdAttn(g_i, k_i) \). To fuse the encoding, we linearly transform the text and tabular encoding to the size of the audio encoding and employ the use of multi-headed self-attention (Vaswani et al., 2017). The text, audio and table features are multiplied by their softmax-ed weights (\( W' = \sigma(g_{W_{att}} + b_{w_{att}})VT,TA \), summed \( S = W'_T + W'_A + W'_TA \), and weighted averaged to get attention weights \( W_T, W_A, W_{TA} = W'_T, W'_A, W'_TA \), which are added to get the fused embeddings \( X_{fused} = gW_T + hW_A + LW_{TA} \). We augment \( X_{fused} \) with the speaker information \( s \) by concatenation (represented by \( \oplus \)) and the position embeddings \( pos \) by addition as \( X_{final} = (X_{fused} + pos) \oplus s \). The augmented document features, called DocEmbedding, can be used by an encoder (recurrent, attention-based or Transformer) for processing to produce the task predictions.

3 Experiments

Datasets: We train and test several baseline and state-of-the-art models that utilize the multimodal input on two datasets: Multimodal Aligned Earnings Call (MAEC) Dataset (Li et al., 2020) and Multimodal Multi-Speaker Merger& Acquisition Call Financial Forecasting (M3A) Dataset (Sawhney et al., 2021b), both containing aligned text transcripts and audio recordings of their respective types of conference calls. We collect the most recently filed 10-K and 10-Q documents before the date of the call\(^2\) and parse the HTML content to retrieve all tables with at least 10 rows. We describe the dataset statistics in Table 1 and Table 2. We tune all hyper-parameters using Grid Search and implement all methods with Keras\(^3\). We use training/validation/testing splits released by respective datasets.

4 Results and Discussion

Effect of Tabular Datat on Financial Predictions: Table 3 shows the performance of several baseline and SOTA models for predicting price movement and stock volatility for Merger & Acquisition calls on the M&A dataset. Table 4 reports the volatility prediction performance on the MAEC dataset. We report average MSE and F1 scores for volatility and price movement prediction, respectively. We observe significant gains (8-12%) in both tasks across attention based (MDRM, VoLT-AGE, MMFTR) and Transformer models (M3A)

\(^1\)https://pypi.org/project/opensmile/
\(^2\)Using https://github.com/jadchaar/sec-edgar-downloader
\(^3\)https://keras.io/
Table 3: Mean τ-day volatility (MSE) and price movement prediction (F1 score) results for Merger & Acquisition calls (M&A dataset) across several models. * indicates result is significantly better than the M3ANet under Wilcoxon’s Signed Rank test. Adding DocEmbedding outperforms base methods across all tasks and intervals.

<table>
<thead>
<tr>
<th>Model</th>
<th>Volatility Prediction</th>
<th>Price Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE (3 \downarrow)</td>
<td>MSE (7 \downarrow)</td>
</tr>
<tr>
<td>RoBERTa + LSTM (Liu et al., 2019)</td>
<td>0.78 (0.009)</td>
<td>0.58 (0.009)</td>
</tr>
<tr>
<td>GloVe + LSTM (Pennington et al., 2014)</td>
<td>0.80 (0.005)</td>
<td>0.60 (0.004)</td>
</tr>
<tr>
<td>F1nBERT + LSTM + (Araci, 2019)</td>
<td>0.78 (0.008)</td>
<td>0.60 (0.004)</td>
</tr>
<tr>
<td>MDRM (Qin and Yang, 2019)</td>
<td>0.78 (0.005)</td>
<td>0.58 (0.003)</td>
</tr>
<tr>
<td>MDRM + DocEmbedding</td>
<td>0.76 (0.006)</td>
<td>0.55 (0.001)</td>
</tr>
<tr>
<td>M3ANet (Sawhney et al., 2021b)</td>
<td>0.79 (0.020)</td>
<td>0.61 (0.012)</td>
</tr>
<tr>
<td>M3ANet + DocEmbedding</td>
<td>0.73 (0.006)</td>
<td>0.54 (0.012)</td>
</tr>
</tbody>
</table>

Table 4: Mean τ-day MSE for stock volatility prediction for Earnings Calls (MAEC dataset) across several methods. * indicates result is significantly better than the VoLTAGE under Wilcoxon’s Signed Rank test. Our approach of augmenting with DocEmbeddings outperform corresponding base methods across 3, 7,15-day intervals.

<table>
<thead>
<tr>
<th>Model</th>
<th>Volatility Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE ( \downarrow)</td>
</tr>
<tr>
<td>Vpast</td>
<td>2.99</td>
</tr>
<tr>
<td>LSTM</td>
<td>1.97</td>
</tr>
<tr>
<td>HAN (GloVe)</td>
<td>1.43</td>
</tr>
<tr>
<td>MDRM (Qin and Yang, 2019)</td>
<td>1.37</td>
</tr>
<tr>
<td>MMTFR (Sawhney et al., 2021b)</td>
<td>0.60</td>
</tr>
<tr>
<td>MMTFR + DocEmbedding</td>
<td>0.58</td>
</tr>
<tr>
<td>VoLTAGE (Sawhney et al., 2020b)</td>
<td>0.63</td>
</tr>
<tr>
<td>VoLTAGE + DocEmbedding</td>
<td>0.61</td>
</tr>
<tr>
<td>M3A (Sawhney et al., 2021b)</td>
<td>0.59</td>
</tr>
<tr>
<td>M3A + DocEmbedding</td>
<td>0.57 *</td>
</tr>
</tbody>
</table>

4.1 Bias Reduction through Company Filings

We evaluate the gender bias in SOTA M3A model by quantifying the error disparity in MSE/F1 score between male and non-male speakers (\(\Delta G = MSE_F - MSE_M / F1_M - F1_F\)) for individual text, audio and table inputs and their combinations across 3, 7 and 15-day intervals in Table 6. We observe that the table modality has the least error disparity. Audio modality has consistently higher error individually as well as in combination with either of the other modalities, while it significantly drops when considering just text and table data. The primary reason for the observation tends to be the imbalance in the male and female distribution in speakers of earnings calls. In our case, since female examples are very less in comparison to the male counterparts (only 7% in earnings calls and 12% in M&A calls identify as females), the model discriminates between male and female examples by inferring insufficient information beyond its source and learns imperfect generalizations between the attributes and labels.

4.2 Audio vs Tabular Information

While audio input modality certainly improves model performance, it adds unintended model bias due to the differences in acoustic features for males and females. Audio clips require processing-heavy algorithms such as forced alignment (Sakoe and Chiba, 1978) to extract meaningful features from linguistic and acoustic utterances as opposed to semi-structured information in tables that can be utilized with minimal processing. Replacing audio clips with tabular data from company filings leads to a reduction of data processing time and data storage requirements by over 90% and 50%,
As evident from Table 5 and 6, tabular information can substitute audio input to reduce gender bias in multimodal financial prediction tasks.

respectively for both MAEC and M&A datasets. As evident from Table 5 and 6, tabular information preserves model performance while avoiding unwanted stereotypes arising due to gender-specific audio features such as shimmer and jitter. Hence, we propose to utilize tabular information as an effective substitute for audio input for multimodal financial prediction tasks.

Table 5: Ablation analysis of M3A model augmented with DocEmbedding for each modality for volatility (MSE) and price movement prediction (F1 score) tasks across Earnings Calls and M&A calls datasets (mean and stdev. of 5 runs for each approach). Combining audio, text and tabular data gives best performance (see bold). Green shade highlights that substituting company filings instead of its audio counterpart in conjunction with text transcripts does not significantly deteriorate model performance.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Merger &amp; Acquisition Calls</th>
<th>Earnings Calls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volatility Prediction</td>
<td>Price Prediction</td>
</tr>
<tr>
<td></td>
<td>MSEi ↑</td>
<td>MSEi ↓</td>
</tr>
<tr>
<td>✓ ✓ ✓</td>
<td>0.79 (0.003)</td>
<td>0.65 (0.005)</td>
</tr>
<tr>
<td>✓ ✓ ✗</td>
<td>0.80 (0.001)</td>
<td>0.64 (0.000)</td>
</tr>
<tr>
<td>✓ ✗ ✓</td>
<td>0.72 (0.002)</td>
<td>0.62 (0.007)</td>
</tr>
<tr>
<td>✓ ✗ ✗</td>
<td>0.78 (0.009)</td>
<td>0.61 (0.007)</td>
</tr>
<tr>
<td>✗ ✓ ✓</td>
<td>0.77 (0.009)</td>
<td>0.57 (0.009)</td>
</tr>
<tr>
<td>✗ ✗ ✓</td>
<td>0.74 (0.005)</td>
<td>0.55 (0.012)</td>
</tr>
<tr>
<td>✗ ✗ ✗</td>
<td>0.69 (0.008)</td>
<td>0.54 (0.012)</td>
</tr>
</tbody>
</table>

Table 6: Modality specific $\Delta G$ i.e. the difference between the MSE and F1 for volatility prediction in Earnings Calls dataset and price prediction in M&A calls, respectively for 3, 7, and 15 days over 5 runs. We use SOTA M3A model for experiments. Here A stands for Audio only, T for Text only and Tab for Tabular modality. We show that tabular information can substitute audio input to reduce gender bias in multimodal financial prediction tasks.

Table 6: Modality specific $\Delta G$ i.e. the difference between the MSE and F1 for volatility prediction in Earnings Calls dataset and price prediction in M&A calls, respectively for 3, 7, and 15 days over 5 runs. We use SOTA M3A model for experiments. Here A stands for Audio only, T for Text only and Tab for Tabular modality. We show that tabular information can substitute audio input to reduce gender bias in multimodal financial prediction tasks.

5 Conclusion and Future Work

In this work, we show that combining tabular data from financial semi-structured documents with text transcripts and audio recordings improves stock volatility and price movement prediction by 5-12% along with reduction in gender bias learned by audio-based neural networks by over 30%. We empirically show that our approach is generic and extensible to recurrent, attention-based and Transformer models. Future work can utilize advances in document-NLP to extract temporal information extraction (Mathur et al., 2021), temporal dependency parsing (Mathur et al., 2022c), and NLI (Mathur et al., 2022b) for better contextual understanding of financial reports. Predicting the correct layout can also help align audio with transcripts (Mathur et al., 2022a).

6 Limitations

We acknowledge the presence of gender bias in our study, given the imbalance in the gender ratio of speakers of the calls. We also acknowledge the demographic bias (Sawhney et al., 2021a) in our study as the companies are organizations within the public stock market of the United States of America and may not generalize directly to non-native speakers. At the same time, we extensively study the components causing gender bias and propose ways to fix it in the current contributions.

7 Potential risks

Our contributions are meant as exploratory research in the financial domain and no part of the work should be treated as financial advice. All financial investment decisions should be made after extensive testing. Practitioners should check for various biases (demographic, gender, modeling, randomness) before attempting real-world use cases.

8 Ethical Considerations

Examining a speaker’s tone and speech in conference calls is a well-studied task in past literature (Qin and Yang, 2019; Chariri, 2009). Our work focuses on conference calls for which companies publicly release transcripts and audio recordings. The data used in our study corresponding to M&A and Earnings conference calls is open-sourced and available for download. The company document filings we use to extract tabular data are publicly available, open source and devoid of human intervention at its source. We do not collect any personalized data or violate any privacy laws in using, storing or releasing the company filing data for financial analysis.
References


Anis Chariri. 2009. Ethical culture and financial reporting: Understanding financial reporting practice within javanese perspective*. Issues In Social And Environmental Accounting, 3.


Puneet Mathur, Franck Dernoncourt, Quan Hung Tran, Juixiang Gu, Ani Nenkova, Vlad Morariu, Rajiv Jain, and Dinesh Manocha. 2020a. Doctime: A document-level temporal de-


Puneet Mathur, Vlad Morariu, Verena Kaynig-Fittkau, Juixiang Gu, Franck Dernoncourt, Quan Hung Tran, Ani Nenkova, Dinesh Manocha, and Rajiv Jain. 2022c. Doctime: A document-level temporal de-


A Dataset Details

We download the company filings data according to the train/test/val splits given by (Li et al., 2020; Sawhney et al., 2021b).

B Hyperparameter Tuning

Hyper-parameters for our model were tuned on the respective validation set to find the best configurations for different datasets. We summarize the range of our model’s hyper parameters such as: size of the transformer’s feed-forward layer and size of the linear layers ∈ {4, 16, 64}, dropout δ ∈ {0.0, 0.1, 0.25, 0.5}, learning rate λ ∈ {0.01, 0.001, 0.0001}. We used grid search to choose the best set of training configurations across each dataset. We run 5 rounds of hyper-parameter search trials and report average of observed results.

C1: Citation to creators of artifacts:

We use two datasets: (i) earnings calls dataset (MAEC (Li et al., 2020)) and (ii) M&A calls (Sawhney et al., 2021b) dataset. We augment both using past 10-K and 10-Q reports (filled annually and quarterly, respectively) found on the Securities and Exchange (SEC) website \(^4\). All datasets and documents are publicly available.

C2: License and terms for use of data artifacts:

Both the datasets are available to use for research purposes. 10-K and 10-Q documents from SEC are already in public domain due to mandatory release as per government norms for public good.

C3: Intended use of data artifacts:

The intended use of financial datasets is to enable investors take informed financial decisions, research and development to foster progress of AI methods and financial modeling for public good.

C4: Steps taken to protect / anonymize names, identities of individual people or offensive content:

We do not use any identifiable user data for any experiments. All persons mentioned in the financial reports and conference calls are publicly known and consent to release their names and data as part of SEC guidelines.

C5: Coverage of domains, languages, linguistic phenomena, demographic groups represented in data:

Our work uses conference calls and financial documents data in English language. Adaptation to other languages may need appropriate processing.

C6: Data statistics (train/test/dev splits):

The data statistics are given in Table 1 and Table 2. We download the company filings data according to the train/test/val splits given by (Li et al., 2020; Sawhney et al., 2021b).

D1: Total computational budget and computing infrastructure:

We trained the models on Nvidia GeForce RTX 2080 GPU clusters.

D2: Experimental setup, hyperparameter search and best-found values:

Hyper-parameters for our model were tuned on the respective validation set to find the best configurations for different datasets. We used grid search to choose the best set of training configurations across each dataset. We run 5 rounds of hyper-parameter search trials and report average of observed results.

D3: Descriptive statistics about results (e.g., error bars around results, summary statistics from sets of experiments):

We performed Wilcoxin’s signed rank test to establish statistical significance of the best results against the baselines.

D4: Implementation Software and Packages:

We implemented our solution in Python 3.6 using Keras framework. We used Huggingface’s implementation for BERT transformer. We used OpenSMILE\(^5\) and Praat (Boersma and Van Heuven, 2001) for audio processing. We collect the most recently filed 10-K and 10-Q documents before the date of the call\(^6\)

\(^4\)https://www.sec.gov/
\(^5\)https://pypi.org/project/opensmile/
\(^6\)Using https://github.com/jadchaar/sec-edgar-downloader