OPEN SOURCE TOOLS FOR HR SENTIMENT ANALYSIS

Marian Pompiliu CRISTESCU PhD Professor,

Lucian Blaga University of Sibiu, Romania

Dumitru Alexandru MARA, PhD Student,

Lucian Blaga University of Sibiu, Romania

Lia-Cornelia CULDA, PhD Student,

Lucian Blaga University of Sibiu, Romania

Raluca Andreea NERIȘANU, PhD,

Lucian Blaga University of Sibiu, Romania

Abstract: *In today's business world, it's more important than ever to understand how employees feel. This research explores open-source tools to analyze the sentiment of employee feedback. We developed a cost-effective Python script tailored to do sentiment analysis on text written by employees. For our methodology we used the TextBlob library to process textual data and evaluate the sentiment polarity. Initial tests confirm the script's effectiveness and precision. The findings of this research not only underscore the potential of free tools in sentiment analysis but also emphasize the importance of continuous feedback mechanisms in organizational settings.*

Keywords*: HR analytics, Python, sentiment analysis*

JEL code: *C63*

INTRODUCTION

In the evolving landscape of the corporate world, understanding employee sentiments has become a pivotal aspect of human resource management. Sentiment analysis, often referred to as opinion mining, is a computational study of people's opinions, sentiments, evaluations, attitudes, moods, and emotions (Liu, 2017). This technique has gained traction for its ability to minimize computational expenses while enhancing accuracy in data analysis (Kumar et al., 2021). In the realm of human resources, sentiment analysis can serve as a powerful tool to evaluate the mood and feelings of employees, thereby assisting organizations in making informed decisions.

HR SENTIMENT ANALYSIS

The importance of understanding and addressing employee sentiments cannot be overstated. The digital age has brought forth a multitude of tools and techniques to assess these sentiments, with sentiment analysis emerging as a frontrunner. Sentiment analysis, often interchangeably used with the term 'opinion mining', is a computational approach to deciphering people's opinions, sentiments, evaluations, attitudes, moods, and emotions from written text (Liu, 2017).

Historically, human resource departments relied heavily on traditional methods such as surveys, face-to-face interviews, and feedback forms to understand employee sentiments. While these methods provided valuable insights, they were often time-consuming, prone to biases, and lacked real-time feedback. The advent of sentiment analysis has revolutionized this paradigm. By leveraging natural language processing and machine learning algorithms, sentiment analysis offers a more efficient, unbiased, and instantaneous way to understand employee sentiments (Rajan et al., 2021).

The integration of sentiment analysis into HR systems has had profound implications for firm performance. Choi (2019) highlighted that HR systems significantly influence the performance of firms. More importantly, a positive perception of these HR systems by employees mediates the relationship between HR and performance. This underscores the importance of sentiment analysis as a tool to estimate and enhance this positive perception.

Furthermore, the rise of social media and online platforms has provided HR departments with vast amounts of unstructured data, such as employee reviews, comments, and discussions. Sentiment analysis tools have become indispensable in mining these vast datasets to extract meaningful insights about employee sentiments, concerns, and suggestions (Sudhir & Deshakulkarni Suresh, 2021).

Employing sentiment analysis in HR offers numerous paramount benefits, transcending traditional methods and ushering in a new era of efficient human resource management. For starters, traditional methods, like surveys or interviews, have been known to suffer from time lags. In contrast, sentiment analysis delivers real-time insights into employee sentiments. This immediacy enables HR professionals to promptly address any concerns, as highlighted by Rocca, Giacomini, & Zola (2020).

Another compelling advantage of sentiment analysis is its capacity for objective analysis. Whereas human biases might infiltrate conventional sentiment assessment methods, sentiment analysis ensures that evaluations of employee feelings remain impartial. Such an approach guarantees insights that are both accurate and devoid of personal biases, a sentiment echoed by Rajan et al. (2021).

Further, sentiment analysis steers HR departments towards data-driven decision-making. Instead of relying on intuition, HR professionals can base their decisions on tangible data. As Sudhir & Deshakulkarni Suresh (2021) emphasize, this data-centric method aligns HR strategies and interventions closely with the genuine needs and sentiments of employees.

Additionally, monitoring employee sentiments continuously leads to enhanced employee engagement. Organizations that heed to sentiment feedback create avenues for a more engaged workforce. While positive feedback affirms successful HR practices, negative feedback acts as a beacon, guiding HR towards areas needing improvement, a perspective shared by Choi (2019).

From an economic standpoint, sentiment analysis tools, particularly automated ones, can be more wallet-friendly in the long haul than their traditional counterparts. As pointed out by Rajan et al. (2021), these tools trim computational expenses and can sift through vast data pools with minimal human intervention.

Beyond these immediate benefits, sentiment analysis also holds strategic significance. As Liu (2017) suggests, the insights procured from such analysis can serve as cornerstones for HR's strategic planning. Recognizing the broad strokes of sentiment trends equips HR with the foresight needed for crafting effective policies, curating training programs, and spearheading organizational change.

In our study, we performed sentiment analysis on data sourced from the internet, particularly from indeed.com. This endeavor necessitated web content mining, which is fundamental to the process (Atanasova et al., 2010). Data mining, as described by Bâra & Lungu (2012), is the extraction of patterns, relationships, and insights from vast datasets, offering valuable insights for businesses.

Sentiment analysis, at its core, leverages data mining techniques. Stancu et al. (2020) emphasized the importance of data extraction in improving decision support systems. By integrating sentiment analysis and data mining, companies can utilize textual data to enhance decision-making, thereby gaining a competitive edge. This approach is not limited to business analytics but extends to financial market research and decision-making.

The advancements in technology, especially in the domains of machine learning and big data, have augmented the capabilities of predictive analysis (Stoica & Stoica 2021). For instance, when sentiment analysis is fused with big data analytics, it has the potential to inform stock market predictions. Yet, it's important to note that the impact of sentiment analysis isn't uniformly effective across all industries.

In the financial sector, sentiment analysis offers a more comprehensive understanding of market sentiment than traditional financial metrics alone, as highlighted by Cristescu et al. (2022). Beyond finance, Vancea et al. (2017) mentioned that political events can influence financial markets. Such events might also affect employee sentiment about their workplace and societal roles, as well as corporate reactions to these political shifts.

Thus, sentiment analysis is not merely a tool; it's a transformative mechanism for HR professionals. It holds the key to decoding the pulse of an organization. By harnessing its multifaceted benefits, HR departments can sculpt a workplace that's not only harmonious and engaged but also brimming with productivity.

USEFUL OPEN-SOURCE TOOLS FOR SENTIMENT ANALYSIS

In data science, open-source tools help organizations understand and use their data (Provost, F., & Fawcett, T., 2013) and they can become pivotal in understanding employee sentiments. Python modules such as Pandas, NumPy, and scikit-learn, along with R and its packages, offer pliable data processing and visualization for HR sentiment analysis. The affordability of open-source technologies, in comparison to pricey proprietary software, allows HR professionals to employ advanced methods, algorithms, and models in their sentiment analysis tasks. Owing to the collaborative nature of the open-source community, these tools are routinely updated, enabling HR departments to stay updated with the latest advancements in sentiment analytics (Peng, R. D., & Matsui, E., 2016). By leveraging these open-source tools, HR professionals can derive meaningful insights into employee sentiments, aiding in informed decision-making and enhancing workplace culture.

Free platforms like R, Python, and Weka have gained traction in HR sentiment analytics due to their accessibility and versatility (Wickham & Grolemund, 2017).

Various techniques exist for sentiment analysis in HR contexts. One method is utilizing TextBlob, esteemed for its simplicity, which operates on a lexicon-based sentiment foundation. An alternative method is the adoption of the BERT model, an avant-garde, context-sensitive tool, showing expertise in discerning complex textual dynamics (Devlin et al., 2018; Loria, n.d.; Sun et al., 2019).

METHODOLOGY

The primary data for this research was collected from indeed.com, where 150 reviews were amassed. The reviews were specific to three companies: Lidl, Kaufland, and Carrefour in Romania. Once the reviews were collected, they were arranged in an Excel file with separate sheets for each company, ensuring a structured format for data processing.

For sentiment analysis, the open-source library TextBlob was chosen due to its efficiency and ease of use in processing textual data (Loria, n.d.). TextBlob provides a simple API to return polarity and subjectivity of a text, which helps in understanding its sentiment value.

Before the actual sentiment analysis could be performed, the Excel file containing the reviews was read into a Python environment using the pandas library.

Upon completion of sentiment analysis for all reviews across the sheets, the modified data, now containing the original reviews alongside their respective sentiment polarity scores, was written back to a new Excel file named 'reviews_with_sentiment_textblob.xlsx'. This ensured that the processed data was preserved for any future references or analyses.

RESULTS

The script starts by importing the necessary Python libraries as seen in Figure 1.

```python
import pandas as pd
from textblob import TextBlob
import matplotlib.pyplot as plt
```

Figure 1. Importing the necessary Python libraries

Following the imports, the Excel file (reviews.xlsx) containing the reviews distributed across three sheets corresponding to the three companies is read into the environment via the script presented in figure 2.

```python
def main():
    # Read Excel file with multiple sheets
    xls = pd.ExcelFile('reviews.xlsx')
    average_polarities = {}
```

Figure 2. Reading the Excel file

Figure 3 presents a dedicated function, analyze_sentiment, that was created to evaluate the sentiment polarity of the provided review text using TextBlob.

The polarity value, ranging between -1 and 1, denotes the sentiment's nature, with values closer to -1 being negative, values closer to 1 being positive, and values around 0 being neutral.

```python
def analyze_sentiment(text: str):
    analysis = TextBlob(text)
    return analysis.sentiment.polarity

def analyze_reviews_for_sheet(sheet_name, xls):
    reviews = xls.parse(sheet_name)
    polarities = [analyze_sentiment(review) for review in reviews['Review']]
```

Figure 3. Functions to analyze the sentiment of the texts

Also, as seen in figure 3, a function named analyze_reviews_for_sheet is developed to facilitate sentiment analysis for reviews on each sheet.

For each sheet or company, the script computes the sentiment polarity for every review. After obtaining the polarities for all reviews in a sheet, the average polarity is calculated to give a consolidated sentiment score for the reviews of that specific company.

As seen in figure 4, the script includes a visualization section where the average sentiment polarities of all three companies are plotted as a bar chart, offering a comparative view:

```python
    # Plotting average sentiment polarity of all sheets
    plt.bar(average_polarities.keys(), average_polarities.values())
    plt.title('Average Sentiment Polarity by Sheet')
    plt.ylabel('Average Polarity')
    plt.xticks(rotation=45, ha='right')
    plt.tight_layout()
    plt.show()
```

Figure 4. Plotting the average sentiment polarity

Once sentiment analysis was performed, the resulting polarities were integrated into the original dataset and into a new Excel file via the script presented in figure 5.

This allowed for easy cross-referencing between the review text and its corresponding sentiment score.

```python
# Analyze sentiment for each sheet and save results to a new Excel file
with pd.ExcelWriter('reviews_with_sentiment_textblob.xlsx') as writer:
    for sheet_name in xls.sheet_names:
        analyzed_reviews, avg_polarity = analyze_reviews_for_sheet(sheet_name, xls)
        average_polarities[sheet_name] = avg_polarity
        analyzed_reviews.to_excel(writer, sheet_name=sheet_name, index=False)
```

Figure 5. Writing the results in a new Excel file

As can be seen in figure 6, the new Excel file contains a new „Polarity" column.

	A	B	C
1	Review	Polarity	
2	The employer is ok, serious and the v	0,122222222	
3	Unfortunately, you have people who	0,191146245	
4	Regularity, professionalism, kindness	0	
5	The worst experience and the most s	-0,157142857	

Figure 6. The resulted excel file

The bar chart, rendered by the script and presented in figure 7, provided a visual comparative sentiment assessment across the three companies.

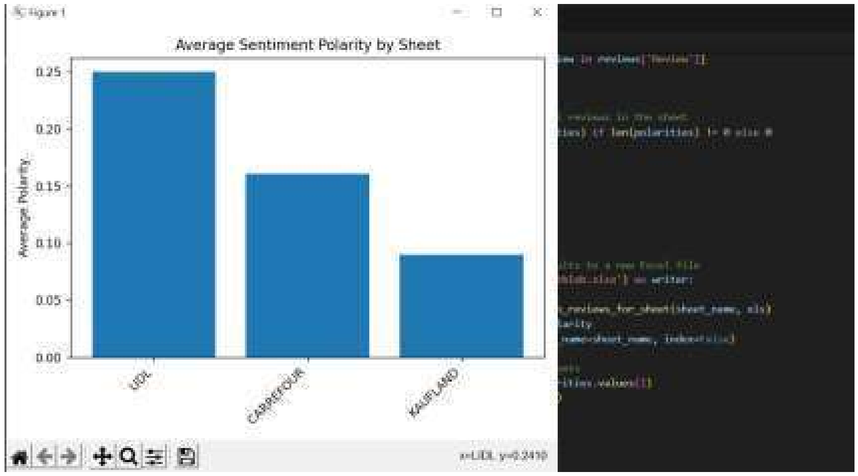

Figure 7. Visualising the sentiments of the employee reviews for the three companies

Thus, the script for sentiment analysis processed the evaluations effectively, and the results were systematically stored and displayed. Both the Excel file and the bar chart are tangible outcomes of this analysis, aiding in assessing and comparing employee sentiments for the three Romanian retail companies.

CONCLUSION

The significance of sentiment analysis in various domains has been increasingly recognized in recent years. This study's exploration of open-source tools for sentiment analysis, specifically focusing on employee feedback from three companies, aligns with the broader trend in the research community. Several studies have delved into the effectiveness, reliability, and application of sentiment analysis tools.

For instance, Novielli et al. (2020) evaluated the performance of sentiment analysis tools in a software engineering context and found that lexicon-based tools often outperform supervised approaches in certain settings. This underscores the importance of choosing the right tool based on the specific context and data at hand. Another study by Rocca et al. (2020) highlighted the potential of sentiment analysis in supporting environmental reporting, emphasizing its cost-effectiveness and speed compared to traditional surveys Rocca et al., 2020.

However, it's crucial to approach sentiment analysis with caution. Boukes et al. (2019) concluded that off-the-shelf sentiment analysis tools might be unreliable for specific languages and domains, emphasizing the need for manual validation. This resonates with the methodology of our study, where we tailored a Python script specifically for employee feedback.

In the area of open-source software, sentiment analysis has been utilized frequently, with numerous tools utilizing support-vector machines. This further validates our decision to conduct our analysis using the open-source tool TextBlob.

Finally, while sentiment analysis is a strong tool for interpreting emotions and views from textual data, the instrument and its application must be context-specific. Our research contributes to this growing field by showcasing the potential of open-source tools in analyzing employee feedback, emphasizing the importance of continuous feedback mechanisms in organizational settings. Future research could focus on developing these tools for even more particular scenarios or investigating their interaction with other data analysis techniques.

REFERENCES

ATANASOVA, T., KASHEVA, M., SULOVA, S., & VASILEV, J. (2010). Analysis of the possible application of Data Mining, Text Mining and Web Mining in business intelligent systems. p. 1297.

BOUKES, M., VAN DE VELDE, B., ARAUJO, T., & VLIEGENTHART, R., (2020). What's the Tone? Easy Doesn't Do It: Analyzing Performance and Agreement Between Off-the-Shelf Sentiment Analysis Tools. Communication Methods and Measures, 14(2), pp. 83–104.

CHOI, J., (2019). What one thinks determines one's actions: The importance of employees' perception in implementing HR systems. Asia Pacific Journal of Human Resources, 57(1), pp. 85–102.

CRISTESCU, M. P., NERISANU, R. A., MARA, D. A., & OPREA, S.-V. (2022). Using Market News Sentiment Analysis for Stock Market Prediction. Mathematics, 10(22), Article 22.

KUMAR, R. S., SAVIOUR DEVARAJ, A. F., RAJESWARI, M., JULIE, E. G., ROBINSON, Y. H., & SHANMUGANATHAN, V., (2022). Exploration of sentiment analysis and legitimate artistry for opinion mining. Multimedia Tools and Applications, 81(9), pp. 11989–12004.

LIU, B., (2017). Many Facets of Sentiment Analysis. A Practical Guide to Sentiment Analysis, 5, pp. 11–39.

NOVIELLI, N., CALEFATO, F., DONGIOVANNI, D., GIRARDI, D., & LANUBILE, F., (2020). Can We Use SE-specific Sentiment Analysis Tools in a Cross-Platform Setting? Proceedings of the 17th International Conference on Mining Software Repositories, pp. 158–168.

ROCCA, L., GIACOMINI, D., & ZOLA, P., (2021). Environmental disclosure and sentiment analysis: State of the art and opportunities for public-sector organisations. Meditari Accountancy Research, 29(3), pp. 617–646.

SUDHIR, P., & SURESH, V. D., (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. Global Transitions Proceedings, 2(2), pp. 205–211.

VANCEA, C., KAMER-AINUR, A., & DUHNEA, C. (2017). Political uncertainty and volatility on the financial markets—the case of Romania. Transformations in Business and Economics, 16, 457.

Monographs:

PENG, R. D., & MATSUI, E., (2016). The Art of Data Science. A Guide for Anyone Who Works with Data. Skybrude Consulting, LLC.

PROVOST, F., & FAWCETT, T. (2013). Data Science for Business. O'Reilly Media, Inc.

WICKHAM, H., & GROLEMUND, G., (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc.

Articles in conference proceedings:

BÂRA, A., & LUNGU, I. (2012). Improving Decision Support Systems with Data Mining Techniques. In A. Karahoca (Ed.), Advances in Data Mining Knowledge Discovery and Applications. InTech.

DEVLIN, J., CHANG, M. W., LEE, K., & TOUTANOVA, K., (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, pp. N/A.

STANCU, A.R., POMPILIU, C.M., & STOYANOVA, M. (2020). Data Mining Algorithms for Knowledge Extraction. In Springer Proceedings in Business and Economics. Springer.

STOICA, F. & STOICA, L. F. (2021). Integrated Tool for Assisted Predictive Analytics. In D. Simian & L. F. Stoica (Eds.), Modelling and Development of Intelligent Systems. Springer.

SUN, C., QIU, X., XU, Y., & HUANG, X., (2019). How to fine-tune BERT for text classification? China National Conference on Chinese Computational Linguistics, pp. 194-206.

Internet sources:

LORIA, S., (n.d.). TextBlob: Simplified Text Processing. [online]. [2023-09-18]. Accessible: https://textblob.readthedocs.io/en/dev/