NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF
ECONOMICS

FACULTY OF SOCIO-ECONOMIC AND COMPUTER SCIENCES
SOFTWARE ENGINEERING

PROJECT PROPOSAL

SITE DEVELOPMENT FOR AUTOMATIC COLLECTION, ANALYSIS
AND VISUALIZATION OF COMPANY ETHICAL BEHAVIOR

Roman Solomatin, SE-19-1

Supervisor: PHD, Professor of
the Department of Information
Technologies in Business Perm
HSE, A. V. Buzmakov,

PERM

2023

Abstarct

Introduction

Background. The ethics of companies have long been of concern to stakeholders, particularly with regard to their actions in contentious situations and their delivery of customer-centric services. In recent years, there has been a growing emphasis on evaluating the ethical standards of companies, particularly in the banking sector and through the lens of Environmental, Social and Governance (ESG) factors(Murè, Spallone, Mango, Marzioni, & Bittucci, 2021, Miralles-Quirós, Miralles-Quirós, & Redondo Hernández, 2019, Climent, 2018). The need for such assessments has become increasingly urgent as society continues to grapple with the consequences of corporate misconduct and the broader impact of corporate activities on society as a whole.

Assessing a company's ethical standards is a complex process that involves evaluating various aspects of a company's operations, such as its business practices, policies, and overall culture. In addition to traditional financial metrics, ESG factors play a critical role in determining a company's overall ethical standing. All of this provides valuable insight into a company's ethical standards.

Currently, there are a number of services that claim to evaluate a company's ethics, but these evaluations are often based on court cases and other official records rather than customer feedback. This has resulted in a scenario where individuals must conduct their own research to determine the ethics of a particular company. This research often involves reviewing customer feedback from various websites, which can be time-consuming and may not always provide a comprehensive or accurate picture of a company's ethical practices.

To address this issue, there have been recent calls for the development of a system that would collect and analyze customer feedback from multiple websites to provide a more comprehensive view, and this system could also incorporate other data sources such as financial reports, sustainability reports, and news articles to provide a more holistic view of a company's ethical practices. The system could also incorporate machine learning algorithms, such as sentiment analysis and topic modeling, to analyze customer feedback and extract valuable insights. These insights could then be used to generate a score or rating that provides an overall assessment of a company's ethical practices.

The field of natural language processing has seen a significant advancement in recent years, largely due to the emergence of neural network-based algorithms. These algorithms represent text data in a more nuanced and complex manner, allowing for a deeper understanding of the underlying semantics and meaning. They can help to analyze semantics of texts.

Literature Review

Etics Measurment

The importance of ethical behavior in business cannot be overstated. As evidenced by various studies and research (Climent, 2018, Murè, Spallone, Mango, Marzioni, & Bittucci, 2021), companies that prioritize ethical behavior tend to achieve greater financial success and better business performance over the long term than those that engage in unethical practices.

Evaluating a company's ethical standards can be approached from several perspectives. From the perspective of the company itself, various factors can be considered, including the level of capitalization to ensure that it is not at risk of bankruptcy, the impact it has on the surrounding environment, and the direction of its investments (Harvey, 1995). On the other hand, customers may focus on the quality of customer service (Brunk, 2010) and the degree to which a company's services are intrusive (Mitchell, Lewis, & Reinsch, 1992).

It is important to note that evaluating a company's ethics is not a one-time endeavor. Rather, it is an ongoing process that requires continuous monitoring and evaluation of the company's actions, policies and practices over time. This is particularly important given the constantly evolving nature of business conduct and the need to stay abreast of emerging issues and trends. In addition, it is important to consider the broader societal impacts of corporate activities and to evaluate companies not only on their financial performance, but also on their environmental, social and governance (ESG) factors.

Text Analyis

Machine learning algorithms for text analysis have been widely used to extract information from unstructured data using large annotated datasets. Among the various methods used, several algorithms have proven to be particularly effective in this area. These include the bag of words (Harris, 1954), TF-IDF (Jones, 1972), Word2Vec(Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) , ELMO (Peters, Neumann, Iyyer, Gardner, Clark, Lee, & Zettlemoyer, 2018), GPT (Radford, Wu, Child, Luan, Amodei, & Sutskever, 2019), and BERT (Devlin, Chang, Lee, & Toutanova, 2018). Each of these algorithms has unique characteristics that make it well suited for specific applications.

The bag of words model represents text data by assigning a unique number to each word in a document. This method is easy to implement, but does not take into account the order of words in a sentence. On the other hand, the TF-IDF model represents text data by considering both the frequency of a word in a document (TF) and its rarity across all documents in the corpus (IDF). This approach can be used to determine the importance of a word in a given document and is commonly used in information retrieval and natural language processing tasks, but these algorithm do not understand full context of words.

Word2Vec utilizes a vector representation of words, which enables the algorithm to capture the meaning of words in similar contexts. This allows for a more accurate and sophisticated representation of the relationships between words, leading to improved performance in tasks such as text classification and sentiment analysis.

ELMO, GPT, and BERT, on the other hand, are based on the transformer architecture, in which each sentence is represented by a vector of numbers, commonly known as an embedding. This representation allows for a more comprehensive and holistic understanding of the text, as it takes into account the context of the entire sentence or text.

Of these algorithms, BERT is considered to be the most advanced and powerful(Devlin, Chang, Lee, & Toutanova, 2018), as it is able to consider the context of the entire sentence or text, whereas GPT and ELMO only consider a one-sided context.

This allows BERT to achieve state-of-the-art performance in a wide range of NLP tasks, including text classification, named entity recognition, and question answering.

Results Anticipated

Conclusion

References

Brunk, K. H. (2010). Exploring origins of ethical company/brand perceptions—a consumer perspective of corporate ethics. *Journal of Business Research, 63*(3), 255–262.

Climent, F. (2018). Ethical versus conventional banking: A case study [Number: 7 Publisher: Multidisciplinary Digital Publishing Institute]. *Sustainability, 10*(7), 2152.

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*.

Harris, Z. S. (1954). Distributional structure. *WORD, 10*(2-3), 146–162.

Harvey, B. (1995). Ethical banking: The case of the co-operative bank. *Journal of Business Ethics, 14*(12), 1005–1013.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems, 26*.

Miralles-Quirós, M. M., Miralles-Quirós, J. L., & Redondo Hernández, J. (2019). ESG performance and shareholder value creation in the banking industry: International differences [Number: 5 Publisher: Multidisciplinary Digital Publishing Institute]. *Sustainability, 11*(5), 1404.

Mitchell, W. J., Lewis, P. V., & Reinsch, N. (1992). Bank ethics: An exploratory study of ethical behaviors and perceptions in small, local banks. *Journal of Business Ethics, 11*(3), 197–205.

Murè, P., Spallone, M., Mango, F., Marzioni, S., & Bittucci, L. (2021). ESG and reputation: The case of sanctioned italian banks [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/csr.2047]. *Corporate Social Responsibility and Environmental Management*, *28*(1), 265–277.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.