

Summary of:

title={Identification of fraudulent financial statements using linguistic credibility analysis}, author={Humpherys, Sean L and Moffitt, Kevin C and Burns, Mary B and Burgoon, Judee K and Felix, William F}, journal={Decision Support Systems},

The authors find 101 known fraudulent 10-K filings, and 101 likely non-fraudulent 10-K filings, and use Agent99 Analyzer to extract features from the texts of the MD&A sections. Statistical and Machine Learning algorithms are found to classify the fraudulent and non-fraudulent filings more accurately than human judges typically perform in deception detection tasks. It is also found that the false positive and false negative rates are more balanced than for human judges, indicating less bias from the statistical and machine learning algorithms. The authors discuss a wide range of approaches to analysis and detection of deception, outlining the underlying assumptions of each approach. Some approaches focus more on the deceiver's intentional additions that are intended to deceive, whereas others focus on the unintentional 'leakage' of information arising from the deceiver's attempt to deceive. The authors note that larger companies are likely to afford more resources to producing misleading or deceptive MD&As. The research study itself may serve fraudulent companies in adjusting their methods of deception. However, believing forcing companies to dedicate more resources in order to successfully deceive is a societal good. It means that more people within the company will be privy to the deception, which increases the chance of whistleblowers stepping forward, and it simply increases the cost of deception, which discentivises it. The linguistic/textual features extracted primarily use counts or ratios of word categories in the text as proxies for metalinguistic properties such as syntactic/lexical complexity, verbosity, imagery, deference, uncertainty.