This paper examines the sentiments of different threats to employment. Specifically, this paper studies twitter data, using a data set of 150.000 twitters. For each topic, 30.000 tweets are exacted, using keywords. Subsequently, the tweets are cleaner. Keywords, numbers, websites, special signs, symbols and whitespaces are removed. Subsequently, stemming is used to derive the stem of the words. By doing so, one can categories more related words, allowing for a more accurate match with the lexicon. For each topic, the frequency plots of three lexicons are studied. Specifically, “nrc”, “afinn” and “bing” are the three lexicons used here. “nrc” allows for the highest degree of specification when it comes to sentiments. It provides insights into “fear”, “joy”, “positive”, “negative”, “trust”, “disgust”, “anger”, “anticipation”, “surprise”, and “sadness”. “afinn” examines values from

One can use either the lexicon method or a machine learning based text classification tool.

In the first section there is a benchmark analysis of 3 popular lexicon-based text classification tools. For our further descriptive statistics, we focus on the “nrc” lexicon because it provides the most versatile and holistic sentiment framework relative to “afinn” and “bing”.

There a numerous method to undertake text classification. Here we are focusing on sentiment analysis and topic modelling. For sentiment analysis we either need supervised machine learning or a lexicon method without machine learning. For the topic modelling we need unsupervised machine learning.

The study provides benchmark results for the different methods for the different topics. Further, for each topic sentiments, frequency, and polarization are analyzed. Additionally, Lochard Ditricht Allocation (LDA) and text classifiers for each topic are used. Last but not least, we provide a basic illustration of how to create a twitter bot algorithm. Future research could examine topic specific bots.

Frequency plots: Plots with the most frequent words for both positive and negative sentiments.

LDA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| Exposure | -0.22\*\*\*  (0.03) | -0.16\*\*\*  (0.03) | -0.04\*\*\*  (0.01) | -0.14\*\*\*  (0.02) |
| Offshorability | -2.29\*\*\*  (0.50) | 2.02\*\*\*  (0.55) | -0.87\*\*\*  (0.28) | 2.66\*\*\*  (0.53) |
| Medium education | 9.52\*\*\*  (1.67) | -1.20\*\*\*  (1.54) | 11.80\*\*\*  (0.93) | 6.19\*\*\*  (1.35) |
| High Education | 27.73\*\*\*  (2.01) | 14.42\*\*\*  (2.40) | 32.75\*\*\*  (1.24) | 22.26\*\*\*  (1.91) |
| Wage | -0.07\*\*\*  (0.01) | 0.04\*\*\*  (0.00) | -0.07\*\*\*  (0.00) | 0.03\*\*\*  (0.00) |
| Wage squared | 0.00\*\*\*  (0.00) | -0.00\*\*\*  (0.00) | 0.00\*\*\*  (0.00) | 0.00\*\*\*  (0.00) |
| R2 – Adjusted  Industry FEs  Observations | 0.163  Yes  6,708 | 0.147  Yes  14,065 | 0.168  Yes  18,975 | 0.210  Yes  36,070 |

All regressions control for industry fixed effects. The first regression shows the exposure of robots on wages. One can see a consistently significant, negative effect on wages. Offshorablility also provides a negative coefficient. The education variable is split into two categorical variables. Moreover, the wage variable is also split into wage and wage squared.

The second regression examines the exposure of robots on employment.

The number of observations is significantly higher for the scores of software on wages and employment.

Irrespective, all coefficients are significant with a p-value of .

**Text Classification Benchmarking**

To further illustrate robustness of the findings, we used supervised machine learning to detect sentiments in the tweets. Specifically, this section will show the usage of the random forest, naïve Bayes and linear tf-tif model.

**LDA Method**

Latent Dirichlet Allocation (LDA) is a predominant method within the field of topic modelling for text mining. LDA is an unsupervised machine learning method, hence a method used for prediction of an un-specified pattern.

Math model:

For robustness, other methods such as LSA and (XXX) are illustrated.

The controversial release of Capital in the 21st century by Thomas Piketty (2013), provided a narrative for the dynamics of inequality. This book was heavily inspired by the general laws of capitalism and historical materialism introduced by Karl Marx, providing a theory of history suggesting that our material conditions will shape our history. Piketty argues that capital accumulation and technology will disrupt the forces of production and economic/political life resulting in a more inequal society (Acemoglu and Robinson, 2014). Acemoglu and Robinson (2014) argue that this theory fails to incorporate the endogenous evolution of technology, accommodating for the fact that technology and its impact are severely dependent on the accompanying institutions and politics. Further, they claim that one needs to disentangle the multifaceted nature of technology.

Further research by Acemoglu and Restrepo (2018) provide such an analysis. They point towards numerous countervailing forces of technological change. Specifically, they emphasize the role of the displacement effect, Productivity effect and excessive automation. The displacement effect suggests that workers are replaced through automation, because technology occupies their prior positions. Inevitable this suggest a disentanglement of the relationship between wages and output, with a declining share of labor.

The productivity effect claims that due to the reduction in production cost, economic expansion will result in accelerated demand for labor for non-automated tasks. Acemoglu and Restrepo argue that this embodiment could emerge in non-automated sectors or the sector which is undergoing automation.

Moreover, Acemoglu and Restrepo suggest that automation occurs at both the extensive margin, replacing tasks performed by workers, and the intensive margin, replacing tasks performed by machines, deepening automation. They argue that this would also lead to further productivity but countervail the displacement effect. Following this argument, the reinstatement effect, the emergence of novel tasks, take the same direction. Subsequently, Acemoglu and Restrepo discuss the mismatch between technology and skills. This mismatch is created by a lack of accommodation of new technologies in the educational system, creating a lag in productivity.

To complement this research, two hypotheses of former research are introduced (Acemoglu and Restrepo ,2016). Firstly, there is crowding out of growth opportunities because of a unidirectional focus on specific technologies. Secondly, the authors argue that excessive automation, caused by misguided US tax incentives has led to the socially disruptive innovation, deteriorating productivity growth.

Following this debate, Webb (2020) suggests that the impact of inequality deviates between different technologies. Historical analysis of software and robots illustrate evidential support robots have challenged the numerous low skill jobs. This view is supported by e.g. Ayres and Miller (1983) and Acemoglu and Restrepo (2017), claiming that jobs such as manufacturing (machining, welding, painting, assembly) have been exposed to industrial robots since the sizable breakthrough in the 1980s. Furthermore, Webb suggests that software has challenges jobs of middle-income workers. For explanation Webb points here to Autor, Levy and Murnane (2003).