

H2D – COMPANY SEARCH USING TEXT-MINING AND IDENTIFICATION OF REGIONAL CHARACTERISTICS

Yuri Campbell, Karl Trela, Frank Pothen, Laura Brock (Research Fellows)
 Miriam Armin, Satiyabooshan Murugaboopathy (Research/Student Assistants)

Text-Mining Firms in the Hydrogen Economy

Establishing a green hydrogen economy is viewed as being instrumental in achieving carbon neutrality by 2050. By splitting water into hydrogen and oxygen, electrolyzers are components of a well-functioning hydrogen economy. To identify regional factors that have an impact on the growth of the hydrogen economy several steps were performed. First we conducted an initial search for producers of hydrogen electrolyzers and companies related to hydrogen in general. From these, we derived an algorithm that would rank 1.1 million companies from the Hoppenstedt company database according to their likeliness to produce electrolyzers. For this, we gathered a sampling of homepages from companies with the highest and lowest likelihood of being related to hydrogen in general. Following this, we used the lower ranked companies and the anchor examples to train a text classification model. The trained classifier scored the candidate companies according to the likelihood that the company was involved with hydrogen electrolysis. A detailed investigation of the high ranked firms can return precise results.

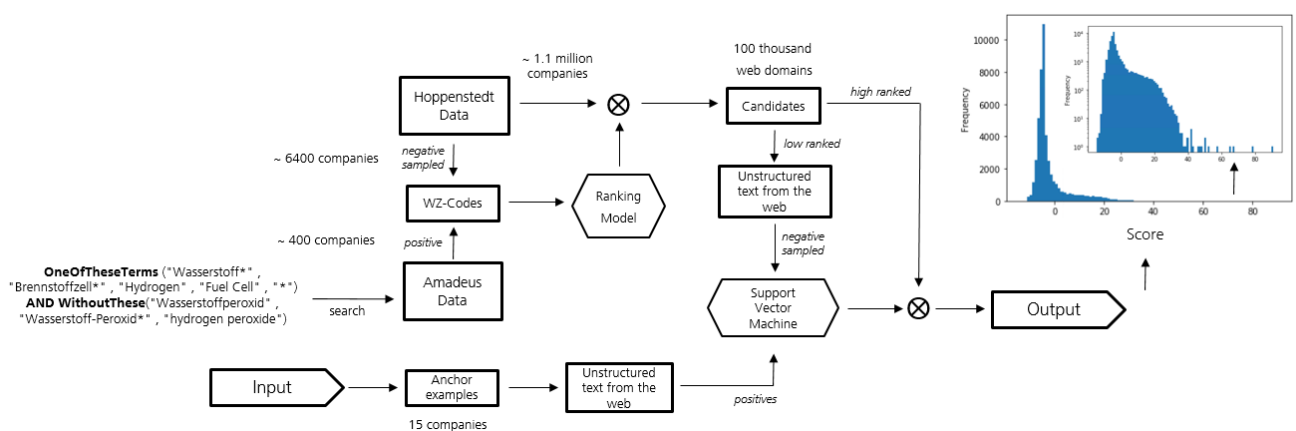


Figure 1: Hoppenstedt & Web-based Approach

Hydrogen and Electrolysis in Germany

By using an analysis score of at least 30, we identified 423 companies that could be in the hydrogen electrolysis business. We visualised these on the map of Germany to the right. Hydrogen Electrolyser firms are mainly found in West Germany with locations forming a crescent spanning from Hamburg over Cologne, Frankfurt and Stuttgart to Munich. Eastern Germany is underrepresented, but exhibits smaller clusters in Berlin and Saxony. An analysis over time of the number of firms may result in interesting findings for the development of individual clusters.

Location Choices of Hydrogen Electrolyser Firms

In order to identify regional characteristics that have an impact on the number of firms in the industry, we correlated the normalised number of firms given by the Hoppenstedt algorithm with an analysis score of over 50 to different economic factors in the different states over a ten year period from 2010-2019. Figure 3 displays the number of firms per municipality adjusted for population size in relation to GDP, the unemployment rate, turnover from chemical products and turnover from metal production and processing for the year 2019

The below graphs display a regression line that uses the Poisson distribution to estimate a generalised linear model. The p value for the Z-statistic indicates that the presence of the chemical industry and the metal production and processing industry are significant factors that determine the presence of a green hydrogen industry. Using the Nagelkerke method to determine the pseudo R2 values for the respective regressions we obtain an $R^2 = 0.058$ for the regression for the turnover on chemicals, an $R^2 = 0.034$ for the regression on the turnover for metal production and processing, an $R^2 = 0.006$ for the regression against the unemployment rate and an $R^2 = 0.045$ for the regression against GDP. This indicates that these factors are relevant but we have not found the key determinants for the emergence of the green hydrogen industry.

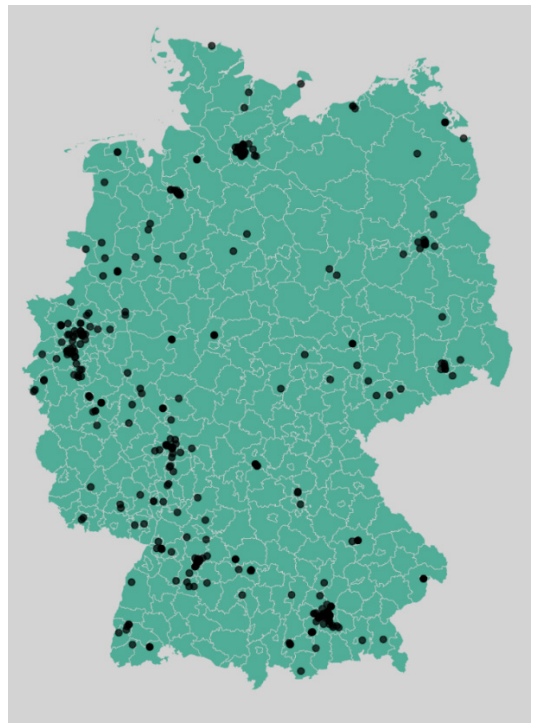


Figure 2: Distribution of Electrolysers

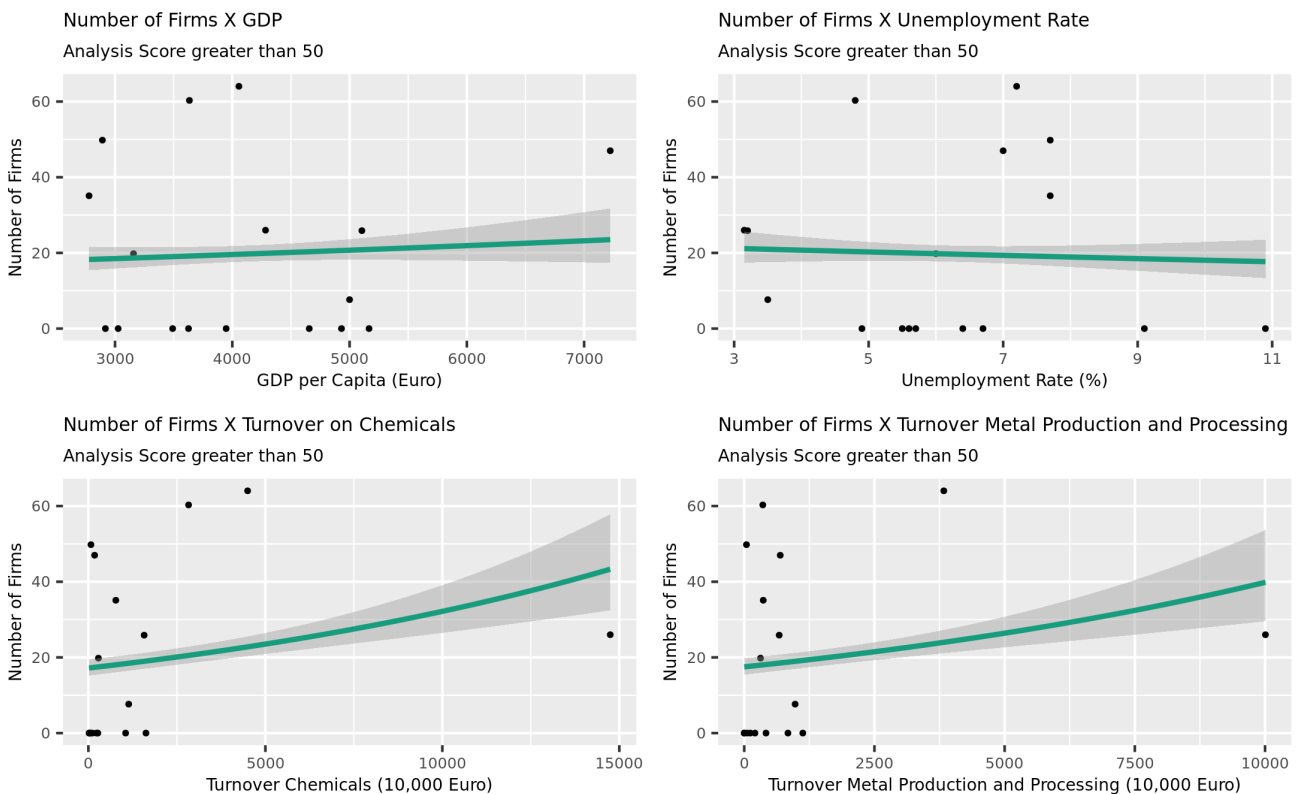


Figure 3: Poisson Regression of Location Factors