So far, models based on several financial ratios have been created to evaluate the financial situation of businesses operating in Visegrad group (Czech Republic, Hungary, Poland and Slovakia) and the models have been verified based on small research sample. The authors of these models called them early warning systems. Korol (2010) called this a kind of abuse. The research will be used not only financial and non-financial ratios, but also off-balance sheet data.

In the literature, there are many works related to textual analysis but most of them concern US companies (see, e.g., Li, 2006, 2010; Davis et al., 2011; Ferris et al., 2012; Dyer et al., 2017; Mai et al., 2019). Few works study non-US companies, especially Central and Eastern European Countries, using textual analysis (e.g. Lang, Stice-Lawrence, 2015), one concerns Denmark (Matin et al. 2019). This may be due to the availability of mandatory filings and disclosures (e.g., 10-Ks, 10-Qs, 8-Ks annual reports, IPO prospectuses, RNS, etc). So, the methods for textual analysis will be assessed for non-US companies on a huge research sample in the project.

The main goal of the project is to develop early warning models for companies operating in the Visegrad Group countries using artificial intelligence. The research contributes to the theory of economy by investigating the possibility of utilizing off-balance sheet data in early warning models for assessing financial situation of a company. The project covers two areas:

1. Building up early warning models using artificial intelligence, which aim at the assessment of the impact of non-financial indicators on the level of correct recognition of financial situation of companies. In the research, among others Boosted trees trained with Extreme Gradient Boosting (XGB) will be applied.

2. Investigating the possibility of utilizing off-balance sheet data in early warning models for assessing financial standing of a company using Long Short Term Memory networks (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) models.

The project will extension of the research conducted by Zięba, Tomczak, Tomczak (2016). The project will specify at least 90 financial ratios and several non-financial ratios, such as the business sector, place of business (voivodship, city), legal form, year of establishment, number of employees. From 435000 available unique companies in the database at least 100 000 companies will be analyzed from the Visegrad group countries, including still operating businesses, enterprises have been liquidated and / or which have gone bankrupt, economical bankrupt companies. The sample covers both large, medium-sized and small businesses. Only those enterprises with access to off-balance-sheet data will be selected.

Next, off-balance sheet information will be used for the assessment of financial standing of businesses that are most useful for the point of view of managers and potential investors. It is estimated that approximately 80% of information is stored in the form of text documents. This fact creates the need to construct tools that support people in obtaining and processing information from text documents, including the development of methods and tools for textual analysis (Lula, 2005).

In the literature, there is lack of studies concerning on the use of text data to evaluate the financial standing of companies from Central and Eastern European Countries as well as others emerging countries. Data from text documents, off-balance information, are very important for potential investors and managers because they provide additional knowledge about the company's situation. Mai et al. (2018) and Matin et al. (2019) as ones of the few, used deep learning models to predict the bankruptcy of enterprises using textual data. In their research, they pointed out the possibility of using LSTM for forecasting. In these studies, models such as LSTM as well as BERT-based models will be constructed in which text information can be included. Python and R will be used in this study. For numerically intensive calculations we will utilize local workstations (supplied by the employing entity – WUST) and – if necessary – resort to the computational power of the WCSS at WUST. Also, for the purpose of training and fine-tunning models, it is planned to use Google Colab platform, that offer a free access to the GPU/TPU resources that significantly speed up the learning process.