

Dotychczas do oceny kondycji finansowej przedsiębiorstw funkcjonujących w krajach Grupy Wyszehradzkiej (do której należą Czechy, Polska, Słowacja i Węgry) tworzono modele oparte na kilku wskaźnikach finansowych i modele są weryfikowane na podstawie niewielkiej próbki badawczej. Autorzy tych modeli nazywali je systemami wczesnego ostrzegania. Korol (2010) nazwał to pewnego rodzaju nadużyciem. A właśnie wskaźniki pozabilansowe znacząco zwiększają sprawność modeli w przewidywaniu upadłości.

W literaturze jest wiele prac związanych z analizą tekstu, ale większość z nich dotyczy firm amerykańskich (np. Li, 2006, 2010; Davis i inni, 2011; Ferris i inni, 2012; Dyer, i inni, 2017; Mai i inni, 2019). Niewiele prac poświęconych jest analizie tekstu w oparciu o przedsiębiorstwa nieamerykańskie, zwłaszcza jeśli chodzi o przedsiębiorstwa z Europy Środkowo-Wschodniej (np. Lang, Stice-Lawrence, 2015), jedna praca dotyczy rynku duńskiego (Matin i in., 2019). Może to być spowodowane dostępnością danych obligatoryjnych (np., 10-Ks, 10-Qs, 8-Ks rocznych raportów, IPO prospekty, RNS, itd.). Dlatego w projekcie zostaną wykorzystane metody analizy tekstu do oceny firm spoza Stanów Zjednoczonych na podstawie dużej próby badawczej.

Głównym celem projektu jest opracowanie modeli wczesnego ostrzegania dla przedsiębiorstw działających w krajach Grupy Wyszehradzkiej przy wykorzystaniu sztucznej inteligencji. W badaniu wyróżnione zostały dwa obszary, które zarówno włączają się w nurt badań podstawowych (teoria ekonomii) jak i stanowią interesujący głos w dyskusji nad wpływem wykorzystania informacji pozabilansowych w rozpoznawaniu sytuacji finansowej przedsiębiorstwa. Są to:

1. Budowa modeli wczesnego ostrzegania z wykorzystaniem sztucznej inteligencji, których celem jest ocena wpływu wskaźników niefinansowych na poziom prawidłowego rozpoznawania sytuacji finansowej przedsiębiorstw. W badaniach zastosowane zostaną między innymi wzmacniane drzewa Extreme Gradient Boosting (XGB).
2. Zbadanie możliwości wykorzystania informacji pozabilansowych w modelach wczesnego ostrzegania do oceny sytuacji finansowej przedsiębiorstwa przy użyciu modeli typu Long Short Term Memory networks (LSTM) oraz Bidirectional Encoder Representations from Transformers (BERT).

Projekt będzie rozszerzeniem badań prowadzonych przez Zięba, Tomczak, Tomczak (2016). W projekcie wyszczególnionych będzie co najmniej 90 wskaźników finansowych i kilka wskaźników niefinansowych, takich jak sektor prowadzonej działalności, miejsce prowadzenia działalności (województwo, miasto), forma prawna, rok założenia przedsiębiorstwa, liczba pracowników. Z dostępnych 435000 przedsiębiorstw w bazie danych przeanalizowanych zostanie co najmniej 100 000 przedsiębiorstw działających w krajach Grupy Wyszehradzkiej, w tym nadal działające przedsiębiorstwa, przedsiębiorstwa, które zostały zlikwidowane i / lub które ogłosiły upadłość, zbankrutowane przedsiębiorstwa w sensie ekonomicznym. Próba obejmuje zarówno duże, średnie i małe przedsiębiorstwa. Wybrane zostaną tylko te przedsiębiorstwa do których będzie dostęp do danych pozabilansowych.

Następnie, informacje pozabilansowe zostaną wykorzystane do oceny finansowej przedsiębiorstw, które są bardziej przydatne z punktu widzenia menedżerów i potencjalnych inwestorów. Szacuje się, że około 80% informacji przechowywanych jest w postaci dokumentów tekstowych. Fakt ten stwarza potrzebę konstruowania narzędzi wspierających ludzi w pozyskiwaniu i przetwarzaniu informacji z dokumentów tekstowych, w tym rozwój metod i narzędzi do analizy tekstu (Lula, 2005).

W literaturze brakuje opracowań dotyczących wykorzystania tekstu do oceny kondycji finansowej przedsiębiorstw z Europy Środkowo-Wschodniej. Dane z dokumentów tekstowych, informacje pozabilansowe, są bardzo ważne dla potencjalnych inwestorów i menedżerów, ponieważ dają dodatkową wiedzę o sytuacji przedsiębiorstwa. Mai i in. (2019) oraz Matin i in., (2019) jako jedni z nielicznych zastosowali głębokie modele uczenia do predykcji upadłości przedsiębiorstw przy wykorzystaniu danych tekstowych. W swoich badaniach zaznaczyli o możliwości wykorzystania LSTM do prognozowania. W obecnych badaniach zostaną skonstruowane modele takie jak LSTM oraz BERT, w budowie których można uwzględnić informacje tekstowe. W tym badaniu zostaną wykorzystane oprogramowania Python i R. Dodatkowo autorzy zakładają możliwość wykorzystania platformy Google Colab, która udostępnia za darmo zasoby typu GPU oraz TPU, w celu przyspieszenia procesu uczenia wybranych modeli.