Zgłoszenie tematu pracy dyplomowej **:: STUDIA II STOPNIA ::**

na rok akademicki 2022/23

Promotor:	**dr hab. Jozef Kapusta, prof. UP**
Temat pracy magisterskiej (j. polski oraz j. angielski):	Comparison of Word Embeddings Methods for Fake News Classification *Porównanie metod osadzania słów w celu klasyfikacji fałszywych wiadomości*
Zakres i oczekiwane rezultaty pracy:	The amount of fake news, which purposely try to manipulate people's opinion, is exponentially increasing. Given the massive amount of Web content, automatic fake news detection is a practical natural language processing and machine learning problem. Machine learning models take vectors (arrays of numbers) as input. When working with text, the first thing you must do is come up with a strategy to convert strings to numbers (or to "vectorize" the text) before feeding it to the model. Word embeddings give us a way to use an efficient, dense representation in which similar words have a similar encoding. In the theoretical part: The thesis aims to summarise and comparison of the "traditional" methods for word embedding (Word2vec, Tf-Idf, GloVe). The author will select word embedding suitable for fake news classification methods. Exploring the new experimental methods for word embeddings from current scientific sources is expected. In the practical part: The practical part of the thesis will focus on creating text classifier and impact assessment word embedding methods for performance measures of the models (accuracy, precision, recall, f1-score).
*Aspekt naukowy, problemowy pracy:	definition and implementation of a documents models, implementation methods natural language, implementation machine learning methods.
Literatura	• Bird, S., Klein E., and Loper, E. (2009). Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media. • Bengfort, B., Ojeda, T., Bilbro, R. (2018). Applied Text Analysis with Python: Enabling Language - Aware Data Products with Machine Learning, O'Reilly Media, 332 p. • Natural Language Toolkit, online: https://www.nltk.org/ • Chen, JT., Tao, YB., and Lin, H. (2018). Visual exploration and comparison of word embeddings. In. Journal of Visual Languages and Computing, Vol. 48, p. 178-186
**Oprogramowanie, język programowania, środowisko systemowe:	Jupyter Notebook Environment (Python)

Zgłoszenie tematu pracy dyplomowej **:: STUDIA II STOPNIA ::**

na rok akademicki 2022/23

**Środowisko uruchomieniowe:	Windows or Linux
Dodatkowe wymagania i uwagi:	English language

.

* Regulamin studiów § 35 2. Praca dyplomowa na profilu praktycznym, podobnie jak praca inżynierska, powinna mieć charakter aplikacyjny, badawczy, projektowy lub oceniający praktykę w świetle teorii.
** pola opcjonalne