

8-10-2020

## **An Analysis of the Effectiveness of Applying a Machine Learning Approach for Classification of Technical Documents in Knowledge Discovery Systems**

Roman Melnyk  
*Nova Southeastern University, romanm321@gmail.com*

Martha Snyder  
*Nova Southeastern University, smithmt@nova.edu*

Alexander Verner  
*KLA, verner.alexander@gmail.com*

Follow this and additional works at: [https://aisel.aisnet.org/treos\\_amcis2020](https://aisel.aisnet.org/treos_amcis2020)

---

### **Recommended Citation**

Melnyk, Roman; Snyder, Martha; and Verner, Alexander, "An Analysis of the Effectiveness of Applying a Machine Learning Approach for Classification of Technical Documents in Knowledge Discovery Systems" (2020). *AMCIS 2020 TREOs*. 58.  
[https://aisel.aisnet.org/treos\\_amcis2020/58](https://aisel.aisnet.org/treos_amcis2020/58)

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# An analysis of the effectiveness of applying a machine learning approach for classification of technical documents in knowledge discovery systems

TREO Talk Paper

**Roman Melnyk**

Nova Southeastern University  
rm1731@mynsu.nova.edu

**Martha M. Snyder**

Nova Southeastern University  
smithmt@nova.edu

**Alexander Verner**

KLA Corporation  
alexander.verner@kla-tencor.com

## Abstract

An important component of knowledge management (KM) is the organization of documents for quick and easy access. One advantageous and effective way of organizing these documents is to group them by a fixed set of specific knowledge categories. For large-scale technical teams, the number of categories can reach thousands or even tens of thousands, which makes the aforementioned cataloging especially useful.

Text classification (TC) is a sophisticated process that involves data pre-processing, transformation, dimensionality reduction, application of classification techniques, classifier evaluation, and classifier validation. TC remains a prominent research topic and still depends on human work rather than on machine learning (ML). It is a relatively new area of research and remains in a premature phase.

The goal is to develop and evaluate a prototype model that uses ML algorithms to classify technical documentation in a KM system for technical teams of financial institutions involved in software development projects. This research contributes to the field of KM by determining whether an ML approach constitutes a feasible solution for TC in knowledge discovery.

## REFERENCES

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *ArXiv preprint arXiv:1707.02919*.
- Fromm, H., Wambsganss, T., & Söllner, M. (2019). Towards a Taxonomy of Text Mining Features. *Twenty-Seventh European Conference on Information Systems (ECIS2019)*, Stockholm-Uppsala, Sweden, (pp. 1-12).
- Kao, A., Quach, L., Poteet, S., & Woods, S. (2003, November). User assisted text classification and knowledge management. *In Proceedings of the twelfth international conference on Information and knowledge management* (pp. 524-527).
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018a). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733-765.
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018b). Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, 21(3), 766-799.