**UDC 004.89**

# VIDEO KEYWORD GENERATION FROM TEXT DESCRIPTION BASED ON ARTIFICIAL INTELLIGENCE

**Mykola Yemets**, student, yemets.m.o@nmu.one,
Dnipro University of Technology
**Kostiantyn Khabarlak**, Ph.D., Assistant Professor,
khabarlak.k.s@nmu.one, Dnipro University of Technology

In the digital age, the exponential growth of video content has led overwhelming need for efficient methods to organize, search, and retrieve this vast amount of information [1]. One of the fundamental challenges in this domain is the accurate annotation of videos with relevant keywords or tags, enabling users to quickly identify and access desired content. Traditionally, this task has been labor-intensive and time-consuming, often relying on manual annotation processes that are not scalable to the ever-expanding volume of video data.

Artificial Intelligence (AI) emerges as a transformative solution to address these challenges, offering capabilities in natural language understanding and semantic inference. Leveraging machine learning algorithms and deep neural networks, AI enables the automated generation of descriptive keywords directly from textual descriptions associated with videos.

The purpose of this work is to develop an innovative application leveraging artificial intelligence techniques to assist users in identifying relevant keywords for a video based on its textual description. By automating the process of keyword generation, the application aims to enhance the efficiency and accuracy of video annotation.

The implementation of this application will utilize the Python Scikit-learn library, leveraging its rich set of tools and functionalities for machine learning tasks. Specifically, the Random Forest and Gradient Boosting methods will be employed to train predictive models for keyword generation from textual descriptions associated with videos. Random Forest and Gradient Boosting are ensemble learning techniques that offer robust performance in handling high-dimensional data and capturing complex relationships between input features and target variables. The core idea is to analyze the textual descriptions of videos and automatically generate corresponding keywords.

During training, the model receives textual descriptions of videos as input and learns the relationship between these descriptions and the keywords that reflect their content. By understanding this relationship, the model can accurately predict keywords based on new textual descriptions it encounters.

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the mode of the classes

(classification) or the mean prediction (regression) of the individual trees.Each tree in the ensemble is built from a randomly sampled subset of the training data and a random subset of features at each split, which introduces diversity among the trees. During prediction, each tree independently provides a class prediction or regression estimate, and the final prediction is determined by aggregating the predictions of all trees. Random Forest is known for its robustness to overfitting, scalability to large datasets, and ability to handle high-dimensional data [3].

Gradient Boosting is also ensemble learning technique that builds a predictive model in a sequential manner by combining the predictions of multiple weak learners, typically decision trees. Unlike Random Forest, where trees are built independently, Gradient Boosting sequentially trains new trees to correct the errors made by the previous ones.

In each iteration, the model fits a new tree to the residuals (the differences between the predicted and actual values) of the previous model, gradually reducing the error with each successive tree [4]. By iteratively improving the model's predictions, Gradient Boosting effectively minimizes loss and produces a strong predictive model [5].

The developing application aims to provide new way users interact with and explore video content by providing an efficient and intuitive solution for keyword generation based on textual descriptions.

By automatically extracting relevant keywords from video descriptions, the application simplifies the process of annotating and categorizing videos, enabling users to quickly identify and access the content they are interested in. This functionality enhances the overall user experience by streamlining content discovery and facilitating more accurate search results, saving users time and effort in navigating through vast collections of multimedia content.

At this stage, we have completed the initial phase of information gathering and conducted preliminary analysis to lay the groundwork for the implementation of machine learning techniques. The researchers analyzed the dataset, which consists of a description of the characteristics of 359,165 videos that were trending on YouTube between 2020 and 2023.

Based on the data obtained, the distribution of all videos by category and the distribution of the number of tags globally and in the most popular categories were determined (fig. 1).

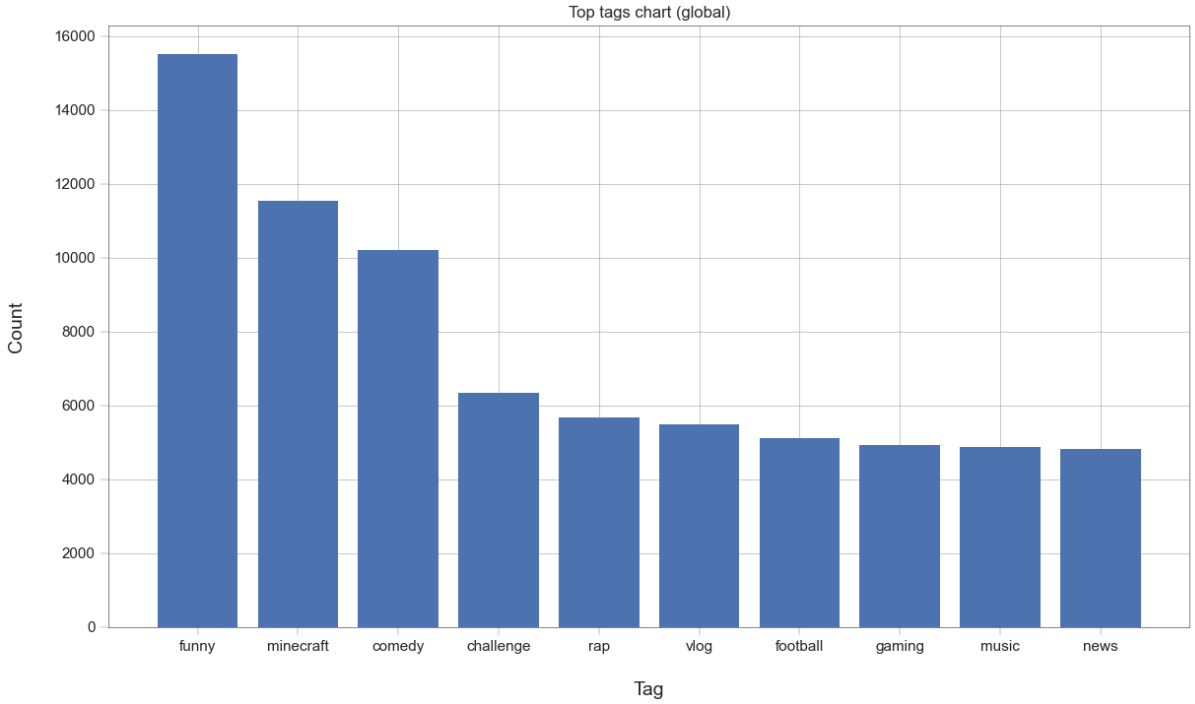


Figure 1 - Diagram of the distribution of tags in all categories

**Conclusion**. Overall, the developed application will provide users with a powerful tool that will simplify the video annotation process, improve search capabilities, and enhance the overall user experience when accessing and interacting with video content.

## References

1. Snoek C G, Worring M, Smeulders A W, Geusebroek J M. The challenge problem for automated detection of 101 semantic concepts in multimedia. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 421-430.
3. What is Random Forest? | IBM. [Online] URL: https://www.ibm.com/topics/random-forest.
4. Elith J, Leathwick J R, Hastie T. A working guide to boosted regression trees. Journal of Animal Ecology. 2008. 77(4):802-813.
5. Understanding Gradient Boosting | Medium. [Online] URL: https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764.