

References

- [1] Soloviev V.D., Chernov S.V. Artificial intelligence and intelligent systems. – M.: Hotline – Telecom, 2020. – 368 p.
- [2] Bogatyrev Yu.V., Matveev A.V. Economic security in the digital economy. – St. Petersburg: Piter, 2021. – 304 p.
- [3] Chen M., Tworek J., Jun H., et al. Evaluating large language models trained on code // arXiv preprint arXiv:2107.03374. – 2021.
- [4] Wang W., Lee R., Wang C., et al. Software vulnerability detection using deep learning: A survey // IEEE Access. – 2021. – Vol. 9. – P. 10486-10504.
- [5] Harman M., Mansouri S.A., Zhang Y. Search-based software engineering: Trends, techniques and applications// ACM Computing Surveys. – 2012. – Vol. 45(1).
- [6] Panichella S., Oliveto R., Di Penta M., et al. How developers use static analysis tools in their daily work // IEEE Int. Conf. on Program Comprehension, 2015. – P. 1–10.
- [7] Mittelstadt B., Russell C., Wachter S. Explaining explanations in AI// Communications of the ACM. – 2019. – Vol. 61(5). – P. 56-65.

ON THE QUESTION OF THE APPLICABILITY OF MODEL THEORY TO THE CHOICE OF THE BEST CLASS OF MACHINE LEARNING METHOD

Popova N.¹, Issayeva A.², Samoylova I.³, Pankin Y.⁴

^{1,2,3}Karaganda Buketov University, Karaganda, Kazakhstan

⁴Kazaktelekom, Karaganda, Kazakhstan

¹E-mail: popovanv092024@gmail.com

²E-mail: isaevaaiga@gmail.com

³E-mail: irinasam2005@mail.ru

The choice of a suitable machine learning method is traditionally based on empirical methods such as cross-validation, bootstrap, and heuristics. However, with the rapidly increasing complexity of models and the variety of tasks, there is a need for theoretically sound approaches. One of the potential directions is the use of concepts from the theory of models of first-order logic, which allows us to formalize the behavior of learning algorithms through the properties of the classes of functions described by them [1].

The concept of VC dimension (Vapnik–Chervonenkis dimension) characterizes the ability of a class of functions to implement all possible partitions of a finite set of points into positive and negative examples. A class of functions is considered learnable if its VC dimension is finite [2]. This means that there is a limit to its complexity, providing a controlled generalizing ability.

The VC-dimension is actively used in establishing upper bounds on generalization error within the framework of PAC learning. For example, linear classifiers in R^n have a VC – dimension of $n + 1$, which makes them learnable and predictable when the number of features is limited.

The Non-Independence Property (NIP) describes the behavior of first-order theories in which it is impossible to construct formulas capable of independently encoding infinite data sets. From a formal perspective, NIP excludes the possibility of shattering infinite sets and is associated with

the absence of formulas that can realize all possible membership patterns. This means that in a theory with NIP, any formula only generates "ordered" families of sets [3].

The connection between VC and NIP is established as follows: NIP classes form a subclass of classes with finite VC-dimension, which makes them particularly interesting from the standpoint of resistance to overfitting [4].

The use of VC-dimension and NIP enables classification of models according to their ability to generalize [4]:

Model Class	VC-Dimension	NIP	Comment
Linear classifiers	Low ($n+1$)	Yes	Suitable for small datasets
Decision trees (limited depth)	Medium	Yes	Controlled complexity
Neural networks	High	No (in general)	Require large datasets
SVM with non linear kernel	Medium / High	May lack NIP	Depends on the kernel used

Table 1: The use of VC-dimension and NIP

In tasks with limited data, preference is given to NIP models [3]. They demonstrate predictable behavior and offer the possibility of interpretation and verification – for example, in medicine, education, and law.

To illustrate the applicability of this approach, consider several scenarios where model selection can be justified through properties such as VC-dimension and NIP.

In text classification tasks, where data is often represented as sparse high-dimensional vectors but the training set size is limited, the best choice is typically linear models with regularization – such as logistic regression with L1 or L2 regularization, or a linear SVM. These models have finite VC-dimension, satisfy the NIP property, and offer good generalization while maintaining controlled complexity.

In computer vision tasks, where the data volume is large and the underlying dependencies are complex and high-dimensional, deep convolutional neural networks (CNNs) are applied. These architectures have high VC-dimension and generally do not satisfy NIP, but with large-scale datasets and the use of regularization techniques (such as Dropout or Batch Normalization), stable generalization and high accuracy can be achieved.

In research contexts – such as educational or sociological studies, where model interpretability is more important than absolute accuracy, preference is given to simple algorithms like shallow decision trees or naive Bayes classifiers. These models have finite VC-dimension, fall into the NIP class, and are easily interpretable, which makes them especially valuable in applications that demand explainability of the results.

Despite the strength of the theoretical foundations, the application of VC-dimension and NIP in practical AutoML remains limited:

- For many modern models (e.g., transformers), VC-dimension estimates are unknown;
- Determining NIP-related properties requires model-theoretic tools that are not available in standard machine learning frameworks;

- A high VC-dimension does not always lead to overfitting if the data has a specific structure (e.g., low effective dimensionality).

Model theory offers a powerful formal language for analyzing classes of machine learning models. VC-dimension and the NIP property serve as key indicators of a model's generalization ability and can be used for theoretically grounded algorithm selection based on data volume and interpretability requirements.

The further development of interdisciplinary approaches – at the intersection of logic, probability theory, and machine learning – may lead to the creation of new, explainable, and provably robust ML systems.

References

- [1] Kaplan, I., Ramsey, N. (2023). Model Theory and Machine Learning: A Survey. <https://arxiv.org/abs/2301.08731>
- [2] Vapnik, V. N., Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2), 264–280.
- [3] Kaplan, I., Shelah, S. (2020). On the structure of NIP theories. *Journal of Symbolic Logic*, 85(4), 1623–1650. <https://doi.org/10.1017/jsl.2020.61>
- [4] Livni, R., Simon, P. (2020). VC Theory in Machine Learning: A Model-Theoretic Perspective. arXiv:2005.04372. <https://arxiv.org/abs/2005.04372>
- [5] Guide to NIP Theories. Retrieved from <https://math.univ-lyon1.fr/~simon/guideNIP.pdf>

DEVELOPMENT OF A VIRTUAL ASSISTANT USING PYTHON

Samoilova Irina Alexeevna¹, Spirina Yelena Aleksandrovna², Smirnova Marina Aleksandrovna³, Popova Nadegda Viktorovna⁴

^{1,2,3,4}Karaganda Buketov University, Karagandy, Kazakhstan

¹E-mail: irinasam2005@mail.ru

²E-mail: sea_spirina@mail.ru

³E-mail: smirnova_marina_alex@mail.ru

⁴E-mail: dandn@mail.ru

This article aims to explore and develop technologies for creating a virtual assistant using the Python programming language, with the Discord online platform as a case study.

Python was chosen as the development tool because it is a powerful and versatile programming language widely used in various areas of software development, including the creation of bots (virtual assistants, conversational agents) for online platforms. The objective of this work was to develop an innovative solution that meets the needs of modern users and contributes to the development of online communities on the Discord platform.

Virtual assistants on Discord are programs designed to interact with users in an online environment. These bots possess various functions and capabilities that allow them to participate in discussions, provide information, entertain, and assist server members. Bots are software applications created to automate tasks and perform specific actions within chats. The role of a