

PHILOLOGY AND LINGUISTICS

Modern data science in philology

Krasniuk Svitlana¹

¹ Senior Lecturer;
Department of philology and translation,
Kyiv National University of Technologies and Design; Ukraine

Abstract. Modern Data Science in philology is an important and relevant field of research that uses computational methods to analyze linguistic data. It combines statistics, computer science and linguistics, which allows you to gain new knowledge about language, text and their use in different contexts. Here are some key aspects that highlight the relevance and importance of modern data science in philology. The modern science of data in philology is extremely important and relevant. It not only provides new tools for the analysis of language data, but also promotes the development of interdisciplinary research, adaptation to new technologies and forms of communication. In light of rapid changes in society and technology, data science in philology is becoming key to understanding language as an important social phenomenon.

Keywords: *Data Science, philology, Data Mining, machine linguistics.*

Introduction.

Data Science in education plays an important role in the transformation of educational processes, increasing the efficiency of educational systems and providing a personalized approach to learning. The use of big data, machine learning, analysis of educational achievements and individual characteristics of students makes it possible to create new learning models, optimize the teaching and management processes of educational institutions [1, 2].

It is worth noting that Data Science is an interdisciplinary field that combines knowledge of programming, mathematics, statistics and industry knowledge to analyze and interpret data, including big data. Data science focuses on the process of processing Big Data, from its collection to obtaining analytical conclusions and building forecasts [3, 4, 5, 6]. In this process, it actively uses machine learning, artificial intelligence, forecasting models, data visualization and other tools [7].

Machine Learning is a key tool in Data Science, as it allows you to automatically detect patterns in data and make predictions based on these patterns [8, 9]. Thanks to machine

PHILOLOGY AND LINGUISTICS

learning, analysts can work with large volumes of data and obtain important information from them for decision-making [10, 11].

Deep Learning is an important part of Data Science and is actively used to solve complex tasks that require processing large volumes of data and identifying complex patterns [12, 13]. Due to its ability to work with large data (Big Data) and learn complex patterns automatically, deep learning has become a key tool in many areas of data science.

The author emphasizes that Data Mining and Data Science are related, but distinct concepts that interact in the process of analyzing large volumes of data (in particular, large volumes of philological data).

Considering the above, the author makes a preliminary conclusion that Data Science in philology opens up new opportunities for analysis and interpretation of linguistic data. This approach integrates traditional methods of linguistics with modern technologies [14, 15], such as machine learning, natural language processing, and statistical models.

The Main Part.

Data Science in philology is an important field of research that combines traditional methods of linguistic data analysis with modern technologies such as machine learning, natural language processing (NLP), statistical models, and big data analysis. This approach allows you to gain new knowledge about language, text and their use in society.

The main directions of data science in philology:

1. Natural Language Processing (NLP)

Background and Purpose: NLP is a field that combines computer science and linguistics to analyze and process human language. The use of NLP in philology allows you to automate the analysis of texts, identify patterns, translate and interact with users.

NLP tasks:

Tokenization: Breaking the text into individual words or phrases.

Lemmatization and stemming: Reducing words to their basic forms to simplify analysis.

Syntactic and semantic analysis: Studying the structure of sentences and the meaning of words in context.

2. Analysis of texts

Lexical analysis: Study of the frequency of use of words, idioms and phraseological units in texts. This can be useful

PHILOLOGY AND LINGUISTICS

for understanding language changes, stylistic features, and general trends in speech.

Stylistic analysis: Applying statistical methods to study the style of authors, including the use of metaphors, allegories, and other rhetorical figures. For example, analysis can reveal how writing style changes in different contexts or historical periods.

3. Corpus linguistics

Creation of corpora: Development of large sets of texts (corpora) for the analysis of linguistic phenomena. Corpus can contain literary works, newspaper articles, scientific articles, etc.

Corpus Analysis: Using software to automatically analyze corpora, which allows researchers to identify patterns in language use. For example, corpus analysis can be used to study how certain words or phrases are used in different genres of texts.

4. Analysis of social media and online data

Sociolinguistic analysis: Studying language on social media platforms to identify language changes, new words and phrases that appear in users' speech. For example, you can analyze how the language of young people differs from the language of older generations.

Discourse research: Using data from social media to analyze how different social groups communicate and what topics they are interested in. It can help to study public opinion on various topics.

5. Analysis and generation of language data

Automatic text generation: Using machine learning models to generate text based on input data. This may include the generation of news, literary works or even poetry.

Translation models: Using neural networks to automatically translate texts from one language to another, which can improve access to information and aid in language learning.

6. Analysis of historical texts

Digital Humanities: Using data science technologies to analyze historical texts to study language evolution, cultural change, and social structures. For example, researchers can analyze changes in vocabulary and grammatical structures in works from different eras.

Visualization techniques: Using visualizations to present analysis results that help understand complex language phenomena.

PHILOLOGY AND LINGUISTICS

7. Training and assessment.

Automated grading: Development of systems to automatically grade student writing, analyze grammatical errors, and provide feedback. This can increase the effectiveness of teaching and reduce the burden on teachers.

Adaptive learning: Using data analytics to create personalized learning programs that take into account the knowledge level and interests of students.

Application of data science in philology:

- **Academic research:** Using data science methods to study literary works, analyze language trends, create lexicographic resources, etc.

- **Translation and localization:** Development of tools for automated translation and adaptation of content for different language cultures.

- **Social Studies:** Analysis of language data to study social phenomena such as identity, culture and social interaction.

- **Education:** Using data analytics to improve the language learning process, including the creation of interactive language learning platforms.

Conclusions.

Data science is becoming an important component of modern education, contributing to improving the quality of the educational process, individualizing the approach to students, and making informed decisions in the management of educational institutions. Thanks to these technologies, education becomes more flexible, efficient and accessible.

Big data and data science work together to solve complex problems and unlock new opportunities in various fields. Thanks to powerful tools and technologies, big data analysis helps extract useful knowledge, increase the efficiency of organizations and create innovative solutions in various industries.

Machine learning is an integral part of Data Science, helping to automate the process of analyzing large volumes of data, making predictions, classifying information and discovering hidden patterns. This allows businesses and scientists to make informed decisions quickly and efficiently.

Deep machine learning is a powerful tool in data science that helps solve complex tasks of processing large volumes of data. Its ability to learn complex patterns and automatically find hidden patterns makes it indispensable for image

PHILOLOGY AND LINGUISTICS

processing, natural language, recommender systems, medical diagnostics, and many other fields. Deep learning also plays an important role in big data analysis, making it possible to make more accurate predictions and understand data that was previously difficult to process with traditional methods.

The author reiterates the thesis that Data Mining is an important part of data science, which focuses on finding hidden patterns in data. Data Science, in turn, includes not only the extraction of such regularities, but also data processing, building models and decision-making based on analysis [18, 19, 20]. This makes Data Science a broader and more comprehensive discipline that is used to solve complex problems in various fields such as business, medicine, education and others.

Data science in philology offers new approaches to the study of language, text, and their impact on society. The integration of modern technologies and methods of analysis allows researchers to gain new knowledge about language phenomena, as well as to improve learning and communication processes. This interdisciplinary approach opens up new horizons for research in philology, particularly in literary studies, sociolinguistic analysis, translation, and language teaching.

Discussion and prospects for further research.

The author puts forward the thesis that Genetic Algorithms in combination with deep machine learning constitute a powerful tool for solving various problems in linguistics [21]. Genetic algorithms, which are part of evolutionary computing, use the mechanisms of natural selection and genetics to optimize decisions [22, 23], and deep learning provides opportunities for working with large volumes of data and discovering complex patterns. Genetic algorithms in combination with deep machine learning represent a promising direction in linguistics, which allows the development of new methods of analysis and generation of language data. The use of these technologies opens up new opportunities for the study of language, its structure and functioning in modern society.

References:

- [1] Tetiana Tsalko, Svitlana Nevmerzhytska, Svitlana Krasniuk, Svitlana Goncharenko, Liubymova Natalia (2024). Features, problems and prospects of data mining and data science application in educational management. *Bulletin of Science and Education*, № 5(23), 2024. pp.637-657

PHILOLOGY AND LINGUISTICS

- [2] Краснюк Світлана. (2024) Data Science у освітньому менеджменті // Діалог культур у Європейському освітньому просторі [Електронний ресурс] : Матеріали IV Міжнародної конференції, м. Київ, 10 травня 2024р. Київський національний університет технологій та дизайну / упор. С. Є. Дворянчикова. – К. : КНУТД, 2024. – С. 119-124.
- [3] Науменко, М. (2024). Аналіз та аналітика великих даних в маркетингу та торгівлі конкурентного підприємства. *Grail of Science*, (40), 117-128. <https://doi.org/10.36074/grail-of-science.07.06.2024.013>
- [4] Maxim Krasnyuk, Svitlana Nevmerzhytska, Tetiana Tsalko. (2024). Processing, analysis & analytics of big data for the innovative management. *Grail of Science*, #38, April 2024. pp. 75-83. <https://www.journal-grail.science/issue38.pdf>
- [5] Krasnyuk M., Krasniuk I. Big data analysis and analytics for marketing and retail. *Штучний інтелект у науці та освіті: збірник тез Міжнародної наукової конференції (AISE) (1-2.03.2024 р.)*, Київ, 2024.
- [6] Maxim Krasnyuk, Dmytro Elishys (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. *Science and technology today*, #4 (32) 2024. pp. 833-857
- [7] Krasnyuk, M., & Krasniuk, S. (2020). Application of artificial neural networks for reducing dimensions of geological-geophysical data set's for the identification of perspective oil and gas deposits. *Збірник наукових праць ЛОГОС*, 18-19. <https://doi.org/10.36074/24.04.2020.v2.05> DOI: <https://doi.org/10.36074/24.04.2020.v2.05>
- [8] Науменко, М. (2024). Ефективне застосування класичних алгоритмів машинного навчання при прийнятті адаптивних управлінських рішень. *Наукові перспективи*, 2024, #5 (47). [https://doi.org/10.52058/2708-7530-2024-5\(47\)-855-875](https://doi.org/10.52058/2708-7530-2024-5(47)-855-875)
- [9] Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. *ЛОГОС*. 2020. P. 55-57.
- [10] Krasnyuk M., Tkalenko A., Krasniuk S. Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. *ЛОГОС*. 2021.
- [11] Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. *ЛОГОС*. 2021.
- [12] Maxim Krasnyuk, Svitlana Krasniuk, Svitlana Goncharenko, Liudmyla Roienko, Vitalina Denysenko, Liubymova Natalia. Features, problems and prospects of the application of deep machine learning in linguistics // *Bulletin of Science and Education*, № 11(17), 2023. pp.19-34. <http://perspectives.pp.ua/index.php/vno/article/view/7746/7791>
- [13] Науменко, М. (2024). Оптимальне використання алгоритмів глибокого машинного навчання в ефективному управлінні підприємством. *Успіхи і досягнення у науці*, 2024, #4 (4). [https://doi.org/10.52058/3041-1254-2024-4\(4\)-776-794](https://doi.org/10.52058/3041-1254-2024-4(4)-776-794)
- [14] Krasniuk, S., & Goncharenko, S. (2024). MODERN MATHEMATICAL LINGUISTICS. *Collection of Scientific Papers «SCIENTIA»*, (September 13, 2024; Pisa, Italia), 115-119. Retrieved from <https://previous.scientia.report/index.php/archive/article/view/2060>
- [15] Krasniuk, S. (2024). ADVANCED TEXT MINING IN PHILOLOGY. *Collection of Scientific Papers «SCIENTIA»*, (September 27, 2024; Stockholm, Sweden), 68-72. Retrieved from

PHILOLOGY AND LINGUISTICS

- <https://previous.scientia.report/index.php/archive/issue/view/27.09.2024>
- [16] Goncharenko, S., & Krasniuk, S. (2024). Innovative architecture of large language models. In *Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування*. Національний авіаційний університет.
- [17] Krasniuk, S., & Goncharenko, S. (2024). Ethics of using large language models in machine linguistics. *Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування*. Національний авіаційний університет.
- [18] Науменко, М. (2024). Інтелектуальний аналіз бізнесових даних як фактор посилення конкурентної позиції підприємства. *Успіхи і досягнення у науці*, 2024, 5 (5). [https://doi.org/10.52058/3041-1254-2024-5\(5\)-746-762](https://doi.org/10.52058/3041-1254-2024-5(5)-746-762)
- [19] Krasnyuk M.T., Hrashchenko I.S., Kustarovskiy O.D., Krasniuk S.O. (2018) Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. *Economies' Horizons*. 2018. No. 3(6). pp. 121-136
- [20] Kulynych Y., Krasnyuk M., Krasniuk S. Knowledge discovery and data mining of structured and unstructured business data: problems and prospects of implementation and adaptation in crisis conditions. *Grail of Science*. 2022. (12-13). pp. 63-70.
- [21] Krasnyuk, M., & Krasniuk, S. (2024). Evolutionary technologies and genetic algorithms in machine translation. *European Science*, 3(sge30-03), 91-98. <https://doi.org/10.30890/2709-2313.2024-30-00-025>
- [22] Науменко, М., & Краснюк, М. (2024). Ефективне застосування генетичних алгоритмів у вирішенні багатоекстремумних оптимізаційних задач в менеджменті конкурентного підприємства. *Grail of Science*, (41), 65-73. <https://doi.org/10.36074/grail-of-science.05.07.2024.008>
- [23] Kulynych, Y., Krasnyuk, M., & Krasniuk, S. (2022). Efficiency of evolutionary algorithms in solving optimization problems on the example of the fintech industry. *Grail of Science*, (14-15), 77-84. <https://doi.org/10.36074/grail-of-science.27.05.2022.010>