метричні характеристики реалізованого коду. Це моделювання сприяло модифікації алгоритмів та підтвердженню гіпотези про граф, у якого значення об'єму першої компоненти канонічної визначальної пари є максимальним.

ПЕРЕЛІК ПОСИЛАНЬ

1. I.S. Grunskii, I.A. Mikhaylova, S.V. Sapunov (2012). Domination on the vertices of labeled graphs. *Algebra and Discrete Mathematics,* Vol. 14, Is. 2, p. 174-184.
2. О.С. Сенченко, М.І. Притула, О.А. Середа (2022). Представлення детермінованих графів визначальною парою слів. *Праці ІПММ НАН України*, том 36, №2, с. 104-116. doi: 10.32782/1683-4720-2022-36-09..
3. This repository presents the implementation of algorithms for the linguistic representation of deterministic graphs, 2023. URL: https://github.com/NickIT87/IAMM_proceedings.
4. NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks, 2023. URL: https://networkx.org/documentation/.
5. Introduction to Code Metrics. URL: https://radon.readthedocs.io/en/latest/intro.html.

UDC 004.42 (8)

Artem Vizniuk[1], Ivan Laktionov[1], Grygorii Diachenko[1]
[1]Dnipro University of Technology, Dnipro, Ukraine

STUDY OF THE EFFECTIVENESS OF ML ALGORITHMS IN PREDICTING THE PROBABILITY OF CROP DISEASE OCCURRENCE

Abstract. The aim of this research is to study the effectiveness of ML algorithms in predicting the probability of the corn disease «Fusarium Head Blight» in Dnipro region of Ukraine. Linear regression, feedforward neural networks and random forest models were considered for prediction. The random forest model obtained the best metric score on the testing set: R^2=0.965, RMSE=3.44. Directions for further research were substantiated in the given subject area.

***Keywords**: ML, regression, time series, disease prediction, crops, agriculture.*

Introduction. According to the agricultural production statistics, provided by the globally recognized FAO [1], cereal crops, which have tripled in production over the past 20 years, are the most widely cultivated agricultural crops in open field conditions. Based on the statistical analysis of the data, accumulated by the FAO [2], it was established that the most cultivated cereal crops (according to harvested areas) in Eastern and Southern Europe, and Ukraine, in particular, are wheat, corn and barley. Trend dynamic of harvested areas over a period from 2012 to 2021 for wheat, corn and barley in Ukraine is shown in Fig. 1.

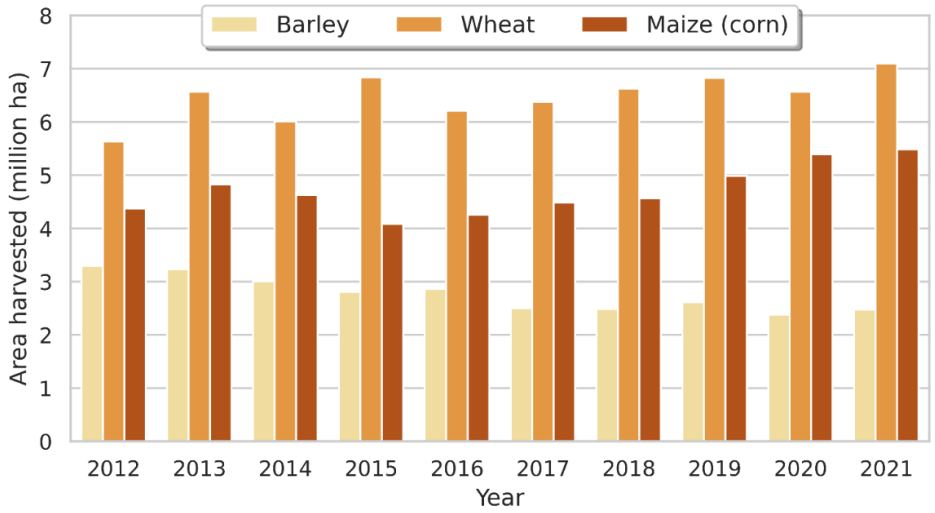

Figure 1. Area harvested for barley, wheat and corn crops in Ukraine over a period from 2012 to 2021

As can be seen in Fig. 1, wheat is the most commonly cultivated cereal crop in Ukraine, followed by corn and barley.

By analysing the last articles and research papers on the topic of crop disease prediction it was established, that fluctuations in climatic parameters, such as air temperature, relative humidity, precipitation, wind speed etc., have the biggest impact on crop disease occurrence [3–5]. Among the reviewed approaches to crop disease prediction based on climatic data ML algorithms demonstrated significant effectiveness in processing climatic data for predicting the probability of crop disease occurrence [3–5].

Problem statement. The aim of this research is to evaluate the effectiveness of ML algorithms in predicting the probability of the «Fusarium Head Blight» disease occurrence in corn in Dnipro region of Ukraine and to determine the most effective approach (according to R^2 and RMSE metrics) that will be integrated into a software application to provide real-time predictions of the disease probabilities of the observed crops. To achieve this aim, the following objectives were formulated and solved:

1. To substantiate the methods and means of research to evaluate the effectiveness of ML algorithms in predicting the probability of the «Fusarium Head Blight» disease in corn.
2. Determine the most effective regression model according to the R^2 and RMSE metric score on the testing set and the respective hyperparameters of this model.
3. Programmatically implement an algorithm for predicting the probability of corn disease using the most performant regression model.

Methodology. Climatic data with the «Fusarium Head Blight» disease probabilities for Dnipro region were downloaded from the METOS by Pessl Instruments weather station using the FieldClimate platform, which was provided by Metos Ukraine LLC. The values of air temperature, precipitation, air humidity, leaf wetness time, and disease probability were measured hourly from September 2022 to September 2023. Descriptive statistics of the collected data are shown in Table 1.

Table 1

Descriptive statistics of the collected data

	Air temperature (°C)	Precipitation (mm)	Humidity (%)	Leaf wetness duration (min)	Infection probability (%)
Count	8658				
Mean	10.62	0.07	72.8	8.7	4.84
Standard deviation	9.74	0.34	17.32	21.13	17.61
Median	9.65	0	76	0	0
Min	–10.9	0	19	0	0
Max	37	14	100	60	100

To account for the history of changes in climate parameters and disease probabilities, an additional hyperparameter timestamps is introduced, which determines how many previous values of climate parameters and disease probabilities are taken into account. For a timestamps value of 1, current climate parameters along with the climate parameters and disease probability an hour ago are used as input. For a timestamps value of 2, the climatic parameters and disease probability 2 hours ago are also provided as inputs.

Certain ML models, such as neural networks, are trained using a gradient descent algorithm, which converges faster when input values are on a similar scale. Therefore, the input values were transformed as follows:

– leaf wetness duration (min) was divided by the maximum value of 60;

– humidity (%) was divided by the maximum value of 100;

– precipitation (mm) and air temperature were standardized by the following equation: $x' = (x - \bar{x})\sigma^{-1}$, where x' – is the standardized value of a numerical feature x, $\bar{x}$ – is the mean value of x on the training set, and σ – is the standard deviation of x on the training set;

– disease probability (%) was divided by the maximum value of 100.

The dataset was partitioned into training, validation and testing sets with a distribution of 70:15:15, corresponding to 6058 records for training, 1298 records for validation and 1299 records for testing purposes.

To predict the probability of the «Fusarium Head Blight» disease the following regression models were considered in this paper: linear regression, neural network and random forest. The research was conducted in the Google Colab environment using the programming language Python 3.10.12. Regression ML models hyperparameters which resulted in the best metric score on the testing set are presented in Table 2.

Table 2

Results of the considered ML models

Model	**Hyperparameters**	**Results on the testing set**
Linear regression	timestamps=3	R^2=0.96, RMSE=3.61
Random forest	timestamps=3, max_depth=5, n_estimators=10	R^2=0.965, RMSE=3.44
Feedforward neural network	timestamps=3, optimizer=AdamW(learning_rate=0.001, weight_decay=0.003,beta_1=0.9,beta_2=0.999), layers=[Dense(10, activation=ReLU), Dense(5, activation=ReLU), Dense(3, activation=ReLU), Dense(2, activation=ReLU), Dense(1, activation=Linear)], batch_size=256, epochs=500	R^2=0.962, RMSE=3.59

In Table 2 it is shown that all considered ML models performed the best for the hyperparameter value timestamps=3. The random forest model demonstrated the best metric score on the testing set: R^2=0.965, RMSE=3.44.

The **scientific novelty** of the results obtained lies in the establishment of a set of input climatic data and the evaluation of the effectiveness of machine learning algorithms in predicting the probability of the «Fusarium Head Blight» disease in corn, which makes it possible to optimize the choice of the approach when designing specific types of systems, taking into account the R^2 and RMSE metrics.

Conclusions. In the course of the research on predicting the probability of the «Fusarium Head Blight» disease in corn, it has been established that the climatic parameters and disease probabilities in the last 3 hours along with the current climatic parameters should be taken into account for the most accurate predictions. The random forest model with hyperparameters max_depth=5 and n_estimators=10 has achieved the best metric score on the testing set. Future studies on this topic would consider recurrent neural network (RNN) architectures to address the problem of crop disease prediction based on climatic data.

REFERENCES

1. Agricultural production statistics 2000 – 2021. URL: https://www.fao.org/3/cc3751en/cc3751en.pdf (Accessed on 30.11.2023).
2. FAOSTAT: Crop and livestock products. URL: https://www.fao.org/faostat/en/#data/QCL (Accessed on 30.11.2023).

3. Fenu G., Malloci F. M. An application of machine learning technique in forecasting crop disease. ACM International Conference Proceeding Series. 2019. No. June P. 76-82. DOI: 10.1145/3372454.3372474.
4. Xiao, Q., Li, W., Kai, Y., et al. Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. BMC Bioinformatics. 2019. Vol. 20, No. Suppl 25. P. 1–15. DOI: 10.1186/s12859-019-3262-y.
5. Patil, R. R., Kumar, S., Rani, R. Comparison of Artificial Intelligence Algorithms in Plant Disease Prediction. Revue d'Intelligence Artificielle. 2022. Vol. 36, No. 2. P. 185–193. DOI: 10.18280/ria.360202.

УДК 519.8

С.А.Ус[1], О.Ю.Хархула[1]
[1]Національний технічний університет «Дніпровська політехніка», Дніпро, Україна

ДОСЛІДЖЕННЯ ТА РАНЖУВАННЯ АЛЬТЕРНАТИВ ПЛАСТИКОВОЇ УПАКОВКИ

Анотація. Проведено дослідження наявних альтернатив пластиковому пакуванню. Розв'язано задачу багатокритеріального вибору.

***Ключові слова**: пластикове забруднення, пакування, вторинне використання, задача вибору, метод аналізу ієрархій*

Вступ. Проблема пластикового забруднення сьогодні є дуже актуальною. Об'єми пластику, що використовуються у промисловості і споживанні людьми, дедалі зростають і немає жодних передумов вважати, що ця тенденція буде змінюватися. При цьому наслідки пластикового забруднення є досить значущими і завдають великої шкоди навколишньому як водному так і наземному середовищу. Тому для відповідального виробництва і споживання пошук альтернативних варіантів пакування і їх впровадження є актуальною задачею.

Метою даної роботи є аналіз переваг та недоліків альтернативних варіантів пакування і вибір раціонального варіанта для впровадження.

Основний зміст роботи. У доповіді «Подолання пластикової хвилі: комплексна оцінка шляхів припинення забруднення океану пластиком» [1] проведено оцінювання існуючої ситуації у сфері пластикових відходів й розглянуто кілька сценаріїв її розвитку.

Головним завданням визначено скорочування споживання пластику, інакше через 20 років щорічні обсяги його відходів з 11 млн. метричних тон зростуть до 29 млн. метричних тонн. Тобто на кожному метрі берегової лінії на земній кулі лежатиме 50 кг пластикового сміття, а в океані плаватиме приблизно 600 млн. тон таких відходів.

Для виходу із ситуації запропоновано чотири варіанта. Найоптимістичніший із них – «Зміна системи» (SCS) – скорочення виробництва