ISSN 2708-2032
e-ISSN 2708-2040

INTERNATIONAL
UNIVERSITY

INTERNATIONAL JOURNAL OF INFORMATION & COMMUNICATION TECHNOLOGIES

Volume 2, Issue 2
June, 2021

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ IT МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ
INTERNATIONAL UNIVERSITY

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Том 2, Выпуск 2
Июнь, 2021

Издание зарегистрировано Министерством информации и общественного развития Республики Казахстан. Свидетельство о постановке на учет № KZ82VPY00020475 от 20.02.2020 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Выходит 4 раза в год.

УЧРЕДИТЕЛЬ:
АО «Международный университет информационных технологий»

ISSN 2708-2032 (print)
ISSN 2708-2040 (online)

СОДЕРЖАНИЕ

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

CONTENTS

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

INFORMATION AND COMMUNICATION NETWORKS AND CYBERSECURITY

SMART SYSTEMS

DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

МАЗМҰНЫ

БАҒДАРЛАМАЛЫҚ ҚАМТАМАНЫ ӘЗІРЛЕУ ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУІПСІЗДІК

ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР

ЭКОНОМИКА ЖӘНЕ БАСҚАРУДАҒЫ САНДЫҚ ТЕХНОЛОГИЯЛАР

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 2.Is. 2. Number 06 (2021). 112–119
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2021.06.2.015

УДК 004.62

Alimkhan A.M.

International Information Technology University, Almaty, Kazakhstan

PREDICTING BASKETBALL RESULTS USING DEEP LEARNING ALGORITHMS

Abstract. With the development of information technology and an ever-expanding statistical base, the possibilities for forecasting are expanding, and the dependencies of the calculated indicators on the result are considered. In this article we compare 3 most widely spread game result prediction models, and namely, Support Vector Regression, K-Nearest Neighbor model and the Linear Regression model in terms of their prediction accuracy and experimentally demonstrate the advantages of the latter.

Keywords: deep learning, machine learning, prediction models, regression technique, data analysis

Introduction

Today, there are many new basketball metrics used by fans and analysts around the world to compare and measure players. For example, there is a Player Performance Rating (PER) that considers achievements such as field goals, free throws, 3 pointers, assists, rebounds, blocks and interceptions, and negative results such as missed shots, turns and personal fouls. Some of these advanced basketball metrics, such as PER, are great for measuring a player's performance in a basketball game. Yet, does this mean that the player will bring success to his team? May be. But the only way to see if a player succeeds is his victory. Many sports betting fans are interested in predicting the results. It is also the subject of sports analytics. Since for the most reliable forecast it is necessary to consider and correctly analyze a large complex of sports characteristics, the use of various AI methods is quite logical.

When developing your own software tool, where you can consider all the nuances of the problem like the prediction result of the game being solved, you should adjust it for yourself. However, it is obvious that this path requires much more effort and time. But it is also obvious that the result obtained will be more accurate. Naturally, to go down this path, you need not only to have sufficient qualifications in the field of artificial intelligence, but also to program well yourself. Also, it is necessary to know about deep learning, machine learning, the prediction models like Linear Regression, Support Vector Machine, and k-Nearest Neighbors Regression Model etc., to use them for predicting the result of a basketball game.

Data analysis and exploration

Analysis of the source data. Data Preparation is a very laborious iterative process that takes up to 80% of all resource and time costs in the Data Mining life cycle and includes the following tasks for processing the initial ("raw") data [12]:

- Data examining - choice of highlights (or indicators) and objects, considering their significance for the purposes of Information Mining, quality, and specialized imperatives (volume and sort).
- Data cleansing - expelling typos, erroneous values (for illustration, a number in a string parameter, etc.), lost values (Lost Values or NA), barring copies and diverse depictions of the same protest, reestablishing uniqueness, judgment, and coherent associations.
- Feature Generation - creation of the determined highlights and their change into vectors for the Machine Learning show, as well as introduction of changes to enhance the precision of machine learning calculations.

- Integration - consolidating information from different sources (data frameworks, tables, conventions, etc.), counting their conglomeration, when unused values are calculated by summing data from a set of existing records.
- Formatting - syntactic changes that do not alter the meaning of the information, but are required for modeling devices, for sorting cases in a particular arrangement or expelling superfluous accentuation in content areas, trimming "long" words, adjusting genuine numbers to integrability, etc.

Model Selection and Testing

To improve the accuracy of predictive models, it is necessary to apply appropriate machine learning methods. Depending on the nature of the business problems under consideration, different approaches are used, considering the types and volumes of data. This section introduces the categories of machine learning. There are such algorithm models used in building a prediction program as:

- Linear Regression
- Support Vector Regression
- K-Nearest Neighbour

Linear Regression. Linear Regression is meant to predict continuous numeric variables. In expansion, the adjusted adaptation of straight relapse actualized within the Deductor analytical platform also allows of solving the classification problem.

Linear Regression is maybe one of the foremost well-known and popular calculations and insights in machine learning.

Predictive modeling is essentially concerned with decreasing model errors or, in other words, making predictions as accurate as possible. We will borrow calculations from different areas, counting measurements, and utilize them for this reason.

Benefits of Linear Regression:

- The quickness and ease of obtaining the model.
- Interpretability of the show. The straight show is straightforward and justifiable for the examiner. The gotten relapse coefficients can be utilized to judge how a specific calculation influences the result, and to draw extra valuable conclusions on this basis.
- Wide appropriateness. Numerous genuine forms in financial matters and trade can be depicted by direct models with adequate exactness.

Knowledge of this approach. For Linear Regression, typical problems (for example, multicollinearity) and their solutions are known, tests for assessing the static significance of the resulting models are developed and implemented.

SVM. SVM is a supervised learning algorithm used to solve classification problems. Here the most thought-of the strategy is to exchange the beginning vectors to a space of higher measurement and hunt for an isolating hyperplane with the biggest hole in this space. Two parallel hyperplanes are developed on both sides of the hyperplane isolating the classes. The isolating hyperplane will be the hyperplane that makes the most prominent remove to two parallel hyperplanes. The calculation accepts that the more noteworthy the contrast or remove between these parallel hyperplanes, the smaller the normal blunder of the classifier will be.

Advantages of SVM:

- Fast classification method.
- The method is reduced to solving a quadratic programming problem in a convex domain, which usually has a unique solution.
- The method allows for more confident classification than other linear methods.

Disadvantages of SVM:

- They have a long learning curve, so in practice they are not suitable for large datasets. Another disadvantage is that SVM classifiers do not work well with overlapping classes.

The important concepts in SVM are as follows:

- Support vectors - The data points that are closest to the hyperplane are called support vectors. The dividing line is defined using these data points.
- Hyperplane - It is a decision plane or space that is divided between a set of objects that have different classes.
- Margin - This can be defined as a gap between two lines at the data points of a cabinet of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margins are considered good margins, and small margins are considered bad ones.

The main purpose of SVM is to subdivide datasets into classes to find the maximum limit hyperplane (MMH), and this can be done in the following two steps:

- First, the SVM iteratively generates hyperplanes that best separate classes.
- It will then select the hyperplane that separates the classes correctly.

K-Nearest Neighbour. kNN stands for k Nearest Neighbor or k Nearest Neighbors - it is one of the simplest classification algorithms, sometimes used in regression problems. Due to its simplicity, it is a good example from which to start exploring the field of Machine Learning. This article describes an example of writing the code of such a classifier in Python, as well as visualizing the results obtained.

The allocation issue in machine learning is the issue of allotting a question to one of the predefined classes based on its formalized highlights. Each of the objects in this issue is referred to as a vector in an N-dimensional space, each measurement in which may be a depiction of one of the object's qualities. Let us say we need to classify monitors: the measurements in our parameter space will be the diagonal in inches, the aspect ratio, the maximum resolution, the presence of an HDMI interface, cost, etc. The case of text classification is somewhat more complicated, for this purpose the term-document matrix is usually applied.

Literature Review

The articles discuss the most interesting trends in machine learning and artificial intelligence formed at the beginning of 2018 outside of specific mathematical methods for optimization, processing, and analysis of data. There is an increasing attention of researchers to the question of methodologies, or metamodels (from the English metamodel): the principles of using, combining, and choosing specific models and methods of machine learning. A long-term progress in the development of machine learning methods has spawned not only a variety of mathematical, software, and even hardware solutions designed for predictive and generative data analysis tasks in various fields, but also encountered many difficulties and obstacles along the way.

The main difficulty that a person faces in the process of getting to know the field of machine learning is a huge number of disparate methods, each of which has its own features, area of use and benefits. However, this diversity sometimes baffles sophisticated researchers. With the development of mathematical and algorithmic methods it becomes more and more difficult to navigate well in all the nuances of the applied algorithms. Unfortunately, the methodological base lags far behind the rapid process of developing new learning algorithms, and the process of choosing a trainable model sometimes comes down to a simple search.

Despite the rapid development of machine learning in the last decade, artificial intelligence remains a very vague concept. It includes many subject areas, from time series prediction to generating plausible images on a specific topic. Machine learning methods, which form the computational basis of artificial intelligence technologies, remain highly specialized for each specific task.

Support vector machines - the concept of the algorithm is, as in the case of logistic regression, in the search for a dividing plane (or several planes), however, the way to search for this plane in this case is different - a plane is searched so that the distance from it to the nearest points - representatives of both classes – is as long as possible, for which methods of quadratic optimization are usually applied.

Lazy classifiers (Lazy learners) - a special kind of classification algorithms, which, instead of pre-building a model and then, based on it, making decisions about assigning an object to a particular class, hinge on the idea that similar objects most often have one and the same class. When such an algorithm receives an object for classification as an input, it searches among previously viewed similar objects and, using information about their classes, forms its prediction regarding the class of the target object.

In their article "Machine learning approaches to predicting basketball game outcomes" Sushma Jain and Harman deep Kaur discuss the difference between the execution of the NBAME model with conventional machine learning classifiers and show that the NBAME demonstrates the next expectation exactness rate (74.4%). To illuminate the inadequacies of SVMs, Pai et al. and Kaur et al. utilized SVMs combining separate choice rules and fluffy rules, to create unused prescient coordinate result models. The result illustrated that the models accomplished higher precision than did the routine SVMs.

In article of Chenjie Cao entitled "Sports Data Mining Technology Used in Basketball Outcome Prediction" the author focused on two machine learning models, SVM and hybrid fuzzy-SVM (HFSVM) to predict NBA games. Their dataset is restricted to the regular season 2015-2016, which is split into training and test sets. The hybrid Fuzzy-SVM model combines the advantages of the fuzzy model and the SVM for a basketball match result prediction. They claim to have been able to get better game results in terms of their accuracy, it is observed that this study, with its 79.2% accuracy rate achieved by using the suggested CNFS model, has the highest accuracy rate among all the studies.

You can see that classification algorithms can have a variety of ideas in their basis and, of course, for different types of problems they show different efficiency. So, for problems with a small number of input features, rule-based systems can be useful, if it is possible to calculate some similarity metric for the input objects quickly and conveniently using lazy classifiers. As for the problems with a very large number of parameters, which are difficult to identify or interpret, such as image or speech recognition, neural networks are becoming the most appropriate classification method.

Experimental part

There is a process of creating prediction models below with math statistics like the mean squared error, the mean absolute error and the variance score. So according to the metrics of each model, we will have a chance to select the best prediction model.

Figure 1 illustrates the process of creating the Linear Regression model:

```
# Create the Linear Regression model

linReg = linear_model.LinearRegression()
linReg.fit(x_train, y_train)

linReg.predict(x_test)

y_lin_pred = linReg.predict(x_test)

print('Score: %.3f' % linReg.score(x_train, y_train))
print('Mean squared error: %.3f' % mean_squared_error(y_test, y_lin_pred))
print('Mean Absolute error: %.3f' % mean_absolute_error(y_test, y_lin_pred))
print('Variance score: %.3f' % r2_score(y_test, y_lin_pred))
```

```
Score: 0.910
Mean squared error: 1.007
Mean Absolute error: 0.761
Variance score: 0.880
```

Figure 1 - Linear Regression Model

Figure 2 shows the creation of the Support Vector Regression model with math points

```
# Create the Support Vector Regression model

svr = SVR(kernel='rbf', gamma=1e-3, C=150, epsilon=0.3)
svr.fit(x_train, y_train.values.ravel())

y_svr_pred = svr.predict(x_test)

print('Score: %.3f' % svr.score(x_train, y_train))
print("Mean squared error: %.3f" % mean_squared_error(y_test, y_svr_pred))
print('Mean Absolute error: %.3f' % mean_absolute_error(y_test, y_svr_pred))
print('Variance score: %.3f' % r2_score(y_test, y_svr_pred))
```

```
Score: 0.936
Mean squared error: 1.562
Mean Absolute error: 1.038
Variance score: 0.814
```

Figure 2 - Support Vector Regression Model

Figure 3 demonstrates the creation of the k-Nearest Neighbors Regression Model model with math points.

```
# Create the k-Nearest Neighbors Regression Model

knn = neighbors.KNeighborsRegressor(n_neighbors = 7, weights = 'uniform')
knn.fit(x_train, y_train)

y_knn = knn.predict(x_test)

print('Score: %.3f' % knn.score(x_train, y_train))
print("Mean Squared Error: %.3f" % mean_squared_error(y_test, y_knn))
print('Mean Absolute error: %.3f' % mean_absolute_error(y_test, y_knn))
print('Variance Score: %.3f' % r2_score(y_test, y_knn))
```

```
Score: 0.823
Mean Squared Error: 2.132
Mean Absolute error: 1.215
Variance Score: 0.747
```

Figure 3 - k-Nearest Neighbors Regression Model

After testing 25% of the information (the rest of the information is being prepared for training), the chosen models showed the following results:

	Model	Mean Squared Error	Mean Absolute Error	Variance Score
0	Linear	1.020	0.761	0.879
1	Support Vector	1.562	1.038	0.814
2	k-Nearest Neighbors	2.132	1.215	0.747

Figure 4 - Selected models

And, definitely, the winner of these models is the Linear Regression Model which has a lower root mean square error (the lower the better), and a higher estimate of variance (the higher the better). This does not mean that the other two models - the Pivot Vector and k-Nearest Neighbors

Regression can be ignored, because they still have very impressive results, but not as strong as the Linear Regression Model.

Prediction

There are several datasets of 2017 and 2018 seasons. So here the prediction will be about the individual player win share.

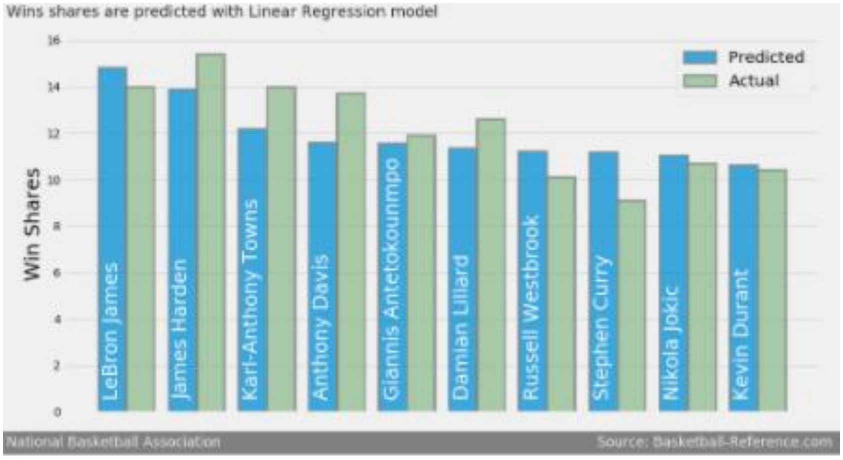

Figure 5 - Win Share Linear Regression

Based on the predictions of the Linear Regression Model (Figure 5), James LeBron "LABron" came out on top! The model predicted his lead in the NBA with 14.81. However, in fact, LeBron (14 winning shares) came in second after 2017-2018. MVP James Harden (15.4 winning shares) was second in the forecast. Not a bad forecast! Karl-Anthony Towns and Anthony Davis came in 3rd and 4th, respectively, in both the linear regression prediction and the last NBA season.

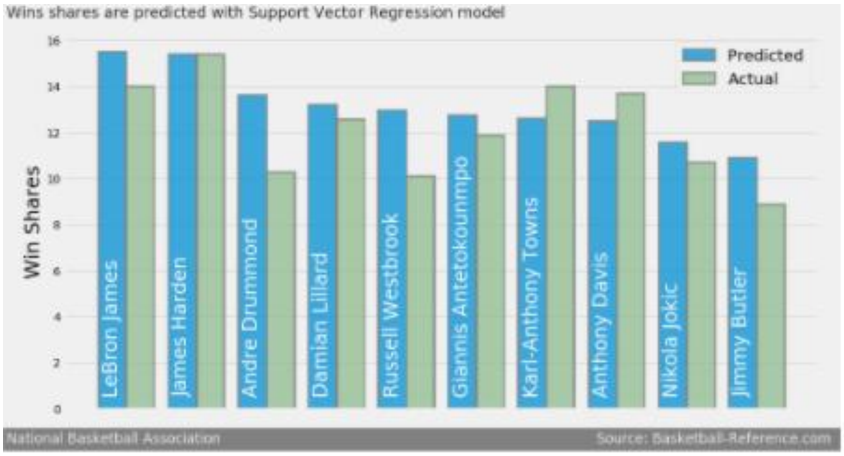

Figure 6 - Support Vector Regression

Win share predictions using the support vector regression model.

The Support Vector Regression Model produced unusual results which are shown in Figure 6. Based on the predictions of this model, LeBron (15.5 winning shares) came out on top again, while Harden came in second (15.4 wins). The most exciting result of this was that the model accurately predicted Harden's winning shares! He led the 2017–2018 NBA season with 15.4 winning shares, in line with the projected value. André Drummond, who was not in the top 10 players with predicted payoff using a Linear Regression model, came in third in terms of payoff using a Support Vector

Regression Model. Towns and Davis made the list of predictions, while Stephen Curry and Kevin Durant didn't even make the top ten.

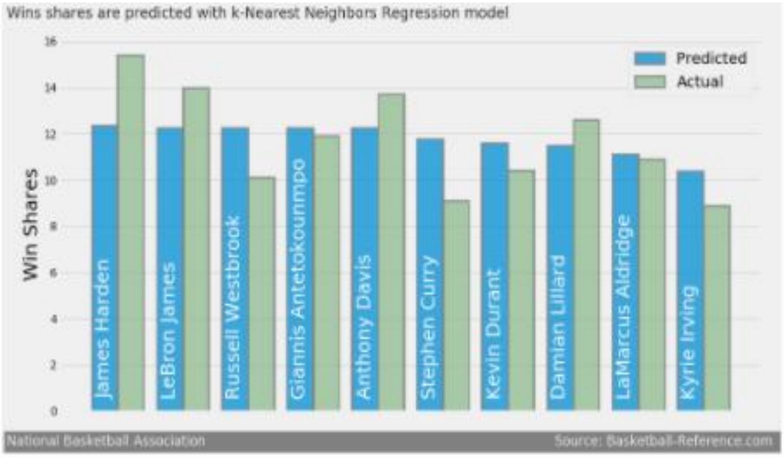

Figure 7 - k-Nearest Neighbors Regression

Win predictions using the k-Nearest Neighbors Regression Model

Winner share predictions using the k-Nearest Neighbors model were much less accurate than those of the previous models. For this model (Figure 7), Harden's predicted winning shares were the highest at just 12.3. Another bizarre result of the projected values was that LeBron, Russell Westbrook, Giannis Antotokumpo, and Davis were all associated with 12.24 winning shares!

Conclusio

In conclusion, the experimental comparison of the three prediction models has proved that the Linear Regression Model has got a high variance score, minimum scores in error tables and high calculation speed. The Linear Regression Model is also transparent and understandable for the analyst. Based on the obtained regression coefficients, one can judge how a particular factor affects the result and draw additional useful conclusions on this basis. Besides, there are many real processes in the economy and business that can be described with sufficient accuracy by Linear Regression Models for which typical problems (for example, multicollinearity) and their solutions are known, tests for assessing the static significance of the resulting models are developed and implemented.

REFERENCES

1. Regression Tree Model for Predicting Game Scores for the Golden State Warriors in the National Basketball Association [Electronic resource] URL: https://www.mdpi.com/2073-8994/12/5/835/htm (accessed: 16.05.2020)
2. Regression Techniques you should know [Electronic resource] URL: https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/ (accessed: 14.09.2015)
3. Building My First Machine Learning Model | NBA Prediction Algorithm [Electronic resource] URL: https://towardsdatascience.com/building-my-first-machine-learning-model-nba-prediction-algorithm-dee5c5bc4cc1 (accessed: 9.07.2020)
4. K-Nearest Neighbors Algorithm Using Python [Electronic resource] URL: https://www.edureka.co/blog/k-nearest-neighbors-algorithm/ (accessed: 20.11.2020)
5. Implementing SVM and Kernel SVM with Python's Scikit-Learn [Electronic resource] URL: https://stackabuse.com/implementing-svm-and-kernel-svm-with-pythons-scikit-learn/ (accessed: 24.04.2020)

Әлімхан А.М.

Терең оқыту алгоритмдерін қолдана отырып, баскетбол нәтижелерін болжау

Аңдатпа. Бұл мақалада ақпараттық технологиялардың дамуымен және үнемі өсіп келе жатқан статистикалық базамен болжау мүмкіндіктері кеңейіп, есептелген көрсеткіштердің нәтижеге тәуелділігі ескерілетіні жазылған. Сондай-ақ нәтижені болжау барысында болжаудың әртүрлі басқа модельдерін қолдануымыз керек және болжау моделінің нәтижесінде көп ұпайлары бар модельді таңдау үшін оларды салыстырамыз.

Түйінді сөздер*:* терең оқыту, машиналық оқыту, болжау модельдері, регрессия әдісі, деректерді талдау.

Әлімхан А.М.

Прогнозирование результатов игры в баскетбол с использованием алгоритмов глубокого обучения

Аннотация. В статье говорится о том, что с развитием информационных технологий и постоянно пополняющейся статистической базой расширяются возможности прогнозирования, учитываются зависимости рассчитываемых показателей от результата. Нам необходимо использовать разные модели прогнозирования, и мы сравниваем их, чтобы выбрать ту модель, которая дает возможность получить наиболее точные результаты.

Ключевые слова: глубокое обучение, машинное обучение, модели прогнозирования, метод регрессии, анализ данных.

Автор туралы мәлімет:

Әлімхан Абайхан Мамытханұлы, «Компьютерлік инженерия және ақпараттық қауіпсіздік» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті.

Сведения об авторе:

Әлімхан Абайхан Мамытханұлы, магистрант кафедры «Компьютерная инженерия и информационная безопасность», Международный университет информационных технологий.

About the author:

Alimkhan Abaikhan Mamitkhanuly, master student, Department of Computer Engineering and Information Security, International Information Technology University.

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Далабаева Айсара Касымбековна Джоламанова Балия Джалгасбаевна Медведев Евгений Юрьевич
Компьютерная верстка	Туратауова Айжаркын Ахметовна
Компьютерный дизайн	Туратауова Айжаркын Ахметовна

Редакция журнала не несет ответственности за
недостоверные сведения в статье и
неточную информацию по цитируемой литературе

Подписано в печать 26.06.2021 г.
Тираж 500 экз. Формат 60x84 1/16. Бумага тип.
Уч.-изд.л. 10.1. Заказ №165

Издание Международный университет информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59