

The object of the study is the automotive industry of the Republic of Kazakhstan. The subject of the study is the management of the decision-making process in assessing the consumer capabilities of potential customers of car dealerships, the process of forecasting car pricing.

A method using a global search engine optimization algorithm, a forest conveyor line with a random forest model with Bayesian optimization (RFBO), is proposed.

The algorithm of the method is as follows:

- obtaining and processing initial data taking into account the degree of uncertainty;*
- formation of the optimization vector;*
- creation of descendant vectors;*
- ordering of vectors in descending order;*
- reducing the dimension of the feature space;*
- knowledge base training.*

In the presented work, data from websites www.m.Kolesa.kz, www.Cars.com and the average values of the median salary in the Republic of Kazakhstan were used to create a knowledge base, the program code of the platform was created using the Visual Studio Code in the Python language.

The task to be solved was to predict car prices and assess the consumer capabilities of potential car dealership customers.

We evaluate our solution based on a dataset that was created by analyzing several car classified sites and data on potential customers. Our results show that the accuracy of the model training was 92.1 %, and the accuracy of forecasting car prices and evaluating the consumer capabilities of potential customers was 87.3 % – this is primarily due to lower prediction errors than those of the estimated regressors using the same set of input data, high-quality object mapping and a more competitive RFBO algorithm, superior to simple linear models.

The developed software solution should be used for making automated management decisions by car dealerships and credit organizations

Keywords: big data, intelligent system, potential customer, purchasing power, automation of managerial decision-making

DEVELOPMENT OF AN INTELLIGENT SYSTEM AUTOMATING MANAGERIAL DECISION-MAKING USING BIG DATA

Karshyga Akishev

Candidate of Technical Sciences, Honorary Professor*

Amandos Tulegulov

Candidate of Physical and Mathematical Sciences, Head of Department, Honorary Professor*

Aslan Kalkenov

Master's Student 2nd Years of Study*

Kapar Aryngazin

Candidate of Technical Sciences, Professor Department of Architecture and Design

Toraighyrov University

Lomov str., 64, Pavlodar, Republic of Kazakhstan, 140008

Zhadira Nurtai

Corresponding author

PhD, Associate Professor*

E-mail: zhadira_nurtai@mail.ru

Dastan Yergaliyev

Candidate of Technical Sciences, Professor Department of Aviation Engineering and Technology

Academy of Civil Aviation

Zakarpatskaya str., 44, Almaty, Republic of Kazakhstan, 050039

Yergesh Manas

Doctoral Student of the 1st Year of Study

Department of Information Technology

L. N. Gumilyov Eurasian National University

Satpayeva str., 2, Astana, Republic of Kazakhstan, 010008

Ainura Jumagaliyeva

Senior Lecturer, Master's Degree*

*Department of Information Technologies

Kazakh University of Technology and Business

Mykhamedkhanov str., 37a,

Astana, Republic of Kazakhstan, 010000

Received date 22.09.2023

Accepted date 15.12.2023

Published date 29.12.2023

How to Cite: Akishev, K., Tulegulov, A., Kalkenov, A., Aryngazin, K., Nurtai, Z., Yergaliyev, D., Manas, Y., Jumagaliyeva, A. (2023). Development of an intelligent system automating managerial decision-making using big data. *Eastern-European Journal of Enterprise Technologies*, 6 (3 (126)), 27–35. doi: <https://doi.org/10.15587/1729-4061.2023.289395>

1. Introduction

According to the American marketing agency Hedges & Company, today the total number of cars in the world for 2023 is 1.43 billion units. The dynamics of car sales are shown in Fig. 1 [1].

At the same time, the number of used cars in the world is growing year by year, while the volume of new car production is falling for objective reasons.

To date, according to the Statista source, the global used car market has exceeded 1.27 trillion, and the annual growth is 3 %.



Fig. 1. Dynamics of car sales in the world for the period 1998–2020 (the data is obtained from the source «Statista») [1]

The trend that has developed since the pandemic shows the dynamics of car growth through Internet sites, while prices on such sites are based on the conclusions of site owners and do not quite meet the wishes of potential customers.

In this regard, research is needed to develop software tools using various algorithms, methods to improve the efficiency and reliability of determining car prices taking into account a wide range of input data, allowing to exclude the influence of the human factor.

Thus, studies that are devoted to forecasting car prices and evaluating the consumer capabilities of potential customers have scientific significance and are important because the decision to purchase a car must have a strong justification for any consumer.

2. Literature review and problem statement

The process of pricing cars and evaluating the consumer capabilities of potential customers is quite an interesting task, which depends on various factors and is solved by various approaches.

For example, in developed countries, car prices are guided by demand and standard of living, various marketing campaigns are held to attract buyers, discounts are offered on non-popular models [2, 3].

For less developed countries, there are more complex problems associated with the purchase of cars.

In [4], the method of supervised learning, random forest, was used to predict prices for used cars, a random forest with 500 decision trees was created for data processing, training accuracy was 95.82 %, and testing accuracy was 83.63 %. The disadvantage of the study is that forecasting is performed only for used cars, and the sample is limited. There is no methodology for evaluating the purchasing power of potential customers.

In [5], machine learning and object detection methods are used. A combination of several adaptive learning methods (deep learning) has been proposed, which can determine the characteristics of a car by license plates. The disadvantages of the study are the limited data set, there is no data on predictive estimates, and further research is required.

In [6], the methods of automatic machine learning are used, the forecast for a used car using 4 algorithms. The disadvantage of the study is that it does not have an algorithm for the operation of the interface, does not specify qualitative and quantitative indicators, the effectiveness of predictive estimates based on the algorithms used in the study.

In [7], the Kaggle platform is used for machine learning and open source data analysis to predict used car prices using controlled algorithms. The disadvantages of the study are a controlled algorithm, the absence of influence of external factors. The work requires further research.

In [8], multiple linear regression is used to predict car prices. The disadvantages of the study include the limited data set, the lack of a decision-making function, and there is no data on the effectiveness of a predictive assessment.

In [9], Shapley’s calculations are presented using the Apache algorithm based on clustering predictive estimates of car prices. The disadvantage is the complexity of the calculation algorithm, possible errors in the distribution of prices by class and a limited data set.

In [10], experiments were conducted using various algorithms to estimate forecasted car prices, on the basis of which it was revealed that the random forest algorithm is the most effective. The disadvantage of the study is the lack of practical implementation of the study, the use of an experimental data set.

In [11], three machine learning methods (artificial neural network, support vector machine, and random forest) are used in a complex to predict car prices. Using a web scraper written in the PHP programming language. The forecasting model has been integrated into a Java application. The model’s prediction accuracy was 87.38 %. The disadvantage of the study is the need for large computing resources and a limited amount of input data.

All this allows us to assert that it is advisable to conduct research on problems related to forecasting car prices with an expanded data set, evaluating customer’s consumer capabilities using the most effective algorithms and automated management decision-making, since none of these functions are considered in any of the works.

3. The aim and objectives of the study

The aim of the study is to develop a platform for assessing the solvency of potential buyers of car dealerships and car dealers.

To achieve the aim, the following objectives were accomplished:

- to develop a flowchart of the program;
- to develop the software architecture of the platform;
- to evaluate the capabilities of the developed platform.

4. Materials and methods

The object of the study is the automobile market of the Republic of Kazakhstan. The study uses a random forest model with Bayesian optimization (RFBO) to evaluate the effectiveness of machine learning models and assess the consumer capabilities of potential car dealership customers.

An algorithm for generating packets using a random forest and Bayesian optimization is proposed, which is a promising solution for modeling and surpasses traditional data balancing methods and increases the accuracy of model balancing and forecasting. We assume that the choice of the proposed method is justified when working with disparate and nonequilibrium datasets. At the same time, the limitations associated with optimizing model parameters using Bayesian optimization do not affect the assessment of consumer capabilities of potential car dealership customers.

The first classifier puts its imprint on the re-selected copies of the training sample without any new information in the training sample. The main random forest method used in the study is that if classification accuracy is considered as the main goal of the prediction model, then a separate classifier can be completely excluded and an ensemble method can be used. Later, a DELL Optiplex 5270 AIO computer with an Intel Core i5 processor, 8GB, 512 GB, Windows 11, Python software, and a Jupyter laptop were used.

5. Evaluation of research results

5.1. Development of the program flowchart

The program code was created using VisualStudioCode software tools in the Python language. The flowchart of the program is shown in Fig. 2.

The flowchart describes the sequence of actions that can be performed in the program:

- 1) the beginning;
- 2) working with the database;
- 3) data cleaning (cleaned data);
- 4) application of the random forest model: for data analysis or forecasting;

- 5) saving the conveyor: for later use;
- 6) loading pipeline: for performing additional operations;
- 7) launching Streamlet (ru stream lit);
- 8) exploratory data analysis (eda);
- 9) user input (user_input);
- 10) condition check (if x==name_col true, otherwise false);
- 11) status check (if input_data==8 true, otherwise false);
- 12) status check (if car_cost>initial_payment is true, otherwise false);
- 13) completion (end).

5.2. Development of the software architecture of the platform

Fig. 3 shows the basic structure of the project.

Create a special file in the «spiders» directory (Fig. 4). This file contains all the code and instructions needed to extract data from a web page. To create a database, data is downloaded from the «Koleso» website.

We need to make a note that it is advisable to understand the page codes that we will use for subsequent processing (identifiers, classes, tags and other attributes). In particular, for the Kolesa.kz website, this looks like Fig. 5. Since the site is russian-language, the data in the database is presented in russian.

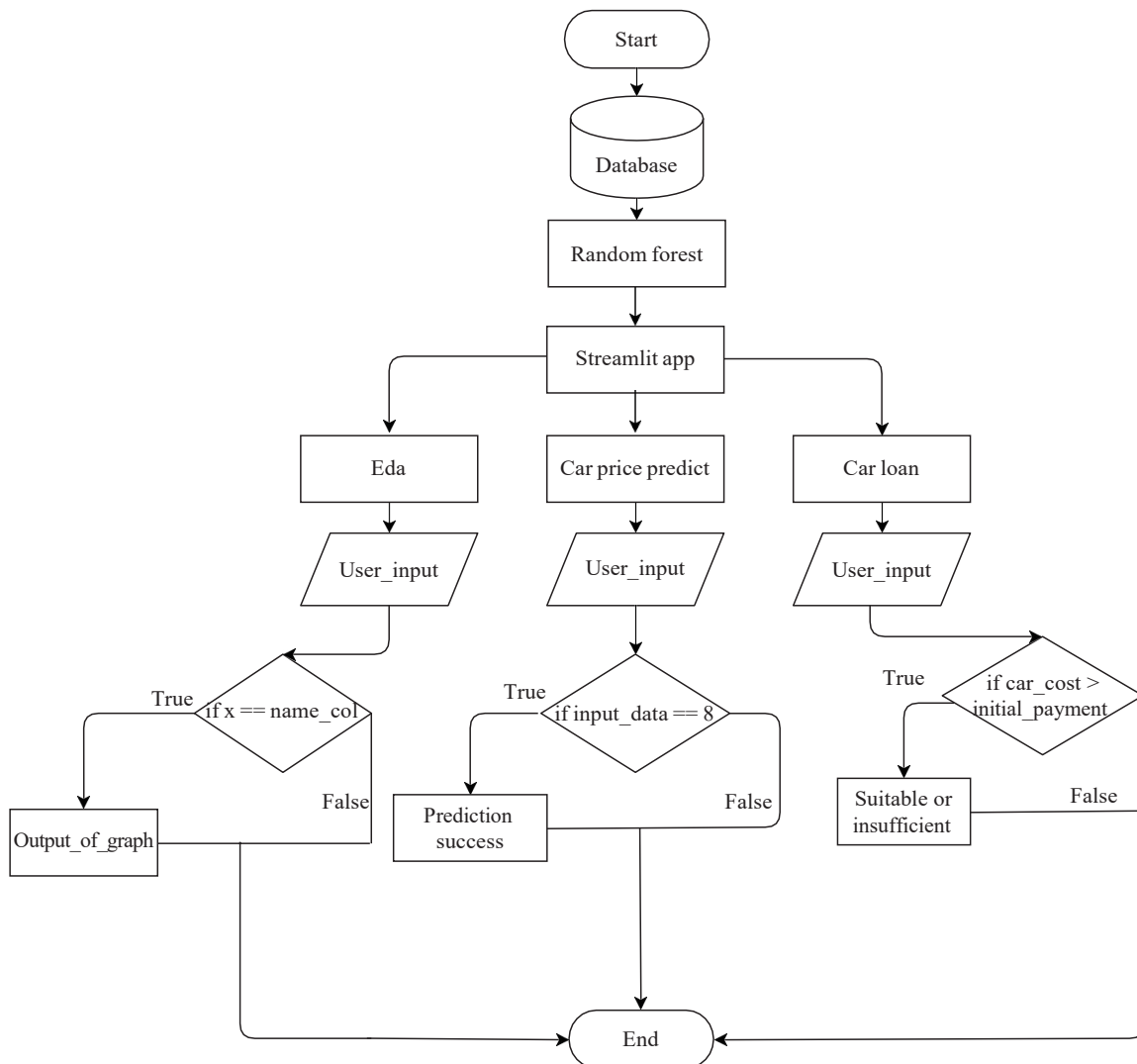


Fig. 2. Program flowchart

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
PS C:\Users\User\OneDrive\Рабочий стол\dismag> scrapy startproject kolesa
New Scrapy project 'kolesa', using template directory 'C:\Users\User\AppData
tes\project', created in:
    C:\Users\User\OneDrive\Рабочий стол\dismag\kolesa

You can start your first spider with:
    cd kolesa
    scrapy genspider example example.com
PS C:\Users\User\OneDrive\Рабочий стол\dismag>
    
```

Fig. 3. Basic structure for the project

```

PS C:\Users\User\OneDrive\Рабочий стол\dismag> cd .\kolesa\kolesa\
PS C:\Users\User\OneDrive\Рабочий стол\dismag\kolesa\kolesa> ni kolesaspider.py

Каталог: C:\Users\User\OneDrive\Рабочий стол\dismag\kolesa\kolesa

Mode                LastWriteTime         Length Name
----                -
-a----             02.09.2023   23:26             0 kolesaspider.py

PS C:\Users\User\OneDrive\Рабочий стол\dismag\kolesa\kolesa>
    
```

Fig. 4. Create a working file

Fig. 5. HyperText Markup Language (HTML) page code for kolesa.kz

Scrapy Shell speeds up the process of developing and debugging code. Python code for data mining – developed using the Scrapy framework to collect car sales data from the kolesa.kz website.

The SpiderkolesaSpider class is a subclass of scrapy. Spiders, the basic structure of scraper, for attributes, name, allowed_domains and start_urls.

Parse_page for processing, parse_page(self, response) extracts links and passes them to the parse_contents method.

Parse_contents (self, response) data for each car.

CSS selectors extract information from the HTML code on the technical characteristics and the place of sale of the car and the price (Fig. 6).

The received data is recorded in a Comma-Separated Values (CSV) file, which is located in the current working directory and will contain information about the sale of cars (technical specifications, price, etc.) (Fig. 7).

The CSV file can be easily imported and used for data analysis using Python or other data science tools. This makes

the scraper a useful tool for collecting and analyzing information, for example, about the sale of cars from the kolesa.kz website (Fig. 8).

The program being developed is created in VisualStudioCode, enter «CreateNewJupyterNotebook» on the dropdown list (Fig. 9). Create a Jupyter workbook, open inside VS Code, add code to describe the data processing process, model training, car price forecasting.

In our work, we use the NumPy, Pandas, Matplotlib, pyplot libraries.

Seaborn is a library, an addition to Matplotlib for creating informative statistical graphs.

The program created a variable called "kolesadata" (Fig. 10), and the variable "df" was nested in it. The variable "df" is a data structure containing information about various cars. The variable "kolesadata" can now be used in the program to access and process this data. This is convenient because it allows you to temporarily work with vehicle information and perform various analyses and operations with this data.

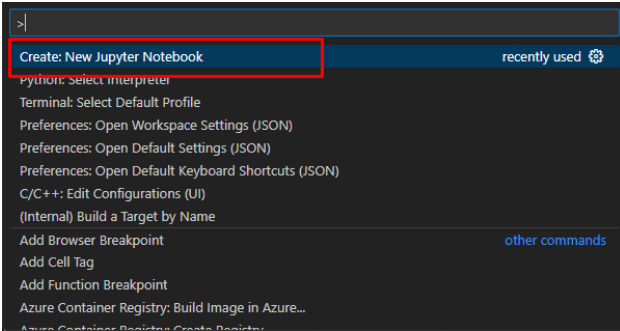


Fig. 9. Creating a Jupyter notebook project

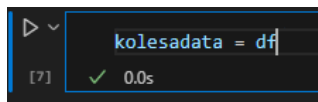


Fig. 10. Creating a new container

The function `kolesadata.isnull().sum()` analyzes the data in the `kolesadata` object to identify and count missing values for making decisions on processing missing values (Fig. 11).

The machine learning method associated with forecasting car prices consists of the following components:

- printing (`kolesadata["body"].unique()`);
- print (`kolesadata["transmission"].unique()`);
- printing(`kolesadata["wheel"].unique()`);

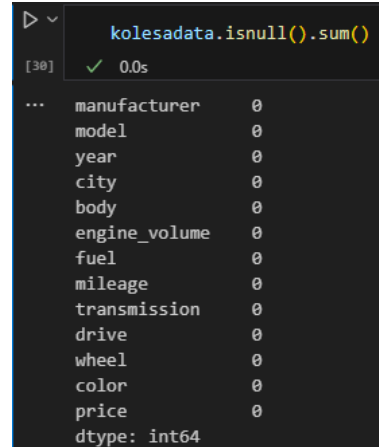


Fig. 11. Checking for missing values using `kolesadata.isnull().sum()`

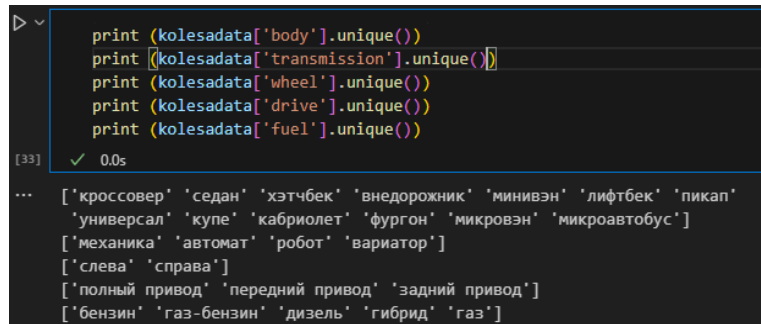


Fig. 12. Model training algorithm

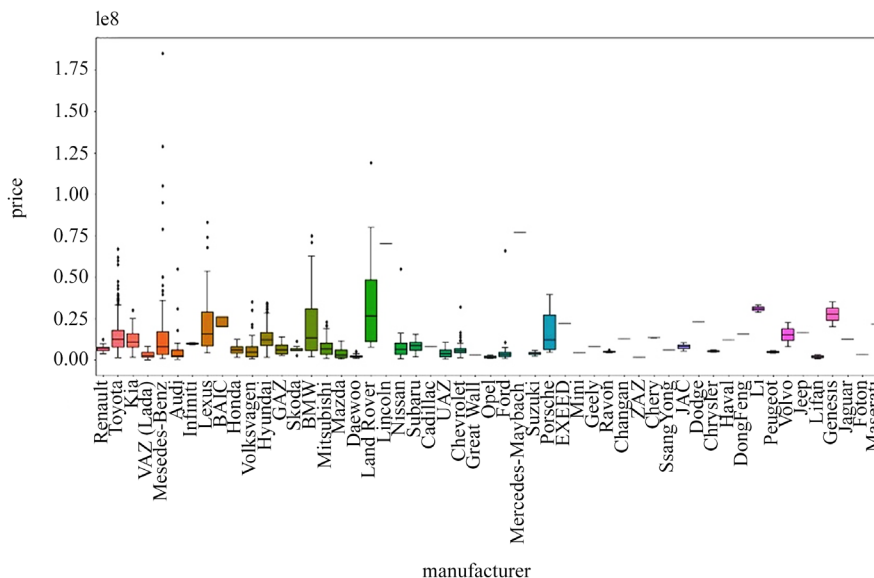


Fig. 13. Price distribution depending on transmission type

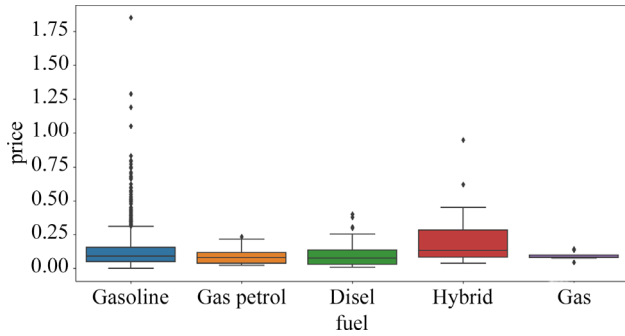


Fig. 14. Comparison of prices for cars by fuel type

In one paper, it is impossible to describe the entire methodology for developing a software tool for assessing the purchasing power of a potential client of a car dealership, so some details will be omitted.

5. 3. Evaluation of the capabilities of the platform «Evaluation of the purchasing power of car dealership customers»

By executing the stream let run command.\car_price_pred.py, the web application is automatically launched, the system generates a message about the availability of the web application in the browser, navigating to the Uniform Resource Locator (URL) and working with the web application (Fig. 15).

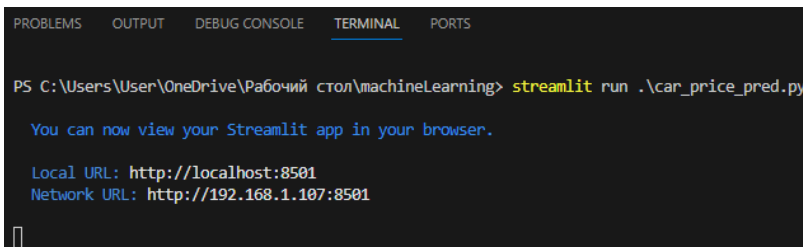


Fig. 15. Application launch process

A general view of the program menu for assessing the purchasing power of potential customers of car dealerships is shown in Fig. 16. The program menu consists of 4 directories.

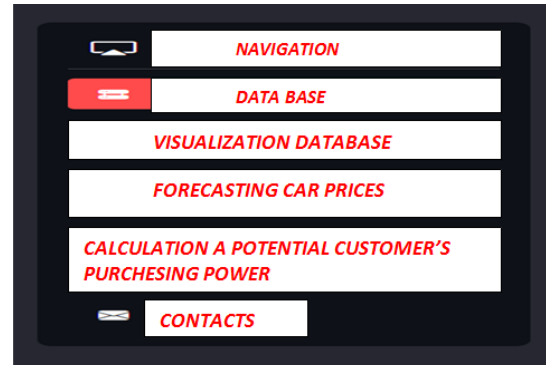


Fig. 16. Menu of the program «Assessing the purchasing power of potential customers of car dealerships»

The database contains data on manufacturers, models, years of manufacture, transmission type, fuel types and prices for cars (Fig. 17).

The calculation of the purchasing power of a potential buyer is carried out automatically with the appropriate characteristics and decisions about the possibility or impossibility of selling (Fig. 18) the car to the buyer (data are entered in russian).

The possibility of making a decision by the program ensures the exclusion of the influence of the human factor and depends only on the quality of information from the knowledge bases and the algorithm of the program.

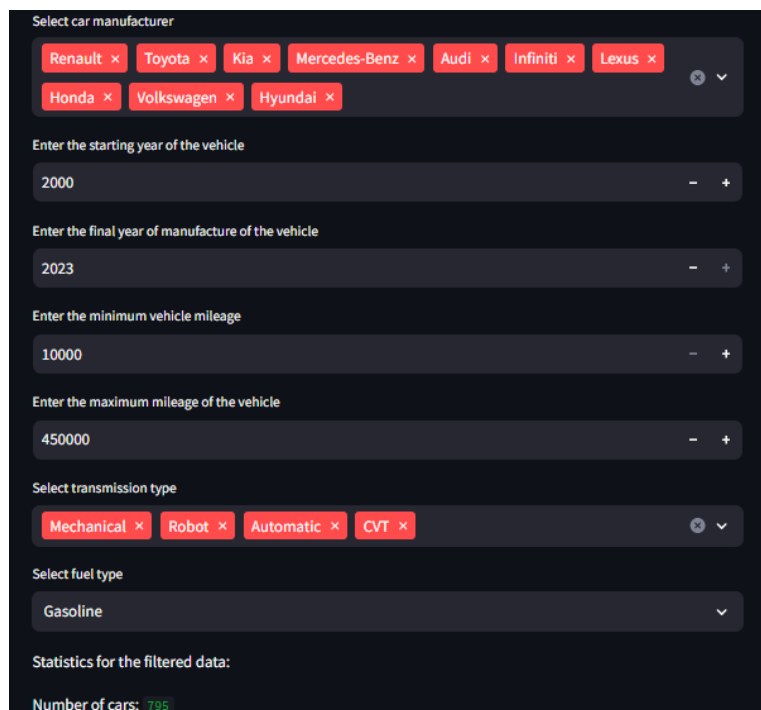


Fig. 17. Car price forecasting

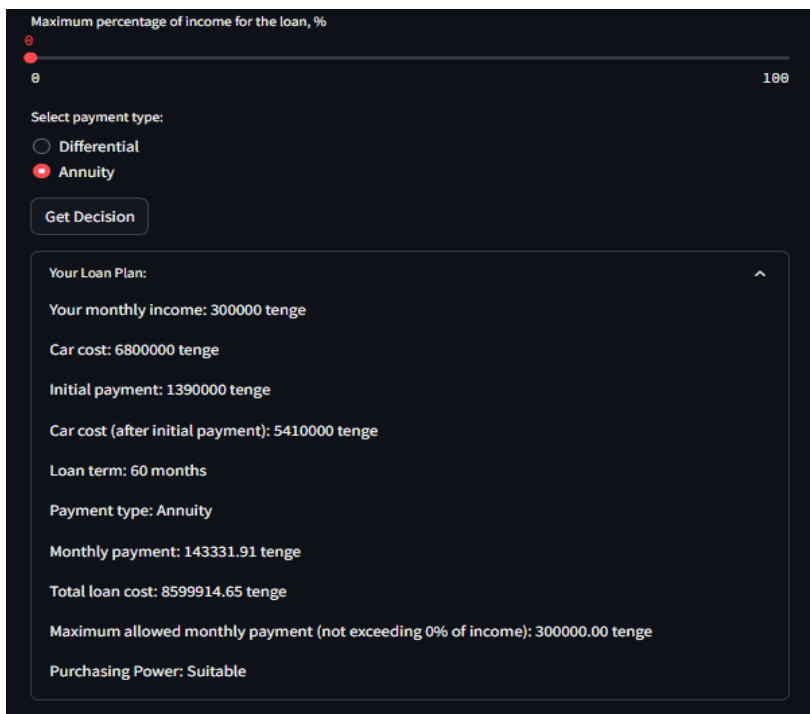


Fig. 18. Option for making a decision by the program

6. Discussion of the potential of the proposed platform «Assessment of consumer capabilities of potential car dealership customers»

The main obstacles to the promotion or commercialization of the research results may be the unwillingness of car dealerships to use unfamiliar developments, the need for marketing research.

An analysis of the work on the subject of the study shows that the solutions presented in the paper have novelty not only in terms of filling the knowledge base, but also in assessing managerial decision-making by potential clients. The platform presented in [9] and the work [12] have the functionality for making managerial decisions, but they have a number of functional limitations.

The results are achieved by providing the following qualitative indicators:

- firstly, the platform algorithm provides the most complete and high-quality display of data on cars sold in the Republic of Kazakhstan, optimal parameters for assessing the purchasing capabilities of potential customers (Fig. 16–18);
- secondly, the platform algorithm implements the possibility of obtaining traffic data and the dynamics of wages in the country;
- thirdly, the influence of the human factor and risks is excluded.

The disadvantages of the developed platform are limited access to the databases of automakers, the availability of stable data transfer to knowledge bases.

Limits of application – automakers, car dealerships, second-tier banks, microfinance organizations. The conditions for applying the results obtained are a qualitative analysis of the consumer capabilities of potential customers of car dealerships based on big data and the use of a knowledge base.

In future research, it is possible to study the effectiveness of these methods compared to the algorithm used.

7. Conclusions

1. The proposed data collection algorithm, random forest with Bayesian optimization (RFBO), surpasses traditional methods of forecasting car prices by using an expanded set of input data showing an increase in accuracy compared to other traditional models, while the accuracy of model training was 91.2 %, and the accuracy of forecasting car prices and evaluating the consumer capabilities of potential customers was 84.3 %.

2. The developed architecture of the program makes it possible to use data collection algorithms more effectively to predict car prices for any category.

Thus, it can be concluded that the proposed algorithm using the random forest method and Bayesian optimization is a promising solution for solving predictive problems on car prices, allowing such results to be obtained with 91.2 % accuracy in model training, and 84.3 % accuracy in predicting car prices and evaluating the consumer capabilities of potential customers. Moreover, it has been determined that fine-tuning hyperparameters using Bayesian optimization improves the equilibrium of the model and the accuracy of forecasting.

3. We evaluate our solution based on an expanded dataset that was created by analyzing several car classified sites and potential customer data. Our results show that the accuracy of the model training was 91.2 %, and the accuracy of forecasting car prices and evaluating the consumer capabilities of potential customers was 84.3 % – this is primarily due to lower prediction errors than those of the estimated regressors using the same set of input data, high-quality object mapping and a more competitive RFBO algorithm, superior to simple linear models.

The developed software solution should be used to make automated management decisions by car dealerships and credit organizations.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and the results reported in this paper.

Financing

The study was performed without financial support.

Data availability

All data are available in the main text of the manuscript.

Acknowledgments

We express special gratitude to the Honorary Professor of the Kazakh University of Technology and Business Akishev Karshyga for professional advice and competent guidance at all stages of our research.

We sincerely hope that our research will allow small and medium-sized businesses in the Republic of Kazakhstan to find reliable and loyal customers.

Use of artificial intelligence

The authors confirm that artificial intelligence was not used in the development of the platform.

References

1. Mirovoy rynek poderzhannyh avtomobiley. Available at: <https://dzen.ru/a/YcyFhayywxZ8S5on>
2. Rekordy avtoprodazh stavyat v Kazahstane uzhe vtoroy mesyats. Available at: <https://informburo.kz/novosti/rekordy-avtoprodazh-stavyat-v-kazahstane-uzhe-vtoroi-mesyac>
3. V Kazahstane nablyudaetsya defitsit kachestvennyh i dostupnyh avto. Available at: https://forbes.kz/economy/auto/v_kazahstane_nablyudaetsya_defitsit_kachestvennyh_i_dostupnyh_avto/
4. Pal, N., Arora, P., Kohli, P., Sundararaman, D., Palakurthy, S. S. (2018). How Much Is My Car Worth? A Methodology for Predicting Used Cars' Prices Using Random Forest. *Advances in Information and Communication Networks*, 413–422. doi: https://doi.org/10.1007/978-3-030-03402-3_28
5. Yadav, A., Kumar, E., Yadav, P. K. (2021). Object detection and used car price predicting analysis system (UCPAS) using machine learning technique. *Linguistics and Culture Review*, 5 (S2), 1131–1147. doi: <https://doi.org/10.21744/lingcure.v5ns2.1660>
6. Gaikwad, D. A., Suwarnakar, P. S., Mahajan, Y. R., Petkar, A. U., Theurkar, S. G. (2023). Used Car Price Prediction Using Random Forest Algorithm. *International Journal For Multidisciplinary Research*, 5 (3). doi: <https://doi.org/10.36948/ijfmr.2023.v05i03.3308>
7. Abishek, R. (2022). Car Price Prediction Using Machine Learning Techniques. *International Research Journal of Modernization in Engineering Technology and Science*, 4 (2), 54–50. Available at: https://www.irjmets.com/uploadedfiles/paper//issue_2_february_2022/18785/final/fin_irjmets1643874132.pdf
8. Dholiya, M., Tanna, S., Balakrishnan, A., Dubey, R., Singh, R. (2019). Automobile Resale System Using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*, 06 (04), 3122–3125. Available at: <https://www.irjet.net/archives/V6/i4/IRJET-V6I4666.pdf>
9. Kundu, S., Maulik, U. (2021). Cloud deployment of game theoretic categorical clustering using apache spark: An application to car recommendation. *Machine Learning with Applications*, 6, 100100. doi: <https://doi.org/10.1016/j.mlwa.2021.100100>
10. Carvalho, G., Viana, D., Barbosa, L., Ren, T. I. (2023). Combining Structured and Unstructured Data using Co-Attention for Car Price Prediction. *Procedia Computer Science*, 222, 646–655. doi: <https://doi.org/10.1016/j.procs.2023.08.202>
11. Gegic, E., Isakovic, B., Keco, D., Masetic, Z., Kevric, J. (2022). Car Price Prediction Using Machine Learning Techniques. *TEM Journal*, 8 (1), 113–118. doi: <https://doi.org/10.18421/TEM81-16>
12. Dudnyk, V., Sinenko, Y., Matsyk, M., Demchenko, Y., Zhyvotovskiy, R., Repilo, I. et al. (2020). Development of a method for training artificial neural networks for intelligent decision support systems. *Eastern-European Journal of Enterprise Technologies*, 3 (2 (105)), 37–47. doi: <https://doi.org/10.15587/1729-4061.2020.203301>