International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 37 العدد
Part 2 المجلد
المجلة الدولية للعلوم والتقنية
International Science and Technology Journal
ISTJ

http://www.doi.org/10.62341/lbqf5637

Received	**2025/11/23**	**تم استلام الورقة العلمية في**
Accepted	**2025/12/14**	**تم قبول الورقة العلمية في**
Published	**2025/12/16**	**تم نشر الورقة العلمية في**

Evaluating Automobile Quality Using the Random Forest Algorithm

Aisha Muftah Eghwila[1], Eman Bashir Alghwil[2]

Alasmarya Islamic University - Zliten city - Libya
1- a.ghuayla@asmarya.edu.ly, 2- e.alghwil@asmarya.edu.ly

Abstract

Machine Learning uses ensemble methods, where the smaller models that are combined to form the ensemble model are referred to as basic models. Ensemble methods often result in considerably higher performance than any of the individual base models could achieve. The problem with the study is that the results of previous studies of car quality prediction using the K-means algorithm was by classifying the six characteristics of car specifications (buying, maint, doors, persons, lug_boot, safety), so that its quality is evaluated and the result using the K-means algorithm is Only 56%. so, this study presents a description of the quality assessment of cars based on the random forest algorithm by classifying the six characteristics of the car specifications, so that their quality is evaluated by the Python program. In this research, the ensemble method is called bagging, which is short for bootstrap aggregating. Bagging builds multiple base models with resampled training data with replacement. During the analysis and comparison of the results of the random forest algorithm with the results of the k-means algorithm, the results of the random forest algorithm were better than the k-means algorithm with the same data set, the accuracy of the random forest algorithm was 98.07%, while the accuracy of the k-means algorithm was 56%.

Keywords: random forest, automobile quality, K-means algorithm.

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 37 العدد
Part 2 المجلد
http://www.doi.org/10.62341/lbqf5637

تقييم جودة السيارات بالاعتماد على خوارزمية الغابة العشوائية

عائشة مفتاح اغويله
كلية تقنية المعلومات الجامعة الأسمرية الإسلامية
a.ghuayla@asmarya.edu.ly

ايمان بشير الغويل
كلية تقنية المعلومات الجامعة الأسمرية الإسلامية
e.alghwil@asmarya.edu.ly

الملخص:
التعلم الآلي يستخدم أساليب المجموعة، حيث يُشار إلى النماذج الأصغر التي يتم دمجها لتكوين نموذج المجموعة بالنماذج الأساسية. غالبًا ما تؤدي أساليب المجموعة إلى أداء أعلى بكثير مما يمكن أن يحققه أي من النماذج الأساسية الفردية. تكمن مشكلة الدراسة في أن نتائج الدراسات السابقة للتنبؤ بجودة السيارة باستخدام خوارزمية k–means كانت من خلال تصنيف الخصائص الستة لمواصفات السيارة (الشراء، الصيانة، الأبواب، الأشخاص، صندوق الأمتعة، السلامة)، بحيث يتم تقييم جودتها وتكون النتيجة باستخدام خوارزمية k–means 56% فقط . لذا، تقدم هذه الدراسة وصفًا لتقييم جودة السيارات بناءً على خوارزمية الغابة العشوائية من خلال تصنيف الخصائص الستة لمواصفات السيارة، بحيث يتم تقييم جودتها عن طريق برنامج بايثون. في هذا البحث، تسمى طريقة التجميع بالتعبئة، وهي اختصار لتجميع التمهيد. تقوم عملية التعبئة ببناء نماذج أساسية متعددة مع بيانات تدريب مُعاد تشكيلها مع الاستبدال. أثناء تحليل ومقارنة نتائج خوارزمية الغابة العشوائية مع نتائج خوارزمية k–means، كانت نتائج خوارزمية الغابة العشوائية أفضل من خوارزمية k–means التي لها نفس مجموعة البيانات، وبلغت نسبة دقة خوارزمية الغابة العشوائية 98.07% بينما بلغت دقة خوارزمية k–means 56%.

الكلمات المفتاحية :الغابة العشوائية، جودة السيارات، خوارزمية.K-means

1. Introduction

The area of computer science called Artificial Intelligence (AI) is dedicated to developing systems that can perform functions that would normally need human intelligence. intelligence, including skills like decision-making, problem-solving, trend identification, natural language comprehension, and learning from experience. In

http://www.doi.org/10.62341/lbqf5637

order to replicate human-like cognitive processes, analyze huge amounts of data, and identify patterns, AI systems employ algorithms to accomplish this. The goal of artificial intelligence is to create robots that can solve problems, adapt to different situations, and function autonomously [1][2]. In general, AI technologies may be divided into two categories: those that are general AI and those that seek to carry out any cognitive task. a limited kind of AI that can perform specific functions that humans can accomplish, such as picture analysis or voice identification. AI has a lot of potential, but it also brings up important ethical and social issues related to privacy, prejudice, and employment implications [3][4][5].

2. Related Work

Lu, Yiqi, et al., 2018, the study presents a forecasting method of electric vehicle EV which charging load based on Random forest algorithm and the load data of a single charging station. This method is accomplished by the classification and regression tree (CART) algorithm to realize short-term forecast for the station. According to the simulation results, the prediction can reach the mean absolute percentage error a MAPE of 9.76% and the root mean square error a RMSE of 2.27. It can be used as a charging prediction method to provide reference for various EV charging load control strategies. The proposed charging prediction algorithm of station group can effectively track the estimated daily charging capacity of different charging stations, based on the actual recorded data. According to the simulation results of the optimized forest structure, the prediction can reach a MAPE of 10.83% and a RMSE of 39.59. Also, it can be used for practical application [6].

Collard, Marcus, 2022, the study compares the performance of linear regression, ridge regression, lasso regression, and random forest regression ML algorithms in predicting the price of used cars. The study has been conducted with a large public dataset of used cars. The results show that random forest regression demonstrates the highest price prediction performance across all metrics used. It was also able to represent average depreciation much more closely than the other algorithms, at 13.7% predicted annual geometric depreciation for the dataset independent of vehicle age [7].

http://www.doi.org/10.62341/lbqf5637

Putra, Purwa Hasan, et al., 2023, the researchers used random forest and decision tree methods to predict car prices. The results using the random forest and decision tree methods have different percentage. When using the random forest method there is an accuracy = 72.13%, whereas with the analysis of the decision tree method gives an accuracy 67.21%. So it can be concluded that the random forest method has a better analytical accuracy than the decision tree method [8].

Yang, et al., 2023. The study examines the use of auto traffic accident data from 2018 to 2020 in China's national auto accident in-depth investigation system. Random forest was used to classify the significance of 12 incident features, and 7 significant incident features were finally adopted. By comparing the algorithms and improving the results, the traffic accident degree prediction model with higher accuracy was finally obtained [9].

3. Dataset and Preprocessing

The stages of the suggested algorithm design are as follows: input and data set definition as shown in Table 1, six distinct features (purchase, maintenance, doors, people, Lug_boot, and safety) each of which has a set of values as shown in Table 2, and finally, the number of instances: 1728 (the attribute space is entirely covered by instances) [10][11].

Table 1: Attributes Used by The Model

Attributes	The meaning
buying	buying price
maint	price of the maintenance
doors	number of doors
persons	capacity in terms of persons to carry
lug_ boot	the size of luggage boot
safety	estimated safety of the car

Table 2: Attribute Values

Attribute Values				
Buying	v-high	High	Med	Low
maint	v-high	High	Med	Low
doors	2	3	4	5-more
Persons	2	4	More	\
lug_ boot	Small	Med	Big	\
Safety	Low	Med	High	\

http://www.doi.org/10.62341/lbqf5637

4. Research Design

Figure 1 shows the proposed system for predicting the car quality using six features of the vehicle specs.

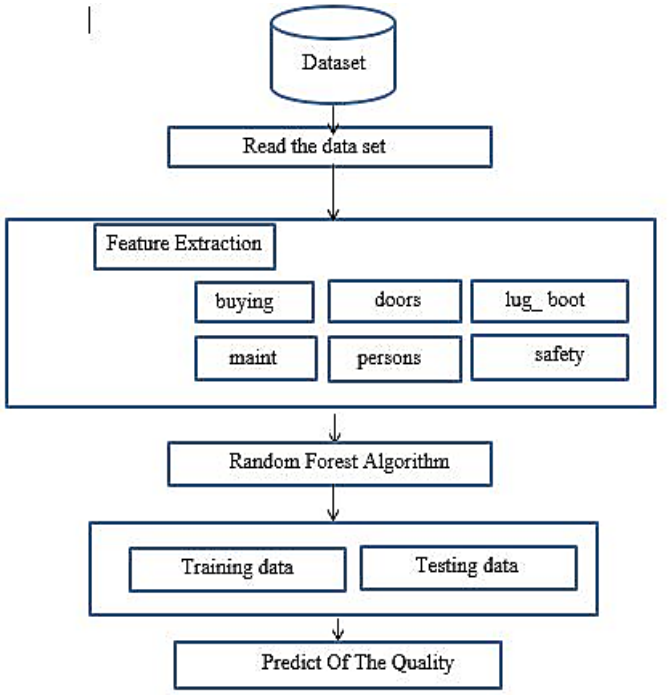

Figure 1: Diagram of Proposed System

5. Predict the quality

After predicting the car quality using six features of the vehicle specs, the random forest algorithm's accuracy was computed and the degree of class categorisation in percentage terms is displayed in Table 3. The class will be Unacc if, for instance, 1210 records out of 1728 are classified.

Table 3: Class Distribution

Class	N	N[%]
Unacc	1210	(70.023%)
Acc	384	(22.222%)
Good	69	(3.993%)
v-good	65	(3.762%)

http://www.doi.org/10.62341/lbqf5637

6. Python Implementation

The implementation offered by the scikit-learn Python machine learning framework will be used in this study to create a basic random forest model for predicting car quality given other automotive variables[12][13][14] (figure 2).

```
      buying  maint  doors  persons  lug_boot  safety
580        1      1      1        1         1       1
1214       2      3      0        2         2       2
1665       3      3      1        2         0       0
661        1      2      0        1         1       1
587        1      1      1        2         0       2
...      ...    ...    ...      ...       ...     ...
835        1      3      2        2         2       1
1216       2      3      1        0         0       1
1653       3      3      1        0         2       0
559        1      1      0        2         0       1
684        1      2      1        1         0       0

[1209 rows x 6 columns]        buying  maint  doors  persons  lug_boot  safety
1318       3      0      0        2         1       1
124        0      1      0        1         2       1
648        1      2      0        0         0       0
249        0      2      1        0         2       0
1599       3      2      3        0         2       0
...      ...    ...    ...      ...       ...     ...
906        2      0      1        1         2       0
309        0      2      3        1         1       0
838        1      3      3        0         0       1
1076       2      1      3        2         1       2
1453       3      1      1        2         1       1

[519 rows x 6 columns] 580     0
1214    2
1665    0
661     0
587     1
       ..
835     1
1216    0
1653    0
559     0
684     0
Name: class, Length: 1209, dtype: int32 1318    0
124     0
648     0
249     0
```

```
1599    0
       ..
906     0
309     0
838     0
1076    1
1453    1
Name: class, Length: 519, dtype: int32
>>>
```

Figure 2: Data Randomly into 70% Training and 30% Test

7. Result and Discussion

With a consistent, task-oriented interface, Scikit-learn displays a wide variety of machine learning approaches, both supervised and unsupervised, making comparison easy. methods for a specific application. Its use of the scientific Python ecosystem makes it simple to integrate into applications outside of the traditional field of Statistical data analysis. The equations from Chapter Three,

http://www.doi.org/10.62341/lbqf5637

where the random forest algorithm could attain a, were used to get the results. The accuracy equation presented in Chapter Three achieved a classification accuracy of 98. 07% on the pinned set. Only 10 samples were misclassified, while the remaining ones were correctly classified. In comparison to the k-means method, the algorithm misclassified 759 out of 1728 samples with an accuracy of over 56%. The thesis outperformed the k-means method, which was run on the same dataset, in terms of accuracy and results.

8. Conclusion

The performance of the random forest algorithm in evaluating the car's quality was examined in this study, and then the model's performance was compared to that of other models. The recommended method was assessed using the correct scale.

The findings of the random forest method were compared to those of the k-means algorithm, and the results of the random forest approach were analyzed and compared. The forest algorithm performed significantly better than the k-means algorithm using the same dataset. With an accuracy of 98. 07%, the random forest technique performed well, but with an accuracy of just 56%, the k-means algorithm performed poorly.

The value of features is determined by the random forest method, which is really helpful.

Lastly, when evaluating vehicle quality, the suggested score was (42. 07%) higher than the k-means method.

References

[1] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, “An ensemble random forest algorithm for insurance big data analysis,” Ieee access, vol. 5, pp. 16568–16575, 2017.

[2] S. K. Sudhanshu , P. B. Negi, S. Maurya, P. Juneja, and A. Rana, “Vehicle Quality Prediction by k-means Algorithm, India during Lockdown Period of Covid-19 Pandemic,” in 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, 2021, pp. 1254–1259.

[3] P. H. Putra, A. Azanuddin, B. Purba, and Y. A. Dalimunthe, “Random forest and decision tree algorithms for car price prediction,” J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. 1, vol. 3, no. 2, pp. 81–89, 2023.

http://www.doi.org/10.62341/lbqf5637

[4] J. Yang, S. Han, and Y. Chen, “Prediction of Traffic
[5] K. Liu, X. Hu, H. Zhou, L. Tong, W. D. Widanage, and J. Marco, “Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification,” IEEE/ASME Trans. Mechatronics Accident Severity Based on Random Forest,” J. Adv. Transp., vol. 2023, 2023., vol. 26, no. 6, pp. 2944–2955, 2021.
[6] “Car Evaluation - UCI Machine Learning Repository.” Accessed: Jan. 06, 2024. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Car+Evaluation
[7] Y. Lu et al., “The application of improved random forest algorithm on the prediction of electric vehicle charging load,” Energies, vol. 11, no. 11, p. 3207, 2018.
[8] S. K. Sudhanshu , P. B. Negi, S. Maurya, P. Juneja, and A. Rana, “Vehicle Quality Prediction by k-means Algorithm, India during Lockdown Period of Covid-19 Pandemic,” in 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, 2021, pp. 1254–1259.
[9] M. Collard, “Price Prediction for Used Cars: A Comparison of Machine Learning Regression Models.” 2022.
[10] P. H. Putra, A. Azanuddin, B. Purba, and Y. A. Dalimunthe, “Random forest and decision tree algorithms for car price prediction,” J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. 1, vol. 3, no. 2, pp. 81–89, 2023.
[11] J. Yang, S. Han, and Y. Chen, “Prediction of Traffic Accident Severity Based on Random Forest,” J. Adv. Transp., vol. 2023, 2023.
[12] U. L. Yuhana, V. R. Oktavia, C. Fatichah, and A. Purwarianti, “Automatic Assessment of Answers to Mathematics Stories Question Based on Tree Matching and Random Forest.,” Int. J. Intell. Eng. Syst., vol. 15, no. 2, 2022.
[13] H. AlSagri and M. Ykhlef, “Quantifying feature importance for detecting depression using random forest,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 5, 2020.
[14] R. Garreta, G. Moncecchi, T. Hauck, and G. Hackeling, Scikit-learn: machine learning simplified: implement scikit-learn into every step of the data science pipeline. Packt Publishing Ltd, 2017.