



Article

Estimation of Real-World Fuel Consumption Rate of Light-Duty Vehicles Based on the Records Reported by Vehicle Owners

Isabella Yunfei Zeng¹, Shiqi Tan², Jianliang Xiong^{3,*} , Xuesong Ding⁴, Yawen Li⁴ and Tian Wu⁵ ¹ UK-China (Guangdong) CCUS Centre, Guangzhou 510663, China; isabellazeng04@gmail.com² Department of Automation, Tsinghua University, Beijing 100084, China; tsq19@mails.tsinghua.edu.cn³ School of Economics and Management, Tsinghua University, Beijing 100084, China⁴ School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China; dingxuesong2000@bupt.edu.cn (X.D.); warmly0716@bupt.edu.cn (Y.L.)⁵ Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; wutian@amss.ac.cn

* Correspondence: xjl19@mails.tsinghua.edu.cn

Abstract: Private vehicle travel is the most basic mode of transportation, so that an effective way to control the real-world fuel consumption rate of light-duty vehicles plays a vital role in promoting sustainable economic growth as well as achieving a green low-carbon society. Therefore, the factors impacting individual carbon emissions must be elucidated. This study builds five different models to estimate the real-world fuel consumption rate of light-duty vehicles in China. The results reveal that the light gradient boosting machine (LightGBM) model performs better than the linear regression, naïve Bayes regression, neural network regression, and decision tree regression models, with a mean absolute error of 0.911 L/100 km, a mean absolute percentage error of 10.4%, a mean square error of 1.536, and an R-squared (R^2) value of 0.642. This study also assesses a large pool of potential factors affecting real-world fuel consumption, from which the three most important factors are extracted, namely, reference fuel-consumption-rate value, engine power, and light-duty vehicle brand. Furthermore, a comparative analysis reveals that the vehicle factors with the greatest impact are the vehicle brand, engine power, and engine displacement. The average air pressure, average temperature, and sunshine time are the three most important climate factors.

Keywords: real-world fuel consumption rate; machine learning; big data; light-duty vehicle; China



Citation: Zeng, I.Y.; Tan, S.; Xiong, J.; Ding, X.; Li, Y.; Wu, T. Estimation of Real-World Fuel Consumption Rate of Light-Duty Vehicles Based on the Records Reported by Vehicle Owners. *Energies* **2021**, *14*, 7915. <https://doi.org/10.3390/en14237915>

Academic Editors: Wei Pan and Thanikanti Sudhakar Babu

Received: 12 October 2021
Accepted: 23 November 2021
Published: 25 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tightening the control of oil consumption has always been among the urgent focuses of building a greener city. In the context of climate worsening and China's commitment to achieve carbon peak in 2030 and carbon neutrality in 2060, regulations on fuel consumption are increasingly becoming stricter. Recently, a new round of investigation into fine particle sources in Beijing was officially released. The results revealed that coal combustion is no longer the main source of PM_{2.5} in Beijing, and mobile sources such as vehicles have become the primary source of inhalable pollutants. To date, China has implemented a series of measures to control the fuel consumption rate of vehicles. In September 2019, the Ministry of Industry and Information Technology (MIIT) of the People's Republic of China and other relevant ministries issued the "Decision on Amending the Measures for the Parallel Management of Average Fuel Consumption of Automobile Enterprises and New Energy Vehicle Score". The objective of introducing the automobile enterprise fuel consumption score is to promote the sustainable development of China's new energy vehicle industry, accelerate the transformation of the energy structure, upgrade the traditional gasoline vehicle industry, and achieve a set of other goals in accordance with carbon neutrality. To improve the performance and accuracy of the fuel consumption score, which aims at reducing fuel consumption, the most effective method is to expand the production of purely electric and plug-in hybrid electric vehicles.

Currently, the fuel consumption score is calculated from the fuel consumption database provided by the MIIT, and can be roughly divided into the following steps. The first step is to calculate the average fuel consumption of each automobile enterprise according to the national standard (GB27999-2014). The calculation is based on the weighted average of the output of each vehicle and the fuel consumption value specified in the standard. The fuel consumption of each vehicle is closely related to the vehicle's curb weight, which varies significantly for different vehicles. Therefore, the required fuel consumption standard is not unified. The second step is to calculate the fuel consumption reported by each automobile enterprise for the corresponding vehicle types, according to the MIIT. The third step is to calculate the difference between the 2018 standard and the 2018 actual fuel consumption (the fuel consumption reported by the MIIT) multiplied by the output, which is the actual fuel consumption score for the specific vehicle enterprise.

The "Limits and Measurement Methods for Emissions from Light-Duty Vehicles (CHINA 6)" guidelines, which are issued jointly by the Ministry of Ecology and Environment and the General Administration of Quality Supervision, Inspection and Quarantine, require that all sold and registered light vehicles shall satisfy the standards, starting from 1 July 2020. According to the "Energy Conservation and New Energy Automobile Industry Development Plan (2012–2020)", the average fuel consumption rate of passenger vehicles in China should be reduced to 5.0 L/100 km by 2020. The MIIT has promulgated the "Measures for the Parallel Administration of the Average Fuel Consumption and New Energy Vehicle Credits of Passenger Vehicle Enterprises", which was implemented on 1 April 2018. The promulgation and implementation of these policies imposes stricter requirements for energy saving and emissions reduction technology in the automobile industry. To solve the current energy and environmental problems and achieve carbon neutrality in the near future, it is of great significance to estimate the real-world fuel consumption of light-duty vehicles and to identify its impact factors.

At present, the most direct approach to determining the fuel consumption rate of a vehicle is to check the reference fuel consumption information provided by the MIIT, which may be far different from the actual case. Since the implementation of a vehicle emissions test standard, China has adopted the New European Driving Cycle (NEDC) working conditions to test fuel consumption and emissions. However, some problems have arisen after years of practice. The test results under the NEDC working conditions are quite different to the real-world driving situation in China, and the gap between them in terms of fuel consumption is approximately 26%; roughly consistent with the calculation result of 30% found by Liu et al. [1]. However, this value shows a considerable difference in different vehicle models. Specifically, the gap between the NEDC and the real fuel consumption of Geely Auto's 2016 Boyue series (1.8T, 184 horsepower, L4) is about 37%, while that of MG's 2013 MG3 series (1.5 L, 109 horsepower, L4) is about 4.3%. This discrepancy not only interferes with judgement in terms of understanding the actual driving state, but also does harm to the government's credibility from the perspective of vehicle drivers.

The problem of the NEDC working condition has three main aspects. First, the NEDC working condition is very different to the driving characteristics of automobile vehicles in China. This difference is particularly evident in the emissions performance, fuel consumption, and optimized calibration value based on the NEDC. Second, this divergence directly affects the implementation of China's energy conservation and emissions reduction policies, which has a negative impact on the government's reputation to some extent. Third, the existing NEDC working condition method underestimates the energy saving effect of new energy vehicles.

In fact, the NEDC condition is too ideal from three perspectives. First, there is a large gap between the laboratory simulation conditions and the actual road conditions. Specifically, China has a vast territory and the road conditions in different regions vary greatly, and this fact is neglected by the NEDC conditions. Second, the NEDC working condition test ignores the influence of external factors such as air pressure and temperature, which may influence the fuel consumption to a certain degree. Third, the NEDC working

condition test does not consider the actual behaviors of drivers, such as their driving habits and the use of air conditioners.

Hence, the China Automotive Test Cycle (CATC) was launched in 2015. Compared with the NEDC (European Fuel Consumption and Emissions Assessment Standard) adopted in the fifth phase of emission regulations (CHINA 5), and the Worldwide Harmonized Light-Duty Vehicle Test Cycle (WLTC) working conditions adopted by the CHINA 6 standard, CATC's working conditions more realistically reflect the actual conditions of China's roads. The successful introduction of this project enabled an independent basic standard system to be established for the Chinese auto industry.

Three different types of data are collected with regard to China's working conditions. First, the collection of real-time and synchronized large-scale driving data for different vehicles in different regions is realized using CAN+GPRS technology. Second, geographic information system all-road low-frequency dynamic big data are used to calculate the actual turnover of the vehicle and its coefficient at different speed intervals, reflecting the macroscopic distribution of the vehicle more accurately and objectively. Third, the driving behavior characteristics, air-conditioner usage characteristics, and other vehicle-level characteristics are also considered.

The data used in this study were obtained from the BearOil app (www.xiaoxiongyouhao.com (accessed on 3 February 2021)), which has already been downloaded six million times with more than 800 thousand active monthly users. The accumulated mileage of active vehicles in 31 different provincial regions of China has exceeded 23 billion kilometers, and the real-world fuel consumption rate records have exceeded 51 million. Moreover, this study also takes into consideration vehicle factors such as vehicle brand, engine power, and engine displacement, as well as climate and environmental factors such as average air pressure, average temperature, and sunlight hours. Therefore, this study aimed to discover the most important factors that impact on the real-world fuel consumption rate of vehicles.

The rest of the paper is organized as follows. First, the related literature is reviewed in Section 2. The data source, the extracted real-world fuel consumption rate, and the climate factors are discussed in Section 3. Section 4 describes the experiments, including the model selection and model training. In addition, Section 4 also reports the results, including the comparison of different models and the assessment of the most important features. Section 5 discusses the assessment of feature importance. Section 6 presents the conclusions and the implications with regard to policy.

2. Literature Review

Considering the large proportion of environmental pollution that can be ascribed to automobile sources [2,3], as well as the constraints on fossil fuel production, it is important to obtain relatively accurate fuel consumption information to adjust energy allocation appropriately. Furthermore, the application of artificial intelligence in the field of business intelligence has risen gradually [4]. Therefore, models for estimating the real-world fuel consumption rate and assessments of impact factors are being proposed at an increasing pace.

In the field of vehicle fuel consumption, extensive research has been carried out around the world, and this can be roughly divided into three aspects.

Firstly, models that characterize the actual fuel consumption status have been proposed in some studies. Among the existing models which have been proposed to estimate fuel consumption rate, the HDM-4 fuel consumption model is one of the most widely utilized. Many studies have used this model and then carried out calibration, which is a necessary step in this methodology [5,6]. The accuracy of the HDM-4 fuel consumption model and the need for further calibration were discussed in [7]. This study was based on a limited set of tests, wherein only a small number of vehicles were tested at constant speed on selected sections under limited weather conditions. In addition, Li et al. [8] used a multilayer perceptron (MLP) method and considered parameters including external environmental factors, the manipulation of vehicle companies, and the driving habits of drivers. It was found that the multilayer perceptron method could classify their nonlinear dataset

in the most reasonable way under sensitivity analysis. Some studies used a two-level clustering model to determine the driving patterns of electric vehicles. However, this model only focused on simple, static vehicle parking patterns and did not consider other traffic information or weather conditions [9]. In a similar study, Wu et al. [10] predicted the fuel consumption rate by learning from the real-world data of vehicle owners. Via a Pearson coefficient correlation analysis with the help of data mining, Yamashita et al. [11] selected the driving behavior indicators including speed, acceleration, and left/right/U-turns that were proved to be highly correlated with fuel consumption. Through neural network modeling and regression analysis, these indicators generated more than 12 aggregation models, and the best mean absolute percentage error value among them was below 5%. These classifications of driving behavior and the mean absolute percentage errors of the proposed models provide a certain reference for the assessment of driving behavior. Ahn et al. [12] used a microscopic fuel consumption and emissions model to predict the fuel consumption of normal light-duty vehicles based on the instantaneous vehicle speed and acceleration levels, and Lei et al. [13] introduced compound acceleration variables to this model to capture the effects of the interaction between the historical acceleration and the current speed on emissions and fuel consumption, producing reasonable estimates compared with the actual measurements. In some studies, instantaneous Global Positioning System (GPS) speed measurements enabled models to be applied to estimate fuel consumption and emissions directly [14].

Secondly, some countries have developed certain vehicle fuel consumption measurement tools in practice, typically including the Vehicle Energy Consumption Calculation Tool (VECTO) and the Motor Vehicle Emission Simulator (MOVES). VECTO is software developed by the European Commission. When vehicles enter the market, VECTO helps to estimate their fuel consumption and carbon emissions. The main vehicle properties such as mass, air drag, tire rolling resistance, axle and gearbox torque loss maps (torque loss as a function of input torque and speed), and engine maps (maximum torque, motoring torque, and fuel consumption) are the inputs of this software. In this system, the instantaneous engine power depends on three factors: the power demand at the wheels, the power demand of the auxiliaries, and the efficiency of each component in the powertrain. The fuel consumption is measured through interpolation in the fuel consumption map, together with the instantaneous engine torque and speed. Stijn et al. [15] proved that the fuel consumption was predicted with an error of less than 1.5% for individual trips by VECTO, and less than 0.5% when averaged over various repetitions. MOVES is a new generation emission model developed by the US EPA since 2001, and MOVES3 has now been released. The model uses the vehicle specific power (VSP) variable independently of vehicle weight in the calculation of power demand [16,17], and uses cluster analysis to characterize the relationship between VSP and fuel consumption. The vehicle power demand formula used in MOVES takes vehicle speed, acceleration, and gradient as independent variables [18]. In these fuel consumption prediction models, the estimation of automobile engine power is an important part of the model. Due to the basis of traditional dynamics, there are physical errors in such estimates. It is an urgent task of current research to design fuel consumption prediction models based on actual conditions rather than physical formulas, in order to improve the accuracy of predictions.

Thirdly, existing studies have shown that real-world fuel consumption rate is influenced by various factors. The objective characteristics of the road and vehicle, such as the road surface [19,20], road width [21,22], traffic congestion and speed limits [23,24], energy management strategy [25–27], and fuel-tank status monitoring technology [28], play significant roles. Ejsmont [20] handled the above-mentioned factors by investigating the relationship between the surface texture and the rolling resistance of light and heavy vehicle types. He used the mean profile depth as a proxy parameter for the road surface, and the results revealed that, although a correlation exists, it cannot be explained in absolute terms because the regression between the mean profile depth and the rolling resistance is not linear. Kono [21] considered many factors, including traffic information, geographic

information, vehicle parameters, and driver behavior, to analyze and predict fuel consumption for ecological routes. Comparing the results to those obtained by the traditional time-priority route search method and a driving experiment, the author concluded that it is important to propose an indicator of fuel reduction effectiveness for future emissions reduction technologies, including ecological route searches. Brundell-Freij [23] reported that the speed, the acceleration, and the type of gears influenced fuel consumption. His study results revealed that the influence of the street and traffic environment on the driving behavior was dependent on driver variables and vehicle performance. Furthermore, subjective characteristics such as driving velocity [23,28] and driving acceleration [29,30] also affect real-world fuel consumption rates and are used to describe the temporal characteristics of driving patterns. Xu, Chen, and Li [28] reported that speed has a remarkable effect on fuel consumption, particularly when vehicles travel on urban roads where there are many traffic signals. Hence, to reduce fuel consumption, the authors proposed a double-layer speed optimization method with real-time computation, and obtained the optimal real-time speed, demonstrating the potential of the double-layer speed optimization method for improving fuel consumption and reducing travel time. By tracking the driving speed of cars in 11 cities in China, Wang [29] inferred that the infrastructure of roads and the sizes of cities are vital factors affecting the heterogeneity of driving behavior. Another study used a vehicle-specific fuel consumption model based on a PEMS application to estimate fuel consumption under different driving patterns. The vehicle fuel consumption per unit time exhibited a strong positive correlation with the cruise speed. When the vehicle accelerated, the fuel consumption rate significantly increased, but changed only slightly when the vehicle decelerated [31].

Real-world fuel consumption rate is affected by climate. For example, winter has been related to a decrease of 20% in fuel efficiency [25]. Other studies have established the relationship between temperature and driving environment [26,27]. Zahabi [25] investigated fuel efficiency, and then compared vehicle performance to that of a standard gasoline vehicle in a cold Canadian urban environment. He considered many different factors including the driving conditions, temperature, and speed. In his results, a low temperature below 0 °C in winter was identified as a factor exerting a detrimental influence on fuel consumption. Specifically, it was found that fuel efficiency decreased by 20% in winter compared with that in summer. In the present study, the climate environment is also considered an important factor, and the temperature factor is discussed in detail. Weilenmann, Favez, and Alvarez [26] proposed that cold starting, which refers to the internal temperature of vehicles, can reduce the emission of modern gasoline and diesel passenger cars. Alvarez and Weilenmann [27] proposed that low ambient temperatures affect hybrid electric vehicles in terms of fuel consumption and investigated these characteristics in five in-use hybrid electric vehicle models.

However, a small dataset and number of vehicle models means that it is not clear whether the estimations from the above models reflect the actual fuel consumption under realistic driving conditions. Modeling formulas based on traditional physics may have large errors with respect to reality. Even though various indicators, especially those of speed, were considered and inputted to the models, their weights were still not assessed, and this greatly limits the practical application of the models. In addition, climate factors have been proved to be highly related to fuel consumption rate, but they are still not introduced into forecasting models. To expand the applicability, research should consider multiple factors comprehensively.

To fill the research gap, we collected big data on many vehicle models, their actual fuel consumption, and the local climate conditions to describe the driving reality as much as possible. In addition, we comprehensively assessed the impact of vehicle performance parameters and local climate parameters on fuel efficiency. Our research proposed five models, namely, linear regression, naïve Bayes regression, neural network regression, decision tree regression, and LightGBM models, to estimate the real-world fuel consumption rate of light-duty vehicles in China. After being trained on large amounts of data, the

deep learning method greatly optimized the prediction of vehicle engine performance, thereby helping to improve the accuracy of the prediction. The results obtained by these five models were compared to determine the optimal model. Additionally, this study assessed 18 different factors and ranked the importance of all the factors.

3. Materials and Methods

3.1. Data

The data used in this study were obtained from two sources: the real-world fuel consumption rate records reported by vehicle owners in the BearOil app and the monthly dataset of the surface climate and road grade in some regions of China.

The real-world fuel consumption rate data are generated as follows. Users record information on the refuel time, the total mileage of the vehicle, liters of fuel, the oil price, and the gas payment via the BearOil app (iOS/Android versions) and the applets (WeChat applet, Amap applet, Alipay applet) each time they refuel. After that, they mark whether the tank is full or not and then save the record. Once a user fills up the tank, the liters of fuel the user adds to fill up the tank next time represents the fuel consumption during the trips between refueling. The average fuel consumption is a weighted average based on the fuel consumption of a single trip, with the weight being the mileage of the trip. Each time a user saves a record, it is automatically uploaded to the BearOil cloud server (after informing the user and obtaining consent and authorization). We filtered the records of different users of the same car model on the cloud server to exclude samples with obvious recording errors, and then filtered out the samples with large deviations by considering the distributions of the average fuel consumption of different users of the same car model, taking the remaining samples as valid samples. Then, we took the arithmetic means of the average fuel consumptions of all valid samples for each car model. Finally, we matched the data with the corresponding external environmental data, including temperature, humidity, pressure, wind speed, precipitation, sunshine, etc., based on the user's location and time information.

3.1.1. Fuel Consumption Rate Information

In this study, about 2 million records of real-world fuel consumption rates reported by vehicle owners in 17 provincial capitals of China in the period 2013–2017 were extracted from the BearOil app. Examples of the real-world fuel consumption rate data are shown in Table 1. To protect user privacy, the user number (User_ID) only shows the last eight digits of the true value.

Table 1. Raw data example of real-world fuel consumption rate information from BearOil app.

Feature Name	Instance 0	Instance 1	Instance 2	Instance 3	...
User_ID	65961294	17424034	28206249	78105203	...
City	Hangzhou	Shanghai	Wuhan	Guangzhou	...
Date	28 June 2017	26 June 2013	24 July 2013	26 June 2019	...
Brand Name	BMW	ROEWE	SKODA	TOYOTA	...
Series Name	BMW X1	ROEWE 350	FABIA	LEVIN	...
Version Year	2016	2011	2011	2016	...
Engine	1.5	1.5	1.4	1.8	...
Gearbox	L/136ps/L3	L/109ps/L4	L/86ps/L4	L/99ps/L4	...
refConsumption (L/100 km)	AMT-6	MT-5	MT-5	E-CVT	...
realConsumption (L/100 km)	6.1	7.8	6.5	4.2	...
	11.8	9.6	7.0	4.6	...

The User_ID in the sample is the unique ID of a BearOil app user. Therefore, the same User_ID corresponds to several samples, and this was used to record the time-varying relationship between the user's real-world fuel consumption rate, including the reporting

time and the city in which the user lives, and the fuel consumption rate measured by the user.

The relevant information of the vehicle is given in the sample, including the vehicle brand, series, and version. Because the exact versions of different vehicles brands are quite different, only the version year of each example is shown here. Additionally, the sample features include information on the vehicle engine and transmission. The engine parameters include the displacement (unit: L), power (unit: ps), and cylinder number. The transmission parameters indicate the type of transmission, including manual transmission (MT), automatic transmission (AT), automated manual transmission (AMT), continuously variable transmission (CVT), direct shift gearbox (DSG), and so on.

Additionally, our dataset also includes a reference value for the fuel consumption rate (refConsumption) of the corresponding vehicle, which is provided by the MIIT of the People's Republic of China. The fuel consumption rate measurement method adopted by the MIIT refers to the second stage of the NEDC. However, various problems exist, such as incompatibility with the current vehicle power and the overall quality and large differences in the actual driving conditions. In addition, owing to the impact of different climate conditions, driving behaviors, and other factors, the reference fuel consumption rate is often a poor proxy for the real-world fuel consumption rate.

Table 2 presents the descriptive statistics and Figure 1 shows the sample distribution for engine displacement. As shown in Table 1 and Figure 1, most of the samples were vehicles with displacement coverages of 0.8 L–2.5 L. For the samples with displacement coverages of 0.8 L–1.6 L, the average real-world fuel consumption rate records reported by vehicle owners in the BearOil app (realConsumption) was 8.052, with a standard deviation of 1.476 and a 75th percentile of 8.913. For the samples with displacement coverages of 1.6 L–2.5 L, the average realConsumption was 9.489, with a standard deviation of 2.042 and a 75th percentile of 10.734.

Table 2. Descriptive statistics.

		Engine Displacement (L)				
		$x \leq 0.8$	$0.8 < x \leq 1.6$	$1.6 < x \leq 2.5$	$2.5 < x \leq 4$	$x > 4.0$
Standard deviation	refConsumption	0.755	0.757	1.173	1.308	4.479
	realConsumption	1.298	1.476	2.042	2.607	6.015
Min	refConsumption	1.5	1.2	2	2.9	2.4
	realConsumption	2.355	0.898	1.288	3.246	2.025
Max	refConsumption	7.8	9.8	12.5	15.7	17.3
	realConsumption	11.836	16.076	22.365	24.319	28.064
P25	refConsumption	5.200	5.900	6.800	9.900	13.200
	realConsumption	5.756	7.066	8.073	11.362	13.995
Median	refConsumption	5.600	6.400	7.600	10.700	14.400
	realConsumption	6.438	7.924	9.288	13.072	17.045
Mean	refConsumption	5.653	6.447	7.631	10.681	12.917
	realConsumption	6.584	8.052	9.489	13.120	15.479
P75	refConsumption	5.800	6.900	8.400	11.600	15.600
	realConsumption	7.317	8.913	10.734	14.742	18.468
Observation		2177	48,570	89,008	2179	71

Note: x denotes engine displacement.

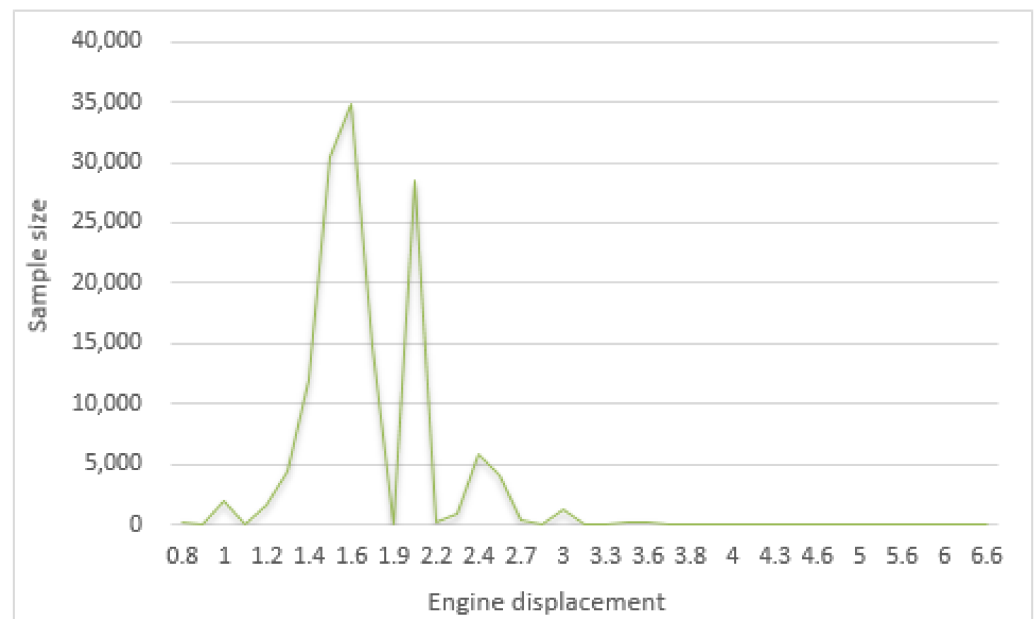


Figure 1. Sample distribution of engine displacement.

Moreover, because some information is often omitted by app users in the process of data uploading, there are many missing values in the original dataset. The corresponding processing methods are introduced in the data preprocessing section of this paper.

3.1.2. Environment Information

The climate information data were extracted from the monthly reports on surface meteorological observations provided by the meteorological departments of the provincial regions in China. In this study, climate data from 2013 to 2017 were collected, which is consistent with the spatial and time range of the fuel consumption rate data.

Each climate data item contains the station number of the observation area and the corresponding annual and monthly statistical information. The relevant climate characteristics of feature number, specific feature name, and units of measurement are listed in Table 3.

As can be seen, the climate information includes the temperature, barometric pressure, precipitation, sunlight, and other information. The climate information of different regions during the sample period also exhibits great variation, which has a non-negligible impact on the real-world fuel consumption rate of automobiles.

Because the climate of a certain region exhibits a relatively fixed pattern within a certain month, this study averaged the climate conditions within each month in different cities, as climate factors. For the wind direction, an Arabic number 1 was assigned to a north wind, and this number was increased by 1 for every 22.5 degrees clockwise. Additionally, if the wind speed was less than or equal to 0.2 m/s, conditions were considered to be calm, which corresponds to the Arabic number 17. Therefore, there are in total 17 wind direction categories successively numbered from 1 to 17.

Table 3. Factors of meteorological data and units of measurement.

Feature Number	Feature Name	Unit	Measurement Resolution
V10004	Average pressure	0.1 hPa	
V10201	Extreme maximum pressure	0.1 hPa	
V10202	Extreme minimum pressure	0.1 hPa	
V13004	Mean vapor pressure	0.1 hPa	
V12001	Average temperature	0.1 °C	
V12011	Extreme maximum temperature	0.1 °C	
V12012	Extreme minimum temperature	0.1 °C	
V12211	Mean maximum temperature	0.1 °C	
V12212	Mean minimum temperature	0.1 °C	
V12201	Average temperature anomaly	0.1 °C	
V13003	Mean relative humidity	1%	
V13007	Minimum relative humidity	1%	Monthly
V11002	Average wind speed	0.1 m/s	
V11042	Maximum wind speed	0.1 m/s	
V11041	Extreme maximum wind speed	0.1 m/s	
V11212	Maximum wind direction	azimuth	
V11043	Extreme maximum wind direction	azimuth	
V13011	Average precipitation	0.1 mm	
V13052	Maximum daily precipitation	0.1 mm	
V13212	Precipitation anomaly percentage	1%	
V13012	Daily precipitation ≥ 0.1 mm days	1 day	
V14033	Sunshine percentage	1%	
V14032	Sunshine time	0.1 h	
V15001	Road grade		

3.2. Factor Extraction

3.2.1. Factor Extraction of Fuel Consumption Rate Information

The fuel consumption rate information obtained from the BearOil app mainly contains the following information: the vehicle factors, reference fuel consumption rate, and real-world fuel consumption rate.

For a given user who drives the same car, there are always certain fluctuations in the fuel consumption rate reported each time, and this is attributed to different driving behaviors and driving environments at different times. To clarify the objective of our research, we aimed to predict the average real-world fuel consumption rate of specific vehicle types under specific climate conditions. Therefore, the real-world fuel consumption rate reported by a specific app user in different cities and months was averaged and treated as the prediction target.

There were significant differences among fuel consumption rates for different vehicle brands, engine parameters, and transmission parameters. Therefore, this study selected the above factors as the model input. The displacement and power characteristics of the engine parameters are continuous variables, while the other characteristics are discrete variables. Because the number of vehicle series belonging to different brands was too large in our data set, there would have been too many dimensions if we had employed one-hot encoding. Since a certain correlation exists between the proposed parameters and the exact vehicle series, the vehicle series was not used as an input.

Moreover, although many studies have reported that the reference value provided by the MIIT and the actual fuel consumption rate are quite different, the reported official data can still act as a reasonable reference for the real-world fuel consumption rate, and could also be used to reduce errors from abnormal fuel consumption values uploaded by the app users. This study, therefore, incorporates the reference consumption given by the MIIT, which is a continuous variable, as an input.

3.2.2. Factor Extraction of Climate Information

The available climate factors are listed in Table 3. We merged the fuel consumption information with the corresponding climate information in different cities and for different dates, and combined them to be used as input variables in our models. Additionally, to prevent multicollinearity originating from strong correlations between climate characteristics, it was necessary to test the correlation coefficients between these input variables. The climate variable pairs with correlation coefficients above 0.8 are listed in Table 4.

Table 4. Correlation table of climate factors.

Feature A	Feature B	Pearson Correlation
V12012	V12212	0.99296
V12012	V12211	0.96847
V12012	V12001	0.98527
V12012	V13007	0.70025
V12012	V13004	0.95931
V12012	V12011	0.93381
V14032	V14033	0.90305
V12212	V12211	0.975585
V12212	V12001	0.99426
V12212	V13004	0.95482
V12212	V12011	0.95523
V13212	V13012	0.99583
V10202	V10004	0.99927
V10202	V10201	0.99769
V10004	V10201	0.99944
V12211	V12001	0.99237
V12211	V13004	0.93057
V12211	V12011	0.97633
V11042	V11041	0.83803
V12001	V13004	0.94535
V12001	V12011	0.97182
V13007	V13003	0.86545
V13004	V12011	0.88903

As can be seen, there is a strong positive correlation between many climate-related variable pairs, and this required us to select a proper set of corresponding characteristics. For each variable pair with a strong correlation, only one characteristic was selected, and all the remaining environment factors that were chosen to be inputs in the estimation process are listed in Table 5.

Table 5. Selected environment factors.

Feature Number	Factor Name	Unit
V10004	Average pressure	0.1 hPa
V12001	Average temperature	0.1 °C
V12201	Average temperature anomaly	0.1 °C
V13003	Mean relative humidity	1%
V11002	Average wind speed	0.1 m/s
V11212	Maximum wind direction	azimuth
V11043	Extreme maximum wind direction	azimuth
V13011	Average precipitation	0.1 mm
V13012	Daily precipitation \geq 0.1 mm days	1 day
V14032	Sunshine time	0.1 h
V15001	Road grade	°

From the above analysis, it was found that there is a strong correlation between the average and extreme values of the climate-related variables, e.g., between the average temperature and average minimum temperature. For factor pairs with a strong correlation,

this study preferred to select the average value as the input. The main reason is that extreme values only represent the climate conditions over a short period, while the average value is more representative of the climate condition over a given period of time, namely, one month in our research.

3.3. Model Selection and Criteria

The objective of this study was to predict the real-world fuel consumption rate of vehicles according to the vehicle factors and climate conditions. The selection of the model's input factors was described in the above section. The proposed models and model selection procedure are introduced in this section.

The proposed regression models included linear regression, naïve Bayes regression, neural network regression, decision tree regression, and LightGBM models. There were three main criteria for model selection. Firstly, fuel consumption prediction was targeted for the continuous dependent variable, and therefore regression models were selected. Secondly, in addition to the accuracy of fuel consumption prediction, our study was also concerned with identifying the important factors affecting fuel consumption, which involves interpretability. That is, we needed to know whether the results of each model were easy to explain. In this case, linear regression and decision tree regression are suitable choices, since linear regression is good at obtaining linear relationships, while decision tree models are developed for obtaining nonlinear relationships in the data set. Thirdly, fuel consumption prediction has a strong practical application scenario, and the prediction speed of the fuel consumption is often of great concern in car owners' use. Therefore, we referred to the practices in studies by Sousa et al. [32], Pattekari and Parveen [33], Alsalmán et al. [34], and Chen et al. [35] and added LightGBM models, naive Bayes models, and neural network models, which have shown good performance in practice and have high prediction speed, like linear regression and decision tree regression.

The criteria for model selection included the mean absolute error ($MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|$), mean absolute percentage error ($MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y'_i|}{y_i}$), mean squared error ($MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2$), and R squared ($R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$).

In the above formula, y_i denotes the true value, y'_i denotes the predicted value, and \bar{y}_i denotes the mean actual value. Smaller MAE, MAPE, and MSE, and larger R^2 values mean that the error between the predicted and the actual value is smaller, indicating that the model fits well and performs better.

The selection criteria for the architecture of the models can be summarized as follows. (1) Neural network regression: the neural network consists of three fully connected layers. In the input layer, the number of neurons is set as 64 and the activation function is set as ReLU ($f(x) = \max(0, x)$), which is conventionally used in DNNs [36]. In the hidden layer, the number of neurons is set as 64 based on the geometric pyramid rule proposed by Masters [37] and the activation function is set as ReLU. The L2 regularizer and L1 regularizer are selected for the weighting matrix and the output matrix, respectively, and lambda is set as 0.05 to reduce overfitting [38]. In the third layer, the number of neurons is set as 1. RMSprop is selected as the optimizer [39]. The loss function is set as MSE and the metric function is set as MAE [40,41]. (2) Decision tree regression: the maximum tree depth is set as 4, the maximum number of leaf nodes is set as 200 and the minimum number of sample leaves is set as 2. Parameter selection is based on a genetic algorithm to obtain the optimal parameter settings that give maximum accuracy [42]. (3) LightGBM: the number of leaf nodes is set as 25, the learning rate is set as 0.01, and the number of iterations is set as 5000 for better accuracy [43].

4. Results

4.1. Model Training and Experiment Results

The raw data consisted of 2,424,379 records from 2002 to 2020 for 194,516 users. During data preprocessing, 53,852 records containing missing values and 106,223 records outside

the observation period were removed, leaving 2,264,304 records. After excluding the detected fuel consumption outliers, 1,453,299 data records remained. Finally, we took the average value of the fuel consumption for the same car models and obtained 142,005 items of data.

After removal of the missing values and outliers, and standardization from the original data, 70% of the data were randomly selected as the training dataset, and the remaining 30% of the data were used as the test dataset, according to common methods used for predicting engine performance based on machine learning models [44–46]. In this paper, the input data to the training set consisted of a randomly selected 70% sample. The models were reproducible using the same software, program, and data. However, the results were non-repeatable due to uncontrolled randomization which causes computational variation that cannot be removed for the learning libraries used, such as Keras (with TensorFlow), in this study. In fact, this unique type of uncontrolled randomization is a significant and common challenge for machine learning methods [47].

In Table 6, the ‘refConsumption’ row represents the result from directly using the MIIT reference fuel consumption rate for model prediction. As can be seen, the error between the reference fuel consumption rate value and real-world fuel consumption rate value is quite large. The remaining rows represent the training and prediction errors of the four regression models, respectively.

Table 6. Results of regression model training and testing.

Model	Training Data				Testing Data			
	MAE	MAPE	MSE	R2	MAE	MAPE	MSE	R2
refConsumption	1.896	26.4%	5.351	−2.245	1.899	26.4%	5.386	−2.241
Linear regression	1.024	11.8%	1.840	0.575	1.028	11.8%	1.854	0.577
Naïve Bayes	1.024	11.8%	1.840	0.575	1.028	11.8%	1.854	0.577
Neural network	0.963	11.3%	1.630	0.624	0.985	11.5%	1.707	0.610
Decision tree	1.079	12.4%	2.013	0.536	1.086	12.5%	2.048	0.532
LightGBM	0.858	9.8%	1.343	0.690	0.914	10.4%	1.552	0.646

We referred to studies by Liu et al. [1] and Liu et al. [48] to present the average fuel consumption results for each model by displacement distribution, as shown in Table 7. The criteria for displacement distribution are based on the Chinese national standard GB3730.1-88. Clearly, LightGBM shows the best prediction, with the smallest difference from the actual fuel consumption.

Table 7. Average fuel consumption results for each model by displacement distribution.

Model	Engine Displacement (L)				
	$x \leq 1.0$ (1.530%)	$1.0 < x \leq 1.6$ (34.203%)	$1.6 < x \leq 2.5$ (62.680%)	$2.5 < x \leq 4.0$ (1.534%)	$x > 4.0$ (0.050%)
refConsumption	5.653 (−0.932)	6.447 (−1.605)	7.631 (−1.859)	10.681 (−2.439)	12.917 (−2.562)
Linear regression	7.014 (+0.884)	8.185 (+0.006)	10.171 (−0.040)	11.420 (+0.160)	17.586 (−0.656)
Naïve Bayes	7.0137 (+0.884)	8.185 (+0.006)	10.171 (−0.040)	11.420 (+0.160)	17.587 (−0.655)
Neural network	6.282 (+0.152)	8.118 (−0.061)	10.204 (−0.007)	11.194 (−0.066)	17.785 (−0.458)
Decision tree	7.573 (+1.443)	8.214 (+0.035)	10.125 (−0.085)	11.362 (+0.102)	16.680 (−1.563)
LightGBM	6.590 (−0.034)	8.056 (+0.007)	9.516 (−0.003)	13.161 (−0.029)	17.408 (−0.835)

Note: The values in brackets below the displacement range indicate the percentage for each displacement range, while the values in brackets below the model results are the differences between the average of the predicted values and the corresponding average of the actual fuel consumption.

4.2. Comparison and Analysis of Different Models

In this section, we compare the five different models according to our proposed criteria.

Figure 2 shows the mean absolute error (MAE) between the model prediction and the actual values of each model. As can be seen, the mean absolute error from the reference fuel consumption rate provided by the MIIT was 1.899 L/100 km, while the mean absolute error using our dataset and including vehicle factors and climate conditions was approximately 1 L/100 km. Among the models, the mean absolute error of the LightGBM model (0.914 L/100 km) was the lowest.

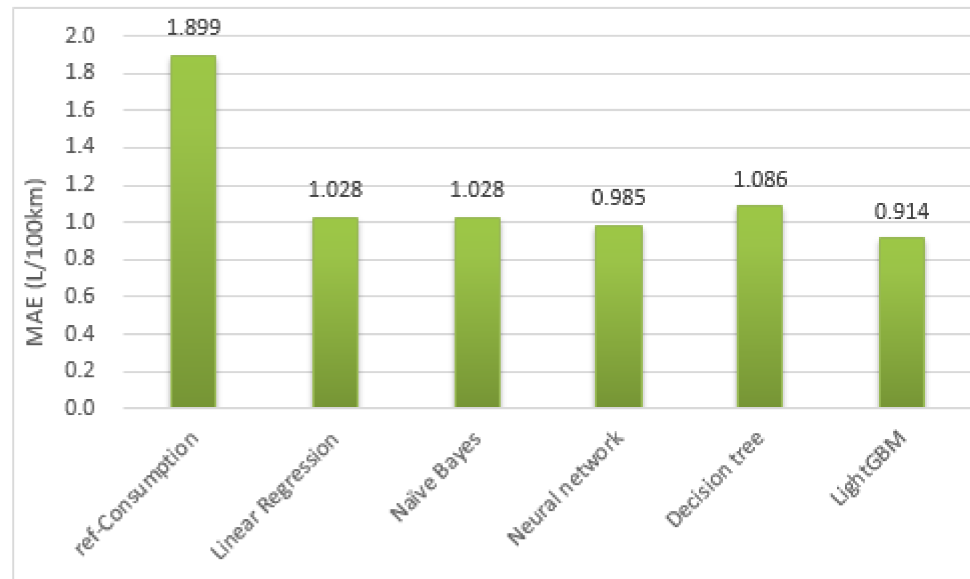


Figure 2. MAE of different models.

However, the MAE only indicates the absolute value of the deviation and may not sufficiently reveal the magnitude of the relative deviation from actual values. Therefore, Figure 3 shows the mean absolute percentage error (MAPE) between the model prediction and the actual values. The results reveal that the MAPE between the reference fuel consumption rate and the real-world fuel consumption rate was approximately 26.4%. The best prediction model was still LightGBM, with a corresponding MAPE of 10.4%, which is better by 16% compared with the reference rate. This demonstrates that our proposed prediction model could be applied practically in the prediction and revision of vehicle fuel consumption, thus enhancing the credibility of the fuel consumption score.

Additionally, the mean square errors (MSE) of the model prediction and actual values of different models are shown in Figure 4. The results reveal that the mean square error of the LightGBM model is the lowest. This verifies our proposition that LightGBM performs best among the five regression models.

R^2 is an index measuring the degree to which a set of independent variables explain a dependent variable in a regression model. From Figure 5, it can be seen that the reference fuel consumption rate given by MIIT explains the real-world fuel consumption rate to a very low degree, while the highest R^2 value is that of the LightGBM model, followed by that of the neural network model. The results indicate that the dependent variables in the LightGBM model explain 64.6% of the variation in the independent variable.

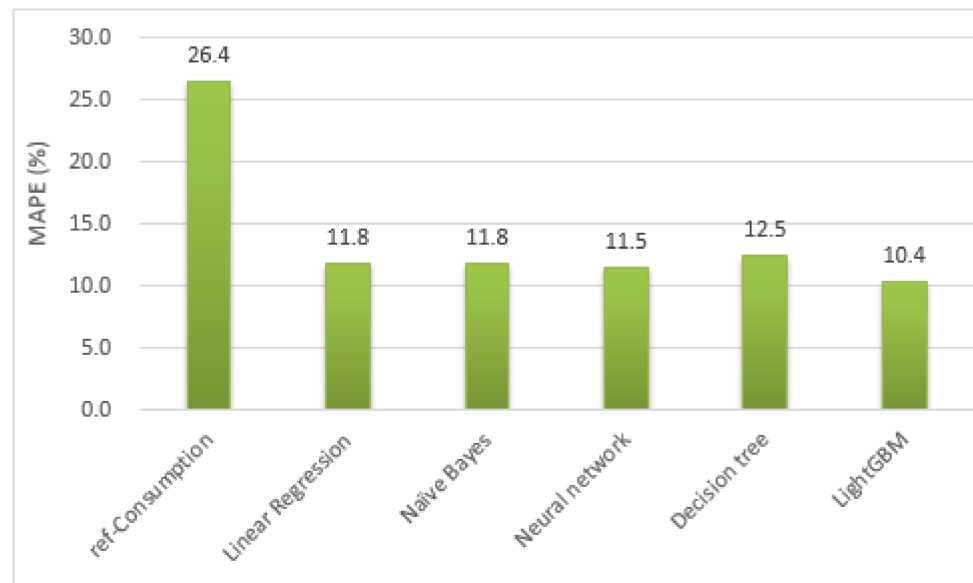


Figure 3. MAPE of different models.

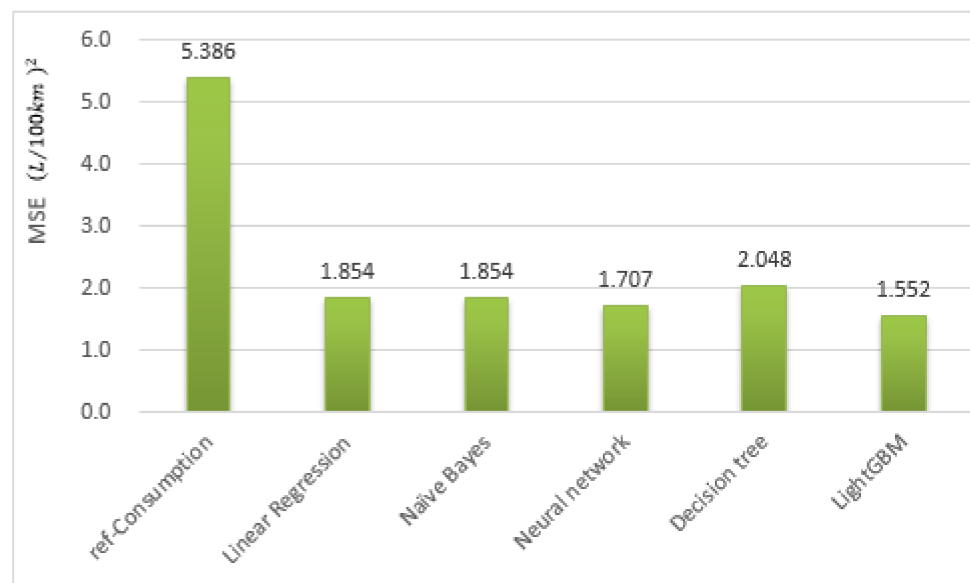


Figure 4. MSE of different models.

From the above results, it can be seen that the prediction errors of the real-world fuel consumption rate by the five regression models based on the vehicle and climate parameters were much lower than that of the reference fuel consumption rate value provided by the MIIT. The MAPE can be reduced by 16% by most of them, showing that the proposed prediction model has practical significance and can be used in real-world applications to produce much more precise estimates of the fuel consumption rate.

Moreover, the comparison between the results obtained by the different models revealed that the LightGBM regression model was the optimal model with the best performance in reducing the prediction error.

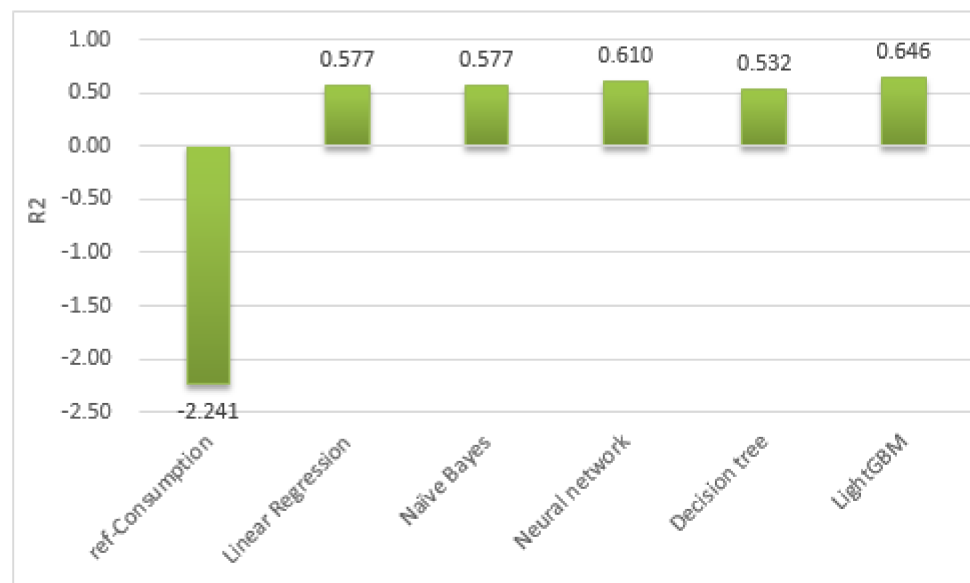


Figure 5. R² of different models.

5. Discussion

The above results reveal that the LightGBM model achieved the best performance. The estimated weights of the input parameters in the LightGBM model are shown in Figure 6.

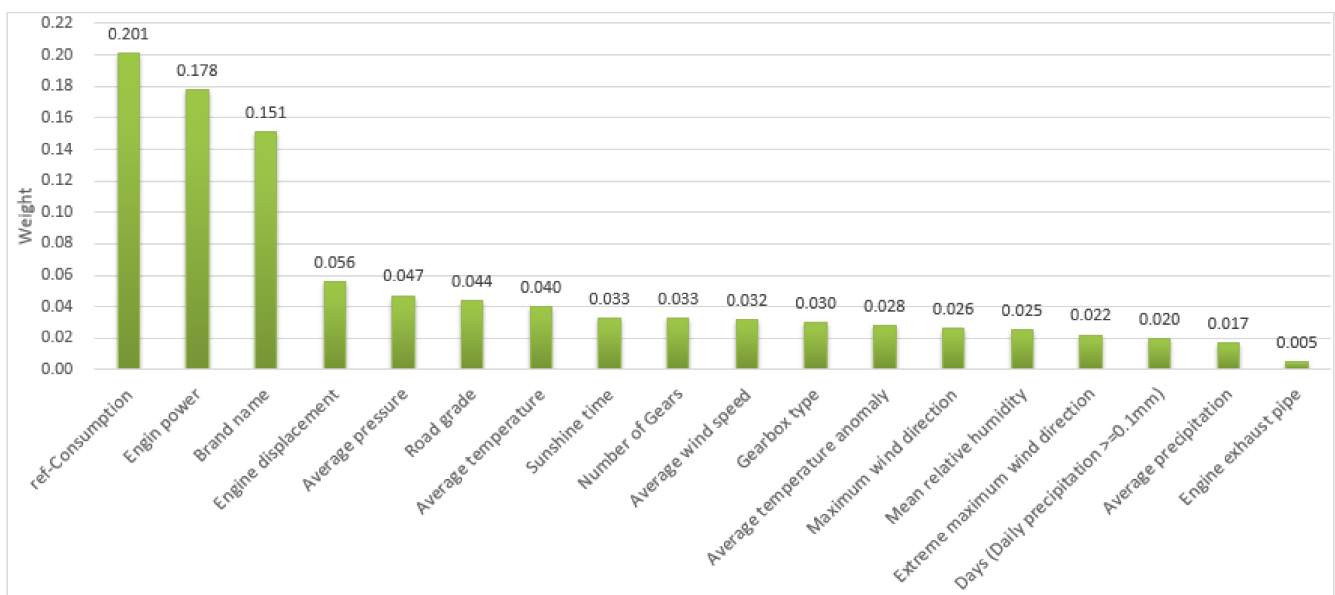


Figure 6. Weights of 18 factors in LightGBM model.

From the relative weights of the different factors, it can be seen that the reference fuel consumption rate is the most significant input characteristic, which is consistent with the reality that the MIIT reference value could act as a proxy for a large proportion of the actual fuel consumption.

The engine power and vehicle brand are input factors with weights exceeding 0.1, which indicates that vehicle parameters, as well as the exact brand, could impact the real-world fuel consumption to a non-negligible degree. The extra explanatory power brought by engine power and vehicle brand could be attributed to unrealistic testing conditions, which do not capture the different fuel consumption rates of different vehicles.

Additionally, the engine displacement is an input factor with a weight exceeding 0.05, which indicates the great effect of atmospheric pressure on the combustion efficiency of gasoline fuel.

Moreover, among the environment factors, the average pressure (V10004), road grade (V15001), average temperature (V12001), average wind speed (V11002), and sunshine time (V14032) also have a great impact on the real-world fuel consumption rate.

In summary, in this part of the study we carried out a comparative analysis of the vehicle and climate factors in our dataset and found that, in addition to the reference fuel consumption rate, the vehicle factors that had the greatest impact on the real-world fuel consumption rate were the vehicle brand, engine power, and engine displacement. The climate factors that had the greatest influence on real-world fuel consumption rate were the average air pressure, average temperature, and sunshine time.

According to the research in this paper, the parameters required for fuel consumption prediction are of two types: one is the vehicle parameter information and the other concerns the matching with the corresponding external environment data, according to the user's location and time information. These include temperature, humidity, pressure, wind speed, precipitation, sunshine, road grade, etc. Therefore, the models in this paper could be embedded into the BearOil app as a forecasting module. Users would only need to input vehicle parameters and location information, and the corresponding fuel consumption prediction results would be obtained. Furthermore, we extracted information from millions of users and their corresponding car models in our study. Therefore, each car model corresponded to different records from different users, even if it was extracted for the same date. In the future, we will try to track each user for a longer period of time and extract refueling data from the user for each month to construct a complete time series dataset. For this purpose, we could refer to the study of Bokde et al. [49] and apply the Monte_Carlo function of the ForecastTB package for better model comparison, since this requires the input of a time series dataset.

6. Conclusions

With the ongoing innovation and development in information technology, artificial intelligence (AI) will greatly accelerate technological progress in our increasingly digital and data-driven world. In this paper, we utilized five regression models, namely, linear regression, naïve Bayes regression, neural network regression, decision tree regression, and LightGBM models, to estimate the real-world fuel consumption rate of light-duty vehicles in China, based on a large sample of individual real-world driving and fuel consumption data.

The MAE, MAPE, MSE, and R^2 values between the real-world fuel consumption rate and the value predicted using the vehicle and climate factors were far better than when only referring to the fuel consumption rate provided by the MIIT of China. The comparison of the different models revealed that the LightGBM regression model performed best among the candidate models according to all our criteria (MAE = 0.914 L/100 km, MAPE = 10.4%, MSE = 1.552, and R^2 = 0.646).

This study also assessed 18 different factors and determined the priority ranking of each factor. From the relative weight of each factor in the LightGBM model, it can be seen that the three most important factors were the reference fuel consumption rate, engine power, and vehicle brand.

In our study, China's light-duty vehicles show higher fuel consumption rates compared to the NEDC, which provides a meaningful reference for other countries. Therefore, a proper light-duty vehicle test cycle should be adopted for the reality of the situation in China, while the role of the NEDC must be seriously evaluated.

Achieving cleaner transportation also matters greatly to China. In the traditional vehicles industry, manufacturers should emphasize the technological improvement of key components such as the engine. For instance, dual-fuel engines that are promising for balancing superior power with combustion efficiency can be focused upon. Based on the

assessments of the United States and the European Union [50,51], waste heat recovery (WHR) can effectively reuse produced heat as engine work, reducing fuel consumption. That is worth learning.

Additionally, the government should tighten the regulation of vehicle fuel consumption, prevent overconsumption from entering the market, and support heterogeneous vehicle models in different ways. It is wise to adopt policy tools such as taxes and credits to guide consumer preferences to certain vehicle brands that show strong environmental responsibility and fuel-saving potential, especially new energy vehicles.

Author Contributions: Conceptualization, I.Y.Z., J.X. and T.W.; methodology, I.Y.Z. and S.T.; formal analysis, I.Y.Z. and S.T.; resources, Y.L. and T.W.; data curation, T.W.; writing—original draft preparation, I.Y.Z.; writing—review and editing, X.D. and J.X.; funding acquisition, Y.L. and T.W. All authors have read and agreed to the published version of the manuscript.

Funding: The project was sponsored in part by the National Natural Science Foundation of China (71804181, 71774095, 61902037), in part by the Fundamental Research Funds for the Central Universities under Grant 500419804, and in part by the National Center for Mathematics and Interdisciplinary Sciences, CAS.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: For helpful comments and discussions, we thank Yushan Yang.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Liu, Y.; Zhou, H.; Xu, Y.; Kongjian, Q.; Yu, H. *Feasibility Study of Using WLTC for Fuel Consumption Certification of Chinese Light-Duty Vehicles*; SAE Technical Paper 2018-01-0654; SAE International: Warrendale, PA, USA, 2018.
- Peng, T.; Ou, X.; Yuan, Z.; Yan, X.; Zhang, X. Development and application of China provincial road transport energy demand and GHG emissions analysis model. *Appl. Energy* **2018**, *222*, 313–328.
- Liang, X.; Zhang, S.; Wu, Y.; Xing, J.; He, X.; Zhang, K.M.; Wang, S.; Hao, J. Air quality and health benefits from fleet electrification in China. *Nat. Sustain.* **2019**, *2*, 962–971. [[CrossRef](#)]
- Jiang, S.; Chen, H. Examining patterns of scientific knowledge diffusion based on knowledge cyber infrastructure: A multi-dimensional network approach. *Scientometrics* **2019**, *121*, 1599–1617. [[CrossRef](#)]
- Ko, K.H.; Moon, B.K.; Lee, T.W.; Lee, W.H.; Yoo, I.Y.; Lee, S.Y.; Han, D.S.; Jeong, S.H. An economic calibration method for fuel consumption model in HDM4. *Wirel. Pers. Commun.* **2016**, *89*, 959–975. [[CrossRef](#)]
- Jiao, X.; Bienvenu, M. Field measurement and calibration of HDM-4 fuel consumption model on interstate highway in Florida. *Int. J. Transp. Sci. Technol.* **2015**, *4*, 29–46. [[CrossRef](#)]
- Perrotta, F.; Parry, T.; Neves, L.C.; Buckland, T.; Benbow, E.; Mesgarpour, M. Verification of the HDM-4 fuel consumption model using a big data approach: A UK case study. *Transp. Res. D Transp. Environ.* **2019**, *67*, 109–118. [[CrossRef](#)]
- Li, Y.; Tang, G.; Du, J.; Zhou, N.; Zhao, Y.; Wu, T. Multilayer perceptron method to estimate real-world fuel consumption rate of light duty vehicles. *IEEE Access* **2019**, *7*, 63395–63402. [[CrossRef](#)]
- Li, X.; Zhang, Q.; Peng, Z.; Wang, A.; Wang, W. A data-driven two-level clustering model for driving pattern analysis of electric vehicles and a case study. *J. Clean Prod.* **2019**, *206*, 827–837. [[CrossRef](#)]
- Wu, T.; Han, X.; Zheng, M.M.; Ou, X.; Zhang, X. Impact factors of the real-world fuel consumption rate of light duty vehicles in China. *Energy* **2019**, *190*, 116388.
- Yamashita, R.J.; Yao, H.H.; Hung, S.W.; Hackman, A. Accessing and constructing driving data to develop fuel consumption forecast model. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *113*, 12217. [[CrossRef](#)]
- Ahn, K.; Rakha, H.; Trani, A.; Van Aerde, M. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *J. Transp. Eng.* **2002**, *128*, 182–190. [[CrossRef](#)]
- Lei, W.; Chen, H.; Lu, L. Microscopic emission and fuel consumption modeling for light-duty vehicles using portable emission measurement system data. *World Acad. Sci. Eng. Technol.* **2010**, *66*, 918–925.
- Rakha, H.; Medina, A.; Sin, H.; Dion, F.; Van Aerde, M.; Jenq, J. Traffic signal coordination across jurisdictional boundaries: Field evaluation of efficiency, energy, environmental, and safety impacts. *Transp. Res. Rec.* **2000**, *1727*, 42–51. [[CrossRef](#)]
- Broekaert, S.; Grigoratos, T.; Savvidis, D.; Fontaras, G. Assessment of waste heat recovery for heavy-duty vehicles during on-road operation. *Appl. Thermal Eng.* **2021**, *191*, 116891. [[CrossRef](#)]
- Jiménez, J.L.; McLintock, P.M.; McRae, G.J.; Nelson, D.D.; Zahniser, M.S. Vehicle specific power: A useful parameter for remote sensing and emissions studies. In Proceedings of the Ninth CRC on-Road Vehicle Emissions Workshop, San Diego, CA, USA, 21 April 1999.

17. Frey, H.C.; Roupail, N.M.; Zhai, H. Speed-and facility-specific emission estimates for on-road light-duty vehicles on the basis of real-world speed profiles. *Transp. Res. Rec.* **2006**, *1987*, 128–137. [CrossRef]
18. U.S. Environmental Protection Agency (EPA). Overview of EPA’s MOtor Vehicle Emission Simulator (MOVES3). 2021. Available online: <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1011KV2.pdf> (accessed on 5 November 2021).
19. Ma, H.; Xie, H.; Huang, D.; Xiong, S. Effects of driving style on the fuel consumption of city buses under different road conditions and vehicle masses. *Transp. Res. D Transp. Environ.* **2015**, *41*, 205–216. [CrossRef]
20. Ejsmont, J.A.; Ronowski, G.; Świeczko-Żurek, B.; Sommer, S. Road texture influence on tyre rolling resistance. *Road Mater. Pavement Des.* **2017**, *18*, 181–198. [CrossRef]
21. Kono, T.; Fushiki, T.; Asada, K.; Nakano, K. Fuel consumption analysis and prediction model for “eco” route search. In Proceedings of the 15th World Congress on Intelligent Transport Systems and ITS America’s 2008 Annual Meeting ITS America ERTICOITS Japan Trans Core, New York, NY, USA, 16–20 November 2008.
22. Hu, J.; Wu, Y.; Wang, Z.; Li, Z.; Zhou, Y.; Wang, H.; Bao, X.; Hao, J. Real-world fuel efficiency and exhaust emissions of light-duty diesel vehicles and their correlation with road conditions. *J. Environ. Sci.* **2012**, *24*, 865–874. [CrossRef]
23. Brundell-Freij, K.; Ericsson, E. Influence of street characteristics, driver category and car performance on urban driving patterns. *Transp. Res. D Transp. Environ.* **2005**, *10*, 213–229. [CrossRef]
24. Wang, S.; Zhang, X.; Cao, J.; He, L.; Stenneth, L.; Yu, P.S.; Li, Z.; Huang, Z. Computing urban traffic congestions by incorporating sparse GPS probe data and social media data. *ACM Trans. Inf. Syst.* **2017**, *35*, 30. [CrossRef]
25. Zahabi, S.A.H.; Miranda-Moreno, L.; Barla, P.; Vincent, B. Fuel economy of hybrid-electric versus conventional gasoline vehicles in real-world conditions: A case study of cold cities in Quebec, Canada. *Transp. Res. D Transp. Environ.* **2014**, *32*, 184–192. [CrossRef]
26. Weilenmann, M.; Favez, J.Y.; Alvarez, R. Cold-start emissions of modern passenger cars at different low ambient temperatures and their evolution over vehicle legislation categories. *Atmos. Environ.* **2009**, *43*, 2419–2429. [CrossRef]
27. Alvarez, R.; Weilenmann, M. Effect of low ambient temperature on fuel consumption and pollutant and CO₂ emissions of hybrid electric vehicles in real-world conditions. *Fuel* **2012**, *97*, 119–124. [CrossRef]
28. Xu, B.; Chen, X.; Li, K.; Hu, M.; Bian, Y.; Yu, Q.; Wang, J. Double-layer speed optimization for reducing fuel consumption with vehicle-to-infrastructure communication. *J. Intell. Transp. Syst.* **2019**, *23*, 513–524. [CrossRef]
29. Wang, Q.; Huo, H.; He, K.; Yao, Z.; Zhang, Q. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. *Transp. Res. D Transp. Environ.* **2008**, *13*, 289–297. [CrossRef]
30. Gong, Q.; Midlam-Mohler, S.; Marano, V.; Rizzoni, G. An iterative markov chain approach for generating vehicle driving cycles. *SAE Int. J. Engines* **2011**, *4*, 1035–1045. [CrossRef]
31. Wang, H.; Fu, L.; Zhou, Y.; Li, H. Modelling of the fuel consumption for passenger cars regarding driving characteristics. *Transp. Res. D Transp. Environ.* **2008**, *13*, 479–482. [CrossRef]
32. Sousa, S.I.V.; Martins, F.G.; Alvim-Ferraz, M.C.M.; Pereira, M.C. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ. Model. Softw.* **2007**, *22*, 97–103. [CrossRef]
33. Pattekari, S.A.; Parveen, A. Prediction system for heart disease using Naïve Bayes. *Int. J. Adv. Comput. Math. Sci.* **2012**, *3*, 290–294.
34. Alsaman, Y.S.; Halemah, N.K.A.; AlNagi, E.S.; Salameh, W. Using decision tree and artificial neural network to predict students academic performance. In Proceedings of the 2019 IEEE 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 11–13 June 2019; pp. 104–109.
35. Chen, C.; Zhang, Q.; Ma, Q.; Yu, B. LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemometr. Intell. Lab. Syst.* **2019**, *191*, 54–64. [CrossRef]
36. Agarap, A.F. Deep Learning using Rectified Linear Units (ReLU). *arXiv* **2019**, arXiv:1803.08375.
37. Masters, T. *Practical Neural Network Recipes in C++*; Morgan Kaufmann: San Mateo, CA, USA, 1993; pp. 242–379.
38. Ying, X. An overview of overfitting and its solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [CrossRef]
39. Zou, F.; Shen, L.; Jie, Z.; Zhang, W.; Liu, W. A sufficient condition for convergences of adam and rmsprop. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11127–11135.
40. Christoffersen, P.; Jacobs, K. The importance of the loss function in option valuation. *J. Financ. Econ.* **2004**, *72*, 291–318. [CrossRef]
41. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model. Dev.* **2014**, *7*, 1247–1250. [CrossRef]
42. Sarkar, S.; Patel, A.; Madaan, S.; Maiti, J. Prediction of occupational accidents using decision tree approach. In Proceedings of the 2016 IEEE Annual India Conference (INDICON), Bangalore, India, 16–18 December 2016; pp. 1–6.
43. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
44. Sayin, C.; Ertunc, H.M.; Hosoz, M.; Kilicaslan, I.; Canakci, M. Performance and exhaust emissions of a gasoline engine using artificial neural network. *Appl. Thermal Eng.* **2007**, *27*, 46–54. [CrossRef]
45. Najafi, G.; Ghobadian, B.; Tavakoli, T.; Buttsworth, D.R.; Yusaf, T.F.; Faizollahnejad, M. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. *Appl. Energy* **2009**, *86*, 630–639. [CrossRef]
46. Yusaf, T.F.; Buttsworth, D.R.; Saleh, K.H.; Yousif, B.F. CNG-diesel engine performance and exhaust emission analysis with the aid of artificial neural network. *Appl. Energy* **2010**, *87*, 1661–1669. [CrossRef]

47. Alahmari, S.S.; Goldgof, D.B.; Mouton, P.R.; Hall, L.O. Challenges for the repeatability of deep learning models. *IEEE Access* **2020**, *8*, 211860–211868. [[CrossRef](#)]
48. Liu, Y.; Wu, Z.X.; Zhou, H.; Zheng, H.; Yu, N.; An, X.P.; Li, J.Y.; Li, M.L. Development of China Light-Duty Vehicle Test Cycle. *Int. J. Automot. Technol.* **2020**, *21*, 1233–1246. [[CrossRef](#)]
49. Bokde, N.D.; Yaseen, Z.M.; Andersen, G.B. ForecastTB—An R package as a test-bench for time series forecasting—Application of wind speed and solar radiation modeling. *Energies* **2020**, *13*, 2578. [[CrossRef](#)]
50. Zhou, F.; Joshi, S.N.; Rhoté-Vaney, R.; Dede, E.M. A review and future application of Rankine Cycle to passenger vehicles for waste heat recovery. *Renewable and Sustainable Energy Reviews* **2017**, *75*, 1008–1021. [[CrossRef](#)]
51. Zacharof, N.G.; Fontaras, G. *Report on VECTO Technology Simulation Capabilities and Future Outlook*; Publications Office of the European Union: Luxembourg, 2016.