Article

Development of Patent Technology Prediction Model Based on Machine Learning

Chih-Wei Lee 1, Feng Tao 1, Yu-Yu Ma 2 and Hung-Lung Lin 3,*

1 Institute of Industrial Economics, Jinan University, Guangzhou 510632, China; alanlee99983@gmail.com (C.-W.L.); fengtao@jnu.edu.cn (F.T.)
2 School of Education Science, Minnan Normal University, No. 36 Shi Qian Zhi St., Zhangzhou 363000, China; missuperyoyo@gmail.com
3 School of Economics and Management, Sanming University, No. 25 Ching-Tung Rd., Sanming 365004, China
* Correspondence: hsa8936.hsa8936@msa.hinet.net

Abstract: Intellectual property rights have a great impact on the development of the automobile industry. Issues related to the timeliness of patent applications often arise, such as the inability of firms to predict new technologies and patents developed by peers. To find the proper direction of product development, the R&D departments of enterprises need to accurately predict the technology trends. Machine learning adopts calculation through a large amount of data through mathematical models and methods and finds the best solution at the fastest speed through repeated simulation and experiments, to provide decision makers with a reference basis. Therefore, this paper provides accurate forecasts through established models. In terms of the significance of management, the planning of future enterprise strategy can be divided into three stages as a short-term plan of 1–3 years, a medium-term plan of 3–5 years, and a long-term plan of 5–10 years. This study will give appropriate suggestions for the development of automobile industry technology.

Keywords: patent technology; intellectual property; automobile industry; artificial neural network; machine learning; ensemble learning

MSC: 68Uxx; 68Wxx

1. Introduction

With the development of China’s economy, China’s automobile industry has transformed from original equipment manufacturer (OEM) services between 2000 and 2010, to a new stage of independent research, development, and production of domestic brand cars through drawing lessons from the world’s famed factories. According to the survey report of the China Association of Automobile Manufacturers, the sales volume of China’s brands has increased year by year from USD 2.943 million in 2011 to USD 7.749 million in 2020, with an increase rate of about 263% [1]. In addition, the market share of China’s domestic automobile brands has grown year by year from 29% in 2011 to 38.4% in 2020. This shows that the innovation and R&D achievements of Chinese domestic automobile brands through OEM and technical cooperation have won the recognition of market consumers.

Because of the progress and development of information technology, many products have been designed to provide customers with a combination of Internet Plus and the intellectualized experience of new-generation products. Just like the theory of the product lifecycle (PLC), every product must go through stages of development, growth, maturity, and decline [2]. To survive in the competitive environment, every enterprise must launch new products through innovation and R&D before reaching the final stage of its lifecycle. As stated by Pryshlakivsky and Searcy, enterprises should adopt different measures to launch new products or services and give new value to products or services at different
stages of the product lifecycle to hold advantage in the highly competitive market and achieve the goal of sustainable operation [3].

Similarly, in the rapidly changing market environment of science and technology, the automobile industry needs to keep innovating. Based on the analysis of 13 years’ patent data of 39 innovative enterprises in China’s telecommunications, electrical machinery, automobile, and pharmaceutical industries, the cooperation breadth of employees has a positive impact on the innovation performance of enterprises [4]. The development of new automobile products refers to a series of decision-making processes from the research for selecting products that meet the needs of the market through to product design, and on to manufacturing process design, until normal production. Product development involves a wide range of aspects including design, engineering analysis, trial production, and experimentation with the components or technologies of new vehicles, such as improvement of automobile engine power, vehicle crash tests, computer-aided AEB automatic brakes, and so on [5]. Hence, the investment in automobile products must allocate limited resources to the projects which need to be developed effectively to achieve the best results.

The key process of the development of new automobile products is to accurately determine the direction of a new product’s development. Automobile products differ from other products in the characteristics of its products, in addition to both innovation and science and technology, and the safety requirements are critical. As a consequence, it is very important to select the right direction to reduce the risks in development. Many studies have pointed out the risks of new product development, which can be manifested in the following aspects [6–8]:

1) Technology development risk:

This refers to the requirement that the developed technology conform to scientific principles, but the reasons for development failure are complicated. For example, it is difficult to complete the research or meet technical difficulties because of the immature technology at the present stage. Chin et al. [9] proposed several new product development risks, which are described as follows. (a) The production risk refers to the failure to meet production requirements within a predetermined time. (b) The R&D risks caused by a person who cannot complete the product specification design within the expected time. (c) Supplier risk means that the supplier may not provide good materials or may not provide them within the expected period. (d) Product reliability risk refers to the risk that the expected performance will not be achieved under normal manufacturing procedures. It may be that inadequate conditions lead to the failure of the research, or it may be that the preconceived ideas of the participants turn out to be wrong and unworkable.

2) Market competition risk:

This refers to the risk caused by market competition in the market. According to Allayannis et al. [10] and Guay and Kothari [11], the risk of market competition is composed of the following factors. (a) The scale of market competition: the greater the competitive power and cost of competitors, the greater the market risk will be (Allayannis et al., 2001). (b) The intensity of market competition: this is mainly reflected in the competition for market share to improve sales and profitability (Guay and Kothari, 2003). (c) R&D competition: this happens when the technology is still in progress and has been successfully developed by other researchers.

3) Risk of an objective environment:

Due to changes in the objective social, economic, and technological environment, the original technological development is out of date or no longer necessary. The automobile industry is a typical intellectual-property-intensive industry, and its development depends heavily on intellectual property rights. In particular, the global automobile industry is at a critical moment of industrial restructuring and technological transformation. As an important force in the global automobile industry, intellectual property rights are indispensable.
Generally speaking, before developing a new technology or innovative products, the enterprise will first check the patent application data to determine whether there is similar technologies or products extant in the industry. In the process of patent inquiry, the timeliness of an application for patent certification is crucial; otherwise, the technology or product in the certification or research and development process cannot be queried correctly. For example, it is nearly impossible to know whether other companies are developing the same innovative technology or product, or whether the competition is still applying for certification. Finally, after the completion of technology or product development, the patent application often leads to disputes, thus causing significant damage to enterprises. According to the State Intellectual Property Office [12], the number of intellectual property rights and competition dispute cases in the automobile industry showed an increasing trend year by year from 2009 to 2018, with an average annual increase rate of 28.22%. Among the intellectual property and improper competition disputes in the automobile industry (manufacturing) in 2019, intellectual property infringement disputes accounted for 76.83%, unfair competition disputes accounted for 13.26%, and intellectual property contract disputes accounted for 9.45%. Presently, the global automobile industry is in a critical moment of industrial restructuring and technological transformation. As an important force in the global automobile industry, China’s intellectual property rights of new patented technologies must be a top priority, whether this involves setting up joint ventures with famous overseas automobile enterprises, carrying out technology research and development cooperation projects with all parties, or designing and implementing OEM models. In the process of new product technology research and development, new patent disputes are an important issue worthy of attention. For example, in 2020, the proportion of patent infringement in China is as high as 10.8% [12].

In the automobile industry, many scholars have proposed effective methods to predict the technology, production, inventory, sales, and market conditions of the automobile industry, such as Yuan and Cai [13]. The growth curve method and entropy method were proposed to predict the future situation of vehicle power energy. The research results show that hybrid electric vehicles have the most promising future development prospects, followed by battery electric vehicles and traditional internal combustion engine vehicles. The development of fuel cell electric vehicles is slow. Hanggara [14] used the method of moving average combined with market supply and demand to forecast the automobile production in Indonesia, to reasonably estimate the production volume and solve the problem of overproduction. Babai et al. [15] put forward Bayesian parametric frequency and non-parametric methods for empirical evaluation of the demand of about 3000 inventory units in the automobile industry, and compared the proposed method with other methods. The results also proved that the proposed method could provide a more accurate decision-making reference for the inventory management plan of the automobile industry. Wan et al. [16] proposed the integration of principal component regression with neural network, support vector machine, and other methods, aiming to formulate a sales prediction model for electric vehicles. Wan et al., (2021) proposed the integration of principal component regression with a neural network, a support vector machine, and other methods for a sales prediction model of electric vehicles (EV). After an example analysis, it was verified that the integrated model had a good practical forecasting effect. The principal component regression–back propagation (PCR-BP) model and the principal component regression–support vector machine (PCR-SVM) model are better than a single model, such as the support vector machine model alone. Tsang et al. [17] proposed the fuzzy-based battery lifecycle prediction framework (FBLPF) to effectively manage the automobile market. This framework integrated the multi-responses Taguchi method (MRTM) and the adaptive neuro-fuzzy inference system (ANFIS), and the market forecast of the product lifecycle of electric vehicles can be carried out through the integrated method. The research results prove the accuracy of the proposed method and put forward corresponding plans and countermeasures for the sustainable development of energy and environmental protection in the automobile industry. Although these methods can effectively forecast
the market profile and development prospects of the automobile industry, they must also reduce the risk involved in the automobile industry for enterprises in the process of product production, sales, and R&D. However, for professional managers or decision makers, the mathematical calculation process is complicated and tedious. Based on this point of view, many researchers have adopted machine learning methods to predict the technology development in the automobile industry. For instance, Lee M. [18] adopted the text mining model method of machine learning. Research on AI algorithm classification using patent data submitted from 1980 to 2017 can demonstrate the dynamic change pattern of the fusion of AI and EV technology. Wang et al. [19] stated that new product development in China’s automobile industry can be predicted by machine learning methods. From 2001 to 2014, they obtained 1088 valid sample datasets from the Chinese automobile industry to construct an evaluation indicator system. Choi et al. [20] raised a hybrid method, which takes into account expert opinions, patents, and machine learning methods, analyzes the results, and combines semi-supervised learning with active learning to effectively find emerging and promising technologies. Lee et al. [21] helped pharmaceutical technology identify emerging patented technologies through machine learning and multilayer neural network model calculation and simulation. Teng et al. [22] used a VSM (vector space model) and K-MEAN to solve the technical problem of batteries developed by new technologies through machine learning. Suhail et al. [23] used machine learning to integrate random forest and decision tree to carry out calculations and simulations to help dentists make decisions during tooth extraction, avoiding errors caused by human judgment. Barrera-Animas et al. [24] used machine learning to conduct simulation and prediction, employing linear regression and a support vector model, etc., to solve the expensive and complex problems faced by the existing five major cities in the UK in meteorological forecasting. Ensafi et al. [25] adopted machine learning to predict the sales of seasonal goods with ARIMA and convolutional neural networks (CNN). Its advantage is that it can find out the most accurate trend of commodity sales through repeated calculation, help enterprises accurately forecast sales volume, and avoid the problem of overproduction and inventory in order to provide decision makers with a quick method for judgment. Machine learning adopts calculation through a large amount of data using mathematical models and methods and finds the best solution at the fastest speed through repeated simulation and experiment, providing decision makers a reference basis. Many studies also show the practicability and value of machine learning methods. Therefore, this paper proposes a new prediction method—“The concept of technology maturity combined with machine learning model”. This method can model and forecast the status quo of patent technology in the automobile industry, aiming to provide decision makers and managers in the automobile industry with an accurate prediction of the trends of new technologies or products in the future market, before investing in research and development. Finally, the method reduces the significant losses caused by the patent disputes mentioned above. In addition, this study takes a case study of patented body technology in China’s automobile industry for modeling and analysis, and compares the model accuracy of traditional time series and non-traditional machine learning methods, respectively, to prove that the proposed method is more accurate and stable than others, thus proving the stability and applicability of the model. Importantly, the proposed method can provide a systematic and scientific reference for decision makers and managers in the automobile industry to initiate technology or product R&D plans, and put forward valuable suggestions for researchers and enterprises.

To sum up, the background and purpose of this study are introduced in the introduction, and the main research structure and core are specifically divided into the following four parts. The literature review in Section 2 includes the technology forecasting method and the research on innovation for R&D and patent market forecasting. Section 3 mainly concerns the construction of the trend model of the R&D patent market, including a machine-learning-building, integrated-learning prediction model. Section 4 presents a case study analysis, including research analysis, model verification and discussion, and predic-
tion of future development trends. The conclusion, presented in Section 5, summarizes the management and academic aspects of the proposed method.

2. Literature Review

2.1. Technology Prediction Methods for Innovation and R&D

A prediction is an estimate or calculation made about a future outcome that people are concerned about, or about an uncertain event that people want to comprehend in advance [26]. Making predictions is very important for business operations. Through different forecasting methods, decision makers of enterprises can understand the economic development or the future changes of the market to form the goals or decisions of their enterprises. Through scientific management methods, the risks and costs of enterprise operation can be reduced, and the enterprise objectives can be achieved smoothly [21]. Commonly used innovation and R&D technology forecasting methods can be divided into the following categories.

(1) Quantitative prediction method [21–27].

Quantitative prediction methods include trend extrapolation, analogies, causal models, and so on. Trend extrapolation is similar to the autoregressive integrated moving average model (ARIMA). Analogies are similar to support vector machines (SVM). Causal models are similar to linear regression (LR). Quantitative analysis has been successfully applied to different fields and to solve related problems. For example, it has been used to analyze the new patented technology of mobile communication in South Korea, based on the ARIMA model, and the information system of the Korean Intellectual Property Office (KIPO) was adopted for data collection [27]. According to the International Patent Classification (IPC), 20,294 patents were classified into 152 categories. Finally, Korea’s major mobile communication technologies were classified into four categories. This provides an important reference standard for decision makers in government departments and related industries when investing in R&D. Researchers have used metrology and patent analysis to analyze the S curve in the logistic growth curve model of hydrogen energy and fuel cell technology, and they determined the best patent strategy for the fuel cell industry [28]. The results show that the S curve is an efficient means of quantifying a method of predicting the technology of cumulative published patent numbers. Researchers have used regression analysis to evaluate weapons technology in the defense industry, and proposed a method of constructing a technology map, which divides technologies into four categories according to their technical effects [29].

These studies confirmed the advantages of quantitative analysis, as follows: (a) different standards or variables can be considered at the same time, that is, an approach can include different standard variables in experiments and analyses in different environments to obtain the best results; (b) technology forecasts can be adapted to different industries; (c) quantitative forecasts can be applied to different products.

(2) Qualitative prediction methods and the combination of qualitative and quantitative methods [30–32].

Expert group judgment has different applications, such as the Delphi technique, interview, brainstorming, and nominal group techniques. The Delphi technique is an expert judgment method often used in technical forecasting, which is especially applicable when historical data are insufficient and require objectivity and independence of expert judgment compared with the other three methods.

(3) The combination of qualitative and quantitative methods [30–32].

By and large, the Delphi technique, focus group interviews, and brainstorming are commonly used to cope with multicriteria decision-making methods of quantitative analysis, such as analytic hierarchy process (AHP), analytic network process (ANP), entropy, and the technique for order preference by similarity to an ideal solution (TOPSIS). This combination has been successfully applied in different fields and solved related prob-
lems. For example, researchers used the expert interview method and ANP to predict the warehousing operation of out-stock and in-stock of the logistics center of a chain supermarket and optimized its warehousing classification management [30]. Researchers have used the Delphi technique and AHP to evaluate the factors that affect the success of start-up companies when rice bran polysaccharide is used in the Taiwan venture capital industry [31]. Researchers have used expert groups combined with entropy and TOPSIS to classify and forecast the warehouse management of green plant e-commerce vendors, and they developed methods for warehouse optimization [32]. These studies confirm the advantages of combining qualitative and quantitative methods, for example, (a) obtaining a variety of different but valuable perspectives; (b) being able to apply these perspectives to long-term and new products’ forecasting.

To sum up, technological innovation prediction is the premise and basis of enterprise technological innovation decision making. Through the evaluation of innovation or research and development, enterprises can obtain an accurate sense of future technological development and the changing trends. This provides a scientific basis for enterprises to reduce subjectivity and blindness in processes of technological innovation decision making. In the competitive market and complicated environments, an enterprise’s technological innovation determines its survival and development. Therefore, to ensure the correctness of technological innovation, enterprises should choose appropriate forecasting methods according to different environmental factors, such as time and place, to reduce the risks involved in enterprise operation.

2.2. Research on Innovation, R&D, and Patent Market Prediction

There have been many studies using various forecasting methods in patent R&D demand or market demand as follows.

(1) Research on traditional forecasting methods in patent technology and market demand.

Researchers have proposed that the future market trend and new patents of the home appliance industry can be predicted by combining the pearl curve with related indicators of home appliance isolation technology, and the results showed that the proposed method is effective in application [33]. Researchers classified and predicted the patent of “coherent light generator” based on bass and the ARIMA time series model, they and proposed a new classification standard for this technology (mainly divided into the first class and subclasses) [34]. Finally, viewpoints and countermeasures have been put forward for the future trend of first-class patents and subclasses of technologies through the analysis results. The authors of [35] proposed the use of LR and clustering technology to predict the future trend of new product development, supply, and demand in many global industries. The research revealed that the innovation of technology will accelerate the development of PLC for the uncertainty of products and sales demand, and proved that the proposed method can be used as an effective tool for decision making. The authors of [36] proposed the application of patent analysis with the concept of growth curve and technology maturity to predict the development of spare wind turbine technology.

The results show that the technology of jet engine wind turbines is in the early-maturity stage, the gearless wind turbine is at the end of its growth curve, and the airborne wind turbine is at the end of the maturity stage on the growth curve, which proves the effectiveness of the proposed method. The authors of [37] used support vector machines (SVMs) to conduct progressive analysis of difficult classification problems of patents. The results indicated that the proposed method can effectively classify patents and provide an important reference standard for inventors or lawyers when facing related problems. The authors of [38] used the S-curve and LR method to analyze the data of the United States Patent and Trademark Office (USPTO). New patents for unmanned vessel technologies (UVTs) were studied and the current technology stage of UVT was determined. The result reflected that UVTs are in the growth stage of their technology lifecycle and represents an emerging technology with future investment value.
(2) Research on machine learning in patent technology prediction.

The authors of [39] stated an improved method of machine learning to predict emerging technologies and verified it with the patent data provided by the United States Patent and Trademark Office. The research presented that the proposed method can effectively predict the future development of new technologies with an accuracy of up to 70%, which helps enterprises reduce costs and risks in the process of innovation and R&D, and enables enterprises to effectively carry out strategic investment. The authors of [40] applied patent and machine learning methods to design a new method combining coding and tag coding based on existing research on patent grant term prediction. The results show that the proposed method can effectively confirm the patent application grant period in the data of the Indian Patent Office. The authors of [41] used patent-related data provided by the United States Patent and Trademark Office to predict patents in the healthcare industry and classified different technologies by using the standard of cooperative patent classification. This study assessed the potential of different technology clusters in foreign countries to provide a reference for decision makers or managers of enterprises or national regulatory authorities regarding future investment in innovative R&D.

Cho et al., (2021) first constructed a communication network with association rules. Machine learning methods were then used to predict the future using various link prediction indices, and finally latent Dirichlet assignment (LDA) topic modeling was used to identify keywords related to the technology that is expected to converge [42]. The analysis of patent data of 2012–2014 from the US Patent and Trademark Office in the chemical engineering and environmental technology fields showed that the random forest model in machine learning has the best prediction effect on a 4-year interval. By predicting the new technology fields that may emerge in the future, the study could provide direction suggestions for companies focused on technological advances. The authors of [43] analyzed complex patent problems by combining self-organizing map (SOM), principal component analysis (PCA), and support vector machine calculus with machine learning methods, then compared it with a single machine learning method. The results showed that the proposed integrated machine learning method was more accurate and saved more resources than the single machine learning method. Using a machine learning and multilayer neural network method, researchers selected 18 input and 3 output indicators from the database of the United States Patent and Trademark Office for pharmaceutical technology and explored the nonlinear relationship between input and output indicators [21]. The result indicated that the multiple patent indicators can be used to identify whether a drug is worth developing at an early stage, before it was developed into a new pharmaceutical technology. The authors of [44] put forward the method of machine learning and semantic analysis of patented text information to judge the patented technology of vehicle signal and electronic message transmission, and to predict the trend of future development.

In summary, traditional patent and market demand forecasting has proved that the proposed methods can measure the utility value of new products or technologies in the input process and reduce the risk of enterprises, such as the dispute cases of new patents. However, these traditional methods also have many disadvantages or deficiencies, such as cumbersome and slow calculation processes, difficult data collection, uncertain information, and other problems [45–47]. Therefore, many studies put forward machine learning to replace traditional prediction methods, and the above research has proved the effectiveness of machine learning as a prediction method. Based on the existing research, this paper proposes various prediction and integration algorithms of machine learning, compared the time series methods, and proposed the feasibility of an innovative patent prediction method after a comprehensive comparison. Table 1 presents a comparison of the advantages and disadvantages of the proposed method and other model methods.
Table 1. Comparison of the advantages and disadvantages of the proposed method and other model methods.

| Aspects   | Proposed Method                                                                 | Other Methods (Qualitative/Quantitative)                                                                 |
|-----------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| **Advantages** | (1) It can be combined with other classification and regression algorithms to improve its accuracy and stability and avoid overfitting by reducing the variation of results. | (1) A variety of different and valuable points of view can be gained. |
|           | (2) It is composed of processing nodes similar to human brain neurons. The greatest advantage of a neural network is that it can accurately predict complex problems. | (2) It is suitable for long-term prediction and prediction of new products, and can be used when historical data is insufficient. |
|           | (3) The support vector machine method can effectively solve the classification and regression problems of high-dimensional features. | (3) This method can make up for the lack of basic information. |
| **Disadvantages** | (1) This method is prone to overfitting. | (1) It is less reliable for product prediction by region. |
|           | (2) This method is sensitive to missing data. | (2) Qualitative advice is sometimes incomplete or impractical. |
|           | (3) There are many neural network parameters in this method. | (3) Generally, it is only applicable to the prediction of the total amount, but it has poor reliability when applied to regions, customers, and product categories. |

**Summary**

After comparison, there are three reasons for choosing this scheme: (1) the calculation will be faster; (2) the obtained model will be more accurate; (3) it is suitable for a large amount of data and for the method of applying mathematics to assist in making decisions.

3. Model Construction of R&D Patent Market Trend

This paper will build the model in separate three stages. The first stage is “machine learning—the construction of the ensemble learning prediction model”. The theory of the model and the constructing procedure will be explained in this stage. The second stage is “the model validation”. In this stage, the data of car body patent applications in China’s automobile industry will be taken as a case study. In this paper, the relevant data collected by the Chinese Intellectual Property Office are used for modeling and analysis, and the errors between the proposed model and the traditional prediction model are compared to prove the accuracy and applicability of the proposed method. The third stage is “the forecast for future trends”. Some suggestions and countermeasures are put forward for the analysis of market demand information of the automobile industry, providing a reference for use by the relevant personnel of the automobile industry and scholars. The construction process of this research model is described as follows, and the specific research framework is illustrated in Figure 1.

3.1. Stage 1: Machine Learning—The Construction of the Ensemble Learning Prediction Model

According to the research framework in Figure 1, the first stage, Stage 3.1, is machine learning—the construction of the ensemble learning prediction model. It includes data mining and the construction of the ensemble learning model and the ensemble learning prediction model.
3.1.1. Data Mining and Machine Learning—Ensemble Learning Model

First proposed by Breiman [46], bagging (bootstrap aggregating) is an ensemble learning algorithm for machine learning. Bagging combines the prediction results of each learning model through voting rules to make the final classification prediction. First, train all the other algorithms with the data, and then take all the predictions of the other algorithms as additional input. In general, bagging is an isomorphic model, and the same model is used for training in other learning model algorithms. After that, hard voting or soft voting rules are used to combine the prediction input of the other algorithms mentioned above to obtain the final prediction classification result.

Breiman [46] proposed the bagging (bootstrap aggregating) method, which combines multiple different prediction models by voting or averaging. Although each prediction model uses the same learning algorithm, they all adopt different training datasets. A schematic diagram of the bagging algorithm is shown in Figure 2.

Figure 1. Research framework diagram.
The principle of the bagging algorithm is as follows: Given a training set, \( D \), with size \( N \), the bagging method selects \( m \) subsets \( D_i \) with size \( N \) uniformly, and with the return (using the self-help sample method) as a new training set. \( M \) models can be obtained by using classification and regression algorithms on \( m \) training sets, and then the results of the bagging method can be obtained by taking average values and majority votes. In the end, the accuracy and stability are improved, while the variance of results is reduced to avoid the occurrence of overfitting.

The principle of the Bagging algorithm is described as follows. When a given dataset is \( L = \{ (x_1, y_1), \ldots, (x_m, y_m) \} \), the basic learner is \( h(x, L) \). If the input is \( x \), then \( Y \) is predicted by \( h(x, L) \). Suppose that there is a dataset \( \{ L_K \} \) sequence, each consisting of \( M \) independent observations from the same distribution as \( L \). The task was to obtain better learning results by using \( \{ L_k \} \), which was stronger than learning \( h(x, L) \) in a single dataset. This requires the use of learning \( \{ h(x, L_K) \} \) sequences. If \( y \) is a number, then the process is to replace \( \{ h(x, L_K) \} \) with the average of \( \{ h(x, L_K) \} \) over \( K \), that is, \( h_A(x) = E_L h(x, L) \). Where \( E_L \) represents the mathematical expectation of \( L \), and the subscript \( A \) of \( h_A \) represents a composition. If \( h(x, L) \) predicts class \( j \in \{ 1, \ldots, J \} \), then one way to synthesize \( h(x, L_K) \) is by voting. Let \( M_j = \#\{ K, h(x, L_K) = j \} \) so that will be \( h_A(x) = \arg \max_j M_j \).

Step 1: Determining the target objects.

Firstly, the number of global body patents is taken as the target to predict the number of future patents.

Step 2: Collecting data.

Next, the patent data provided by the Chinese Intellectual Property Office are used to screen the patent data, patent classification, and the patent pool through retrieval. A total of 46 years of statistical data from 1974 to 2020 were used to collect and preprocess the data.

Step 3: Analyzing data.

Finally, WEKA software is applied to perform data mining tasks. After the process of data preprocessing, clustering, classification, and testing the model and the parameters, the machine learning technology is applied to automatically perform a calculation to obtain the model and the parameter values.
3.1.2. On the Basis of Machine Learning—The Construction of the Ensemble Learning Prediction Model

Ensemble learning, also known as multiple classifier systems, is composed of multiple base learners whose spirit is to gather the “wisdom of crowds”. The generalization ability of ensemble learning is usually stronger than that of a base learner. The base learner can be generated by substituting the base learning algorithm into the training sample, and such a base algorithm includes a decision tree, a neural network, etc., though most ensemble learning methods use a single base.

A learning algorithm is used to produce homogeneous base learners; there are still other methods that use different learning algorithms to produce heterogeneous learners. Additionally, because there is no single base learning algorithm, basic learners can also be used as component learners or individual learners.

In principle, ensemble learning is divided into two steps. First, several basic learners are generated in parallel or sequential patterns. Later, all the basic learners are used together, and common merging methods include the concept of majority voting (classification problems) and the concept of weighted averaging (regression problems). Generally speaking, to attain a good ensemble learning model, the base learner should be as accurate as possible, but also as diverse as possible. The accuracy of a learner can be measured using cross-validation or hold-out tests, but there is no rigorous measure of diversity.

There are many approaches to ensemble learning, three of which are described in detail here.

1. Leo Breiman proposed bagging, also known as bootstrap aggregation or bootstrap, as a simple and powerful ensemble learning method. Meanwhile, many homogeneous weak learners are considered, and these weak learners are independent and parallel-constructed; their respective results are determined by averaging or voting [46].

2. Boosting, first put forward by Freund [47], is also a weak learner with a good deal of homogeneity. Unlike bagging, these basic models adapt and learn sequentially and combine the results in a deterministic strategy.

3. Stacking is a weak learner using heterogeneity. It can construct the respective models in parallel and combine the prediction results of different weak learners to train a metamodel and draw conclusions.

Ensemble learning is a kind of supervised learning. This method establishes multiple hypotheses by multiple learning algorithms, and combines them into a whole hypothesis by way of weight, so as to make a reasonable prediction of the test data. Many studies have shown that prediction using ensemble learning is more accurate than a single hypothesis.

In the learning algorithm, the training data need to be set up first. Each training material is made up of a special vector and a category tag, \( Y \). Secondly, the real function is computed. Suppose the real function, \( f \), exists, such that the identity \( y = f(x) \) is true. Finally, the learning algorithm and hypothesis are verified. The goal of the learning algorithm is to find a hypothesis, \( h \), such that \( h \approx f \) formula.

The ensemble learning model consists of a set of hypotheses \( \{h_1, h_2, \ldots, h_n\} \) and a set of hypothesis weights \( \{W_1, W_2, \ldots, W_n\} \), as shown in Formula (1):

\[
h(x) = W_1h_1(x) + W_2h_2(x) + \ldots + W_nh_n(x) \tag{1}
\]

where \( h(x) \) is the ensemble learning model, \( \{h_1, h_2, \ldots, h_n\} \) is a set of hypotheses constructed by multiple learning algorithms, \( \{W_1, W_2, \ldots, W_n\} \) is the corresponding weight of each hypothesis, and the final prediction is obtained by combining the weight of each hypothesis and individual hypotheses. For example, first, apply the WEKA software, then select the decision stump classifier, and select tenfold cross-validation for test option training and evaluation. Next, select the AdaboostM1 classifier, which is an ensemble learner using the lifting algorithm. To compare with the decision stump classifier, the base classifier of AdaboostM1 is set as the decision stump classifier. After confirmation, select the training button for training, and there are multiple classifiers to choose from. The number of
iterations in the parameter setting is set to 10 by default, that is, the training will carry out
the decision stump classifier 10 times. The schematic diagram of the ensemble learning
model of this paper is shown in Figure 3.

![Figure 3. The schematic diagram of the ensemble learning model.](image)

4. Research Analysis—Car Body Patent Forecasting for the Automobile Industry

The main content of Section 4 includes the following three parts. Section 4.1 is the
research analysis. This part mainly explains the source and basis of the collected data.
Section 4.2 is the model verification and discussion. This part focuses on analyzing each
model based on the data in Section 4.1 to verify the accuracy of the proposed model and
each model. Section 4.3 is the forecast of the future trend.

4.1. The Research Analysis

This paper takes vehicle safety collisions as samples. The main reason is that in
addition to the safety of the car body, it also includes seat belts, airbags, safety seats, and
automatic emergency braking, etc. The development and manufacture of these safety
protection measures will be based on the main consideration of car body safety collision.
For example, when the car body design is not secure, safety belts, airbags, and seats cannot
be functional in protecting life safety [48]. In addition, the body structure has also been
affected by environmental protection policies in recent years. For example, strengthening
the rigid structure of the body will increase the weight of the body, which may cause more
fuel or electricity consumption. Therefore, this paper will take the body collision as the
research sample, and is expected to put forward specific development countermeasures
and suggestions for future patent research and development of the automobile industry
through the examination results.

Step 1: The collection of sample data.

The patent database provided by the Intellectual Property Office of the People’s
Republic of China was used as the basis for data analysis. It is difficult to search for the
correct patent data of safety car bodies. This is mainly because of the large amount of
literature, and the number of preliminarily searched studies reached 30,000. In addition,
the International Patent Classification (IPC) standard used does not classify according to
different models (such as trucks, small buses, buses, etc.), and there is no unified standard
for some key components (such as beams, plates, columns, etc.). Therefore, the classification
systems of Japan and the United States were adopted in this study as the main retrieval
basis, and then the keywords of IPC classification were used for retrieval (as shown in Table 2). Finally, the recall rates of the three famous Japanese automobile companies, including Toyota Auto Corporation, Mazda Motor Corporation, and Mitsubishi Auto Industry Corporation, were taken as data samples. For American studies, 90 studies were randomly selected for manual reading, and 83 studies related to the retrieval subject were obtained, with an accuracy of about 92%. Finally, the accuracy was 100% by manual work. The scope of patent data is a total of 46 years from 1974 to 2020.

Table 2. Table of automobile crash safety technology.

| Level 1 Classification | Level 2 Classification | Level 3 Classification            |
|------------------------|------------------------|-----------------------------------|
| Safe car body          | Car body that reduces front impact damage (B62D 21/00; B62D 23/00; B62D 25/00) | Front cross member               |
|                        |                        | Front rail                        |
|                        |                        | Impact energy absorbing device    |
|                        |                        | A pillar                          |
|                        |                        | Upper rail                        |
|                        |                        | Door panel                        |
|                        |                        | Front floor                       |
|                        |                        | Front panel                       |
|                        |                        | Subframe                          |
|                        |                        | Splash shield stiffener           |
|                        |                        | Combinatorial optimization and others |
| B pillar               | Car body that reduces side impact damage (B62D 21/00; B62D 23/00; B62D 25/00) | Lower rail                       |
|                        |                        | Door panel and guard assay        |
|                        |                        | Floor assembly                    |
|                        |                        | Roof member                       |
|                        |                        | Combinatorial optimization and others |
| C pillar               | Car body that reduces rear impact damage (B62D 21/00; B62D 23/00; B62D 25/00) | Back floor                       |
|                        |                        | Back rail                         |
|                        |                        | Back cross member                 |
|                        |                        | Combinatorial optimization and others |

Step 2: Pretest the predicted data.

The classification of automobile industry technologies in this study is shown in Table 2. The first level of classification is the safe car body, and the second level of classification is subdivided into the front-collision-damage-reduction car body, side-collision-damage-reduction car body, and rear-collision-damage-reduction car body, and finally corresponds to the parts of the third level of classification.

The development of the retrieval strategy, using the International Patent Classification (IPC) as a large category, is coordinated with Table 2. For example, when the keyword B62D corresponds to the safe car body, it can combine the standard classification with the actual terms used in the automobile industry.

The data were taken as samples from 1974 to 2011, and the linear regression method in machine learning was used to establish the mathematical model and judge whether the fitting was feasible. For example, Rsq > 80%: if the model is established, then patent data can be further predicted. First, the accumulative number of automobile body patents is predicted to be 4772 in 2012 and 5484 in 2016. It can be judged that the model is regular and predictive. Second, the cumulative number of predicted patents for the front collision of the car body is 2856 in 2012 to 3144 in 2016, showing that the model is regular and predictive. Third, the accumulative number of side collision patents was 1511 in 2012, but it could not predict the number in 2016, so it showed that the model did not have regularity and predictability. For the fourth, the accumulative number of rear collision patents for the car body was 454 in 2012, but the number in 2016 could not be predicted, so the model can be considered to not have regularity and predictability.
Therefore, the car body patent data using International Patent Classification (IPC) and level 1 classification are regular and predictive. The data of the International Patent Classification (IPC) and the level 2 classification, such as front, side, or back collision, are not regular and predictable, so it is easy to misjudge the prediction of the technology lifecycle. Consequently, the subsequent use of data will mainly be International Patent Classification (IPC) and level 1 classification of the car body patent.

Step 3: The consistency test.

The model adaptability was tested with the level 1 data of the safe car body according to the results of the pretest in Step 2 in this paper. The general methods of ensemble learning include voting, boosting, and bagging. In terms of effect, the AdaBoostM1 algorithm of boosting is the most effective. The idea of the bagging method is to train multiple classifiers with random sampling which are put back so that the “lower-level” classifiers pay more attention to the misclassified data of the “upper level”. Finally, the result of each classifier is weighed and combined to make the decision. Voting applies multiple classifiers for optional combinations, but the disadvantage is that the majority rules can only avoid the worst-case scenarios. Therefore, in this study, boosting and bagging were used to test adaptability, and voting was excluded.

A decision stump was used as a classifier to verify the adaptability of boosting. The accuracy rate of the test results of the first training was 76.59%. The higher the accuracy rate is, the better result it is, and it is considered practical when the accuracy rate is over 80%. If it cannot meet the standard, then data training needs to be redone to improve the accuracy rate. In the method of improving the accuracy of data, AdaBoostM1 was used to train the data, and the parameters were set to 10 iterations. The final classification accuracy was 100%, which represented the adaptability of the boosting method in this paper.

In this paper, folding cross-validation was used as the test option for training and evaluation of bagging adaptability, and the classification accuracy rate was 96.69%. The accuracy of the bagging method meets the requirements as long as it reaches 80%, but in this paper, the classifier training was carried out for the second time in pursuit of higher accuracy and the Automobile WEKA was used to improve the performance of the ensemble learner. The random forest classifier was used as the comparison standard, and the final classification accuracy rate was 98.88%, which showed that the results could make a correct judgment.

In brief, the results of the above two stages showed that the adaptability test of this paper corresponds to the hypothesis of the theory, so the results can provide a relatively reliable guarantee for the subsequent prediction results.

Step 4: The error result prediction.

Based on the test results of Step 3, this study used the ensemble bagging method to make a prediction and compared the errors of each period through the historical data from 2001 to 2020. The prediction results of each period are shown in Table 2. The average absolute error was 55.25, which was more accurate than other single prediction models. In order to verify the accuracy of the proposed model, this paper analyzes and compares the traditional prediction methods in Section 4.2 to prove the feasibility and applicability of the proposed model in patent prediction.

4.2. Validation and Discussion of the Model

In order to verify the accuracy of the model, this study is explained in three parts, as follows. The first part is to compare the accuracy with different prediction methods. In this study, absolute error was used to verify the accuracy of the proposed model and other methods, such as pearl curve, the ARIMA method, the regression method, the support vector machine (SVR) method, and the ensemble bagging method of neural networks (BPR), to prove the validity of the proposed method. The second part is to compare the accuracy with the posterior error test. In this study, the method proposed by Julong, D [49]
was used to calculate the error of the prediction results, to judge the reliability of the prediction model.

The third part is the co-integration test (CI) and the error correction model (ECM). The CI method was proposed by Engle and Granger [50], and it mainly conducts unit root test on the residual of the regression equation. If the residual sequence is stationary, then it indicates that there is a co-integration relationship between the variables of the equation, otherwise there is no co-integration relationship. The ECM, proposed by Davidson et al. [51], is mainly the influence of short-term fluctuations of variables. Secondly, variables deviate from the long-term equilibrium relationship in short-term fluctuations. These three parts are described below.

(1) The comparison of the accuracy with different prediction methods.

To verify the accuracy of the proposed model, pearl curve, ARIMA, the regression method, the support vector machine (SVM) method, the neural network (BPR), ensemble learning (bag method), and other methods are compared, respectively. The actual value, the theoretical value, the absolute error, and the mean absolute error of each period from 2001 to 2020 are presented in Table 3. According to the mean absolute error analysis, the ensemble learning (bagging method) is 55 and the pearl curve is 73.8, which shows that the bagging method model proposed in this paper is more accurate than the other models.

Table 3. Predictive performance and absolute error of global body patents.

| Year | Patent | Pearl Curve | ARIMA | Regression | Support Vector Machine | Neural Network | Ensemble (Bagging) |
|------|--------|-------------|-------|------------|------------------------|---------------|-------------------|
|      | Actual Value | Predictive Value | Error | Predictive Value | Error | Predictive Value | Error | Predictive Value | Error | Predictive Value | Error | Predictive Value | Error |
| 2001 | 209 | 206 | 3 | 217.5 | 9 | 185 | 24 | 333 | 124 | 217 | 8 | 189 | 20 |
| 2002 | 178 | 222 | 44 | 230.8 | 53 | 182 | 4 | 251 | 73 | 251 | 73 | 176 | 2 |
| 2003 | 241 | 234 | 7 | 244.9 | 4 | 180 | 61 | 233 | 8 | 212 | 29 | 181 | 60 |
| 2004 | 243 | 243 | 0 | 258.1 | 15 | 178 | 65 | 343 | 100 | 215 | 28 | 188 | 55 |
| 2005 | 358 | 247 | 111 | 271.5 | 87 | 175 | 183 | 246 | 112 | 224 | 134 | 191 | 167 |
| 2006 | 304 | 246 | 58 | 284.4 | 20 | 173 | 131 | 372 | 68 | 227 | 77 | 232 | 72 |
| 2007 | 364 | 240 | 124 | 297.2 | 67 | 171 | 193 | 198 | 166 | 250 | 114 | 254 | 110 |
| 2008 | 405 | 231 | 174 | 309.7 | 95 | 169 | 236 | 256 | 149 | 224 | 181 | 286 | 119 |
| 2009 | 368 | 217 | 151 | 321.9 | 46 | 167 | 201 | 483 | 115 | 225 | 143 | 320 | 48 |
| 2010 | 376 | 201 | 175 | 334 | 42 | 165 | 211 | 322 | 54 | 228 | 148 | 349 | 27 |
| 2011 | 219 | 184 | 35 | 345.8 | 127 | 163 | 56 | 418 | 199 | 227 | 8 | 165 | 54 |
| 2012 | 69 | 221 | 152 | 357.3 | 288 | 161 | 92 | 408 | 339 | 240 | 171 | 155 | 86 |
| 2013 | 124 | 205 | 81 | 368.6 | 245 | 159 | 35 | 421 | 297 | 224 | 100 | 145 | 21 |
| 2014 | 89 | 187 | 98 | 379.7 | 291 | 157 | 68 | 450 | 361 | 224 | 135 | 168 | 79 |
| 2015 | 95 | 170 | 75 | 390.6 | 296 | 155 | 60 | 457 | 362 | 224 | 129 | 168 | 73 |
| 2016 | 91 | 152 | 61 | 401.2 | 310 | 153 | 62 | 444 | 353 | 224 | 133 | 98 | 7 |
| 2017 | 110 | 136 | 26 | 411.5 | 302 | 151 | 41 | 434 | 324 | 229 | 119 | 102 | 8 |
| 2018 | 92 | 120 | 28 | 421.7 | 330 | 149 | 57 | 440 | 348 | 225 | 133 | 113 | 21 |
| 2019 | 105 | 106 | 1 | 431.5 | 327 | 148 | 43 | 438 | 333 | 225 | 120 | 93 | 12 |
| 2020 | 164 | 92 | 72 | 441.2 | 277 | 146 | 18 | 415 | 251 | 226 | 62 | 100 | 64 |

| Mean absolute error | NA | NA | 73.8 | NA | 161.39 | NA | 92.05 | NA | 206.8 | NA | 102.25 | NA | 55.25 |

It can be further observed from Figure 4 that the red curve is the number of existing patents (the actual value). After comparing with the theories predicted by other models, it can be concluded that ARIMA is more accurate in the early stage (2001–2009), and the proposed model (ensemble bagging) is more accurate in the middle stage (2010–2012) and the latter stage (2013–2020). However, they will lose accuracy in different intervals, and the method proposed in this paper has stronger stability. In addition, from the analysis results of the trend criterion, it can be observed that the bagging method stage result is better, and the stability of the three stages will be observed in the following part.
Figure 4. The trend of the actual patent number and various prediction methods.

As shown in Figure 4, the results of comparing the three intervals from the average absolute errors of different intervals are as follows. The first interval is from 2001 to 2009, in which ARIMA is the most accurate, with an error of 44. The second interval is from 2010 to 2012, during which the proposed model (the true ensemble bagging method) is the most accurate, with an error of 196. The third segment is from 2013 to 2020, which is also the most accurate model (the true ensemble bagging method), with an error value of 36. By comprehensive observation of Figures 4 and 5, although each method has its accuracy interval or period in different intervals or periods, the accuracy of the model proposed in this study (the true ensemble bagging method) is more accurate than other models in the overall trend (different intervals or periods).

Figure 5. Mean absolute error (lower means more accurate).

(2) The posterior error test method.

After establishing the integrated bagging model, the usability and reliability of the model were tested. In this paper, the posterior error test method was adopted. Let
\[ \delta_i = f_i - \bar{f}_i \ (i = 1, 2, 3, \ldots, n), \] where \( f_i \) is the number of patent applications in a certain year, \( \bar{f}_i \) is the estimated amount calculated by the integrated bagging model, and \( \delta_i \) is the residual.

\[ \bar{f} = \frac{1}{n} \sum_{i=1}^{n} f_i; \quad \bar{\delta} = \frac{1}{n} \sum_{i=1}^{n} \delta_i; \] Standard deviation of raw data \( S_1 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i - \bar{f})^2} \);

Standard deviation of residual \( S_2 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\delta_i - \bar{\delta})^2} \).

Calculate variance ratio \( c = \frac{S_2}{S_1} \) and small error probability \( p = p\{ |\delta_i - \bar{\delta}| < 0.6745S_1 \} \).

According to the values of \( C \) index and \( p \) index, the model level is determined as shown in Table 4. According to Table 5 data, \( f = 210.2, \bar{f} = 26.55, S_1 = 115.74, S_2 = 65.86 \). The posterior error ratio \( C = 0.56 \) and the small error probability \( p = 1 \) were calculated. By comparing the posterior test table (see Table 4), it can be concluded that the prediction of patent application volume by the integrated bagging model is level 3 (generally satisfied). The verification of the actual application data above shows that the bagging model has high reliability for patent prediction.

**Table 4.** Posterior error test table.

| \( p \) Index | \( C \) Index | \( p \) Index | Model Class |
|---------------|---------------|---------------|-------------|
| \( >0.95 \)   | \( <0.35 \)   | \( >0.9 \)   | Level 1 (very satisfied) |
| \( >0.8 \)    | \( <0.5 \)    | \( >0.7 \)   | Level 2 (satisfied) |
| \( >0.7 \)    | \( <0.65 \)   | \( >0.7 \)   | Level 3 (generally satisfied) |
| \( <0.7 \)    | \( \leq 0.7 \) | \( \leq 0.7 \) | Level 4 (unqualified) |

**Table 5.** Actual and forecast number of patents.

| Year | Quantity (Actual Value) | Forecast Quantity | Residual Error |
|------|-------------------------|-------------------|---------------|
| 2001 | 209                     | 189               | 20            |
| 2002 | 178                     | 176               | 2             |
| 2003 | 241                     | 181               | 60            |
| 2004 | 243                     | 188               | 55            |
| 2005 | 358                     | 191               | 167           |
| 2006 | 304                     | 232               | 72            |
| 2007 | 364                     | 254               | 110           |
| 2008 | 405                     | 286               | 119           |
| 2009 | 368                     | 320               | 48            |
| 2010 | 376                     | 349               | 27            |
| 2011 | 219                     | 165               | 54            |
| 2012 | 69                      | 155               | 86            |
| 2013 | 124                     | 145               | 21            |
| 2014 | 89                      | 168               | 79            |
| 2015 | 95                      | 168               | 73            |
| 2016 | 91                      | 98                | 7             |
| 2017 | 110                     | 102               | 8             |
| 2018 | 92                      | 113               | 21            |
| 2019 | 105                     | 93                | 12            |
| 2020 | 164                     | 100               | 64            |

(3) Co-Integration and Error Correction Model (ECM).

The main theoretical basis of CI and ECM is that in many time series studies, the data fluctuation is not a stationary phenomenon, but a random process. In this method, difference methods (DM) are used to change the original unstable sequence into a stable one. For example, the equilibrium degree of short-term and long-term fluctuations is used to provide the model with higher prediction accuracy [52]. Therefore, this paper will verify the accuracy of the model again through this method, and the main analysis process is described below.
Step 1: The first step is to perform a unit root test on the actual value (variable A).

First of all, the trend chart is made for the actual value (variable A) data, as shown in Figure 6, from which the phenomenon of data containing the trend can be judged. Subsequently, augmented Dickey–Fuller (ADF) was used for testing. It can be seen from Table 6 that the insignificant \( p \)-Value in the original sequence test column (0.6869) means that the actual value (variable A) was non-stationary and there a unit root, so difference processing was required. Finally, the first-order difference sequence unit root test was performed on the actual sequence value (variable A). It can be seen from Table 6 that the \( p \)-Value in the column of the first-order difference sequence was significant (0.0430), which means that the sequence data of the actual value (variable A) was stationary. If the \( p \)-Value of the original sequence test was not significant, then the difference method would continue to process until the \( p \)-Value became significant.

![Figure 6. Trend diagram of real values (variable A).](image)

| Process                        | Level          | t-Statistic | Prob. * |
|--------------------------------|----------------|-------------|---------|
| Original sequence test         | ADFTS          | 1%          | −1.752625 | 0.6869  |
| Test critical values           |                |             | −4.532598 |         |
| First order difference sequence| ADFTS          | 1%          | −3.775465 | 0.0430 *|
| Test critical values           |                |             | −4.571559 |         |

Note 1: Augmented Dickey–Fuller Test Statistic (ADFTS). Note 2: \( * = p < 0.05 \).

Step 2: The second step is to perform unit root test on the theoretical predicted value (variable B).

This step was tested in the same way as the previous step, with a unit and test for the theoretically predicted value (variable B). Figure 7 shows that the data of the theoretically predicted value has a tendency, so the ADF test was carried out. The results of the ADF test are shown in Table 7. The insignificant \( p \)-Value (0.6280) in the sequence test column indicates that the theoretically predicted value (variable B) was a non-stationary phenomenon with a unit root, so differential processing was required. Finally, the first-order difference sequence unit root test was performed on the predicted value of sequence theory (variable B). It can be seen from Table 7 that the \( p \)-Value in the column of the first-order difference sequence
was significant (0.0245), which means that the sequence data of the theoretically predicted value (variable B) was a stationary phenomenon.

![Graph showing trend diagram of predicted value (variable B).](image)

**Figure 7.** Trend diagram of predicted value (variable B).

**Table 7.** Unit root test results of actual values (variable B).

| Process                      | ADFTS     | Level  | t-Statistic | Prob. * |
|------------------------------|-----------|--------|-------------|---------|
| Original sequence test       | ADFTS     | 1%     | −1.874321   | 0.6280  |
| Test critical values         |           |        | −4.532598   |         |
| First order difference sequence | ADFTS    | 1%     | −4.087372   | 0.0245 *|
| Test critical values         |           |        | −4.571559   |         |

Note 1: Augmented Dickey–Fuller Test Statistic (ADFTS). Note 2: * = p < 0.05.

Step 3: Third step is to test the stationarity of the residual sequence.

First, both the actual value and the theoretical prediction (variable A and variable B) are first-order differences, so a regression model can be established for co-integration analysis. Then, the least square method is used to estimate the regression model, and the residual sequence value can be obtained. Figure 8 shows the trend diagram of residual sequence values in each period. Finally, the unit root of residual error between the actual value and the theoretical value (variables A and B) was tested, and the results are shown in Table 8. It can be seen from Table 8 that the p-Value of ADF test result was significant (0.0000), which means that the residual sequence data were stable and the co-integration relationship between variables existed.
Step 4: Error correction model (ECM).

It is generated according to the first-order autoregressive distributed lag model (ECM), which describes the short-term fluctuation $\Delta y_t$ of the explained variable. $\left( \frac{\beta_0}{\beta_1 - 1} + y + \frac{\beta_2 + \beta_3}{\beta_1 - 1} x_t - 1 \right)$ is the error correction term, and $(\beta_1 - 1)$ is the coefficient of the error term, also known as the adjustment coefficient, which reflects the degree to which the short-term fluctuation of the variable deviates from the long-term equilibrium. Since there was a co-integration relationship between variables A and B, the ECM model could be established, and the residual sequence of the regression model obtained was the value of the error correction term ECM. We inputted the variables of the error correction term model, selected OLS for estimation, and obtained the error correction model results, as shown in Table 9. The coefficient of the error correction term ECM(−1) was estimated to be significant at the 20% test level, reflecting the extent to which the short-term error fluctuation of $-0.107$ deviates from the long-term equilibrium.

Table 9. Error correction model and equation estimation.

| Process          | Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|------------------|----------|-------------|------------|-------------|-------|
| EECM (short-term error level) | C        | -2.503389   | 12.14467   | -0.206131   | 0.8397|
|                  | D(INC02) | 0.359942    | 0.177777   | 2.024687    | 0.0624|
|                  | ECM(−1)  | -0.107066   | 0.331128   | -0.323337   | 0.7512|
| EECM (long-term error level) | C        | -3.844246   | 10.35557   | -0.371225   | 0.7151|
|                  | D(INC02) | 0.354652    | 0.156139   | 2.271378    | 0.0364|

Note: estimation of error correction model (EECM).

As shown in Table 9, the coefficient of short-term error correction term [ECM(−1)] verified in this study was $-0.107$, taking the absolute value, so the estimated value of test
error tended to 11%. Similarly, the coefficient with a long-term error level of \([D(INC02)]\) was 0.354652, so the estimated test error was close to 35%. In summary, according to Lin et al. [53], as long as the estimated short-term test error value is less than 30% and the estimated long-term test error value is less than 37%, the model is feasible. However, the long-term estimated value of this study was 35% and the short-term estimated value was 10%. Therefore, the model error of this study is feasible.

In summary, there are three error verification methods in this study, the first is the accuracy comparison of different prediction methods, the second is the method using the posterior error test, and the third is the co-integration and error correction model. The results of the above three kinds of test errors show that the prediction method proposed in this study is reliable in error precision determination.

We adopted three error verification methods. The first is the accuracy comparison of different prediction methods, the second is the method of posterior error test, and the third is the co-integration and error correction model. According to the analysis results, the first ensemble (bagging) method was more accurate, the second model was generally satisfied, and the third model had a short-term fluctuation of \(-0.10 < -0.30\). According to Lin et al. (2011), as long as \(<0.3\), the model is an adaptation. Therefore, the model error of this paper is acceptable.

4.3. Forecast the Future Development Trend

In this study, the validity of the proposed method was verified by the accuracy of the validation results, so the global car body patent was forecast for the next 10 years. It can be observed from the predicted results that the number of patents decreased year by year from 161 in 2020 to 144 in 2030, with a projected decline rate of 15.28% over the next 10 years (Figure 9).

Figure 9. Prediction of number of patents by ensemble learning bagging method.

From 161 cases in 2021 to 154 cases in 2024, and to 148 cases in 2027, a decline is predicted in the rate of every three years. The total number of cases will drop 15.28 percent to 144 cases by 2030. This shows that the product technology lifecycle moves from maturity period to decline period. The results indicate that the research and development of rigid strength of car body are in a mature stage. At present, the development patents of body patent technology mostly take body rigidity as the main research and development technology. Although the body rigidity is strengthened to meet the requirements of safety materials (such as steel plates as structural parts and cover parts), it also increases the load
of the car itself. It may cause problems in environmental protection and energy-saving that do not conform to economic effects and carbon emissions. Therefore, new and other technologies (such as energy-saving engines) should be developed at the same time to overcome such problems. In the research conclusion, this study will put forward future suggestions on the analysis results.

5. Conclusions

In this study, a new forecasting method, the “concept of technology maturity combined with machine learning model” was proposed to model and forecast the status quo of patented technology in the automobile industry. It is expected that the proposed method can provide decision makers and managers in the automotive industry with an accurate prediction of the future trend of new technologies or products in the industry market before investing in new technologies or product research and development, so as to reduce the significant losses of enterprises caused by patent disputes. In addition, this study was modeled and analyzed by a case study of body patent technology in China’s automobile industry, and the results prove that the proposed model has stability and applicability in patent prediction results. What is important is that this method can provide a systematic and scientific decision-making reference for decision makers or managers in the automotive industry to use when making new technology or product research and development plans, and bring value to academic and practical circles. Finally, according to the results of this study, suggestions for the industry and academia are put forward as follows.

First, our advice to the industry is as follows: According to the S-curve of the theory of inventive problem-solving theory (TIPS) [54], the technological evolution of products can be divided into infancy (initial stage), the growth stage, the maturity stage, and the decline stage. The lifecycle of the technical system can be judged according to the curve characteristics of the product.

According to the comparison of patent quantity and the S-shaped curve in Figure 10, patents in the whole region started between 1980 and 2030. From 1980 to 1990 is the beginning period. From 2001 to 2010, the number of cases increased from 209 to 376, belonging to the growth stage. From 2011 to 2020, the number of cases decreased from 219 to 164, belonging to the mature stage. The number of cases is expected to decrease from 164 to 144 between 2020 and 2030. The maturity period is characterized by a slight decrease in the number of patents every year, and a slight increase in performance parameters and the invention level, which belong to the first level. Therefore, it is judged that the maturity period of automobile body technology will gradually decline until 2030. The predicted value is the same as the PLC theory (S-shaped curve), so specific countermeasures are proposed for the future. It is suggested to follow the TIPS theory to improve the ideal degree law, improve the system parameters at the present stage, and reduce the production cost. Where possible, patents can cross-license patents to other companies and predict the patented technology and efficiency of other parts of the car. For example, for the front-body collision technology, the application of the technology route can plan the product technology blueprint and the layout of key patents in the next 5–10 years, so as to increase the company’s research and development competitiveness, and can also authorize patents to increase corporate profits.
Based on the results of the analysis, the load problems related to the current body rigidity design and research and the development objectives of the automotive industry in the next 10 years are described as follows.

(1) **Strategic objectives for short-term development (1–3 years):**

Starting from 2021, the short-term trend is 161–156 patents. It is suggested that aluminum, magnesium alloy, and fiber-reinforced composite materials should be selected appropriately for short-term strategic development. In terms of design, optimization design should be carried out according to material characteristics and performance requirements. Cold forming should be the main process, and hot forming, roll forming, and laser welding should be the minor methods. The short-term goal is to reduce body weight by 18%.

(2) **Strategic objectives of medium-term development (1–5 years):**

In the medium term, the number of patents went from 161 to 152. In the mid-term strategic development, the application of aluminum, magnesium alloy, and carbon fiber-reinforced composite materials in the car body should be expanded. Structured materials with performance-integrated lightweight multi objective collaborative optimization designs should be adopted. In terms of technology, hot forming, warm forming, and internal high-pressure forming should be the main processes, and extrusion forming, bending, and thermosetting fiber material should be supplementary processes. The mid-term goal is to reduce body weight by 30%.

(3) **Strategic objectives of long-term development (1–10 years):**

In the long term, the number of patents went from 161 to 144. In the long-term strategic development, the selection of materials should be mainly fiber composite materials, supplemented by light alloy and high strength steel. The design can be integrated with the requirements of the manufacturing process and cost control. In terms of technology, thermoplastic fiber material forming, extrusion forming, bending forming, warm forming, and hot forming should be considered supplementary. The long-term goal is to reduce body weight by 40%. Automobile body patent technology may have produced core technologies, so the number of patent applications will decrease. Hence, patent types will be mainly related to application methods in the future.

Second, our advice to academics is as follows: In this study, traditional statistical methods and machine learning were used to predict the number of patents, while contin-
uous quantification was also used to predict the number of patents. It is suggested that the application of text-based machine learning to patent analysis could be further studied. Secondly, in recent years, the growth of new energy vehicles has been substantial, and the technical problems encountered by the development of traditional fuel vehicles in the automobile body are worth studying.

**Author Contributions:** Conceptualization, H.-L.L.; methodology, H.-L.L. and C.-W.L.; formal analysis, C.-W.L.; writing—original draft preparation, C.-W.L.; writing—review and editing, H.-L.L.; and Y.-Y.M.; supervision, H.-L.L. and F.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

**References**

1. Author Group of China Automobile Industry Association. *Economic Operation of Automobile Industry in 2021, China*; China Automobile Industry Association: Beijing, China, 2022. Available online: http://www.caam.org.cn (accessed on 1 January 2020).
2. Vernon, R. The product cycle hypothesis in a new international environment. *Oxf. Bull. Econ. Stat.* 1979, 41, 255–267.
3. Pryshlakivsky, J.; Searcy, C. Life Cycle Assessment as a decision-making tool: Practitioner and managerial considerations. *J. Clean. Prod.* 2021, 309, 127344. [CrossRef]
4. Zhang, G.; Tang, C. How could firm’s internal R&D collaboration bring more innovation? *Technol. Forecast. Soc. Change* 2017, 125, 299–308.
5. Canuto da Silva, G.; Kaminski, P.C. Proposal of framework to managing the automobilemotive product development process. *Cogent Eng.* 2017, 4, 1317318.
6. Mousavi, S.A.; Seiti, H.; Hafezalkotob, A.; Asian, S.; Mobarraghi, R. Application of risk-based fuzzy decision support systems in new product development: An R-VIKOR approach. *Appl. Soft Comput.* 2021, 109, 107456. [CrossRef]
7. Fontem, B.; Smith, J. Analysis of a chance-constrained new product risk model with multiple customer classes. *Eur. J. Oper. Res.* 2019, 272, 999–1016. [CrossRef]
8. Adler, T.R.; Pittz, T.G.; Meredith, J. An analysis of risk sharing in strategic R&D and new product development projects. *Int. J. Proj. Manag.* 2016, 34, 914–922.
9. Chin, K.S.; Tang, D.W.; Yang, J.B.; Wong, S.Y.; Wang, H. Assessing new product development project risk by Bayesian network with a systematic probability generation methodology. *Expert Syst. Appl.* 2009, 36, 9879–9890.
10. Allayannis, G.; Ofek, E. Exchange rate exposure, hedging, and the use of foreign currency derivatives. *J. Int. Money Financ.* 2001, 20, 273–296. [CrossRef]
11. Guay, W.; Kothari, S.P. How much do firms hedge with derivatives? *J. Financ. Econ.* 2003, 70, 423–461. [CrossRef]
12. Gan, S.; Ge, S.; Han, X.; Yan, S.; Liu, J. 2020 China Patent Investigation Report; Intellectual Property Development Research Center of State Intellectual Property Office: Beijing, China, 2021.
13. Yuan, X.; Cai, Y. Forecasting the development trend of low emission vehicle technologies: Based on patent data. *Technol. Forecast. Soc. Change* 2021, 166, 120651. [CrossRef]
14. Hanggara, F.D. Forecasting Car Demand in Indonesia with Moving Average Method. *J. Eng. Sci. Technol. Manag. JES-TM* 2021, 1, 1–6.
15. Babai, M.Z.; Chen, H.; Syntetos, A.A.; Lengu, D. A compound-Poisson Bayesian approach for spare parts inventory forecasting. *Int. J. Prod. Econ.* 2021, 232, 107954. [CrossRef]
16. Wan, J.P.; Xie, L.Q.; Hu, X.F. Study on the electric vehicle sales forecast with TEI® I methodology. *Int. J. Knowl. Eng. Data Min.* 2021, 7, 1–38. [CrossRef]
17. Tsang, Y.P.; Wong, W.C.; Huang, G.Q.; Wu, C.H.; Kuo, Y.H.; Choy, K.L. A fuzzy-based product life cycle prediction for sustainable development in the electric vehicle industry. *Energies* 2020, 13, 3918. [CrossRef]
18. Lee, M. An analysis of the effects of artificial intelligence on electric vehicle technology innovation using patent data. *World Pat. Inf.* 2020, 63, 102002. [CrossRef]
19. Wang, X.; Zeng, D.; Dai, H.; Zhu, Y. Making the right business decision: Forecasting the binary NPD strategy in Chinese automotive industry with machine learning methods. *Technol. Forecast. Soc. Change* 2020, 155, 120032. [CrossRef]
20. Choi, Y.; Park, S.; Lee, S. Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data. *Scientometrics* 2021, 126, 5431–5476. [CrossRef]
21. Lee, C.; Kwon, O.; Kim, M.; Kwon, D. Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technol. Forecast. Soc. Change* **2018**, *127*, 291–303. [CrossRef]

22. Teng, F.; Sun, Y.; Chen, F.; Qin, A.; Zhang, Q. Technology opportunity discovery of proton exchange membrane fuel cells based on generative topographic mapping. *Technol. Forecast. Soc. Change* **2021**, *169*, 120859. [CrossRef]

23. Suhail, Y.; Upadhyay, M.; Chhibber, A. Machine learning for the diagnosis of orthodontic extractions: A computational analysis using ensemble learning. *Bioengineering* **2020**, *7*, 55. [CrossRef]

24. Barrera-Animas, A.Y.; Oyedele, L.O.; Bilal, M.; Akinosho, T.D.; Delgado, J.M.D.; Akanbi, L.A. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Mach. Learn. Appl.* **2022**, *7*, 100204. [CrossRef]

25. Ensaifi, Y.; Amin, S.H.; Zhang, G.; Shah, B. Time-series forecasting of seasonal items sales using machine learning—A comparative analysis. *Int. J. Inf. Inf. Manag. Data Insights* **2022**, *2*, 100058. [CrossRef]

26. Lin, S.; Lin, R.; Sun, J.; Wang, F.; Wu, W. Dynamically evaluating technological innovation efficiency of high-tech industry in China: Provincial, regional and industrial perspective. *Socio-Econ. Plan. Sci.* **2021**, *74*, 100939. [CrossRef]

27. Kim, J.E.; Cho, Y.S.; Kim, Y.R. The Major Common Technology Field Analysis of Domestic Mobile Carriers based on Patent Information Data. *J. Korea Acad.-Ind. Coop. Soc.* **2017**, *18*, 723–737.

28. Chen, Y.H.; Chen, C.Y.; Lee, S.C. Technology forecasting and patent strategy of hydrogen energy and fuel cell technologies. *Int. J. Hydrogen Energy* **2011**, *36*, 6957–6969. [CrossRef]

29. Altuntas, S.; Derecli, T. A Regression-Based “Patent Data Analysis” Approach: A Case Study for “Weapon Technology” Evaluation Process. *IEEE Trans. Eng. Manag.* **2021**, 1–13. [CrossRef]

30. Lin, H.L.; Ma, Y.Y. A New Method of Storage Management Based on ABC Classification: A Case Study in Chinese Supermarkets’ Distribution Center. *SAGE Open* **2021**, *11*, 2158244211023193. [CrossRef]

31. Lin, C.W.; Kao, C.S.; Chien, C.S. The Evaluation of Venture Capital in the Biotech Investment of Taiwan Rice-Bran Polysaccharide. *Adv. Manag. Appl. Econ.* **2020**, *10*, 151–172.

32. Lin, H.L.; Cho, C.C.; Ma, Y.Y.; Hu, Y.Q.; Yang, Z.H. Optimization plan for excess warehouse storage in e-commerce–based plant shops: A case study for Chinese plant industrial. *J. Bus. Econ. Manag.* **2019**, *20*, 897–919. [CrossRef]

33. Karakan, G.; Koc, T. Technology forecasting methods and an application to white goods sector. *J. Ind. Eng. Turk. Chamb. Mech. Eng.* **2008**, *20*, 29–38.

34. You, H.; Li, M.; Hipel, K.W.; Jiang, J.; Ge, B.; Duan, H. Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Sientometrics* **2017**, *111*, 297–315. [CrossRef]

35. Jose Basallo-Triana, M.; Rodriguez-Sarasty, J.A.; Dario Benitez-Restrepo, H. Analogue-based demand forecasting of short life-cycle products: A regression approach and a comprehensive assessment. *Int. J. Prod. Res.* **2016**, *55*, 2336–2350. [CrossRef]

36. Muthuraman, M.; Kwon, D.; Kim, W.; Cho, S.; Moon, K.; Park, J.; Kim, J.; Yoon, J. A novel approach to forecast promising technology through patent analysis. *Research and Technology Management in the Electricity Industry*; Springer: London, UK, 2013; pp. 189–211.

37. Ong, J.; Geum, Y. Automobilemated classification of patents: A topic modeling approach. *Comput. Ind. Eng.* **2020**, *147*, 106636. [CrossRef]

38. Altuntas, F.; Yilmaz, M. Patent analizi ile teknoloji aqlarının olusturulmas. *J. Entrep. Innov. Manag.* **2017**, *6*, 97–129.

39. Kyebambe, M.N.; Cheng, G.; Huang, Y.; He, C.; Zhang, Z. Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technol. Forecast. Soc. Change* **2017**, *125*, 236–244. [CrossRef]

40. Dutt, R.; Rathi, P.; Krishna, V. Novel mixed-encoding for forecasting patent grant duration. *World Pat. Inf.* **2021**, *64*, 102007. [CrossRef]

41. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Change* **2017**, *117*, 228–237. [CrossRef]

42. Cho, J.H.; Lee, J.; Sohn, S.Y. Predicting future technological convergence patterns based on machine learning using link prediction. *Sientometrics* **2021**, *126*, 5413–5429. [CrossRef]

43. Chang, P.C.; Wu, J.L.; Tsao, C.C.; Lin, M.H. A patent quality classification system using a kernel-pca with svm. In *Proceedings of the ADVCOMP 2015: The Ninth International Conference on Advanced Engineering Computing and Applications in Sciences*, Nice, France, 19–24 July 2015.

44. San Kim, T.; Sohn, S.Y. Machine-learning-based deep semantic analysis approach for forecasting new technology convergence. *Technol. Forecast. Soc. Change* **2020**, *157*, 120095. [CrossRef]

45. Lin, C.C.; Lin, C.L.; Shyu, J.Z. Hybrid multi-model forecasting system: A case study on display market. *Knowl.-Based Syst.* **2014**, *71*, 279–289. [CrossRef]

46. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

47. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In *Proceedings of the ICML’96: Proceedings of the Thirteenth International Conference on Machine Learning*, Bari, Italy, 3–6 July 1996; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1996.

48. Xiao, F.; Gao, X. The ring-shaped body structure design and evaluation method. In *Proceedings of the FSITA 2012 World Automotive Congress*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 447–461.

49. Julong, D. Introduction to grey system theory. *J. Grey Syst.* **1989**, *1*, 1–24.

50. Engle, R.F.; Granger, C.W.J. Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica* **1987**, *55*, 251–276. [CrossRef]
51. Davidson, J.E.; Hendry, D.F.; Srba, F.; Yeo, S. Econometric modelling of the aggregate time-series relationship between consumers’ expenditure and income in the United Kingdom. *Econ. J.* 1978, 88, 661–692. [CrossRef]

52. Duan, L.; Liu, Z.; Yu, W.; Chen, W.; Jin, D.; Li, D.; Sun, S.; Dai, R. Modeling Analysis and Comparision of Neural Network Simulation Based on ECM and LSTM. *J. Phys. Conf. Ser.* 2021, 2068, 012041. [CrossRef]

53. Lin, Y.; Chen, J.; Chen, Y. Dynamic relationships of knowledge creation activities in supply chains: Evidence from patent data in the US auto industry. *Afr. J. Bus. Manag.* 2011, 5, 12563–12576.

54. Altshuller, G.S.; Shapiro, R.B. On the psychology of inventive creativity. *Quest. Psychol.* 1956, 6, 37–49. (In Russian)