Machine Learning-Based Web System for Predicting and Classifying Financial Incentives in the Automotive Sector

Antony Jesus Ramirez Rivas, Rosalynn Ornella Flores-Castañeda
Facultad de Ingeniería y Arquitectura, Universidad César Vallejo, Lima, Perú

***Abstract*—This research presents the development of a web-based system using machine learning to predict and classify financial incentives in the automotive sector, contributing to Sustainable Development Goal 9 (Industry, Innovation and Infrastructure) and SDG 12 (Responsible Consumption and Production). The main objective was to design and implement an intelligent system that enhances decision-making regarding incentives such as exemptions (EXEM), natural gas subsidies (GNT), and tax benefits (TAX). The study employed a quantitative approach, applied type, and pre-experimental design, assessing model performance through accuracy, error rate, and response time metrics. Results showed an accuracy of 93.44%, a 45.12% reduction in error rate, and an average response time of 0.13 seconds. It is concluded that the proposed system significantly improves efficiency in predicting financial incentives, positioning itself as a viable technological tool for the automotive sector and economic sustainability.**

Keywords*—*Artificial intelligence; fiscal policy; forecasting; sustainable development; automobile

I. Introduction

Currently, the automotive sector is showing a growing preference for SUVs (Sport Utility Vehicles), which account for 80% of new car purchases in Canada, thanks to their association with safety, spaciousness, and prestige. However, buyers prefer smaller vehicles if they are provided with economic incentives or disincentives for SUVs, highlighting the urgency of implementing policies that promote more sustainable alternatives and support decarbonization goals [1].

Financial incentives provide benefits and economic support from manufacturers, dealers, or governments, intending to promote the purchase, financing, or marketing of an asset [2]. In addition, aggressive financial incentives for monthly targets can lead dealers to rush sales, especially at the end of the month, resulting in risky loans for low-income customers and higher default rates, without significantly affecting lenders [3]. For further understanding, a vehicle is the second most expensive purchase for a typical household, after housing, and most of these purchases are financed through loans or lease agreements managed directly by dealerships [4]. In addition [5], they noted that U.S. families who financed their vehicles through credit unions accessed lower interest rates compared to those who did so through banks, with an average difference of 0.70 percentage points on new cars and 1.40 on used cars.

In this study, the machine learning problem is defined as a supervised classification task, in which the system predicts the type of financial incentive for a vehicle based on its characteristics and the market context. The input variables include vehicle attributes and market conditions, while the output corresponds to the incentive category, enabling the accurate and automated classification of the incentives available in the automotive sector.

For this reason, the development of a web-based system based on machine learning was proposed as an alternative. As stated in [6], machine learning techniques make it possible to predict and classify data relating to the automotive sector, such as financial incentives. In this way, models can be applied to identify relevant patterns in automotive market data [7].

The fundamental purpose of the current study is to determine how a web-based machine learning system improves the prediction and classification of financial incentives in the automotive sector. Specific objectives include: (a) Determining how a web-based machine learning system improves the accuracy of predicting and classifying financial incentives in the automotive sector. (b) To determine how a web-based machine learning system reduces the error rate in the prediction and classification of financial incentives in the automotive sector. (c) To determine how a web-based machine learning system reduces the average response time in the prediction and classification of financial incentives in the automotive sector.

II. Related Work

On the other hand [8], they used a preprocessing and machine learning method to examine the predictive power of granular data on vehicle equipment in considering the resale price of used cars. Ultimately, they integrated records on specific equipment, thereby improving predictive performance by 3.27%, which allowed retailers to more accurately predict the values of used cars. As a result, predictions were improved, and retailers' financial returns were increased by better adjusting resale prices.

As noted [9], they conducted a study in which they predicted the demand for electric vehicles (EVs) by applying long-term and short-term memory (LSTM) models and ConvLSTM convolutional networks. Regarding the findings, they indicated that the hybrid model outperformed the other models in key measures such as mean absolute error, normalized root mean square error, and R-squared. Finally, the model was able to predict the market share of EVs with a passable mean absolute error of 3.5%, which represents a significant improvement in the accuracy of predictions.

From the perspective of [10], they used learning models, such as XGBoost and Random Forest, to predict the value and duration of car sales in secondary markets, using a database of more than 5,386,116 cars. In this regard, the researchers acknowledged that XGBoost outperformed other models in terms of accuracy, exceeding a value of r2=0.94 in price forecasting. Finally, the model manages to boost vehicle sales, reducing risks for sellers and providing excellent information to buyers.

According to [11], the Croatian used car market was analyzed using supervised learning to predict prices. The study collected data from the "Njuškalo" portal, considering attributes such as year, mileage, and manufacturer. After cleaning out outliers, linear regression, Random Forest, and SVM models were trained. The results showed that linear regression achieved the highest accuracy (R^2 = 0.95), outperforming other algorithms. In addition, an average increase of €1,391 was detected over three months, validating the upward trend. In summary, the study demonstrates that data mining and supervised models optimize price estimation and reduce transaction risks.

For [12], they patented an electric car sales prediction model based on machine learning algorithms and the partial least squares method, using environmental, economic, and human development factors from 31 countries. The algorithms are shown to achieve high levels of accuracy and predictive performance, as well as recognizing factors that have a positive relationship with electric vehicle sales, such as CO_2 emissions, PM2.5, the consumer price index, renewable energy, and life expectancy. They concluded that their model can be used by governments globally as a tool to support the formulation of policies that encourage the adoption of electric vehicles.

III. Methodology

A. Phase 1: Requirements Analysis

The model applied is Random Forest, due to its ability to handle categorical data and its good performance in classification tasks. For this reason, three categories of financial incentives were used: EXEM (tax exemptions), GNT (subsidies or grants), and TAX (tax incentives such as deductions, credits, or preferential rates).

- EXEM: refers to the legal exclusion of certain bases, income, or activities from tax calculations, intending to reduce the tax burden on beneficiaries or promote specific activities [13].
- GNT: refers to non-repayable transfers to support specific activities, with no obligation to repay, commonly used in innovation efficiency studies [14].
- TAX: corresponds to a legal measure granted by the government that alters the standard tax regime for certain individuals, investments, or specific activities to reduce their tax burden [15].

Table I details the essential functional requirements for the successful development and implementation of the prediction and classification system using machine learning.

TABLE I. TABLE OF FUNCTIONAL REQUIREMENTS

Code	Requirement	Description
RF001	User Management	User registration and authentication, including login form.
RF002	Incentive Program Management	Create, edit, and delete programs to predict incentives; display on the web interface.
RF003	Prediction and Classification	Integration of the machine learning model to predict incentives and display results.
RF004	Web Interface	Main dashboard with a list of incentives and forms for entering data.
RF005	Background Processing	Use of Celery and Redis for asynchronous tasks (execution of the ML model).
RF006	REST API	Django backend to expose endpoints that manage data and predictions.
RF007	Data Registration	User registration and authentication, including a login form.

Table II shows the non-functional requirements.

TABLE II. TABLE OF NO-FUNCTIONAL REQUIREMENTS

Code	Requirement	Description
RF001	Scalability	Use of Docker containers to enable modular deployment.
RF002	Availability and persistence	PostgreSQL database for secure storage.to predict incentives; display on the web interface.
RF003	Performance	Efficient processing through asynchronous tasks with Celery and Redis.
RF004	Security	Environment variables for credentials (DB_NAME, DB_USER, DB_PASSWORD, SECRET_KEY) and private network (app-network).
RF005	Maintainability	Code organized into modules (backend, frontend, worker) and use of Docker Compose for reproducibility.
RF006	Portability	System executable in any environment with Docker.

B. Phase 2: System Design

1) Definition of the software architecture technologies:

- Python and Django (Backend)
- Vue.js (Front-End)
- Docker (Deployment)
- Pandas (for history management and export to Excel)

Fig. 1 shows the final interface design.

Fig. 1. Final design of the system when detecting glass.

2) Methodological architecture: For the development of the proposed web system, the waterfall development methodology was used, which is characterized by a sequential approach in which each phase must be completed before starting the next one. This methodology was selected for its structural clarity and for facilitating project planning and documentation. In addition, the logical and technical architecture of the system is defined, ensuring that the components interact correctly to meet functional and non-functional requirements.

Fig. 2 shows the methodological architecture.

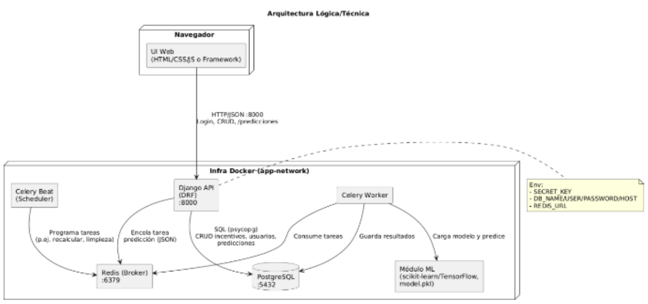

Fig. 2. Technological architecture for system development.

IV. RESULTS

This chapter presents the results obtained from training and evaluating the machine learning model applied to the web system. The dataset consisted of 5,000 financial records linked to financial incentives, extracted from the U.S. government's AFDS (Alternative Fuels Data Center). Of these, 3,500 were used for training, 750 for validation, 500 for testing, and 250 for the system's post-test, representing 75% for training, 15% for validation, and 15% for testing. The model used was a Random Forest Classifier, selected for its ability to handle categorical data and its strong performance in classification tasks. Based on these results, the hypotheses proposed in the research were subsequently tested.

A. Pre-Test

Table III shows the results obtained by [16], who evaluated the performance of nine classification models applied to loan data. This table presents the accuracy, recall, F1-score, and error rate metrics, calculated as ER = 1 − Accuracy.

TABLE III. RESULTS OF CLASSIFICATION MODELS

Models	Accuracy	Error	Recall	F1-Score	AUC
Random Forest	88.05	11.95	79.99	0.62	0.93
Stacking	86.76	13.24	82.8	0.61	0.93
Decision Tree	84.77	15.23	87.41	0.59	0.85
KNN	83.19	16.81	74.38	0.52	0.85
XGBoost	82.45	17.55	81.5	0.53	0.87
MLP	69.18	30.82	44.44	0.26	0.63
SVM	55.83	44.17	51.22	0.22	0.56
Logistic Regression	51.35	48.65	57.04	0.22	0.55
Gaussian NB	31.84	68.16	80.63	0.23	0.55

Table IV shows the average response time results (in milliseconds) reported by [17], with the proposed model (A) performing best with an average of 139.8 ms, followed by CNN (151 ms), Decision Tree (422.6 ms), and SVM (366.6 ms).

- Model A: Long Short-Term Memory (LSTM)
- Model B: Convolutional Neural Network (CNN).
- Model C: Decision Tree.
- Model D: Support Vector Machine (SVM).

TABLE IV. AVERAGE RESPONSE TIME RESULTS

Test		Time (ms)					
		1	2	3	4	5	Average
1	A	102	158	183	147	109	139.8
	B	234	327	512	323	447	368.6
	C	397	399	419	549	447	442.2
	D	359	439	464	371	480	422.6
2	A	151	164	195	111	134	151
	B	458	579	584	486	529	527.2
	C	647	854	354	264	367	497.2
	D	547	347	316	764	345	463.8
3	A	155	170	106	124	148	140.6
	B	564	477	594	524	356	503
	C	456	378	399	576	405	442.8
	D	834	456	347	345	457	487.8

B. Post-Test

Tables V, VI and VII show the metrics by class (EXEM, GNT, and TAX), including Recall, F1-score, as well as overall accuracy, overall error rate, and average processing time.

TABLE V. POST-TEST INDICATORS RESULTS

C	TP	FP	FN	TN	Recall	F1	Accuracy	Error	Time
EXEM	13	2	2	44	0.87	0.87	93.44%	6.56%	0.13
GNT	36	8	0	17	1.00	0.90	86.89%	13.11%	0.13
TAX	2	0	8	51	0.20	0.33	86.89%	13.11%	0.13

TABLE VI. COMPARISON OF RESULTS

Indicator	Random Forest	Web System
Accuracy (%)	88.05%	93.44%
Error (%)	11.95%	6.56%

TABLE VII. COMPARISON OF AVERAGE RESPONSE TIME

Indicador	LSTM	Web System
Average Response Time	0.14	0.13

C. Hypotheses

The results confirm a positive improvement of 6.12% in the overall accuracy of the proposed model compared to the baseline model. This analysis shows that the machine learning-based web system improves the accuracy of financial incentive classification, accepting the alternative hypothesis (HE_{11}).

The results indicate a 45.12% reduction in the total error rate of the proposed system compared to the baseline model. This demonstrates that the system significantly reduces errors in the classification of financial incentives, thus supporting the alternative hypothesis (HE_{21}).

The results show a 7.22% reduction in average response time, demonstrating an improvement in the system's temporal performance. This finding indicates that the machine learning-based web system contributes to optimizing the process of predicting and classifying financial incentives, supporting the alternative hypothesis (HE_{31}).

HG_0: A web-based machine learning system does not improve the prediction and classification of financial incentives. HG_1: A web-based machine learning system improves the prediction and classification of financial incentives. By verifying the specific hypotheses HE1, HE2, and HE3, the general hypothesis is accepted.

V. Discussion

The first specific objective was to determine how a web system based on machine learning improves accuracy in the prediction and classification of financial incentives in the automotive sector. Accordingly, the post-test yielded an overall accuracy of 93.44%, representing a relative increase of 6.12% compared to the Random Forest model. This indicates that, within the evaluated dataset, the proposed system achieves a higher percentage of correct global predictions, suggesting a real improvement in the classifier's ability to distinguish between incentives that are granted and those that are not, in contrast to [11], interesting comparisons were observed when applying a Logistic Regression model to predict prices of used cars, which reported an accuracy of 83.08% on the test set after data cleaning and selection of relevant attributes. Therefore, compared to this precedent, the proposed system surpasses it by approximately 10 percentage points, which can be attributed to the more domain-specific design of the automotive–financial dataset, improved feature engineering, and more careful class balancing.

These comparisons suggest that the achieved accuracy is not an isolated value but is aligned with the performance standards of robust models in related domains. Thus, the superiority over the used-vehicle study may be due to the more specialized nature of the dataset (financial incentives) and more rigorous preprocessing. Likewise, its closeness to the reported parking system performance shows that, under adequate data conditions, Random Forest produces high-level accuracy scores.

However, it must be acknowledged that the conditions of each study differ: the nature of the dataset, variable quality, class balancing, noise levels, and number of instances can strongly influence the results. Therefore, the 93.44% accuracy achieved in this work is interpreted as a very strong performance within the investigated context, with comparative merit relative to the reviewed literature.

Regarding the second specific objective, the goal was to determine how the system reduces the error rate in predicting and classifying financial incentives. The global error rate dropped from 11.95% (baseline model) to 6.56% in the proposed system, representing a significant absolute and relative reduction (≈ 45.12%). This decline confirms that the system not only increases correct predictions (accuracy) but also reduces the relative number of incorrect predictions.

In contrast, studies on financial risk or bankruptcy often involve more heterogeneous datasets with highly imbalanced classes, which tend to increase prediction errors, especially in minority classes.

On the other hand, the third specific objective evaluated whether the web system reduces the average response time in predicting and classifying incentives. The system's average response time was 0.13s compared to 0.14s reported by the optimized LSTM model in the literature (≈ 7.22% relative improvement). Although the absolute difference is small (0.01s), its consistency suggests an operational advantage for the Random Forest model under equivalent evaluation conditions.

Furthermore, the reduction in latency indicates that the proposed model does not sacrifice efficiency for accuracy or error reduction; rather, it achieves a very good balance between precision and speed. In web applications, even small differences in response time can noticeably impact user experience or continuous interaction flows.

According to [15], inference times were compared across several models: Random Forest, XGBoost, LSTM, and a hybrid model (TANEA). They found that LSTM has an inference time of 7.8 MS per sample, while Random Forest achieves approximately 3.0 Ms. This shows that tree-based models can offer much lower latency than deep recurrent neural networks.

Similarly, [16] investigated inference times of neural networks on mobile GPUs, noting substantial latency differences depending on the architecture and model size, especially for lightweight models. Although specific times vary, latencies above 0.1 seconds were reported for heavier models, placing the system's 0.13s latency as competitive when compared with deep models optimized for mobile devices.

Thus, the low response time can be attributed to Random Forest relying on simple comparisons and summations, which demand less computational load on CPU architectures, unlike neural networks that require heavier matrix multiplications. Additionally, well-designed preprocessing and variable encoding can reduce latency by minimizing transformations during inference. However, the 0.01 difference may not be significant in environments with specialized hardware or under concurrent load, meaning its advantage could diminish in more demanding real-world settings.

The general objective of determining how a web system based on machine learning improves the prediction and classification of financial incentives in the automotive sector is supported by the results: an increase in global accuracy

(93.44%), a substantial reduction in error rate (45.12% relative), and a slight improvement in latency (0.13s vs. 0.14s).

These discussions are consistent with the referenced studies, as they demonstrate the strong performance of ensemble techniques in automotive tabular problems and in research on incentive optimization and predictive modeling in the sector. They also show robust performance across different environments, such as FinTech, which requires the ability to handle complex data and improve predictions of financial behavior [18]. For example, [19] developed a model combining Random Forest and Deep Learning to analyze purchase intentions for hybrid vehicles, considering financial incentives as a key variable in consumer decision-making, and achieved high levels of accuracy and stability when classifying factors associated with automotive incentive policies. Likewise, [20] state that various financial incentive programs applied to hybrid and automated vehicles significantly stimulate technological adoption and reduce barriers to market entry.

Thus, the studies examined in the discussion reinforce the validity of the results obtained, demonstrating that combining machine learning models with incentive schemes constitutes an effective strategy to optimize prediction, classification, and the design of financial policies in the automotive sector.

VI. Conclusion

The results confirm that the machine learning-based web system improves the prediction and classification of financial incentives in the automotive sector. Empirically, the optimized model achieved an overall accuracy of over 93%, significantly outperforming the baseline model, and reduced the error rate from 11.95% to 6.56%, indicating a notable improvement in the reliability of predictions. Furthermore, the average response time was reduced from 0.14 s to 0.13 s, demonstrating the system's efficiency in a web environment. These findings reflect the effectiveness of the Random Forest algorithm, and the preprocessing and validation techniques applied.

From a broader perspective, these results suggest that an automated financial incentive classification system can support decision-making at dealerships and in automotive marketing strategies by providing more accurate and timely information on available benefits. However, while the results are promising, the implications for commercial policies and financial strategies should be interpreted with caution, as they depend on data availability and the specific context of each market.

References

[1] J. Axsen and Z. Long, "Sport utility vehicles and willingness-to-downsize: A mixed-method exploration of functional, symbolic, and societal consumer perceptions in Canada," Energy Res. Soc. Sci., vol. 92, p. 102776, Oct. 2022, doi: 10.1016/J.ERSS.2022.102776.

[2] S. Sasaki, T. Kubo, and S. Kitano, "Prosocial and financial incentives for biodiversity conservation: A field experiment using a smartphone app," Ecological Economics, vol. 230, p. 108506, Apr. 2025, doi: 10.1016/J.ECOLECON.2024.108506.

[3] M. Jansen, L. Pierce, J. Snyder, and H. Nguyen, "Product Sales Incentive Spillovers to the Lending Market: Evidence from Subprime Auto Loan Defaults," Manage. Sci., vol. 70, no. 8, pp. 5463–5480, Aug. 2024, doi: 10.1287/mnsc.2023.4935.

[4] C. Latino, L. Pelizzon, and M. Riedel, "How to green the European auto ABS market? A literature survey," European Financial Management, vol. 31, no. 2, pp. 615–638, Mar. 2024, doi: 10.1111/EUFM.12515;Journal:Journal:1468036X;Wgroup:String:Publication.

[5] J. van Rijn, S. Zeng, and P. Hellman, "Financial institution objectives and auto loan pricing: Evidence from the survey of consumer finances," Journal of Consumer Affairs, vol. 55, no. 3, pp. 995–1039, Sep. 2021, doi: 10.1111/JOCA.12392;JOURNAL:JOURNAL:17456606.

[6] M. Hanafy and R. Ming, "Machine Learning Approaches for Auto Insurance Big Data," Risks 2021, Vol. 9, Page 42, vol. 9, no. 2, p. 42, Feb. 2021, doi: 10.3390/RISKS9020042.

[7] F. Miconi and G. M. Dimitri, "A machine learning approach to analyse and predict the electric cars scenario: The Italian case," PLoS One, vol. 18, no. 1, p. e0279040, Jan. 2023, doi: 10.1371/JOURNAL.PONE.0279040.

[8] S. Bergmann and S. Feuerriegel, "Machine learning for predicting used car resale prices using granular vehicle equipment information," Expert Syst. Appl., vol. 263, p. 125640, Mar. 2025, doi: 10.1016/j.eswa.2024.125640.

[9] S. Afandizadeh, D. Sharifi, N. Kalantari, and H. Mirzahossein, "Using machine learning methods to predict electric vehicles penetration in the automotive market," Sci. Rep., vol. 13, no. 1, p. 8345, Dec. 2023, doi: 10.1038/S41598-023-35366-3.

[10] V. Shymanskyi and V. Liaskovets, "Cascade Model for Price and Time of Car Sales Prediction," CEUR Workshop Proc., vol. 3641, pp. 152–167, Jan. 2023, Accessed: May 06, 2025. [Online]. Available: https://www.scopus.com/record/display.uri?eid=2-s2.0-85185886742&origin=resultslist&sort=plf-f&src=s&sot=b&sdt=b&cluster=scoexactkeywords%2C%22Automotive+Sector%22%2Ct%2C%22Automotive+Industry%22%2Ct&s=TITLE-ABS-KEY%28sales+car%29&sessionSearchId=66e5f27fda5dc8e9a38ac4cd7a80df29&relpos=66

[11] L. Bukvić, J. Pašagić Škrinjar, T. Fratrović, and B. Abramović, "Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning," Sustainability 2022, Vol. 14, Page 17034, vol. 14, no. 24, p. 17034, Dec. 2022, doi: 10.3390/SU142417034.

[12] J.-Y. Yeh and Y.-T. Wang, "A Prediction Model for Electric Vehicle Sales Using Machine Learning Approaches," Journal of Global Information Management, vol. 31, no. 1, doi: 10.4018/JGIM.327277.

[13] J. J. Castro Maldonado, L. K. Gómez Macho, and E. Camargo Casallas, "La investigación aplicada y el desarrollo experimental en el fortalecimiento de las competencias de la sociedad del siglo XXI," Tecnura, vol. 27, no. 75, pp. 140–174, Jan. 2023, doi: 10.14483/22487638.19171.

[14] I. Lishner and A. Shtub, "Enhancing Strategic Planning of Projects: Selecting the Right Product Development Methodology," Information 2023, Vol. 14, Page 632, vol. 14, no. 12, p. 632, Nov. 2023, doi: 10.3390/INFO14120632.

[15] D. B. J. Rao, P. Kalpana, G. S. N. Kumar, and N. M. Atcha, "A Comparative Analysis of Software Development Models: Waterfall, Agile and DevOps," Smart Innovation, Systems and Technologies, vol. 417, pp. 589–599, Jan. 2025, doi: 10.1007/978-981-97-8355-7_51.

[16] P. Givord, R. Rathelot, and P. Sillard, "Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines program," Reg. Sci. Urban Econ., vol. 43, no. 1, pp. 151–163, Jan. 2013, doi: 10.1016/J.REGSCIURBECO.2012.06.006.

[17] B. Qian, "Financial subsidies, tax incentives, and new energy vehicle enterprises' innovation efficiency: Evidence from Chinese listed enterprises," PLoS One, vol. 18, no. 10, p. e0293117, Oct. 2023, doi: 10.1371/JOURNAL.PONE.0293117.

[18] H.; Wang, J. ; Yang, N. Zhu, H. Wang, J. Yang, and N. Zhu, "Does Tax Incentives Matter to Enterprises' Green Technology Innovation? The Mediating Role on R&D Investment," Sustainability 2024, Vol. 16, Page 5902, vol. 16, no. 14, p. 5902, Jul. 2024, doi: 10.3390/SU16145902.

[19] H. Li and W. Wu, "Loan default predictability with explainable machine learning," Financ. Res. Lett., vol. 60, p. 104867, Feb. 2024, doi: 10.1016/J.FRL.2023.104867.

[20] Q. Wang, Z. Chen, Y. Zhou, Z. Liu, and Z. Peng, "Real-time monitoring and optimization of machine learning intelligent control system in power data modeling technology," Machine Learning with Applications, vol. 18, p. 100584, Dec. 2024, doi: 10.1016/J.MLWA.2024.100584.