

ФОРМУВАННЯ СТРУКТУРИ ВХІДНИХ ДАНИХ ДЛЯ БАЗОВОГО ПРОГНОЗУ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ

Вінницький національний технічний університет

Анотація

У роботі розглянуто особливості формування структури вхідних даних для базового прогнозу вартості вживаних автомобілів. Проаналізовано доцільність використання ринкових даних з онлайн-оголошень, технічних характеристик транспортних засобів та їх поєднання в межах комбінованого підходу. Визначено, що ринкові оголошення дозволяють врахувати актуальну ситуацію на вторинному автомобільному ринку, однак потребують очищення та перевірки релевантності. Технічні характеристики дають змогу точніше описати конкретний автомобіль, але без ринкової складової не відображають поточну цінову ситуацію. Обґрунтовано, що найбільш доцільним є використання комбінованої структури даних, яка поєднує ринкові та технічні параметри й може бути використана як основа для подальшого прогнозування вартості.

Ключові слова: вживані автомобілі, вхідні дані, базовий прогноз, прогнозування вартості, ринкові оголошення, технічні характеристики, комбінована структура даних.

Abstract

The paper considers the features of forming the input data structure for a baseline forecast of used car value. The feasibility of using market data from online listings, vehicle technical characteristics, and their combination within a combined approach is analyzed. It is determined that market listings make it possible to take into account the current situation in the used car market, but require cleaning and relevance checking. Technical characteristics allow a more accurate description of a specific vehicle, but without the market component they do not fully reflect the current price situation. It is substantiated that the most appropriate solution is the use of a combined data structure that includes both market and technical parameters and can be used as a basis for further value forecasting.

Keywords: used cars, input data, baseline forecast, value forecasting, market listings, technical characteristics, combined data structure.

Вступ

Існують різні підходи до формування структури вхідних даних для базового прогнозу вартості вживаних автомобілів, що відрізняються складом, точністю та сферою використання. Одні підходи ґрунтуються переважно на аналізі ринкових оголошень, інші — на технічних характеристиках транспортного засобу, а також існують комбіновані підходи, що поєднують обидва джерела інформації. У роботі розглядаються основні групи вхідних даних, що можуть використовуватися для формування базового прогнозу вартості вживаних автомобілів, а також аналізується доцільність їх поєднання в межах комбінованого підходу.

Результати дослідження

У межах задачі формування базового прогнозу вартості вживаних автомобілів доцільно виділити кілька основних підходів. Найпоширенішими серед них є формування оцінки на основі ринкових аналогів, прогнозування за технічними характеристиками транспортного засобу та комбінований підхід, який поєднує обидва джерела інформації. В такому випадку ринок вживаних автомобілів характеризується неоднорідністю, а сама вартість залежить як від поточного стану ринку, так і від індивідуальних параметрів конкретного автомобіля [1].

Для цього доцільно розглядати дві основні групи вхідних даних: ринкові дані з оголошень та технічні характеристики автомобіля. Ринкові дані відображають актуальні ціни, пропозицію та стан вторинного ринку, тоді як технічні характеристики описують індивідуальні параметри конкретного транспортного засобу. Поєднання цих груп даних дозволяє сформуванню більш повну інформаційну основу для подальшого прогнозування.

Перша група даних пов'язана з використанням цін подібних автомобілів, представлених у ринкових оголошеннях. Його основою є ринковий підхід до оцінювання, за яким вартість об'єкта визначається шляхом порівняння з ідентичними або подібними, для яких доступна ринкова інформація. Ринковий підхід є доцільним за наявності достатньої та релевантної інформації про порівнювані об'єкти. Неточні, неочищені та відсутність даних по моделі авто роблять результати програми менш надійними [1, 2].

	ad_id [PK] bigint	brand text	model text	title text	price_usd integer	year integer	mileage_km integer	region text	engine_l numeric (4,1)	fuel text	transmission text
1	24413934	Ford	Focus	Ford Focus	7500	2013	130000	Київ	1.0	Бенз	Мех
2	24941794	Ford	Focus	Ford Focus	12500	2017	116000	Чернівці	1.5	Диз	Мех
3	27557198	Ford	Focus	Ford Focus	4300	2005	220000	Кривий Ріг	[null]	Бенз	Мех
4	28372669	Ford	Focus	Ford Focus SE 2.0 Flex Fuel	9800	2016	95000	Павлоград	2.0	Бенз	Роб
5	29665649	Ford	Focus	Ford Focus	8900	2013	205000	Київ	1.6	Диз	Мех
6	29899015	Ford	Focus	Ford Focus	8888	2017	148000	Київ	2.0	Бенз	Не вказано
7	30048941	Ford	Focus	Ford Focus	5300	2008	200000	Вінниця	[null]	Не вказано	Не вказано
8	30545806	Ford	Focus	Ford Focus	5600	2006	206000	Васильків	1.6	Бенз	Мех
9	30852878	Ford	Focus	Ford Focus	11000	2017	100000	Луцьк	2.0	Бенз	Авт
10	31337075	Ford	Focus	Ford Focus	9800	2016	71000	Одеса	[null]	Електро	Авт
11	32101510	Ford	Focus	Ford Focus	3999	2002	314000	Запоріжжя	1.8	Диз	Не вказано
12	32295238	Ford	Focus	Ford Focus ST	11900	2018	87000	Кривий Ріг	2.0	Бенз	Мех
13	32311019	Ford	Focus	Ford Focus	8250	2016	99000	Київ	2.0	Бенз	Авт
14	32344740	Ford	Focus	Ford Focus	7800	2011	302000	Луцьк	1.6	Диз	Мех
15	32654180	Ford	Focus	Ford Focus	6800	2014	276000	Чернігів	1.6	Диз	Мех

Рисунок 1 – Приклад оголошень Ford Focus у базі даних

Переваги підходу:

- спирається на поточну ринкову ситуацію та відображає актуальні цінові тенденції;
- є відносно простим для практичного застосування, особливо для поширених моделей із великою кількістю аналогів;
- дозволяє використовувати реальні дані з онлайн-майданчиків продажу автомобілів.

Недоліки:

- точність залежить від якості оголошень, наявності дублікатів, викидів і нерелевантних пропозицій;
- подібні оголошення можуть суттєво відрізнитися за технічним станом, комплектацією та модифікацією, що ускладнює пряме порівняння;
- підхід гірше працює для рідкісних автомобілів і нетипових комплектацій, де бракує коректних аналогів.

Друга група даних базується на аналізі параметрів самого транспортного засобу. У такому випадку вихідна вартість визначається за ознаками, які описують конкретний автомобіль: марка, модель, рік випуску, пробіг, тип кузова, трансмісія, тип пального, об'єм двигуна та інші характеристики. Найбільш важливими ознаками для прогнозування ціни вживаного автомобіля є бренд, модель, рік, трансмісія, тип кузова, тип пального, об'єм двигуна та пробіг. Задача прогнозування вартості вживаних авто є багатфакторною, а результат суттєво залежить від вибору методу регресії [2, 3].

	car_id [PK] integer	mod [PK] text	enginemodel [PK] text	gearbox [PK] text	power [PK] double precision	yearstart [PK] integer
47	1903	1.0 EcoBoost (125 Hp) Automatic	B7DA	8 gears, automatic transmission	125	2018
48	1903	1.0 EcoBoost (125 Hp) Mild Hybrid	B7DC	6 gears, manual transmission	125	2020
49	1903	1.0 EcoBoost (155 Hp) Mild Hybrid	M0DC	6 gears, manual transmission	155	2020
50	1903	1.0 EcoBoost (85 Hp)	SFDC	6 gears, manual transmission	85	2018
51	1903	1.5 EcoBlue (120 Hp)	ZTDB	6 gears, manual transmission	120	2018
52	1903	1.5 EcoBlue (120 Hp) Automatic	ZTDB	8 gears, automatic transmission	120	2018
53	1903	1.5 EcoBlue (95 Hp)	ZTDB	6 gears, manual transmission	95	2018
54	1903	1.5 EcoBoost (150 Hp)	YZDA	6 gears, manual transmission	150	2018
55	1903	1.5 EcoBoost (150 Hp) Automatic	YZDA	8 gears, automatic transmission	150	2018
56	1903	1.5 EcoBoost (182 Hp)	Y1DA	6 gears, manual transmission	182	2018
57	1903	2.0 EcoBlue (150 Hp)	YLDA	6 gears, manual transmission	150	2018
58	1903	2.0 EcoBlue (150 Hp) Automatic	YLDA	8 gears, automatic transmission	150	2018
59	1903	ST 2.0 EcoBlue (190 Hp)	YMDA	6 gears, manual transmission	190	2019
60	1903	ST 2.3 EcoBoost (280 Hp)	N3DA	6 gears, manual transmission	280	2019
61	1903	ST 2.3 EcoBoost (280 Hp) Automatic	N3DA	7 gears, automatic transmission	280	2019

Рисунок 2 – Приклад технічних характеристик двигунів Ford Focus у базі даних

Переваги:

- дозволяє врахувати індивідуальні особливості конкретного автомобіля;
- краще підходить для аналізу різних модифікацій і комплектацій, ніж просте усереднення ринкових аналогів;
- дає можливість формувати оцінку навіть тоді, коли кількість релевантних оголошень є обмеженою.

Недоліки методу:

- без ринкової складової майже неможливо відобразити поточну кон'юнктуру ринку;
- залежність ціни від характеристик є складною та часто нелінійною, через що прості моделі можуть давати занижену точність.

Комбінована структура даних поєднує ринкові оголошення та технічні характеристики автомобіля.

Ідея полягає в тому, що ринкова інформація використовується як основа для визначення стартової вартості, а технічні параметри – як засіб її уточнення. Це є найбільш обґрунтованим у задачах, де важливо одночасно врахувати актуальний стан ринку та індивідуальні особливості транспортного засобу. Включення детальної інформації про комплектацію та оснащення автомобіля зменшує похибку прогнозування МАЕ на 3,27 %, тобто додаткові характеристики реально підвищують якість передбачення вартості [4].

Переваги підходу:

- поєднання актуальності ринкових даних і точність технічних характеристик;
- краще адаптується до відмінностей між модифікаціями та комплектаціями автомобілів;
- забезпечує більш обґрунтовану стартову оцінку для подальшого прогнозування зміни вартості.

Недоліки підходу:

- потребує більш складної підготовки та узгодження даних з різних джерел;
- вимагає очищення ринкових оголошень і коректного кодування технічних ознак;
- є складнішим у реалізації порівняно з окремим використанням лише одного джерела інформації.

Отже, проведений аналіз показує, що підхід на основі ринкових аналогів є простим і добре відображає поточну ринкову ситуацію, але чутливий до неоднорідності оголошень. Підхід на основі технічних характеристик краще враховує параметри конкретного автомобіля, однак без ринкової складової може втрачати актуальність. Найбільш доцільною структурою даних для базового прогнозу вартості вживаних автомобілів є комбінований підхід, оскільки він поєднує переваги двох попередніх і забезпечує більш обґрунтовану початкову оцінку, хоча й вимагає більших та чітких даних.

Висновки

Формування базового прогнозу вартості вживаних автомобілів доцільно здійснювати на основі структурованих вхідних даних, що поєднують ринкову та технічну інформацію. Ринкові оголошення дозволяють врахувати актуальну ситуацію на вторинному автомобільному ринку, однак потребують очищення, перевірки релевантності та усунення неточних пропозицій. Технічні характеристики автомобіля дають змогу деталізувати оцінку з урахуванням марки, моделі, року випуску, пробігу, типу кузова, двигуна, трансмісії та інших параметрів. Найбільш обґрунтованою є комбінована структура даних, оскільки вона поєднує актуальність ринкової інформації з точністю опису конкретного транспортного засобу та створює надійну основу для подальшого прогнозування зміни вартості.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. International Valuation Standards Council. IVS 105: Valuation Approaches and Methods. 2016.
2. Amik F. R., Lanard A., Ismat A., Momen S. Application of Machine Learning Techniques to Predict the Price of Pre-Owned Cars in Bangladesh. Information. 2021.
3. Lessmann S., Voß S. Car Resale Price Forecasting: The Impact of Regression Method, Private Information, and Heterogeneity on Forecast Accuracy. International Journal of Forecasting. 2017.
4. Bergmann S., Feuerriegel S. Machine Learning for Predicting Used Car Resale Prices Using Granular Vehicle Equipment Information. Expert Systems with Applications. 2025.

Пуцал Ігор Ігорович – студент групи 2КН-226, кафедра комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: igorek12102004@gmail.com.

Богач Ілона Віталіївна – к.т.н., доцент кафедри автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: ilona.bogach@gmail.com.

Pushchal Ihor Ihorovych – student of the 2CS-22b group, Department of Computer Science, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: igorek12102004@gmail.com.

Bogach Ilona Vitaliivna – associate Professor of Automation and Intelligent Information Technologies Department, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: ilona.bogach@gmail.com.