

УДК 622.276.054

Копей В. Б., канд. техн. наук, доцент

Івано-Франківський національний технічний університет нафти і газу, vkopey@gmail.com

ПРОГНОЗУВАННЯ ЧАСТОТИ ВІДМОВ КОЛОН НАСОСНИХ ШТАНГ ЗА ДОПОМОГОЮ АНСАМБЛІВ ДЕРЕВ РІШЕНЬ

Експлуатація свердловинних штангових насосних установок для видобування нафти ускладнена відмовами штангової колони, які вимагають проведення затратних ловильних і спуско-підіймальних ремонтних робіт. Сучасні статистичні методи обробки даних про відмови можуть бути використані для прогнозування частоти відмов колони за різними параметрами свердловини. Метою праці є обґрунтування ефективності застосування методів ансамблів дерев рішень (випадкового лісу та градієнтного бустінга дерев регресії) для прогнозування частоти відмов колон насосних штанг. Ці популярні методи ґрунтуються на використанні множини дерев рішень для створення більш точної моделі. Вони достатньо універсальні, не потребують попереднього масштабування даних і володіють високою правильністю [1]. Моделі розробляли на основі статистичних даних про відмови колон в НГВУ “Долинанафтогаз” [2]. Моделі враховують наступні параметри свердловини: діаметр насоса *Pump*, приведені напруження в верхній частині колони *Stress*, газовий фактор *Gas*, відсутність або наявність викривленості свердловини *Curv*, відсоток води в продукції *Water*, довжина колони *H*, продуктивність свердловини *Product*, відсутність або наявність парафіноутворень *Paraffin*, довжина секцій штанг різного діаметра (*H19*, *H22*, *H25*), місяць відмови *Month*, кількість відмов за три роки *Kv*. Кількість даних 563, відмови штока не враховували. Для моделювання використовували бібліотеку Python для машинного навчання scikit-learn (версія 0.19.1) [3].

Спочатку виконували випадкове перемішування даних та їх поділ на дві частини - для пошуку моделі (80%) і її перевірки (20%). Для оцінки якості регресійних моделей використовували коефіцієнт детермінації R^2 . Найкращі значення параметрів моделей шукали в певних інтервалах (для RandomForestRegressor $n_estimators=[50, 150]$, $max_depth=[3, 5]$, для GradientBoostingRegressor $n_estimators=[50, 150]$, $learning_rate=[0,01, 0,6]$, $max_depth=[3, 5]$) за допомогою функції для стохастичної оптимізації на основі генетичних алгоритмів differential_evolution з пакету SciPy (версія 1.0.0) [4]. Ця функція шукала максимум результату (середнього значення R^2) 7-блокової перехресної перевірки моделі. Результати пошуку для RandomForestRegressor: $n_estimators=133$, $max_depth=5$. Перехресна перевірка моделі на тестових 20% даних показала середнє значення правильності $R^2=0,74$. Результати пошуку для GradientBoostingRegressor: $n_estimators=120$, $learning_rate=0,15$, $max_depth=4$. Середнє значення правильності $R^2=0,89$. За допомогою функції feature_importances_ моделі GradientBoostingRegressor обчислили важливість ознак: *Product* - 0,19, *Gas* - 0,14, *H* - 0,12, *H25* - 0,12, *H19* - 0,10, *H22* - 0,09, *Water* - 0,09, *Stress* - 0,08, *Month* - 0,03, *Pump* - 0,02, *Paraffin* - 0,02, *Curv* - 0,01. З рис. 1 помітно, що модель прогнозує дещо завищені значення частоти відмов для $Kv < 4$, і дещо занижені для $Kv > 4$.

Розглянута також задача бінарної класифікації. Свердловини з $Kv \leq 4$ віднесено до класу низькоаварійних свердловин ($Kvcat=0$), а свердловини з $Kv > 4$ - до високоаварійних ($Kvcat=1$). Таким чином отримано приблизно збалансовані класи за кількістю відмов. Знайдені значення параметрів найкращої моделі RandomForestClassifier ($n_estimators=121$, $max_depth=5$) та моделі GradientBoostingClassifier ($n_estimators=120$, $learning_rate=0,38$, $max_depth=5$).

В таблиці 1 наведено результати перехресної перевірки моделей на тестових даних. Жирним виділені матриці помилок розміром 2×2 . В матриці помилок літера *N* означає клас $Kvcat=0$, літера *P* - клас $Kvcat=1$. Літера *T* означає, що прогноз відповідає дійсності, а

F - не відповідає. В якості показника якості моделі використовують правильність $(TN + TP)/(TN + TP + FN + FP)$, точність $TP/(TP + FP)$, повноту $TP/(TP + FN)$, F1-міру (середнє гармонічне точності і повноти) [1]. Якщо необхідно знизити кількість прикладів FN (свердловин, які помилково віднесені до низькоаварійних), то в якості показника якості моделі слід використовувати повноту. Шляхом перехресної перевірки обчислена середня точність класифікатора (площа під кривою точності-повноти) RandomForestClassifier (0,94) та GradientBoostingClassifier (0,99). Важливість ознак для GradientBoostingClassifier: *Product* - 0,09, *H* - 0,08, *Gas* - 0,07, *Stress* - 0,07, *H25* - 0,06, *H22* - 0,06, *Water* - 0,05, *H19* - 0,04, *Pump* - 0,01, *Month* - 0,01, *Curv* - 0,01, *Paraffin* - 0,01.

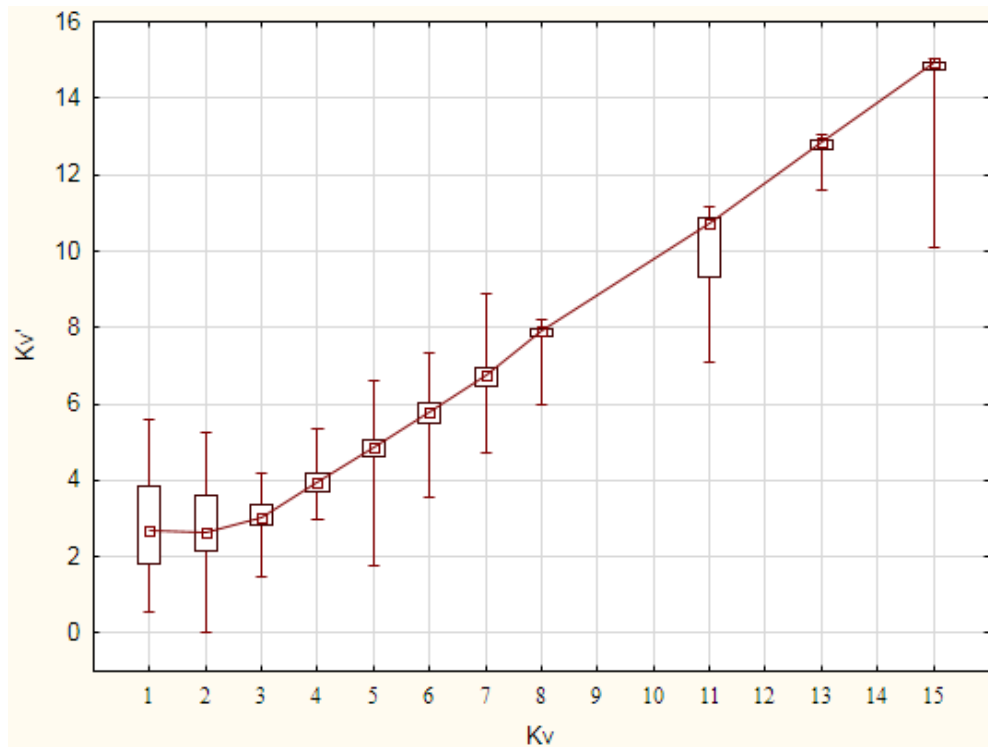


Рис. 1 – Реальні Kv і прогнозовані Kv' значення частоти відмов для моделі GradientBoostingRegressor: □ - медіана; прямокутник - 25%-75%; відрізок - 1%-99%

Таблиця 1 – Середні значення результатів перехресної перевірки моделей RandomForestClassifier і GradientBoostingClassifier на тестових даних

Модель		Класифіковано як		Точність	Повнота	F1-міра	Всього
		0	1				
RandomForest	Фактично 0	TN=51	FP=10	0,89	0,84	0,86	61
	Фактично 1	FN=6	TP=46	0,82	0,88	0,85	52
GradientBoosting	Фактично 0	TN=56	FP=4	0,95	0,93	0,94	60
	Фактично 1	FN=3	TP=50	0,93	0,94	0,93	53

Помітно, що моделі GradientBoosting володіють дещо вищою правильністю (0,89 для регресії і 0,94 для класифікації) і їх застосування дозволить ефективно ідентифікувати високоаварійні свердловини ще на етапі проектування.

Список посилань

1. Мюллер, А. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными [Текст] / Андреас Мюллер, Сара Гвидо; Пер. с англ. – СПб.: ООО "Альфа-книга", 2017. – 480 с.
2. Копей, В.Б. Аналіз відмов колон насосних штанг в НГВУ “Долина нафтогаз” [Текст] / В.Б. Копей, І.І. Стеліга // Розвідка та розробка нафтових і газових родовищ. – № 4(5). – 2002. – С.78-80.

3. Documentation scikit-learn: machine learning in Python – scikit-learn 0.19.1 documentation [Electronic resource]. – Mode of access: <http://scikit-learn.org/0.19/documentation.html>

4. SciPy – SciPy v1.0.0 Reference Guide [Electronic resource]. – Mode of access: <https://docs.scipy.org/doc/scipy-1.0.0/reference/>

УДК 005.311.2:004.94

Павленко П. М., докт. техн. наук, професор
Національний авіаційний університет, м. Київ, petrpav@nau.edu.ua
Балушок К. Б., канд. техн. наук, доцент
заступник головного технолога ПАТ Мотор Січ, м. Запоріжжя
Захарчук Т. М.
Темніков А. В.
Національний авіаційний університет, м. Київ, temnikoff@ukr.net

МЕТОДИ ТА ТЕХНОЛОГІЯ ІНТЕГРАЦІЇ ВИРОБНИЧИХ ДАНИХ

Сучасне промислове виробництво успішно використовує інтегровані інформаційні системи виробничого призначення: CAD/CAM/CAE, PDM, ERP та інші. Але їх повноцінна інформаційна інтеграція до даного часу залишається як науковою, так і прикладною проблемою [1, 2].

Провідні IT-компанії вирішують проблему інтеграції засобами своєї PDM-системи, яка створює інтегроване інформаційне середовище реалізації функцій проектування, управління, планування та інше. Але конкретне підприємство, як правило, одночасно експлуатує різні за своїм функціональним призначенням програмні продукти. Таке різноманіття зумовлено великою кількістю об'єктивних причин: об'єднанням компаній, що використовують різні інформаційні системи, значною вартістю сучасних виробів, різною вартістю інформаційних систем, тощо.

Проблемні питання інтеграції і неузгодженості даних у інформаційних системах виробничого призначення виникають також при використанні власних програмних розробок спеціалістів підприємств. Використання API-інтерфейсів і універсальних форматів даних не завжди реалізують надійну інтеграцію даних. Тобто необхідні нові наукові й проектні рішення з інтеграції даних інформаційних систем виробничого призначення.

Такі розробники інформаційних систем виробничого призначення, як, наприклад, Dassault Systemes (Франція) та Siemens PLM Software (Німеччина) тощо, як правило, використовують декілька різних методів при інтеграції даних своїх систем. Для діючих промислових підприємств (особливо дискретного типу – машинобудівних, авіаційних тощо) необхідні нові рішення, які базуватимуться на спільному використанні більшості методів інтеграції, що не суперечить базовим вимогам організації взаємодії та функціонування існуючих (впроваджених) інформаційних систем виробничого призначення [3].

Для реалізації такої системної інтеграції створюються програмні засоби системи інформаційної підтримки (СІП) інтеграції отриманих проектних та виробничих рішень, які дозволяють здійснювати обмін даними з можливістю контролю та корекції трансльованих даних на основі вихідних структурних описів з урахуванням похибки побудови та відображення даних, що передаються та підтримують інтеграцію інформаційних систем технічної підготовки і виготовлення виробів.

Авторами запропонований новий спосіб інтеграції даних інформаційних систем виробничого призначення та даних, які генерують підсистеми реалізації розроблених методів прийняття оптимальних проектних та виробничих рішень, реалізованих на базі універсальної PDM-системи за допомогою міжмодульних програмних інтерфейсів.