

ФОРМУВАННЯ НАВЧАЛЬНОЇ ВИБІРКИ ТА ОБҐРУНТУВАННЯ АЛГОРИТМІВ ПОБУДОВИ МОДЕЛЕЙ ML ДЛЯ ПРОГНОЗУВАННЯ РЕСУРСІВ ІТ-ПРОЄКТІВ

Горносталь Є.М., Челак В.В., Горносталь О.А.
*Національний технічний університет
"Харківський політехнічний інститут", м. Харків*

Прогнозування ресурсів для реалізації ІТ-проектів, зокрема мобільних застосунків, є складною багатофакторною задачею. Якість результату безпосередньо залежить від повноти, коректності та релевантності навчальної вибірки, а також від обраного алгоритму машинного навчання. У роботі детально розглянуто процес побудови навчальної вибірки для задачі прогнозування ресурсів: кількості спеціалістів, їхніх рівнів старшинства, потреби в ліцензіях, обладнанні, фінансових та часових витратах.

Для формування вибірки зібрано дані про минулі ІТ-проекти, що містять такі ознаки: тип застосунку, кількість функцій, платформи, тривалість етапів, склад команди, загальний бюджет, використані технології та зовнішні сервіси. Проведено попередню обробку даних: очищення, нормалізація, усунення пропущених значень, кодування категоріальних змінних та генерація похідних ознак.

Для вибору моделей розглянуто класи алгоритмів, здатних працювати як із числовими, так і з категоріальними змінними:

- Регресійні моделі: лінійна регресія, *Lasso*, *Ridge* – для базового прогнозування часових і фінансових ресурсів;
- Рішення деревоподібної структури: *Decision Tree Regressor*, *Random Forest*, *XGBoost* – для врахування взаємодії між ознаками, важливості факторів, гнучкої роботи з неідеальними даними;
- Глибоке навчання: нейронні мережі (*MLP*) – для виявлення складних залежностей між вхідними параметрами та вихідними метриками;
- Методи ансамблю: поєднання кількох моделей для зниження варіативності та покращення загальної точності прогнозу.

Проведено порівняльний експеримент із використанням крос-валідації та метрик *MAE*, *RMSE*, R^2 . Результати демонструють, що ансамблеві моделі (зокрема, *Gradient Boosting*) забезпечують найвищу точність у завданні багатофакторного прогнозування ресурсів, тоді як прості регресії можуть слугувати базовими *oracles* для швидкої оцінки. Таким чином, формування якісної навчальної вибірки в поєднанні з правильним вибором алгоритмів машинного навчання створює основу для побудови надійної системи підтримки управлінських рішень у сфері ІТ-проектів.