

## ОПТИМІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АНАЛІЗУ ВЕЛИКИХ ДАНИХ ТЕЛЕКОМУНІКАЦІЙНОЇ КОМПАНІЇ

Лавренюк А.М., Лавренюк С. І., Тульчинський П.Г.

*Фізико-технічний інститут НТУУ «КПІ ім. Ігоря Сікорського», Україна;*

*Інститут кібернетики ім. В. М. Глушкова НАН України, Україна*

*E-mail: lsi@bigmir.net*

### SOFTWARE OPTIMIZATION FOR TELECOMMUNICATION COMPANY BIG DATA ANALYSIS

Using optimal architecture development for software that created using python programming language, we have shown the solution of one of the Big data analysis problem for telecommunication company churn prediction.

We have shown decision-making system creation with flexible modular structure and with software development time and resource costs optimization.

На прикладі розробки оптимальної архітектури програмного забезпечення створеного на мові програмування python показано рішення однієї із задач аналізу великих даних для прогнозування відтоку клієнтів (churn predict) телекомунікаційної компанії. Показано створення системи прийняття рішень з гнучкою модульною структурою та з оптимізацією часових та ресурсних витрат на розробку програмного забезпечення.

**Вступ.** При рішенні задачі прогнозування відтоку клієнтів телекомунікаційної компанії проводиться аналіз великого об'єму даних, адже телекомунікаційні компанії:

- мають клієнтську базу від 1000000 і більше, дані необхідно брати за останні три місяці;
- для побудови моделей та тестування прогнозування на першому етапі достатні дані для телекомунікаційних компаній, що містяться в відкритих джерелах [1, 2];
- при розробці програмного забезпечення та його впровадженні необхідно збалансувати три компоненти, що значним чином впливають на результат:  $T$ -час,  $C$ -ресурси (люди, гроші),  $Q$ -якість;
- програмне забезпечення повинно бути: масштабоване, мультиплатформне, мати невелику сукупну вартість володіння;
- необхідність роботи програмного забезпечення на різних пристроях (робочі станції, смартфони, планшети...), а інтерфейс програмних систем повинен бути розрахований на роботу різних категорій співробітників.

Одній групі користувачів необхідна більша інтерактивність з можливістю проводити самостійно різне представлення результатів та розрахунків додаткових показників. Для другої групи необхідно швидко отримувати точні результати розрахунків основних показників для прийняття управлінських рішень. Тому інтерфейс повинен містити тільки необхідні дані, працювати швидко на різних пристроях та системах.

Якщо розробляти різні модулі на різні операційні системи та пристрої для різних груп користувачів будуть значно збільшені  $T$ ,  $C$  при  $Q=const$ :

$$\begin{aligned} T &= T_0 + \Delta T, C = C_0 + \Delta C, \\ \text{де } \Delta T &> 1, \Delta C > 1 \end{aligned} \quad (1)$$

**Запропонований підхід.** Виходячи з наявного досвіду побудови програмного забезпечення для прогнозування даних [3] та для обробки великих об'ємів даних [4], провівши ряд експериментів, запропоновано для рішення задачі аналізу великих даних використати мову програмування python, та бібліотеки, що необхідні для машинного навчання та прогнозування: pandas, numpy, sklearn та інші.

Доцільно використати триярусну архітектуру: сервер баз даних, сервер додатків та робочі станції користувачів. Основну задачу по обробці даних, машинному навчанню та прогнозуванню візьме на себе сервер додатків. Так, як для машинного навчання в python на базі бібліотеки sklearn використовується не просто бібліотека numpy, а numpy+mkl, то буде забезпечена достатня швидкодія, а django до python дає можливість побудувати веб-додаток (web-application) з високою швидкодією, гнучкістю та мультиплатформністю. Веб додаток проводить аналіз даних, буде інтерфейс користувача для другої групи користувачів, та правильно працює на різних пристроях. Jupiter notebook веб додаток для створення інтерфейсу для першої групи користувачів (рис. 1).

Ядро системи – обробка та аналіз даних, також спільні процедури доступу до даних винесені в окремі модулі та є спільними для двох систем. Таким чином, вірно виділивши спільні процедури, функції, та з'єднавши їх із першою та другою системами ми отримуємо:

$$\begin{aligned} T &= T_0 + \Delta T_1, C = C_0 + \Delta C_1, \\ \text{де } \Delta T_1 &< \Delta T, \Delta C_1 < \Delta C, Q = const, \Delta T_1 \Rightarrow 0, \Delta C_1 \Rightarrow 0 \end{aligned} \quad (2)$$

Слід зазначити, що технології глибинного навчання та методи машинного навчання успішно використовуються у великих об'ємів даних при проведенні аерокосмічних спостережень [5, 6].

Тому запропонований авторами підхід можна буду успішно використати в інших галузях.



Рис. 1. Архітектура програмного забезпечення.

Проведені експерименти показали високий результат якості прогнозування та швидкості обробки даних.

#### **Висновок.**

- мова програмування python та наявні компоненти дають можливість побудови оптимальних програмних систем для аналізу великих об'ємів даних, для машинного навчання, та для побудови мультиплатформних веб додатків або додатків робочого столу (desktop app) не критичних до великої швидкодії;

- запропонований підхід до побудови програмного забезпечення для аналізу великих даних базі мови програмування python зменшує витрати на створення програмного забезпечення та його супровід в подальшому.

#### **Література**

1. Telco Customer Churn data set, <https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/>.
2. Machine Learning Repository, [http://archive.ics.uci.edu/ml/?cm\\_mc\\_uid=15134185542514827482720&cm\\_mc\\_sid\\_50200000=1485174656](http://archive.ics.uci.edu/ml/?cm_mc_uid=15134185542514827482720&cm_mc_sid_50200000=1485174656).
3. Автоматическое построение числовых прогнозов / [Лавренюк С.И., Перевозчикова О.Л., Тульчинский В.Г., Харченко А.В.]/Компьютерная математика.–2010.–№ 2. – С. 52–61.
4. Лавренюк А.Н., Лавренюк С.И., Назаренко Е.В. Подход к оптимизации программы обработки больших объемов сейсмических данных на примере 3D миграции дуплексных волн//Компьютерная математика., сборн. научн. трудов–2014 №1–Р. 69-75.
5. Лавренюк М., Новіков О. Технологія глибинного навчання для великих об'ємів даних // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки, GEO-UA 2016.
6. Новіков О., Шелестов А., Лавренюк М. Розробка методів машинного навчання в проєкті Google Earth Engine // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки, GEO-UA 2016. — 2016.— Р. 20-21.