

РОЗРОБКА СИСТЕМИ ДЛЯ АНАЛІЗУ ДАНИХ ПЛАТФОРМИ ВИВЧЕННЯ ІНОЗЕМНИХ МОВ

Кошарний Є. Ю.

Науковий керівник – д.т.н., проф. Перова І. Г.

Харківський національний університет радіоелектроніки, каф. СТ

м. Харків, Україна

e-mail: yevhenii.kosharnyi@nure.ua

This work is devoted to developing a system aimed at providing a pleasant user experience on a foreign language learning platform while empowering businesses with valuable insights into user behavior. Through comprehensive data analysis, the system enables data-driven decision-making. Leveraging Google Cloud Platform services ensures scalability and efficient handling of user tracking data. Integration of dbt facilitates data transformation for extracting actionable insights. Furthermore, a reporting layer with dashboards, such as Tableau or alternatives, enhances accessibility to the transformed data, facilitating informed decision-making processes. This thesis evaluates the system's efficacy in enhancing user experience and supporting data-driven decisions for businesses.

У сучасному цифровому світі різко зріс попит на ефективні платформи для вивчення іноземних мов, що зумовлено глобалізацією та зростанням важливості володіння кількома мовами в різних галузях. Тому розробка системи для аналізу даних з такої платформи є надзвичайно актуальною і допоможе в розвитку бізнесу у сфері онлайн освіти. Ця робота спрямована на розробку комплексної системи аналізу даних з платформи для вивчення іноземних мов, з основним акцентом на покращення користувацького досвіду та надання цінної інформації для прийняття обґрунтованих рішень бізнесом. Розробка та впровадження системи передбачає багатоетапний підхід, починаючи з ретельного дослідження освітнього ландшафту та аналізу продуктів конкурентів, щоб визначити потреби ринку та виявити потенційні сфери для вдосконалення та покращення. Після цього було ретельно визначено функціональні вимоги до платформи, які окреслюють ключові функції, необхідні для мінімально життєздатного продукту (MVP).

На наступному етапі розроблено дизайн платформи, який ретельно продумано, щоб забезпечити зручність користування та візуальну привабливість як на комп'ютерах, так і на мобільних телефонах. Для цього було використано такі інструменти, як Figma. Процес розробки враховує відгуки зацікавлених сторін для досягнення оптимального користувацького досвіду та привабливого графічного інтерфейсу. Одночасно окреслено високорівневу архітектуру системи з визначенням необхідних сервісів, включно із front-end серверами, back-end API,

операційною базою даних, сервісом доставки повідомлень, сервісом обробки подій користувачів (user events consumer), рішенням для зберігання даних та обробки даних (data warehouse), сервісом для створення аналітики. Технологічний стек для кожного компоненту системи ретельно підібрано відповідно до вимог проекту, щоб забезпечити безперешкодну інтеграцію в систему.

У цьому проекті використано поєднання сучасних технологій у різних компонентах системи для забезпечення надійної та ефективної роботи.

Front-end платформи для вивчення іноземних мов розроблені з використанням React та JavaScript для створення динамічних користувацьких інтерфейсів, і стилізований за допомогою HTML і CSS.

На back-end стороні використано Python фреймворк Django для створення API (Application Programming Interface).

Як операційну базу даних використано PostgreSQL, яка буде розташована в сервісі Cloud SQL в хмарі Google Cloud Platform.

Як сервіс доставки повідомлень використано Google Cloud Pub/Sub.

Компонент для обробки подій користувачів в реальному часі написано на Python та розгорнуто в сервісі Cloud functions.

BigQuery data warehouse слугує надійним сховищем даних, бо цей сервіс легко масштабується, може оброблювати великі об'єми даних та є оптимізованим для зберігання та обробки аналітичної інформації.

Трансформація даних на потужностях BigQuery виконується за допомогою dbt (Data Build Tool), інтегрованого в GitHub Actions для автоматизації робочих процесів.

Нарешті, Tableau використовується для візуалізації даних, надаючи зацікавленим сторонам інтуїтивно зрозумілі візуалізації та інсайти, отримані на основі перетворених даних, що сприяє процесам прийняття обґрунтованих рішень бізнесом. Описані вище компоненти системи зображено на верхньорівневій діаграмі архітектури зображеній на рисунку 1.

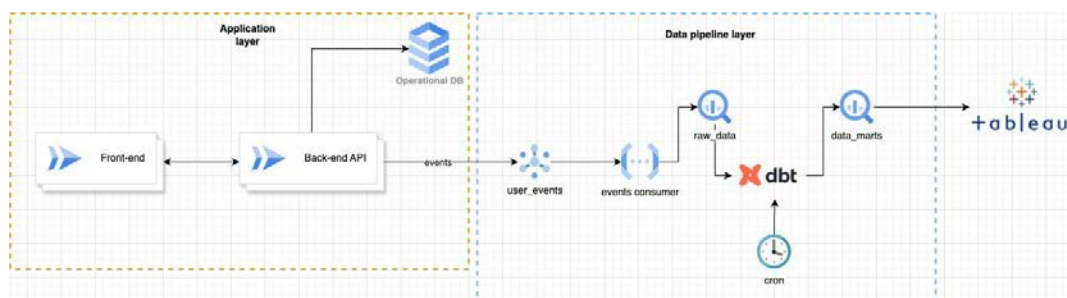


Рисунок 1 – Діаграма архітектури системи

В основі системи лежить процес трансформації даних за допомогою інструменту dbt (Data Build Tool). Цей інструмент дозволяє описати комплексні трансформації даних за допомогою чистого SQL, автоматизувати оновлення моделей даних, забезпечити перевірку якості

даних шляхом написання тестів. Цей важливий крок передбачає перетворення необроблених даних користувачів на дієві інсайти, що дозволить компанії робити цінні висновки з усього обсягу зібраної інформації.

Для подальшого розширення доступу до цих даних у систему інтегровано аналітичні дашборди для візуалізації даних. Для реалізації використано інструмент Tableau. Ці дашборди пропонують зацікавленим сторонам візуальне представлення ключових показників і тенденцій на основі зібраних даних, що дає їм змогу приймати обґрунтовані рішення у реальному часі. У сфері платформ для онлайн навчання та вивчення іноземних мов розробка системи аналізу даних може стати першим кроком до створення моделі машинного навчання, яка в майбутньому може бути використана для покращення користувацького досвіду, прогнозування кількості учнів, або використання методу кластеризації високовимірних даних для отримання корисної інформації з великого об'єму даних.

Протягом усього життєвого циклу розробки застосовуються процедури тестування для забезпечення надійності, масштабованості та продуктивності системи. Приймальне тестування (User acceptance testing) перевіряє функціональність і зручність платформи, в той час як тестування продуктивності (performance testing) оцінює її швидкість реагування та ефективність під різними навантаженнями.

Таким чином, у цій роботі представлено комплексний підхід до розробки системи аналізу даних з платформи для вивчення іноземних мов. Використовуючи передові технології та хмарні рішення, система має на меті забезпечити безперебійний користувацький досвід, одночасно надаючи бізнесу цінну інформацію для прийняття обґрунтованих рішень в освітній галузі.

Список використаних джерел:

1. What is Cloud Run [Електронний ресурс] // Google Cloud. – 2024. – Режим доступу до ресурсу: <https://cloud.google.com/run/docs/overview/what-is-cloud-run>. (дата звернення: 02.03.2024).
2. What is dbt? [Електронний ресурс] // dbt Labs, Inc.. – 2024. – Режим доступу до ресурсу: <https://www.getdbt.com/product/what-is-dbt>.
3. BigQuery overview [Електронний ресурс] // Google Cloud. – 2024. – Режим доступу до ресурсу: <https://cloud.google.com/bigquery/docs/introduction>.
4. Bodyanskiy Ye., Perova I., Zhernova P. Online fuzzy clustering of high – dimensional data based on ensembles in data stream mining tasks, Сучасний стан наукових досліджень та технологій в промисловості. 2019. No1(7). С. 16-24.