**Problem Statement**

Build a machine learning model that can **identify whether a session of visited websites belongs to Alice or not** based on sequences of website IDs and timestamps. **Binary classification** challenge predicts if a session belongs to Alice (target = 1) or not (target = 0).

**About Dataset**

Web browsing sessions from various users, and the goal is to build a machine learning model to identify sessions that belong to **Alice**.

**Files:**

1. **train\_sessions.csv:** contains labeled data for model training.

**Each row** = 1 session (up to 10 websites visited).

**Columns:**

**sites ->** IDs of visited websites.

**time** **->** Timestamps of visits.

**target ->** 1 if the session belongs to Alice, 0 otherwise.

1. **site\_dic.pkl:** A dictionary mapping (site\_id to site\_name)

Use this to understand which site each ID refers.

**Feature Engineering**

Use time-based patterns to extract features that help identify whether a session belongs to **Alice** or not.

1. Convert Time Columns to Datetime Format. Before we can work with time, we must convert the raw strings into datetime objects.
2. Time Differences Between Consecutive Site Visits. (Subtract time1 from time2, time 2 from time3, so on). Each of these differences gives you how many seconds passed **between two steps** in the user’s session.
3. Extract Date and Time Features. This will help detect **when** the session started.
4. Prevent Outliers, captures time between pages within a session. (How many seconds passed between **page 1 and page 2** in the same session).
5. Create Session-Level Features, means how many **pages** the user visited in that session.
6. One-Hot Encode Time-Based Features. Turn categorical time values into one-hot encoded binary features.