

3.1.2 Sequence Data

We use the Yahoo server traffic datasets in the Yahoo Webscore program [19] and the NASA Shuttle dataset in the UCI ML repository [20]. The Yahoo server traffic datasets contain real and synthetic univariate time series, each of which contains 1,420 time ordered records with one time-dependent attribute called *traffic_value*. These datasets contain time series with random seasonality, trend and noise. Each data-point represents one hour’s worth of traffic data. Table 3.5 shows the Schema of the Yahoo server traffic table. Anomalies in the *traffic_value* indicate potential security threats to the Yahoo user’s data.

Table 3.5: Schema of *Yahoo* Server Traffic Table

Attribute Name	Data Type	Description
ID	Numeric	Unique
Time	DateTime	Nullable
Traffic_value	Float	Nullable

The NASA Shuttle multivariate dataset contains 58,000 time-ordered records with seven time-dependent numerical attributes, namely, *Rad_flow*, *Fpv_close*, *Fpv_open*, *High*, *Bypass*, *Bpv_close*, and *Bpv_open*. Table 3.6 shows the Schema of the NASA Shuttle table. Incorrect values in these attributes have consequences for aerospace industry that researches, designs, manufactures, operates, or maintains the spacecrafts.

Table 3.6: Schema of *NASA Shuttle* Table

Attribute Name	Data Type	Description
ID	Numeric	Unique
Time	DateTime	Nullable
Rad_flow	Float	Nullable
Fpv_close	Float	Nullable
Fpv_open	Float	Nullable
High	Float	Nullable
Bypass	Float	Nullable
Bpv_close	Float	Nullable
Bpv_open	Float	Nullable