Signal Type Definitions

Bucketing

Bucketing features involve grouping the data by a specific entity and aggregating a column's values across categories of a categorical column. Each feature value is represented as a dictionary with category names as keys and the corresponding aggregated values as values. For example, a shopping basket of items (a count or value of items purchased) is a cross-aggregation feature.

Similarity

Similarity features compare an individual entity's attributes or past activity to a group, e.g., a ratio of a person's income to the average income of people with the same role, or a comparison of a person's purchases to those of all people living in the same geographic region. In addition, similarity features can compare the aggregation statistics or activities of a group that is related to the entity being analyzed with those of a different group.

Stability

Stability features compare recent data values and events to those of earlier periods, e.g. the similarity of a customer's most recent shopping bask items versus those of the previous 6 months.

Attribute Change

Change features measure the occurrence or magnitude of changes to slowly changing attributes, e.g., the number of times a customer changed address over the past year.

Diversity

Diversity features measure the variability of data values, e.g., the coefficient of variation of invoice amounts over the past 6 months, or the variety of products purchased as measured by entropy.

Location

A location feature measures the static or dynamic location of an event or entity, e.g., the distance between a shop and the customer address.

Timing

A timing feature relates to the timing of events or the time period between events. Seasonality features measure the timing of events, e.g., whether a customer tends to shop on the same days of the week. Inter-event signals measure a statistic on the time period between events, e.g., the maximum time period between two consecutive invoices for a

customer over the latest 4-week period. Clumpiness features measure the occurrence of long gaps in time between bursts of multiple events. For example, binge-watching Netflix TV series is a clumpy behavior, while washing a car every Saturday is not.

Recency

A recency feature measures the time since the latest event to occur, e.g., the last product purchased, or elapsed time since the last purchase.

Latest Event

A recency feature captures an attribute of the latest event to occur, e.g., the latest product purchased, or total spend on the latest purchase.

Frequency

A frequency feature counts the occurrence of events over a time window, e.g., the number of international phone calls a customer made over the past 4 weeks.

Attribute

An attribute feature is an entity's point-in-time attribute value that rarely or never changes, e.g., a customer's year of birth or address. It can also be a feature derived from a lookup, e.g., a customer's current age.

Aggregate Statistic

A statistic feature is any aggregation feature that doesn't match the definition of any other signal type, e.g., the maximum temperature of a hospital patient over the past 24 hours.