Appendix – I

Glossary of relevant terms

- **Data Mining**: Data mining is concerned with analyzing large volumes of, often unstructured data, to automatically discover interesting regularities, patterns or relationships among data items.

- **Temporal Data mining**: The field of temporal data mining is the analyzing large ordered or sequential data streams.

- **Sequential Data**: Sequential data streams can be time ordered data as well as those that are ordered with respect to some index other than time.

- **Temporal Patterns**: The temporal patterns consist of two components: a set of state intervals and a set of relationships between those state intervals that represent the order of states within the pattern. These relationships can be before(b), meets(m), overlaps(o), is-finished-by(fi), contains(c) and starts(s).

- **Sequential Patterns**: A sequential pattern is a large and maximal sequence among the set of all large sequence. If a sequence $S$ of itemsets is denoted by $\langle s_1, s_2, ..., s_n \rangle$, where each $s_j$ is an itemset, it is called an $n$-sequence.

- **Episodes**: Episodes are patterns that occur sufficiently often in the data presented as a single long sequence [36]. This single event sequence is denoted by $\langle (E_1, t_1), (E_2, t_2), ... \rangle$, where $E_i$ takes values from a finite set of event types $E$, and $t_i$ is an integer denoting the time stamp of the $i$th event and $\forall i, t_i \leq t_{i+1}$.
• **Interestingness of a Rule:** A Rule is said to be interesting if it enforces or contradicts our belief system. Some of the interestingness measures defined are *conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility, and actionability.*

• **Post-Processing of Rules:** Post-processing of Rules encompasses managing the generated knowledge after the data mining activity has been performed on the accumulated data. This may involve ranking the rule according to interestingness measures, grouping or pruning them.

• **Set-valued Attributes:** Attributes which can take a set of values for each instance of the entity they are associated with.

• **Signature:** A signature is a superimposed bit pattern generated from the values of the attributes. The target signature is obtained by the bitwise union of all the element signatures of the target set.

• **Signature Based Index:** Signature based index is an index which is built on set-valued attribute, by generating a signature for each instance of that attribute and then inserting <signature, reference-id> pair in the signature file. This signature file is then used as index for retrieval purposes.

• **Bitmaps:** Bitmaps are used to store one slice of bits per distinct attribute value of a record. So there will be as many bitmaps as the cardinality of the attribute which needs to be indexed in the record. Each bit of the slice is mapped to a record or a data object.

• **Temporal Bitmaps:** Temporal bitmap is a variation of the basic bitmap where along with showing presence or absence of a value, the order of that value in the set is also indicated.