1 Data Collection, Build, and Pre-Processing Tasks

1.1 Stage 1 – Data Review

- Selecting a context-expressing dataset.
- Ensuring graph compatibility.
- Sorting features by importance.

1.2 Stage 2 – Feature Importance

- Defining primary columns that are critical in each dataset.
- Identifying columns of secondary importance in each dataset.

1.3 Stage 3 – Feature Relationships

- Specifying feature properties that can be constructed with other datasets, such as relationships (e.g., specialization).
- Analyzing feature importance versus feature relationships.

1.4 Stage 4 – Data Building

- Building a unified dataset with rich relationships that explains a specific effect.
- Creating multiple versions of the unified dataset.
- Comparing dense versus sparse graph structures.

1.5 Stage 5 – Data Pre-Processing

- Cleaning the data.
- Splitting the dataset into training and test sets (edges).
- Applying feature scaling.