

■ Data Preprocessing in Data Mining Architecture (for Smart Ambulance & Hospital Alert System)

- 1. Data Sources** - Accident reports (location, severity, injured count) - Ambulance GPS & availability - Hospital details (beds, ICU, doctors) ➡ **■** These are the raw data inputs into the data mining architecture.
- 2. Data Cleaning** - Remove noise (e.g., duplicate accident entries). - Handle missing values (unknown ambulance location = "null" or imputed). - Correct inconsistent data (e.g., "Karur bus stand!!" → "Karur Bus Stand"). **■** Ensures accuracy & reliability.
- 3. Data Integration** - Merge multiple sources into a single dataset. - Accident data + ambulance data + hospital data combined into a central database. **■** Prepares data for global view.
- 4. Data Transformation** - Convert attributes into proper format: - "02" → 2 (integer injured count). - Severity → categorical values (Low=1, Medium=2, High=3). - GPS coordinates standardized for distance calculation. **■** Makes data uniform & usable.
- 5. Data Reduction** - Keep only important attributes: - AccidentID, location, severity, nearest ambulance, hospital, bed availability. - Drop unnecessary details (like ambulance driver name). **■** Improves processing efficiency.
- 6. Data Mining** - Apply techniques on clean dataset: - Clustering → Find accident hotspots in a city. - Classification → Predict severity category. - Association rules → "If highway accident + high severity → ICU needed."
- 7. Pattern Evaluation** - Check if the mined patterns are useful & valid. - Example: "Most accidents on highways occur between 8–10 PM."
- 8. Knowledge Presentation** - Show results in dashboards, charts, and hospital alerts. - Example: Heatmaps of accident-prone zones, average response time graphs.

How to Explain to Guide (Short Answer) "In our project, we follow the data mining architecture. Accident, ambulance, and hospital data are first collected, cleaned, and integrated. Then we transform and reduce the data into useful formats. Finally, we apply data mining techniques like clustering accident-prone areas, classification of severity, and association rules to improve hospital preparedness. The processed results are shown in dashboards and reports for decision-making."