

Data Problems in Real World

Too Much Data	Too Little Data	Fractured Data
<ul style="list-style-type: none">- Corrupt and noisy data- Feature extraction- Irrelevant data- Very large data sizes- Numeric/Symbolic	<ul style="list-style-type: none">- Missing attributes- Missing attribute values- Small amount of data	<ul style="list-style-type: none">- Incompatible data- Multiple data sources- Data from multiple levels of granularity

Data Preprocessing Techniques

Data Transformation	Information Gathering	Generation of New Information
<ul style="list-style-type: none">- Data filtering- Data ordering- Data editing- Noise modelling	<ul style="list-style-type: none">- Data visualization- Data elimination- Data selection- Principal components analysis- Data sampling	<ul style="list-style-type: none">- Data engineering/new features- Time series analysis- Data fusion- Simulation/creation of unmeasurable attributes- Dimensional analysis- Constructive induction

Data Analysis

- Knowledge discovery
- Prediction
- Modelling (process/system)
- Performing intelligent tasks (decision support, process planning, monitoring, diagnosis, etc).
- Building knowledge-based systems