

Data Cleaning	Data Transformation	Data Reduction	
Missing Data 1. Ignore The Tuplet 2. Fill The Missing Values (manually, by mean or by most probable value)	Normalization Attribute Selection Discretization Concept Hiererchy Generation	Data Cube Aggregation Attribute Subset Selection Numerosity Reduction Dimensionality Reduction	
Noisy Data 1. Binning Method 2. Regression 3. Clustering			
Unsupervised Learning	Dimensionality Reduction <ul style="list-style-type: none"> Feature Elicitation Meaningful Compression Structure Discovery Big data visualization 		Unsupervised learning Input data is unlabeled Has no feedback mechanism Assigns properties of given data to classify it Divided into Clustering & Association Used for analysis Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm A unknown number of classes
	Clustering <ul style="list-style-type: none"> Recommender Systems Targeted Marketing Customer Segmentation 		
Supervised Learning	Classification <ul style="list-style-type: none"> Identity Fraud Detection Image Classification Customer Retention Diagnostics 		Supervised learning Input data is labeled Has a feedback mechanism Data is classified based on the training dataset Divided into Regression & Classification Used for prediction Algorithms include: decision trees, logistic regressions, support vector machine A known number of classes
	Regression <ul style="list-style-type: none"> Population Growth Prediction Estimating life expectancy Market Forecasting Weather Forecasting Advertising Popularity Prediction 		
Reinforcement Learning	<ul style="list-style-type: none"> Real-time decisions Game AI Robot Navigation Learning Tasks Skill Acquisition 		