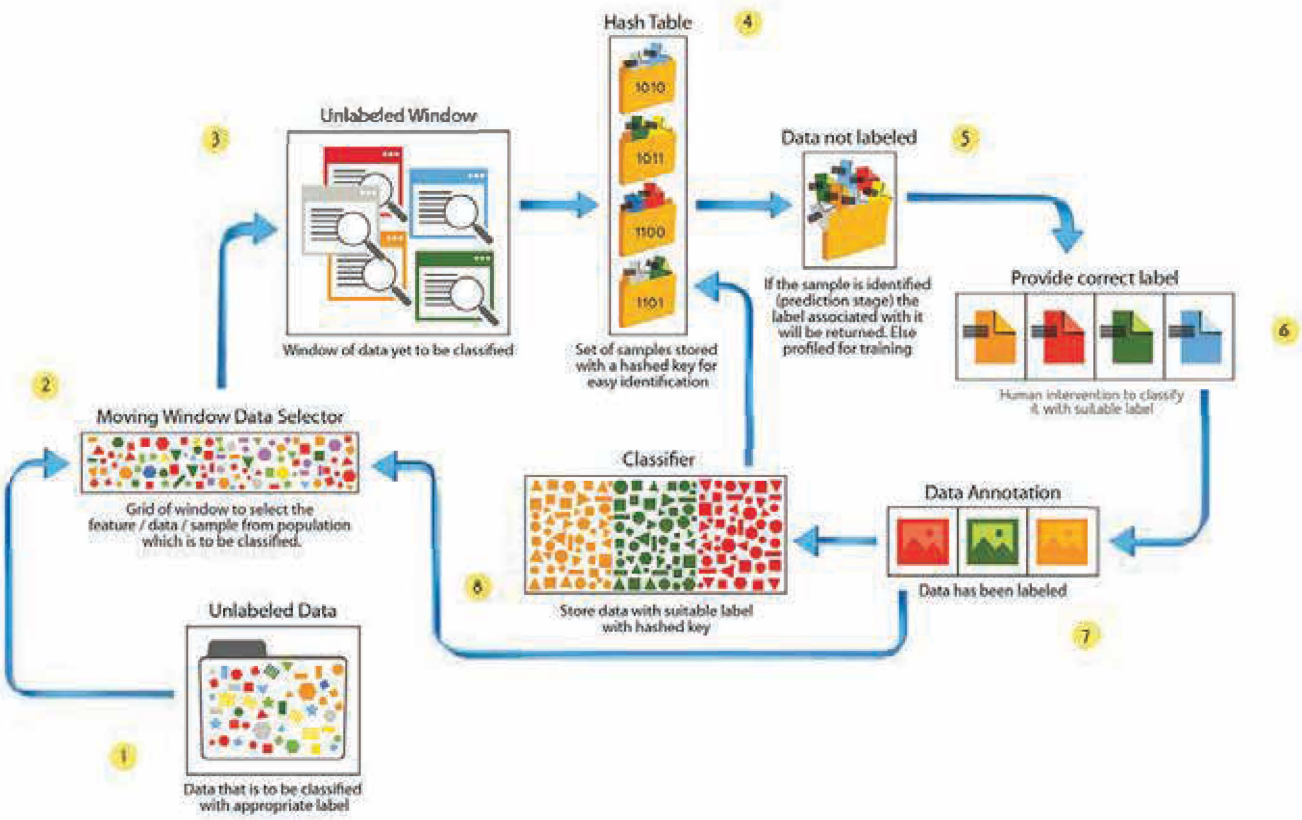
brandidea.ai
Granular Analytics Platform

Data Science Concepts

Gartner 2013 CoolVendor

Gartner 2014 CoolBusiness Model

Active Learning



Active Learning algorithm can attain accuracy with fewer training labels if it carefully selects the data from which it learns. An active learner may submit queries, typically as unlabeled data instances to be labeled by the user. In situations where unlabeled data is abundant or easily obtained, and manual labeling is difficult, expensive and time consuming, learning algorithm can actively query the user for labels.

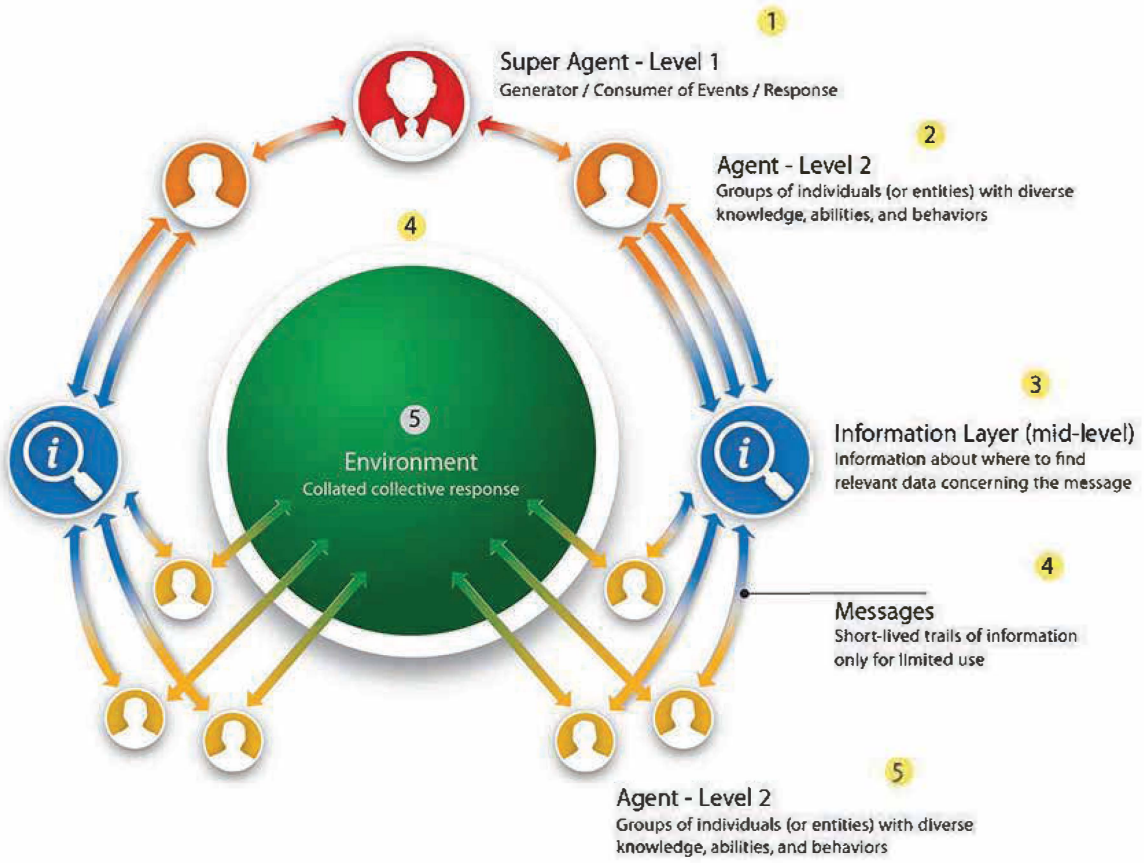
Goals:

Generate a robust classifier, without having to mark up and source the learner with more data than needed. It works towards keeping the human labeling effort to minimum, only requiring guidance where the training utility of the outcome of such a query is high.

BrandIdea's Implementation:

Product recognition & tagging (image analysis).

Agent-based Modeling



Agent-based Modeling is an individual-centric, decentralized approach to model design. The modeler distinguishes the active entities, the agents (people, companies, vehicles, cities, animals, products, etc.) characterizes their behavior (states, reactions, etc.), places them in a certain environment, sets up connections, and runs the simulation. As the consequence of interactions of many individual behaviors, a 'global' behavior emerges.

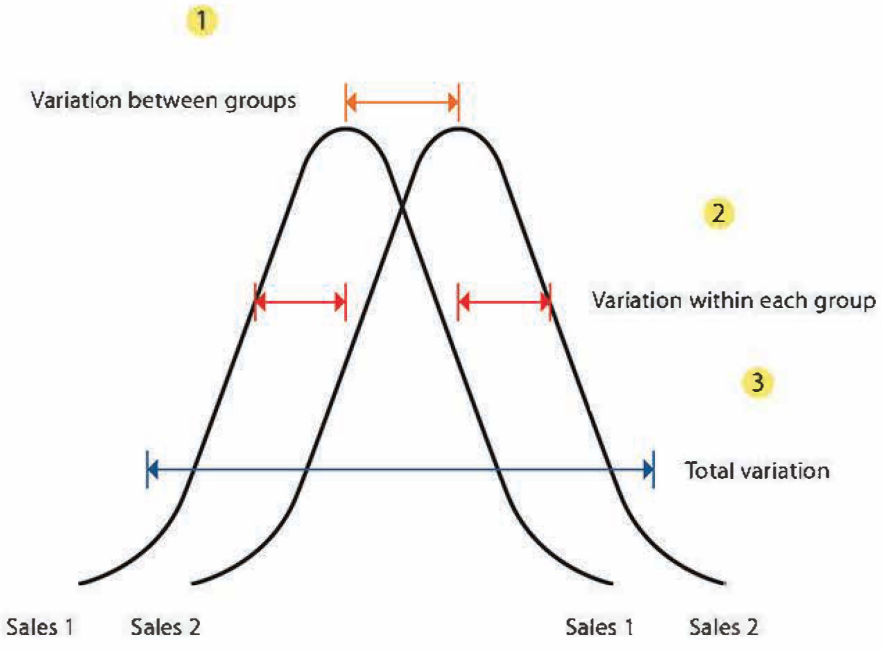
Goals:

Easy, up-to-date, and precise way to model, compare scenarios, optimize, and forecast a strategy.

BrandIdea's Implementation:

Consumer Behavior Analysis, Supply Chain

ANOVA (Analysis of Variance)



Sales 1	Sales 2
150	170
155	162
157	177
145	192
130	184
170	169
165	155

Two sample groups of sales data

A statistical technique to test the degree to which two or more groups of data vary or differ in a research. Large variance usually indicates that there was a significant finding from the experiment.

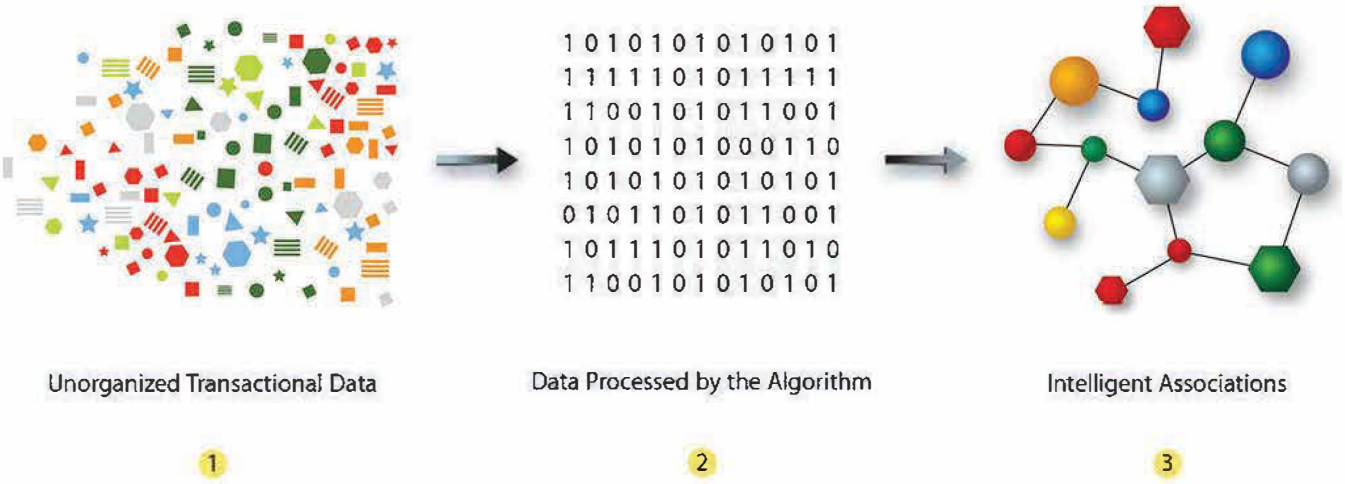
Goals:

Test and determine whether the difference between the data groups exists, over simple incidental likelihood.

BrandIdea's Implementation:

To test the population which is normally distributed, test and deal with outliers, test for homogeneity of variances.

Market Basket Analysis



Association rules are concocted by examining the data for frequent patterns and using the criteria support and confidence to distinguish the most significant associations. Support is an evidence of how frequently the entries appear in the database. Confidence reveals the number of times the statements have been established to be true.

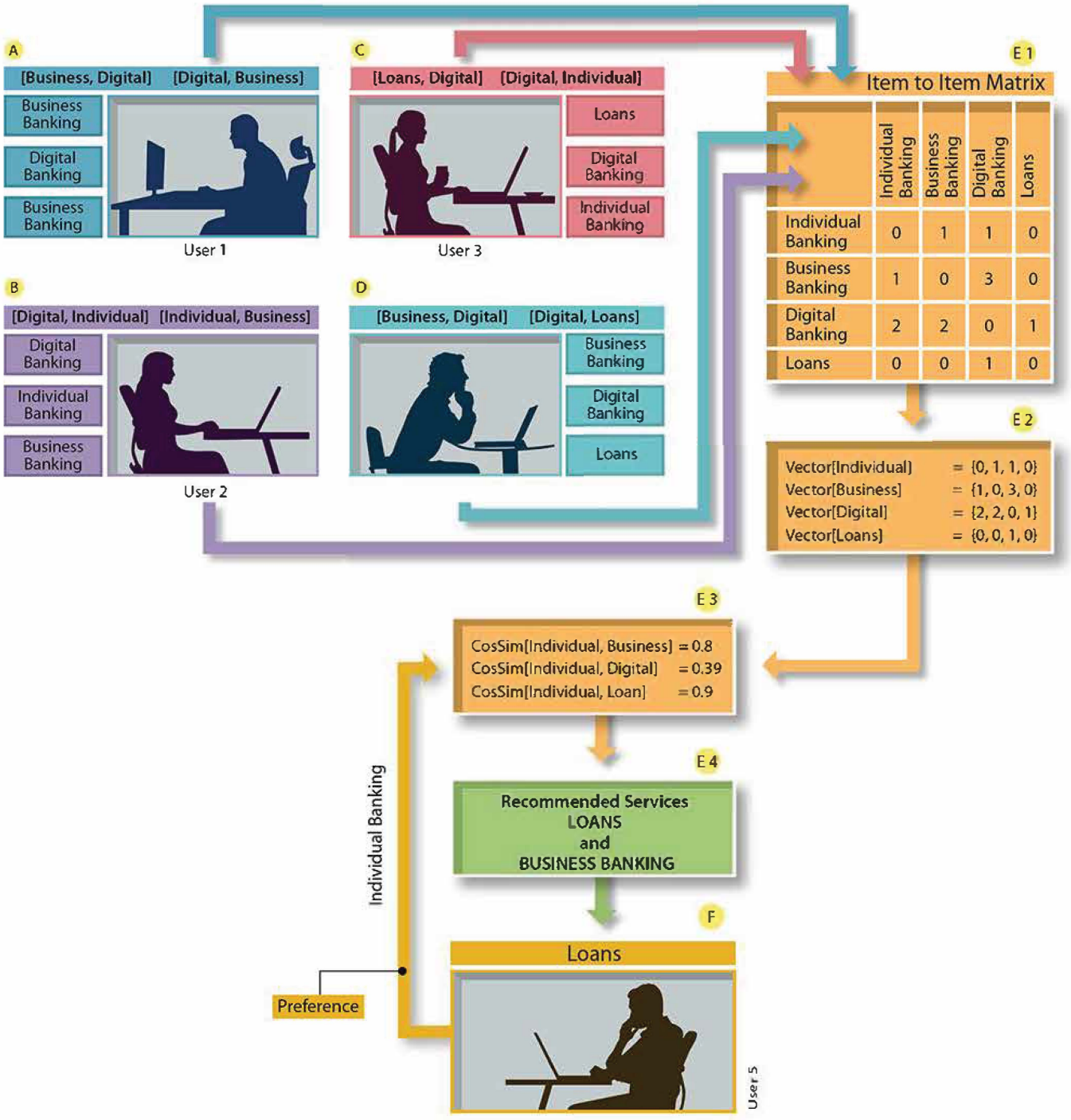
Goals:

Enable the user to find trends and interesting patterns in the data

BrandIdea's Implementation:

Predicting Consumer Behavior, Product Clustering, Market Basket Data Analysis

Collaborative Filtering



Banking service's preference histories of different users of a bank are retrieved and accumulated as an Item to Item Matrix as shown in E(1).

Matrix normalization is performed by applying the formula

$$itemNorm_{i,j} = \frac{item_{i,j}}{\sum_{k=1}^{n} \sqrt{item_{i,k}}}$$

Then compute Cosine similarities of an Item or a Service with the formula

$$CosSim_{A,B} = \sum_{i=1}^{n} A_{i,j} * B_{i,j}$$

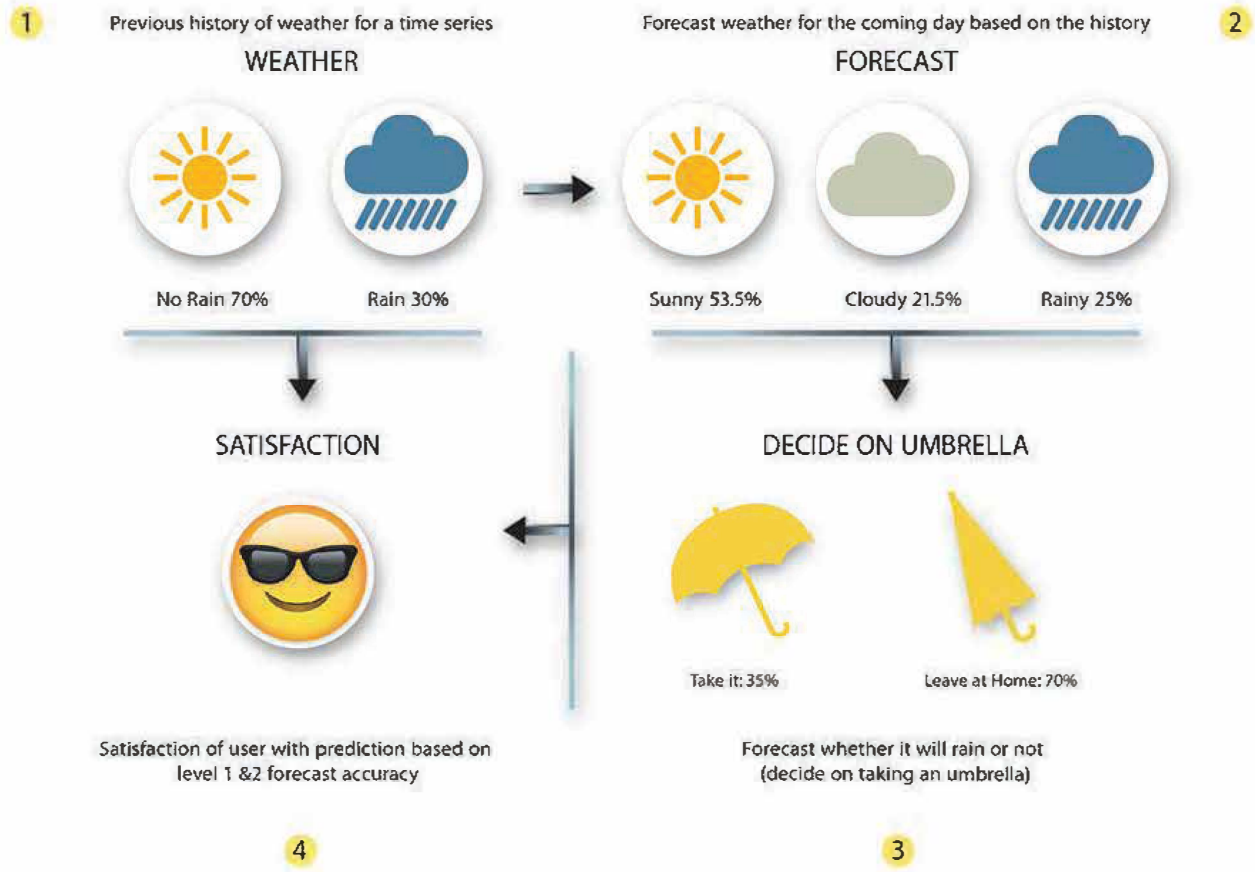
as shown in E(3)

Rank the calculated cosine similarity values and recommend (Cross sell / Up Sell) a service(s) to the user currently aspiring a service. (E 4)

Goals:
Find similarity among values.

BrandIdea's Implementation:
Market Basket Analysis.

Bayesian Network (Classifier)



Bayesian Networks are a graphical representation of structuring the probabilistic model, i.e., in ways the random variables may be dependent on each other. They intuitively represent domains with a causal structure, and the edges in the graph determine which variables directly influence which other variables. It can be equivalently regarded as a representation of factorized structure of the joint probability distribution, or as encoding a set of conditional independence hypotheses on the distribution.

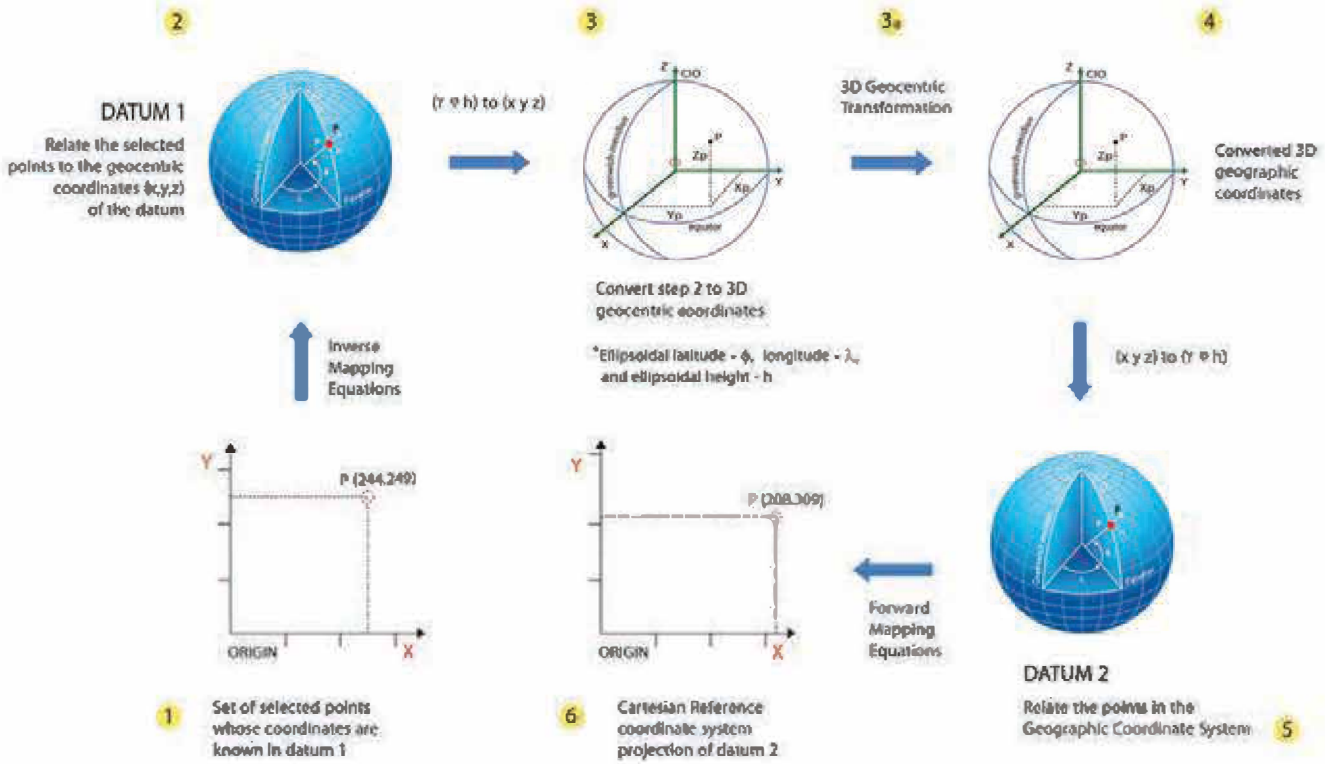
Goals:

Application of probability and statistics to Machine Learning

BrandIdea's Implementation:

Product Recognition using Image Processing

Coordinate Transformation

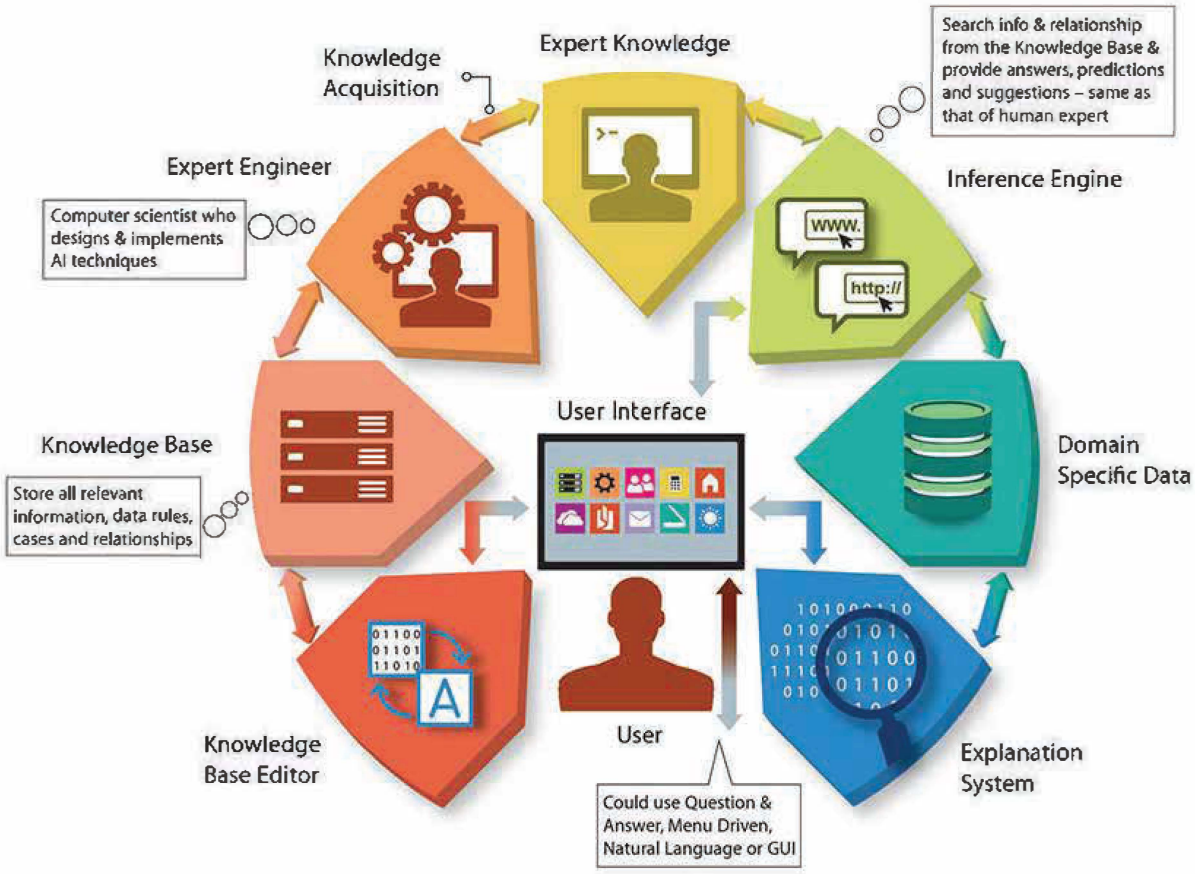


Map and GIS users encounter transformation from one two-dimensional coordinate system to another. This includes the transformation of polar coordinates into Cartesian map coordinates or the transformation from the 2D Cartesian (x, y) system of a particular map projection to another 2D Cartesian (x, y) system of a specified map projection.

Goals:

Identify rotations in the plane, Apply rotation formulas to figures on the coordinate plane.

Expert Systems



Expert Systems are computer systems that imitate the decision-making capabilities of a human expert. They are intended to resolve complicated problems by reasoning about knowledge,delineated primarily as if–then rules rather than through traditional procedural code.

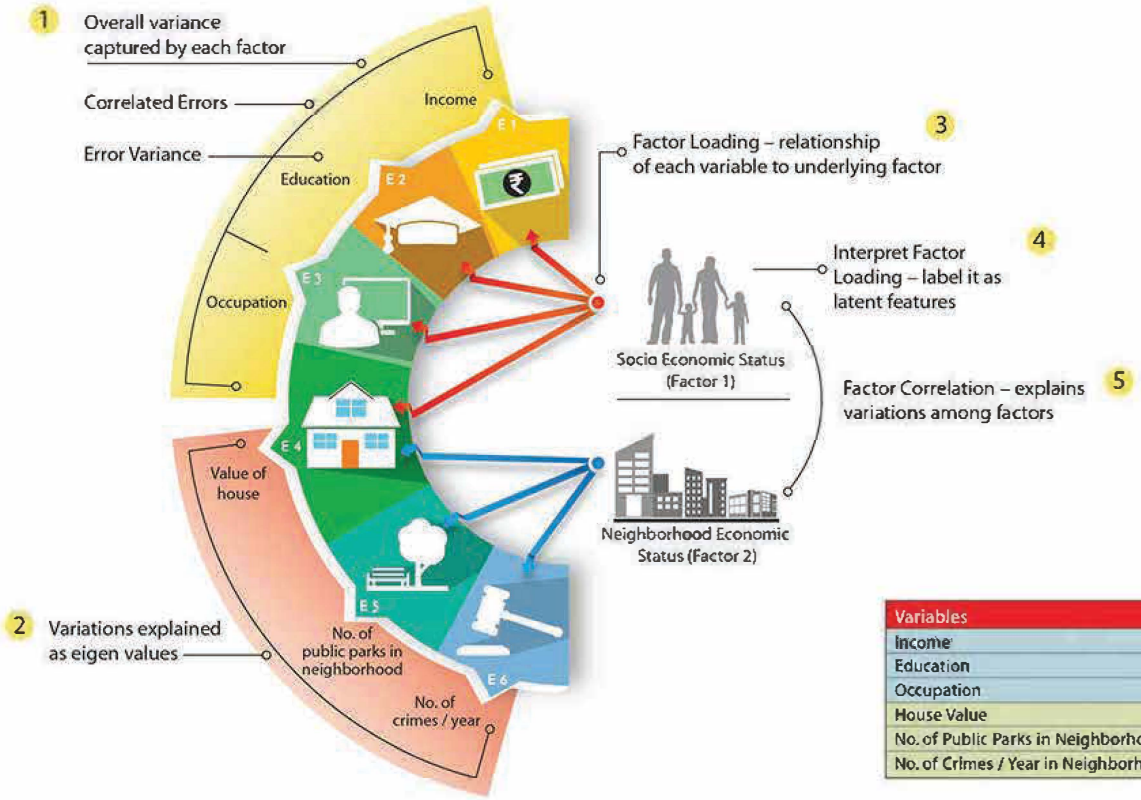
Goals:

Low error rate, steady response.

BrandIdea's Implementation:

SEC Affinity to Sales

Factor Analysis



Variables	Factor 1	Factor 2
Income	0.75	0.13
Education	0.65	0.24
Occupation	0.56	0.20
House Value	0.42	0.65
No. of Public Parks in Neighborhood	0.19	0.59
No. of Crimes / Year in Neighborhood	0.25	0.57

There exist the same number of factors as there are variables. Certain amount of the overall variance will be captured by each factor in the observed variables and the variations are always listed in the order of how much variation they explain as eigen values (is a of measure of variance in the observed variables). A factor with an eigen value ≥1 explains more variance than a single observed variable.

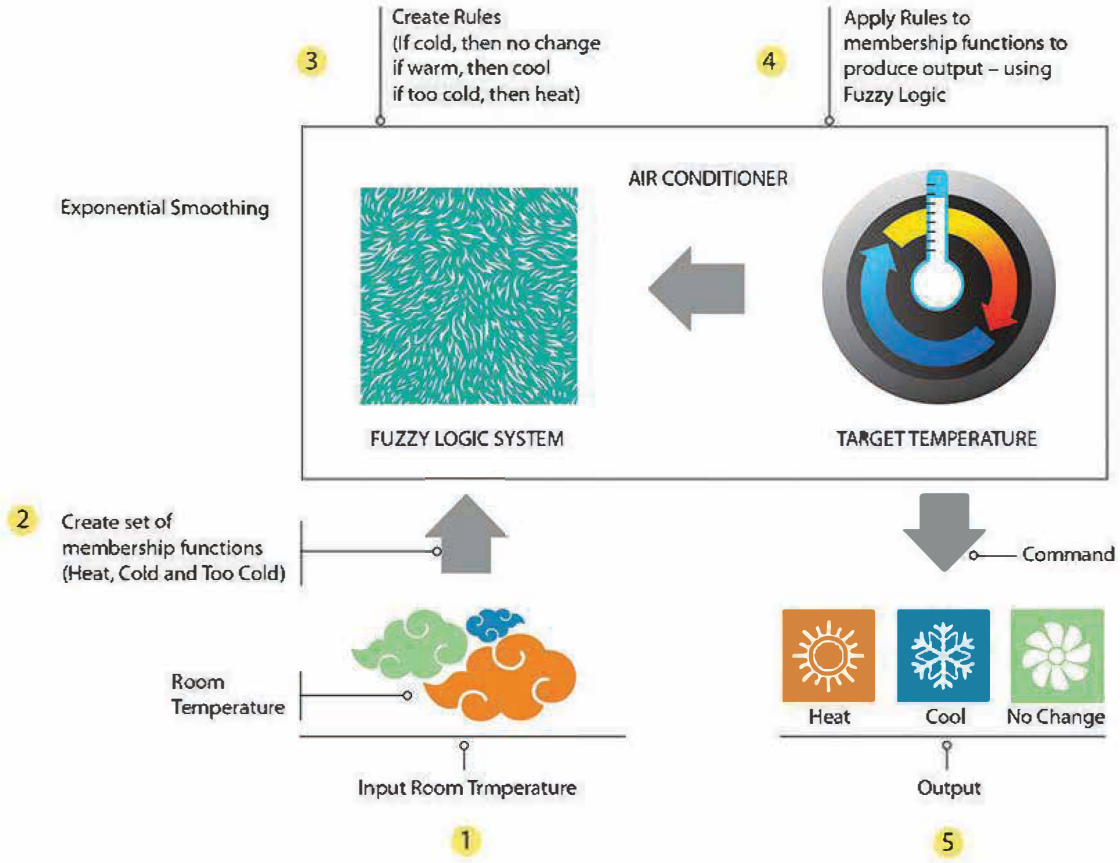
Goals:

To identify otherwise not-directly observable factors based on a set of observable variables

BrandIdea's Implementation:

Socio Economic Classification, SKU Recommendations

Fuzzy Logic



Accumulate data and create a number of partial truths, which are substantially aggregated into higher truths which in turn, when some thresholds are exceeded, stimulate certain further results.

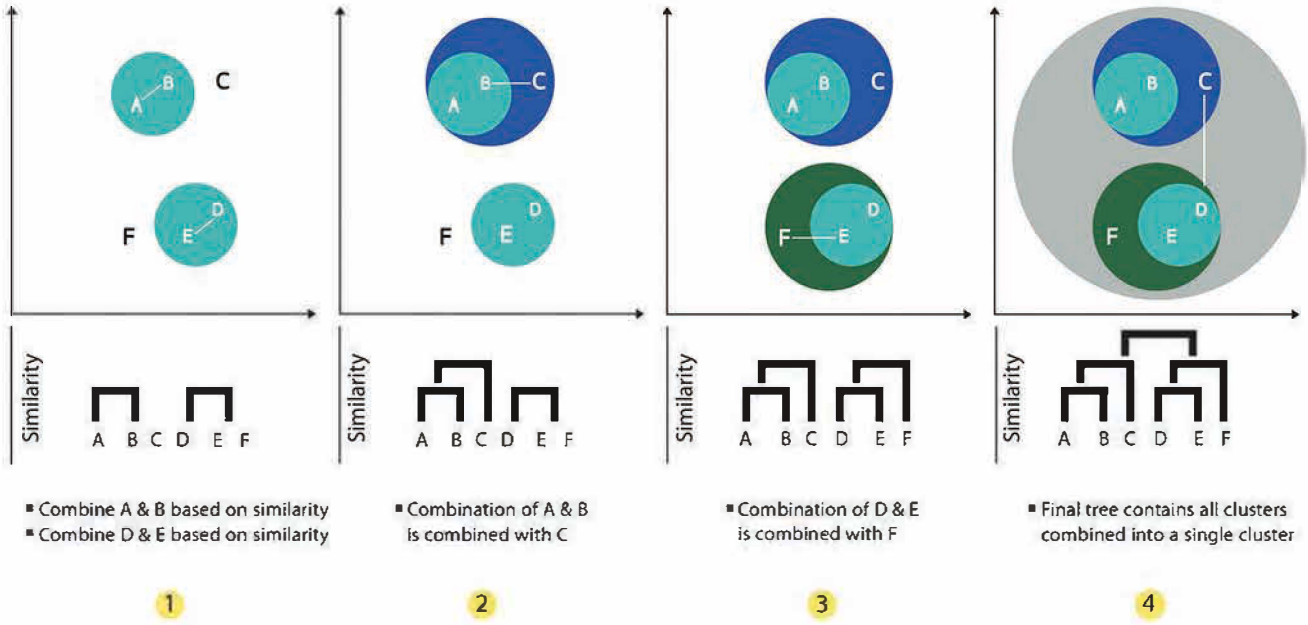
Goals:

More powerful approach to the study of intelligent systems.

BrandIdea's Implementation:

Image Recognition - counting objects from an image

Hierarchical Clustering



It constructs a binary tree of the data that consecutively combines related ensembles of points. The graphical representation of the resultant hierarchy is a tree-structured graph named dendrogram. Visualizing this tree serves an extremely valuable summary of the data.

Goals:

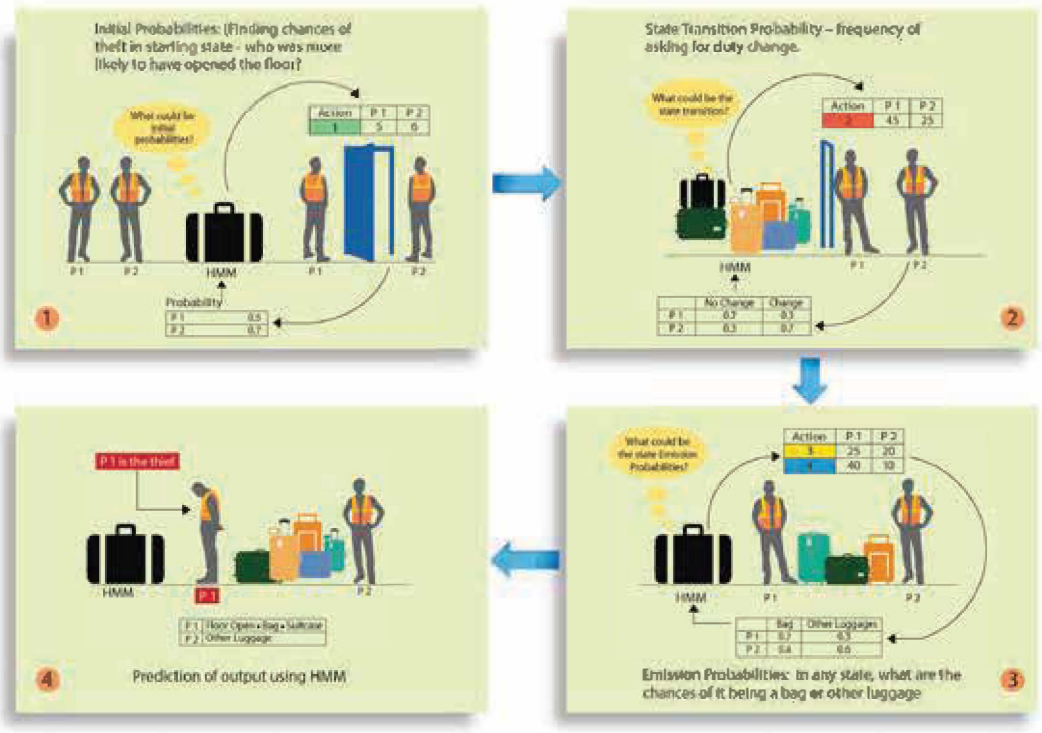
Degree of similarity or dissimilarity between the individual objects being clustered.

BrandIdea's Implementation:

Socio-Economic Classification

Hidden Markov Model (HMM)

Predicting whether Baggage Handler 1 or 2 Stole the Bag



Given Observations:

Action		Person 1 (P1)	Person 2 (P2)
1	Floor Opening Frequency	5 Days / Week	6 Days / Week
2	Duty Change Frequency %	45%	25%
3	Bags out of 50 Luggage's	25	20
4	No of Bags transported /day (on an average 50)	40	10

A finite set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, and not the state,which is visible to an external observer and therefore states are ``hidden'' to the outsiders.

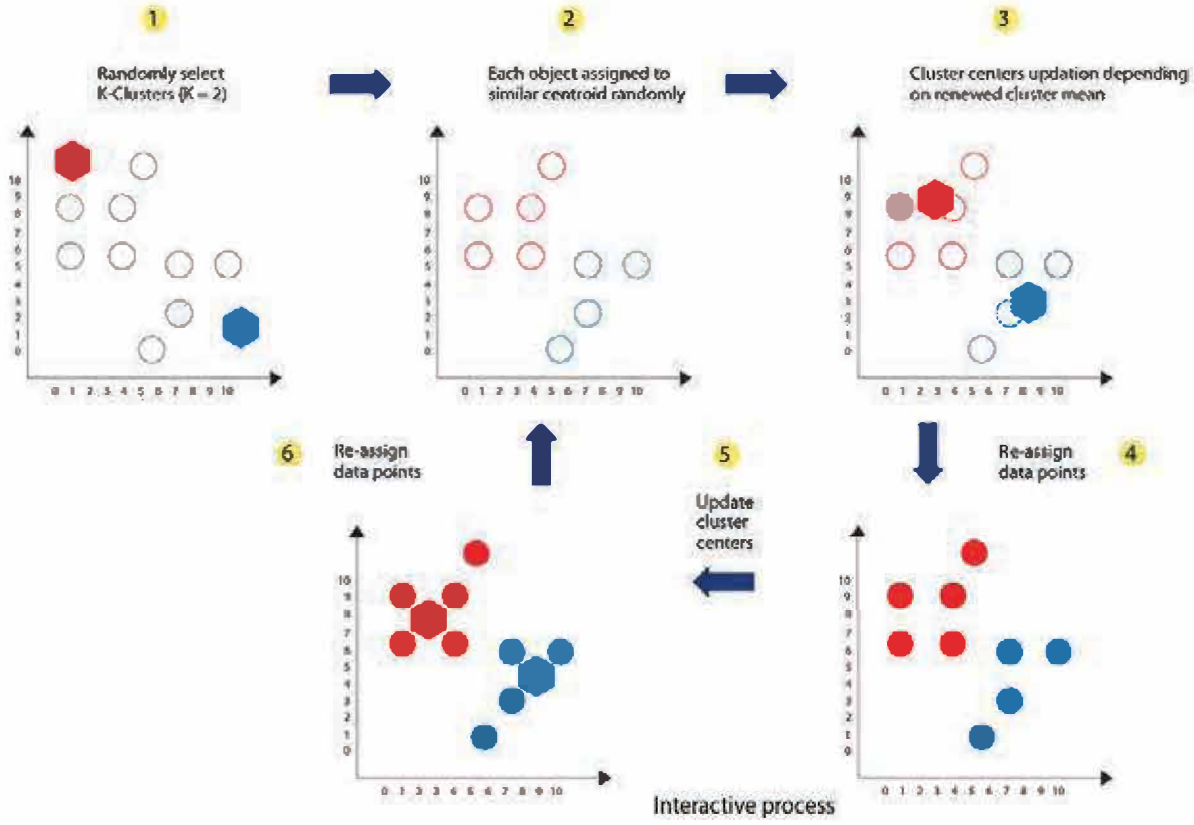
Goals:

Figure out the state sequence given the observed sequence of feature vectors

BrandIdea's Implementation:

Product Recognition using images

K-Means and X-Means Clustering



Clustering methods are used to group the data/observations into a few segments so that data within any segment are alike while data across segments are different. Cluster centroids are chosen randomly through a fixed number of K-clusters. The algorithm partitions the given data into K-clusters, each one having its own cluster membership and assigns each data point to the closest centroid. It then recomputes the centroid using current cluster association and if the clustering does not converge, the process will be repeated until a specified number of times. X-means clustering is a variation of K-means clustering that treats cluster allocations by repetitively attempting partition and keeping the optimal resultant splits, until some criterion is reached.

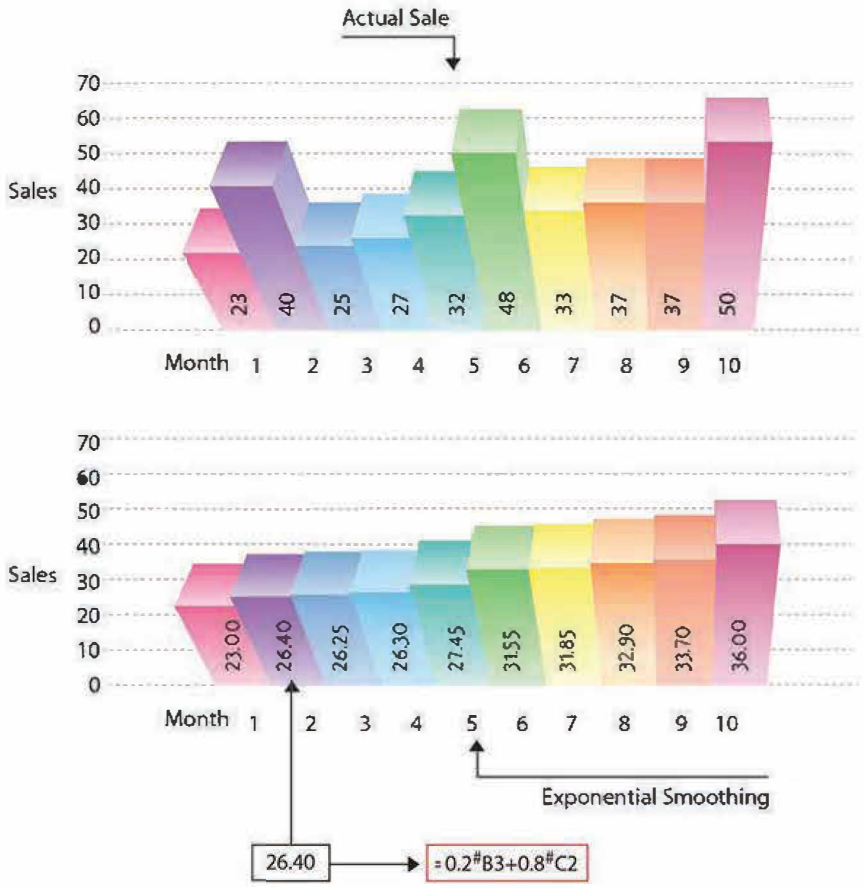
Goals:

Determine intrinsic grouping in a set of unlabeled data. Provide a fast and efficient way to cluster unstructured data, use of concurrency speeds up the process of model construction and the use of the Bayesian Information Criterion gives a mathematically sound measure of quality.

BrandIdea's Implementation:

Consumer Segmentation, Geo-demographic Segmentation, SEC Affinity, Progressive Index

Exponential Smoothing



Month	Actual Sales	Predicted Sales
1	23	23.00
2	40	26.40
3	25	26.25
4	27	26.30
5	32	27.45
6	48	31.55
7	33	31.85
8	37	32.90
9	37	33.70
10	50	36.00

Statistical technique for detecting significant changes in data by ignoring the irrelevant fluctuation in which, older data is given progressively-less relative importance (weight) while newer data is given progressively-greater weight. Furthermore, averaging is employed in making short-term forecasts.

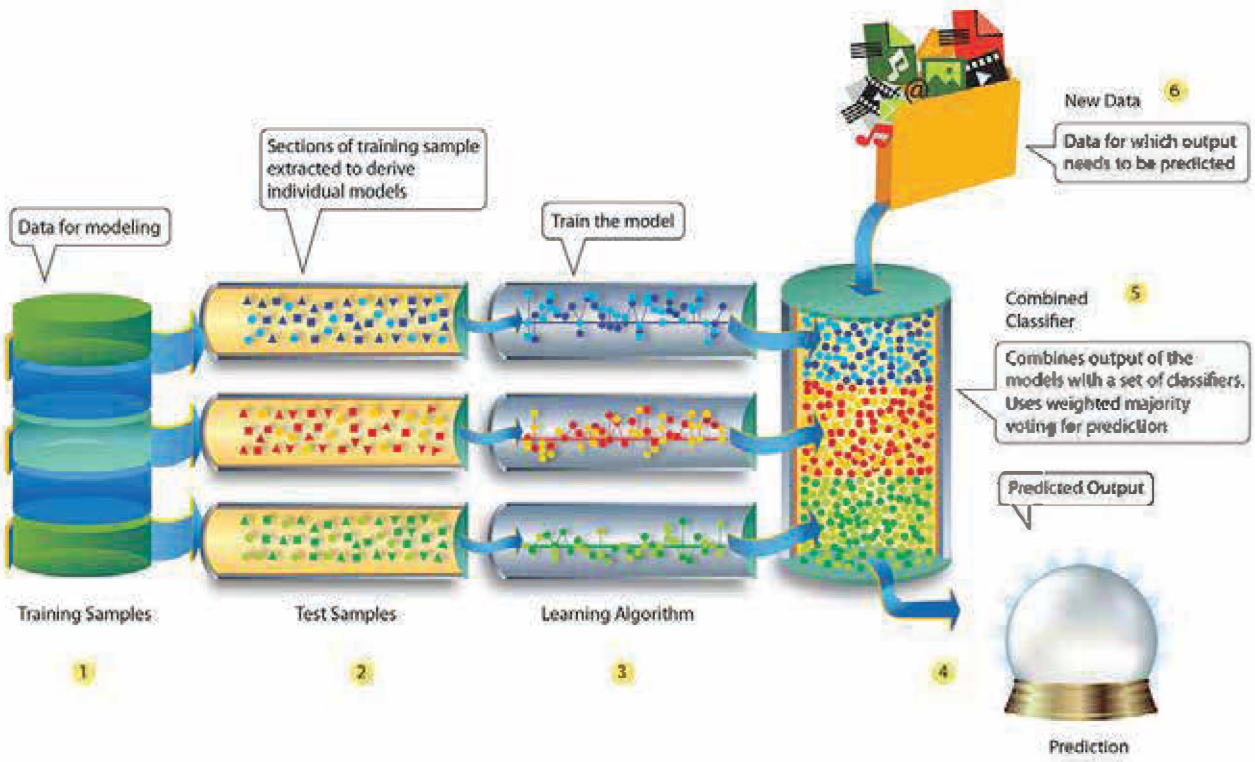
Goals:

Minimize errors

BrandIdea's Implementation:

Sales and Demand Forecasting

Ensemble Learning



A number of classifiers are strategically generated and combined to solve a particular computational intelligence problem. Given a set of training examples, a learning algorithm outputs a classifier which is an hypothesis about the true function F that generates the label values Y from the input sample values X. Given new X values, the classifier predicts the corresponding Y values.

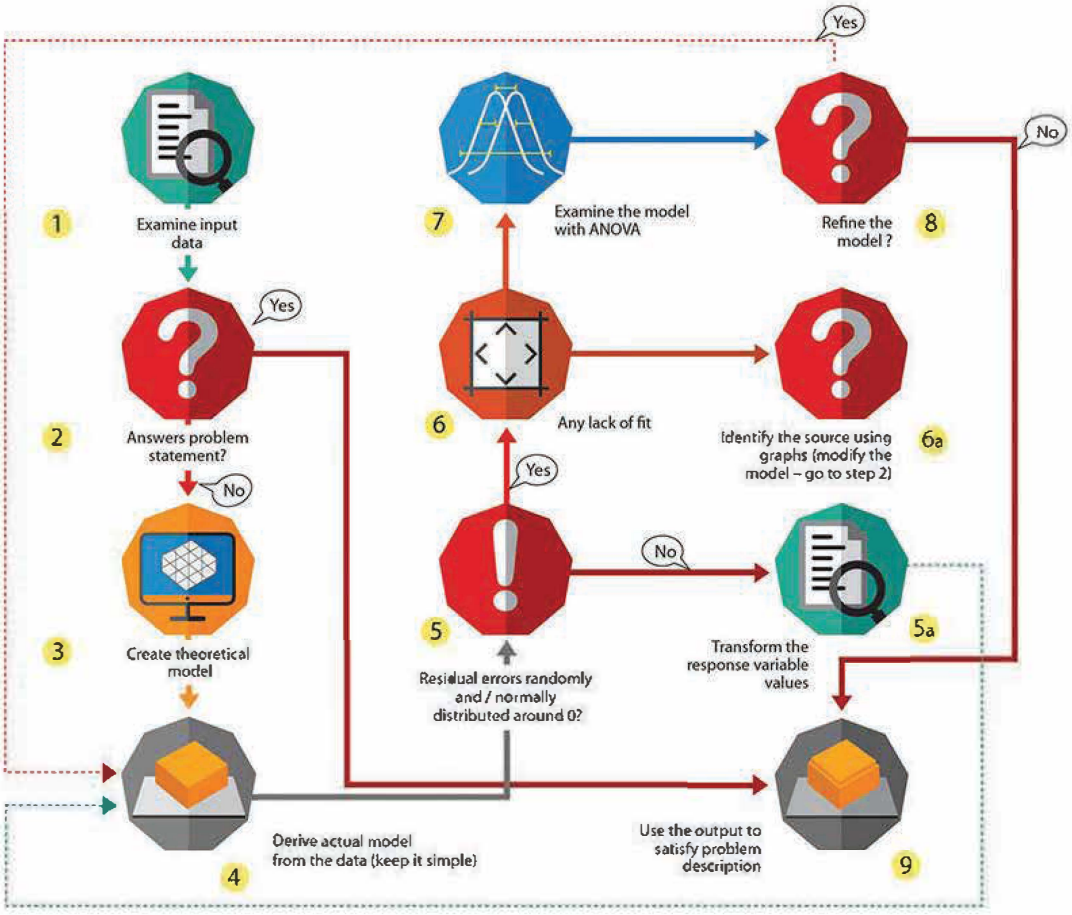
Goals:

Improve the (classification, prediction, function approximation, etc.) performance of a model, or reduce the likelihood of an unfortunate selection of a poor one.

BrandIdea's Implementation:

Handwriting recognition, Image processing, Image segmentation.

Design of Experiments



Methodical technique carried out under regulated environment to detect a hidden effect, test or establish a hypothesis, or demonstrate a known cause. When examining a process, experiments are often used to gauge which process inputs have a significant impact on the output, and what target level of those inputs is required to achieve a desired result.

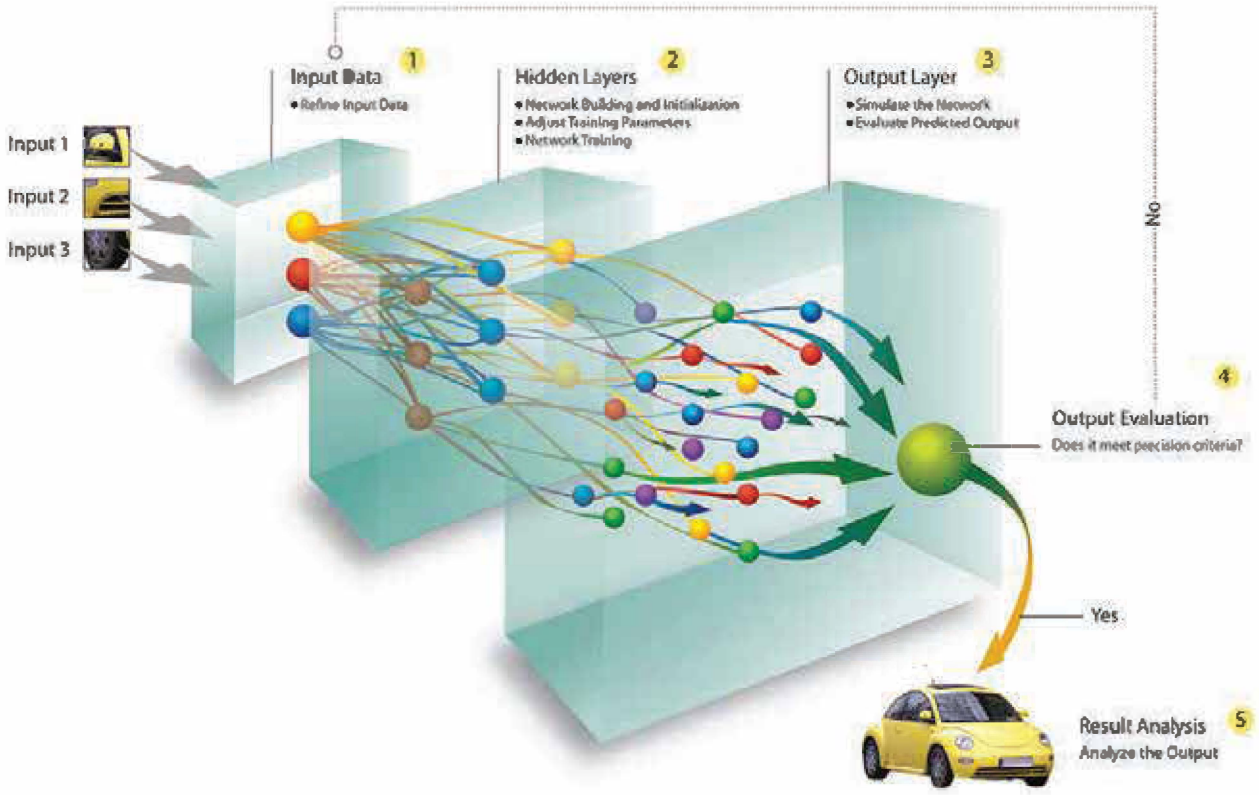
Goals:

Compare alternatives, identify significant inputs, achieve optimal output, reduce variability.

BrandIdea's Implementation:

Calculation of a sample statistic, computation of sampling distribution, and using the sampling distribution to draw inferences about statistical hypotheses - for SEC Affinity, Progressive Index etc.

Neural Networks



Neural Network has a collection of artificial neurons arranged in a sequence of layers, each of which links to other layers. The input units are intended to obtain different forms of information from the external world that the network attempts to learn from, recognize, or otherwise process. The output units sit on the opposite side of the network and signal how it responds to the information it has learned. One or more layers of hidden unit form the artificial brain. The edges are associated with a weight; more the weight, the more influence it has on the output. Every unit sums up the inputs it takes in and if the result exceeds a certain threshold value, the unit "fires" and activates the units it's associated with.

Goals:

Learning without programming, computation in parallel, finding network parameters automatically.

BrandIdea's Implementation:

SKU Recommendation System: RESKUR