DESCRIPTIVE VS. PREDICTIVE ANALYSIS

SYRACUSE UNIVERSITY
School of Information Studies

DESCRIPTION VS. PREDICTION

Predictive analysis

- Uses some variables to predict unknown or future values of other variables: classification, regression

Descriptive analysis

- Derives patterns (average, correlations, trends, clusters, and anomalies) that summarize the underlying relationships in data

Sometimes the difference between descriptive and predictive analysis is not black and white

- Trends, clusters, anomalies

SAMPLE DATA PROBLEM

The marketing department of a financial firm keeps records on customers, including demographic information and number of type of accounts. When launching a new product, such as a Personal Equity Plan (PEP), a direct mail piece advertising the product is sent to 500 existing customers, a sample of its 1 million customers, and a record kept as to whether each customer responded and bought the product. Based on this store of prior experience, the managers decide to use data mining techniques to build customer profile models, which will be used to decide which of the 1 million customers are likely to buy a PEP and thus should receive the advertisement.

http://facweb.cs.depaul.edu/mobasher/classes/csc478/Assignments/bank-data.html

DATA DESCRIPTION

ID	A UNIQUE IDENTIFICATION NUMBER
age	age of customer in years
sex	MALE/FEMALE
region	Inner city/rural/suburban/town
income	income of customer
married	Is the customer married (YES/NO)
children	number of children
car	Does the customer own a car (YES/NO)
save_acct	Does the customer have a saving account (YES/NO)
current_acct	Does the customer have a current account (YES/NO)
mortgage	Does the customer have a mortgage (YES/NO)
pep	Did the customer buy a PEP after the last mailing (YES/NO)

SAMPLE DATA

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	age	sex	region	income	married	children	car	save_act	current_a	mortgage	pep
2	ID12201	54	MALE	INNER_CI	26707.9	YES	1	NO	YES	YES	YES	YES
3	ID12202	27	FEMALE	INNER_CI	11604.4	YES	2	YES	YES	YES	NO	NO
4	ID12203	42	MALE	INNER_CI	15499.9	YES	0	YES	NO	YES	YES	YES
5	ID12204	43	MALE	TOWN	33088.5	NO	0	NO	YES	YES	YES	NO
6	ID12205	64	FEMALE	INNER_CI	34513.6	YES	1	NO	YES	YES	NO	YES
7	ID12206	43	MALE	TOWN	32395.5	YES	3	YES	YES	YES	NO	NO
8	ID12207	49	MALE	RURAL	46633	YES	0	YES	YES	NO	NO	NO
9	ID12208	23	MALE	INNER_CI	13039.9	YES	0	NO	NO	YES	NO	NO
10	ID12209	23	MALE	INNER_CI	12681.9	NO	0	NO	YES	YES	NO	YES
11	ID12210	30	FEMALE	INNER_CI	24031.5	YES	2	YES	YES	YES	YES	NO
12	ID12211	36	MALE	TOWN	37330.5	NO	2	NO	YES	YES	NO	YES
13	ID12212	34	MALE	INNER_CI	25333.2	YES	3	YES	NO	NO	YES	NO
14	ID12213	51	FEMALE	INNER_CI	37094.2	YES	0	YES	NO	YES	NO	NO
15	ID12214	36	MALE	TOWN	33630.6	NO	2	YES	YES	YES	NO	YES
16	ID12215	56	MALE	INNER_CI	43228.2	YES	1	YES	YES	YES	NO	YES
17	ID12216	54	FEMALE	INNER_CI	47796.8	YES	0	NO	YES	YES	NO	NO
18	ID12217	56	FEMALE	TOWN	21730.3	YES	2	NO	YES	NO	NO	NO
19	ID12218	26	MALE	INNER_CI	10044.1	YES	3	NO	YES	YES	YES	NO
20	ID12219	39	MALE	TOWN	17270.1	NO	0	YES	NO	NO	NO	YES
21	ID12220	64	FEMALE	RURAL	45765	YES	3	YES	YES	YES	NO	YES
22	ID12221	46	MALE	RURAL	29525.5	NO	2	NO	YES	NO	YES	NO
23	ID12222	62	FEMALE	RURAL	54863.8	YES	1	YES	YES	YES	NO	YES

DESCRIPTIVE ANALYSIS QUESTIONS

What are the average age and income of the customers?

Is there correlation between age and income?

How many people have 0, 1, 2, 3, or more children?

Is there correlation between the number of children and the decision to buy a PEP?

PREDICTIVE ANALYSIS QUESTIONS

Given a customer's demographic profile, what is the chance that he or she would buy the bank product PEP?