Dove Valentine Mailing Campaign

Course Name: Business Analytics Using Data Mining

Submitted by: *(Student names)*

Group Members (8A)
Harneet Chawla
Ankit Sobti
Kanika Miglani
Varghese Cherian
Saad Khan

Note: Considering our client is an FMCG, each technique mentioned below has been explained in detail ensuring thorough/easy understanding.

Executive summary

Business problem

We have been hired by our client, a reputed FMCG conglomerate, Unilever as data mining consultants. Our client has a range of products in the Personal Care Category that comprises of soaps etc. One of the brands that our client happens to own is the Dove brand of soap. For the first time the client is formulating a Valentine mai-in-coupon scheme to be rolled out in the month of February (next year). The scheme has the following business objectives:

1) Understand the customer profile of those customers who buy Dove soap.
2) Based on the customer profile understanding, predict for next year, new customers who are most likely to buy Dove.
3) Send out a mail-in-discount coupon to those respective customers of Hypermarket.
4) Client will be conducting a promotional campaign for which it will be incurring substantial costs thus it wants to ensure that next year when the campaign is rolled out, the coupons are sent out to customers who are most likely to avail them.
5) Client wants to increase customer loyalty towards Dove soap, considering the intense competition and low switching costs in the category. Our client believes that executing this promotion would boost sales of Dove in February.

Benefits

The mailing campaign that the client plans to execute will be targeted at new customers from various cities who are similar in demographics and other purchasing behavior to those purchasing Dove in the previous year. This will help expand the pie, create awareness and acquire new customers, thereby increasing the profits for the firms on Dove sales.

Data Mining Objective (Supervised Learning)

To develop a model whereby the client will be able to predict if a customer of Hypermarket in February is likely to purchase Dove. The dataset used for modeling is for the month of January in previous year of all customers (8,519) for all transactions during the month. Running this model on the customer & transaction data from January last year would enable us to target potential customers of Dove next year by looking at the new customer's January 2013 transaction data and based on that, predict which customers are most likely to buy Dove during February and send them the promotional mailing accordingly.

The rationale for using the month of January for the input data is:

1) Customers most frequent purchase data will help predict purchase behavior during February.
2) On average, customers visit once a month in the Personal Care category.

Data Preparation

Customer_No	DOB	Age	SEX	ENROLLMENT_SALE_DATE	Enrollment Duration	MARITAL_STATUS	CLEAN_CITY	CLEAN_STATE	Avg Basket UniqueCount(Sub_Deparment)	Avg Basket (UniqueCount(Sku_Number))	Avg. Basket Price-Jan	Avg. Basket Quantity - Jan	Number of Baskets- Jan	Dove	Dove Buy or No Buy
3000008858	9/20/1982	30	F	4/14/2009	3	N	Mumbai	Maharashtra	6	25.5	469.2575	7.1175	0.166666667	0	No Buy
3000039911	8/19/1965	47	M	5/3/2009	3	Y	Mumbai	Maharashtra	5.6	21.6	418.5341667	7.015833333	0.833333333	0	No Buy
3000051676	3/13/1964	48	M	5/17/2009	3	N	Mumbai	Maharashtra	8	29	333.425	7.943333333	0.083333333	0	No Buy
3000061766	8/12/1975	37	M	5/26/2009	3	NA	Mumbai	Maharashtra	5	21	663.4325	6.875	0.166666667	0	No Buy
3000065270	4/21/1978	34	F	5/31/2009	3	Y	Mumbai	Maharashtra	3	4	75.74	1.75	0.083333333	0	No Buy
3000069629	12/2/1977	35	F	6/4/2009	3	Y	Mumbai	Maharashtra	5	11	296.3175	2.655833333	0.083333333	0	No Buy
3000079917	9/24/1971	41	M	6/12/2009	3	N	Mumbai	Maharashtra	3.92	9.42	139.5808333	2.645833333	1	0	No Buy
3000117519	11/3/1985	27	M	9/6/2009	3	N	Ahemdabad	Gujarat	1	1.67	51.375	0.416666667	0.25	0	No Buy
3000117535	4/7/1968	44	M	9/7/2009	3	NA	Ahemdabad	Gujarat	4.13	27.27	487.8425	10.21416667	1.25	0	No Buy
3000117543	12/10/1981	31	M	9/7/2009	3	NA	Ahemdabad	Gujarat	5	26	296.415	6.471666667	0.083333333	0	No Buy
3000117584	1/7/1971	41	F	9/6/2009	3	NA	Ahemdabad	Gujarat	5	14.6	264.4116667	3.935833333	0.416666667	0	No Buy
3000117709	6/3/1958	54	M	9/6/2009	3	Y	Ahemdabad	Gujarat	3.88	16.5	285.1658333	6.051666667	0.666666667	0	No Buy
3000117717	7/4/1987	25	M	9/21/2009	3	N	Ahemdabad	Gujarat	8	17	353.4525	4.163333333	0.083333333	0	No Buy
3000117766	7/23/1960	52	F	9/26/2009	3	NA	Ahemdabad	Gujarat	5	23	680.9175	8.25	0.083333333	0	No Buy
3000117782	7/28/1970	42	F	9/21/2009	3	NA	Ahemdabad	Gujarat	5	22.75	207.9133333	4.641666667	0.333333333	0	No Buy

The input variables included in the data set were:

Customer Data
• Member No. • Age • Sex (Binary variable) • Marital Status (Binary variable) • City (Binary variable) • State (Binary variable) • Enrollment Duration

Basket Level Transactions
• Average Price (January) • Average Quantity (January) • Average SKU Count (January) • Average Sub Department Count (January) • Average number of baskets (January)

The source of the data was transactional level data for the month of January. The data was clubbed for each customer for the month of January. A combination of both customer demographic data & transactional data (basket level) was used. Since the campaign will be rolled out for the first time, the input variables list was exhaustive to ensure less error in the prediction models. The output variable was whether the customer will buy Dove or not buy (binary variable). For data preparation:

1) Demographics data was added to the customer level data (Age, Sex, Marital status, City, State). Enrollment duration was calculated based on the date from when the customer was issued the Hypermarket membership.

2) For the month of January (previous year); for all transactions of a customer, average price, average quantity, average SKU count & average sub department count was calculated based on the number of transactions each customer conducted during the month. If the customer visited more than once, average numbers of baskets were also calculated for each unique customer.

3) Missing values were found for each customer in sex, marital status, city & state. For sex & marital status a 3rd option (N/A) was introduced. For city & state, since the number of rows was very less, those customers were removed.

4) Dummy variables were created for categorical variables (sex, marital status, city, state). Later on the Naïve Bayes technique, 2 bins were created for each numerical variable (age, enrollment duration, average basket unique count, average sub departments, average quantity, average price & average number of baskets). Data was partitioned into training & validation set (50% each).

Benchmark

The simplest rule of classifying a customer is to ignore all predictors' information and classify based on the majority class. Naïve rule when applied to the customer data will predict the customer as **Dove No Buy**. As shown below, out of the total 4,259 customers in the training set, 87.5% are Dove No Buy and the remaining are Dove Buy. Thus based on Naïve rule, the customer will be classified as a Dove No Buy with an error of (for Dove Buy) of 12.5%.

Dove	**Count of Customer No**
Buy	533
No Buy	3,726
Grand Total	**4, 259**

The assumption is that based on the output binary variable (Dove Buy/No Buy), if the classification is done the result will be as shown above. Naïve rule is to be taken as a baseline for evaluating the performance of more complicated classifiers (see below). It can be safely assumed that generally any classifier which uses external predictor information can outperform Naïve rule.

Data Mining Solutions: Methods Applied & Appropriate Performance

a) K-NN Classification

More sophisticated classifiers such as K-NN will use predictor's data to classify a new customer as Dove Buy/No Buy. The technique will identify K observations in the training set which are similar to the new record that is to be classified. Using the similar K neighbors, the new record is classified assigning the new record to the predominant class among the K neighbors. K-NN does not make assumptions about the relationship between the Dove Buy/No Buy dependent variable and the predictors mentioned above, thus it's a nonparametric method as it does not involve estimation of parameters in an assured function. The method draws information from similarities between the predictor values of the records in the data set.

The distance between the records is measured based on the predictors using the Euclidean distance method. After computing the distance between the new data point (which is to be classified) and the existing records, a class is assigned to the new record based on the classes of the K neighbor's classification. When a K-NN classification was conducted on the partitioned data set (using 'Buy Dove' as the success class), a cut off value of 0.4 was assigned. **(Exhibit 1)**

The most intriguing question is to decide the number of K nearest neighbors to which the new record is to be compared to decide its classification. The simplest case is K=1, where we look for the record which is the closest neighbor however the misclassification error is high when K= 1 and error goes down as K is increased. The advantage of using K>1 is that higher values of K provide smoothing that reduces the risk of over fitting due to noise in the training data. However, if K is too high the method loses its ability to capture the local structure in the data set. If K = the total number of records, then we are down to Naïve rule again as the new record will again be assigned to the majority class in the training set; this will result in over smoothing. Thus there needs to be a balance between over fitting to the predictor information and over smoothing. For the model, the K value came out to be 9 nearest neighbors. The method to choose K is the value which has the highest classification performance. The training data is used to classify the records in the validation data and the errors are computed for various K values. The misclassification rate (of the validation set) is used to decide the number of K neighbors.

The advantages of using K-NN method of classification are its simplicity and lack of parametric assumptions and when the data set is large, it performs really well. However, if the training data set is extremely large (i.e. many predictors) finding the nearest neighbor can be tough and the number of records in the training set (to qualify as large); increase exponentially when the number of predictors increase. This is to avoid the *curse of dimensionality.*

K-NN model is performing better as observed above; the error rate for validation data is 6.76% which is less than the error %age of Naïve rule.

b) Naïve Bayes

It is a more sophisticated method than Naïve rule as it integrates the information in the predictors into the Naïve rule thus now the probability of a record belonging to either Dove Buy/No Buy is not only based on the prevalence of the majority class (No Buy) but also on the additional information in the predictors mentioned above. To convert numerical variables into categorical, 2 bins were created for each numerical variable (age, enrollment duration, average basket unique count, average sub departments, average quantity, average price & average number of baskets). Classification task is to estimate the probability of membership to each class (Dove Buy/No Buy) given a certain set of predictor variables i.e. conditional probability. Naïve Bayes assumes independence between the predictors which simplifies the calculation of conditional probabilities.

Although Naïve Bayes gives more accurate conditional probabilities vs. Exact Bayes rule, both often lead to the same classification for a cutoff of around 0.5 or above. Thus to ensure lower error in the model, the cutoff of 0.4 was chosen. Naïve Bayes assumes that the predictors are all mutually independent within each class. There is no information loss due to such assumption as what is important is not the exact probability estimate but only for the new record, the prediction in comparison to validation set.

Based on the results (with 0.4 as the cutoff probability of success i.e. Buy Dove), it is observed that the overall error was

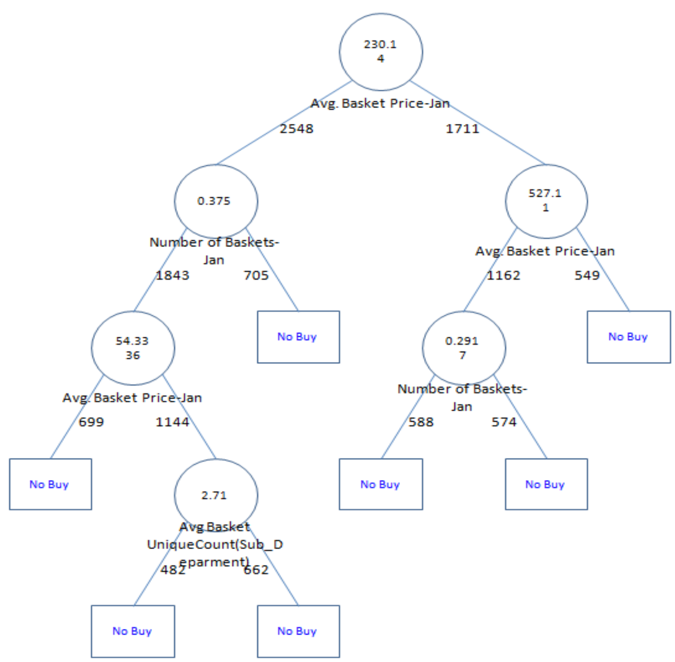
around 17.3% for validation set, however if the cut off value is increased to 0.6, the error %age for validation data goes down to 10.68%. Thus at this cutoff the model is performing better than Naïve rule. Still the model is not recommended. **(Exhibit 2)**

The advantages of Naïve Bayes include its simplicity, computational efficiency and good classification performance. However, it often requires a large data set and if a predictor category is not present (e.g. a third category was introduced except the predictors mentioned above when predicting for new records, Naïve Bayes will assume that the new record with that category of predictor has zero probability. Naïve Bayes technique will work for Hypermarket as the goal currently for the store is to classify customers into Dove Buy/No Buy. Had the goal been to estimate the exact probability of low variety, the method would've produced very biased results. Still based on the analysis above, the recommendation for Hypermarket is to avoid using Naïve Bayes model.

When the two techniques (K-NN & Naïve Bayes) are compared, looking at the error terms for validation data, we can safely assume that Naïve Bayes (with approx. 17.3% error for validation set) will result in a comparatively lesser accurate classification vs. K-NN, where the error for the validation data was 6.76% at the cutoff probability of 0.4 for success i.e. Buy Dove.

c) Classification Tree Technique

Classification tree model was developed on the same inputs as those used in K-NN and the following visualization of full classification tree was observed:



As is clear from the full tree diagram shown above, this model was not able to help understand the relationship between the predictors and the output classification too well. None of the decision nodes split the predictors into two homogenous classes. What was also observed was that the tree did not account for demographic information which might mean that sex, marital status and age did not hold any value in terms of this prediction.

Given the above scenario, it is not possible to choose a best split because none of the nodes are maximizing the reduction in the impurity of the node. Each split takes you to the same class, 'Dove No Buy' hence showing that these are all pure nodes with zero entropy and a zero Gini Index.

Since the output model does not help detect good predictors, this is probably not the best model for such a prediction. The reason why the results are as shown above is due to very limited 'Dove Buy' profiles in the data versus the 'Dove No Buy' **(Exhibit 3)**.This becomes clear when the training & validation errors are analyzed. As observed, model gets a 100% error in predicting the 'Buy' in both the training and validation set. This means the model does not have enough data to learn from and is probably using something similar to a majority Naive rule to classify all given cases as 'Dove No Buy'. Hence, this is probably not a very robust model to classify customers next year into Dove Buy/No Buy.

Conclusions - Limitations and Operational Recommendations

Following are the recommendations for the client:

1) Deploy K-NN Model for the Dove Valentine Mailing Campaign based on Dove Buy/No Buy customer classification.

2) Buy customer and January transaction data from Hypermarket for predicting future Dove purchases for the month of February. However next to next year, buy data for all customers including those (additional columns) to which the coupons were sent & those who availed so they can have a baseline on campaigns that have already run.

3) The model will be revised on the basis of buying behavior & demographics of those who availed coupon and who didn't i.e. the dependent variable will not be Dove Buy/No Buy but will be changed to Dove Valentine Coupon availed or not.

4) The current model is not viable to additional categories outside the Personal care or within Personal care for any other product as it might turn out that for any other product (even if its soap): a) buying cycle is different based on things such as SKU size etc. or b) customer profile based on demographics is not the same as for Dove.

5) Based on how much it would cost the firm to execute the coupon mailing, Unilever should conduct a cost-benefit Analysis (incremental sales vs. cost of mailing) to decide whether or not doing this promotional mailing campaign is even a good idea at the first place or not.

Exhibit 1. K-NN model output for Dove Buy/No Buy as the output variable.

Validation error log for different k

Value of k	% Error Training	% Error Validation	
1	0.02	11.97	
2	5.19	16.25	
3	5.33	7.82	
4	5.42	9.06	
5	5.92	6.88	
6	6.06	7.23	
7	6.01	6.53	
8	6.08	6.62	
9	6.18	6.34	<--- Best k
10	6.18	6.41	

Training Data scoring - Summary Report (for k=9)

Cut off Prob.Val. for Success (Updatable)	**0.4**	(Updating the value here will NOT update value in detailed report)

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	23	241
No Buy	20	3975

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	264	241	91.29
No Buy	3995	20	0.50
Overall	4259	261	6.13

Validation Data scoring - Summary Report (for k=9)

Cut off Prob.Val. for Success (Updatable)	**0.4**	(Updating the value here will NOT update value in detailed report)

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	7	262
No Buy	26	3964

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	269	262	97.40
No Buy	3990	26	0.65
Overall	4259	288	6.76

Exhibit 2. Naïve Bayes model output for Dove Buy/No Buy as the output variable.

Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	**0.4**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	97	167
No Buy	554	3441

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	264	167	63.26
No Buy	3995	554	13.87
Overall	4259	721	16.93

Validation Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	**0.4**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	92	177
No Buy	560	3430

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	269	177	65.80
No Buy	3990	560	14.04
Overall	4259	737	17.30

Exhibit 3. Classification Tree model output for Dove Buy/No Buy as the output variable.

Training Data scoring - Summary Report (Using Full Tree)

Cut off Prob.Val. for Success (Updatable)	**0.4**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	0	264
No Buy	0	3995

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	264	264	100.00

No Buy	3995	0	0.00
Overall	4259	264	6.20

Validation Data scoring - Summary Report (Using Best Pruned Tree)

Cut off Prob.Val. for Success (Updatable)	**0.4**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	0	269
No Buy	0	3990

Error Report			
Class	**# Cases**	**# Errors**	**% Error**
Buy	269	269	100.00
No Buy	3990	0	0.00
Overall	4259	269	6.32