Article can be accessed online at http://www.publishingindia.com

A Comparative Analysis of Support Vector Machines & Logistic Regression for Propensity Based Response Modeling

K. V. N. K. Prasad*, G.V.S.R. Anjaneyulu**

Abstract

Increasing cost of soliciting customers along with amplified efforts to improve the bottom-line amidst intense competition is driving the firms to rely on more cutting edge analytic methods by leveraging the knowledge of customer-base that is allowing the firms to engage better with customers by offering right product/service to right customer. Increased interest of the firms to engage better with their customers has evidently resulted into seeking answers to the key question: Why are customers likely to respond? in contrast to just seek answers for question: Who are likely to respond?This has resulted in developing propensity based response models that have become a center stage of marketing across customer life cycle. Propensity based response models are used to predict the probability of a customer or prospect responding to some offer or solicitation and also explain the drivers – why the customers are likely to respond. The output from these models will be used to segment markets, to design strategies, and to measure marketing performance.

In our present paper we will use support vector machines and Logistic Regression to build propensity based response models and evaluate their performance.

Keywords: Response Modeling, Propensity, Logistic Regression, Support Vector Machines

Introduction

A Propensity Model is a statistical scorecard that is used to predict the behaviour of customers or prospects base. Propensity models are extensively used in marketing arena to build list for solicitation and also act as a robust tool in creating tailored campaigns that are best received by customer. They help in developing analytical infrastructure that helps in identification of prospective opportunities and issues across the customer lifecycle, thus acting as a platform in understanding the dynamics of customer lifecycle.

Propensity models help in identification and generalisation of the "natural inclination or tendency" among the customer base for a given treatment. The identification and generalisation will help in understanding two important aspects – a) who are likely to respond when solicited? b) Why are the solicited customers likely to respond?The outcome of the propensity models will help to a larger extent in designing an optimal marketing strategy "reaching out to right customer with right product at right time through right channel at right price".

In our current paper we will use support vector machines and Logistic Regression to build propensity based response models and evaluate their performance.

Problem Statement

Logistic regression has been the workhorse for developing propensity models in marketing and risk management areas from long time, but for last few years there has been an enormous progress in statistical learning theory and machine learning – providing opportunity to use more robust and less restrictive algorithms to obtain much better results than traditional methods. In the present paper, we will use support vector machines and logistic regression to build propensity based response models and evaluate their performance and also highlight certain positive and negative aspects of the techniques under study.

[*] Department of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.
E-mail:kota.prasad.krishna@gmail.com

[**] Department of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

Literature Review

Increasing cost of marketing is driving companies to use analytics as corner stone to gain deep understanding of consumer behaviour. Amidst intense competition and dynamic shifts in consumer behaviour, the pressure of improving bottom lines has created enormous emphasis on propensity based response models. Using propensity based response models, one can identify a subset of customers who are likely to respond than others, and also generalise the need for response.

Companies use the knowledge of consumer behaviour to segment, to design marketing strategies, and to measure marketing performance (Schiffman & Kanuk, 1991). The use of SVM is rare in both CRM and customer response model, with exceptions (Viaene *et al*., 2001). Response models have been proven to be highly profitable tool in fine-tuning marketing strategies (Elsner *et al* ., 2004). SVMs have great generalisation ability and have strong performance when compared to traditional modeling approaches, but applications of SVMs in marketing are scant (Cui & Curry, 2005). The main purpose of response modeling is to improve future return on investment on marketing (Shin & Cho, 2006). Coussemet & Poel (2006) have used SVM in a newspaper subscription contest, and have proved that SVM have good generalisation ability when compared to logistic regression and random forest. Lately, companies are increasingly deluged by data and sophisticated data mining techniques are available to marketers (Ngai *et al*., 2009).

Support Vector Machines

The Support Vector Machines (SVMs) are supervised learning models with associated learning algorithms that are used for pattern recognition, classification and regression problems. Support Vector Machine (SVM) was introduced by Boser, Guyon & Vapnik in 1992. In 1995, soft margin classifier was introduced by Cortes & Vapnik and the algorithm was extended to problem of regression by Vapnik. Support Vector Machine (SVM) is generalisation of maximal margin classifier.

Maximal Margin Classifier

The maximal margin classifier is defined as the separating hyper plane for which the margin is largest-that is, it is the hyper plane that has the farthest minimum distance to the training observations 1.

Consider class of training observations $x_1 \dots \dots, x_n \in R^p$ and the respective associated class labels $y_1 \dots y_n \in \{-1, 1\}$. The maximal hyperplane is defined as the solution to the optimisation problem:

$$\mathit{maximize}\ \beta\beta_0, \ldots.\ \beta_{p\|\beta\| = 1}$$

Subject to $y_i\,(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M\ \forall\ i = 1, \ldots, n$

$$y_i\,(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M$$

The constraint ensures that each incoming observation will be on the correct side of the hyper plane at least at a distance M from the hyper plane and is called the margin. In above optimisation problem one tries to choose β_o, $\beta_1 \dots \ldots \ldots \beta_p$ to maximise the distance M.

Soft Margin Classifier

In many real world problems constructing linear separable classifiers is not always possible implying that maximum margin classifiers are no longer valid. In 1995, Corinna, Cortes & Vapnik suggested a modified maximum marginal classifier, in which a new classifier is achieved by relaxing the constraints a little to accommodate small amount of misclassification. The generalisation of the maximal margin classifier to the non-separable case is known as the support vector classifier. The soft margin classifier is defined as the solution to the optimisation problem:

$$\mathit{maximize}\ \beta_{0, \ldots\ \beta p,\ \in 1\ \cdots\ \in n\ \|\beta\| = 1}\ M$$

Subject to $y_i\,(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M\,(i - \in_i)$ $\forall\ i = 1, \ldots, n$

$$\text{where }\ \xi i \geq 0 \text{ and } \sum\nolimits_{i=1}^{n} \xi i < C$$

Where C is a nonnegative tuning parameter, M is the margin and one seeks to maximise the margin as much as possible and ξ_i are slack variables which measures of misclassification of the data x_1. If $\xi_i > 0$ then the i^th observation is on the wrong side of the margin, and we say that the i^th observation has violated the margin. If $\xi_i > 1$ then we conclude that i^th observation is on the wrong side of the hyper plane.

Support Vector Machine (SVM) representation

Given a training set of instance-label pairs $(X_i. y_i)$, i = 1, … l where $X_i \in R^n$ and {1, – 1}, let us assume that patterns with $y_i = 1$ belong to class 1 while with $y_i = -1$ belong to class 2. Then training support vector machines (SVM) require the solution for the following optimisation problem:

$$min_{w,b,\xi} \frac{1}{2} W^T W + C \sum_{i=1}^{l} \xi_i$$

$$\text{Subject to } y_i \left(W^T \phi(X_i) + b\right) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \ldots l$$

The above optimisation problem is most general Support Vector Machine (SVM) formulation allowing both non-separable and non-linear cases. The ξ_i are slack variables which measure misclassification of the data and C > 0 is the penalty parameter of the error term. In the above optimisation problem the training vectors X_i are mapped into a higher dimensional space by the function implicitly by employing kernel functions thus, Support Vector Machine(SVM) tries to find a linear separating hype plane with the maximal margin in this higher dimensional space.

Kernel Trick and Kernel Functions

In many real world problems finding a linearly separable hyper plane is not possible, to accommodate non-linearity kernels are used, and the input data are non-linearly mapped into high dimensional space. Consider a vector x in the input space can be represented as $\phi(x)$ in the higher dimensional space H, the mapping of data into higher dimensional space makes it possible to define a similarity measure on the basis of the dot product. If there is a kernel function K such that

$$K(x_1 x_2) = \phi(x_1).\ \phi(x_2)$$

then mapping is provided by

$$\langle x_1 x_2 \rangle \leftarrow K(x_1 x_2) = \langle \phi(x_1)\ .\ \phi(x_2) \rangle$$

Thus, if a kernel function K can be constructed, a classifier can be trained and used in the higher dimensional space without knowing the explicit functional form of mapping. In simple, the kernel trick enables one to find linearly separable hyper plane in feature space for the underlying training data, provided the underlying training data is not linearly separable in input space.A kernel that can be used to construct a SVM must satisfy Mercers condition. The kernel function plays a pivotal role in training SVM and its performance and is based on reproducing Kernel Hilbert Spaces.

Table 1 shows the list of little important kernel function used in practice.

Features Scaling

Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalisation. Scaling of input features will help in overcoming the numerical difficulties during training SVMs, since kernel values

Table 1: Important Kernel Function

Linear kernel	It is the simplest kernel function and is equivalent to principal component analysis. It is the inner product of input features plus an optional constant c	$k(x, y) = x^T y + c$
Polynomial kernel	Polynomial kernel are most popular method for non-linear modeling and are well suited where all the training data is normalized	$k(x, y) = (\alpha x^T y + c)d$ α is the slope c is constant and d is polynomial degree
Gaussian kernel	The Gaussian kernel is an example of radial base kernel. The adjustable parameter sigma plays an pivotal role in performance of SVM and should be carefully fine-tuned,	$K(x, y) = exp\left(-\frac{\lVert x - y \rVert^2}{2\sigma^2}\right)$
Exponential Kernel	It is also a radial base kernel; the exponential kernel is closely related to the Gaussian kernel, with only the square of the norm left out. The exponential kernel produces a piecewise linear solution and	$(x, y) = exp\left(-\frac{\lVert x - y \rVert}{2\sigma^2}\right)$

depend on the dot product of the input feature vectors, large feature values might cause some problems. Thus the input features are scaled between [-1 +1] or [0 1].

Logistic Regression

Consider the following simple linear regression setting with ‘r’ predictor and binary response variable

$$y_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_r x_r + \epsilon_i,\ i = 1,2, \ldots n$$

Where y_i is the binary response variable, $\epsilon_i \sim N\left(0, \sigma_\epsilon^2\right)$, and are independent.

Let P_i denote the probability that $y_i = 1$ and $x_i = x$

$$P_i = P(Y_i = 1 \mid X_i = X) = \frac{1}{(1+e^{-z})}$$

Where $Z = \beta_0 + \beta_1 x_i + \ldots + \beta_r x_r$

Or

$$\text{Logit}(p)\ \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \ldots + \beta_r x_r$$

The above equitation is called logistic regression: a statistical method in which we model the logit (p) in terms of explanatory variables that are available to modeler. It is non-linear in the parameters β_0, β_1,………… β_r The response probabilities are modeled by logistic distribution and estimating the parameters of the model constitutes fitting a logistic regression.

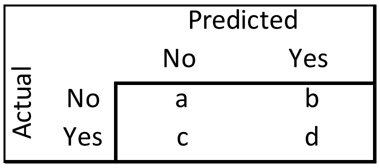
Performance Evaluation Measures

The following are various methods for assessing the discriminating ability of the trained model and in-time validation dataset.

Confusion matrix

A confusion matrix (also known as an error matrix) is appropriate when predicting a categorical target; confusion matrix helps one to evaluate the quality of the output of the classifier

Figure 1: Confusion Matrix



		Predicted	
		No	Yes
Actual	No	a	b
	Yes	c	d

Accuracy Ratio

It shows the proportion of the total number of predictions that were correctly classified.

$$AR = \frac{(a+d)}{(a+b+c+d)}$$

Precision

It is the proportion of the predicted positive cases that were correctly classified.

$$P = \frac{d}{b+d}$$

Kolmogorov-Smirnov (KS)

This measures the maximum vertical separation (deviation) between the cumulative distributions of goods and bads and is defined as follows

$$KS = MAX\left|F_G^{(s)} - F_B^{(s)}\right|$$

The higher the KS value the better is the models ability for separation.

Lift Curve

Lift is a measure of effectiveness of a predictive model and it is defined as the ratio between the results obtained with and with-out use of the predictive model. The lift curve will help analyse the amount of true responders discriminated in each subset. This is extremely helpful for any marketing team for making optimum decisions.

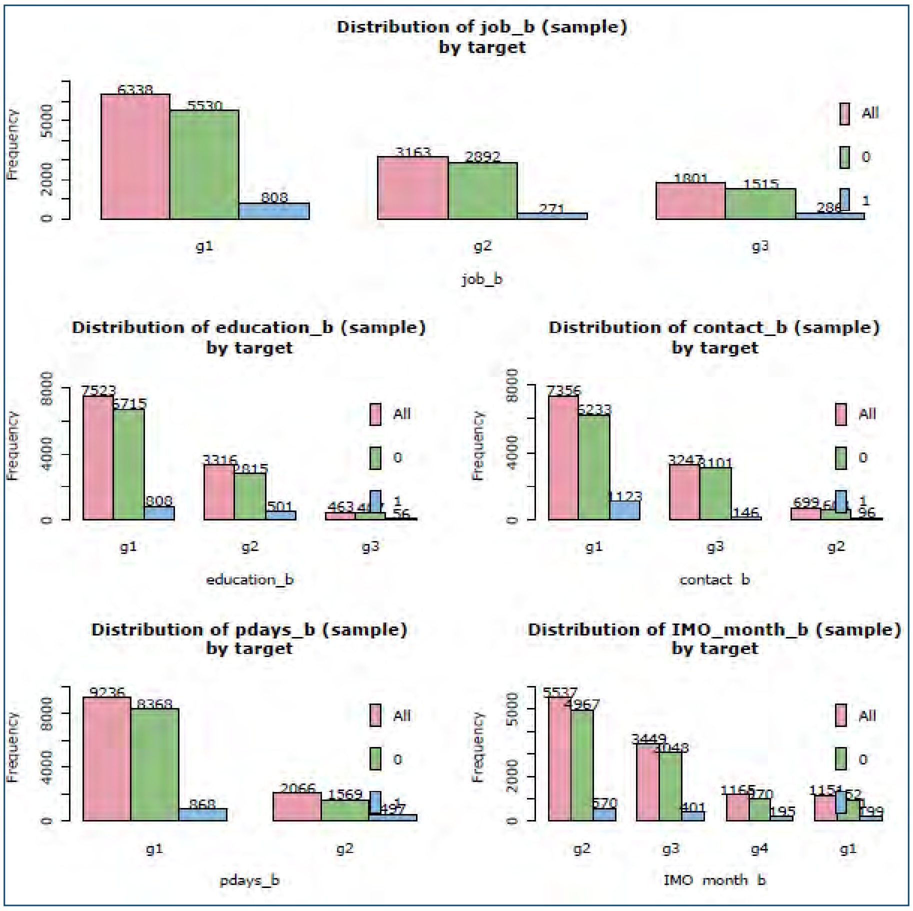
Data Description

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be (or not) subscribed. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Dataset Preparation

The dataset consist of 45,211 instances. We have used random sampling to construct a dataset of 22,605

Figrure 2: Distribution of Variables



instances.This dataset was used for training and validating logistic regression and support vector machines.

Data Cleaning

As a part of data cleaning exercise, all the prospective variables are subjected to univariate and bivariate analysis; the missing values for numeric variables are imputed with the median and in case of the discreet variables the missing values are imputed by mode.

Variable Transformations

All the numerical variables are scaled between [0, 1] using min-max scaling method, the categorical variables are binned into smaller groups based on the response rates.

Basic Statistics

Table 2 summarizes the basic statistics for the numerical variables (raw andscaled).

Logistic Regression vs Support Vector Machine Performance comparisons

Confusion Matrix

The confusion matrix results for the classifier obtained by both models are shown in Figure 3.

The comparison indicates that SVM are marginally better in development and in validation they perform equivalently well with the logistic model

Rank-ordering

The scores obtained by the classifier (Logistic and SVM) are used to rank-order consumers -how likely they are to respond when solicited. The following table provides the

Table 2: Basic Statistics for the Numerical Variables (raw andscaled)

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
age	22605	40.95067	10.62267	925690	18	95
balance	22605	1343	2935	30347863	-3058	102127
day	22605	15.79889	8.28949	357134	1	31
duration	22605	257.4868	255.6189	5820489	1	4918
campaign	22605	2.74059	3.07033	61951	1	63
pdays	22605	40.47184	100.9995	914866	-1	871
previous	22605	0.56912	1.88134	12865	0	55
rand	22605	0.24761	0.14432	5597	2.74E-05	0.5005
target	22605	0.11878	0.32354	2685	0	1
age_scaled	22605	0.29806	0.13796	6738	0	1
balance_scaled	22605	0.08499	0.02664	1921	0.04504	1
duration_scaled	22605	0.05236	0.05198	1184	0.000203	1
previous_scaled	22605	-0.00207	0.00684	-46.7818	-0.2	0

Figrure 3: Confusion Matrix Results for the Classifier

SVM model

		Predicted		
		Yes	No	
Actual	Yes	9799	138	9937
	No	1013	352	1365
		10812	490	

Accuracy	89.82%
Error	10.18%
Recall	25.79%
precision	71.84%

Logistic model

		Predicted		
		Yes	No	
Actual	Yes	9749	188	9937
	No	1068	297	1365
		10817	485	

Accuracy	88.89%
Error	11.11%
Recall	21.76%
precision	61.24%

SVM model

		Predicted		
		Yes	No	
Actual	Yes	9813	170	9983
	No	1044	276	1320
		10857	446	

Accuracy	89.26%
Error	10.74%
Recall	20.91%
precision	61.88%

Logistic model

		Predicted		
		Yes	No	
Actual	Yes	9781	202	9983
	No	1037	283	1320
		10818	485	

Accuracy	89.04%
Error	10.96%
Recall	21.44%
precision	58.35%

rank-ordering ability of the models in train and test data. In training and test the SVM classifier performs better than the logistic classifier.

The maximum KS for SVM occurs in 2nddecile in train and test , while the maximum KS for logistic regression classifier occurs in 3rd decile in train and test.

ROC Curves

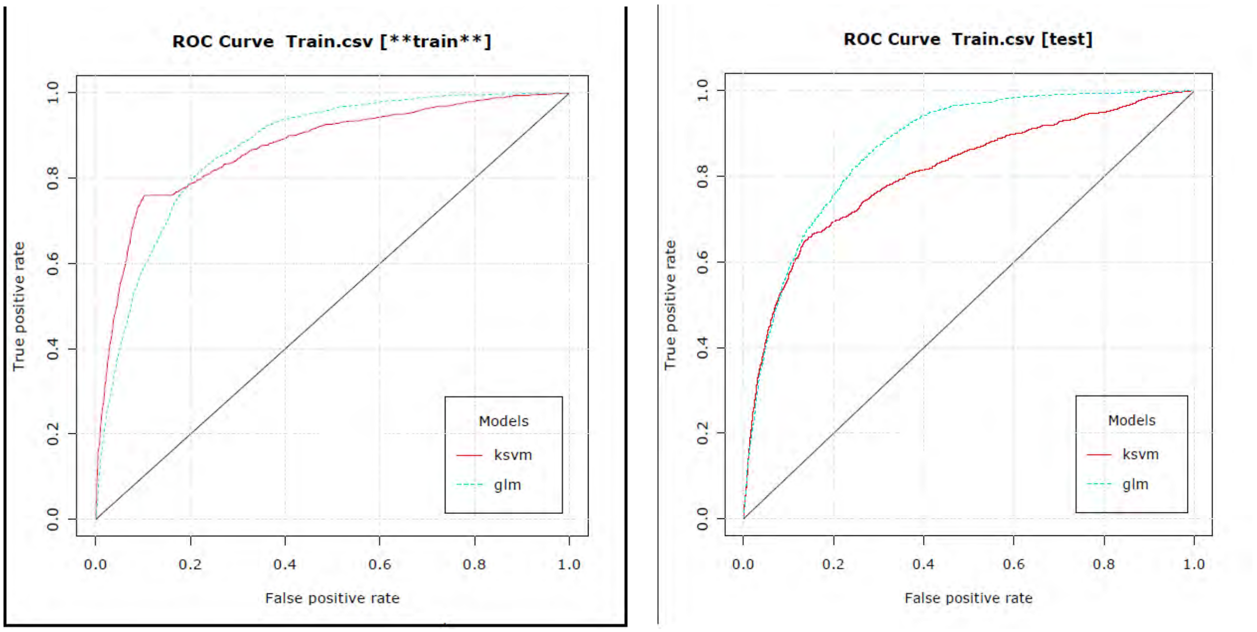
The ROC obtained by the classifier constructed from both the model are populated below, the SVM perform better in marginally better in the training, while in the test it underperform when compared to the logistic regression.

Table 3: Scores Obtained by the Classifier (Logistic and SVM)

	Training								Test							
Decile	Leads	Logistic	SVM	Logistic _ Response Rate	SVM_Response Rate	Logistic_KS_Training	SVM_KS_Training		Leads	Logistic	SVM	Logistic _ Response Rate	SVM_Response Rate	Logistic_KS_Test	SVM_KS_Test
0	**1,130**	575	685	50.88%	60.62%	36.54%	45.70%	**1,130**	565	583	50.00%	51.59%	37.14%	38.69%
1	**1,130**	334	353	29.56%	31.24%	53.00%	63.75%	**1,130**	318	278	28.14%	24.60%	53.10%	51.21%
2	**1,130**	216	63	19.12%	5.58%	59.62%	57.62%	**1,131**	176	85	15.56%	7.52%	56.87%	47.18%
3	**1,131**	101	84	8.93%	7.43%	56.66%	53.24%	**1,130**	129	102	11.42%	9.03%	56.61%	44.60%
4	**1,130**	69	56	6.11%	4.96%	51.04%	46.54%	**1,130**	74	63	6.55%	5.58%	51.64%	38.69%
5	**1,130**	30	35	2.65%	3.10%	42.16%	38.08%	**1,131**	24	58	2.12%	5.13%	42.37%	32.34%
6	**1,131**	20	25	1.77%	2.21%	32.45%	28.78%	**1,130**	19	44	1.68%	3.89%	32.68%	24.79%
7	**1,130**	14	32	1.24%	2.83%	22.24%	20.08%	**1,131**	8	38	0.71%	3.36%	22.04%	16.72%
8	**1,130**	3	24	0.27%	2.12%	11.12%	10.71%	**1,130**	4	43	0.35%	3.81%	11.06%	9.09%
9	**1,130**	3	8	0.27%	0.71%	0.00%	0.00%	**1,130**	3	26	0.27%	2.30%	0.00%	0.00%
Total	**11,302**	1,365	1,365	12.08%	12.08%			**11,303**	1,320	1,320	11.68%	11.68%		

Figrure 4: ROC Curves

ROC Curve		
	SVM	Logistic
Train	87.15%	87.14%
Test	80.76%	86.87%



Lift Charts

The lift obtained by the classifier constructed from both the model are populated below, the SVM perform better in better in the training, while on the test it is performs equivalent when compared to the logistic regression.

Propensity Profile

The output from the SVM will provide no means to profile the prospective lead list, on the other hand theoutput from the logistic model can be use to create customer profiles as shown in Table 4.

Figrure 5: Lift Charts

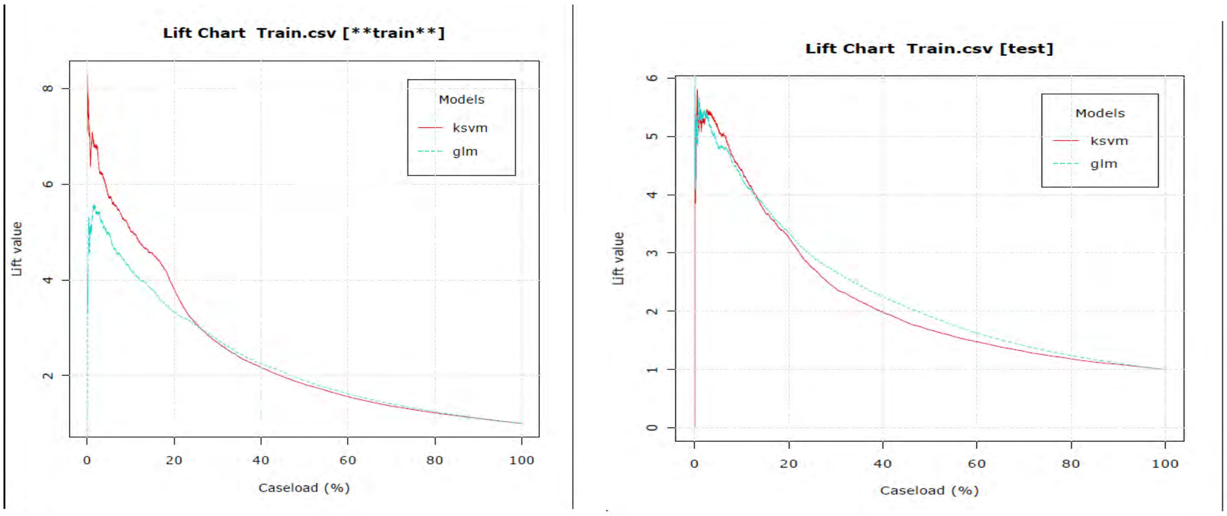


Table 4: Propensity Profile

Bins	Bin's Description	Estimate	Std. Error	z value	Pr(>|z|)	Sig
(Intercept)		-3.47017	0.17144	-20.241	< 2e-16	***
marital	01: Married	-0.2561	0.10625	-2.41	0.015942	*
	02: Single	0.06415	0.11195	0.573	0.566659	
	03: divorced*	0				
default	01: Yes	-0.31495	0.34155	-0.922	0.356471	
	02: No*	0				
loan	01: Yes	-0.62362	0.11023	-5.657	1.54E-08	***
	02: No*	0				
job	01: 'admin.','management','self-employed','technician', 'unemployed','unknown'	0				
	02: 'blue-collar','entrepreneur','hou	-0.34072	0.09203	-3.702	0.000214	***
	03: Other	0.32656	0.09044	3.611	0.000305	***
education	01: 'primary', 'secondary'	0				
	02: 'tertiary'	0.23425	0.0775	3.022	0.002508	**
	03: 'unknown'	0.07789	0.17258	0.451	0.651752	
contact	01:'cellular'	0				
	02:'telephone'	-0.11832	0.1371	-0.863	0.388134	
	03:''unknown'	-1.27635	0.12128	-10.524	< 2e-16	***
pdays	01: Not contacted earlier	0				
	02: Contacted	0.89447	0.09551	9.365	< 2e-16	***
month	01: Jan,Feb, Mar	0				
	02: Apr, May, Jun	-0.10571	0.10622	-0.995	0.319628	
	03: Jul, Aug,Sep	-0.27022	0.11022	-2.452	0.014223	*
	04:Oct, Nov, Dec	-0.01262	0.12531	-0.101	0.919761	
balance_scaled		4.08877	1.09629	3.73	0.000192	***
duration_scaled		20.27482	0.59724	33.947	< 2e-16	***
previous_scaled		-10.41849	4.92502	-2.115	0.034394	*

The prospects who are Single, not defaulted earlier, have no previous loans, having tertiary education, contacted previously on mobiles are more likely to respond to campaigns

SVMs vs Logistic Regression: Pros & Cons

The current study indicates that SVMs perform better than logistic regression on the performance evaluation parameters that we have used to evaluate the classifiers, but still logistic regression continues to be work-horse in response modeling due to the following reasons

Pros

- The SVMs have good generalisation in both in-samples, hold-out and out-of-sample by choosing appropriate parameters.
- The concept of kernels encompasses non-linear transformations, so no prior assumption is made about the functional form of the transformation.
- SVMs are robust to outliers.

Cons

- Unlike logistic regression - SVMs is the lack of transparency of results.
- The choice of kernel is another shortcoming
- Unlike logistic regression - SVMs has high algorithmic complexity and requires extensive memory requirements.

Conclusion

The current study indicates that SVMs perform better than logistic regression on the performance evaluation parameters that we have used to evaluate the classifiers. But lack of transparency of results, extensive memory requirements, issues in implementation in production system for regular scoring, no comparable standards to monitor SVMs on an ongoing basis to track performance make's logistic regression models, the preeminent choice due to extensive theory around regression framework, ease of understanding and implementation, sensible results along with actionable insights could be used to identify generic and niche segments that enable the marketing teams to develop more tailored campaigns.

References

Athanassopoulos, A. D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Resear ch,* 47, 191–207.

Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167.

Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah? SIGKDD Explor. Newsl. 2, 1–13

Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science,* 24, 595–615.

Cui, G., Wong, M. L., & Lui, H. K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science,* 52(4), 597–612.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines, Software Retrieved from http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Elsner, R., Krafft, M., & Huchzermeier, A. (2004). Optimizing Rhenania's Direct Marketing Business through dynamic Multilevel Modeling (DMLM) in a Multicatalog-Brand Environment. *Marketing Science*, 23(2) 192-206.

Hosmer, D.W., & Lemeshow, S. (1989). *Applied logistic regression,* New York: John Wiley & Sons, Inc

Moro, S., Laureano, R., & Cortez, P. (2011).Using data mining for bank direct marketing: An application of the CRISP-DM methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.

Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Syst Appl, 36, 2592–2602. doi:10.1016/j.eswa.2008.02.021

Kim, D., Lee, H., & Cho, S. (2008). Response modelling with support vector regression. *Expert Systems with Applications,* 34, 1102–1108.

Hsu, C., Chang, C., & Lin, C. (2010). *A practical guide to support vector classification*. Department of Computer Science and Information Engineering, National Taiwan University.

Viaene, S., Baesens, B., Van Gestel, T., Suykens, J., Van den Poel, D., Vanthienen, J., De Moor, B., & Dedene, G. (2001). Knowledge discovery in a direct marketing case using least squares support vector machines. *International Journal of Intelligent System*, 16, 1023–1036

Vapnik, V. (2000). *The Nature of Statistical Learning Theory*, Springer, New York.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). The Elements of Statistical Learning. Springer, 2001.

Schhiffman J. B., & Kanuk, L. L. (1997). Consumer Behavior published by Prentice Hall Sixth edition, 446.

Shin, H. J., & Cho, S. (2006). Response modeling with support vector machine. *Expert Systems with Application,* 30(4), 746-760.

Williams, G. J. (2011). Data mining with Rattle and R: The art of excavating data for knowledge discovery, Use R!, Springer.