Predicting Customer Online Shopping Adoption - an Evaluation of Data Mining and Market Modelling Approaches

Sven F. Crone
Department of Management Science
Lancaster University, UK
s.crone@lancaster.ac.uk

Didier Soopramanien
Department of Management Science
Lancaster University, UK
d.soopramanien@lancaster.ac.uk

Abstract - **Accurate prediction of shopping channel preferences has become an important issue for retailers seeking to maximize customer loyalty. In data mining, novel approaches such as neural networks (NN) have been proposed to predict the probability of class memberships in addition to statistical methods from marketing modelling. However, Data Mining suggests new approaches to data preprocessing in order to maximise predictive accuracy, such as rebalancing skewed class distributions of the target variable. Conflicting best practices exist in data mining and market modelling, without diffusion between the two disciplines. To reflect this discrepancy, we evaluate the predictive accuracy of balanced versus imbalanced classification of consumer online shopping behaviour using logistic regression and NN. Experimental predictions are computed using socio-demographic, product and Internet shopping specific variables to classify consumers into "online shoppers", "browsers" and "non-Internet shoppers" using UK sample survey data. Our findings suggest that rebalancing data increases accuracy for both methods. In addition, NN provide superior classification accuracy and limited interpretation of explanations for class membership.**

I. INTRODUCTION

The continuous development of the Internet is challenging marketers to analyse the heterogeneous behavioural effects of consumers. Marketing research indicates that the Internet is changing the way in which consumers use different information and shopping channels before purchasing a product [1]. While some customer segments search for information and purchase online, others only search online but purchase in conventional stores while still others do not use the internet for shopping at all, indicating heterogeneous groups of customers at different stages of adoption of Internet shopping. Seeking to maximize customer retention, retailers need to predict and manage the shopping channel preferences of their customers in order to align their marketing strategies and to offer them an adequate service [2, 3].

Gathering information on homogeneous customer behaviour and predicting their decision processes is traditionally conducted in the domain of marketing modelling and – more recently – in data mining (DM) [4, 5]. While marketing modelling focuses on descriptive and normative modelling to provide explanations of the impact and relationships of independent variables upon class membership within a priori defined models, data mining seeks to identify novel relationships directly from the data to facilitate predictive modelling [6, 7], often lacking any explanation of the discovered causal relationships. To reflect evident differences in heritage, context and objectives, both domains routinely apply different methods with varying predictive accuracy to model the probability of consumer choice or class membership from identical datasets. In particular, novel DM methods from computational intelligence such as artificial neural networks (NNs) promise enhanced predictive accuracy and attractive features such as model building and universal approximation directly from the presented data [8].

However, both disciplines apply similar models on identical datasets to support similar decision processes, such as predicting consumer behaviour. In addition, the accuracy of predicting consumer shopping behaviour may not only be contributed to the performance of the individual methods, but also to differences in the model building process. DM has developed particular best practices in data preprocessing, scaling and sampling or weighting [9, 10]. They digress significantly from established practice in market modelling – even for the use of identical models such as multinomial logistic regression (MLR). In the frequent empirical case of imbalanced classification with symmetric misclassification costs, where one class is underrepresented in the population and the sample, best practice in DM suggests resampling the minority class to achieve balanced class distributions of the predicted variable [11, 12], while market modelling seeks to match the distribution of the target variable within sample to its distribution in the population to be described, i.e. all households in the UK.

Intrigued by the gap in research on contradictory best practices, we seek to evaluate the impact of alternative resampling schemes on both domains, evaluating the impact of imbalanced classification versus rebalanced classification through oversampling for two methods prominent in each domain, MLR and NN. We base our experiments on an empirical dataset, aiming to predict consumer adoption of internet shopping from socio-demographic, product specific and internet behavioural factors. Following a brief introduction to the domains of market modelling and data mining for predictive classification, section 3 briefly introduces the two prominent methods of logistic regression and NN. This is followed by a description of the evaluated dataset and the experimental design, followed by results for the proposed rebalancing approach from data mining in section 4. Conclusions are given in section 5.

II. MODELING CUSTOMER ONLINE SHOPPING BEHAVIOUR FOR DATA MINING

With the increasing popularity of the Internet as a new source of information, innovative services such as online banking and shopping became available in the early 1990s, offering the opportunity for location-independent around the clock pre-sales, sales and post-sales services to retailers and wholesalers. This has challenged marketers to analyze changing consumer shopping behaviour and novel decision making processes within a new, rapidly developing medium. Online marketing research suggests that the Internet is changing the way in which people use shopping channels to purchase products [1]. Whilst consumers may not immediately switch from offline to online shopping, they may add or use the Internet as part of their buying process. Research indicates three groups of consumers at different stages and levels of adoption of Internet shopping. Faced with a purchase need, certain consumers prefer to (1) use the Internet to search or browse for product and price information and buy online or (2) use the Internet only to search for product information but to buy in-store or (3) does not use the internet in their shopping process at all.

Behavioural information on whether and how consumers purchase different products online will influence the marketing strategy, purchasing processes and design of online or e-retailers distributing products and services through the Internet. Distinguishing product and customer segments with particular high or low probability to purchase online may determine whether retailers extend their current distribution channels to provide online product information and / or shopping services to their customers. Specific sites may be tailored to manage the shopping channel preferences of individual classes, specializing on browsers, shoppers or both in order to offer increased service [2, 3]. In addition, predicting each online customers' probability to purchase a product during the browsing stage could lead to customized online purchasing processes, providing consumers likely only to browse with additional information and incentives to reduce the perceived risks and to buy online, therefore increasing sales revenue and market share.

The related issues of modelling stochastic consumer behaviour are taken up by the research areas of marketing modelling and data mining alike. Marketing or market modelling (MM) refers to building formalized, quantitative models originally derived from management science and operational research to support Marketing decisions [6, 7]. Data Mining (DM), in the marketing domain named database marketing, refers to the a step in the non-trivial process of knowledge discovery in data, seeking to identify valid, novel, potentially useful and ultimately understandable patterns in data [13]. While both research areas apply statistical methods and quantitative data to derive models of consumer behaviour, they pursue a different primary purpose in model building. MM focuses on descriptive and normative models to describe decisions and processes of consumer behaviour [6], such as internet shopping behaviour, often based upon consumers' perceived utility of alternative choices using conventional statistical methods. On the other hand, DM models focus on predictive modelling [6], often applying purely data-driven methods from computational intelligence on identical datasets as MM. Particularly NNs promise attractive features to predictive analytics, being a data driven learning machine, permitting universal approximation [8] of arbitrary linear or nonlinear functions from examples without a priori assumptions on the model structure. While novel DM methods such as NN routinely outperform conventional statistical methods such as MLR in predictive accuracy, they lack the power to explain the underlying model for the data generating process. Essentially, MM and DM contrast the traditional statistical research approach of formal model building, formulating formal hypotheses and testing them with empirical data, versus a new paradigm of data driven research, integrating theory and heuristics to build models directly from the data [14]. While conventional models are suitable for structured problems and small datasets, the recent increase in available data has supported the rise of data driven methods for unstructured problems in large datasets. Nevertheless, resistance of market modellers persists to use data-driven approaches in model specification even where sufficient data is available [7, 15], leading to limited applications and research on NN in market modelling [16-18].

In a given problem domain as Internet shopping behaviour, both approaches need to incorporate the same relevant independent variables into their models to derive valid and reliable decisions. Essentially, both DM and MM use identical datasets but different methods, leading to results of different predictive accuracy. However, differences in performance may not only be contributed to the different methods used. Both domains have developed dissimilar best-practices for data preprocessing, using distinct sampling and coding schemes of variables to facilitate predictive accuracy or explanatory value. For example, it would be a routine procedure in DM of datasets with imbalanced target value distributions to rebalance the dataset [11, 12, 19] and to recode ordinal and nominal variables in a particular manner to facilitate supervised classification [20, 21]. However, these approaches are not evaluated in MM research, although they could lead to enhanced predictive accuracy of the MLR models as well. These issues will be addresses in the following experiments, following a brief introduction to the methods.

III. PREDICTIVE CLASSIFICATION METHODS

A. Logistic Regression for Classification

Multinomial logit regression (MLR) models are routinely applied as a means of classifying individuals or instances in various domains including marketing and market modelling [22, 23]. With regard to modelling class membership, MLR assumes that each consumer chooses the outcome that maximizes utility which each individual derives from a covariate vector of observable, independent variables Z_i and a vector of random/stochastic components η_{ij}. The individual's utility $U_i(J=j)$ from any choice j, $j = 1,2,3,\ldots J$, is explained by the characteristics defined in the vector Z_i, such as age, gender and income, as well as information on how these choices are perceived by the individual, e.g. how risky the individual perceives Internet shopping. Assuming that the utility of outcome j is linearly related to the vector of independent variables we assume

$$U_i(J = j) = \alpha_j Z_i + \eta_{ij} \quad , \tag{1}$$

The individual chooses outcome j over outcome k based upon the probability $\mathrm{P}(\bullet)$ that the utility from outcome j is higher than that derived from outcome k, resulting in

$$\mathrm{P}(J=j)=\mathrm{P}\left(\alpha_j Z_i+\eta_{ij}>\alpha_k Z_i+\eta_{ik}\right), \tag{2}$$

The specification of a probability model depends on the statistical distribution assumption of the error components η_{ij} and η_{ik} respectively. If the error terms associated with each choice η_{ij} are identically distributed as a Weibull distribution $F(\eta_{ij})=\exp(-e^{-\eta_{ij}})$, the probability that the individual chooses outcome j is defined [24] as

$$\mathrm{P}(J=j)=\frac{\exp\left(\alpha_j Z_i\right)}{\sum_{j=1}^{J}\exp\left(\alpha_j Z_i\right)} . \tag{3}$$

B. Multilayer Perceptrons for Classification

Multilayer perceptrons (MLPs) represent a prominent class of NNs for classification, implementing a supervised, feedforward, and hetero-associative paradigm [25]. MLPs consist of several layers of nodes u_j fully interconnected through weighted acyclic arcs w_{ij} from each preceding layer to the following, without lateral connections or feedback. Each node output calculates a transformed weighted linear combination of its inputs of the form $f_{act}\left(\vec{w}^T\vec{o}\right)$, with $\vec{o}$ the vector of output activations o_j from the preceding layer, $\vec{w}^T$ the transposed column vector of weights w_{ij}, and f_{act} a bounded non-decreasing non-linear function, such as the linear threshold or the sigmoid, with one of the weights w_{0j} acting as a trainable bias θ_j connected to a constant input $o_o=1$ [26]. Fig.2 gives an example of a MLP with a [3-4-1] topology:

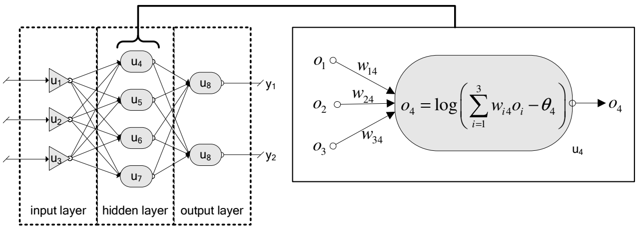

Fig. 1. Three layered MLP showing the information processing within a node, using a weighted sum as input function, the logistic function as sigmoid activation function and an identity output function.

For pattern classification, MLPs partition the input space through linear hyperplanes. To separate distinct classes, MLPs approximate a function $g(\vec{x}):X\rightarrow Y$ through adapting the free parameters $\vec{w}$ to minimize an objective function $e\left(\vec{x}\right)$ on the training data, which partitions the X space into polyhedral sets or regions, each one being assigned to one out of the m classes of Y. Each node has an associated hyperplane to partition the input space into two half-spaces. The combination of the linear node-hyperplanes in additional layers allows a stepwise separation of complex regions in the input space, generating a decision boundary to separate the different classes [26, 27]. The orientation of the node hyperplanes is determined by $\vec{w}$ including threshold θ_j modelled as an adjustable weight w_{0j} to offset the node hyperplane along $\vec{w}$ for a distance $d=\theta_j\|w\|$ from the origin for a more flexible separation. [26] The node non-linearity f_{act} determines the output change as the distance from x to the node hyperplane. In comparison to hard-limiting activation functions with a binary class border, the hyperplanes associated with sigmoid nodes implement a smooth transition from 0 to 1 for separation, allowing a graded response depending on the slope of the sigmoid function and the size of the weights.

The representational capabilities of a MLP are determined by the range of mappings it may implement through weight variation. [28] Single layer perceptrons are capable of solving only linearly separable problems, correctly classifying data sets where the classes may be separated by one hyperplane [26]. MLPs with three layers are capable to approximate any desired bounded continuous function. The units in the first hidden layer generate hyperplanes to divide the input space in half-spaces. Units in the second hidden layer form convex regions as intersections of these hyperplanes. Output units form unisons of the convex regions into arbitrarily shaped, convex, non-convex or disjoint regions [26], as in Fig. 2.

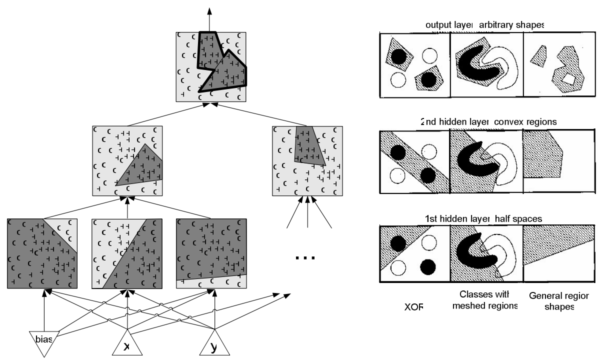

Fig. 2. Partitioning of the input space by linear threshold-nodes in a MLP with two hidden layers, one output node and examples of separable decision regions.

Given a sufficient number of hidden units, a MLP can approximate any complex decision boundary to divide the input space with arbitrary accuracy, producing a (0) when the input is in one region and an output of (1) in the other [29]. This property, known as a universal approximation capability [30], poses the essential problems of adequate model complexity in depth and size, i.e. the number of nodes and layers, and controlling the network training process to prevent overfitting.

MLP offer various degrees of freedom in modelling class membership of instances through different numbers of outputs. The desired output as a binary class membership is often coded with one output node $y_i=\left\{(0;1);(-1;1)\right\}$ and the logistic activation function, resembling nonlinear logistic regression. For multiple classification, n nodes with $t_i=\left\{(0,1);(1,0);...\right\}$ are used respectively [28]. In order to allow ranking of instances, MLPs can model probabilities of class membership trough an extending the logistic function to a softmax output function

$$y_k = \frac{e^{I_k}}{\sum_{l=1} e^{I_l}} \quad , \qquad (4)$$

which normalizes the outputs so that the sum of activations of all units in the output vector representing the predicted probability of class membership is equal to 1, $\sum_i P(y_i \mid \vec{x}_i) = 1$.

IV. EXPERIMENTAL SETUP

A. Data Description and Analysis

We use sample data of 685 Internet users from a recent survey of 5500 UK households to model choice of online shopping channels. The survey questions were determined from previous research. For example, research indicates that older consumers, those in the lower income groups and females are less likely to shop online. Also, shopping channel choice is influenced by the product that the consumer intends to buy. Less predictable products or products where inspection and touching is important are less likely to be bought online, commonly referred to as the perceived product specific risk of online shopping [3, 31]. In the context of our research, for some branded durable products consumers search for product information on the Internet, test the product in-store then go online again in order to find the best deal for the product they intend to buy, e.g. electronic devices such as video cameras For other products as books, CDs or software, the consumer tends to search for information and to buy on the Internet. In addition, the technology acceptance model (TAM) [32] derived from the theory of reasoned action suggests that acceptance and integration of technology within a person's habits and lifestyle may facilitate other activities online, such as shopping, indicating a generic utility derived from using the Internet and computers [33-35]. Each survey consisted of 73 questions on factors related to internet shopping, product, socio demographics plus a derived factor score, of which only a relevant subset was used for model building, seeking to capture and control a wide variety of interacting variables in our models.

Internet shopping specific factors of ordinal scale relate to individual questions on behavioural perception measured on a scale of 1 to 5, with (1="strongly disagree" to 5="strongly agree"), namely "I would buy online if products are Branded" [Brand], "Internet shopping is very convenient" [OSC] and "Going to the shops is as convenient as Internet shopping" [OSGs]. Product specific factors of binary scale were used to indicate the perceived risk or preference to shop for specific products as CDs [PreCD], grocery [PreGR], clothes [PreCL] and travel arrangements [PreTR] in-store or online (with in-store=1; online=0). In addition, information on the Age [Age], annual income [AnInc] and gender [Gender], with (1=male, 0=female) is used to control for the impact of socio-demographic effects. The factor score of Internet utility [ICTutil] measures the perceived ease of use, entertainment and usefulness of both computers and the Internet in the context of the TAM, reducing the dimensionality of 6 correlated variables through factor analysis.

The dependent variables represent three classes $Y = \{y_1, y_2, y_3\}$, with y_1 representing consumers who use the internet to search for information and buy online, y_2 of those who use the Internet only to search for product information but buy in-store and y_3 for the class that does not use the internet in the buying process at all. Fig. 1 gives an overview of the experiment design.

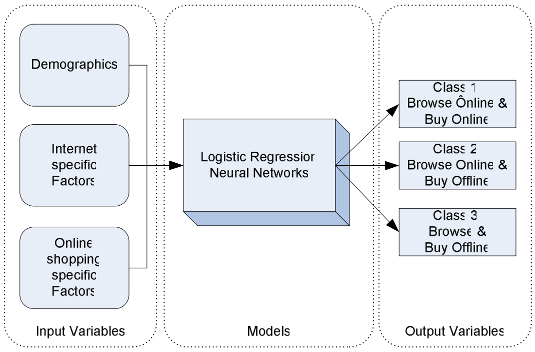

Fig. 1. Experimental setup of dependent and independent variables for alternative models in market modelling and data mining

The dataset has been described, analyzed and validated in [36] and is considered to provide valid and reliable results.

B. Data Sampling

While rebalancing of datasets to achieve homogeneous distributions of the predicted variables is considered imperative in DM to compute valid and reliable results for certain machine learning methods such as NN, it is not used for MLR in the market modelling domain. Therefore, we create two datasets, where dataset A represents an imbalanced distribution of the class memberships $\{y_1; y_2; y_3\}$ with $y_1 > y_2 > y_3$, and dataset B is rebalanced for equal number of instances in each class $y_1 = y_2 = y_3$, as displayed in Fig. 2. Consequently, dataset A versus B reflects a marketing modelling versus a data mining best practices approach to classification.

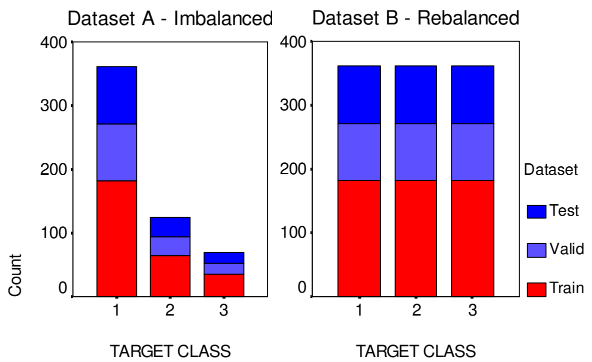

Fig. 2. Frequency distribution of classes for the original and rebalanced datasets for the target class, with [1=browsers & buyers, 2= online browsers but buy in-store & 3=non-internet shoppers]

The datasets A and B were split up into a 50% training and 25% validation subset for parameterization of the methods and a 25% test dataset for out of sample evaluation. For the imbalanced classification A, instances were sampled in relation to their distribution in the original dataset, assuring a [65.0%;25.5%;25.5%] relationship in each data subset.

To create dataset B for balanced classification of the dependent variable, data was rebalanced by randomly duplicating records from the underrepresented classes 2 and 3 until their number equalled the amount of instances in the majority class 1, resulting in equal class distributions [11, 12]. We then split the data into three datasets for training, validation and out of sample testing, resulting an equal number of instances with each class membership across all datasets. Tab. 1 gives an overview.

Tab. 1. Distribution of classes for data-subsets for dataset A and B

	Instance Distribution of Imbalanced Dataset				
Dataset	Class 1	Class 2	Class 3	Sum	in %
A-Training	181	64	35	280	50.0%
A-Validation	90	30	17	137	25.0%
A-Test	90	31	17	138	25.0%
A-Sum	361	125	69	555	100%
A-Sum in %	65.0%	22.5%	12.5%	100%	
B-Training	181	181	181	543	50.0%
B-Validation	90	90	90	270	25.0%
B-Test	90	90	90	270	25.0%
B-Sum	361	361	361	1083	100%
B-Sum in %	33.3%	33.3%	33.3%	100%	

C. Data Preprocessing

While no method specific preprocessing of the datasets was required for the logistic regression, training of MLPs normally recommends adequate preprocessing of data through rescaling of ordinal and nominal input and output vectors to facilitate learning [37]. In order to permit direct comparison of the experiments with MLR we refrain from rescaling and use variables of ordinal or metric scale during learning. However, we normalize the data to avoid computational problems, meet algorithm requirements and facilitate network learning through quicker training [38]. We select a simple normalisation from a variety of different scaling schemes to scale the data

$$y_t' = \frac{y_t}{y_{\max} + h} \quad . \qquad (\mathbf{5})$$

Variables of interval and ordinal scale are scaled to [-1,1], despite strong incentives in the data mining domain to rescale ordinal variables into nominal scale. We refrained from rescaling ordinal variables in order to allow a direct comparison with best practices in market modelling, modelling a presumed equidistant relationship of each ordinal measurement level. Variables of binary scale are coded using and n-1-scaling into a single predictor of $\{-1;1\}$ to depict class membership. Class membership in the output vector is n-coded as a binary vector of $\{(1,0,0);(0,1,0);(0,0,1)\}$ for class 1, 2 or 3 respectively.

V. SIMULATION EXPERIMENT

A. Logistic Regression Experimental Setup

We estimate a MLR model using transformed likelihood function to model the log likelihood method

$$\ln L = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} \ln \mathrm{P}(J = j) \quad , \qquad (\mathbf{6})$$

with the dummy variable $d_{ij} = 1$ if an individual i chooses choice j and $d_{ij} = 0$ otherwise, with P(J=j) specified as in (**3**). The method of maximum likelihood consists of finding coefficients that will maximize the log likelihood function for our MLR. To estimate the parameters that maximize the likelihood function, we apply an iterative, heuristic Newton method using the software Limdep 8. We estimate the MLR model by normalizing on one outcome to remove indeterminacy, producing two parameter matrices to predict three class membership outcomes. The coefficients are related to the probability of an outcome relative to the normalized outcome. They are then used to calculate the predicted probabilities to determine class membership by the highest predicted probability of each class for each individual in the training and test dataset. For a comprehensive review we refer to [24].

B. MLP Experimental Setup

We consider a set of fully connected feed forward architectures of multilayer perceptrons without shortcut connections. To determine an appropriate NN topology for each dataset, we evaluate a candidate set of 10 NN topologies with one and two hidden layers using 11 nodes in the input layer representing the input variables and 3 output nodes identifying the individual class membership, specifying each network topology as 11-n-3. We evaluate a maximum of 25 nodes over all hidden layers to balance the maximum number of 543 training observations with the maximum degrees of freedom through trainable weights, aiming to avoid topologies prone to overfitting. We evaluate each topology using two activation functions, the hyperbolic tangent (TanH) and the logistic function, using the summation as the input function. With regard to the multiple classification objective, we apply the softmax activation function in the output layer.

We train each NN with the Delta-Bar-Delta learning rule of adaptable learning parameters for each weight. The network weights w_{ij} of each topology are initialized 10 times with randomized starting weights of $-0.3 \geq \mathrm{rnd}(w_{ij}) \leq 0.3$, to account for local minima, training the network for a maximum of 10000 epochs of 280 or 543 iterations respectively and saving the best weight matrix by mean classification rate on the validation set. To limit computational time, we apply an early stopping mechanism to abort learning after no improvement of the validation error for 100 epochs. Each network trained for less than 2 minutes on a Pentium IV with 3.8GHz using NeuralWorks Professional. We create, train and evaluate 200 networks and select the topology with the highest mean classification rate on the validation dataset, resulting in a 11-10-3 topology with a logistic activation function. No test data was used to estimate or select a NN model.

C. Experimental Results

We analyze the results of the experiments of predicting consumer shopping behaviour using MLR versus NN across two datasets A and B of imbalanced and rebalanced classification. As the MLR is parameterized on a joint training and validation set, only those results are presented below.

As an error metric, we provide the classification rate of each class and the arithmetic mean classification rate (MCR) in percent, as alternative, less biased performance metrics such as ROC-curves or lift-charts are not applicable to more than two class problems and other metrics are contain an even stronger bias for imbalanced classification tasks. The confusion matrix in Tab. 2 presents the individual classification rate of each predicted class versus the true class membership across training and test set for the MLR and the best NN topology (A-C1-03).

Tab. 2. Classification rates on the imbalanced Dataset A

True Values by Method	Predicted Values of Methods in %					
	Training Data Set			Test Data Set		
	C 1	C 2	C 3	C 1	C 2	C 3
MLR C1	**93.3**	5.2	1.5	**88.9**	7.8	3.3
MLR C2	62.8	**23.4**	13.8	48.4	**22.6**	29.0
MLR C3	36.5	17.3	**46.2**	35.3	29.4	**35.3**
NN C1	**94.5**	4.4	1.1	**92.2**	4.4	3.3
NN C2	64.1	**21.9**	14.6	54.8	**22.6**	22.6
NN C3	28.6	20.0	**51.4**	47.1	11.8	**41.2**

The MLR seems incapable of separating all three classes. While separating classes one and three, of those consumers who browse and buy online against those who do not use the Internet at all, the MLR fails to accurately predict class 2 of those who browse online but don’t proceed to shop online even on the training data. This presents an intuitively appealing result, suggesting that the class of consumers in an intermediate state of Internet adoption – in a DM interpretation the class lying close to the decision boundary between the original classes of online buyers and non-buyers - is harder to differentiate than the others. While the mean classification rate for MLR of 55.9% and 51.9% on training and test set suggests an enhanced performance in comparison to pure random selection, this is largely contributed to the bias towards overpredicting membership of the class 1 for each of the three classes, failing to predict class membership 2 and showing limited ability to separate class 3.

As should be expected from previous DM research, even the best NN topology is incapable of separating all three classes due to the dominance of the majority class in the imbalanced dataset. While both methods fail to separate all 3 classes, the NN outperforms the MLR with an MCR on the test set of 52.1% versus 48.9%, predicting classes 1 and 3 with higher accuracy than MLR. We therefore consider NN superior regarding the prediction of class membership for imbalanced classification, despite its shortcomings in predicting all three classes and neglecting NN limitations in explaining the coefficients of the covariates.

Tab. 3 represents the results for the rebalanced dataset B for the best NN (V.B-C1-01) and MLR respectively. The accuracy of the MLR increases from 48.9% to 54.4% MCR, showing valid and reliable prediction of all three classes significantly larger than 33.3% by chance. This seems particularly interesting, as previous research and best practices in MM do not suggest rebalancing of empirical datasets for increased predictive accuracy. However, further analysis is required to analyze the impact of rebalancing on the coefficients, to estimate the relevance of increased predictive accuracy versus descriptive validity.

Tab. 3. Classification rates on the rebalanced Dataset B

True Values by Method	Predicted Values in %					
	Training Set			Test Set		
	C 1	C 2	C 3	C 1	C 2	C 3
MLR C1	**68.3**	24.3	7.4	**74.4**	16.7	8.9
MLR C2	30.6	**43.9**	25.5	31.1	**36.7**	32.2
MLR C3	17.0	19.9	**63.1**	6.2	27.8	**52.2**
NN C1	**70.2**	21.5	8.3	**74.5**	17.7	7.8
NN C2	31.5	**40.88**	27.62	27.8	**38.9**	33.3
NN C3	19.3	12.1	**68.5**	26.7	15.6	**57.8**

The predictions of the NN also improve from 52.1% to 57.0% MCR, showing robust separation of all classes and therefore supporting the best practices recommendation of DM to rebalance classes in the case of homogenous misclassification costs. Both methods are able to distinguish the previously inseparable class of Internet browsers, a class of particular interest and therefore with potentially significant implications to e-Business practice.

The predictions of the MLR are consistently lower, across training and test set, for both datasets and for each of the 6 class memberships. On MCR, the NN outperforms MLR on the test of 57.0% versus 54.4%, a 5% significant increase in predictive accuracy.. These pre-eminent results suggest a clear dominance of the NN method over MLR, albeit only valid for this dataset in this application domain, but regardless of DM or MM recommendations.

VI. CONCLUSION

We analyse the predictive performance of MLR and NN across two distinct preprocessing schemes of rebalancing skewed datasets through oversampling as proposed in DM research versus using the imbalanced datasets representing the true class distribution of the target variable in the population, as proposed by marketing modelling. Our results indicate that the class of Internet browsers, those that only gather information online but shop offline, is hardest to separate from the traditional classes of online buyers and non-online users. In any setup, NN consistently outperform MLR on all pre-processing schemes and across all classes of consumer behaviour. What is more interesting is that MLR as well NN profit from rebalancing imbalanced datasets to increase correct classification of unseen instances. This indicates the potential for models and methods routinely applied in MM to draw upon the expertise gathered in DM to excel at predictive tasks.

However, adjusting class distributions may lead to altered coefficients of the MLR model, which needs to be analysed in depth regarding changes in sign and impact with statistical significance. Therefore we seek to extend our experiments towards a detailed analysis of the effect on explanatory modelling as opposed to predictive modelling. Considering the experimental setup, we seek to extend the evaluation towards using uncompressed data of the correlated variables used to derive the factor score of internet utility, omitting preceding linear transformations of variables and to extend evaluation towards imbalanced out of sample datasets. Another extension of research may decompose the multiple classification problem into two consecutive binary classifications, i.e. the use of internet versus no use, and the use of the internet to browse versus browsing and shopping, to allow for the application of novel, binary methods such as support vector machines. In addition, we seek to incorporate additional discrepancies between data mining and market modelling practices, such as rescaling of ordinal variables for binary input vectors as recommended for NN. Furthermore, we seek to train NN directly upon raw questionnaire data in order to limit the effects on nonlinearities in ordinal coding of questions on arbitrary scales, further extending the use of data mining methods towards marketing.

REFERENCES

[1] "A perfect market. Survey: E-commerce," in *Economist*. London, 2004.

[2] M. M. Montoya-Weiss, G. B. Voss, and D. Grewal, "Determinants of online channel use and overall satisfaction with a relational service provider," *Journal Of Academy Of Marketing Science*, vol.31,pp.448-458, 2003.

[3] R. A. Peterson and M. C. Merino, "Consumer information search behavior and the Internet," *Psychology & Marketing*, vol. 20, pp. 99-121, 2003.

[4] M. H. Dunham, *Data mining introductory and advanced topics*. Upper Saddle River, N.J.: Prentice Hall, 2003.

[5] I. H. Witten and E. Frank, *Data mining: practical machine learning tools and techniques*, 2nd ed. Boston, MA, 2005.

[6] P. S. H. Leeflang and D. R. Wittink, "Building models for marketing decisions: Past, present and future," *Int. Journal of Research in Marketing*, vol. 17, pp. 105, 2000.

[7] J.-B. E. M. Steenkamp, "Introduction: Marketing Modeling on the Threshold of the 21st Century," *International Journal of Research in Marketing*, vol. 17, pp. 99, 2000.

[8] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, pp. 359 - 366, 1989.

[9] D. J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*. Cambridge, Mass.: MIT Press, 2001.

[10] S. F. Crone, S. Lessmann, and R. Stahlbock, "Empirical comparison and evaluation of classifier performance for data mining in customer relationship management," 2004.

[11] N. Japkowicz and S. Stephen, "The class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis*, vol. 6, pp. 429-450, 2002.

[12] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM SIGKDD Explorations*, vol. 6, pp. 7 - 19, 2004.

[13] U. M. Fayyad, *Advances in knowledge discovery and data mining*. Menlo Park, Calif. [u.a.]: AAAI Press [u.a.], 1996.

[14] T. W. Miller, *Data and Text Mining - A Business Apllications Approach*. New Jersey: Pearson, 2005.

[15] M. G. Dekimpe and D. M. Hanssens, "Time-series models in marketing: Past, present and future," *International Journal of Research in Marketing*, vol. 17, pp. 183, 2000.

[16] J. J. Merelo, A. Prieto, V. Rivas, and J. L. Valderrabano, "A neural net-based model for decision making in marketing," *Accounting, Management and Information Technologies*, vol. 8, pp. 237, 1998.

[17] T. S. Gruca and B. R. Klemz, "Using Neural Networks to Identify Competitive Market Structures from Aggregate Market Response Data," *Omega*, vol. 26, pp. 49, 1998.

[18] C. G. Dasgupta, G. S. Dispensa, and S. Ghose, "Comparing the predictive performance of neural network model with some traditional market response models," *International Journal of Forecasting*, vol.10, pp.235, 1994.

[19] M. Kubat and S. Matwin, "Addressing the Curse of Imbalanced Training Sets: One-Sided Selection.," presented at Proceedings of the 14th ICML'97, Nashville, TN, U.S.A., 1997.

[20] D. Pyle, *Data preparation for data mining*. San Francisco, Calif.: Morgan Kaufmann Publishers, 1999.

[21] E. Tuv and G. Runger, "Pre-Processing of High-Dimensional Categorical Predictors in Classification," *Applied Artificial Intelligence*, vol. 17, pp. 419, 2003.

[22] H. Hruschka, W. Fettes, and M. Probst, "An empirical comparison of the validity of a neural net based multinomial logit choice model" *European Journal of Operational Research*, vol. 159, pp. 166, 2004.

[23] J. H. Roberts and G. L. Lilien, "Explanatory and predictive models of consumer behavior," *Handbooks in Operations Research and Management Science*, vol. 5, pp. 27, 1993.

[24] W. H. Greene, *Econometric analysis*, 5. ed., internat. ed ed. Upper Saddle River, NJ [u.a.]: Prentice-Hall, 2003.

[25] D. E. Rumelhart, J. L. McClelland, and University of California San Diego. PDP Research Group., *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, Mass.: MIT Press, 1986.

[26] R. D. Reed and R. J. Marks, *Neural smithing: supervised learning in feedforward artificial neural networks*. Cambridge, Mass.: The MIT Press, 1999.

[27] R. Rojas, *Theorie der neuronalen Netze: eine systematische Einführung*. Berlin: Springer, 1993.

[28] K. Koutroumbas, "On the partitioning capabilities of feedforward neural networks with sigmoid nodes," *Neural Computation*, vol. 15, pp. 2457-2481, 2003.

[29] R. Stahlbock, *Evolutionäre Entwicklung künstlicher neuronaler Netze zur Lösung betriebswirtschaftlicher Klassifikationsprobleme*. Berlin: WiKu, 2002.

[30] S. F. Crone, "Training Artificial Neural Networks using Asymmetric Cost Functions," in *Computational Intelligence for the E-Age*, L. Wang, J. C. Rajapakse, K. Fukushima, S.-Y. Lee, and X. Yao, Eds. Singapore: IEEE, 2002, pp. 2374-2380.

[31] K. P. Chiang and R. R. Dholakia, "Factors driving consumer intention to shop online: An empirical investigation," *Journal Of Consumer Psychology*, vol. 13, pp. 177-183, 2003.

[32] F. D. Davis and V. Venkatesh, "A critical assessment of potential measurement biases in the technology acceptance model: three experiments," *International Journal of Human-Computer Studies*, vol. 45, pp. 19, 1996.

[33] H.-P. Shih, "Extended technology acceptance model of Internet utilization behavior," *Information & Management*, vol. 41, pp. 719, 2004.

[34] L. R. Vijayasarathy, "Predicting consumer intentions to use on-line shopping: the case for an augmented technology acceptance model," *Information & Management*, vol. 41, pp. 747, 2004.

[35] A. L. Lederer, D. J. Maupin, M. P. Sena, and Y. Zhuang, "The technology acceptance model and the World Wide Web," *Decision Support Systems*, vol. 29, pp. 269, 2000.

[36] D. Soopramanien, R. Fildes, and A. Robertson, "Internet Usage and Online Shopping Experience as Predictors of Consumer Preferences to Shop Online", Lancaster 07.01.2005.

[37] W. S. Sarle, "Neural Network FAQ - periodic posting to the Usenet newsgroup comp.ai.neural-nets," vol. 2004, 2002.

[38] C. M. Bishop, *Neural networks for pattern recognition*. Oxford: Clarendon Press, 1995.