

# Data Mining for Predicting Customer Satisfaction

Francesco Vivarelli

British Telecom

Segmentation Marketing and Customer Data



# Overview

- About BT
- Motivation
- Data: the BT Customer Survey
- Drivers of Satisfaction
- Model and Lift Charts
- Actions Taken
- Summary

# About BT

- BT is an integrated group of businesses that provide voice and data services in UK and Europe
- BT offers
  - Services to business and residential customers
  - Runs the network and sells network capacity to competitors
  - Offers business services and solutions world-wide
  - One of UK's leading Internet provider
  - helps businesses to gain advantage from communications technology



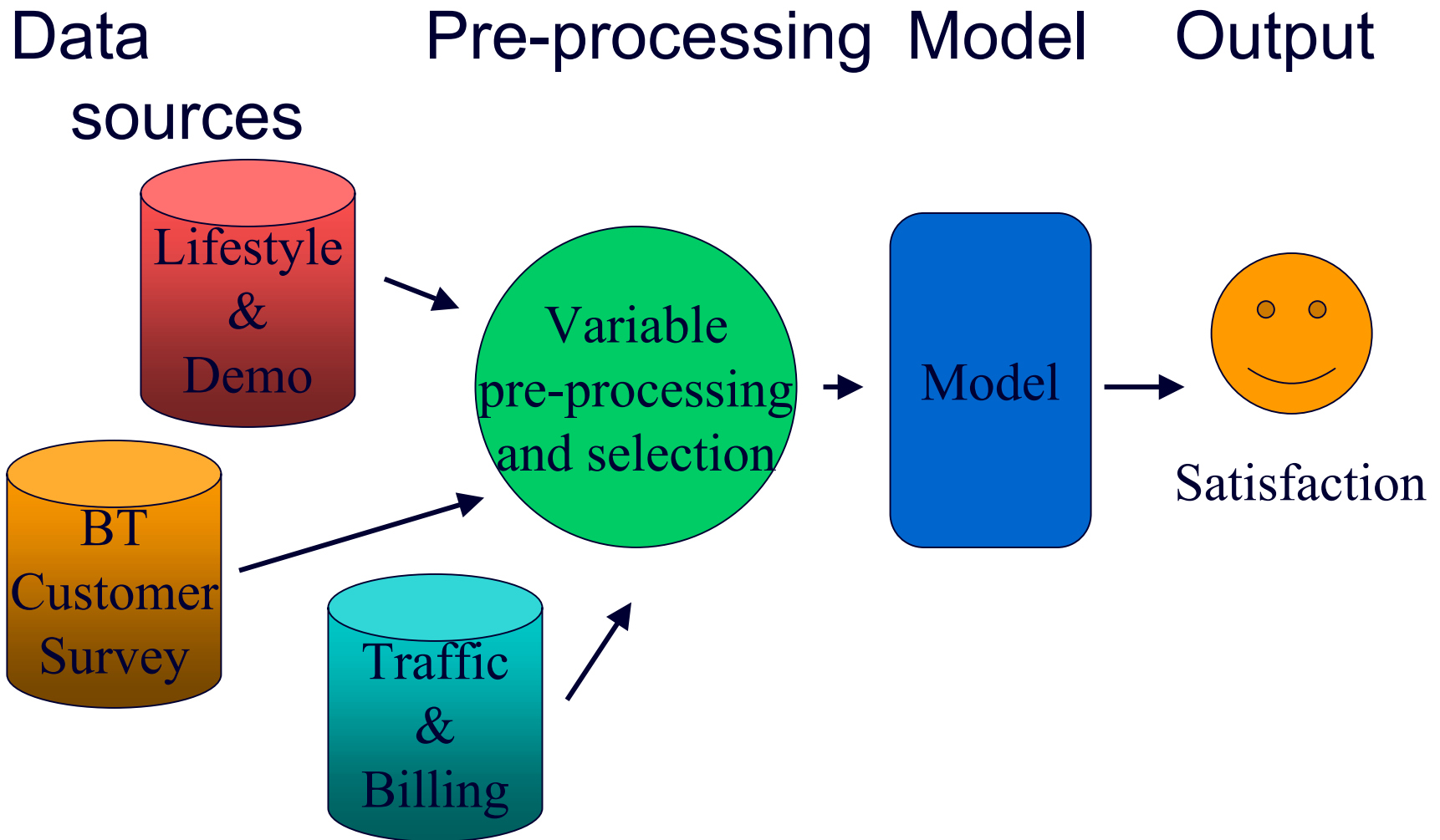
# About BT

- Turnover of about £ 25 bn
- More than 100k employees
- More than 28M UK lines
- 19M residential customers
- More than a million of Broadband connections

# Why Customer Satisfaction?

- Improving Customer Satisfaction is the cornerstone of BT strategy
- Commitment to reducing Dissatisfaction
  - dissatisfied customers more likely to share bad experience with others
- Crucial for Customer-centric enterprise
- Need to understand what drives Satisfaction so that key areas can be identified

# Roadmap of the project



# BT Customer Survey

- March - December 2002, BT carried out the most comprehensive customer survey ever held in the UK
- 19M customers mailed
- Response rate: over 7%
- Binary and categorical data
- 170 fields derived
- Six Sections

# BT Customer Survey - Section 1: You And BT

- Why did you open a new account with BT?
- Why did you choose BT?
- Satisfaction level with BT services
- Satisfaction with Value for Money
- Space for suggestion on possible improvements



# BT Customer Survey - Section 2: You And Your Home Telephone

- Telephone users in the household
- How many telephone lines
- Use of competitor's lines and level of satisfaction
- Bill size
- Type of calls (e.g. local, international,...)
- Use of reseller

# BT Customer Survey - Section 3: You And The Internet

- Where do you access the Internet?
- Use of the Internet (e.g. e-mail, gaming...)
- Hours spent (per week) on the Internet
- Method of payment
- Info about ISP provider
- Satisfaction levels with speed
- Satisfaction levels with quality and reliability

# BT Customer Survey - Section 4: You And Your Mobile Phone

- Number of mobile phone per household
- Bill size per month

# BT Customer Survey - Section 5: You And Your TV

- Info about TV services (e.g. digital, satellite)
- TV + Phone service?
- Monthly bill

# BT Customer Survey - Section 6: You And The Future

- Broadband
- Home PC network
- Video-phone services
- Online computer games
- Digital Equipment
- e-billing
- SMS from fixed line

# Traffic and Billing data

- Collected from transactional summaries of BT customers
- E.g. Package ownership, calls duration and charges
- Snapshot over a given time-window of how customer use the telephone

# Lifestyle and Demographics

- Describe customer at a personal level
- Provided by third party suppliers
- Demographics: Age, gender, occupation
- Financial: Income, credit cards, shares
- Lifestyle data (hobbies)

# Data pre-processing & Model selection

- CS, TB and LD data are collected in one string of data which describe features of each survey's responder
- ~ 800 attributes for each of the 1.5M responders
- Missing values
- Standardisation of input values
- Heuristic classification of comments

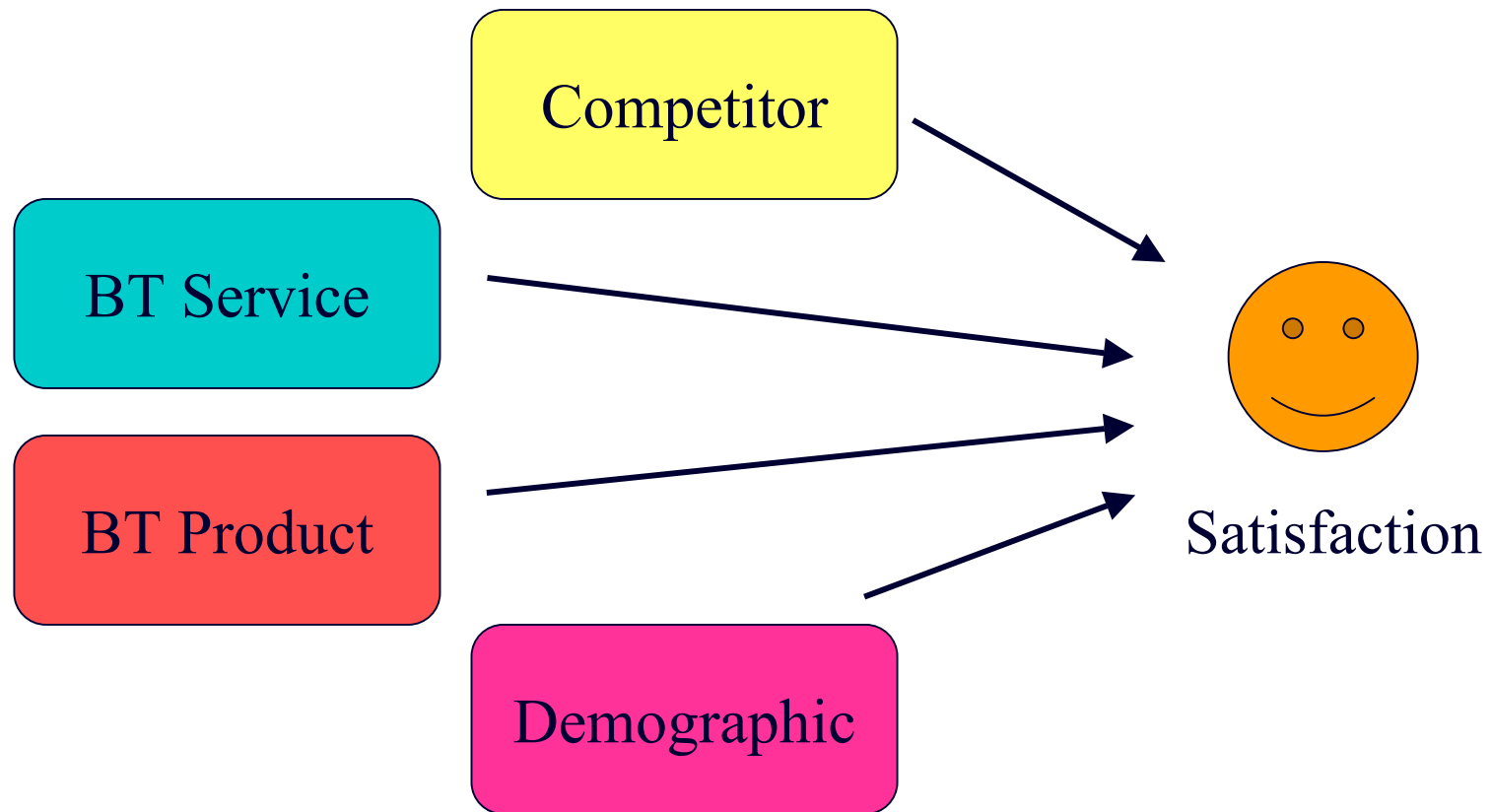


# Data pre-processing & Model selection

- Target is dissatisfaction, i.e. satisfaction levels between 1 and 4 (out of 10)
- Data split in training, validation and test sets
- We tried several models (Trees, NNs, LR)
- Logistic regression was chosen
- Output: probability that a customer is not satisfied with BT's level of service
- What are the drivers of dissatisfaction?

# What drives Satisfaction?

- 4 groups of factors identified

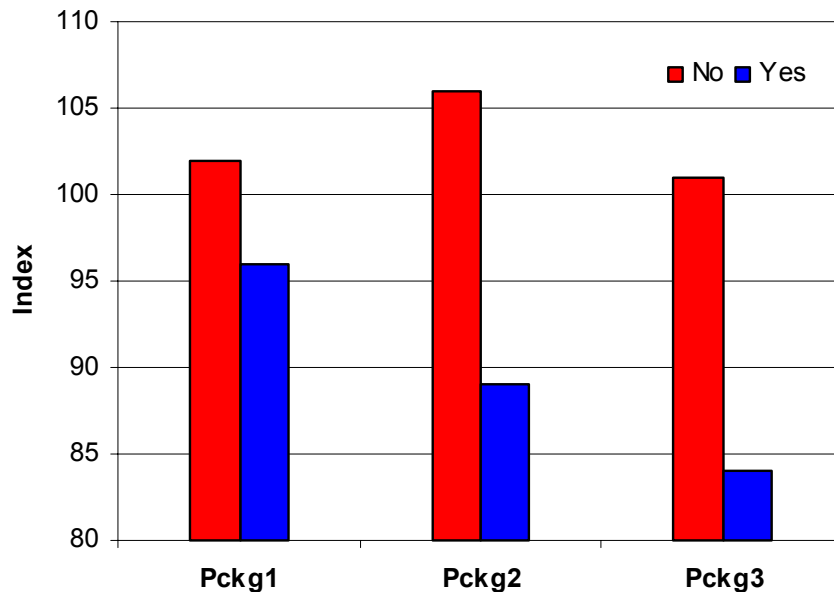


# What drives Satisfaction?

BT products	BT service
BT Quarterly Bill size	Faults
Calls to Internet provider	Interactions
Future Interest in Broadband	Service
Perception of Package complexity	Staff
Package ownership	
Satisfaction with Internet	
Uses of Internet (shopping, banking)	

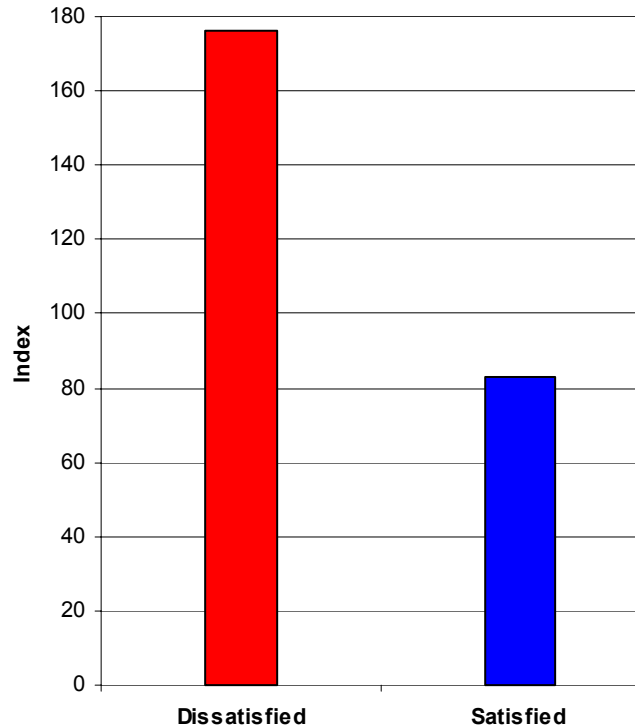
COMPETITOR	DEMOGRAPHIC
Bill with competitor	Age
Mobile phone bill	Family Income
Mobile phones in the family	Lifestage
Satisfaction with competitor	Occupation

# Package ownership



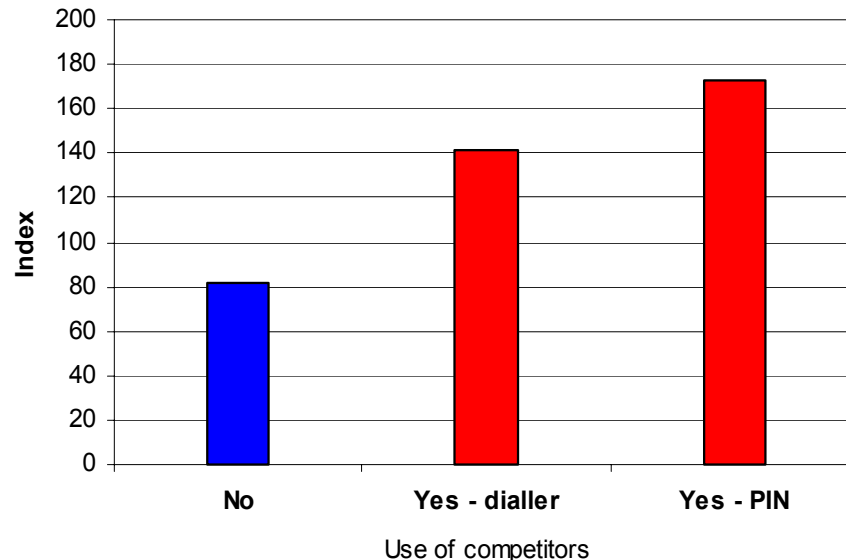
- Satisfaction levels related to Package ownership
- No package, more likely to be dissatisfied
- Can we improve our offer of packages?

# Satisfaction with Internet



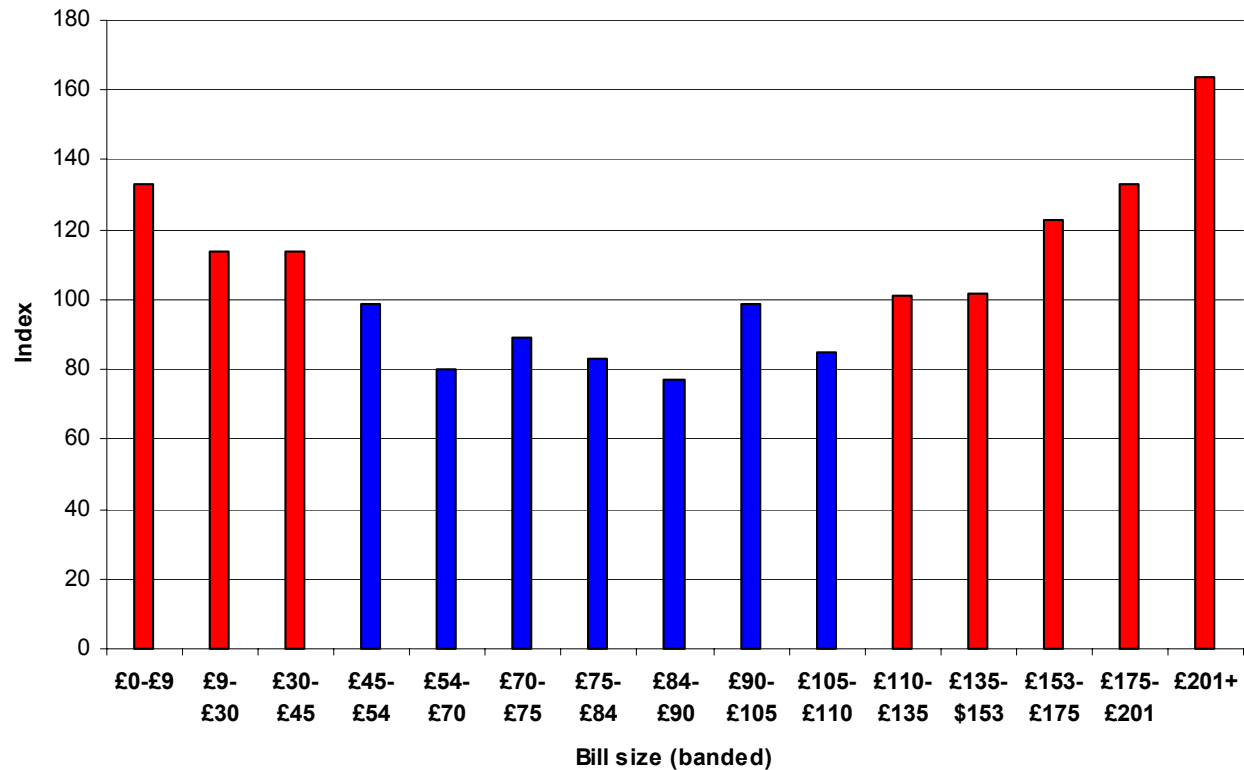
- Dissatisfied with Internet are twice more likely to be dissatisfied with BT

# Use of competitors



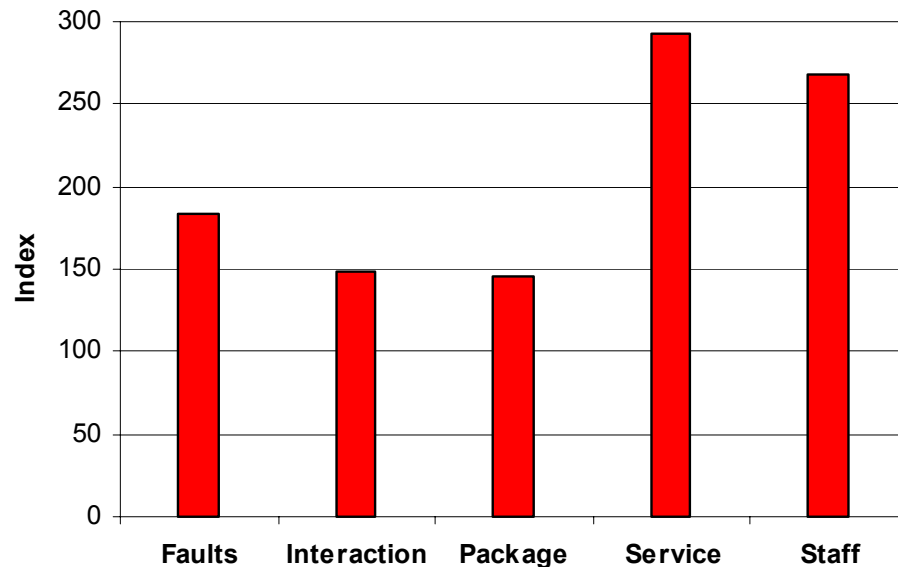
- Satisfaction levels linked to the use of competitors
- Customers who use competitors are the most likely to be dissatisfied

# Bill size



- Satisfaction levels related to bill size

# Comments



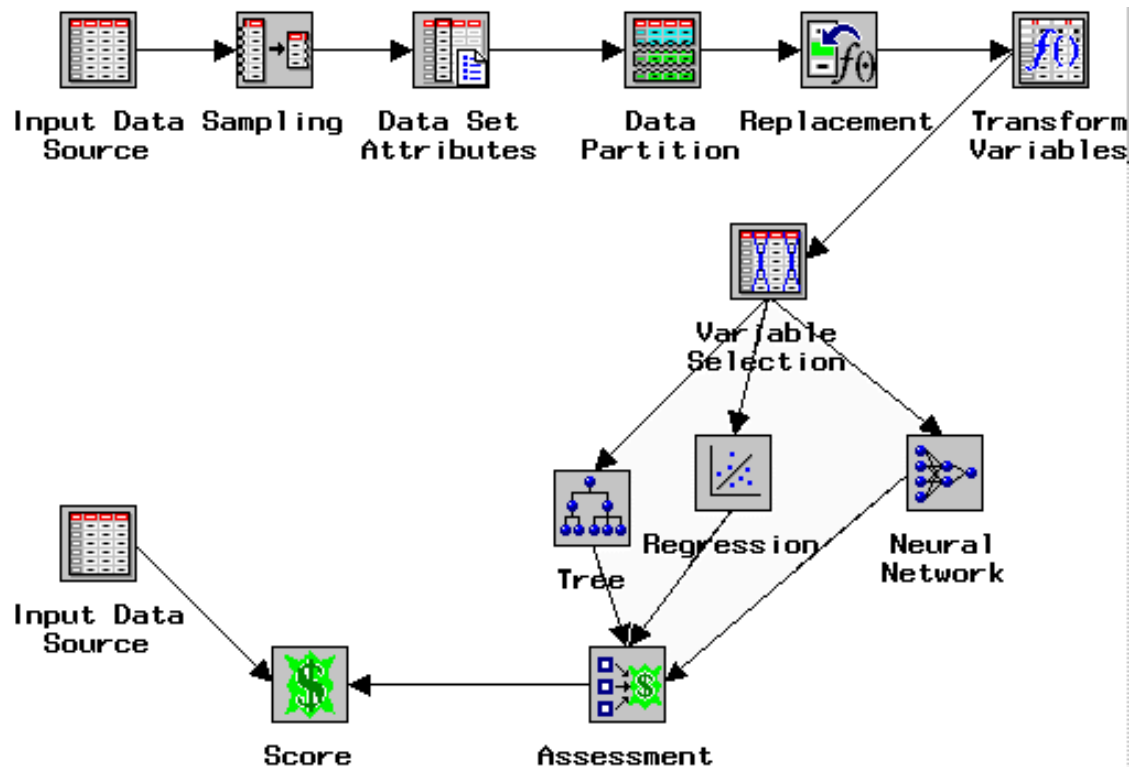
- Customers who write about Staff, Services, Package, Faults and Interaction are more likely to be dissatisfied



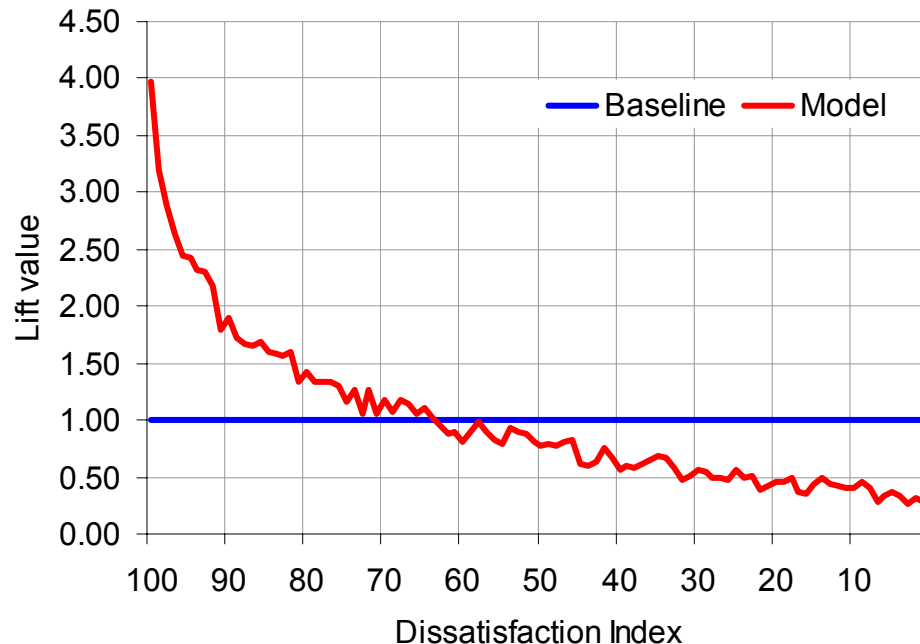
# SAS EM Desktop

- Analysis and modelling carried out with SAS<sup>®</sup> and Enterprise Miner<sup>™</sup>

- SEMMA process
- Score new responders



# Lift chart - Dissatisfaction model



- Model 4 times better than baseline in top percentile
- Cut off point at percentile 63

# Actions taken

- **Broadband** It is at the heart of BT strategy. Connected 1 M customer by June 2003. Target: 5M by 2006. Lower prices, wide range of applications, extending the reach of Broadband to many UK areas
- **Prices** customers asked for simpler packages and better value. The **New BT Together** options respond to that request



Adobe Acrobat  
Document

# Actions taken

- **Improvements in customer contacts**  
Recently BT has launched a new contact centre organisation which significantly improves channel management and customer relationship
- **Data fusion**

# Summary

- We presented a model of customer satisfaction
- Analysis based on the real voice of the customer
- We identified 4 main drivers
- Remarkable uplift
- Future work

# Contact details

- Francesco Vivarelli  
Data mining manager  
BT Segmentation marketing  
Tel: +44 (0)20 74928787  
e-mail: [francesco.vivarelli@bt.com](mailto:francesco.vivarelli@bt.com)

# Data mining for Predicting Customer Satisfaction

*Francesco Vivarelli\**

*BT*

BT has a documented strategic priority “to deliver the highest levels of customer satisfaction performance and reduce the number of dissatisfied customers each year”.

In order to achieve this objective, we need to understand the underlying drivers of satisfaction (and dissatisfaction) so that we can address them. Last year BT conducted the most comprehensive consumer survey ever launched in the United Kingdom; among many questions asked, two related to customers' level of satisfaction with BT - in terms of overall service and value for money.

In this paper we present the Satisfaction model we developed by using data mining techniques on the responses to the customer survey, against transactional summaries, lifestyle and demographic data. We describe the data preprocessing and the key factors that drive customer satisfaction. We present the model analysis and we show the lift obtained. BT's 19 million-customer base can be scored, obtaining a likely satisfaction index for each customer.

Models that estimate the satisfaction of customers can give a significant strategic advantage to the business. In the paper we discuss how to implement and leverage the key findings of the analysis in several business areas.

The analysis is carried out using SAS®-8.2 for Windows-98 and Enterprise Miner™-4.1.

## 1. Introduction

Improving customer satisfaction is the cornerstone of BT strategy as BT's growth is pursued by putting customer's interest first.

A key part in increasing customer satisfaction will be achieved by improving processes, systems, products and tools. At the same time BT is committed to reducing customer dissatisfaction. This is important for two reasons. To begin with, by getting things right for customers the first time, we will not incur the additional costs associated with repairing faults, or dealing with repeated requests for help. Also, it's axiomatic that dissatisfied customers are much more likely to share their bad experiences with others than satisfied customers are their positive experiences.

Our target could be achieved more easily when we get a better understanding of what influences the levels of satisfaction of our customers with the services offered by BT. In this respect the Customer Survey that BT carried out recently is an invaluable source of information in order to identify what are the key areas that need to be improved.

In this paper we present the data mining model that we developed to gain a better understanding of customer satisfaction levels. In particular, we illustrate the steps we followed to build the model, presenting the results obtained and highlighting the most relevant attributes that drive customer satisfaction. As a results of this analysis we will be able to put in place improvements, services and products that will enhance the customer experience of BT.

The structure of the paper is as follows. In Section 2 we describe the three sets of data we analysed in this project; we also discuss the pre-processing carried out on the raw data. A discussion of the model selected is briefly presented in Section 3. The list of the most important attributes that drive customer satisfaction is presented in Section 4, where we also show the uplift obtained in our model. Section 5 illustrates some of the actions taken following the key findings of the project. The last Section summarise the material presented in the paper and suggests directions for future implementation of the main findings.

## 2. Data sources and data pre-processing

For this project we analysed data gathered from three different sources.

**The BT Customer Survey (CS).** In the period March – December 2002 we carried out the most comprehensive customer survey ever held in the UK. We mailed a questionnaire to our customer base<sup>1</sup> of about 19 million customers; the response rate was over 7%.

Most of the data collected in the CS were binary values, some were categorical variables. We also left some empty space where customers were invited to send us few suggestion on how BT services and products could be improved. The CS was divided in six sections:

- (a) **Your satisfaction as BT customer** In this section BT tried to evaluate customer satisfaction levels with regards to services offered and value for money. The satisfaction levels were scaled from 1 to 10.
- (b) **You and your telephone** covers data about the use of the home telephone, e.g. number of users lines and telephone per household, bill size and type of calls. We also asked about the use of competitors' lines and the satisfaction with competitors' services.
- (c) **You and the Internet** collects information about the use of the internet, the places of access, the time spent on web-surfing and satisfaction with access and speed.
- (d) In **You and your mobile phone** we asked about the number of cellular handsets and the expenditure for mobile telephony per household.
- (e) **Communication services** collected information about the TV services owned by customers (e.g. digital or satellite television) and the cost of the service.
- (f) **Future interests** is the section where we considered interests in future technology, such as Broadband, home PC network, video-phone services, online computer games, digital equipment and e-billing.

Overall, for each customer we derived 170 fields.

**Traffic and Billing (TB)** data<sup>2</sup> are collected from transactional summaries of BT customers, e.g. package ownership, calls duration and charges, number of destinations called. These data give a snapshot over a given time-window of how the customer use the telephone. The number of TB variables available for the analysis is about 400.

**Lifestyle and demographics (LD)** data describe each customers at a personal level. They are provided by third party suppliers and have been obtained through media like "shoppers survey" questionnaires and product registration response forms and services. Attributes available include demographic (e.g. primary and partner age, marital status, number and age of children, occupation type), financial (e.g. household income, credit cards, stocks and shares) and lifestyle data (e.g. hobbies and interests, newspapers read, car ownership, home ownership status). The total number of LD variables is 230.

CS, TB and LD data are collected in one input vector (of about 800 variables) which describes features of each respondent to the survey. This huge amount of information has been pre-processed, analysed and modelled by using base SAS<sup>®</sup> and Enterprise Miner<sup>™</sup> 4.1 (see [2]).

In every data mining project, data pre-processing plays a crucial role. Some models (such as the neural networks) omit entire observations from training if any of the input variables are missing. Hence we need to replace missing data with imputed values. We replaced interval missing values with the input's mean. Missing values of a categorical/binary input were replaced with the mode.

The input vector describing each customer is composed by features whose values may differ for several order of magnitude; thus we need to transform each component, obtaining linearly scaled values. We standardised interval variables so that they have mean 0 and variance 1. A binary variable is replaced by a variable which contains values 0 or 1. A nominal variable with  $n$  categories is expanded in  $n$  dummy variables set to 0; only the one corresponding to the level to be coded is set to 1.

In the section **Your satisfaction as BT customer** we have collected a total of 400k customers' opinions. We have read the words and phrases written; each comment has been heuristically classified in one out of 70 binary categories, according to the main topic addressed in the text.

The target of our models is the level of customer dissatisfaction with BT services. The 10 levels of satisfaction have been mapped into a target binary variable, setting the target to 1, when the satisfaction was graded four or lower, and to 0 for satisfaction levels between five and ten. We decided to model the lowest levels of satisfaction as this is a better defined problem and it also allows us to identify areas where we can make the most crucial improvements.

The data collected have been split into three sets, training (containing 40% of the total base), validation and test sets (containing 30% of the data each). The training set is used to estimate the parameters of the model, the validation set to select the best model (e.g. the number of hidden nodes in a neural network) and the test set is used to estimate the generalisation error of the model selected.



<b>BT PRODUCTS</b>	<b>BT SERVICE</b>	<b>COMPETITOR</b>	<b>DEMOGRAPHIC</b>
BT Quarterly Bill size	Faults	Bill with competitor	Age
Calls to Internet provider	Interactions	Mobile phone bill	Family Income
Future Interest in Broadband	Service	Mobile phones in the family	Lifestage
Perception of Package complexity	Staff		Occupation
Package ownership			
Satisfaction with Internet			
Uses of Internet (shopping, banking)			

**Table 1** The list of the most relevant attributes for the Customer Satisfaction model. The category **BT PRODUCTS** groups the attributes linked with BT products' portfolio; **BT SERVICE** lists the most relevant comments expressed by customers about services; attributes related to services offered by BT competitors are grouped in **COMPETITOR**; **DEMOGRAPHIC** collects customer's personal attributes.

### 3. Model selection

We considered modelling customer dissatisfaction by using Logistic Regression (LR) and Neural Networks (NN) models. The section briefly presents these model; a more thorough discussion of these models, can be found in [1].

LR is a linear model which estimates the probability that a customer will behave in a certain way on the basis of one or more independent inputs. It is important to stress the fact that the model is linear, i.e. it can discover only linear relation between a customer's features and behaviour. Non-linear mapping between customer's features and customer's response can be modelled by a feed-forward neural network (NNs) which uses a layer of hidden processing units.

The output of the model is squashed into a sigmoid function; this enables us to interpret the output value of the model as the probability that a customer is not satisfied with the level of service.

The parameters of the LR and NNs have been optimised by a Conjugate gradient; this algorithm is suitable for large problems (memory requirements are linear in the number of parameters) and it is fast in converging to a local minimum (as it uses information about the curvature of the objective function). In our project, we tested NNs with up to 8 hidden nodes. As the generalisation errors reported by NNs were not significantly lower than the one of the LR, we modelled customer satisfaction with a LR.

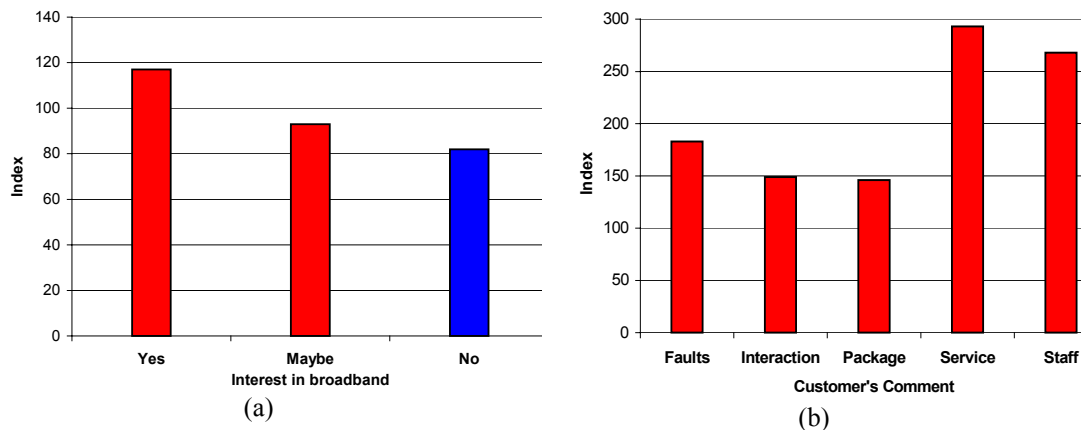
### 4. The Key Attributes and Lift chart

As stated in the introduction, main object of the project is to gain a better understanding of the drivers of customer satisfaction among survey's respondents. In the present section we present the list of key attributes identified during the analysis. We have identified four categories (shown in Table 1):

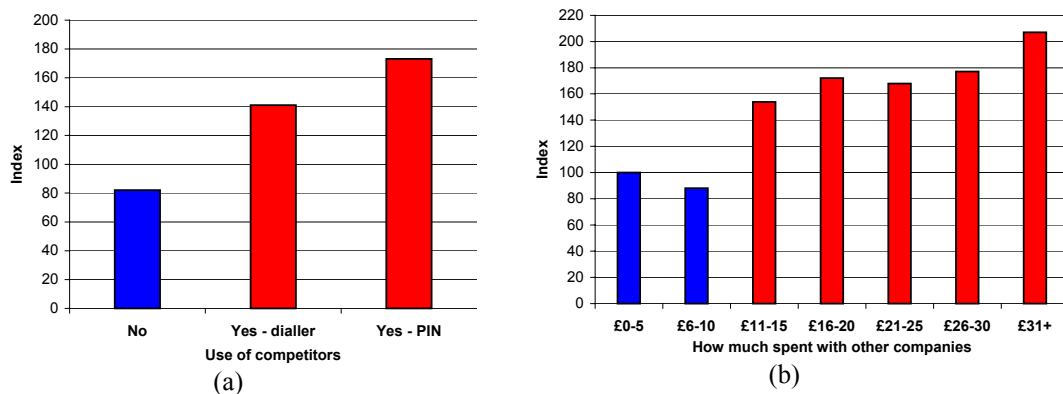
- BT PRODUCTS** This category collects attributes directly linked with BT services, such as the size of quarterly bill, calls made to Internet providers, satisfaction with Internet access, BT package ownership, interest in Broadband. We note that customers with interest in Broadband are more likely to be dissatisfied (see Figure 1 (a)); also high is the percentage of dissatisfied customer without any BT package and those who ask simpler package offers from BT (data not shown).
- BT SERVICE** The comments most associated with a high index of dissatisfied were about Faults handling, Interactions, and Service improvement. Figure 1 (b) shows the index against those comments (the legend explains how we derived the dissatisfied index); we note that, among customers who suggested to improve Service and Staff, percentage of dissatisfied is almost three times higher than the average.
- COMPETITOR** attributes are related to areas such as mobile phones (bill size and number of handsets in the household) and use of competition. The mobile division of BT has been de-merged about two years ago; nevertheless many customers still tend to identify the mobile services as they were offered by BT. We also note that the use of competitors is related to dissatisfaction with BT services; see Figure 2.
- DEMOGRAPHIC** attributes collect customer's personal attributes such as occupation, income, age.

These key attributes have been fed into a LR to evaluate customer's propensity to be dissatisfied with BT's services and products. The lift chart of the model is shown in Figure 3.

In the chart, customers of the test set have been scored and sorted in descending order according to the dissatisfaction scored obtained (reported on the x-axis).



**Figure 1** The histograms show the index of dissatisfied against future interest in Broadband (a) and customers' comments (b). The Index is calculated by the ratio between the percentage of dissatisfied customers in a certain category (e.g. those with interest in Broadband) and the percentage of dissatisfied survey's respondents (in hundredths). A percentage of dissatisfied customer higher (lower) than average corresponds to an index higher (lower) than 100. In both charts dissatisfaction index higher (or lower) than 100 is shown in red (or blue). Interest in Broadband has a high index of dissatisfied customers.



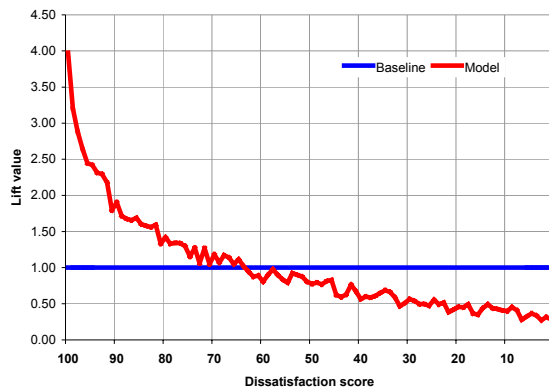
**Figure 2** The histograms show percentage of dissatisfied customers against use of competitors (a) and bills with competitors (b). In both charts, percentage of dissatisfied customer higher (or lower) than the average is shown in red (or blue). We note that higher spenders (£11 or above) in competitors' services are more likely to be dissatisfied with BT services.

The y-axis reports the lift obtained by the model against the baseline (e.g. in the top percentile, the model is able to identify 4 times more dissatisfied customers than the random model). The performance reported by the model is significantly better than the baseline, suggesting that the model can be employed in estimating propensity to be dissatisfied among survey's responders. The cut off point is at dissatisfaction score 63.

## 5. Actions taken

The analysis carried out has confirmed our current understanding of the attributes which drive customer satisfaction and helped us in prioritising the efforts to improve our services. In the following we list some of the actions taken.

- **Renewal of our commitment to Broadband.** Broadband is already at the heart of BT strategy and will continue to be so. Our target is to connect one million broadband customers by summer 2003 and to increase this to five million connections in 2006. This will be achieved through lower prices, a wide range of applications, aggressive marketing techniques as well as by increasing the number of broadband enabled areas. In the near future our focus will be on reducing costs and prices of broadband, developing the broadband applications and extending the reach of broadband to as much to the United Kingdom as is commercially possible.
- **Improvements in customer contacts.** We have identified the need to deliver an outstanding contact centre and customer interaction as key to improve satisfaction with BT.



**Figure 3** The charts show the lift value obtained for the Customer Satisfaction model against the dissatisfaction score. Lift value of the model and the baseline are shown in red and blue, respectively. On the top percentile, the model is able to identify 4 times more customers with a low level of satisfaction than the baseline. Cut off points was at percentile 63. The results reported are obtained on the test set.

Over the past months BT has launched a new contact centre organisation which improves channel management and relationship with customers.

- **Proposition of simpler packages.** Recently we have proposed a revised BT Together package The package apply to local and national calls and offers three options; the first has a flat rate of 6 pence/hour (evening and weekend); the seconds offers free call during evening/weekend; option three introduces free calls anytime. This should address the requests of simpler and clearer packages highlighted in the survey. We have also introduced packages for international calls.

## 6. Summary and Future work

In BT we consider customer satisfaction of paramount importance. In this paper we presented a model of customer satisfaction with BT services and products based on the data collected via the BT customer survey. The analysis has confirmed our current understanding of the attributes which drive customer satisfaction; however for the first time the analysis was based on the “real voice” of our customers.

We have used a LR model based on four families of drivers, i.e. BT PRODUCTS, BT SERVICE, COMPETITOR and DEMOGRAPHIC. To summarise the main findings, customers most likely to be dissatisfied are those with a high level of sophistication, interested in new technologies, and familiar with BT’s competitors. The comments most strongly related to dissatisfaction were those which suggested to simplify BT’s packages, to improve service and faults handling. The model has shown a remarkable uplift of 4 in the top percentile. We have also shown some of the early actions taken as result of this analysis.

In future we will work together with our marketers colleagues to develop new services and packages that will enhance customer experience of BT and improve satisfaction levels. We also plan to extend a similar analysis to the whole BT base, developing a model that can estimate dissatisfaction level of all BT’s customers.

## References

- [1] Berry, M.J.A. and Linoff, G.S. (1997), **Data Mining Techniques for Marketing, Sales and Customer Support**, New York, John Wiley & Sons.
- [2] SAS Institute Inc.(2000), **Getting Started with Enterprise Miner Software, Release 4.1**, Cary, NC, SAS Institute Inc.

---

\* Data Mining manager

BT- Segmentation Marketing and Customer Data

Pp501 Holborn centre, 120 Holborn

London EC1N 2TE

e-mail: francesco.vivarelli@bt.com

<sup>1</sup> We also sent a different survey to non-BT customers; however these data have not been considered for the present exercise.

<sup>2</sup> Note that the use of TB information is subject to rigorous and complex regulation and can be employed only in a limited number of situations; on this occasion we were allowed to use TB data in our analysis. Because we use data that are aggregated and anonymous, we will not be contacting customers as a results of this analysis.