

**Customer Churn Analysis in the Wireless Industry:  
A Data Mining Approach**

**Shyam V. Nath  
Daleen Technologies  
902 Clint Moore Road Suite 230  
Boca Raton, FL 33487**

**Tel: (561) 981-2146  
snath@daleen.com**

**Dr. Ravi S. Behara  
ITOM  
Florida Atlantic University  
777 Glades Road  
Boca Raton, FL 33431, U.S.A.**

**Tel: (561) 297-2778  
rbehara@fau.edu**

# **Customer Churn Analysis in the Wireless Industry: A Data Mining Approach**

## **Abstract**

This paper presents a customer churn study in the wireless telecommunications industry. An Oracle database of fifty thousand real customers was analyzed using the Naïve Bayes algorithm data mining option for supervised learning that was implemented through JDeveloper. The results of the study, and future directions, are then discussed.

## **INTRODUCTION**

During the last five years, the wireless telecomm sector has been one of the fastest-growing businesses in the economy although the technology sector has taken a big hit in recent years. With a unique value proposition – freedom and connectivity – the number of subscribers doubled every two years during the 90's. These events shaped the new telecommunications landscape, as we know it today and more promising developments are to come (Duke Teradata 2002):

- By 2003, 25% of all telephone minutes will be accounted for by wireless services.
- By 2006, the US penetration in the wireless-voice market is expected to hit 189 million subscribers. At the same time the wireless-data market is expected to jump to 38 million subscribers highlighting the rapid growth of wireless data relative to wireless voice.
- Of all wireless customers, 70% are using digital networks that allow carriers to efficiently offer more appealing services.

- Investment in network infrastructure has increased by 17% and the number of cell sites increased by 22.3%, indicating a clear upward trend in US coverage and quality.

Despite the vertiginous levels of growth and promise, serious changes to industry profitability have recently emerged: (a) Consolidation: From the nearly 60 cellular companies, virtually all of them are now bankrupt, bought out, or struggling with heavy debts. Only six big players now account for 80% of the wireless pie. (b) Growth: Subscriber growth rates went from 50% yearly to 15%-20% in 2002, and analysts predict a meager 10% growth rate in 2003. (c) Competition: As an obvious result (and to the consumer delight), firms engaged in a devastating price war that not only eroded revenue growth but also endangered their ability to meet their titanic debts. (d) Customer Strategy: The industry paradigm has arguably changed from one of "make big networks, get customers" to "make new services, please customers." In short, the industry has moved from an acquisition orientation to a retention orientation. (Duke Teradata 2002)

Until now, firms have been able to acquire customers without much effort. Demand for wireless services has been such that if a customer decided to drop the service and switch to another carrier, another new customer was right behind. The priority was to maintain the customer acquisition rate high, often at the expense of customer retention. But this situation has changed (Mozer et al 1999). As the well of new wireless subscribers has begun to run dry, churn - that can be looked as the customer's decision to end the relationship and switch to another company - has become a major concern. Industry studies estimate that one-third of US wireless subscribers will churn annually, or decide to switch from one carrier to another (In-Stat 2002). In 2001 the industry average churn rate was 20%-25% annually, which

translates to approximately 2% churn per month. This means that companies lost 2% of their customers every month. Third quarter, 2001, statistics show annual churn rates in an even higher range, 28%-46% annual churn (Duke Teradata 2002).

The following are the reasons for the high level of churn: (a) many companies to choose from, (b) similarity of their offerings, and (c) cheap prices of handsets. In fact, the biggest current barrier to churn - the lack of phone number portability - is likely to change in the short term. Wireless Local Number Portability (WLNP) or the ability to change mobile carriers and keep one's mobile phone number poses a big challenge to the already reeling Telecommunication Industry. The Federal Communications Commission (FCC) has stated that wireless carriers must implement WLNP in November 2003, giving carriers less than a year to make the necessary changes. WLNP is an added dimension that threatens to significantly impact wireless churn rates (In-Stat 2002, Skedd 2002).

The churn rate for U.S. mobile carriers is 2 % to 3 % monthly, a major expense for the companies, which spend \$400 to \$500 to sign a single customer who typically generates about \$50 in monthly revenue. Companies are now beginning to realize just how important customer retention is. In fact, one study finds that "the top six US wireless carriers would have saved \$207 million if they had retained an additional 5% of customers open to incentives but who switched plans in the past year" (Duke Teradata 2002). Over the next five years, the industry's biggest marketing challenge will be to control churn rates by identifying those customers who are most likely to leave and then taking appropriate steps to retain them. The first step therefore is predicting churn likelihood at the customer level.

This paper presents a study based on predictive modeling using data mining to predict churn rate of subscribers in the wireless industry. The study uses Oracle 9 database technology with the Data Mining capabilities to identify key customer characteristics to predict churn. It is expected that, with a better understanding of these characteristics, managers can develop a customized approach to customer retention activities within the context of their Customer Relationship Management efforts.

## **DATA MINING**

Data Mining is the process of using raw data to infer important business relationships. Once the business relationships have been discovered, they can then be used for competitive advantage. Despite this general consensus of the value of data mining, a great deal of confusion exists as to just exactly what it is. Is data mining simply a fancy name for data analysis? Does it require special tools and special knowledge?

Statistical tools are used to draw conclusions from representative samples taken from larger amounts of data. They are useful for finding patterns and correlations in “small to medium” amounts of data but fall short when the amount of data begins to overwhelm the tool. Typically, when we deal with greater than, say, 25 input variables and tens of thousands of records, traditional statistical regression techniques struggle. Because statistical tools can’t analyze all the data, they force data analysts to use representative samples of the data and to eliminate input variables from the analysis. By throwing out variables and using samples of the data, we are throwing away “information.” (Oracle9i Data Mining 2001)

Data mining doesn’t have these limitations. Data mining goes deep into the data. It uses machine-learning algorithms to automatically sift through each record and variable to

uncover patterns and information that may have been hidden. Although most modern data mining packages still offer the classical statistical approaches, data mining has moved far beyond these first-generation statistical measures to more insightful and powerful approaches that assist in explaining or predicting “what is going on in the data.” Traditional statistical methods of analysis typically build linear models that are inadequate when dealing with non-linear and complex data patterns. In addition, these models are negatively affected data that contain large numbers of outliers. Further business data has non-numerical data. These three characteristics of non-linearity, non-numeric data, and the presence of significant number of outliers requires other data mining tools. Specifically, this study discusses and uses predictive algorithms such as Predictive Association Rules and Transactional Naive Bayes that are suited to create models for churn prediction.

Currently, data mining users are mostly business-to-business marketers and service retailers (Armstrong and Kotler 2001). However, firms from all areas of business are beginning to understand the power of data mining. These firms use their data warehouses to learn as much as possible about their customers and target promotions toward the customers’ needs. For example, Kraft has data on 30 million users of its products. Kraft uses data mining to identify the special interests of these customers to send them targeted information such as recipes, nutritional guides, and coupons. Fingerhut sends surveys to its customers and uses the results to customize catalogs with products most likely to appeal to its consumers. As the use of data mining grows, firms will continue to get more creative in their applications. One example of creativity is American Express’ new customer loyalty initiative. American Express has a database that links cardholders with their respective zip codes. When a new

restaurant opens in a specific zip code, AMEX sends the appropriate cardholders coupons to the new restaurant. Initiatives like these form the future uses of data mining.

## **NAIVE BAYES**

Naive Bayes is a type of supervised-learning module that contains examples of the input-target mapping the model tries to learn. Such models make predictions about new data based on the examination of previous data. Different types of models have different internal approaches to learning from previous data. The Naive Bayes algorithm uses the mathematics of Bayes' Theorem to make its predictions. The algorithm is typically used for:

- Identifying which customers are likely to purchase a certain product
- Identifying customers who are likely to churn
- Predicting the likelihood that a part will be defective

Bayes' Theorem is about conditional probabilities. It states that the probability of a particular predicted event, given the evidence in this instance, is computed from three other numbers: the probability of that prediction in similar situations in general, ignoring the specific evidence (this is called the prior probability); times the probability of seeing the evidence we have here, given that the particular prediction is correct; divided by the sum, for each possible prediction (including the present one), of a similar product for that prediction (i.e., the probability of that prediction in general, times the probability of seeing the current evidence given that possible prediction).

A simplifying assumption (the "naive" part) is that the probability of the combined pieces of evidence, given this prediction, is simply the product of the probabilities of the

individual pieces of evidence, given this prediction. The assumption is true when the pieces of evidence work independently of one another, without mutual interference. In other cases, the assumption merely approximates the true value. In practice, the approximation usually does not degrade the model's predictive accuracy. The supervised-learning modules using Naive Bayes Theorem have the advantages of simplicity and speed.

### Transactional Naive Bayes

"Transactional Naive Bayes" refers to the way the input is formatted; the algorithm is the same. The table below shows an example of traditional data format, with columns for the items (customer, apples, oranges, pears, and bananas), rows for the customers (Joe, Jim, Jeff), and zeroes or ones in each table cell, indicating whether, for example, Joe bought an apple (no), an orange (no), a pear (no), or a banana (yes):

	apples	oranges	pears	bananas
Joe	0	0	0	1
Jim	1	0	0	1
Jeff	0	1	0	0

Traditional data layout often produces a sparse matrix because of all those zeroes; it takes up more space in the database, and therefore, takes more time in calculations. Transaction-based format has basically two columns: customer and "hits."

For Joe, the table cell contains "bananas":

Joe	bananas
Jim	apples, bananas
Jeff	oranges



Transactional format looks like a "shopping basket" rather than a checklist and is better in cases where the customers buy only subsets of products. Transactional format has the advantage of being the way the data is stored in the database for this type of problem.

### **Supervised v/s Unsupervised Learning**

Data mining can be separated into “supervised learning” and “unsupervised learning” techniques. Supervised learning requires the data analyst to identify a target field or dependent variable. The supervised-learning technique then sifts through data trying to find patterns and relationships between the independent variables and the dependent variable. (Oracle9i Data Mining provides the Naïve Bayes data mining algorithm for supervised-learning types of problems.) In “unsupervised learning,” the user does not indicate the objective to the data mining algorithm. Associations and clustering algorithms make no assumptions about the target field. Instead, they allow the data mining algorithm to find associations and clusters in the data independent of any a priori defined business objective. Oracle9i Data Mining provides the Association Rules data mining algorithm for unsupervised-learning problems.

### **DATA MINING WITH ORACLE 9i**

Oracle has embedded data mining within the Oracle9i database with Oracle9i Data Mining (ODM). This feature enables Oracle9i to provide an infrastructure for application developers to seamlessly integrate data mining with the database. Data mining functions, such as model building, scoring functions, and testing are provided via a Java API. The Java API provides complete programmatic control of data mining functions to deliver data mining within the database.

Oracle9i Data Mining simplifies the process of extracting business intelligence from large amounts of data. It eliminates off-loading vast quantities of data to external special-purpose analytic servers for data mining and scoring. With Oracle9i Data Mining, all the data mining functionality is embedded in Oracle9i Database, so the data, data preparation, model building, and model scoring activities remain in the database. Because Oracle9i Data Mining performs all phases of data mining within the database, each data mining phase results in significant improvements in productivity, automation, and integration. Significant productivity enhancements are achieved by eliminating the extraction of data from the database to special purpose data mining tools and the importing of the data mining results back into the database. These improvements are notable in data preparation, which often can constitute as much as 80% of the data mining process. With Oracle9i Data Mining, all the data preparation can be performed using standard SQL manipulation and data mining utilities within Oracle9i Data Mining. “Scoring” to make predictions is greatly simplified now, as the data and the model are in the same location — the database.

### **Major Steps Of Data Mining**

The Data Mining Server (DMS) is the server-side, in-database component that performs the actual data mining operations within an ODM implementation. The DMS also provides a metadata repository consisting of mining input objects and result objects, along with the namespaces within which these objects are stored and retrieved. The DMS is embedded in the Oracle9i database, and, hence, benefits from the scalability and availability features of the Oracle9i database.

We will discuss the data mining operations now. Oracle9i Data Mining supports two data mining functions: *classification* for supervised learning and *association rules* for

unsupervised learning. The mining functions use two algorithms: Naive Bayes and Association Rules. Users can build models on transactional or non-transactional data using these algorithms. For the Naive Bayes algorithm, ODM supports additional operations for building a model on calibration data, testing a model's accuracy on new data, computing lift and applying a model to data (scoring). For Association Rules, ODM, provides the ability to retrieve and analyze model rules.

### **Build Model**

The ODM Java API supports are two levels of settings: *function* and *algorithm*. When the function level settings do not specify particular algorithm settings, ODM can choose an appropriate algorithm, providing defaults for the relevant parameters. In general, model building at the function level makes many of the technical details of data mining transparent to the user. Models are built in the data-mining server (DMS). After a model is built, it is persisted in the DMS and can be accessed by its user-specified unique name.

The typical steps for model building are as follows:

1. Create input data (by associating a mining data object with existing data, for example, a table or file).
2. Create a function settings object.
3. Create a logical data specification and associate it with the function settings.
4. Create a data usage specification and associate it with the function settings.
5. Create algorithm settings (optional).
6. Invoke the build method.

### **Test Model**

Model testing gives an estimate of model accuracy. The data analyst can test classification models, as produced by the Naive Bayes algorithm. After a model is built, model testing computes the accuracy of a model's predictions when the model is applied to a new data set. The test results are stored in a mining test result object. A classification test result includes a *confusion matrix* that allows a data miner to understand the type and degree of classification errors made by the model. The test operation accepts the name of a previously built model and data for testing the model. The test data must conform to the logical data specification used for building the model.

### **Apply Model**

Applying a supervised learning model to data results in *scores* or *predictions* with an associated probability. The analyst can score classification models, as produced by the Naive Bayes algorithm. The data to be scored must have attributes compatible with the training data, that is, it must have a superset of attributes with the same names and respective data types or a suitable mapping. The result of the apply operation is placed in the schema specified by the user.

The ODM user specifies the result content. For example, a user may want the customer identifier attribute, along with the score and probability, to be output in a table for each record in the provided mining data. This is specified using the MiningApplyOutput class. ODM supports the apply operation for a table (a set of records) or a single record (represented by a Java object).

### **Compute Lift**

ODM supports computing lift for a binary classification model, as produced by the

Naive Bayes algorithm where the target attributes takes on exactly two values. In this study these two values are 1 for customers churning and 0 for customers not churning. Given a designated positive and negative value, test cases are sorted according to how confidently they are predicted to be positive instances (most confidently positive come first; most confidently negative come last).

## **DATA MINING FOR CHURN ANALYSIS**

### **Data For Modeling**

The data mart for the study was created using data provided by a wireless telecom company and includes more than 150 variables describing more than 100,000 customers. The data was obtained from The Teradata Center for Customer Relationship Management at Duke University. The data consist of calibration and validation samples of customers from a major wireless telecommunications company. The calibration sample includes observed churn and a set of potential predictor variables. The two validation samples include the same predictor variables, but no churn variable. The data are organized into three data files: Calibration, Current Score Data, and Future Score Data. See table below.

	Calibration	Current Score Data	Future Score Data
Sample Size	100,000	51,306	100,462
# of Predictor Variables	171	171	171
Churn Indicator	Yes	No	No
Customer ID	1,000,001– 1,100,000	2,000,001– 2,051,306	3,000,001– 3,100,462

*Table 1: Nature of Dataset Used for Study*

The Calibration Data contain the dependent variable – churn – as well as several potential predictors. The Future Score Data contain the predictors but not churn. The study develops models on the calibration data and uses these models to predict for the Current and Future Score Data using the Naïve Bayes algorithm. The predictors include three types of variables: *usage data* such as minutes of use, revenue, handset equipment; *company interaction data* such as customer calls into the customer service center; and *customer household demographics*.

Customers were selected as follows: mature customers who were with the company for at least six months, were sampled during July, September, November, and December of 2001. For each customer, predictor variables were calculated based on the previous four months. Churn was then calculated based on whether the customer left the company during the period 31-60 days after the customer was originally sampled. The one-month treatment lag between sampling and observed churn was for the practical concern that in any application, a few weeks would be needed to score the customer and implement any proactive actions.

The actual percentage of customers who churn in a given month is approximately 1.8%. However, churners were over sampled when creating the Calibration sample to create a roughly 50-50 split between churners and non-churners (the exact number is 49,562 churners and 50,438 non-churners). Over sampling was not undertaken in creating the Current Score and Future Score validation samples. This is to provide a more realistic predictive test. The Current Score data contain a different set of customers from the Calibration data, but selected at the same point in time. The Future Score data contain a different set of customers selected

at a future point in time. One interesting aspect of study will be to investigate the accuracy for same-period versus future predictive accuracy.

### **Characteristics of Input Data**

Ultimately, churn occurs because subscribers are dissatisfied with the price or quality of service, usually as compared to the offerings of competing carriers. The main reasons for subscriber dissatisfaction vary by region and over time. We categorize our input variables as follows.

- Demographics: Geographic and population data of a given region.
- Usage level. Call detail records (date, time, duration, and location of all calls), peak / off-peak minutes used, additional minutes beyond monthly prepaid limit etc.
- Quality of Service (QOS): Dropped calls (calls lost due to lack of coverage or available bandwidth), and quality of service data (interference, poor coverage).
- Features / Marketing: Details of service bundle such as email, instant messaging, paging, rate plans offered by carrier and its competitors, recent entry of competitors into market, advertising campaigns, etc.

A subset of these information sources was used in our modeling study. The study was conducted over a fairly short time interval during which the market did not change significantly. More important, the market forces were fairly uniform in the various geographic regions from which the subscribers were selected.

### **Results And Analysis**

For this study an Oracle 9.2 database on Windows 2000 platform with Data mining option and Jserver was setup. The database comes with a user called ODM for Oracle Data mining. This user has to be unlocked using the SQL (Sequential Query Language) command: `ALTER USER ODM ACCOUNT UNLOCK;` Additionally, we need to unlock the user ODM\_MTR.

The data for processing was obtained as text files and it was imported into Oracle database using the SQL Loader. This helps to take the data into the database and create a data mart to allow data mining. The data mining features of Oracle can be invoked using the java calls that require Jdeveloper. The product called DM4J provides the wizards for generating the Java components for the data modeling.

The study developed the Naïve Bayes model and when tested the same against a known set of 2000 customers, showed 68% accuracy. The confusion matrix below provides the basis for the coefficients for the churn management model.

<b>Predicted / Real</b>	0	1
0	<b>34</b>	766
1	56	<b>1144</b>

*Table 2: Confusion Matrix: Horizontal values are real, vertical values are predicted*

For each churn predictor, we obtained an estimate of the probability of churn for each subscriber in the data set, based on decision making that ultimately requires a “churn” or “no churn” prediction, the continuous probability measure were thresholded to obtain a discrete predicted outcome. In the telecommunications industry the outcome is often expressed using a *lift curve*. The lift curve is related to the signal detection theory used in telecommunication.



For a given threshold on the probability of churn, we determine two quantities: the fraction of all subscribers having churn probability above the threshold, and the fraction of all churners having churn probability below the threshold and this information is used to populate the discrete values of 0 and 1 in the churn column.

The top ten attributes for churn in the order of importance were found to be the following. The DM4J also provides a tool to rate the importance of the various predictor variables.

1. DUALBAND type of phone set
2. CARTYPE dominant vehicle lifestyle
3. EDUC1 education level of first house hold member
4. ETHNIC ethnicity
5. TOT\_ACPT total offers accepted from retention team
6. OCCU1 occupation of the first household member
7. AREA geographic area
8. INCOME estimated household income
9. DWLLSIZE dwelling size
10. PROPTYPE property type details

This study also compared the Adaptive Bayes Network and the Naïve Bayes Network for building the data model and the later took less time (3:43 min v/s 1:45min) for a sample set of 5000 records. Finally, the model was used to predict churn. When we apply the model, the predictions are obtained and stored in an output table as shown in Figure 1 below.

*Figure 1: Sample Prediction Data*

Name: Th6\_ClassificationModelApply  
 Start date: 3/29/03  
 Start time: 5:31 PM  
 End date: 3/29/03  
 End time: 5:32 PM

Model name: [Th5\\_ClassificationModelBuild](#)

Input data  
 Schema: ODM\_MTR  
 Table/View: FUTURE

Output data  
 Schema: ODM\_MTR  
 Table/View: Th6\_out

Apply output table:  
 Fetch Size:

PREDICTION	PROBABILITY	CUSTOMER_ID
0	9.2941524029719382259821 ...	2020248
1	9.9999972618918957926043...	2019332
1	6.4758623972545805586244...	2019333
1	9.9999986550717632452205...	2019337
1	5.8554839539631816440402...	2019345

## CONCLUSION

This study successfully developed a working database system based data mining to predict the churn for the wireless industry. The study built and tested a predictive model using wireless customer data. The model obtained a 68% predictive accuracy in the first pass using the Naïve Bayes algorithm. The model was then used to predict the future churn for the wireless customers. Thus the study successfully developed a framework using available database technologies that can help the customer to “mine their own business”. The study

will be extended to devise an intervention strategy based on the churn prediction to reduce the lost revenue by increasing customer retention.

As more and more companies start to “mine their own data,” the industry will have increasing need to make this information available to customer operations and business managers in real time. Predictive real-time analytical tools can then enable organizations to mine their existing customer information to develop models of predictive customer behaviors and link these predictive models to the customer contact system in real time, enabling the customer-facing agents to make cross- and up-sell offers that have the greatest likelihood of acceptance, the lowest risk and the greatest return value to the organization. Web enabling the database servers and data mining engines will facilitate direct delivery of these real time outputs to the customer facing staff.

## REFERENCE LIST

Armstrong, G., and P. Kotler. 2001. *Principles of Marketing*. Prentice Hall New Jersey.

Duke Teradata 2002. Teradata Center for Customer Relationship Management. [On-line]. Retrieved on: Nov 7, 2002. Available:

[http://www.teradataduke.org/news\\_t\\_2.html](http://www.teradataduke.org/news_t_2.html)

In-Stat. 2002. WLNP Threatens to significantly impact wireless churn rates. [Online].

Retrieved on Sep 2002. Available: <http://www.instat.com/newmk.asp?ID=312>

Mozer, Michael, Richard Wolniewicz, Eric Johnson and Howard Kaushansky. 1999. Churn reduction in the wireless industry, *Proceedings of the Neural Information Processing Systems Conference*, San Diego, CA.

Oracle9i Data Mining 2001. An Oracle white paper December 2001. [Online].

Retrieved on: Nov 8, 2002. Available:

[http://otn.oracle.com/products/bi/pdf/o9idm\\_bwp.pdf](http://otn.oracle.com/products/bi/pdf/o9idm_bwp.pdf))

Skedd, Kirsten 2002. WLNP threatens to significantly impact wireless churn rates [On-line].

Retrieved on Sep 14, 2002. Available:

<http://www.instat.com/press.asp?ID=311&sku=IN020258WP>