**RC22473 (W0203-002) March 1, 2002**
**Computer Science**

# IBM Research Report

## A Machine-Learning Approach to Optimal Bid Pricing

**Richard D. Lawrence**
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

# A Machine-Learning Approach to Optimal Bid Pricing

Richard D. Lawrence
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598
ricklaw@us.ibm.com

May 17, 2002

**Abstract**

We consider the problem faced by a seller in determining an optimal price to quote in response to a Request for Quote (RFQ) from a prospective buyer. The optimal price is determined by maximizing expected profit, given the underlying seller costs of the bid items and a computed probability of winning the bid as a function of price and other bid features such as buyer characteristics and the degree of competition. An entropy-based information-gain metric is used to quantify the contribution of the extracted features to predicting the win/loss label. A naive Bayes classification model is developed to predict the bid outcome (win or loss) as a function of these features. This model naturally generates the win probability as a function of bid price required to compute the optimal price. Results obtained by applying this model to a database of bid transactions involving computer sales demonstrate statistically significant lift curves for predicting bid outcome. A method for creating additional synthetic bids to improve computation of the win probability function is demonstrated. Finally, the computed optimal prices generated via this approach are compared to the actual bid prices approved by human pricing experts.

## 1 Introduction

For centuries, sellers have faced the challenge of setting prices for goods and services that will generate maximum profit. Until relatively recently, pricing strategies have focused largely on setting fixed prices for items sold via conventional sales channels like stores and catalogs. The advent of electronic commerce has enabled buyers to readily access information on pricing and other competitive product features, as well as creating the infrastructure for sellers to easily adjust prices to reflect near-term changes in demand and/or competitor's prices. More broadly, dynamic [7] or flexible [2] pricing strategies establish prices that

1. may change over time (temporal dynamic pricing)
2. may differ across different buyers (price differentiation [17])
3. may depend on bundling with other products and services (product differentiation [1]).

The stock market and electronic auctions such as eBay [5] are two everyday examples of dynamic pricing. Other realizations include adjustment of prices as a result of on-line experiments [9] to measure customer price sensitivity for specific commodity-like items such as books. Revenue management strategies [11] introduced in the airline industry in the 1970's led to the first widespread use of differentiated pricing, where different passengers may pay different fares based on booking class. Product differention [1] is crucial in commodity markets: two otherwise identical products may carry different consumer valuations due to bundling with other related products, or with services and delivery terms. The problem addressed in this paper aligns more closely with the second and third issues, namely how do we establish a price on a bundles of goods for different classes of buyers, given a database of similar historical bid transactions and the outcomes of these bids (won or loss). This problem is discussed further in the following section.

## 2 The Bid Pricing Problem

We consider the following problem. A prospective buyer issues a Request for Quote (RFQ) for a product configuration with a set of minimal requirements that must be satisfied by the offering. For example, a buyer may request 1000 laptop computers, with specific requirements on processor speed, memory, hard drive size, and other characteristics, possibly including services such as installation and financing. These laptops may be combined or bundled with a different number of desktop computers with different characteristics. In this paper, we consider only relatively large RFQs, as measured by quantity and total potential revenue; small transactions, such as the purchase of a single laptop computer, are generally not conducted via a formal RFQ process. In a business-to-business environment, the buyer may advertise the RFQ to a select list of suppliers or sellers. Each seller composes a bid response comprised of a specific configuration designed to meet the RFQ requirements, along with pricing information and possibly other terms and conditions. We will use the term *pricer* for the expert within the seller organization who prepares this response.

Figure 1(a) illustrates this process. In this paper, we are in effect acting as the seller, seeking to generate an optimal price in response to the buyer-initiated RFQ. We assume $C$ additional competitive sellers who have received the identical RFQ. The bidding process can be a single, sealed-bid response, but in many cases there may be subsequent rebids in which either the product offered is changed, and/or the price and terms and conditions are modified. This is shown as a series of bids $B_t^c$ in Figure 1(a). Ultimately, the buyer selects a bid winner based on a comparative evaluation of bid characteristics.[1] In multi-attribute bids, it is possible to win a bid without necessarily having the lowest price.

In constrast to an open auction, we are generally unable to observe all of the information shown in Figure 1(a). Figure 1(b) shows the information actually available to the target seller: all bid iterations on behalf of this seller are known, of course, but we generally do not know with any confidence the bid sequences generated by the competitors. Indeed, we know for certain only the outcome (win or loss), and our sequence of bids culminating in this outcome. Even the number and identity of the competitors may not be known with certainty.

A recent review [13] of bid pricing models and supporting tools summarizes some of the

[1] Quantitative techniques for buyer-evaluation of multi-attribute bids are discussed in [6].

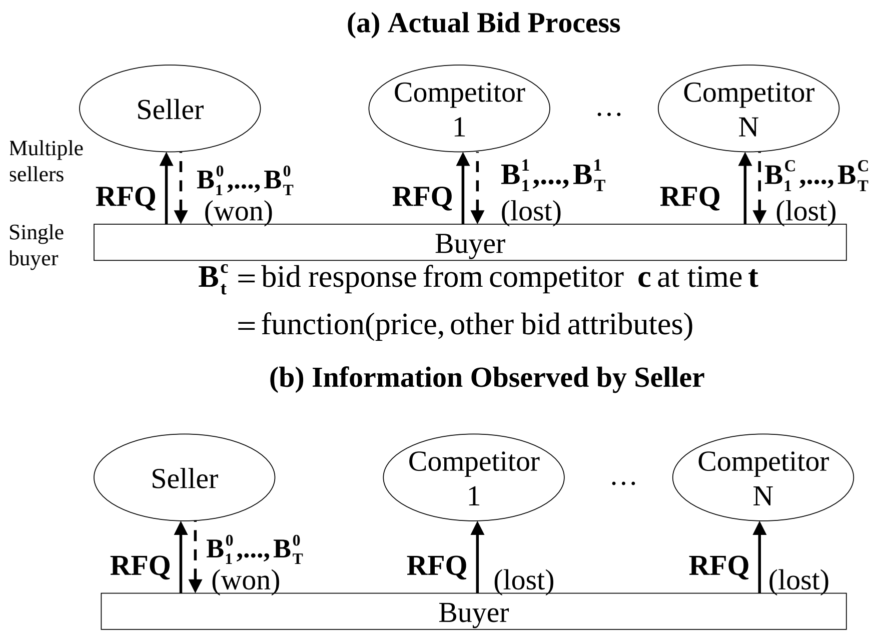


Figure 1: Bid response to a Request for Quote (RFQ).

earliest models in this area. These early statistical models assumed complete knowledge of competitor's bidding history, with the winner determined by the lowest price. The more recent ServPrice model [3][4][13] does not require historical data, but rather utilizes input from human experts to establish the probability of winning under different scenarios.

The pricing methodology presented here follows a more conventional machine-learning approach: rather than relying on statistical models with estimated parameters or expert classification of potential bid scenarios, we attempt to learn directly the probability of winning from a set of bid transactions with known outcomes. One disadvantage of this approach in that we require extensive historical data. On the other hand, we do not need to make explicit assumptions about competitive behavior. Even for experts, predicting bid outcomes can be perplexing and prone to errors, and therefore it is an advantage to rely on actual prior transactions to accurately characterize the probability of winning a given bid.

Section 3 describes the methodology, and Sections 4 and 5 summarize results obtained from the analysis of a set of actual transactions generated in response to RFQs for computer equipment.

# 3 Methodology

In this section, we develop a classification methodology for predicting the bid outcome (win or loss), and then show how the classification method can be used to generate the win probability function required to optimize the offering price.

### 3.1 Price Optimization

We begin by defining a set of random-variable features characterizing each item included in a bid response. Upper-case letters will be used to denote these features, with the values of the features given in lower-case. Let $X_p$ denote the offered price, and $X_m, m = 1, \ldots, M$ denote $M$ non-price features. Combining all features yields the feature vector

$$\mathbf{X} = [X_p, X_1, \ldots , X_M]. \tag{1}$$

Each bid item $n = 1, \ldots , N$ is described by the vector of feature values

$$\mathbf{x}^n = [x_p^n, x_1^n, \ldots , x_M^n], \tag{2}$$

where $x_p^n$ is the offered price for bid item $n$, and $x_m^n$ is the value of the non-price feature $m$ for bid item $n$. Examples of non-price features may include information about the prospective buyer (*e.g.* price sensitivity), identity of the competitors, seller's cost, posted or list price, and seller incumbency with the prospective buyer. Each historical bid is tagged with a binary output variable $w^n$ denoting the target label (won or loss), and thus each bid item $B^n$ can be represented as

$$\begin{aligned} B^n &\equiv \{\mathbf{x}^n, w^n\} \\ w^n &\in [\text{win}, \text{loss}]. \end{aligned} \tag{3}$$

In specifying a price for an item in a bid offering, the seller must effectively balance the enhanced likelihood of winning the bid with a lower price, versus the increased profits at higher offered prices. The expected profit associated with bid item $n$ is

$$\text{Profit}(x_p^n) = P(\text{win}|\mathbf{X} = \mathbf{x}^n) \left[x_p^n - C^n\right], \tag{4}$$

where $C^n$ is the seller's base cost, and $P(\text{win}|\mathbf{X} = \mathbf{x}^n)$ is the conditional probability of winning a bid given feature values $\mathbf{x}^n$, including price $x_p^n$. We adopt the simpler notation $P(\text{win}|\mathbf{x}^n)$ to denote $P(\text{win}|\mathbf{X} = \mathbf{x}^n)$, and so on. The price that optimizes the expected profit is

$$\hat{x}_{p,Opt}^n = \arg\max_{x_p^n} \left\{P(\text{win}|\mathbf{x}^n) \left[x_p^n - C^n\right]\right\}. \tag{5}$$

Computation of the win probability $P(\text{win}|\mathbf{x}^n)$ is the major challenge in the evaluation of equation (5). One approach would be to assume knowledge of the competitor's bidding stategy under circumstances similar to the current bid, and compute the win probability based on the probability of the offered price being lower than that of the competitor:

$$P(\text{win}|\mathbf{x}^n) \equiv P(x_p^n \ < x_p^{n,comp}|\mathbf{x}^n), \tag{6}$$

where $x_p^{n,comp}$ denotes a competitor's price for a bid characterized by $\mathbf{x}^n$. The probability on the right-hand side could be estimated by assuming that the competitor's price distribution is identical to the seller's price distribution for known prior *winning* bids with characteristics "similiar" to $\mathbf{x}^n$ using some appropriate distance metric. One problem with this approach is that it explicitly assumes that the bid will be won with the lowest price. Another problem is that it does not take into account historical losing bids. Finally, it is difficult to assess the accuracy of the resulting win probability because it is not derived as part of a formal prediction methodology. For these reasons, we consider a more formal classification approach in the following section.

### 3.2 Naive Bayes Classification

As suggested in the previous subsection, it is useful to develop and evaluate a classification method for predicting the win/loss label as a means of assessing the validity of the computation of the win probability. Equation (3) immediately suggests this classification problem: given feature vectors $\mathbf{x}^n, n = 1, \ldots, N$, predict the binary win/loss label $w^n$. Any of a number of different classification methods could be used. However, one essential criterion is that the resulting method easily generate the win probability as a function of $x_p^n$, given fixed non-price features $x_1^n, \ldots, x_M^n$. The naive Bayes classifier [8][12][16] is particularly well-suited for this task.

We provide a brief overview of the naive Bayes algorithm. Applied to the specific problem here, Bayes theorem yields

$$P(\text{win}|\mathbf{x}^n) = \frac{P(\mathbf{x}^n|\text{win})P(\text{win})}{P(\mathbf{x}^n)}, \tag{7}$$

where $P(\text{win}|\mathbf{x}^n)$ is the probability of a winning a bid $n$ characterized by feature vector $x_p^n$, $P(\text{win})$ is the prior probability of observing a win, $P(\mathbf{x}^n)$ is the prior probability of observing $\mathbf{x}^n$, and $P(\mathbf{x}^n|\text{win})$ is the conditional probability of observing $\mathbf{x}^n$ given a win outcome. The accurate evaluation of $P(\mathbf{x}^n|\text{win})$ can require a potentially huge number of training examples for even a modest number of discrete feature values. This complexity has led to widespread use of the popular naive Bayes classifier, in which it is assumed that the feature values are conditionally independent given the output or target label. For the specific problem here, the naive Bayes approximation yields

$$\tilde{P}(\text{win}|\mathbf{x}^n) \propto P(\text{win})P(x_p^n|\text{win}) \prod_m P(x_m^n|\text{win}), \tag{8}$$

where we have omitted the denominator in equation (7) because it is independent of the target label. We have added a tilde to $\tilde{P}(\text{win}|\mathbf{x}^n)$ to denote that it is not yet normalized. The target label has only two values, win or loss, and hence we write the loss-analog of equation (8):

$$\tilde{P}(\text{loss}|\mathbf{x}^n) \propto P(\text{loss})P(x_p^n|\text{loss}) \prod_m P(x_m^n|\text{loss}). \tag{9}$$

The output label predicted by the naive Bayes classifer is

$$w_{pred}^n = \arg \max_{(win,loss)} \left[\tilde{P}(\text{win}|\mathbf{x}^n), \tilde{P}(\text{loss}|\mathbf{x}^n)\right]. \tag{10}$$

### 3.3 Calculation of the Win Probability

Equations (8) and (9) can be combined to obtain the normalized win probability required in equation (5):

$$P(\text{win}|x_p^n, x_1^n, \ldots, x_M^n) = \frac{\tilde{P}(\text{win}|\mathbf{x}^n)}{\tilde{P}(\text{win}|\mathbf{x}^n) + \tilde{P}(\text{loss}|\mathbf{x}^n)}, \tag{11}$$

where equation (2) has been used to expand $\mathbf{x}^n$. For the purposes of evaluating equation (5), we need to be able to evaluate $P(\text{win}|x_p^n, x_1^n, \ldots, x_M^n)$ as a function of price $x_p$, while holding the non-price features $x_1^n, \ldots, x_M^n$ constant. Let $\mathbf{x_p}$ contain a sequence of $I$ prices $[x_{p,1}, \ldots, x_{p,I}]$ for which the win probability is to be evaluated; then

$$P(\text{win}|\mathbf{x}_p; x_1^n, \ldots, x_M^n) \equiv [P(\text{win}|x_{p,1}; x_1^n, \ldots, x_M^n), \ldots, P(\text{win}|x_{p,I}; x_1^n, \ldots, x_M^n)]. \tag{12}$$

With this result, equation (5) can be rewritten as

$$\hat{x}_{p,Opt}^n = \arg\max_{x_p} \{P(\text{win}|\mathbf{x}_p; x_1^n, \ldots, x_M^n)\,[x_p - C^n]\}\,. \tag{13}$$

### 3.4 Feature Selection

It is useful to develop a metric for characterizing the information content carried by each feature with respect to predicting the win/loss label. For this purpose, we employ the well-known information gain [12] used in C4.5 [15]. The entropy of the bid examples $B$ relative to the bid outcome is

$$\text{Entropy}(B) = p(\text{win}, B)\log_2[p(\text{win}, B)] + p(\text{loss}, B)\log_2[p(\text{loss}, B)], \tag{14}$$

where, for example, $p(\text{win}, B)$ is the proportion of win outcomes in $B$:

$$p(\text{win}, B) = \frac{\text{freq}(\text{win}, B)}{|B|},$$

and $|B|$ is the total number of bid examples. As above, let $X$ denote any price or non-price feature. Here, we assume that $X$ evaluates to discrete values either because $X$ is a categorical variable, or as a result of a binned discretization of a continuous attribute. The expected entropy after the bid examples have been partitioned according to the possible discrete values of $X$ is

$$\text{Entropy}_X(B) = \sum_{v \in \text{values}(X)} \frac{|B_v|}{|B|}\text{Entropy}(B_v), \tag{15}$$

where $\text{Entropy}(B_v)$ denotes the entropy of the subset of bid examples with feature $X$ carrying label $v$, and $|B_v|$ is the number of such examples. The information gain associated with this feature is the difference of these expressions,

$$\text{InfoGain}_X = \text{Entropy}(B) - \text{Entropy}_X(B), \tag{16}$$

reflecting the reduction of entropy or information by knowing the values of feature $X$.

## 4 Classification Results

In this section, we describe the actual bid data used to generate results, discuss the feature selection process, and then evaluate the accuracy of the classification method used to predict win/loss outcomes. All models were developed in Matlab [10].

### 4.1 Description of the Input Data

The data used in the analysis reported in this section were extracted from a database summarizing the circumstances and outcomes of bid transactions involving computers and associated options such as memory modules and displays. These transactions represent bids made by the seller in response to a Request for Quote (RFQ) from a prospective buyer or customer. Here, buyers are businesses, governmental agencies, or educational institutions; there are no individual consumers represented. The original database contains approximately 50,000 rows, where a database row corresponds to a single transaction. Each transaction summarizes the relevant bid information on a single part, where a part refers to either a computer or an option with a unique identifying number. An RFQ often involves multiple parts reflecting combinations of possibly different computers (*e.g.* desktops and servers) with different options (*e.g.* memory upgrades) possibly included in some subset of these. Each part is marked with a final outcome, either *won*, *loss*, or *pending*. Not all parts included in the same RFQ will necessarily have the same outcome: it is possible to win only a subset of the offered parts. In this particular data set, we do not have access to the bid iterations shown in Figure 1; only the final bid and its outcome are available.

The data were filtered in the following manner. First, transactions with pending status were omitted, and we retained only transactions where we had some reasonable confidence in the win/loss label.[2] In order to obtain some temporal locality, we retained only transactions generated over a 14 week period. For the purpose of building and testing the classification model, we included only transactions involving computers as opposed to options, since the options tend to be much less expensive and there is less motivation to focus on optimizing their offering price. With this filtering, we retained a total of 3744 bid items involving 376 unique desktop and mobile computers, representing bids in response to 1900 RFQs, issued by 941 different buyers.

### 4.2 Feature Selection and Information Gain

A significant challenge in any practical application of data-mining or machine-learning is the extraction of relevant features from the raw input data. In the present case, this involves reducing the data available for each bid part to a set of features as in equation (2). Table 1 shows a subset of features[3] extracted from the data described in Section 4.1. Also shown are the number of unique values for each feature, and the information gain computed using equation (16). For categorical features (*e.g.* Customer Industry Name), the third column represents the number of possible discrete values. Binary features (*e.g.* Incumbency) typically have either *yes* or *no* labels. Continuous features (*e.g.* Bid Price) have been binned using equal-population bins to discretize the feature.

An important issue is how to normalize the unit price, given that absolute unit prices for low-end computers differ significantly from higher-end products. It is reasonable to consider normalizations that reflect the seller's cost ($C$) and the published list price ($LP$). Denoting the unit bid price as $BP$, the top section of Table 1 shows the information gain computed

[2]It should be noted that the final win/loss outcome is entered manually into the database, and is subject to some subjectivity and uncertainty.

[3]Some features are proprietary and have been omitted from this table; these features were not used in analysis reported here.

Table 1: Information Gain for Selected Bid Features

| **Feature Name** | **Feature Description** | **Number of Values** | **Information Gain** |
|---|---|---|---|
| Bid Price [(BP-C)/C] | Cost-based price normalization | 7 (Binned) | 0.0139 |
| Bid Price [(BP-C)/(LP-C)] | Cost- and list-based price normalization | 7 (Binned) | 0.0062 |
| Bid Price [BP/LP] | List-based price normalization | 7 (Binned) | 0.0033 |
| Customer Industry Name | Government, Education, Finance, ... | 9 (Discrete) | 0.0304 |
| Incumbency | Strong current position with this buyer? | 2 (Binary) | 0.0250 |
| Number of Employees | Estimated number in buyer's organization | 5 (Binned) | 0.0102 |
| Profit Margin at List Price | Profit margin if part is sold at list price | 5 (Binned) | 0.0066 |
| Part Quantity | RFQ quantity for each part | 5 (Binned) | 0.0056 |
| Financing Opportunity | Opportunity for seller-based financing in deal? | 2 (Binary) | 0.0055 |
| High-profile Account | Expected high future revenue from this buyer? | 2 (Binary) | 0.0049 |
| Internal Advocate at Buyer | Strong seller advocate within buyer organization? | 2 (Binary) | 0.0034 |
| Part Revenue Opportunity | Potential revenue for this part | 5 (Binned) | 0.0034 |
| RFQ Revenue Opportunity | Potential revenue for total RFQ | 5 (Binned) | 0.0033 |
| Identity of Competitors | Names of primary competitors | 3 (Discrete) | 0.0031 |
| Services Opportunity | Opportunity to sell additional services? | 2 (Binary) | 0.0004 |

using three different normalizations for the offered price:

$$\begin{aligned}
\tilde{x}_{p1} &\equiv \tfrac{BP-C}{C} \quad &\text{(Cost-based)}\\
\tilde{x}_{p2} &\equiv \tfrac{BP-C}{LP-C} \quad &\text{(Cost- and List-based)}\\
\tilde{x}_{p3} &\equiv \tfrac{BP}{LP} \quad &\text{(List-based).}
\end{aligned}$$

Note that the first of these normalizations, where the price is effectively normalized as the fractional profit margin, yields the highest information gain of the three normalizations. For this reason, we retain this normalization for the price feature introduced in equation (2), *i.e.*

$$x_p \equiv \frac{BP - C}{C}. \tag{17}$$

The lower section of Table 1 shows the information gain of various non-price features included in equation (2). Note that the Customer Industry Name and Incumbency carry the most information relevant to predicting the win/loss label.

### 4.3 Model Accuracy

In this section, we evaluate the accuracy of the naive Bayes classification model described in Section 3.2. The immediate objective is to predict the win/loss label, and we employ the conventional approach of training the model against a subset of the bid examples, and then evaluating the accuracy against the remaining test examples. The features shown in Table 1 are used as inputs to the model, with bid price normalized as in equation (17). For each bid in the test set, the win/loss label is computed using equation (10), and the win probability for the bid is obtained via equation (11).

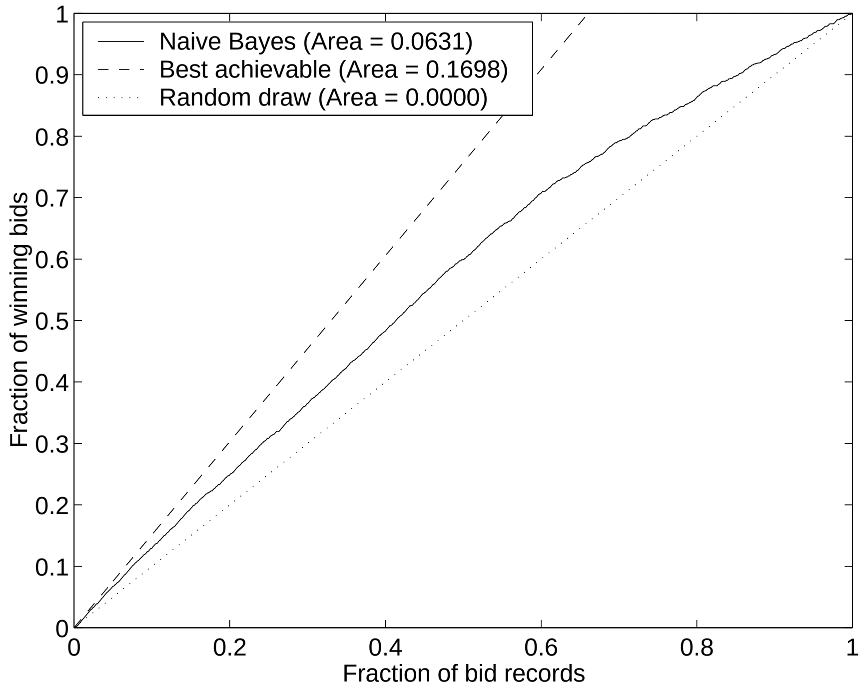


Figure 2: Lift curves computed for test data.

Figure 2 shows the lift curve generated from the win probabilities computed for each bid in the test data. A random 50/50 train/test split was used: the train and test sets consisted of 1868 and 1876 bid records, respectively. The form of the lift curve is conventional [18]: the x-axis shows the fraction of records in the test set, and the y-axis shows the fraction of actual win bids captured as a function of the fraction of test records. If a purely random prediction method were used, such as assuming that some fraction $\alpha$ of the records contain a fraction $\alpha$ of the wins ($\alpha \in [0, 1]$), then the result (for a sufficiently large sample) is simply a straight line as shown in Figure 2. However, if we sort the bids such that the bids with the highest computed win probability are at the top of the list, we expect to observe a fraction $\tilde{\alpha}(> \alpha)$ of the wins in the top fraction $\alpha$ of sorted records. This behavior is indeed observed for the naive Bayes results shown in Figure 2. As a reference, we also plot an upper bound: the best achievable lift curve obtained under the assumption that the win/loss labels of all test bids are known precisely. The naive Bayes lift curve falls roughly midway between the two bounds of a random draw and a perfect classification scheme.

Figure 3 shows the distributions of the computed win probabilities from equation (11), plotted separately for actual wins and losses in the 1876-record test set. Note that the two distributions are quite different, both with respect to their means, and the skewness of the win bids towards the higher computed win probabilities.

A quantitative measure of the obtained lift is given by the area between the lift curve and the straight line generated by a random sample:

$$\text{Area} = \int_0^1 dx[\text{NB}(x) - R(x)], \tag{18}$$

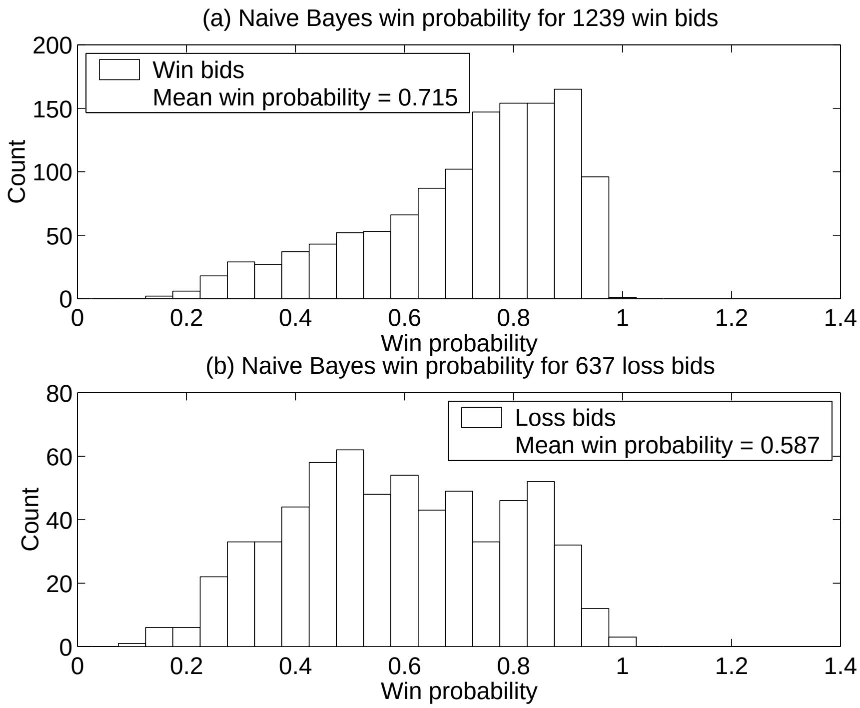


Figure 3: Win probability distributions for test data.

where $x$ denotes the fraction of total records (the x-axis in Figure 2), $\mathrm{NB}(x)$ is the naive Bayes lift curve, and $R(x)$ is the straightline random draw. It is also useful to define the ratio of the area to its optimal value:

$$\mathrm{Ratio} = \frac{\int_0^1 dx[\mathrm{NB}(x) - R(x)]}{\int_0^1 dx[\mathrm{Opt}(x) - R(x)]}, \tag{19}$$

where $\mathrm{Opt}(x)$ is the optimum lift curve assuming perfect knowledge of the test-set output labels.

Results obtained using a single fixed train/split ratio can be potentially misleading, so we also show results obtained using 10-fold cross validation[12]. The total set of 3744 bid items is randomly divided into 10 equal subsets, and 10 different evaluations are run, each using a different single subset as the test set, with the remaining 9 partitions taken as the train set. The results of this analysis are shown in Table 2. Note that the features are sorted in decreasing order by information gain shown in Table 1. We add features to the model one at a time, and monitor the mean, min, and max of the quantities Area and Ratio defined in equations (18) and (19); the statistics are taken over the 10 cross-validation runs. The accuracy, as measured by mean(Ratio), improves monotonically with the addition of each new feature, reaching an asymptote of approximately 38% of the optimum area. Approximately 98% of the asymptotic accuracy is captured by the first seven features (Customer Industry Name $\rightarrow$ Financing Opportunity).

In summary, the results of 10-fold cross validation suggest a statistically significant lift in the naive Bayes model in predicting the win/loss label of a new bid, given the features shown

Table 2: Results of 10-Fold Cross-Validation

| Feature Number | Feature Name | mean (Area) | min (Area) | max (Area) | mean (Ratio) | min (Ratio) | max (Ratio) |
|---|---|---|---|---|---|---|---|
| 1 | Customer Industry Name | 0.0402 | 0.0173 | 0.0604 | 0.2447 | 0.0986 | 0.3616 |
| 2 | Incumbency | 0.0517 | 0.0330 | 0.0717 | 0.3125 | 0.2271 | 0.3975 |
| 3 | Bid Price [(BP-C)/C] | 0.0549 | 0.0350 | 0.0747 | 0.3313 | 0.2412 | 0.4140 |
| 4 | Number of Employees | 0.0584 | 0.0372 | 0.0803 | 0.3522 | 0.2558 | 0.4448 |
| 5 | Profit Margin at List Price | 0.0589 | 0.0385 | 0.0774 | 0.3557 | 0.2649 | 0.4629 |
| 6 | Part Quantity | 0.0607 | 0.0428 | 0.0794 | 0.3664 | 0.2946 | 0.4754 |
| 7 | Financing Opportunity | 0.0620 | 0.0449 | 0.0797 | 0.3751 | 0.3086 | 0.4767 |
| 8 | High-profile Account | 0.0625 | 0.0471 | 0.0811 | 0.3775 | 0.3081 | 0.4856 |
| 9 | Internal Advocate at Buyer | 0.0627 | 0.0475 | 0.0810 | 0.3792 | 0.2899 | 0.4850 |
| 10 | Part Revenue Opportunity | 0.0619 | 0.0442 | 0.0805 | 0.3741 | 0.2800 | 0.4818 |
| 11 | RFQ Revenue Opportunity | 0.0622 | 0.0424 | 0.0807 | 0.3756 | 0.2688 | 0.4830 |
| 12 | Identity of Competitors | 0.0632 | 0.0436 | 0.0810 | 0.3818 | 0.2765 | 0.4754 |
| 13 | Services Opportunity | 0.0630 | 0.0436 | 0.0809 | 0.3809 | 0.2765 | 0.4751 |

in Table 2.

# 5 Computation of Optimal Prices

In this section, we illustrate the computation of the win probability, develop an approach to improve this calculation, and then provide a comparison of computed optimal prices for historical bids with the prices generated by human pricers.

## 5.1 Win-Probability Calculation

Figure 4 shows the win probability as a function of the bid price computed for a single bid item. This curve was computed from equation (12), using all 3744 bid examples[4] described in Section 4.1. As shown in Table 1, the continuous bid price is binned into 7 equal-population bins. The win probability is therefore evaluated at the 7 discrete price values; the widths of these segments are different due to the use of equal-population bins in the normalized price variable. The upward move at segment 5 appears to be a statistical flucuation; it does not appear if the number of bins is reduced from 7 to 5.

A surprising and counter-intuitive characteristic of the win probability is that it *increases* as a function of bid price. One reason for this behavior may be due to implicit strategies employed by the pricers who produced the bid prices in our historical data. It is likely that there are certain competitive situations where a buyer may be inclined to pay a somewhat higher price to the seller because of certain intangibles such as long-term loyalty, superior service, and so on. It is possible that pricing experts are capable of recognizing such situations, and will deliberately price higher to maximize profit. On the other hand, there are converse situations where the seller realistically has little chance of winning the bid, but the pricer is willing to

[4]We do not retain the train/test split introduced in Section 4.3 because we do not seek to evaluate accuracy here, and we prefer to use all available data to improve the statistics of the conditional probabilities.

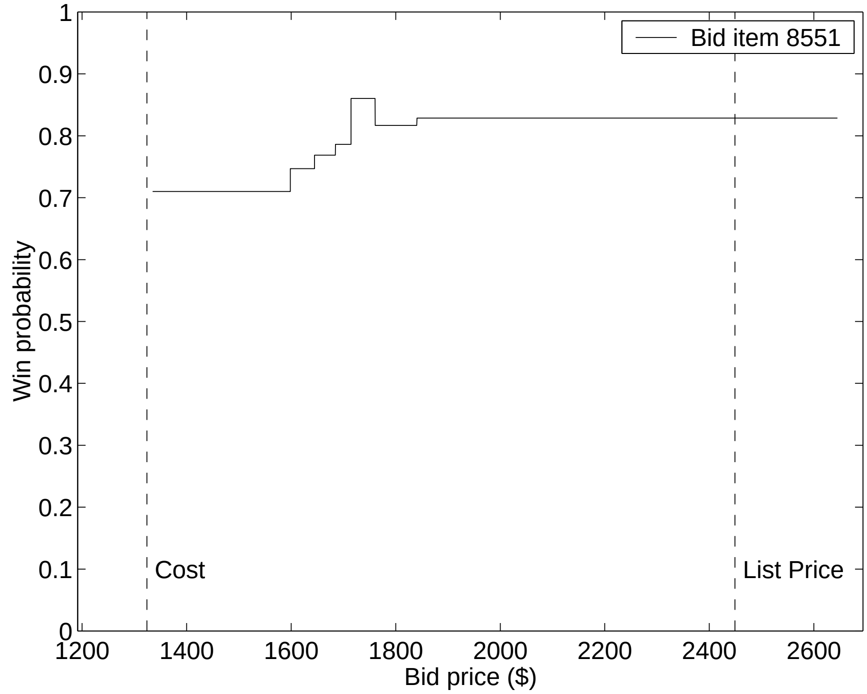


Figure 4: Win probability as a function of bid price.

gamble with a very low price offer. These two scenarios, if they occur with some frequency, will force the observed behavior: the seller will win preferentially at higher prices because pricers exploit these opportunities, and the seller will lose preferentially at lower prices because the pricer is willing to price aggressively to avoid an almost certain loss, that, indeed, is ultimately realized.

A different explanation for the increased win probability as a function of bid price has to do with the way in which a single bid item is presented to the naive Bayes model. Each record explicitly presents an outcome (win or loss) at a single price, namely the price quoted by the pricer. However, this event implicitly represents more information than a simple win/loss at the stated price. Indeed, it is completely reasonable to assume that a win at price $x_p^n$ for bid item $n$ implies a win at all lower prices for this item, and a loss at price $x_p^n$ implies a loss at all higher prices[14].

This observation suggests adding "mirrored" bid items to the existing data that completely replicate all features of an existing bid, with the exception that the bid price is decreased for bids with win labels, and increased for bids with loss labels. Let $x_{p,i}^n$ denote a bid price for item $n$ falling in price bin $i, i = 1, \ldots, I$, where $I$ is the number of bins used to discretize bid price. ($I \equiv 7$ here.) With reference to equation (3), we define a winning bid as

$$B^n(\mathrm{w}in) \equiv \ \{x_{p,i}^n, \mathbf{x}_M^n, \mathrm{w}in\}, \tag{20}$$

where $\mathbf{x}_M^n$ denotes the $M$ non-price features. Consistent with the above discussion, we replicate this bid at all *lower* price bins, retaining the identical non-price features $\mathbf{x}_M^n$ and win label. An analogous procedure is followed for lost bids, yielding the following algorithm for the generation

Table 3: Results of 10-Fold Cross-Validation With Added Mirrored Bid Records

| Feature Number | Feature Name | mean (Area) | min (Area) | max (Area) | mean (Ratio) | min (Ratio) | max (Ratio) |
|---|---|---|---|---|---|---|---|
| 1 | Customer Industry Name | 0.0434 | 0.0282 | 0.0560 | 0.2560 | 0.1595 | 0.3225 |
| 2 | Incumbency | 0.0595 | 0.0489 | 0.0716 | 0.3508 | 0.2948 | 0.4142 |
| 3 | Bid Price [(BP-C)/C] | 0.1043 | 0.0953 | 0.1124 | 0.6157 | 0.5901 | 0.6505 |
| 4 | Number of Employees | 0.1049 | 0.0952 | 0.1130 | 0.6194 | 0.5975 | 0.6542 |
| 5 | Profit Margin at List Price | 0.1106 | 0.0996 | 0.1184 | 0.6526 | 0.6271 | 0.6817 |
| 6 | Part Quantity | 0.1144 | 0.1034 | 0.1226 | 0.6752 | 0.6509 | 0.7056 |
| 7 | Financing Opportunity | 0.1146 | 0.1034 | 0.1232 | 0.6763 | 0.6511 | 0.7093 |
| 8 | High-profile Account | 0.1143 | 0.1020 | 0.1229 | 0.6744 | 0.6423 | 0.7074 |
| 9 | Internal Advocate at Buyer | 0.1135 | 0.1005 | 0.1228 | 0.6697 | 0.6325 | 0.7069 |
| 10 | Part Revenue Opportunity | 0.1133 | 0.1006 | 0.1228 | 0.6686 | 0.6334 | 0.7070 |
| 11 | RFQ Revenue Opportunity | 0.1145 | 0.1015 | 0.1228 | 0.6758 | 0.6388 | 0.7067 |
| 12 | Identity of Competitors | 0.1151 | 0.1021 | 0.1237 | 0.6791 | 0.6426 | 0.7118 |
| 13 | Services Opportunity | 0.1150 | 0.1021 | 0.1236 | 0.6788 | 0.6423 | 0.7116 |

of mirrored bids $\tilde{n}$ from bid $n$:

$$\begin{aligned} \{x_{p,i}^n, \mathbf{x}_M^n, \mathrm{win}\} &\rightarrow \{x_{p,i-1}^{\tilde{n}}, \mathbf{x}_M^{\tilde{n}}, \mathrm{win}\} \ldots \{x_{p,1}^{\tilde{n}}, \mathbf{x}_M^{\tilde{n}}, \mathrm{win}\}, \\ \{x_{p,i}^n, \mathbf{x}_M^n, \mathrm{loss}\} &\rightarrow \{x_{p,i+1}^{\tilde{n}}, \mathbf{x}_M^{\tilde{n}}, loss\} \ldots \{x_{p,I}^{\tilde{n}}, \mathbf{x}_M^{\tilde{n}}, loss\}, \\ \mathbf{x}_M^{\tilde{n}} &\equiv \mathbf{x}_M^n. \end{aligned} \tag{21}$$

A desirable characteristic of this algorithm is that it adds win and loss bids in approximately the same ratio as the original win/loss ratio, and therefore approximately retains the prior probability of observing a win label.

Equation (21) generated 12086 mirrored bid items, and these records were added to the original 3744 examples. The naive Bayes model was regenerated against this aggregated data set of 15830 records, and the 10-fold cross-validation described in Section 4.3 was repeated over the aggregated set. These results are shown in Table 3. In comparison with the initial results in Table 2, these results show significantly improved prediction accuracy: the asymptotic value of the mean(Ratio) increases from 38% in Table 2 to 68% in Table 3. This improvement is due primarily to the enhanced information carried in the price feature $(BP - C)/BP$ as a result of the addition of the mirrored bids.

Figure 5 shows the win probability curve computed with the addition of the mirrored bids. This function shows the expected monotonic decrease with respect to bid price, reflecting the additional information incorporated in the analysis via the addition of the mirrored bid items.

## 5.2 Optimal-Price Calculation

Given the win probability curve as a function of bid price for a specific bid item, the price that optimizes profit for the transaction is readily computed from equation (13). Figure 6 shows a sample calculation of the expected profit [from equation (4)] as a function of bid price. The expected profit is computed over the same bins as used to discretize the bid price, and the

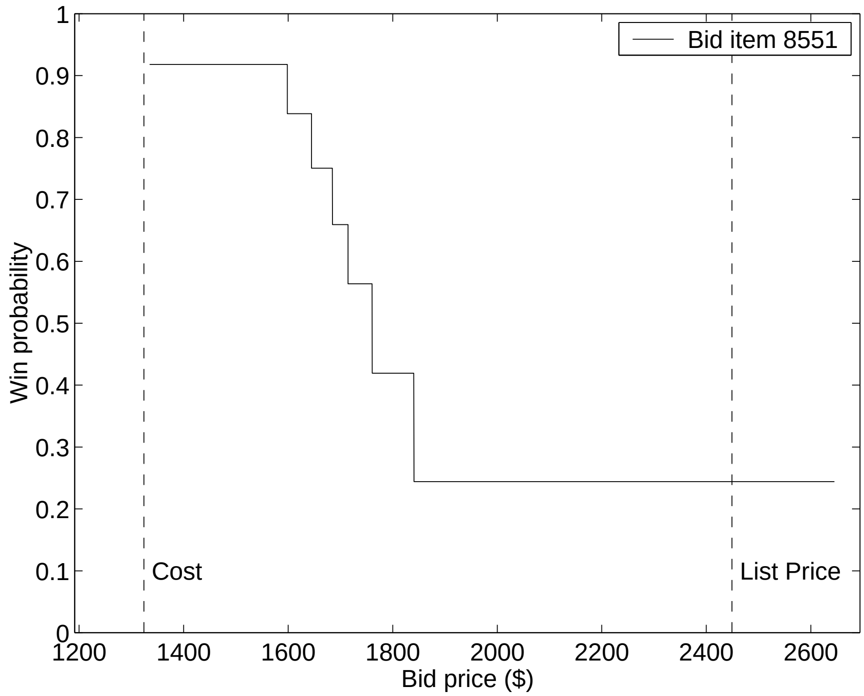


Figure 5: Win probability with added mirrored bid records.

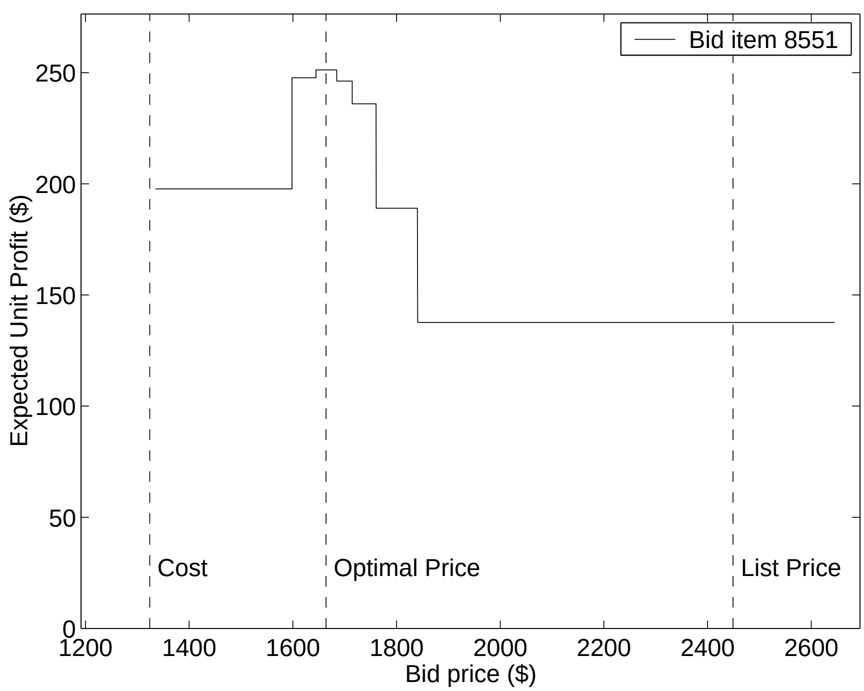


Figure 6: Expected profit as a function of approved price.

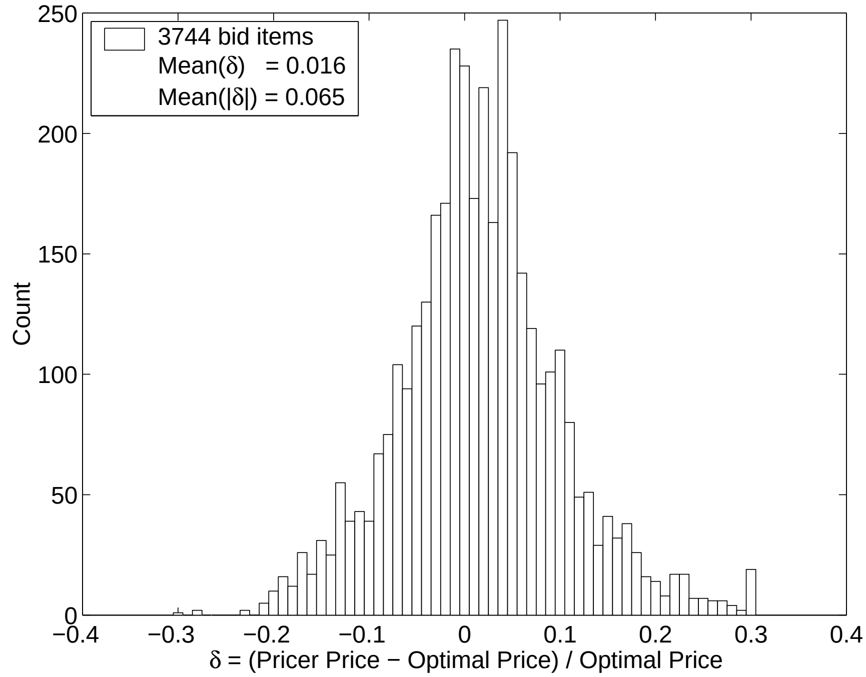


Figure 7: Comparison of computed optimal prices with pricer-generated prices.

optimal price is taken as the mean bid price within the bin that yields the maximum expected profit.

Figure 7 addresses the interesting question of how the computed optimal prices compare with pricer-generated prices. The histogram shown here was generated for the actual data set of 3744 bid items. Interestingly, the optimal prices show little bias relative to the pricer-generated prices: the mean of the difference is only 1.6%, which indicates that the optimal prices are slightly lower (*i.e.* more aggressive) than the human-generated prices. The mean of the absolute difference is 6.5%. Note, however, that our objective here is to compute optimal prices, not to predict human-generated prices, so we do not necessarily expect close agreement between these two sets of results. Moreover, while the performance of a predictive algorithm is straightforward to analyse by assessing the accuracy against test or holdout data, it is much more difficult to quantify the accuracy of the optimal-price computation considered in this paper: we do not have a practical means of rebidding the original RFQs with the optimal prices, and comparing the profit so obtained with the profit generated by the prices quoted by pricing experts.

## 6 Summary and Future Work

We have presented a machine-learning approach to optimizing the bid price in response to a Request for Quote, taking into account seller costs and the probability of winning as a function of various bid features, including offered price. Rather than relying on statistical models with estimated parameters or expert classification of potential bid scenarios, we attempt to learn directly the probability of winning from a set of bid transactions with known outcomes.

A naive Bayes classifier is developed to predict the win/loss outcome of new bid. Results of applying this model to actual bid data demonstrate a statistically significant lift for test data. Independent of the primary objective of optimizing bid pricing, these results suggest a useful capability to prioritize bids (and available supply, if supply is constrained) based on the likelihood of winning a bid.

Win probabilities are computed from the naive Bayes model, with additional synthetic bids systematically added to improve the calculation. Optimal prices generated using this methodology are shown to be slightly more aggressive than human-generated bid prices.

An important issue in the development of the current pricing model is the sparsity of data concerning the competitive circumstances at the time of the bid. It is likely that the robustness of the model could be improved by incorporating additional features characterizing the overall state of market demand at the time of each historical bid. Given more extensive historical data, additional metrics could also be developed, such as enhanced customer price sensitivities and better quantification of price-elasticity curves.

Finally, a related machine-learning problem is to develop a method to predict human-approved bid prices, based on analysis of prices generated by expert pricers. Such a capability would be useful both for training new pricers, as well as for providing an independent check on the results of experienced pricers.

# 7 Acknowledgments

Ed Pednault suggested a significant improvement to the computation of the win probability function. Se June Hong provided useful suggestions on the presentation. We acknowledge useful discussions with Harriet Cao, Roger Gung, Yunhee Jang, and Yingdong Lu on related methodologies. Mark Heid, Bill Onesti, and Lori Payton-Johnson provided key insights into the bid pricing problem.

# References

[1] Y. Bakos. Reducing buyer search costs: implications for electronic marketplaces. *Marketing Science*, 43(12), 1997.

[2] M. Bichler, *et. al.*. Applications of flexible pricing to business-to-business electronic commerce. *IBM Systems Journal*, 41(2), 2002, to appear.

[3] P. Bussey, N. Cassaigne, and M. Singh. Bid pricing - calculating the possibility of winning. *Proc IEEE International Conference on Systems, Man, and Cybernetics*, 5, 4752–4577, 1998.

[4] N. Cassaigne and V. Papaioannou. Knowledge focused bid price setting process. *Proc 11th International Workshop on Database and Expert Systems Applications*, 841–845, 2000.

[5] eBay auction site. http://www.ebay.com.

[6] V.S. Iyengar, J. Lee, and M. Campbell. Q-Eval: evaluating multiple attribute items using queries. *Proceeding of the 3rd ACM Conference on Electronic Commerce EC'01*, 144–153, 2001.

[7] P.K. Kannan and P.K. Kopalle. Dynamic pricing on the Internet: importance and implications for consumer behavior. *Intl Journal of Electronic Commerce*, 5(3):63–83, 2001.

[8] P. Langley, W. Iba, and K.Thompson. An analysis of Bayesian classifiers. *Proceedings of the Tenth National Conference on Artificial Intelligence*, AAAI Press, San Jose, CA, 399–406, 2001.

[9] M. Marshall. Dynamic pricing for e-commerce, an integrated solution. *Proceedings Third International Workshop on Advanced Issues of E-Commerce and Web Based Information Systems*, WECWIS 2001, Third International Workshop on., 192–194, 2001.

[10] *Matlab: the language of technical computing.* http://www.mathworks.com.

[11] J. McGill and G.J. Van Ryzin. Revenue management: research overview and prospects. *Transportation Science*, 33(2):233–256, 1999.

[12] T. Mitchell. *Machine learning.* McGraw-Hill, Boston, Massachusetts, 1997.

[13] V. Papaioannou and N. Cassaigne. A critical analysis of bid pricing models and support tool. *2000 IEEE International Conference on Systems, Man, and Cybernetics*, 3, 2098–2103, 2000.

[14] E. Pednault. IBM Research. Private communication, 2002.

[15] J.R. Quinlan. *C4.5: Programs for machine learning.* Morgan Kaufmann Publishers, San Mateo, California, 1993.

[16] I. Rish, J, Hellerstein, and J. Thathachar. An analysis of data characteristics that affect naive Bayes performance. IBM Technical Report RC21993. IBM Research, Yorktown Heights, NY, 2001.

[17] H.R. Varian. Differential pricing and efficiency. *First Monday (http://www.firstmonday.org)*, 1(2), August, 1996.

[18] I. Witten and E. Frank. *Data mining: Practical machine learning tools and techniques with Java implementations.* Morgan Kaufmann Publishers, San Francisco, California, 2000.