Which Machine-Learning Models Best Predict Online Auction Seller Deception Risk?

By David Lary, Alexey N. Nikitkov and Dan N. Stone

Dr. David Lary
National Aeronautics and Space Administration (NASA) Goddard Space Flight Center
UMBC/JCET NASA/GSFC
Greenbelt, MD 20771, USA
David.J.Lary@nasa.gov

Dr. Alexey Nikitkov
Brock University
Taro 231, Faculty of Business
St. Catharines, ON L2S 3A1, Canada
Phone: (905) 688-5550 ext. 3272
Fax: (905) 688 9779
anikitko@brocku.ca

Contact Author: Dr. Dan N. Stone
University of Kentucky
Von Allmen School of Accountancy
355F Gatton Business and Economics Building
Lexington, KY 40506, USA
Phone: 859-257-3043
Fax: 859-257-3654
dstone@uky.edu

February 14, 2010

Alexey Nikitkov thanks Brock University for the grant supporting this study. Dan Stone thanks the University of Kentucky, the Gatton College of Business and Economics and the Von Allmen School of Accountancy for grants supporting his research. Thanks also to Kevin Park (University of Kentucky) for assistance with the data and to Red Medley (University of Kentucky) for assistance with manuscript preparation and editing.

The authors contributed equally to the project. The archival data are available from public sources.

Keywords: seller deception risk, electronic auctions, electronic markets, machine-learning models.

Which Machine-Learning Models Best Predict Online Auction Seller Deception Risk?

Abstract

How useful are publicly available data, paired with machine-learning models, for assessing seller deception risk in online auctions? The authors use eBay transaction data (n = 1,600) from a quantitative, case-control research design to: (a) identify predictors of seller deception risk and (b) compare the predictive accuracy of five machine-learning models: a naive Bayes classifier, decision trees (DTs), neural networks (NNs), neuro-fuzzy inference (NFI), and support vector machine (SVM) classification, as well as consensus model that aggregates model forecasts. The sample consists of three sub-samples: (a) likely deceptive, i.e., "case" listings (n = 309), (b) a matched control sample (n= 908) of transactions from the same period and listing categories as the deceptive listings and (c) a random control sample from the same market and period (n = 383). The decision tree model has the best overall prediction accuracy (i.e., 76.4%) for a 40% hold-out sample and minimizes error cost when Type I (Alpha), i.e., falsely claiming deception, errors are more costly than Type II (Beta), i.e., falsely claiming no deception, errors. The SVM model minimizes costs in the hold-out sample predictions when Type II errors are more costly than Type I errors. Across differing training and sampling methods, the Decision Tree and Consensus models evidence high prediction accuracy and across-method stability. The results provide among the first large-sample tests of the validity of a diverse set of machine-learning algorithms for predicting online auction seller deception.

Keywords: seller deception risk, accounting controls, electronic auctions, electronic markets, machine-learning models

1. Predicting Online Deception

Internet auction fraud consistently emerges as among the most frequent and costly of online crime (Internet Crime Complaint Center 2008). For example, in 2003, the U.S. Federal Trade Commission reported $100 million in losses due to online auctions (Wingfield 2004). Online fraud imposes multiple social and economic costs. These costs include direct buyer and seller losses to fraudsters, lost sales due to fraud risk, price discounts offered by legitimate online sellers to compensate for fraud risk, and increasingly, the cost of regulating and intervening in online markets.

The high cost of online fraud suggests the potential value of reliably predicting online auction seller deception using publicly available information. Machine-learning models offer the possibility of highly accurate, automated predictions of online auction market deception. This study uses publicly available eBay data cues to test the predictive success of five publicly available, non-proprietary machine-model algorithms. Design and sampling was based on a case-control (also called a case referent) method (Shadish et al. 2002). In a case-control design, units, usually patients in medical research, are selected for the "case" condition, based upon their having a rare, important outcome -- for example, cancer. The case sample is contrasted with one or more "control" samples that are not expected to have this outcome. We apply a case-control research design method to draw three samples. The “case” sample consists of suspected deceptive eBay listings. This sample is then compared with two randomly selected control samples from the same period: (1) matched to the case listing product categories and (2) all eBay listings.

Although more commonly applied in medicine and epidemiology (Armenian and Gordis 1994; Armenian and Lilienfeld 1994), a case-control design affords unique advantages in identifying relevant cues for predicting rare (i.e., outlier), negative outcomes, such as cancer in medicine and deceptive listings in online auctions. Case-control designs allow for selection of negative outcome samples that are both ethical, and, ecologically and externally valid. Case-control designs avoid the potential artificiality of laboratory research and the related necessity of participant deception (Shadish et al. 2002). In addition, they have particular value where the identification of negative outcomes is rare or difficult, as in deception detection. Finally, because they are real-world data, well-executed case-control designs have high ecological and external validity.

eBay, the world's largest and most diverse online auction market, includes approximately 84 million active accounts in 39 countries (Cohen 2002; eBay 2009). Seller fraud and deception are persistent problems for eBay. This paper uses publicly available eBay data to create models that predict seller deception risk. While the present manuscript tests machine-learning models with a specific eBay sample, within a specific time period, the ultimate, pragmatic goal of this project is the conception, design and implementation of a generalized decision aid system, built on machine-learning algorithms, for the prediction of seller deception risk in any online, i.e., web-based market (see Pandit, Chau et al. (2007) for one nascent system with this goal). The present project is one step towards this goal.

1.1 Deception Research: Off-line and Online

Research provides insight into both off-line (or "meat-space"), and online, deception. Detecting and predicting deception is difficult, even with "rich", e.g., verbal and body language, cues. For example, experimental evidence suggests that judgments about whether others are lying are only slightly better than chance (Bond Jr. and DePaulo 2006). Comparisons of truthful and lying verbal content analyses (DePaulo et al. 1997; DePaulo et al. 2003) suggest that liars provide less information, tell less compelling tales, make a more negative impression and are tenser. However, many behaviors that might be expected to predict deception show no, or only a weak, relation to deceit. Research based in interpersonal deception theory (IDT) and channel expansion (CMT) and media richness theory (MRT) has also investigated the role of media "richness" of communications among trading partners in online transactions (Grazioli and Jarvenpaa 2003). Evidence suggests that the difficulty of detecting deception increases with the "leanness" of media. Hence, the difficult problem of deception detection is made more difficult in lean environments, such as in online auctions.

Existing research also provides important insights into online auction deception. One useful body of relevant investigation consists of journalistic, qualitative case accounts of online deception. These accounts include popular press descriptions of fraud and deception (Anonymous 2001a, 2001b; Carlton and Pui-wing 2000; Warner 2003), "how-to" advice to consumers based on archival accounts of one-time online deceptions (Hitchcock and Page 2006; Silver Lake Editors 2006), first-person accounts of one-time deceptions written by deceived buyers (Klink and Klink 2005), and a unique first-person deception description by a convicted fraudulent seller (Walton 2006). Research has also applied social disorganization theory to describe the differing roles of buyers in three online anticrime communities in efforts to control Internet auction fraud (Chua and Wareham 2004, 2007). Further, research has explored individual auction buyers' perceptions of deception and trust in online auction markets (Ba and Pavlou 2002).

1.2 Machine Learning Modeling Research

The importance and pervasiveness of online auction deception, the increased difficulty of deception detection in computer-mediated communication, and the growing efficiency and predictive efficacy of machine learning models, suggests the utility of automating online auction deception detection. A set of key research questions -- that the present paper poses – are as follows:

1. (how) do deceptive eBay listings (i.e., offers of products and services for sale) differ from legitimate eBay listings?
2. can cues derived from publicly available information detect and classify these differences?
3. which machine-learning models evidence the highest, and most stable (across sampling methods), predictive accuracy?[i]

Machine-learning models offer important advantages relative to traditional statistical prediction models (McKee 2009). These advantages include the following considerations: Traditional statistical prediction models are parametric; they specify, a priori, i.e., before examining data, the relation between predictors and outcome. In many cases, however, the form of the relationship is unknown or poorly understood; in addition, parametric models impose potentially limiting assumptions and related simplifications. Machine-learning models offer a nonparametric alternative to parametric modeling; hence, they contain fewer embedded assumptions. In addition, they are capable of learning relations in the data that may not be evident in a priori model specification. Machine-learning models include the capability for identifying and simulating nonlinear relations among variables. Finally, machine-learning models are dynamic; they are particularly suited to prediction in environments with evolving relations between cues and outcomes, such as fraud and deception in the evolving online auction market.

Research, using both archival and experimental data, suggests that machine-learning models may be useful in detecting deception. Table 1 summarizes selected research that tests machine-learning (or related) modeling applications to online auction deception detection using either simulated or archival data sets. Five published studies model some aspect of internet-based fraud detection (See Table 1, Panel A). Ku, Chen et al. (2007) use social network analysis (SNA) to identify internet auction fraud and achieve a "hit rate" (correct cases / total cases) of over 90% for a sample of 100 transactions. Chau, Pandit et al. (2006) (2007) and Zhang, Zhou et al. (2008) use a Markov Random Field (MRF) to detect fraudulent activities, though little validation evidence is provided in these early-stage reports from on-going research. Abbasi and Hsinchun (2009) compare the predictive validity of five machine-learning models for predicting fake-escrow websites; they find that a support vector machine model best predicts fraud outcomes.

Insert Table 1 about here

One commercial product, Auction Inquisitor (Elite Minds Inc 2007) predicts the likelihood of seller deception in individual online auctions. However, the prediction model or models that are embedded in Auction Inquisitor are unspecified; no data or results are provided as to the product's predictive efficacy.[ii] Appendix A summarizes recent machine-learning and related modeling methods to predict credit card (Table A1), and financial statement, telecommunications and insurance fraud (Table A2).

Finally, Zhou, Burgoon et al. (2004b) used machine learning and natural language processing to predict deception in an experimental task. Results indicated that machine learning models offered promise for predicting deception. They found that neural networks exhibited consistent performance across settings. The comparisons also highlighted the importance of variable selection in maximizing classification performance.

To summarize, machine-learning research to date offers some hope for the possibility of robust, assumption-free, nonlinear, data-informed machine-learning predictive models. However to date, most studies focus on testing one or two machine-learning models, while validation has often been with small samples (e.g., n = 10). In contrast, we sought to test a relatively large number of models against a large validation sample. We chose five machine-learning models based on their availability, and, prediction performance in previous research. For example, evidence suggests that the prediction performance of KNN (k-nearest-neighbor) model is generally inferior to that of a SVM model. Hence, we did not include a KNN model in our tests. In addition, we omitted statistical prediction methods that do not include self-learning capabilities (e.g., PCA = principal component analysis; LR = logistic regression; MCLP = multiple criteria linear programming; QD = quadratic discriminant analysis). Some fraud-detection algorithms are proprietary and therefore unavailable, e.g., 2LFS (2-Level Fraud Spotting algorithm) and random forests. Finally, we did not have access to software for testing a MRF = Markov Random Field approach.

We next consider five machine-learning models as potential automated predictors of online auction seller deception risk; more complete descriptions of the methods appear in the cited references.

1.3 Naive Bayes Classifier

The Naive Bayes classification technique applies Bayes' theorem and assumes strong independence of individual features (Domingos and Pazzani 1997; Friedman et al. 1997; Mitchell 1997). That is, it assumes that the presence or absence of a particular feature of a class does not affect the presence or absence of any other feature. In this way, the Naive Bayes classifier can better estimate the means and variances required for accurate classification with less training data than many other classifiers. This makes it particularly effective for datasets containing many predictors or features. Despite a naive design and apparently over-simplified assumptions, Naive Bayes classifiers often accurately predict outcomes in complex real-world applications.

1.4 Decision Tree (DT)

DT (also known as classification tree) methods are a good choice for classification or prediction where one goal is to generate easily understood sets of rules that can be translated into a natural query language. DTs are built through a process known as binary recursive partitioning; this iterative process splits the data into partitions, and then splits it further on each of the DT branches. Classification trees are widely used in applied fields as diverse as medicine (diagnosis), computer science (data structures), botany (classification), and psychology (decision theory). Classification trees lend themselves to graphical displays, making them easier to interpret than numerical displays of outcomes. In creating a DT, we determined a hierarchical set of rules that provided an efficient classification of the dataset. The regression tree algorithm used here is based on Breiman (1984).

1.5 Neural Networks (NNs)

NNs are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. NNs can be trained to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, NNs training results in particular inputs mapping to specific target outputs. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Many such input/target pairs are needed to train a network. Several sources offer good introductions to NNs (Bishop 1995, 1998; Haykin 1999, 2001).

Successful applications of NNs include pattern recognition, identification, classification, speech, vision, and control systems. Financial applications include real estate appraisal, loan advising, mortgage screening, corporate bond rating, credit-line use analysis, credit card activity tracking, portfolio trading program, corporate financial analysis, and currency price prediction (e.g., (Rehkugler and Poddig 1991; Etheridge et al. 2000; Abu-Mostafa et al. 2001; Lam 2004; Mohammadian and Kingham 2005; Boyacioglu et al. 2009)). The NN constructed herein used a pattern recognition network, which was a feed-forward network with tan-sigmoid transfer functions in both the hidden layer and the output layer. There were 20 neurons in one hidden layer.

1.6 Neuro-Fuzzy Inference (NFI)

NFI, or fuzzy inference system (FIS), models are based on the concept of Fuzzy Logic (FL), which was conceived by Zadeh to process data by allowing partial, rather than crisp, set membership or non-membership (Zadeh 1965), hence the term "fuzzy". Zadeh reasoned that people do not require precise, numerical information input, and yet they are capable of highly adaptive control and inference. FL provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. FL is almost synonymous with the theory of fuzzy sets, a theory that relates to classes of objects with vague boundaries in which membership is a matter of degree. FL was conceived as a better method for sorting and handling data but has proven to be an excellent choice for many control system applications since it mimics human control logic but with greater consistency and reliability.

Such models use an imprecise but descriptive language. They are robust, forgiving of operator and data input errors, and often afford highly accurate prediction with little or no tuning.

Fuzzy logic maps an input to an output space; the primary mechanism for this mapping is a list of if-then statements called rules, as opposed to modeling a system mathematically. All rules are evaluated in parallel: rule order is unimportant. The rules themselves are useful because they refer to variables and the adjectives that describe those variables. See (Jang et al. 1997; Mamdani and Assilian 1999; Sugeno 1985) for more complete descriptions of the fuzzy inference system.

1.7 Support Vector Machine (SVM) Classification

SVMs are based on the concept of decision planes that define decision boundaries and were first introduced by Vapnik (1995, 1998, 2000) with subsequent extension by others (Scholkopf et al. 2000; Smola and Scholkopf 2004; Zhou et al. 2006). SVMs are a set of related supervised learning methods used for classification and regression. A decision plane separates objects based on their class memberships. The simplest example is a linear classifier -- which separates objects into groups. However, most classification tasks require more complex structures to achieve accurate separation, i.e., to correctly classify new objects (test cases) on the basis of the known (training case) examples. Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers.

Viewing input data as two sets of vectors in an n-dimensional space, an SVM will construct a separating hyperplane in that space; this hyperline maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the better the generalization error of the classifier. SVM have demonstrated effectively in business applications, for example, in building credit scoring models (Zhou et al. 2006). In this study, we use the SVMs provided by LIBSVM (Chen et al. 2006; Fan et al. 2005).

1.8 Aggregation: Collective Choice Model

In many applied tasks, consensus models outperform individual predictors or models (e.g., in weather forecasting, see Fritsch, Hilliker et al. (2000), for identifying financial statement fraud, see McKee (2009). Accordingly, we compute and report a consensus prediction of the five machine-learning models.

1.9 Error Types and Costs

Two types of error are possible in any prediction task. Surprisingly, the definitions of these error types reverse across literatures and even across authors within literatures.[iii] Following Cleary and Thibodeau (2005), we define the null hypothesis as a prediction that a listing is honest, i.e., not fraudulent. A Type I (or Alpha) error consists of classifying an online market listing as deceptive when it is legitimate. In contrast, a Type II (or Beta) error consists of classifying a listing as honest when it is deceptive. Incurring Alpha versus Beta errors may be differentially costly and these costs may be borne by different stakeholders. For example, Type I errors occur when regulators identify as fraudulent, and take (mistaken) action against, honest sellers. Honest sellers incur Type I error costs to comply with (excessive) regulations, or, when they leave markets due to compliance costs. Buyers incur Type I error costs when they pay more for, and have fewer, product and seller choices due to the removal of honest sellers who are misclassified as deceptive. Regulators commit Type II errors when they fail to identify, and take actions to constrain, dishonest sellers. Type II error costs are borne by defrauded buyers and honest sellers who lose sales due to buyer flight from an emerging "market for lemons" (Akerlof 1970; Lee et al. 2005).

We defined three levels of auction transaction listing outcomes: 3 = a positive (good) outcome, 2 = ambiguous, and, 1 = negative (bad) outcome. A model committed a Type I error when it predicted a more negative outcome than was realized, i.e., where the model prediction = 1 or 2 when the outcome was a 3, or a prediction = 1 when the outcome was a 2. A model committed a Type II error when it predicted a more positive outcome than was realized, i.e., where the model prediction = 2 or 3 when the outcome was a 1, or a prediction = 3 when the outcome was a 2. In assessing model performance, we considered the frequency of Type I (Alpha) and Type II (Beta) errors, and three cost functions related to error occurrence. The absolute cost values, i.e., \$6, \$9, and \$18, assigned to the error occurrences are arbitrary; however, these relative values illustrate specific relations between Type I (Alpha) and Type II (Beta) error costs. Specifically, the three cost functions are:

(1) Type I and Type II errors are equally costly (i.e., \$9 per error),

(2) Type I errors are three times as costly as Type II errors (\$18 vs. \$6 per error), and,

(3) Type II errors are three times as costly as Type I errors (\$18 vs. \$6 per error).

We next describe the research method and data used to compare the predictive value of the methods.

2. Method

2.1 Design

Consistent with a case-control design, we selected a case sample of eBay listings that we expected to have primarily negative transaction outcomes. We contrasted this sample with two control samples of eBay listings. For all listings in the sample, we collected transaction outcomes, which we classified as positive (3), ambiguous (2), or, negative (1). The "case" sample consisted of suspected fraudulent or deceptive eBay listing. “Randomly sampled controls are the exemplar” (Shadish et al. 2002, p. 129) in case-control research designs. We chose two randomly selected control samples. The first control sample was of listings that were matched with, and randomly sampled from, the product categories of “case” sample listings. The second control group was a random sample of all eBay listings that were concurrent with the case sample. Transaction samples were from eBay listings within a 32 day window (from mid-December 2006 to mid-January 2007). Ninety-eight point six percent (98.6%, n = 1,578) of the sampled listings were drawn during the two-week period ending December 31, 2006. Matching a small number of sampled listings (1.4%, n = 22) required extending the sample to mid-January 2007.

2.1.1 Case Sample of Likely Deceptive Listings. Identification of the case sample relied on expert, i.e., one author’s, knowledge of publicly available data. Specifically, an expert in online auction fraud prospectively, i.e., before knowing the outcome, identified 309 eBay listings that he believed to be deceptive. The expert identified suspected deceptive listings by assessments of three categories of potential markers of deception: (a) the seller’s feedback history, (b) changes in account activity, and (c) listing characteristics and acceptable payment methods. Potential deception markers related to the seller’s feedback history included a short or old feedback history, evidence of self- or accomplice-posted feedback (e.g., feedback posted from accounts now listed as “not a registered user anymore” (called NARU) or a private (non-disclosed) feedback history), a history of mostly purchasing or selling low-cost items and a low feedback rating.

Potential markers of deception related to changes in account activity included recent changes from: (a) selling primarily in the non-US to the US market, (b) selling non-brand-name to brand-name merchandise, (c) a few to a large number of items for sale and (d) from selling single items to multiple identical items. Potential deception markers related to listing characteristics and acceptable payment methods included a refusal to accept credit card payment, email communication from the seller indicating that PayPal payments are unacceptable despite the listing stating their acceptability, previous buyer feedback postings to the seller's account alleging that seller has attempted to change payment methods in previous sales, listings that included the posting of a seller email account that was free (e.g., a Gmail or Yahoo account), closed listing after the seller received a buyer's email address through an eBay communication, sellers who indicated that PayPal was an acceptable payment method when the seller was outside of PayPal's registration domain.

2.1.2 Control Sample of Listings Matched, by Product Category, to Case Sample. After training in the characteristics of eBay listings and the desired data, a paid assistant who was blind to the purpose of the study randomly chose up to three listings from the same product category as the suspected deceptive listings. After eliminating listings with incomplete information (n = 19, modal product with missing information = "adult", i.e., pornographic, DVDs), the resulting matched control sample consisted of 908 listings.

2.1.3 Control Sample of Random Listings. A paid assistant who was blind to the purpose of the study randomly chose 455 eBay listings from the same time period as the case sample. After eliminating listings of duplicate products from the same sellers (n=56) and listings with incomplete information (n = 16, e.g., modal product = "adult" DVDs), the resulting sample included 383 listings.

2.2 Listing Outcomes

Listing status was determined in late March, 2007 by a paid assistant who was "blind", i.e., unaware of the purpose of the study. Listings were classified into one of three categories based on the feedback and listing status as of this date:

3 = positive outcome: positive transaction feedback posted by buyer

2 = ambiguous outcome: no or neutral feedback posted by buyer or other undetermined or ambiguous outcome,

1 = negative outcome: negative feedback posted by buyer, transaction deleted by eBay, seller status now listed as Not a Registered User (NARU)

Table 2 summarizes the samples and outcomes. Consistent with expectations, the deceptive (case) sample had mostly negative (77.0%) transaction outcomes. In contrast, ~ 1% of the two control sample listings had negative transaction outcomes. Analysis of variance, and post-hoc comparisons (Bonferroni correction, $p \leq .05$), indicated differences in transaction outcomes between each of the three samples $F(2, 1597) = 658.9$, $p < 0.0001$). The deceptive sample average transaction outcome was 1.25 (sd = 0.485), the matched control sample average outcome was 2.36 (sd = 0.498) and the random sample average outcome was 2.49 (sd = 0.521).

Insert Table 2 about here

2.3 Listing Characteristics and Included Variables

A paid assistant, who was blind to the purpose of the study, collected data on sixty-five attributes of the sampled listings (see Appendix for a data field listing). From these, we identified thirty-eight quantitative listing attributes that were potentially predictive of listing outcomes. We next applied two methods: (a) factor analysis and (b) variable correlation with listing outcomes, to reduce the number of, and aggregate, the number of predictor variables.

Specifically, an exploratory factor analysis identified groupings of variables that potentially explained unique variance in listing outcomes. We chose eleven variables (see Table 3): (a) that correlated with listing outcomes ($|r| \geq .01$) and (b) with no more than two variables from any factor in the factor analysis.[iv] We also tested the sensitivity of the results by applying other methods and procedures for selecting and including model variables. The results of alternative variable selection procedures did not yield models whose predictive performance substantively differed from the models reported herein.

Insert Table 3 about here

2.4 Machine Learning Model Training and Assessment

Evidence suggests that the predictive accuracy of machine learning models differs in their sensitivity to the availability of data for training and testing (De Andrés Suárez et al. 2002); accordingly, we use two different training and assessment approaches to test the sensitivity of the models to data availability and training procedures.

2.4.1 -- 80/20 Data Split – Predictions for All Data. In the first testing process, for four of the five models, i.e., DT, NFI, FIS/NFI, and SVM, we randomly split the data into portions: 80% training and 20% evaluation. For the NN model, we randomly split the available data into three portions: 80%, 10% and 10%. The largest portion containing 80% of the dataset was for initial training. Initial training was iterative; after each iteration, we evaluated the current root mean square (RMS) error of the model output. RMS error was calculated by using the second 10% portion of the data that was not used in the training. In the second phase of training, the RMS error and changes in RMS error with training iteration (epoch), determined the extent of model convergence. When training was complete, the final 10% portion of data was used for independent model validation. This final 10% portion of the data was randomly chosen from the training dataset and is not used in either the training or RMS evaluation. For the first sampling process, the reported results are from predictions generated using the entire data set.

2.4.2 -- 60/40 Data Split – Predictions for Hold-out Sample Only. In the second testing processes, for the same four models as in the first procedure, we divided the sample into 60% training, 40% evaluation (hold-out sample). For the NN model, the sample was split 60% initial testing, 20% iterative testing and 20% final evaluation. For this process, the reported results are for only the evaluation (40% hold-out) sample.

3. Results

3.1 Correlations and Overall Model Results

The eleven selected predictor variables correlate with listing outcomes ($r > .05$, $p < .01$; See Table 4). A US and non-Chinese seller location, higher positive seller feedback and a seller eBay Store correlate with positive listing outcomes; a shorter seller account history, lower buyer feedback, a listing ending on a weekend and a vague product location description correlate with negative listing outcomes.

Insert Table 4 about here

Figures 1 and 2 show the confusion matrices for each method (panels a to e) and the collective inference model (panel f). Figure 1 shows the results for the 80/20 - full sample predictions); Figure 2 shows the results for the 60/40 – hold-out sample predictions.

Insert Figures 1 & 2 about here

Each panel, horizontally, reports the actual class of each data point and vertically the prediction by the machine-learning algorithm. The diagonal squares show the correct classifications while the off diagonal squares show misclassifications. The bottom right cell in each matrix shows the total percent of correctly classified and misclassified (in parentheses) cases; classes 1, 2 and 3 correspond to negative, ambiguous and positive outcomes, respectively.

Any single evaluation metric may inaccurately characterize machine-learning model performance (Provost et al. 1998); accordingly, we use seven metrics to assess model performance.

1. Negative transaction outcome accuracy or hit rate. The percentage of negative (i.e., deceptive or category 1) transactions that are correctly identified (See Table 5),
2. Overall accuracy or hit rate. The overall percentage of accurately categorized listings (See Table 5),
3. Alpha (Type I) error rate. The percentage of legitimate listings that are incorrectly identified as deceptive (See Table 5),
4. Beta (Type II) error rates. The percentage of deceptive listings that are incorrectly identified as legitimate (See Table 5),
5. Cost function 1: Alpha and Beta errors are equally costly ($12 per error; See Table 6),
6. Cost function 2: Alpha errors three times as costly as Beta errors ($18 vs. $6 per error, respectively; See Table 6),
7. Cost function 3: Beta errors three times as costly as Alpha errors ($18 vs. $6 per error, respectively; See Table 6).

Insert Tables 5 & 6 about here

3.2 Naive Bayes Classifier

The Naive Bayes classifier performs well in accurately categorizing class 1 (negative outcome) transactions. However, in both samples, it has the highest Type II error rate. Because of its high Type II error rate, the Naïve Bayes classifier performs poorly in comparisons of error costs in relation to other models. Across both testing methods, the Naïve Bayes classifier is outperformed by other models.

3.3 Decision Trees (DT)

The DT model evidences good stability across performance metrics, about equal Alpha and Beta error rates, and good to excellent performance across the evaluated metrics. In the full-sample analysis, the DT model ranks second to the SVM model in overall accuracy. In addition, the DT model outperforms all models in the hold-out sample results. In short, the DT model emerges as among the best of the evaluated models.

3.4 Neural Networks (NN)

The NN classifier emerges as an average to poor performer relative to the other tested models. Its characteristics include about equal Alpha and Beta error rates and better relative performance in the hold-out in than the full-sample analysis.

3.5 Neuro-Fuzzy Inference (NFI)

The NFI system performs poorly in comparison to the other tested models. It emerges as the poorest predictor in the full sample analysis and the penultimate poorest predictor in the hold-out sample analysis. The proportion of Type II to Type I errors with this model is somewhat unstable across sampling methods: in the full sample this ratio = 1.04 while in the hold-out sample this ratio = 0.71.

3.6 Support Vector Machine (SVM) Classification

SVM classification evidenced the most variable prediction performance across the assessment samples. This model was the best performer with the full sample, with near-perfect overall prediction accuracy (99.6%). With the hold-out sample however, this model tied for fourth place in overall error cost. Hence, SVM performance may depend more on sampling and analysis methods than do the other tested models.

3.7 Consensus Performance

Consistent with other literature which suggests strong predictive performance among consensus models (Archie and Karplus 2009; Mallios 2003; Fritsch et al. 2000), the consensus prediction model was both stable and among the best predictors with either sampling approach. With the full sample method, the consensus model was the third best predictor; with the hold-out sample method, the consensus model was the second best predictor.

3.8 Additional Analyses

We tested the sensitivity of the results to other cue sets and collective aggregation models. For example, we ran the model using different sets of plausible, publicly available cue sets. The results did not substantially differ from those reported. In addition, we tested other combinations of, and methods for selecting, predictor variables. The results of these analyses did not substantively differ from those reported.

4. Summary, Limitations, and Discussion

4.1 Summary

The key research questions posed are:

1. Can cues, derived from publicly available information, reliably discriminate deceptive from honest eBay listing?
2. Which machine-learning models best predict deceptions?

Regarding the first question, the results suggest a set of information cues that reliably discriminate deceptive from honest eBay listings. Some of these cues, for example, a longer seller account duration, and higher seller and buyer feedback ratings, are consistent with intuition and the results of previous research (e.g., (Dawn and Judy 2008; Steiglitz 2007; Gu 2007; MacInnes 2005; Dellarocas 2003; Ba and Pavlou 2002). Others, for example, a listing that ends on a weekend, and an eBay account registered in China or Hong Kong, are less intuitive.

The results also suggest that machine-learning models can reliably predict online deception, though with between-model variability in model stability and prediction accuracy. Three models emerge as candidates for implementing a machine-learning model for online deception prediction: decision tree, SVM, and a consensus model. The decision tree model may be the best single choice since it is the best overall predictor, stable across sampling procedures and more computationally efficient than a consensus model approach. The second best choice may be the SVM model, which performed best in the full sample approach, but which was only fourth-best with the hold-out sample approach. Finally, the consensus model was stable and among the best predictors; however, its use requires computing the results for five model; hence, computation (in)efficiency may make this model pragmatically undesirable.

4.2 Implications for the Creation of a Generalized Automated Decision Aid to Detect Seller Deception Risk

The success of certain machine learning models (e.g., the DT and SVM) in predicting online seller deception in our initial sample would seem to encourage continued exploration of the feasibility of decision aids for predicting online auction seller deception risk. One project plan towards the development of a practical automated system to predict online auction seller deception risk, in the eBay auction market, would include:

1. Creation of a robotic agent to automatically harvest data through the eBay system interface. This process would generate data for testing, analysis and model training.

2. Additional analysis to further validate the factors that best predict the likelihood of seller deception.
3. Training of a machine-learning model or models with a large sample dataset.
4. Further assessment of the model predictive power, and model refinement, with additional samples.
5. Conceptual and physical design of the system to assess the likelihood of fraud for any eBay listing (offer).
6. Implementation of a web-based, prototype system for public testing and system and model refinement.

Figure 3 proposes one possible hardware configuration for implementing this system.

Insert Figure 3 about here

However, the functionality of an automated, widely disseminated system for predicting seller deception in online auctions is potentially limited by the game theoretic, evolutionary implications of such a system (Ba et al. 2000). Consider seller deception in auction markets as a two-person game (i.e., buyers, sellers) where seller strategies evolve partially as a function of the information about seller deception strategies held by buyers. Now introduce an information system that, for example, identifies auction listings that end on weekends as more likely to be deceptive. Given this information system, rational deceptive sellers will shift their deceptive auctions to non-weekend end dates in order to better "mask" them, i.e., hid them among legitimate listings (Boetig 2006; Von Hirsch et al. 2000).

Given this conceptualization, an inherent irony of deceptive prediction systems is decreasing effectiveness with wider dissemination of the knowledge of the predictive cues in the system: the more widely known a deceptive prediction system's cues are among deceivers, the lower the system's effectiveness in predicting seller deception. Deception strategies among rational sellers will evolve to reduce the systems' predictive power in direct proportion to the extent of disclosure of the cues that identify deceptive listings.

4.3 Limitations

The sample and research design give rise to a set of limitations. One limitation is that the sample consists of real-world data that is limited by its location within a specific physical time period and a virtual market. Online auction deception is an evolutionary process (Ba et al. 2000); old deceptions can reappear in new virtual forms. Our case-control sample consists of a set of deceptions that were prevalent at a specific point in time in a specific online market. Therefore, it can be argued that our sample provides strong evidence of the characteristics of deceptive and legitimate listings within the eBay market at a specific time. However, the extent to which the characteristics of our case (i. e., deception) sample generalize beyond the time period, market and period of market formation of their collection is unclear; in fact, theoretical and empirical investigations of the evolution of deception strategies would seem to be an important, though relatively unexplored, research topic that is essential to the creation of sustainable deception detection systems.

A second limitation is that the identification of outcomes in case-control studies can be problematic (Shadish et al. 2002). For example, some buyers in our sample may fear seller retaliation and therefore post positive feedback on listings with negative outcomes. Alternatively, other buyers may be ignorant of the online market or of the products they buy online, and may therefore post negative feedback on listings that more knowledgeable buyers would classify as a positive transaction outcome, or positive feedback on poor transaction outcomes. Such data limitations are inherent in case-control samples and also reflect the ecology of the online market. It may be possible to improve listing outcome detection through the use of automated, natural language processing of, for example, feedback comments. However, such improvements in outcome detection come at the cost: they would complicate the harvesting of relevant cues, and delay the timely delivery of predictive models; but complicating the predictive models would have value in masking, i.e., hiding, the identity and role of the model's predictive cues.

4.4 Conclusion

The results suggest the feasibility of discriminating deceptive from legitimate online listings using only publicly available data. It may be possible to further improve the predictive reliability of machine-learning models by converting the textual data available in the listings, using emerging natural language processing algorithms, into quantitative measures (Zhou et al. 2004a). Evidence suggests that the textual descriptions provided in listings have incremental explanatory value beyond the available quantified variables (Pavlou and Dimoka 2006).

Building buyer trust in online vendors is among the most important impediments to the continued growth of online auctions and commerce. Our analyses suggest that machine-learning models can, in a case-controlled test sample, successfully automate deception detection risk, using only publicly available information. Our results suggest the virtue of proceeding with efforts to generate low-cost, reliable automated decision aids for detecting the likelihood of online auction, seller deception risk.

Table 1
Summary of Machine-Learning and Related Applications to Internet Fraud

Paper	Application	Models	Results
(Ku et al. 2007)	Internet auction fraud	SNA implemented using DT	For n = 100 sample, ~ 90% hit rate for detecting fraud
(Chau et al. 2006)	Internet auction fraud	Proposes 2LFS model	For n = 6 known deceivers, correct classifications
(Pandit et al. 2007)	Internet auction fraud	NetProbe + NetProbe Incremental = MRF and NN	For n = 10 known deceivers, correct classifications
(Zhang et al. 2008)	Model network links among sellers with negative transaction outcomes	Proposes MRF model	None provided
(Abbasi and Hsinchun 2009)	Detect fake escrow websites	SVM, NN, DT, NB, PCA	SVM best predicted frauds
(none – commercial application)	AuctionInquisitor (Elite Minds Inc. 2007)	Not specified	None provided

Key:

Machine-Learning Models:

2LFS = 2-Level Fraud Spotting algorithm
SVM = support vector machine
NN = neural network
DT = decision trees
NB = naïve Bayes
MRF = Markov Random Field

Other Models:

SNA = Social network analysis
PCA = principal component analysis

Table 2

Samples and Outcomes (%s by row; most frequent row & column outcomes shown in **bold**)

			Outcome					
Sample	n		3 = Positive		2 = Ambiguous		1 = Negative	
Deceptive (Case)	309	19.3%	7	2.3%	64	20.7%	**238**	**77.0%**
Matched (Control)	908	56.8%	335	36.9%	**565**	**62.2%**	8	0.9%
Random (Control)	383	23.9%	**190**	**49.6%**	189	49.3%	4	1.0%
Total	1,600	100.0%	532	33.3%	818	51.1%	250	15.6%

Table 3

Descriptions of Predictor Variables in the Machine-Learning Models

Variable Label	Variable Description
1. BdrFeedback	Feedback score of winning bidder
2. FB_Duration	Seller’s feedback tenure: number of days between the first feedback received on seller’s account and the listing date.
3. Seller_location1	Seller's account registered in the US (1) versus not (0)
4. Listing_Location	Listing includes clear description of geographical location (1) versus not (0)
5. Pos180	Number of positive feedback postings in the past 180 days on seller’s account.
6. PosFBPerc	Seller’s % of Positive Feedback since opening their account
7. Seller_Location2	Seller account registered China or Hong Kong (0) versus not (1)
8. Slr_Registration_Days	Number of days since the seller registered their account with eBay
9. SlrFeedback	The seller's overall eBay feedback score per eBay formula
10. Store1	Seller has eBay store (1) or not (0)
11. Weekend	Listing ends on weekend day (1) or not

See Appendix for variable definitions

Table 4

Correlations among Variables

	Slr_ Registration _Days	Seller_ Location_1	Seller_ Location2	Pos 180	PosFB Perc	Slr Feed back	Store1	Week end	FB_ Duration	Listing Location1	Bdr Feed back
Outcome	**-0.153**	**0.448**	**0.574**	**0.054**	**-0.105**	**0.057**	**0.096**	**-0.080**	**-0.156**	**-0.115**	**-0.136**
Slr_Registration_Days		**-0.310**	**-0.293**	**0.132**	**0.187**	**0.215**	-0.048	0.013	**0.998**	**0.054**	**0.071**
Seller_ Location_1			**0.784**	**-0.077**	**-0.103**	**-0.110**	-0.009	**-0.085**	**-0.312**	**-0.050**	**-0.126**
Seller_ Location2				**-0.107**	**-0.095**	**-0.137**	-0.025	**-0.080**	**-0.296**	**-0.139**	**-0.135**
Pos180					-0.008	**0.854**	0.261	0.042	**0.116**	0.058	0.009
PosFBPerc						0.007	**0.073**	0.025	0.225	0.067	**0.037**
SlrFeedback							**0.288**	0.017	**0.201**	**0.100**	0.037
Store1								0.012	**-0.049**	**0.105**	**-0.051**
Weekend									0.014	**-0.026**	0.028
FB_Duration										**0.057**	**0.072**
Listing_Location1											**0.092**

Significant Correlations (p ≤ .05) shown in **Bold**

See Appendix for variable definitions

Table 5

Model Accuracy and Error Rates

Panel A: Full Sample

Model	% Category 1 Cases Accuracy Classified*	% Total Cases Accurately Classified**	Error Type Rates	
			Type I (Alpha)**	Type II (Beta)**
Naïve Bayes	92.1%	64.2%	8.4%	27.4%
Decision Tree	86.8%	91.5%	3.2%	5.2%
NN	73.7%	65.1%	16.6%	18.2%
FIS / NFI	88.6%	78.4%	10.6%	11.0%
SVM	96.5%	99.6%	0.1%	0.3%
Consensus	93.9%	88.0%	2.8%	9.1%

* n = 114

** n = 1,439

Panel B: Hold-Out (40%) Sample

Model	% Category 1 Cases Accuracy Classified*	% Total Cases Accurately Classified**	Error Type Rates	
			Type I (Alpha)**	Type II (Beta)**
Naïve Bayes	95%	63.9%	6.6%	29.5%
Decision Tree	80%	76.4%	12.5%	11.1%
NN	45%	64.2%	16.6%	19.1%
FIS / NFI	70%	57.3%	25.0%	17.7%
SVM	65%	61.5%	36.1%	2.4%
Consensus	80%	69.1%	16.7%	14.1%

* n = 20
** n = 456

Table 6
Comparison of Model Error Costs and Average Rank

Panel A: Full Sample

Model	Equal cost ($12)	Type I 3X Type II ($18, $6)	Type II 3X Type I ($18, $6)	Average Rank
Naïve Bayes	$4.30	$3.16	$5.44	5.7
Decision Tree	$1.01	$0.89	$1.13	2.0
NN	$4.18	$4.08	$4.27	5.3
FIS / NFI	$2.59	$2.57	$2.62	4.0
SVM	$0.05	$0.04	$0.06	1.0
Consensus	$1.43	$1.05	$1.81	3.0

Panel B: Hold-Out (40%)

Model	Equal cost ($12)	Type I 3X Type II ($18, $6)	Type II 3X Type I ($18, $6)	Average Rank
Naïve Bayes	$4.33	$2.96	$5.71	4.0
Decision Tree	$2.83	$2.92	$2.75	1.3
NN	$4.28	$4.13	$4.43	3.7
FIS / NFI	$5.12	$5.56	$4.69	5.3
SVM	$4.62	$6.64	$2.60	4.0
Consensus	$3.70	$3.85	$3.54	2.7

Key:
I 3X II = Type I errors three times as costly as type II errors
II 3X I = Type II errors three times as costly as type I errors

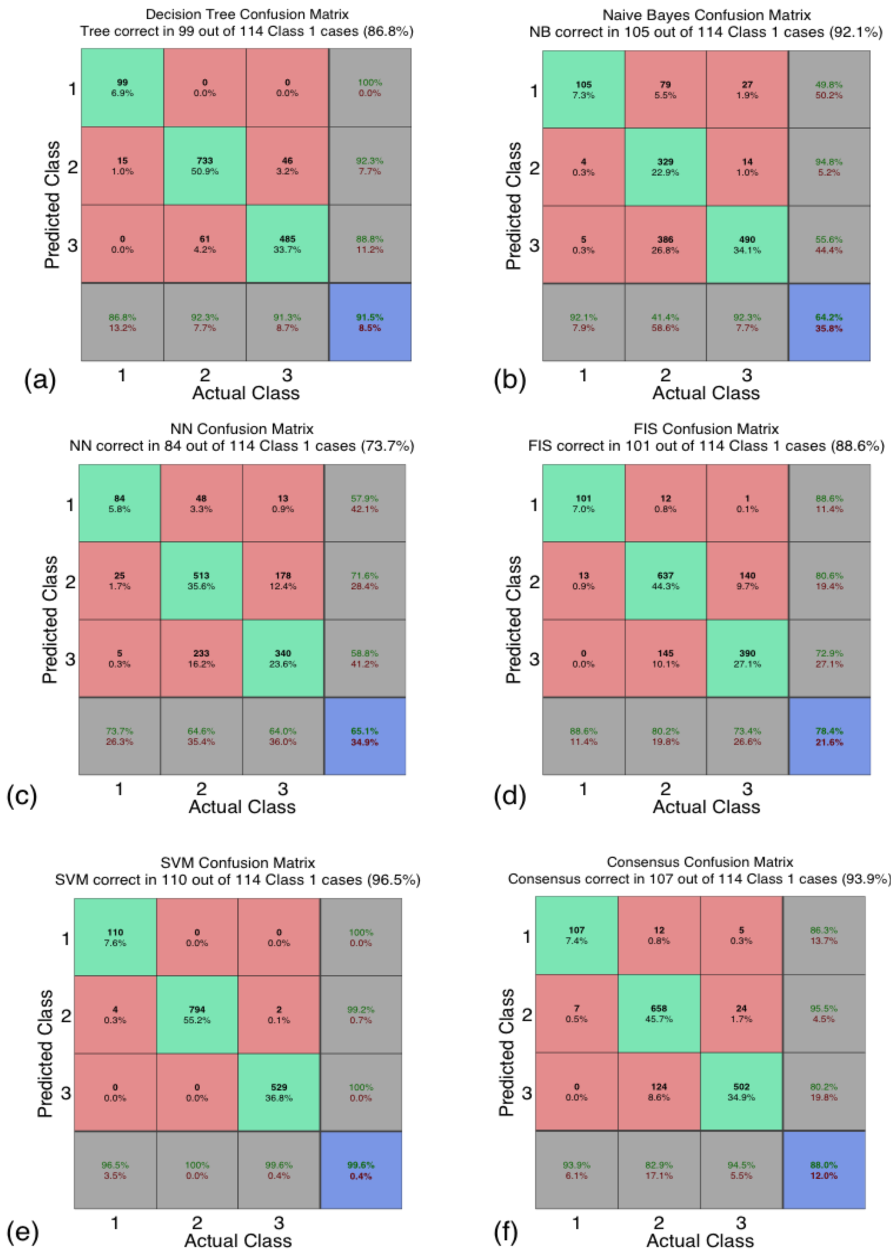


Figure 1. Results using all the data for training. Each panel shows the confusion matrix for five different machine-learning approaches in predicting transaction outcome (panels a to e) and the collective inference among all five methods (panel f). In each panel, horizontally we have the actual class of each data point and vertically the prediction by the machine-learning algorithm. Class 1 corresponds to a negative outcome; class 2 to an ambiguous outcome; class 3 to a positive outcome. The diagonal green squares show the correct classifications and the off diagonal red squares show the misclassifications. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red).

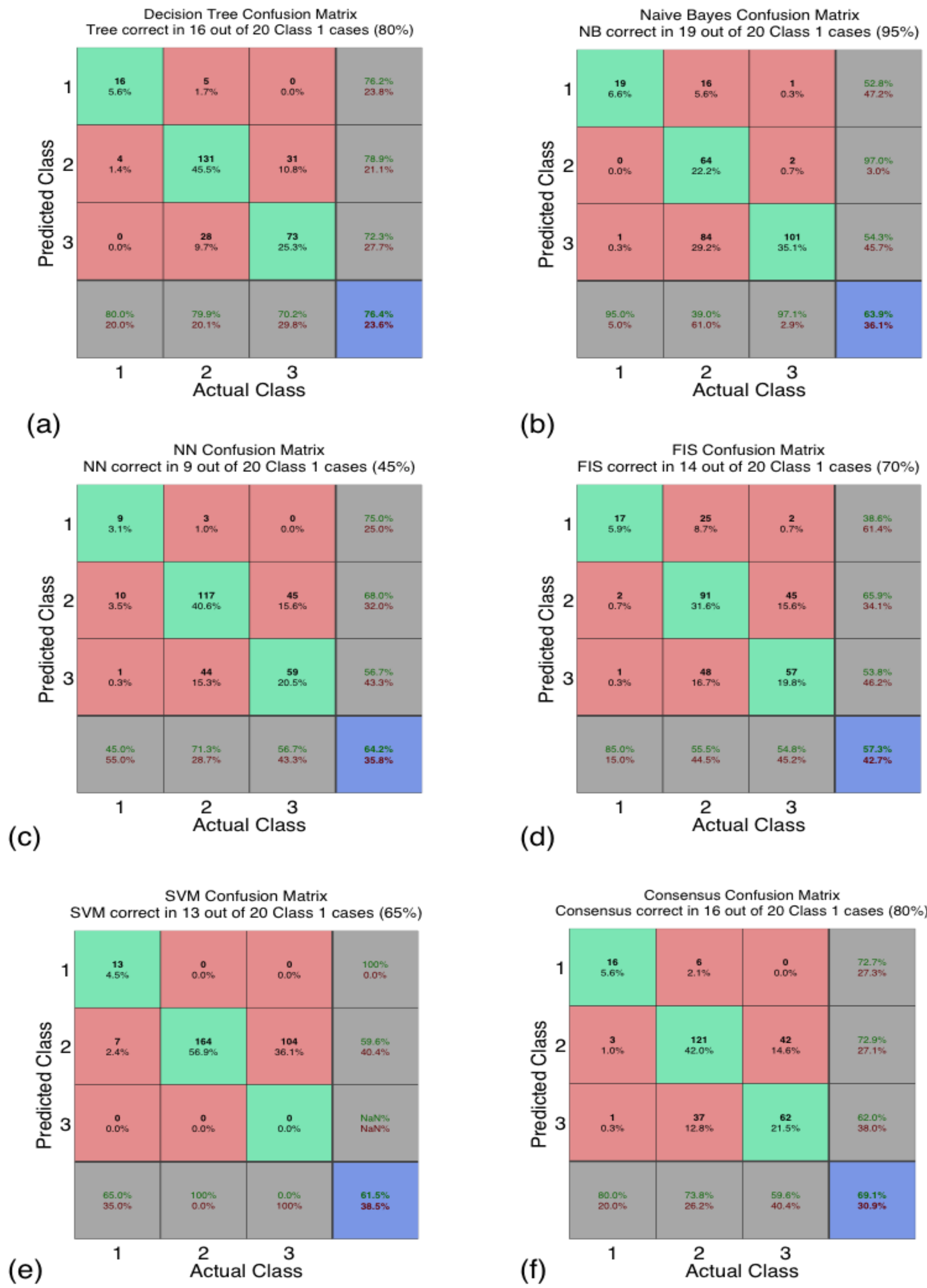
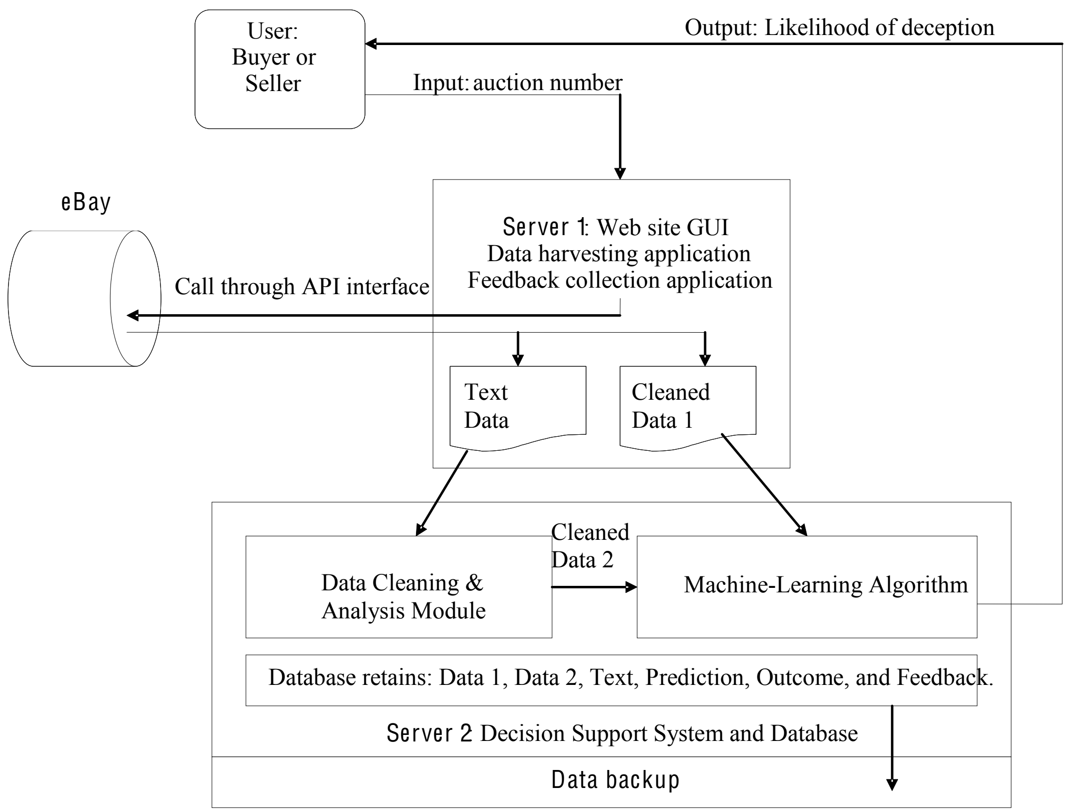


Figure 2. Results using 60% of the data for training and 40% for testing. Each panel shows the confusion matrix for five different machine-learning approaches in predicting transaction outcome (panels a to e) and the collective inference among all five methods (panel f). In each panel, horizontally we have the actual class of each data point and vertically the prediction by the machine-learning algorithm. Class 1 corresponds to a negative outcome; class 2 to an ambiguous outcome; class 3 to a positive outcome. The diagonal green squares show the correct classifications and the off diagonal red squares show the misclassifications. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red).

Figure 3
Proposed Architecture to Automated Detection & Reporting of Online Auction Seller Deception Risk



References

Abbasi, A., and C. Hsinchun. 2009. A comparison of fraud cues and classification methods for fake escrow website detection. *Information Technology & Management* 10 (2/3):19p.

Abu-Mostafa, Y. S., A. F. Atiya, M. Magdon-Ismail, and H. White. 2001. Introduction to the special issue on neural networks in financial engineering. *Ieee Transactions on Neural Networks* 12 (4):653-656.

Akerlof, G. A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84 (3):488-500.

Anonymous. 2001a. Three Are Indicted Over Phony eBay Bids. *Wall Street Journal*, B.6.

———. 2001b. Two Men Plead Guilty To Fraud in Auction Of Painting on eBay. *Wall Street Journal*, B.2.

Archie, J., and K. Karplus. 2009. Applying undertaker cost functions to model quality assessment. *Proteins-Structure Function and Bioinformatics* 75 (3):550-555.

Armenian, H. K., and L. Gordis. 1994. Future perspectives on the case-control method. *Epidemiol Rev* 16 (1):163-164.

Armenian, H. K., and D. E. Lilienfeld. 1994. Applications of the case-control method. Overview and historical perspective. *Epidemiol Rev* 16 (1):1-5.

Ba, S., and P. A. Pavlou. 2002. EVIDENCE OF THE EFFECT OF TRUST BUILDING TECHNOLOGY IN ELECTRONIC MARKETS: PRICE PREMIUMS AND BUYER BEHAVIOR. *MIS Quarterly* 26 (3):243-268.

Ba, S., A. B. Whinston, and H. Zhang. 2000. The Dynamics of the Electronic Market: An Evolutionary Game Approach. *Information Systems Frontiers* 2 (1):31-40.

Bishop, C. M. 1995. *Neural networks for pattern recognition*. Oxford: Oxford University Press.

———. 1998. *Neural networks and machine learning*. Berlin; New York: Springer.

Boetig, B. P. 2006. The Routine Activity Theory. *FBI Law Enforcement Bulletin* 75 (6):12-19.

Bond Jr., C. F., and B. M. DePaulo. 2006. Accuracy of Deception Judgments. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)* 10 (3):214-234.

Boyacioglu, M. A., Y. Kara, and O. K. Baykan. 2009. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications* 36 (2):3355-3366.

Breiman, L. 1984. *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group.

By Ken, B. 2000. EBay Seller Used Aliases, Submitted Phony Bids at Site. *Wall Street Journal*, 1.

Carlton, J., and T. Pui-wing. 2000. Online Auctioneers Face Growing Fraud Problem. *Wall Street Journal*, 1.

Chan, P. K., W. Fan, A. L. Prodromidis, and S. J. Stolfo. 1999. Distributed data mining in credit card fraud detection. *Ieee Intelligent Systems & Their Applications* 14 (6):67-74.

Chau, D. H., S. Pandit, and C. Faloutsos. 2006. Detecting fraudulent personalities in networks of online auctioneers. *Knowledge Discovery in Databases: Pkdd 2006, Proceedings* 4213:103-114.

Chen, P. H., R. E. Fan, and C. J. Lin. 2006. A study on SMO-type decomposition methods for support vector machines. *Ieee Transactions on Neural Networks* 17 (4):893-908.

Chua, C. E. H., and J. Wareham. 2004. Fighting Internet Auction Fraud: An Assessment and Proposal. *Computer* 37 (10):31.

———. 2007. The role of Online Trading Communities in Managing Internet Auction Fraud. *MIS Quarterly* 31 (4):759-781.
Cleary, R., and J. C. Thibodeau. 2005. Applying digital analysis using Benford's Law to detect fraud: The dangers of Type I errors. *Auditing-a Journal of Practice & Theory* 24 (1):77-81.
Cohen, A. 2002. *The perfect store: inside eBay*. 1st ed. Boston: Little, Brown and Co.
Comunale, C. L., and T. R. Sexton. 2005. A Fuzzy Logic Approach to Assessing Materiality. *Journal of Emerging Technologies in Accounting* 2 (1):1-15.
Dawn, G. G., and E. S. Judy. 2008. A typology of complaints about eBay sellers. *Communications of the ACM* 51 (4):69-74.
De Andrés Suárez, D. J., P. Lorca Fernández, and E. F. Combarro. 2002. The Sensitivity of Machine Learning Techniques to Variations in Sample Size: A Comparative Analaysis. *Digital Accounting Research* 2 (4):131-156.
Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49 (10):1407-1424.
DePaulo, B. M., K. Charlton, H. Cooper, J. J. Lindsay, and L. Muhlenbruck. 1997. The Accuracy-Confidence Correlation in the Detection of Deception. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)* 1 (4):346.
DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. Cues to Deception. *Psychological Bulletin* 129 (1):74.
Domingos, P., and M. Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29 (2-3):103-130.
eBay. *eBay Inc.: A Short History* 2009 [cited 1/26/2009. Available from http:/?news.ebay.com/about.cfm.
Elite Minds Inc. 2010. *Auction Inquisitor (TM)* 2007 [cited 2/10 2010]. Available from http://www.auctioninquisitor.com/.
Etheridge, H. L., R. S. Sriram, and H. Y. K. Hsu. 2000. A comparison of selected artificial neural networks that help auditors evaluate client financial viability. *Decision Sciences* 31 (2):531-550.
Fan, R. E., P. H. Chen, and C. J. Lin. 2005. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research* 6:1889-1918.
Fawcett, T., and F. Provost. 1997. Adaptive fraud detection. *Data Mining and Knowledge Discovery* 1 (3):291-316.
Friedman, N., D. Geiger, and M. Goldszmidt. 1997. Bayesian network classifiers. *Machine Learning* 29 (2-3):131-163.
Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky. 2000. Model Consensus. *Weather and Forecasting* 15 (5):571-582.
Grazioli, S., and S. L. Jarvenpaa. 2003. Consumer and Business Deception on the Internet: Content Analysis of Documentary Evidence. *International Journal of Electronic Commerce* 7 (4):93-118.
Gu, J. 2007. Consumer Rights Protection on the Online Auction Website - Situations and Solutions: A Case Study of EBay. In *British and Irish Law, Education, and Technology Association*. Hertfordshire.
Haykin, S. S. 1999. *Neural networks : a comprehensive foundation*. 2nd ed. Upper Saddle River, N.J.: Prentice Hall.

———. 2001. *Kalman filtering and neural networks*. New York: Wiley.
Hilas, C. S., and J. N. Sahalos. 2006. Testing the fraud detection ability of different user profiles by means of FF-NN classifiers.
Hitchcock, J. A., and L. Page. 2006. *Net crimes and misdemeanors : outmaneuvering web spammers, stalkers, and con artists*. 2nd ed. Medford, N.J.: Information Today/CyberAge Books.
Internet Crime Complaint Center 2008. Annual Report. http://www.ic3.gov/media/annualreport/2008_IC3Report.pdf. Access date: 2/14/2010.
Jang, J.-S. R., C.-T. Sun, and E. Mizutani. 1997. *Neuro-fuzzy and soft computing : a computational approach to learning and machine intelligence*. Upper Saddle River, NJ: Prentice Hall.
Kim, M. J., and T. S. Kim. 2002. A neural classifier with fraud density map for effective credit card fraud detection.
Kirkos, E., C. Spathis, and Y. Manolopoulos. 2007a. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* 32 (4):995-1003.
Kirkos, E., C. Spathis, A. Nanopoulos, and Y. Manolopoulos. 2007b. Identifying Qualified Auditors' Opinions: A Data Mining Approach. *Journal of Emerging Technologies in Accounting* 4 (1):183-197.
Klink, E. E., and S. G. Klink. 2005. *Dawn of the ebay deadbeats : true tales of treachery, lies, and fraud from the dark recesses of the world's largest online auction marketplace*. 1st ed. Allendale, NJ: Mooncusser Media LLC.
Koskivaara, E., and B. Back. 2007. Artificial Neural Network Assistant (ANNA) for Continuous Auditing and Monitoring of Financial Data. *Journal of Emerging Technologies in Accounting* 4 (1):29-45.
Kotsiantis, S., E. Koumanakos, D. Tzelepis, and V. Tampakas. 2006a. Financial application of neural networks: Two case studies in Greece. Paper read at ARTIFICIAL NEURAL NETWORKS - ICANN 2006, PT 2.
———. 2006b. Forecasting Fraudulent Financial Statements using Data Mining. *International Journal of Computational Intelligence*
3 (2):104-110.
Ku, Y., Y. Chen, and C. Chiu. 2007. A Proposed Data Mining Approach for Internet Auction Fraud Detection. In *Intelligence and Security Informatics*, 238-243.
Lam, M. 2004. Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. *Decision Support Systems* 37 (4):567-581.
Lee, B.-C., L. Ang, and C. Dubelaar. 2005. Lemons on the Web: A signalling approach to the problem of trust in Internet commerce. 26 (5):607-623.
Lenard, M. J., A. L. Watkins, and P. Alam. 2007. Effective Use of Integrated Decision Making: An Advanced Technology Model for Evaluating Fraud in Service-Based Computer and Technology Firms. *Journal of Emerging Technologies in Accounting* 4 (1):123-137.
MacInnes, I. 2005. Causes of Disputes in Online Auctions. *Electronic Markets* 15 (2):146-157.
Malek, W., K. Mayes, and K. Markantonakis. 2008. Fraud Detection and Prevention in Smart Card Based Environments Using Artificial Intelligence. In *Smart Card Research and Advanced Applications*, 118-132.
Mallios, R. R. 2003. A consensus strategy for combining HLA-DR binding algorithms. *Hum Immunol* 64 (9):852-856.
Mamdani, E. H., and S. Assilian. 1999. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Human-Computer Studies* 51 (2):135-147.

McKee, T. E. 2009. A Meta-Learning Approach to Predicting Financial Statement Fraud. *Journal of Emerging Technologies in Accounting* 6 (1):5-26.
Mitchell, T. M. 1997. *Machine Learning*. New York: McGraw-Hill.
Mohammadian, M., and M. Kingham. 2005. Intelligent data analysis, decision making and modelling adaptive financial systems using hierarchical neural networks. *Knowledge-Based Intelligent Information and Engineering Systems, Pt 2, Proceedings* 3682:775-781.
Pandit, S., D. H. Chau, S. Wang, and C. Faloutsos. 2007. Netprobe: a fast and scalable system for fraud detection in online auction networks. In *Proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Canada: ACM.
Park, L. J. 2005. Learning of neural networks for fraud detection based on a partial area under curve.
Pavlou, P. A., and A. Dimoka. 2006. The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. *Information Systems Research* 17 (4):392-414.
Peng, Y., G. Kou, A. Sabatka, J. Matza, Z. Chen, D. Khazanchi, and Y. Shi. 2007. Application of Classification Methods to Individual Disability Income Insurance Fraud Detection. In *Computational Science – ICCS 2007*, 852-858.
Phua, C., D. Alahakoon, and V. Lee. 2004. Minority Report in Fraud Detection: Classification of Skewed Data. *SIGKDD Explorations* 6 (1):50-59.
Provost, F., T. Fawcett, and R. Kohavi. 1998. The Case Against Accuracy Estimation for comparing Induction Algorithms. Paper read at Proceedings of the Fifteenth International Conference on Machine Learning, at Madison, WI.
Ragothaman, S., and A. Lavin. 2008. Restatements Due to Improper Revenue Recognition: A Neural Networks Perspective. *Journal of Emerging Technologies in Accounting* 5 (1):129-142.
Rehkugler, H., and T. Poddig. 1991. Artificial Neural Networks in Financial Analysis - a New Approach to Forecast Movements of Stocks, Indexes and Currencies. *Wirtschaftsinformatik* 33 (5):365-374.
Scholkopf, B., A. J. Smola, R. C. Williamson, and P. L. Bartlett. 2000. New support vector algorithms. *Neural Computation* 12 (5):1207-1245.
Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
Shawver, T. J. 2005. Merger Premium Predictions Using a Neural Network Approach. *Journal of Emerging Technologies in Accounting* 2 (1):61-72.
Silver Lake Editors. 2006. *Scams and Swindles: Phishing, Spoofing, ID Theft, Nigerian Advance Schemes, Investment Frauds. How to Recognize and Avoid Internet Era Rip-Offs*. Los Angeles: Silver Lake Publishing.
Smola, A. J., and B. Scholkopf. 2004. A tutorial on support vector regression. *Statistics and Computing* 14 (3):199-222.
Steiglitz, K. 2007. *Snipers, shills & sharks : eBay and human behavior*. Princeton: Princeton University Press.
Sugeno, M. 1985. *Industrial applications of fuzzy control*. Amsterdam: North-Holland.
Vapnik, V. N. 1995. *The nature of statistical learning theory*. New York: Springer.
———. 1998. *Statistical learning theory*. New York: Wiley.
———. 2000. *The nature of statistical learning theory*. 2nd ed. New York: Springer.

Vatsa, V., S. Sural, and A. K. Majumdar. 2005. A game-theoretic approach to credit card fraud detection.
Von Hirsch, A., D. Garland, A. Wakefield, and University of Cambridge. Centre for Penal Theory and Penal Ethics. 2000. *Ethical and social perspectives on situational crime prevention*. Oxford: Hart.
Walton, K. 2006. *Fake: How One Man's Con Game Created An International Scandal and Triggered a Nationwide FBI Manhunt*. New York: Simon Spotlight Entertainment.
Warner, M. 2003. eBAY'S WORST NIGHTMARE. *Fortune* 147 (10):89-94.
Whitrow, C., D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams. 2009. Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery* 18 (1):30-55.
Wingfield, N. 2004. Cat and Mouse-Problem for Cops on Ebay Beat. *Wall Street Journal* August 3, A1.
Zadeh, L. A. 1965. FUZZY SETS. *Information and Control* 8 (3):338-&.
Zhang, B., Y. Zhou, and C. Faloutsos. 2008. Toward a Comprehensive Model in Internet Auction Fraud Detection. Paper read at Proceedings of the 41st Hawaii International conference on System Sciences, at Hawaii.
Zhou, L., J. K. Burgoon, J. F. Nunamaker, and D. Twitchell. 2004a. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation* 13 (1):81-106.
Zhou, L., J. K. Burgoon, D. P. Twitchell, D. P. Tiantian Qin, and J. F. Nunamaker Jr. 2004b. A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication. *Journal of Management Information Systems* 20 (4):139-165.
Zhou, L. G., K. K. Lai, and L. A. Yu. 2006. Credit scoring using support vector machines with direct search for parameters selection. Paper read at Soft Computing, Jul 01-08, at Chengdu, PEOPLES R CHINA.

Appendix A

Table A1 - Credit Card Fraud Applications of Machine-Learning Models

Paper	Application	Models	Results
(Chan et al. 1999)	Credit card fraud: data from 2 banks	NB + 3 other methods	Detection improvements & cost savings compared to banks existing detection method
(Kim and Kim 2002)	Credit card fraud	NN with fraud density mapping	More accurate assessments of fraud relative to population proportions
(Park 2005)	Credit card fraud	NN + partial area under curve	Increased sensitivity of NN model to fraud base rates
(Vatsa et al. 2005)	Credit card fraud	Game theoretic model: compare detection & detection avoidance strategies	Mutual learning occurs among human participants
(Malek et al. 2008)	Smart card fraud	Proposes NN architecture	n/a
(Whitrow et al. 2009)	Credit card fraud: data from 2 banks	Compare data aggregation methods: SVM, LR, NB, DT, QD, KNN, random forests	Random forests…"a consistent winner…" (p. 48)

Table A2 - Recent Financial Statement, Telecommunications and Insurance Fraud Applications of Machine-Learning Models

Paper	Application	Models	Results
(Fawcett and Provost 1997)	Cellular phone cloning fraud	Proposed architecture & algorithm (Detector Construction) for detecting	Outperforms previous detection method used by phone company
(Kotsiantis et al. 2006a, 2006b)	Financial statement fraud, bankruptcy	SVM, BPA, NN (with VP), Winnow, RBF NN, LR, Consensus predictor	Consensus model predicts most accurately
(Hilas and Sahalos 2006)	Telecommunications fraud	Feedforward NN	Summary data outperforms detailed data
(Peng et al. 2007)	Health care insurance fraud	DT, NB, MCLP	NB best predicts fraud
(Phua et al. 2004)	Insurance fraud	NB, NN (with BPA), DT	Aggregated model predicts most accurately
(Lenard et al. 2007)	Financial statement fraud in service-based computer &	NFI	76.7% accuracy in identifying fraudulent financial statements

	technology companies		
(Koskivaara and Back 2007)	1. Goal: predict key account balances 2. Develop decision aid, i.e., Artificial Neural Network Assistant (ANNA) 3. Test using data from 10 agencies in a municipality	NN	Predictions outperformed simple analytical review procedures typically used by auditors.
(Kirkos et al. 2007a)	Identify fraudulent financial statements	DT, NN, NB using financial statement ratios	Prediction accuracy depends on training method.
(McKee 2009)	Use “stacked”, i.e., aggregated models to predict financial statement fraud	Create Three-Layer Stack: NN, LR, CT models	Stacked models outperform single-model predictions, e.g., 83 vs. 71.4 accuracy compared to NN

Key

Machine-Learning Models:

NFI = neuro-fuzzy inference
SNA = Social network analysis
SVM = support vector machine
NN = neural network
DT = decision trees
NB = naïve Bayes
KNN = k-nearest-neighbors

Other Models:

PCA = principal component analysis
MRF = Markov Random Field
LR = logistic regression
MCLP = multiple criteria linear programming
QD = quadratic discriminant
LP = linear programming

Training Methods

BeP = belief propagation (a variant on MRF)
BPA = back-propagation algorithm (NN)
VP = voted-preceptron (NN)
CT = classification tree (variation on a DT approach)

Appendix B - Data Collected on Sampled Listings

1. Group Identification, that is, whether the listing is drawn from the suspected deception, matched, or random sample
2. Listing outcome: negative, neutral, positive
3. Listing_date, date the listing appeared on eBay
4. listing_time, time the listing appeared on eBay
5. Item number per eBay numerical system
6. Item description, description of the item for sale
7. Listing Title, title of the item listing
8. Date when listing information was collected as an observation

9. Time when listing information was collected as an observation
10. Seller's eBay identification
11. Seller_reg_date, that is, date the seller registered their account with eBay
12. Slr_Registration_Days, number of days since the seller registered their account with eBay
13. Seller_reg_time, time the seller registered their account with eBay
14. Seller_location1, is the seller's account registered in the US? (0 - US, 1- Other)
15. Seller_Location2, is the seller's account registered in China or Hong Kong? (0 - Other, 1-China or Hong Kong)
16. Seller_Location3, is the seller's account registered in a developed country? (0 – G7, 1 -Other)
17. PosFBPerc, the percentage of positive feedback received by the seller since opening their account
18. FB_Duration, feedback tenure, time elapsed, measured in days, between the first feedback and the day observation was collected.
19. Status, is the seller's account in good standing (confirmed), or temporarily disabled from participating in trade (suspended, Account on hold)?
20. StarRating, the star rating of the seller's eBay account suggesting the volume of trade achieved since inception of the account.
21. UniqueNeg, the number of negative feedback postings to the seller's account from unique user IDs (+2 Neg. feedbacks posted by same user ID: +1 unique Neg. score).
22. PositiveTotal, the total number of positive feedback postings to the seller's account
23. Pos30, the number of positive feedback postings to the seller's account within the past 30 days
24. Pos180, the number of positive feedback postings to the seller's account within the past six months
25. Pos365, the number of positive feedback postings to the seller's account within the past year
26. Neu30, the number of neutral feedback postings to the seller's account within the past 30 days
27. Neu180, the number of neutral feedback postings to the seller's account within the past six months
28. Neu365, the number of neutral feedback postings to the seller's account within the past year
29. Neg30, the number of negative feedback postings to the seller's account within the past 30 days
30. Neg180, the number of negative feedback postings to the seller's account within the past six months
31. Neg365, the number of negative feedback postings to the seller's account within the past year
32. MatchObsNum, auction number for matching observation
33. LastBid, the dollar amount of the final bids on the listing
34. BuyItNow, posted price in addition or instead of the auction format
35. BuyItNow_Presence, whether the item could be purchased using the eBay Buy-it-Now feature
36. Duration, the listing of the duration in days (range: 1 to 30)
37. ListStartDate, the start date of the listing
38. ListStartTime, the start time of the listing
39. ListEndDate, the end date of the listing
40. Weekend, whether the listing ended on a weekend

41. ListEndTime, the end time of the listing
42. Listing location, seller provides clear description of geographical location (1) versus not (0)
43. Listing_Location,, the text description of the location of the product (item for sale)
44. Bid_Count, the number of bids received on the listing
45. QuantitySold, the number of items sold in the listing (range: 0 to 20)
46. ReserveMet, whether the listing reserve was met by the bids received
47. MOCC , whether the seller would accept Money order or Cashier's check as a payment
48. PersCheck, whether the seller would accept personal check as payment
49. Paypal, whether the seller would accept PayPal for payment
50. Escrow, whether the seller would agree to use an escrow service for payment
51. COD, whether the seller would agree to use collect-on-delivery as a payment
52. CC, whether the seller would accept credit card payment
53. RelistedID, the auction number if the item was not sold and was relisted under this auction number
54. SlrFeedback, the seller's overall eBay feedback score per eBay formula
55. SlrLevel, whether seller has achieved Power Seller status
56. Tran_FB_Type, type of the feedback that the seller received for the transaction (Neg, Neu, Pos)
57. FBTime, if the item sold and buyer posted feedback, the date and time of the feedback posting
58. FBComment, if the item sold and the buyer posted feedback, the text of the comment
59. BidderID, if the items sold this is the buyers eBay identification
60. BdrStatus, is the buyer's account in good standing (confirmed), or temporarily disabled from participating in trade (suspended, Account on hold)?
61. BdrPrivate, if the winning buyers' eBay feedback private?
62. BdrFeedback, this is the buyers' eBay feedback score
63. Store1, does the seller created an eBay store? (1 – Yes, 0 – No)
64. SellerEmail, did the seller include an email account information on the item description page
65. SQRTrade, is the seller registered as a "Square Trade" seller?

Endnotes

[i] Although tangential to the present investigation, a substantial research literature investigates accounting applications of machine learning models; published applications include a comparison of three models (i.e., DT, NN, Naïve Bayes) to predict qualified auditor opinions (Kirkos et al. 2007b), two applications of NFI (Comunale and Sexton 2005; Lenard et al. 2007), and applications of NN to:

(a) continuously audit and monitor financial data (Koskivaara and Back 2007)

(b) identify improper revenue recognition (Ragothaman and Lavin 2008), and,

(c) predict bank merger premiums (Shawver 2005).

[ii] Contacting the company to obtain additional information on this product produced no reply.

[iii] For example, Kirkos et al. (2007a) reverse the Type I and II error definitions adopted herein.

[iv] One-hundred-sixty-one observations were identical on the eleven selected variables. Accordingly, the net sample, after eliminating duplicated observations is 1,439 (=1,600 - 161).