Comparison of data classification methods for predictive ranking of banks exposed to risk of failure

Charles A. Worrell, Shaun M. Brady, and Jerzy W. Bala

Abstract— The difficulty of understanding a financial institution's risk of default has been highlighted by multiple recent episodes in both the U.S. and in Europe. This paper describes a study on the empirical comparison of classification techniques for predictive ranking of the 12 month risk of default in banks. This work compares the scoring capabilities of different predictive models. The models compared were induced from past levels of risk exposure observed in historic data. The ranking performance of the models is compared by assessing the highest risk cases, using the left-hand side of the model's ROC curves (i.e., curves representing true positive to false positive rates). Empirical comparisons were performed using FDIC call report data and a one-year-ahead ranking prediction schema. This comparison demonstrates that inductive machine learning techniques can be successfully applied for predictive ranking of default risk. Observed results indicate better performance by symbolic rule or decision tree based models than by traditional modeling techniques based on statistical algorithms.

Index Terms— Machine learning, Supervised learning, Predictive models, Risk analysis

I. INTRODUCTION

In order to minimize the potential economic impact of bank failures, there is a need for early detection of banks that are candidates for some form of early intervention. This paper focuses on the comparison of methods capable of ranking banks that are prone to fail in the future based on their historical financial statements. Prioritization of potentially large volumes of banks is particularly useful for regulators with limited resources. In other words, only banks with a high likelihood of failure, normally a small percentage of the total population, would be considered candidates for some form of rigorous audit and / or regulatory intervention (e.g., re-capitalization). The early and accurate prediction of the risk exposure of banks has value for reducing the financial cost associated with late and / or unnecessary interventions.

The empirical study presented in this paper involves building predictive classification models for identifying banks that might fail 12 months in the future. The data used in this study comes from the Federal Deposit Insurance Corporation (FDIC) Statistics on Depository Institutions (SDI) database, which compiles quarterly data on 8,000 U.S. banks. Each training example is a vector of 30 attribute values (primarily numeric) that corresponds to a single bank.

The Texas Ratio (TR) is a measure of default risk developed by Gerard Cassidy to predict banking failures in Texas and New England during the recessionary periods of the 1980s and 1990s. Positive training examples represent banks for which the Texas Ratio is above a critical threshold.

TR is a static calculation of a ratio that takes existing measurements of a bank's level of bad assets (noncurrent loans and repossessed properties) with its supply of capital and loan loss reserves. In the traditional TR calculation, a bank is more likely to fail when its level of problem assets exceeds the capital it has available to manage the troubled assets. Such a traditional TR calculation has its utility in a reactive mode of intervention.

We extend the use of the traditional TR calculation to a forward looking schema, where the current measurements of a bank's level of bad assets and its supply of capital and loan loss reserves represent arguments in a modeling function that maps these arguments to the risk of achieving an "above the threshold TR" risk level 12 months in the future. The use of TR for modeling expected future states has utility in a proactive mode of intervention.

The classification models that were derived are used to rank banks by their expected risk of default one year in the future. They can be used to generate a rank ordered list of target banks that should be examined. Such a target list can be used to help regulators optimize investigative resources and increase efficiency by focusing interventions on those banks that warrant the most attention due to their expected level of risk.

Since historically only a very small percentage of the highest ranked banks are prone to fail, it is important to maximize the classification precision (or true positive rate) on the highest ranked banks. As a result, the classifier's performance on lower ranked banks becomes almost irrelevant. This problem is similar to that of web search, in

which a search engine might return thousands of web pages for a given query, only a small number of which, typically the top ranked pages, are ever viewed by the user. Thus, it is much more important for an effective search engine to optimize the relevance of top ranked pages than the lower ranked ones.

While existing research offers an array of machine learning algorithms that can accommodate the ranking of classification decisions, these algorithms, and measures such as the Area Under the [ROC] Curve (AUC) that are used to evaluate their performance (e.g., [1] and [2]), generally tend to focus on the entire ranked list. Based on the needs of our application, we propose a simple method for measuring the performance of classification / ranking algorithms that instead of measuring the area under the entire ROC curve, assesses only the left most part of the curve (i.e., the part covering only the top n% of the ranked cases). In addition to measuring the performance, the proposed method may also be utilized as a substitute for accuracy or AUC, to guide the model generation process itself. The aforementioned area includes both the upper and lower regions in the left part of the curve; we have named it LAUC (Left Area Under the Curve). Figure 1 shows the ROC curves obtained from two different learning algorithms. With LAUC as a measure (i.e., to the left of the cutoff point), Algorithm 2 achieves a better performance than Algorithm 1, whereas Algorithm 1 is preferred over Algorithm 2 with the conventional AUC. Assessments using LAUC have recently been performed by other machine learning researchers [3].

The use of the lower left corner of the ROC graph space is different and new from the traditional use of selecting classifiers with points on the ROC curves located closer to the upper left corner of the ROC graph space. Classifiers in the lower left corner are "conservative", i.e., they make positive classifications with strong evidence and with few false positive errors. Their performance for the higher classification score values is better than for the lower scores. Subsequently, they are much more suited for ranking upper percentiles of the scoring target list.

Because only a small proportion of the banks represented in the FDIC SDI data have a high risk of default, the target class distribution in the training data set might become quite imbalanced. Our comparative study tests different machine learning algorithms for their ability to handle the imbalanced class distribution problem and illustrates their performance with measurements of the LAUC.

Section II of this paper describes some previous work on modeling financial risk, Section III describes the underlying approach used here for ranking banks using classification models, Section IV presents an empirical comparison of rankings made with classification models, and finally, Section V describes our conclusions.
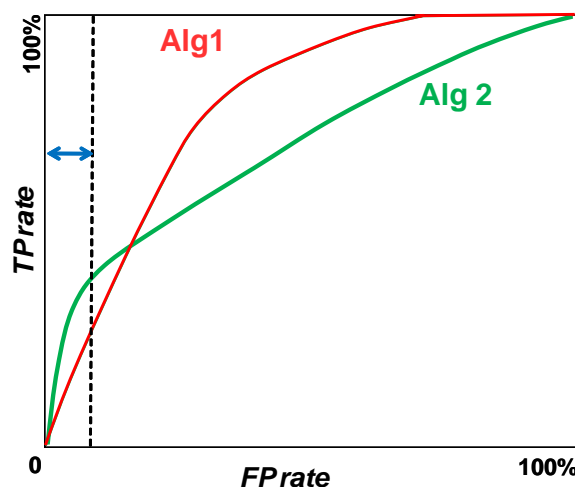

Fig. 1. Example ROC curves- True Positive (TP) Rate versus False Positive (FP) Rate

II. PREVIOUS WORK

A. Financial Risk Modeling

An analyst reasons about individual bank financial condition on the basis of analogy, experience, heuristics, and theory as well as research evidence. This reasoning process is supported by diverse financial data sources. As the amount of this data explodes, analysts are faced with the difficult task of analyzing it (e.g., sifting the banks into meaningful risk groups). One approach to automating this task is to use an expert system with embedded declarative knowledge. Developing an effective expert system can be a challenge if the available experts are unable to articulate the rules that govern their reasoning, which is often the case. This can sometimes be overcome by using supervised learning techniques to uncover the rules in use. Even then, the ambient financial conditions represented in the data can change unpredictably, making it difficult to pre-program all the necessary knowledge to model its complexity in advance. In addition, it is complex and labor-intensive to adequately capture an expert's knowledge; this has become a major bottleneck in the development of financial expert systems. An adaptive approach, such as one based on machine learning algorithms, can provide a means to acquire the knowledge required for automated data-driven decision-making. Examples of such financial risk modeling approaches include the data analysis techniques of multivariate analysis [4], neural networks [5], logistic regression [6], and financial ratio based models [7].

In general, the use of statistical techniques has suffered from the restrictive assumption of distributional normality, which is rarely satisfied by complex financial attributes and target class distributions. For example, distributions of ratios derived from financial statements are complex and skewed. This skew suggests a deviation from strict proportionality between ratio components. This challenge gave birth to various studies on alternative modeling techniques, such as decision trees and K-nearest neighbor [8]. These techniques do not require any underlying probability distribution or

dispersion equality.

Among machine learning approaches, the use of neural network techniques has achieved wide acceptance in bank failure prediction [9]. However, such techniques lack the ability to explain their decision-making process, which is often dissatisfying to expert practitioners (the model comprehensibility challenge). In contrast, the use of rule based modeling techniques provides some insight into the decision-making recommendation [10].

B. Ranking with Classification Models

Ranking examples using classification models, such as rule based models, is not a novel undertaking. There have been many attempts in the past, including those by researchers in the fields of expert systems, e.g., MYCIN [11] and fuzzy logic [12]. There are also several existing studies in Machine Learning e.g., [13]. These studies generally focused on methods for generating partial matching, whereby the scores for individual examples are computed based on how well they match the rules. In these approaches, examples that satisfy all conditions of a rule share the same score.

Extensive studies, on the other hand, have been dedicated to the incorporation of ranking capabilities into the decision tree learning paradigm. Related work generally falls into four groups of methods: learning probability estimation trees [14], geometric methods [15], hybrid trees including: the Perceptron Tree [16] and NBTree [17], and ensembles of trees [18].

Recently, Loterman et.al [19] has reported in the International Journal of Forecasting on a trend of using non-linear techniques that perform significantly better than more traditional linear techniques in modeling financial markets. They also advocate the use of comprehensible model components. The research presented in this paper contributes to this trend.

III. SUPERVISED MODELING FOR RANKING

In this paper we consider a subset of machine learning tasks called supervised learning. Supervised learning infers a classifier model from training examples. The inference process generalizes from training examples to the unseen regions of the larger data set.

A set of training examples describing decision classes (i.e., classification outcome) is input to the learning program in order to derive general descriptions (models) of the decision classes. This step involves the process of learning from training data. A set of training examples annotated by concept membership information is used as the basis for automatically inducing a general description for each decision model. The learned model is correct for the given examples. Since it extends its membership information to unseen parts of the representation space it is assumed to also be a good predictor for the classification of unobserved examples of the concept.

In this paradigm an example can be anything that can be expressed in terms of representational language. An example can be a physical object, a situation, a cause, or a concept. Training examples are usually described in an attribute-based representation. In an attribute-based representation, an example is represented as an n-tuple of attribute values, where n is a number of attributes. All n attributes define the event space. A domain is associated with each attribute used to describe examples. The domain determines the values that the attributes may assume. The values in a domain may be unordered (i.e. nominal), ordered (i.e. linear), or hierarchically structured.

Most machine learning techniques generate model descriptions by detecting and describing similarities among positive examples and dissimilarities between positive and negative examples. Constructing model descriptions from training examples involves the transformation of training examples using a set of refinement operators [20]. A refinement operator is either a specialization or a generalization operator. When applied to a training example, a generalization / specialization operator transforms it into a more general / special description.

Each description characterizes a subset of all examples, while all of the descriptions in a given representational language form a description space. Learning can be viewed as a search process through the description space to find clusters of examples and resulting descriptions of the target concept. Generalization / specialization operators are search operators. Search heuristics contain preference criteria (also called biases). One of the most important description preference criterion is accuracy. Description accuracy depends on the completeness and consistency of the description with regard to the learning examples. Some other preference criteria are simplicity and comprehensibility.

Classification accuracy has been used as a major performance metric for machine learning algorithms. However, many real life machine learning applications involve the ranking of cases rather than their classification. Ranking of cases is usually based on some kind of reliability, likelihood, or numeric assessment of the quality of each classification. In other words, the decision-making process extends the class membership prediction to include an estimate of the reliability of that prediction. For example, in credit application processing the goal is to rank applicants in terms of their likelihood of profitability and / or likelihood of loan default. This is significantly different than simply classifying them into qualified or un-qualified groups. Other decision-making applications where case ranking could be of importance include bankruptcy prediction, medical diagnosis, customer targeting for marketing campaigns, and customer churn prediction. In addition, the use of rule based models facilitates human comprehension of the ranking process, which may be an essential requirement in decision ranking applications. Ranking of cases is also valuable in applications where it is preferable to defer a decision in the absence of certainty, than to make a wrong decision (i.e., medical or military applications).

The Machine Learning community has investigated the incorporation of ranking capabilities in a decision tree learning paradigm [21]. However, not much work has been done for ranking with rules. Although rules are similar to decision trees, there are also some important differences between them

when used for ranking. Separate-&-conquer (covering) techniques of rule learning algorithms may generate rules that overlap, whereas divide-&-conquer techniques with decision trees do not. Rules may not cover some areas of a feature space, but leaf nodes of a decision tree cover the entire area of the feature space. This difference brings both research challenges and opportunities for developing methods of ranking cases with rules. In addition to the above differences, a rule learning algorithm for a two class problem may only learn rules for one class, but a decision tree always includes leaf nodes for both classes. Rule learning algorithms tend to generate fewer rules than leaf nodes of a decision tree.

IV. EXPERIMENTS

A. Data Sets

The financial data used in this empirical evaluation was obtained from the FDIC SDI data repository which contains the Uniform Bank Performance Report (aka Call Report) that each FDIC insured institution is required to file quarterly. The FDIC Institution Directory (ID) provides the latest comprehensive financial and demographic data for every FDIC-insured institution, including the most recent quarterly financial statements, with performance and condition ratios (http://www2.fdic.gov/sdi).

Table 1 lists the attributes of the FDIC data used in experiments. The data represents historically recorded performance measures for a given three month period. A ground truth measure of risk exposure was calculated using the Texas Ratio (TR). TR was developed by RBC Capital Markets' banking analyst Gerard Cassidy as a way to predict bank failures during the state's 1980s recession. The ratio is still widely used throughout the banking industry.

Cassidy's original Texas Ratio formula is:

TR = (Non-Performing Loans + Real Estate Owned) / (Tangible Common Equity + Loan Loss Reserves)

TR is determined by dividing the bank's nonperforming assets (nonperforming loans and the real estate now owned by the bank because it foreclosed on the property,) by its tangible common equity and loan loss reserves. Tangible common equity is equity capital less goodwill and intangibles. As the ratio approaches 1.0, the bank's risk of failure rises. Relatively speaking, the higher the ratio, the higher the bank's risk of default.

TABLE 1
FDIC Data Attributes

Attribute	Definition
1	Yield on earning assets
2	Cost of funding earning assets
3	Net interest margin
4	Noninterest income to earning assets
5	Noninterest expense to earning assets
6	Net operating income to assets
7	Return on assets (ROA)
8	Pretax return on assets
9	Return on Equity (ROE)
10	Retained earnings to average equity
11	Net charge-offs to loans
12	Credit loss provision to net charge-offs
13	Earnings coverage of net charge-offs
14	Efficiency ratio
15	Assets per employee
16	Cash dividends to net income
17	Loss allowance to loans
18	Loan loss allowance to noncurrent loans
19	Noncurrent assets plus other real estate owned to assets
20	Noncurrent loans to loans
21	Net loans and leases to deposits
22	Net loans and leases to core deposits
23	Equity capital to assets
24	Core capital (leverage) ratio
25	Tier 1 risk-based capital ratio
26	Total risk-based capital ratio
27	Average Assets
28	Average earning assets
29	Average equity
30	Average Loans

TRs were calculated for the quarter one year after the quarter used to describe the data with the attributes presented in Table 1. Different TR thresholds were used to annotate the class membership (e.g., high or low risk exposure). The data sets together with their one-year-ahead annotation constituted the training data sets. The training data sets were used to learn different classification models.

B. Classification Models

We compared the predictive ranking performance of the following classification models:

1) Support Vector Machines (SVM): A classification technique that constructs a separating hyperplane in the attribute space that maximizes the margin between the instances of different classes [22]. An SVM engine with a RBF based kernel was used in the experiments.

2) C4.5: An algorithm used to generate a binary decision tree [23]

3) CN2: An algorithm for rule induction learning [24].

4) Naïve Bayes: A simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions [25].

5) Logistic Regression: An algorithm used for predicting the probability of occurrence of an event by fitting data to a logistic curve function [26].

C. Comparison of Predictive Ranking Performance

We conducted a number of experiments using the FDIC data set prepared for training classifiers as described in section IV-A. A three-fold cross-validation was used to evaluate predictive performance using the following measures:

1) True Positive Rate (TPR) defined as TP/(TP+FN), (also called Hit Rate, Recall, Sensitivity)

2) Positive Predictive Value (PPV) defined as TP/(TP+FP), (also called Precision)

3) Area Under the ROC Curve (AUC)

4) True Positive Rate over False Positive Rate Gain in the LAUC Region (TPR/FPR). FPR is defined as FP/(FP+TN)

To assess performance of predictive ranking we use a portion of the AUC measure. Instead of measuring the area under an entire ROC curve, we look only at the left most part of AUC, that is- the part covering only the top 25% of ranked cases. Figure 2 depicts the ROC curves for the entire measurement area. The best "AUC performers" (Table 2) are the C4.5 and Naïve Bayes classifiers. Both also show the TPR measure above 0.6 for the target class.

TABLE 2
Empirically Evaluated Performance Measures*

Method	TPR	PPV	AUC
SVM	0.00	0.99	0.81
C4.5	0.60	0.88	0.82
CN2	0.43	0.92	0.76
Naïve Bayes	0.85	0.65	0.82
Logistic Regression	0.00	0.99	0.81

*These results were obtained for the Q1-2007/Q1-2008 FDIC train / predict data sets (i.e., the Q1-2007 financial descriptors are used to predict the risk exposure level in Q1-2008).

Figure 3 depicts the left most part of the ROC, the part covering the top 25% of ranked cases based on the scores predicted by classifiers. The 25% cut-off point was used because the target class, i.e., banks with an elevated risk exposure, represented 25% of cases in the data. It can be observed now that the best "LAUC performers" are the C4.5 and CN2 classifiers.

The experimental results show clearly that performance measured by LAUC does not necessarily correlate with that measured by the AUC. They also show that classifier models induced by symbolic machine learning (i.e., decision tree and rule-based algorithms) exhibit better predictive ranking performance than the other classifiers tested. Table 3 depicts the True Positive over False Positive Gain in the LAUC (i.e., left-hand side of the ROC curve) region at three different FPR cut-off points. This gives us some feel for how much better we are performing in TP rate over FP rate in the top portion of the ranking list.

TABLE 3
TPR over FPR Gain in the LAUC Region (TPR/FPR)

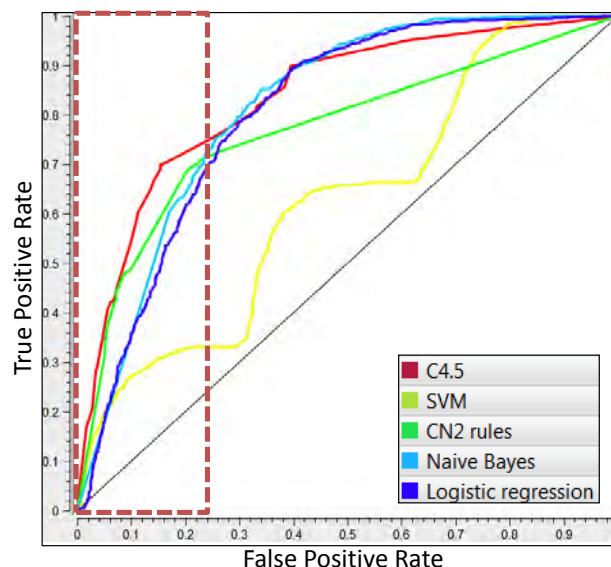| Method	FPR @ 0.05	FPR @ 0.10	FPR @ 0.20	Average Gain
SVM	4	2.5	1.5	2.66
C4.5	8	5.5	3.5	**5.66**
CN2	7	5	3.2	**5.06**
Naïve Bayes	4	3.5	3.1	3.53
Logistic Regression	4	3.5	3	3.50

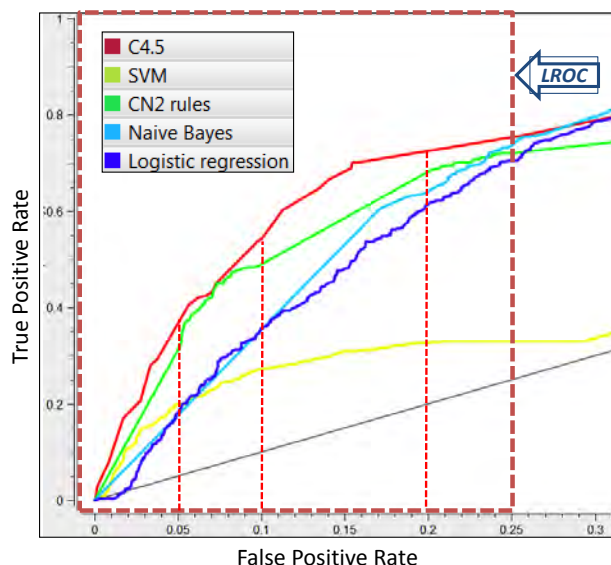Fig. 2. ROC curves for the target class (exposure risk class)

Fig. 3. Left-hand Side of ROC

It can be observed that the best predictive ranking performers for this application are two symbolic learning classifiers - C4.5 and CN2.

V. CONCLUSION

This paper described a study on the empirical comparison of classification techniques for predictive ranking of the risk of default in banks. We concluded that LAUC provides a better measure of classifier performance than AUC for this purpose, and that the performance of LAUC does not necessarily correlate with that of AUC. Specifically, the results demonstrate that inductive machine learning techniques can be successfully applied for predictive ranking of financial risk and point to better performance of symbolic rule or decision tree based models than traditional modeling techniques based on statistical algorithms. The mechanisms for trading off TPR with PPV (Recall with Precision) that are inherent to C4.5 and CN2 as well as other learning decision tree or rule set models

may account for this stronger performance. It is our conjecture here that precise rules or decision tree branches, i.e., the ones with higher PPV rates (Precision), provide better results when assessed with LAUC, while more general rules or branches, i.e., the ones with higher TPR rates (Recall) may work better when assessed with AUC. This TPR-PPV control is especially suited for ranking risk in today's complex financial data repositories (e.g., FDIC), where assumptions of distributional normality are rarely met, both for class and attribute level distributions. The presented research is consistent with the recent findings by Loterman et.al. [19] on the preferable use of non-linear techniques with comprehensible model components for financial data modeling.

In future research, we will empirically validate other methods for assessing predictive ranking performance. Specifically, we will conduct a broader comparative study that will include a larger repertoire of modeling techniques (e.g., parametric statistical models, neural networks, ensemble models), the use of additional financial performance measures and their different thresholds for classification (e.g., Non-current Loan Ratio, Tangible Common Equity Ratio, Tier 1 Risk Based Ratio, etc.), and extended benchmarking data sets. We will also investigate issues related to ranking model understandability (e.g., degradation with the increased complexity of the ranking methods) and interoperability with other modeling frameworks. Finally, the long-term goal of this research is to develop a new class of rule based financial data modeling methods geared towards predictive risk ranking.

REFERENCES

[1] Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159.

[2] Provost F. J. & Domingos, P. (2003). Tree induction for probability-based ranking. Machine Learning, 52(30).

[3] Zhang, J., Bala J., Hadjarian A., Han B. (2010), Ranking Cases with Classification Rules," a book chapter in "Preference Learning, by Springer-Verlag.

[4] Sinkey, F., (1975). A multivariate statistical analysis of the characteristics of problem banks. The Journal of Finance 30

[5] Tam, K., Kiang, Y., (1992). Predicting bank failures: A neural networks approach. Management Science 38, 926-947.

[6] Estrella, A., Park, S., Peristianai, S., (2000). Capital ratios as predictors of bank failure. Federal Reserve Bank of New York Economic Policy Review.

[7] Ohlson, .A. (1980). Financial Rations and the Probabilistic Prediction of Bankruptcy, Journal of Accounting Research, Vol. 18, no. 1, pp. 109-131.

[8] Tam, K.Y., (1991). Neural Network Models and the Prediction of Bank Bankruptcy", OMEGA, Vol.19, no.5, pp. 429-445.

[9] Han, H. Jo, I., and H. Lee, (1997). Bankruptcy prediction using case-based reasoning, neural networks and discriminant analysis, Expert Systems with Applications, Vol.13, no.2, pp. 97-108.

[10] Pazzani M., 1997. Comprehensible Knowledge Discovery: Gaining Insight from Data, First Federal Data Mining Conference and Exposition.

[11] Buchanan and Shortliffe (eds.) (1984). Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming project.

[12] Wang L.X. and Mendel J.M. (1992), Generating Fuzzy Rules by Learning from Examples, IEEE Transactions on Systems, Man and Cybernetics, Volume 22, No 6., pp. 1414-1427.

[13] Zhang J & Michalski, R.S. (1995). An Integration of Rule Induction and Exemplar-Based Learning for Graded Concepts. Machine Learning 21(3): 235-267

[14] Provost F. J. & Domingos, P. (2003). Tree induction for probability-based ranking. Machine Learning, 52(30).

[15] Alvarez I. & Bernard S. (2005). Ranking Cases with Decision Trees: a Geometric Method that Preserves Intelligibility. IJCAI, 635-640.

[16] Utgoff, P. (1988). Perceptron Trees – A case study in hybrid concept representation. Proceedings of the seventh national conference on artificial intelligence, 601-606.

[17] Kohavi R (1996). Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.

[18] Provost, F., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In Proceedings of the 15th International Conference on Machine Learning, pp. 445-453. San Francisco: Morgan Kaufmann

[19] Loterman G., Brown, I., Martens, D., Mues, C., and Baesens, B. Benchamarking Regression Algorithms for Loss Given Default Modeling. International Journal of Forecasting. Volume 28, Issue 1, January–March 2012, Pages 161–170.

[20] Michalski, R. S. (1983). A Theory and Methodology of Inductive Learning. Artificial Intelligence, 20:111-116.

[21] Provost F. J. & Domingos, P. (2003). Tree induction for probability-based ranking. Machine Learning, 52(30).

[22] Corinna Cortes and V. Vapnik, (1995). Support-Vector Networks, Machine Learning, 20.

[23] Quinlan, J. R. C4.5, (1993)Programs for Machine Learning. Morgan Kaufmann Publishers.

[24] Peter Clark and Tim Niblett, The CN2 Induction Algorithm", Machine Learning, Volume 3, Number 4, 261-283, DOI: 10.1023/A:1022641700528.

[25] Domingos, Pedro & Michael Pazzani (1997) On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29:103–137.

[26] Agresti A (2007). Building and applying logistic regression models. An Introduction to Categorical Data Analysis. Hoboken, New Jersey: Wiley. p. 138.

Charles Worrell is a Principal Scientist at the MITRE Corporation in McLean, Virginia where he develops systems based on Bayesian Inference Networks to detect events such as disease outbreaks, accounting fraud, and other illicit activities for customers that range from the U.S. Securities and Exchange Commission to the nation's Intelligence Community. His research interests include modeling systemic risk, automated detection of financial crimes, and simulating human decision making.

He started his career as an officer in the U.S. Navy, serving aboard ships in both the Pacific and Atlantic Fleets and specializing in operations and telecommunications. Since that time, he has worked in Network Operations at Verizon and as Director of Systems Development at the American Automobile Association's (AAA) Response Services Center. He is the lead inventor on the patented DAP-E Method for Decomposing Human Behavior into Quantified Layers of Perception.

Dr. Worrell holds a Ph. D. in Information Technology from George Mason University, as well as degrees from the University of Pennsylvania and the Naval Postgraduate School.

Shaun Brady received his B.Sc in International Finance in 1980 from San Francisco State University in San Francisco, California, and his Masters in Management Information Systems (1992) and Doctorate of Management (2010) from University of Maryland University College, College Park, Maryland. He has been a Principal with the MITRE corporation since 2009 helping develop the strategic plan and research agenda for our newly created Financial Stability and Oversight portfolio. His focus is on leveraging MITRE's extensive defense and intelligence related systems engineering capabilities to help the financial regulatory community assess and implement new data quality and governance processes necessary to identify and monitor systemic risks. Dr. Brady began his professional career with over 15 years in senior leadership positions in banking before transitioning into consulting and spending the next 10 years prior to joining MITRE supporting the risk management and related data governance and information product needs of

many of the world's largest financial institutions. He is a frequent speaker on data quality and risk management and has contributed articles to a number of financial publications and journals. He holds US Patent 6,633,875 on the collection, anonymizaton, and sharing of confidential data over the internet, and has US Patent pending 60/278,446 for a new method for managing the credit risk life cycle.

Jerzy Bala holds a Ph. D. in Information Technology from George Mason University, Fairfax Virginia, as well as MS in Electrical Engineering from AGH University of Science and Technology, Krakow, Poland.
His employment experience includes InferX Corporation, Datamat Systems Research, Inc., and consulting work for MITRE Corporation. His field of interest is in Predictive Analytics, and specifically in Machine Learning, Knowledge Discovery, Data Mining, Text Mining, Information Retrieval and Extraction. During his 20 year career he has served as a Principal Investigator on a number of projects under the aegis of Office of Naval Research, Air Force Research Lab, Defense Research Projects Agency, Missile Defense Agency, National Geospatial-Intelligence Agency, Central Intelligence Agency, Department of Education, and the U.S. Department of Veterans Affairs. He has conceived two patented data mining algorithms on distributed Data Mining and visual knowledge representation. His research resulted in over 80 peer-reviewed papers published in IT conference proceedings and scientific journals. He is the co-author of the milestone book, Machine Learning - A Multistrategy Approach, Morgan Kaufman, San Mateo CA., 1994.
Dr. Bala has received ten Commonwealth of Virginia Outstanding Achievement Awards for success in the U.S. Department of Defense research projects. In 1993, he was awarded a two-year prestigious postdoctoral research grant by the National Science Foundation in Computational Science and Engineering to investigate a class of multistrategy Machine Learning algorithms.