

# Finding Terrorists via Data Mining

---

G. S. Almasi and M. J Rothman

almasi@cloud9.net (914) 232-8195

mjrothman@preference-engine.com (845) 223-5925

# “Data Mining Could Have Prevented 9-11” – Rep. Curt Weldon

---

Philadelphia, August 26, 2002 (by Dan Verton in *Computerworld*)

Legislation that Congress failed to adopt two years ago would have created an interagency data-mining capability that could have detected and helped prevent last September's terrorist attacks, a senior Republican congressman asserted last week.

Speaking near his home district at the Information Sharing & Homeland Security conference here, **Rep. Curt Weldon (R-Pa.)** lambasted the federal government, including Congress, for failing to act on critical **data-mining and intelligence integration** proposals that he and others authored years before the terrorist attacks.

"There are 33 classified agency systems in the federal government, but none of them link their raw data together," said Weldon, chairman of the House Subcommittee on Military Research and Development. **"We could have and should have had a better data-fusion capability on and before 9-11."**

# We have Data Mining Expertise

---

- **Michael J. Rothman**

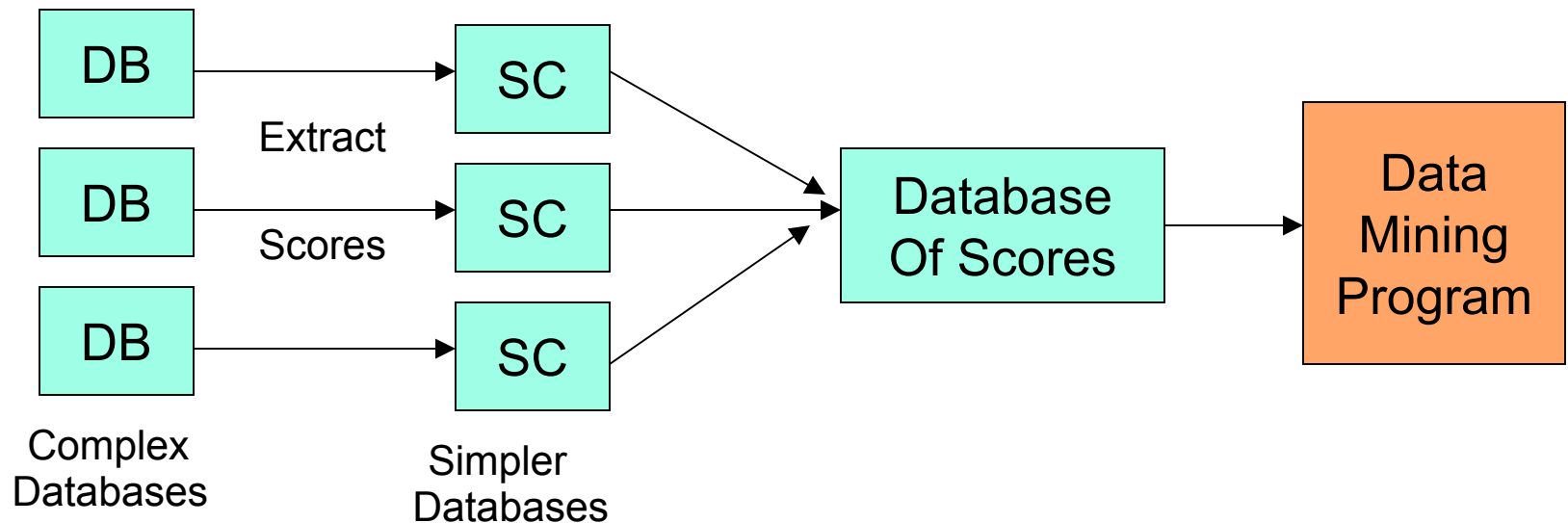
- Founder MR&A, 10 years' experience extracting value from very large consumer and business databases
- SVP for Strategic Information, First USA Bank, integrated many databases to enable the bank to understand customer interests
- Principal, IBM Consulting, Business Intelligence
- PhD, University of Michigan

- **George S. Almasi**

- Technology innovator MR&A, 8 years' experience at IBM developing and applying Data Mining tools
- Expert at visualization, letting users “see” patterns in the data
- Senior Manager at IBM TJ Watson Research Center
- PhD, Massachusetts Institute of Technology

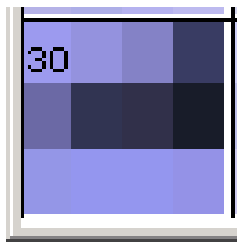
# Our “Score Extraction” Can Yield Early Results

- “Score Extraction” is a technique we developed at First USA for getting early data mining results from disparate data bases without the long delay entailed in fully fusing such data first.



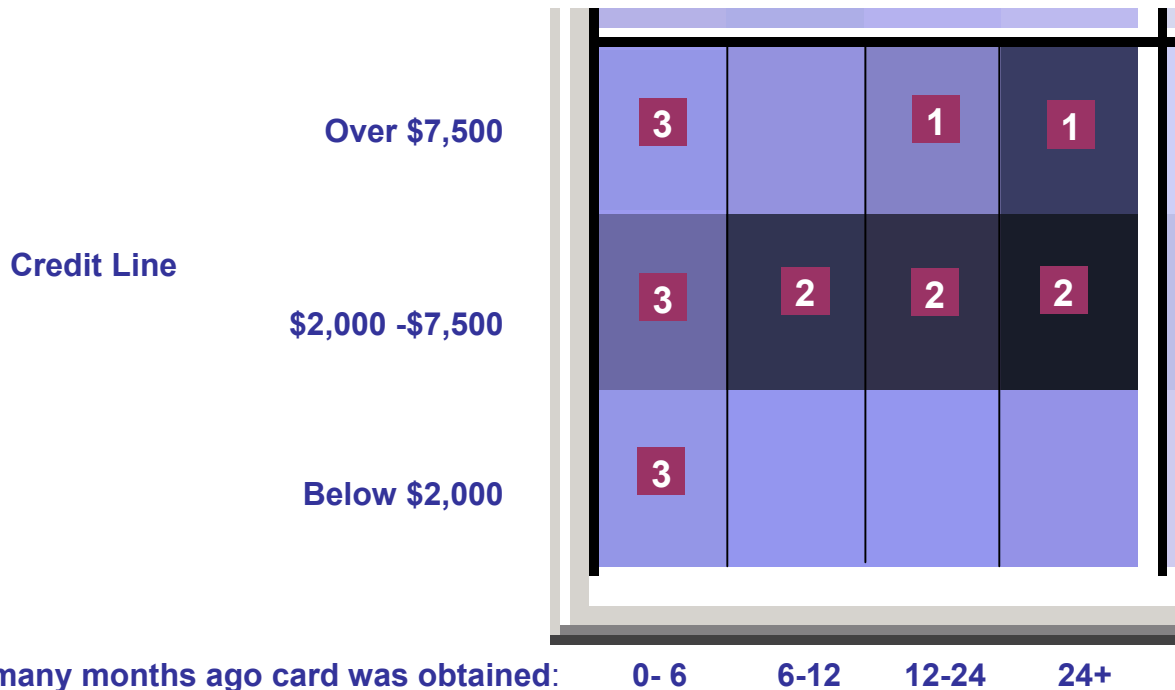
# Example: Bankruptcy Prediction

- **We applied Score Extraction to bankruptcy prediction:**
  1. Complex Credit Bureau data was transformed by filtering out many of the 200+ items per record and by expanding each household's revolving credit balance into 12 categories, based on their cards' ages and credit lines.
  2. A data mining program divided these households into 36 groups, using these new categories.
  3. Checking subsequent bankruptcy history of each group revealed one group with unusually high bankruptcy rate
  4. Placing each group's credit balances in a small 3x4 checkerboard pattern revealed the characteristic "signature" below for the high bankruptcy group:



This is explained in the next chart...

# The Analyst can see a Bankruptcy Signature



## What an analyst sees here:

1. They used to be able to get credit lines over \$7500, but not in the last two years
2. They have a “sweet spot” in credit lines between \$2000 & \$7500... easy to get and they quickly add up to a lot of money
3. They have been unable to get new cards in the last 6 months as their “risk” score elevates

Score (balance / credit line): ■ None ■ Max

efax650\_36 Kmap -- grayest = 29% of 95157 recs, colored by BANKRUPTCIES (mean=0.003), sorted by weights

File View Color By Sort By BB Signatures Exit

0 0 1K(2%) 1 0.005 3K(3%) 2 0.001 3K(3%) 3 0.001 921(0%) 4 0.002 2K(2%) 5 0 28K(29%)

BANK\_BAL02 MORT\_BAL BANK\_BAL02 BANK\_BAL05 BANK\_BAL01 BANK\_BAL04 STUDENT\_BAL INSTFIN\_BAL INSTFIN\_BAL STUDENT\_BAL INSTFIN\_BAL MORT\_BAL OTHDEPT\_BAL

We've developed Special Tools...  
...for spotting patterns fast in the extracted scores

*These 530 households  
(0.6% of the total)  
Have high  
Probability  
Of bankruptcy*

18 0.013 994(1%) 19 0.006 2K(2%) 20 0.003 2K(3%) 21 0.002 3K(4%) 22 0.009 7K(7%)

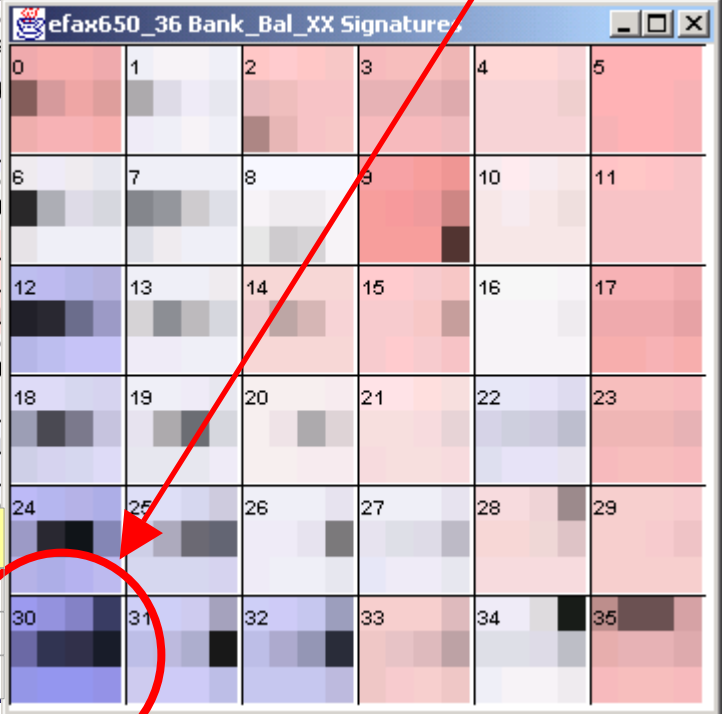
BANK\_BAL05 BANK\_BAL08 BANK\_BAL08 BANK\_BAL11 HOMEFURN BAL AUTO\_BAL MORT\_BAL HOME\_BAL MORT\_BAL BANK\_BAL11

24 0.026 734(0%) 25 0.013 1K(1%) 26 0.006 3K(3%) 27 0.006 1K(1%) 28 0.002 21K(21%)

BANK\_BAL08 BANK\_BAL11 BANK\_BAL11 BANK\_BAL11 POPDEPT\_BAL BANK\_BAL POPDEPT\_BAL BANK\_BAL BANK\_BAL MORT\_BAL

30 0.038 530(0%) 31 0.019 957(1%) 32 0.02 608(0%) 33 0.002 1K(1%) 34 0.005 11K(11%)

BANK\_BAL11 BANK\_BAL08 BANK\_BAL11 BANK\_BAL12 BANK\_BAL08 BANK\_BAL11 POPDEPT\_BAL OTHDEPT\_BAL BANK\_BAL MORT\_BAL BANK\_BAL



	under 6 mo.	6 to 12 mo.	12 to 24	over 24
over \$7500	Bal_03	Bal_06	Bal_09	Bal_12
\$2000 to \$7500	Bal_02	Bal_05	Bal_08	Bal_11
under \$2000	Bal_01	Bal_04	Bal_07	Bal_10

# We can contribute to Homeland Security

---

- We can extract patterns from complex and disparate databases more quickly than current approaches
- The techniques we used for bankruptcy prediction should also work for identifying terrorists, using data from the private sector (credit usage, credit fraud, travel records, etc.) as well as the government sector.
- In concert with the right people in the intelligence community, we can
  - Define a data unification and filtering project, combined with data mining, to find predictors for terrorist activity
  - Build a prototype using sample data
  - Act as advisors on a larger follow-on project



# We need to connect with the appropriate people in Congress in order to contribute to homeland security.

---

- Many agencies are involved in homeland security, but some in Congress have a larger perspective. If we connect, then we will be able to find the right place to take advantage of our approach.
  - Our approach is fitting for this atmosphere of urgency. We don't try to solve the big database problem, but rather move quickly to extract key features from each relevant database, and then combine them so that they can be used by our investigators
- It may be most important for senior officials to hear our message of inexpensive, rapid integration of heterogeneous databases.

## More Information

---

- Our resumes are attached
- Our site <http://www.preference-engine.com> has some of our publications, and also a live demo of the bankruptcy data visualization
- The bankruptcy study is described at <http://www.twocrows.com/largedb.pdf>
- To contact us:
  - George Almasi,  
[almasi@cloud9.net](mailto:almasi@cloud9.net), 914-232-8195
  - Michael Rothman,  
[mjrothman@preference-engine.com](mailto:mjrothman@preference-engine.com), 845-223-3900