# The Top 10 Data Mining Mistakes

The tutorial will reveal the top mistakes we Data Miners can make, from the simple to the subtle, using real-world (often humorous) stories. The topics will be presented from case studies of real projects and the (often overlooked) symptoms that suggested something might be amiss...

The goal will be to learn "best practices" from their flip side -- mistakes. But also, following the introduction of a topic (e.g., bootstrapping) the 3-hour tutorial format will allow for brief summaries of how to do it right -- that is, mini-tutorials on the key principles to keep in mind when using a particular Data Mining technique.

Mistakes to be covered include: Lack data, Focus on Training, Rely on 1 technique, Ask the wrong question, Listen (only) to the data, Accept leaks from the future, Discount pesky cases, Extrapolate (practically and theoretically), Answer every inquiry, Sample without care, Believe the best model.

*Audience:*

Those who seek to find patterns in data. Anyone who's worked with, or plans to work with, real data. The subtler mistakes (e.g., improper sampling) may elude novice practitioners, but the story-symptom-solution format of the tutorial should be accessible to all conference attendees. Experienced statisticians and researchers will hear enough to stay engaged and will be encouraged to contribute stories as well during what is planned to be a very interactive tutorial.

*Coverage:*

The material is centered on how to think rightly about a problem, and not on technical equations or terms. The practical nature of the illustrative problems helps to focus participants on the "uncommon common sense" necessary to practice well the art of Data Mining.

*Background:*

The best background for attendees is to have a problem they want to solve, and to have experience trying any analysis technique. But, non-technical people as well, have commented very positively on earlier, shorter versions of the "Top 10 DM Mistakes" talk.

*Biography of Dr. John Elder:*

Dr. John Elder heads a small Data Mining firm with offices in Charlottesville, Virginia, and Washington, DC. John earned degrees in Electrical Engineering at Rice University, then worked in the Defense consulting industry for 5 years, where he authored an early Data Mining tool for the Air Force which led to improved guidance and flight control

applications. He then earned a Ph.D. in Systems Engineering from the University of Virginia while working as Director of Research for an investment management firm, and wrote an influential tool for global optimization. After two years post-doctoral research at Rice in the Computational and Applied Mathematics Department, John returned to Virginia and started Elder Research, Inc. in 1995, where he's led projects successfully applying Data Mining to a wide variety of financial, commercial, and medical applications -- including cross-selling, customer segmentation, direct marketing, credit scoring, sales forecasting, stock selection, drug efficacy, biometrics, market timing, and fraud detection. Dr. Elder has written several book chapters and articles on pattern discovery techniques, and is a frequently invited conference speaker. He is active on Statistical and Engineering journals and boards, and his popular Data Mining courses are acclaimed for clarity. He has been named to Who's Who in the World for his contributions to the field. Dr. Elder has been honored, since Fall 2001, to serve on Panel formed by Congress to guide critical defense technology for the National Security Agency.