

GIORDANO, HALLERAN & CIESLA

A PROFESSIONAL CORPORATION

ATTORNEYS AT LAW

PLEASE RESPOND TO:

U.S. POSTAL SERVICE ADDRESS:

POST OFFICE BOX 190

MIDDLETOWN, NEW JERSEY 07748

OR:

HAND DELIVERY AND OVERNIGHT SERVICE ADDRESS:

125 HALF MILE ROAD

RED BANK, NEW JERSEY 07701

(732) 741-3900

FAX: (732) 224-6599

www.ghclaw.com

May 2003

Contracting with the Soothsayers: Fundamentals of Statistical Model Licensing By Kurt E. Anderson

Introduction

Statistical models are becoming more and more frequently used in all aspects of commerce. Most people are exposed to statistical models in political elections and when applying for a loan. The major media outlets use statistical models to predict the winners of elections based on exit polls. Similarly, lenders use statistical models to evaluate credit risk associated with applicants for loans. However, statistical models are also being regularly used for other applications including everything from evaluating insurability to predicting the success of marketing campaigns to guiding securities transactions for hedge funds. There is no end to what they may be applied to in the future.

Companies frequently rely on independent contractors to develop statistical models rather than maintaining in-house staff for this purpose. Unfortunately, contract negotiations for the development of statistical models are frequently stymied by a lack of understanding of the intellectual property issues involved. This article will try to elucidate some of the more befuddling concepts.

The Basics.

Statistical models are, in essence, complex mathematical algorithms or formulas. They are derived or “built” by analyzing data. Lots of data. Typically, the “predictiveness” (and thus the true value) of a model is a function of the skill of the developer, the quantity of data available to build the model, the quality of the data, and the degree to which the model will be applied to a narrowly focused situation.

From a value perspective, statistical relationships and the weighting coefficients associated with them form the two principal components of most statistical models.

Each model will also contain several variables which have statistically significant relationships with the subject that the model is formulated to predict. These are the relationships between two or more data elements that are found, by the developer, to be predictive. Let’s use

an example that is very near and dear to my heart: wine. Suppose we were building a statistical model to predict the size of a grape crop for a wine producer. An analysis of the data may reveal that average temperature is statistically associated with the size of the grape crop. If so, we would use this statistical relationship, among others, in our model.

Each statistical relationship will be given a particular weight depending upon the “predictiveness” of that relationship. The predictiveness of the model is primarily a function of the predictiveness of each statistical relationship and the weight given to each statistical relationship within the model. Statistical relationships within a model are given a weight by the use of a coefficient which gives the statistical relationship greater or lesser influence on the model as a whole. In our prior example, if the statistical relationships used in the model included average temperature, air humidity, type of fertilizer and precipitation, it may be that we would give greater weight to temperature and precipitation and less weight to air humidity and type of fertilizer.

The vast majority of statistical relationships that are found in any set of data used to build a model within a given industry are generally known. It is very unusual for a new statistical relationship to be discovered. Statistical relationships are the tools of the trade used by developers to build models. However, the composition of statistical relationships within a given model and, especially the weighting coefficients associated with those statistical relationships, can be unique to each model. Whether all of the weighting coefficients are unique may depend on the industry for which the model is built. For example, in the credit scoring industry, it is not unusual for all of the weighting coefficients in a custom model to be specifically derived. However, in the insurance industry, model developers may use weighting coefficients that are commonly known and consistently used in the industry in combination with weighting coefficients that are specially derived for a particular statistical model.

Statistical models are typically “delivered” in the form of a written report or software. The written report can then be used to write the software to automate the calculations performed by the model. Model developers may or may not provide the computer programming services needed to “code the mode.”

Which type of intellectual property is a statistical model?

Patent. In theory, as an algorithm, a statistical model could be subject to patent protection. However, in order to obtain patent protection, one would be required to disclose the algorithm in the patent application. Since most statistical models are developed for the purpose of obtaining a competitive edge over competitors, disclosing the specifics of the model in a patent application would not be a very effective means of protecting the asset.

Copyright. Attempting to apply concepts of copyright gets even more confusing. Copyright protects the expression of an idea, not the idea itself. Since statistical models are primarily utilitarian in nature, there is usually not a great deal of “creative expression” capable of

GIORDANO, HALLERAN & CIESLA

A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW

125 Half Mile Road, Red Bank, NJ 07701 • (732) 741-3900
441 East State St., Trenton, NJ 08608 • (609) 695-3900
www.ghclaw.com

copyright protection. The contents of the written report or the software containing the model, however, may very well be susceptible of copyright protection. Accordingly, it is important to differentiate between the model itself and the written deliverables or the software used to implement the model.

Trade Secret. In most cases, the true value of a statistical model is that no one else knows it. In other words, it is generally protected as a trade secret. Since the law recognizes trade secrets as a property right (similar to a right to ownership of tangible property), trade secret is an effective means of protecting statistical models while being able to exploit their value.

Who owns the model?

The better question is: Who should own which pieces of the model?

In order to effectively and fairly determine ownership issues in a statistical model development agreement, it is important to remember the component parts of the deliverables and the role of the developer. The deliverables in any such agreement will consist of:

- (1) the model itself (composed of statistical relationships or variables and the weighting coefficients);
- (2) the written report describing the data analysis and the model itself, among other things; and
- (3) potentially, software.

The Model Itself. Since the developer is in the business of building statistical models, the developer should be unwilling to assign all of its interest in the model. To do so may preclude the developer from re-using the statistical relationships found within the data. Conversely, the company for whom the model is developed will want to make sure that the developer does not give the model to competitors. A reasonable compromise in this situation is for the model to be licensed to the company for whom it is developed, provided that the composition of the model, taken as a whole and, in particular, those specially derived weighting coefficients, together constitute trade secrets of the company for whom the model is developed. From the developer's perspective, the only potential exceptions to this rule would be to ensure that the developer is free to re-use statistical relationships contained in the model and that the developer is free to independently develop even an identical model from data provided by another company for that purpose.

Written Report. Model developers may frequently re-use substantial portions of the written material provided to recipients of their models. Accordingly, developers may assert copyright protection in much of the content in the report.

GIORDANO, HALLERAN & CIESLA

A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW

125 Half Mile Road, Red Bank, NJ 07701 • (732) 741-3900
441 East State St., Trenton, NJ 08608 • (609) 695-3900
www.ghclaw.com

Software. The issues with regard to software which implements a statistical model are no different from those associated with any other custom software development agreement. The developer will want to be able to reuse much of the code for subsequent customers. Accordingly, such software should be licensed by the developer to the company for whom the model is developed.

Conclusion

Too often, lack of understanding gives way to the approach of parties wanting to “own it all” because they don’t know what is important for them to own. This frustrates companies trying to procure statistical models and reduces the sales cycle times of developers. Understanding the fundamentals of statistical model development is essential to avoiding unnecessary conflicts in negotiating statistical model development agreements.

GIORDANO, HALLERAN & CIESLA

A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW

125 Half Mile Road, Red Bank, NJ 07701 • (732) 741-3900
441 East State St., Trenton, NJ 08608 • (609) 695-3900

www.ghclaw.com