9B17M178

# CREDIT GUARANTEE CORPORATION: ACCOMMODATING AN EXPANSION STRATEGY—NOTE

*Tuhin Sengupta and Shrestha Pratik wrote this technical note as an aid to instructors in the classroom use of the case Credit Guarantee Corporation: Accommodating an Expansion Strategy, No. 9B17M177. This technical note should not be used in any way that would prejudice the future use of the case.*



This technical note provides further explanation on the following important aspects of the teaching note analysis:

- Statistical models
  - Scorecard model
  - Decision tree
  - Random forest model
  - Neural network model
- Standard model (Basel II framework)
- Rete algorithm

Although these concepts will not be dealt with in detail, this note gives a glimpse into their inner workings. The following list of readings will help readers understand these concepts.

## RELEVANT READINGS

- Iain Brown and Christophe Mues, "An Experimental Comparison of Classification Algorithms for Imbalanced Credit Scoring Data Sets," *Expert Systems with Applications* 39, no. 3 (2012): 3446–3453.
- Jean-Paul Decamps, Jean-Charles Rochet, and Benoît Roger, "The Three Pillars of Basel II: Optimizing the Mix," *Journal of Financial Intermediation* 13, no. 2 (2004): 132–155.
- Charles L. Forgy, "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem," *Artificial Intelligence* 19, no. 1 (1982): 17–37.
- Ernest Friedman-Hill, *Jess in Action: Rule-Based Systems in Java* (Greenwich, CT: Manning Publications, 2003).

- Anil K. Jain and J. V. Moreau, "Bootstrap Technique in Cluster Analysis," *Pattern Recognition* 20, no. 5 (1987): 547–568.
- James Taylor, *Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics* (Boston, MA: Pearson Education, 2012).
- David West, "Neural Network Credit Scoring Models," *Computers & Operations Research* 27, no. 11 (2000): 1131–1152.


## STATISTICAL MODELS

A statistical model is a way to statistically find the characteristics that will most likely result in a final outcome. For example, suppose you want to predict the height of primary class students five years from now. A few characteristics to consider are the students' present height, diet, and gender. These parameters form the characteristics of the model.

To make any statistical model accurate and predictive, two key inputs are required:

**Historical Data**—With accurate and diverse historical data, one can predict the future with greater certainty. Consider the example where you want to predict the future height of a primary class student. Suppose you have a lot of historical data but the data are only about girls (female gender). The model might be highly accurate for female students, but the predictions from the data will be incorrect for male students. Hence, a diverse data set is important. It is also essential to have accurate data, as incorrect data can provide only incorrect predictions.

**Domain Expertise**—Domain expertise is a key aspect of making a highly accurate model. Continuing the previous example, suppose you have found that a student who eats a particular kind of sweet is more likely to grow taller than one who does not. For a non-expert in the field, this characteristic could mean there is a correlation between eating the sweet and the student's future height. However, a domain expert might disagree on the same result. The expert might be able to shed more light on this characteristic by noting that the students who have this sweet all go to a sports club where they are given intense athletic training. Hence, a strong correlation can seem to be irrelevant data to a domain expert.

The following four models are the most widely used in the financial and banking industry to analyze risk for any asset.

### Scorecard Model

A scorecard model is a predictive model, where we try to predict the future by getting a quantitative measure of each characteristic. This is done by dividing each characteristic into different ranges and assigning different scores to these ranges. This assignment is carried out after conducting data analysis and data crunching using software such as R and SAS. Special statistical and machine learning methods such as K-means clustering and bootstrap validation are used to get these scores. A final scorecard model will have the most important characteristics for different ranges along with their respective partial scores (as it is only for one characteristic). The sum of these partial scores forms a final score, used to assess the outcome. Returning to the example of predicting students' height, assume that you found three important characteristics, as shown in Figure 1 below.

- Arm Span: What is the length of the stretch of the arm?

- Present Height: What is the present height of the student?
- Gender: Is the student male or female?

**Figure 1: Scorecard Model for Predicting Future Height**

| Predicting Height Scorecard | | | |
|---|---|---|---|
| **Bins** | **Range** | **Description** | **Score** |
| **Arm Span (a)** | | | |
| Small | 0 < = a < 0.5 | Less than 0.5 m | 20 |
| Medium | 0.5 < = a < 1 | Between 0.5 m and 1.0 m | 35 |
| Large | a > = 1 | Greater than 1 m | 45 |
| **Present Height (b)** | | | |
| Short | 0 < = b < 1 | Less than 1.0 m | 25 |
| Average | 1 < = b < 1.5 | Between 1.0 m and 1.5 m | 30 |
| Tall | b > = 1.5 | Greater than 1.5 m | 35 |
| **Female** | | | |
| Female | TRUE | FEMALE | 15 |
| Male | FALSE | MALE | 32 |

Source: Created by the authors.

As the figure shows, if the arm span for a particular student is between 0.5 and 1.0 metres (m), the partial score for the arm span characteristic is 35. Similarly, if the current height of the student is less than 1 m, the partial score for the present height characteristic is 25, and so on. For the final characteristic, assume that the student is a girl; her partial score would be 30.

Summing the scores from all the characteristics gives a total score, which is indicative of the future height of the student. In the previous example, the total score for the student is 35 + 25 + 30 = 90, as found from the height-predicting score model. This score in itself is not significant, but if we compare it with another student's scores, we can tell if this student will grow taller than the other student.

## Decision Tree

A decision tree, unlike a score model, is a graphical representation of a predictive model using historical data. Here, each characteristic is a node on the branch of a tree, and the result can be found at the leaf (node with no further branching) of the tree. Unlike a score at the end, as in a scorecard, we get an actual predicting value.

Take the same example of finding the future height of students with the same three characteristics, and create a decision tree like the one shown in Figure 2. For example, consider a child who is female, with an arm span of 0.75 m (less than 0.80 m) and present height of 1.35 m (between 1.00 and 1.50 m). Her future height (five years from now), as predicted by the model, would be 1.60 m.