

Chapter 9

Conceptual Approach to Predict Loan Defaults Using Decision Trees

Syed Muzamil Basha
VIT University, India

Dharmendra Singh Rajput
VIT University, India

N. Ch. S. N. Iyengar
Sreenidhi Institute of Science and Technology, India

ABSTRACT

In this chapter, the authors show how to build a decision tree from given real-time data. They interpret the output of decision tree by learning decision tree classifier using really recursive greedy algorithm. Feature selection is made based on classification error using the algorithm called feature split selection algorithm (FSSA), with all different possible stopping conditions for splitting. The authors perform prediction with decision trees using decision tree prediction algorithm (DTPA), followed by multiclass predictions and their probabilities. Finally, they perform splitting procedure on real continuous value input using threshold split selection algorithm (TSSA).

INTRODUCTION

Readers want to understand the pitfalls in handling real time financial data, in which Decision trees are extremely useful in representing the real time scenario of finance industry. Our aim in this chapter is to explore that domain, in particular during evaluating loan applications. Consider, a scenario where an individual needs to buy a house, to which they may not have all the money, where there is a need to take a loan from the bank. The bank is going to look at some properties of one's history, like:

1. Credit

DOI: 10.4018/978-1-5225-4999-4.ch009

Figure 1. Intelligent system



2. What has it been like in the past?
3. How much money does He/She actually make?
4. How long His/Her loan is off?
5. What time He/She is willing to spend before paying it off ?
6. Personal information about him, like gender, age and so on.

The bank will collect the information and try to make a prediction, seeing if sanctioning a loan to the individual is a risky factor or not. Typically, Industry experts look at credit history, which looks at all the other loans taken in the past and credit cards and all that and have paid those off on time.

This chapter is organized as follows: First, we are learning how to build a decision tree from data, followed by Interpreting the output of decision tree. Next, we are learning decision tree classifiers using really recursive greedy algorithm, followed by selection of the best feature to split, based on classification error called feature split selection algorithm, with all possible stopping condition for splitting. Next, we are learning how to perform predictions with decision trees using decision tree prediction algorithm, followed by multiclass predictions and their probabilities and finally, performing splits on real continuous value input using threshold split selection algorithm.

LITERATURE REVIEW

In a direction to understand the amount of research carried in applying the decision tree in predicting the loan risk factor, Mistikoglu et al., (2015) adopted C5.0 and CHAID algorithms to construct decision trees and to extract rules that show the associations between the input and output variables for roofer fall accidents. Important input variables of this models were identified as the fall distance, injury cause, safety training, and construction operation prompting fall, considering that these factors had the best predictive value related to whether a roofer fall accident can result in a fatality (or) nonfatal injury. Bahnsen et al. (2015) proposed an example dependent cost sensitive decision tree algorithm, by incorporating the different example dependent costs into a new cost based impurity measure and a new cost based pruning criteria. The study used three different databases, from three real-world applications as credit

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/conceptual-approach-to-predict-loan-defaults-using-decision-trees/210968

Related Content

Evidence Sources, Methods and Use Cases for Learning Lightweight Domain Ontologies

Albert Weichselbraun, Gerhard Wohlgenannt and Arno Scharl (2011). *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances* (pp. 1-15).

www.irma-international.org/chapter/evidence-sources-methods-use-cases/53878

How to Capitalize Knowledge within Online Communities: An Approach Based on the SECI Model and an Empirical Method of Questioning

Lamia Berkani and Azeddine Chikh (2015). *Knowledge Management for Competitive Advantage During Economic Crisis* (pp. 194-207).

www.irma-international.org/chapter/how-to-capitalize-knowledge-within-online-communities/117849

Intellectual Capital Measurement and Reporting Models

Jamal A. Nazari (2015). *Knowledge Management for Competitive Advantage During Economic Crisis* (pp. 117-139).

www.irma-international.org/chapter/intellectual-capital-measurement-and-reporting-models/117846

Conceptual Approach to Predict Loan Defaults Using Decision Trees

Syed Muzamil Basha, Dharmendra Singh Rajput and N. Ch. S. N. Iyengar (2019). *Sentiment Analysis and Knowledge Discovery in Contemporary Business* (pp. 148-161).

www.irma-international.org/chapter/conceptual-approach-to-predict-loan-defaults-using-decision-trees/210968

Designing Component-Based Heuristic Search Engines for Knowledge Discovery

Craig M. Howard (2002). *Heuristic and Optimization for Knowledge Discovery* (pp. 68-88).

www.irma-international.org/chapter/designing-component-based-heuristic-search/22150