Chapter 9
Predicting Company Bankruptcy Using Machine Learning Techniques: A Step-by-Step Guide

Eunjung Lee
Lewis University, USA

ABSTRACT

In today's business landscape, predicting the financial health of a company is essential for informed decision-making by investors, creditors, and other stakeholders. Using historical financial data and machine learning techniques, it is now possible to predict the likelihood of a company going bankrupt. This chapter provides a step-by-step guide on how to predict company bankruptcy using the "company bankruptcy prediction" dataset available on UCI machine learning repository. The chapter covers data analysis, pre-processing, applying various machine learning algorithms, evaluating model performance, and applying models to new datasets. The aim is to equip readers with the necessary skills to analyze and predict the financial health of companies, making it a valuable resource for investors, creditors, and financial analysts.

1. INTRODUCTION

In today's rapidly changing business landscape, predicting the financial health of a company is crucial for investors, creditors, and other stakeholders to make informed decisions. With the increasing availability of data and advancements in machine learning techniques, it is now possible to predict the likelihood of a company going bankrupt using historical financial data. Bankruptcies can have ripple effects on the economy, and predicting bankruptcies can help prevent these negative impacts.

This chapter provides a comprehensive guide on how to predict company bankruptcy using machine learning techniques, using the "Company Bankruptcy Prediction" dataset available on UCI Machine Learning Repository. Readers will learn how to analyze and pre-process the data, apply various machine

DOI: 10.4018/978-1-6684-8386-2.ch009

learning algorithms, evaluate model performance, and apply models to new datasets. By the end of this chapter, readers will have the necessary skills to analyze and predict the financial health of companies, making it a valuable resource for investors, creditors, and financial analysts.

Machine learning involves using algorithms to identify patterns and make predictions based on data. This chapter will demonstrate how to use Weka, a free analytical software, to predict the likelihood of a company going bankrupt based on financial data. The chapter will also provide an overview of the performance evaluation metrics used to assess the accuracy of the models.

To achieve this, we will first define the research questions and metrics that will be used to predict the financial health of a company. We will then collect and analyze the "Company Bankruptcy Prediction" dataset, which contains financial data from a variety of companies. Finally, we will apply machine learning techniques to the data to predict the likelihood of a company going bankrupt. This chapter is aimed at individuals with a basic understanding of machine learning methodology and evaluation methods. Theoretical explanations of each methodology will be kept to a minimum, with a focus on demonstrating practical experiments on how to analyze and interpret data results.

2. PREDICTING COMPANY BANKRUPTCY: PRACTICE USING WEKA

2.1 Define Business Questions

To predict company bankruptcy, we first need to define the business questions we want to answer. As investors or financial analysts, we may be interested in understanding the factors that impact the likelihood of a company going bankrupt and whether we can use financial data to predict the bankruptcy risk of a specific company. Some potential business questions that we can explore include:

What financial ratios and metrics are most strongly associated with company bankruptcies?
Can we use historical financial data to predict the probability of a company going bankrupt within the next year?
How accurately can we predict the bankruptcy risk of a specific company using machine learning techniques?
Which machine learning algorithms perform best for predicting company bankruptcies?

By answering these business questions, we can gain insights into the financial health of companies and make informed decisions as investors or financial analysts. In the next section, we will explore the "Company Bankruptcy Prediction" dataset and analyze the data to gain a better understanding of the factors that impact bankruptcy risk.

2.2 Collect Data

To begin with, we use a real-world dataset from the Taiwan Economic Journal for the years 1999-2009[1]. We need to collect the "Company Bankruptcy Prediction" dataset from the UCI Machine Learning Repository[2]. This dataset consists of financial ratios and other financial information for various companies. We will be using this dataset to predict the likelihood of a company going bankrupt. The "Company Bankruptcy Prediction" dataset contains 95 financial ratios (Figure 24 in the Appendix) and other fi-