

Dự đoán phá sản sử dụng Lựa chọn đặc trưng bằng Giải thuật di truyền - Máy vector hỗ trợ (GA-SVM) và Stacking

Ngày 22 tháng 4 năm 2024

Ngày công bố:	19/7/2023
Tác giả:	Wiena Faqih Abror, Alamsyah, Muhammad Aziz
Nguồn:	SHM Publisher
Nguồn dữ liệu (nếu có):	
Từ khóa:	Bankruptcy, Genetic algorithm, Support vector machine, Stacking
Người tóm tắt:	Hoàng Văn Nhân

1. Mở đầu:

- Đổ vỡ doanh nghiệp gây rối loạn nền kinh tế toàn cầu và có thể gây thiệt hại đáng kể cho các công ty, nhà đầu tư và chính phủ. Dự đoán phá sản cho phép các bên liên quan thực hiện các biện pháp phòng ngừa. ‘
- Nghiên cứu này đề xuất một mô hình học máy mới để dự đoán phá sản, tận dụng GA-SVM để chọn lọc đặc trưng và học tập ensemble Stacking để có khả năng đạt được độ chính xác cao hơn so với các phương pháp truyền thống.

2. Tổng quan về các nghiên cứu trước đó

- Phương pháp truyền thống: Phân tích tỷ lệ tài chính và mô hình thống kê là các phương pháp phổ biến để dự đoán phá sản.
- Giới hạn của các nghiên cứu trước đây: Các phương pháp truyền thống có thể có vấn đề về độ chính xác và gặp khó khăn với dữ liệu phức tạp.

3. Các vấn đề còn tồn đọng:

Độ chính xác của các phương pháp dự đoán phá sản trước kia có thể bị hạn chế.

4. Mục tiêu và ý nghĩa của nghiên cứu:

- Mục tiêu nghiên cứu: Nghiên cứu này sử dụng một bộ dữ liệu từ Tạp chí Kinh tế Đài Loan với 6.819 mẫu để đào tạo bằng thuật toán học máy sử dụng các kỹ thuật phân loại. Mục tiêu đạt được từ nghiên cứu là có được một kỹ thuật phân loại với kết quả độ chính xác tốt nhất.
- Nghiên cứu này có thể cải thiện đáng kể độ chính xác của việc dự đoán phá sản so với các phương pháp hiện có.

- Dự đoán chính xác hơn có thể mang lại lợi ích cho các nhà đầu tư, chủ nợ và nhà hoạch định chính sách bằng cách cho phép họ đưa ra các quyết định sáng suốt và quản lý rủi ro tài chính hiệu quả.

5. Câu hỏi và giả thuyết của nghiên cứu:

Trong nghiên cứu, các tác giả tin rằng GA-SVM có thể xác định các đặc trưng có liên quan nhất để dự đoán phá sản và Stacking có thể kết hợp sức mạnh của nhiều bộ phân loại khác nhau để đạt được độ chính xác cao hơn.

6. Phạm vi nghiên cứu:

- Giới hạn của nghiên cứu: Nghiên cứu tập trung vào việc áp dụng phương pháp GA-SVM và Stacking cho một tập dữ liệu cụ thể của các công ty niêm yết tại Đài Loan.
- Các yếu tố được bao gồm/loại trừ: Nghiên cứu xem xét dữ liệu tài chính của các công ty nhưng không bao gồm các yếu tố liên quan khác như xu hướng ngành hoặc điều kiện kinh tế.
- Tiềm năng phát triển trong tương lai: Phương pháp này có khả năng được áp dụng cho các tập dữ liệu rộng hơn và có khả năng kết hợp thêm các nguồn dữ liệu khác để đạt được độ chính xác dự đoán thậm chí còn tốt hơn.

7. Phương pháp thử nghiệm và kết quả của nghiên cứu:

- Phương pháp được sử dụng:
Nghiên cứu sử dụng phương pháp ba giai đoạn:
 - Tiền xử lý: Giải quyết mất cân bằng dữ liệu bằng Kỹ thuật lấy mẫu quá mức lớp thiểu số tổng hợp (SMOTE).
 - Chọn lọc đặc trưng: Áp dụng GA-SVM để xác định các đặc trưng có liên quan nhất cho việc dự đoán giúp giảm số đặc trưng cần sử dụng từ 96 xuống còn 44.
 - Huấn luyện dữ liệu: Huấn luyện hai mô hình:
 - * Bộ phân loại (classifier) đơn: k-Nearest Neighbors, Naïve Bayes, Cây quyết định, Cây quyết định sử dụng Gradient Boosting, Light Gradient Boosting Machines.
 - * Stacking Ensemble: Sử dụng các classifier ở trên làm bộ học cơ sở và XGBoost làm bộ học meta.

Both models are validated using confusion matrix and k-fold cross-validation. Cả hai mô hình đều được kiểm định bằng confusion matrix và k-fold cross-validation.
- Tính phù hợp của các phương pháp: Các phương pháp được chọn phù hợp với mục tiêu nghiên cứu. GA-SVM có khả năng chọn các đặc trưng có liên quan và Stacking có thể tận dụng sức mạnh của nhiều phân loại khác nhau.
- Kết quả và tính phù hợp của chúng: Nghiên cứu đạt được độ chính xác cao 99,22% bằng cách sử dụng mô hình Stacking ensemble, cho thấy phương pháp này hiệu quả cho tập dữ liệu này.

8. Tổng quan bài nghiên cứu:

- Nghiên cứu này cho thấy tiềm năng của việc sử dụng kết hợp GA-SVM cho việc chọn lọc đặc trưng và phương pháp học tập ensemble Stacking cho việc đạt được dự đoán phá sản có độ chính xác cao.

- Nghiên cứu này đưa ra một cách tiếp cận đầy hứa hẹn để dự đoán phá sản. Độ chính xác cao cho thấy GA-SVM và Stacking có thể là các công cụ có giá trị để phân tích tài chính. Tuy nhiên, cần nghiên cứu thêm để đánh giá tính tổng quát của phương pháp này bằng cách sử dụng bộ dữ liệu rộng hơn và kết hợp dữ liệu tài chính toàn diện hơn.

AI VIET NAM – RESEARCH TEAM

Bankruptcy Prediction Using Genetic Algorithm-Support Vector Machine (GA-SVM) Feature Selection and Stacking

April 22, 2024

Date of publication:	19/07/2023
Authors:	Wiena Faqih Abror, Alamsyah, Muhammad Aziz
Sources:	SHM Publisher
Data sources (if any):	
Keywords:	Bankruptcy, Genetic algorithm, Support vector machine, Stacking
Summary by:	Hoàng Văn Nhân

1. Introduction:

- Bankruptcy disrupts the global economy and can cause significant losses for companies, investors, and governments. Predicting bankruptcy allows stakeholders to take preventive measures.
- This research proposes a new machine learning model for bankruptcy prediction that leverages GA-SVM for feature selection and Stacking ensemble learning to potentially achieve higher accuracy compared to traditional methods.

2. Previous studies:

- Traditional methods: Financial ratio analysis and statistical models are prevalent methods for bankruptcy prediction.
- Limitations of previous studies: Traditional methods have accuracy issues and struggle with complex data.

3. Unresolved issues:

The accuracy of previous bankruptcy prediction methods might be limited.

4. Objectives and Significance:

- Research objective: In the research that was carried out utilizing a data set from the Taiwan Economic Journal as many as 6,819 to be trained using machine learning algorithms using classification techniques. The goal obtained from the research conducted is to obtain a classification technique with the best accuracy results.
- This research could significantly improve the accuracy of bankruptcy prediction compared to existing methods.

- More accurate predictions can benefit investors, creditors, and policymakers by enabling them to make informed decisions and manage financial risks effectively.

5. Hypothesis:

In the research, the authors believe that GA-SVM can identify the most relevant features for bankruptcy prediction, and Stacking can combine the strengths of multiple classifiers to achieve higher accuracy.

6. Scope:

- Limitations of the research:
The research focuses on applying the GA-SVM and Stacking approach to a specific dataset of listed companies from Taiwan.
- Factors included/excluded:
The study considers financial data from the companies but not include other relevant factors like industry trends or economic conditions.
- Potential for future development:
The approach has the potential to be applied to broader datasets and potentially incorporate additional data sources for even better prediction accuracy.

7. Experimental method and research result:

- Method used: The research uses a three-stage approach:
 - Preprocessing: Addresses data imbalance using Synthetic Minority Over-Sampling Technique (SMOTE).
 - Feature Selection: Applying GA-SVM to identify the most relevant features for prediction helps reduce the number of features needed from 96 to 44.
 - Data Training: Trains two models:
 - * Single Classifiers: k-Nearest Neighbors, Naïve Bayes, Decision Trees, Gradient Boosting Trees, Light Gradient Boosting Machines.
 - * Stacking Ensemble: Uses the above classifiers as base learners and XGBoost as a meta-learner.

Both models are validated using confusion matrix and k-fold cross-validation.
- Suitability of methods: The chosen methods are suitable for the research objective. GA-SVM can potentially select relevant features, and Stacking can leverage the strengths of multiple classifiers.
- Results and their suitability: The research achieved a high accuracy of 99.22% using the Stacking ensemble model, suggesting the approach is effective for this dataset.

8. Overview:

- This research demonstrates the potential of using a combination of GA-SVM for feature selection and Stacking ensemble learning to achieve highly accurate bankruptcy prediction.
- This research offers a promising approach for bankruptcy prediction. The high accuracy suggests that GA-SVM and Stacking could be valuable tools for financial analysis. However, further research is needed to evaluate the generalizability of this approach using broader datasets and incorporating more comprehensive financial data.