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.