Bankrupt Classifier

Abdalla Roshdy 20108691
Mahmoud Abdelhady 211008944
Moiad Abdelhamed 211013046
Qasem Hamada 211008579
Presented to: DR / Nashwa EL-Bendary

Introduction

This project investigates how feature selection using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) impacts the performance of machine learning models. By comparing model accuracy, precision, recall, and F1-score across three scenarios—full feature set, GA-selected features, and PSO-selected features—we evaluate whether evolutionary algorithms enhance predictive performance while reducing dimensionality. The results provide insights into optimal feature selection methods for improved efficiency and model interpretability.

Bankrupt Classifier

Dataset

The Bankruptcy dataset includes financial data with 95 features to predict if a company will go bankrupt (1) or not (0). It's used for binary classification in finance. The data comes from financial statements and helps assess credit risk and prevent financial losses.

Methodology

KNN is a classification algorithm that predicts the class of a data point based on the majority class among its K nearest neighbors in the feature space.
The model calculates distances between the new data point and all points in the training set, then selects the K closest ones to make a prediction.
A smaller K value makes the model more sensitive to noise and outliers, while a larger K smooths decision boundaries and reduces variance.
Unlike other models, KNN does not perform a training phase — it stores the entire training dataset and uses it during prediction, making it computationally expensive for large datasets.
To improve performance and reduce complexity, feature selection techniques are often applied.
In this project, two metaheuristic algorithms were used: Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), to select the most relevant features for classification.
GA works by evolving binary chromosomes that represent feature subsets through selection, crossover, and mutation, aiming to maximize classification accuracy.
PSO uses particles that move through the search space to find optimal feature combinations, updating positions based on personal and global best solutions.
Both methods aim to find a compact subset of features that enhances classification performance while reducing model complexity.
After feature selection, the KNN classifier is trained and tested using only the selected features to evaluate improvements in precision, recall, F1-score, and accuracy.

Results

Dataset

	Bankrupt	Non-Bankrupt
• Percision :	0.91	0.99
• Recall :	0.99	0.91
• F1 score :	0.95	0.95
• Accuracy :	0.95%	

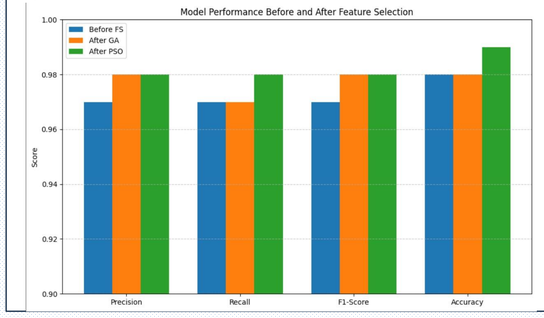
GA:

	Bankrupt	Non-Bankrupt
• Percision :	0.99	0.94
• Recall :	0.94	0.99
• F1 score :	0.96	0.96
• Accuracy :	0.96.13%	

PSO

	Bankrupt	Non-Bankrupt
• Percision :	0.99	0.96
• Recall :	0.96	0.99
• F1 score :	0.98	0.98
• Accuracy :	0.98%	

-This image show Training and Validation process :



References:

1 -https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction/data
2-https://colab.research.google.com/drive/1HMTjK_T_dnYXEPBQYMZfpbDhK3qozK0k?usp=sharing
3-Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study - ScienceDirect