

TẬP DỮ LIỆU PHỤC VỤ CHO BÀI TOÁN PHÂN LỚP

1. Dữ liệu

Dữ liệu trong file Bankruptcy.csv, được upload trên github, thư mục “w05”

Bộ dữ liệu có 250 hàng và 7 cột. Bộ dữ liệu này chứa 6 tham số định tính từ các chuyên gia có thể được sử dụng để dự đoán khả năng phá sản.

2. Từ điển dữ liệu

Column Position	Attribute Name	Definition	Data Type	Example	% Null Ratios
1	Industrial Risk	Industrial Risk (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
2	Management Risk	Management Risk (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
3	Financial Flexibility	Financial Flexibility (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
4	Credibility	Credibility (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
5	Competitiveness	Competitiveness (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
6	Operating Risk	Operating Risk (P: Positive, A: Average, N: Negative)	Qualitative	P, A, N	0
7	Class	Class (B: Bankruptcy, NB: Non-Bankruptcy)	Qualitative	B, NB	0

3. Mô tả các cột

Sau đây là chi tiết về các yếu tố rủi ro định tính đề cập đến các thành phần rủi ro của chúng:

Industry risk (IR): Government policies and International agreements, Cyclicity, Degree of competition, The price and stability of market supply, The size and growth of market demand, The sensitivity to changes in macroeconomic factors, Domestic and international competitive power, Product Life Cycle.

Management risk(MR): Ability and competence of management, Stability of management, The relationship between management/ owner, Human resources management, Growth process/business performance, Short and long term business planning, achievement and feasibility.

Financial Flexibility(FF): Direct financing, Indirect financing, Other financing

Credibility (CR): Credit history, reliability of information, The relationship with financial institutes.

Competitiveness (CO): Market position, The level of core capacities, Differentiated strategy.

Operating Risk (OP): The stability and diversity of procurement, The stability of transaction, The efficiency of production, The prospects for demand for product and service, Sales diversification, Sales price and settlement condition, Collection of A/R, Effectiveness of sale network.

4. Nhiệm vụ:

- Exploratory data analysis
- Data visualization
- **Classification modelling techniques (nhiệm vụ chính)**