**Report on Case Study on Bankruptcy**

**1. Data Understanding and Preprocessing:**

***o Load and inspect the dataset.***

Given Bankruptcy dataset was loaded for analysis and inspected

It was observed that dataset has 6819 rows and 96 columns.

There were no duplicates

There were no null values

First 5 and last 5 rows of data were also seen to get feel of the data

**o Handle missing values appropriately**.

There were hardly any missing values.

***o Detect and handle outliers.***

Some outliers were seen but were not removed ,Boxplot diagram for ‘Net Income to Total Assets’ was drawn using IQR.

2. Exploratory Data Analysis (EDA):

o Generate descriptive statistics.

Descriptive statistics for each column was generated to get summary of the central tendency, dispersion, and shape of the dataset's distribution.

o Visualize feature distributions and target variable.

To visualize features distribution heatmap, corelation diagrams were drawn

o Analyze relationships between features and the target variable.

It was observed that ‘Networth/Assets’, ‘Net income to total assets’,’ Debt Ratio’,’Liability to equity’ and ‘borrower dependency’ affect target variable as can be seen in chart positive and negative correlation is seen

Hypothesis Testing:

o Perform hypothesis testing to identify significant features influencing bankruptcy.

4. Feature Engineering and Selection:

o Create new features to enhance predictive power.

o Use dimensionality reduction techniques if necessary.

Dimensions were reduced and only ' Debt ratio %', ' Net worth/Assets', ' Borrowing dependency' were used to check effect on ‘bankruptcy’

o Select relevant features using methods like RFE or feature importance.

Above selected were important feature as per corelation matrix.

5. Modeling:

o Split the data into training and testing sets.

Data was split into training and test data

o Apply Logistic Regression for classification.

Logistic regression was apllied

o Evaluate model performance using appropriate metrics.

Logit Regression Results

==============================================================================

Dep. Variable: Bankrupt? No. Observations: 5455

Model: Logit Df Residuals: 5451

Method: MLE Df Model: 3

Date: Mon, 17 Jun 2024 Pseudo R-squ.: 0.2028

Time: 15:35:39 Log-Likelihood: -603.48

converged: True LL-Null: -756.97

Covariance Type: nonrobust LLR p-value: 3.081e-66

=========================================================================================

coef std err z P>|z| [0.025 0.975]

-----------------------------------------------------------------------------------------

Debt ratio % 14.4594 nan nan nan nan nan

Net worth/Assets -11.1595 nan nan nan nan nan

Borrowing dependency 1.7512 2.015 0.869 0.385 -2.198 5.700

const 3.2999 nan nan nan nan nan

6. Model Interpretation and Insights:

o Interpret model coefficients to understand feature impact.

Debt ratio has positive corelation with bankruptcy and Networth/Assets have negative corelation.

o Summarize key insights and discuss implications for financial risk management.

Borrowing dependency has positive corelation but as p value is much greater than 0.05 the effect is not significant

o Provide actionable recommendations.

Evaluating bankruptcy risk requires a comprehensive analysis of industry dynamics, financial statements, market conditions, and management strategies. Therefore, while there may be instances where bankruptcy and some of the ratios show a positive correlation, this relationship is nuanced and context-dependent and all the factors need to be balanced for avoiding bankruptcy and running business .