**Project: Bankruptcy Prevention**

**Problem Statement:**

The objective of this project is to create a classification model that predicts the chances of a business facing bankruptcy based on the key feature like Industrial Risk, Management Risk, Financial Flexibility, Credibility, Competitiveness, Operating Risk.

With the help of classification, we will try to predict accurately and give the business an early chance to know about the risk of going bankrupt.

**Steps to follow in project:**

1. **Data Cleaning and Exploratory Data Analysis (EDA):**
2. **Load the Dataset:**

Use pandas to load dataset, we can use (pd.read \_csv),(pd.read\_excel).

1. **Inspect the Data:**

Display the first few rows to get an overview of the data using

head(). Check for missing values using isnull().sum().

1. **Handle Missing Values:**

Handle or remove missing values based on the nature of the data.

Consider using techniques like mean imputation or advanced imputation methods.

1. **Checking Duplicates Values:**

Duplicated records; handling duplicates.

1. **Explore Data Distributions:**

Visualize the distributions of numerical variables using histograms, box plots, or kernel density plots. Utilize seaborn and matplotlib for creating visualizations.

1. **Analyze Relationships Between Variables:**

Use scatter plots, pair plots, or correlation matrices to explore relationships between numerical variables. Consider bar plots or count plots to analyze relationships involving categorical variables.

1. **Data Visualization**
2. Count Plot
3. Pie Plot
4. Bar Plot for relationship between target features and independent features
5. Correlation and Heat Map
6. **Feature engineering**
   1. Label Encoding or One Hot Encoding
   2. Handling Outliers
   3. PPS Score
   4. Features Scaling

* Normalization
* Standardization
  1. Transformations
  2. Feature Selection
  + Uni-Variate Selection
  + **Recursive Feature Eliminator**
  1. Feature Creation if required.

1. **Model Validation**

a. Train and Test Split

b. K-Fold Cross Validation

1. **Model building**
   1. Logistic Regression
   2. Decision Tree
   3. K-Nearest Neighbors
   4. Support Vector Machines
   5. Naïve Bayes
   6. Random Forest
2. **Hyperparameter tuning**
   1. Using GridSearchCV , if required
3. **Model evaluation**
   1. Classification Report
   2. Confusion Matrix
   3. Accuracy Score
   4. F1-Score
4. **Model deployment**

* Using Streamlitdeployment python script which contains best model