**CREDIT RISK ANALYZER**

**WIT­H AI**

**Submitted By**

**ARUNKUMAR A**

**Introduction**

**Credit Analysis**

Credit analysis is a type of analysis an investor or bond portfolio manager performs on companies or other debt issuing entities to measure the entity's ability to meet its debt obligations. The credit analysis seeks to identify the appropriate level of default risk associated with investing in that particular entity.

### **KEY TAKEAWAYS**

* Credit analysis is a type of analysis an investor or bond portfolio manager performs on companies or other debt issuing entities to measure the entity's ability to meet its debt obligations.
* The credit analysis seeks to identify the appropriate level of default risk associated with investing in that particular entity.
* The outcome of the credit analysis will determine what risk rating to assign the debt issuer or borrower.

## **How Credit Analysis Works**

To judge a company’s ability to pay its debt, banks, bond investors, and analysts conduct credit analysis on the company. Using financial ratios, cash flow analysis, trend analysis, and financial projections, an analyst can evaluate a firm’s ability to pay its obligations. A review of credit scores and any collateral is also used to calculate the creditworthiness of a business.

Not only is the credit analysis used to predict the probability of a borrower defaulting on its debt, but it's also used to assess how severe the losses will be in the event of default.

The outcome of the credit analysis will determine what risk rating to assign the debt issuer or borrower. The risk rating, in turn, determines whether to extend credit or loan money to the borrowing entity and, if so, the amount to lend.

## **Credit Analysis Example**

An example of a financial ratio used in credit analysis is the debt service coverage ratio (DSCR). The DSCR is a measure of the level of cash flow available to pay current debt obligations, such as interest, principal, and lease payments. A debt service coverage ratio below 1 indicates a negative cash flow.

For example, a debt service coverage ratio of 0.89 indicates that the company’s net operating income is enough to cover only 89% of its annual debt payments. In addition to fundamental factors used in credit analysis, environmental factors such as regulatory climate, competition, taxation, and globalization can also be used in combination with the fundamentals to reflect a borrower's ability to repay its debts relative to other borrowers in its industry.

## **Special Considerations**

Credit analysis is also used to estimate whether the credit rating of a bond issuer is about to change. By identifying companies that are about to experience a change in debt rating, an investor or manager can speculate on that change and possibly make a profit.

For example, assume a manager is considering buying junk bonds in a company. If the manager believes that the company's debt rating is about to improve, which is a signal of relatively lower default risk, then the manager can purchase the bond before the rating change takes place, and then sell the bond after the change in rating at a higher price.

**What type of information is critical?**

**Does everyone in the organization know what information is critical?** If identifying confidential information is still halfway , do the following:

## **1. Map the data.**

Go through the data handled in different functions. Find out who deal with confidential information and with whom information is exchanged. You should also consider what information can be assembled from small pieces into large and critical entities. Example picture helps to identify the protected data.



## **2. Identify the responsibilities and obligations.**

Data protection is regulated, among other things, by legislation, agreements, NDAs, data protection provisions, internal guidelines, users' expectations, the customer promises. What kind of obligations does your business have to protect information?

## **3. Assess the risks.**

What happens if the information is spread to outsiders? Assess the significance of the information and the effects of data leakage on your business.

## **4. Define security levels.**

Measure the security level and the necessary protection according to criticality of the information. Use instructions to deploy practices to all personnel.

Identification of confidential information is the starting point for the planning of information security. This requires both appropriate technology and practices, and corporate culture. The company's business specific characteristics must be taken into account when planning the solutions. Usability is also essential. Make sure that the selected solutions are also in use – this is the only way the information security becomes a reality.

Some types of loans require more thorough analysis than others. Larger, long-term loans for fixed assets require more thorough analysis than short-term working capital loans. For individual loans, loan analysis and follow-up visits provide most of the guarantee for the institution and thus the analysis is necessarily more extensive. Group loans transfer most of this responsibility to the clients and therefore do not require detailed analysis.

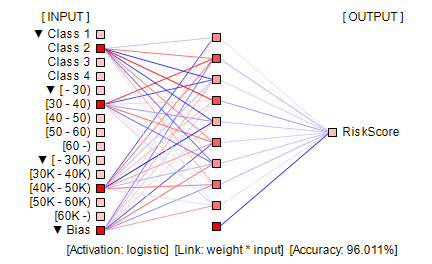
**Credit Risk Predictive Modelling and Credit Risk Predictive By Machine Learning**

If past is any guide for predicting future events, credit risk prediction by Machine Learning is an excellent technique for credit risk management. Prediction models are developed from past historical records of credit loans, containing financial, demographic, psychographic, geographic information, etc. From the past credit information, predictive models can learn patterns of different credit default/delinquency ratios, and can be used to predict risk levels of future credit loans. It is important to note that statistical process requires a substantially large number of past historical records (or customer loans) containing useful information. Useful information is something that can be a factor that differentially affects credit default/delinquency ratios.

### **Why Neural Network and Deep Learning?**

A commonly used method used in risk prediction is regression. Regression works well if information structure is functional and simple. However it does not perform well on complex information with many categorical variables. Another commonly cited method is decision tree classification. Decision tree is not suitable if dependent variables have heavy skews. Credit loan data have this skew. Decision tree can work if default/delinquency rate is about, say, 30% ~ 70%. A commonly used method to overcome this problem is the boosting method which duplicates skewed data. But duplication turns outliers into statistically significant patterns, introducing bogus patterns. It produces lots of false positive or negative predictions. This is a bad approach! In fact, all classification methods suffer from mis-classifications.

This leads Neural Network and Deep Learning based on risk-level scoring methods to be the choice for credit risk predictive modeling. The following figure shows a neural network model;



Neural network arranges information in nodes and weight-links as shown in the above figure. Nodes represent input/output values. Nodes are organized into layers: input layer, (optional) internal layers (normally a single layer as in the figure), and output layer. Input layer nodes accept input values. Values of output layer nodes and internal layer nodes are computed by summing up previous layer nodes multiplied by weight-links' values.

Neural network weight-links are computed in such a way that given input values, network produces certain output value(s) for output layer node(s). This process is called as *network training*. This is performed using past data. Neural network is a heuristic predictive system.

Bias nodes are similar to coefficients in regression. They have value 1 and tend to improve network's learning capability.

In the above chart, positive value weight-links are colored in red. Negative value weight-links are colored in blue. Colors are scaled according to absolute value ratios against the largest absolute value. Absolute value zero is colored in black. Largest absolute value is colored in pure red or blue color. The rest are scaled accordingly.

It is noted that neural network is not good at predicting unseen information. It can make very wild predictions. Thus good comprehensive training data is very important. For more on neural network.

**Objective**

To ensure that loans are made on appropriate terms to clients who can and will pay them back. What analysis is needed and what is the most efficient approach to fulfil that need is primarily determined by the type and nature of the loan.

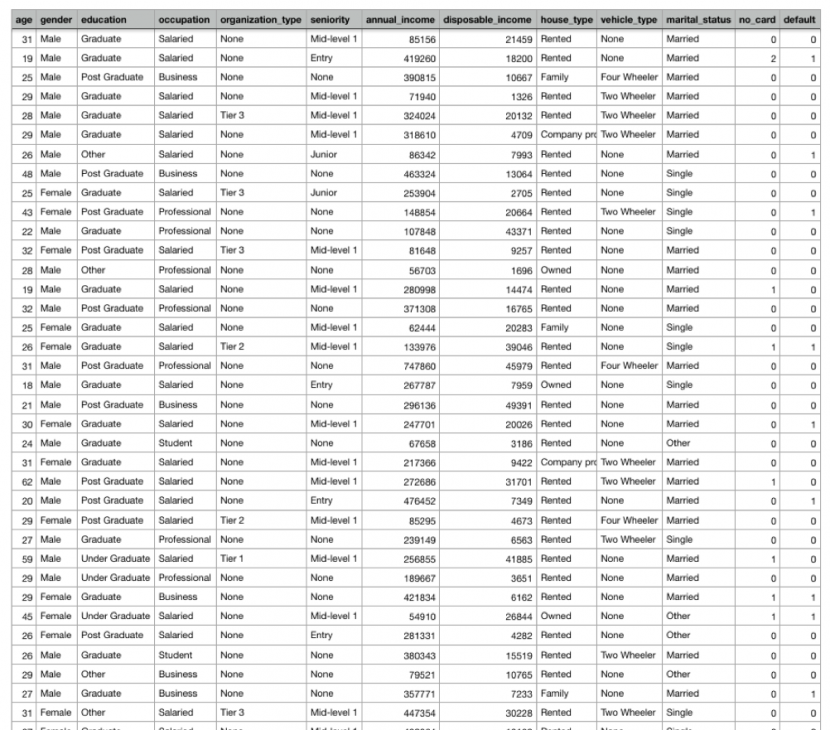
• **To place good and appropriate loans** - can the loan generate income for repayment and will the client repay

• **Determine eligibility of the applicant** - is he/she eligible according to the theprogram criteria

• **Training needs and skills** - to assess the training needs and develop the financial management skills level of the client. (This is the basic principal of programs that integrate their credit and training methodologies.)

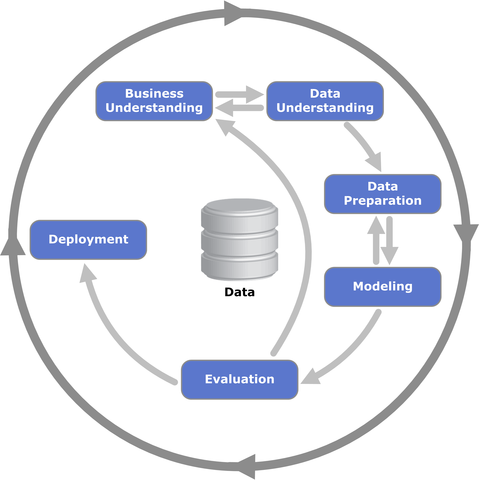
• **Program Indicators** - loan analysis may also be used to generate the indicators that will be used to evaluate the impact of the loan.

**Dataset**



**Methodology andTechniques Used**

**Cross-industry standard process for data mining**



CRISP-DM stands for cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. It is a robust and well-proven methodology. We do not claim any ownership over it. We did not invent it. We are however evangelists of its powerful practicality, its flexibility and its usefulness when using analytics to solve thorny business issues.

**Python Modules Used**

* Scikit-learn (sklearn)
  + DecisionTreeClassifier
  + KFold
  + train\_test\_split
  + accuracy\_score
* Matplotlib
* Pandas
* NumPy

**Decision Tree**

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

A decision tree is a flowchart-like structure in which each internal node represents a “test” on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

There are three commonly used impurity measures used in binary decision trees: **Entropy**, **Gini index**, and **Classification Error**.

**Entropy** (a way to measure impurity):

Entropy=−Sum (p \* log2p)

**Gini index** (a criterion to minimize the probability of misclassification):

Gini=1−Sum (p \* p)

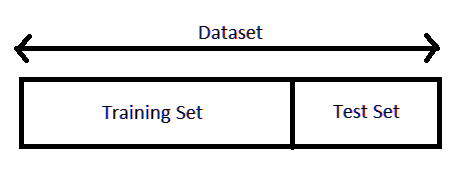
**Classification Error**:

ClassificationError=1−max(p)

where pis the probability of classes.

**Train Test Split**

As we work with datasets, a [**machine learning algorithm**](https://data-flair.training/blogs/machine-learning-algorithm/) works in two stages. We usually split the data around 20%-80% between testing and training stages. Under supervised learning, we split a dataset into a training data and test data in Python ML.



The following section will split the dataset randomly into two groups, training dataset and test dataset. We will use 70% data as training data and remaining 30% as test data.

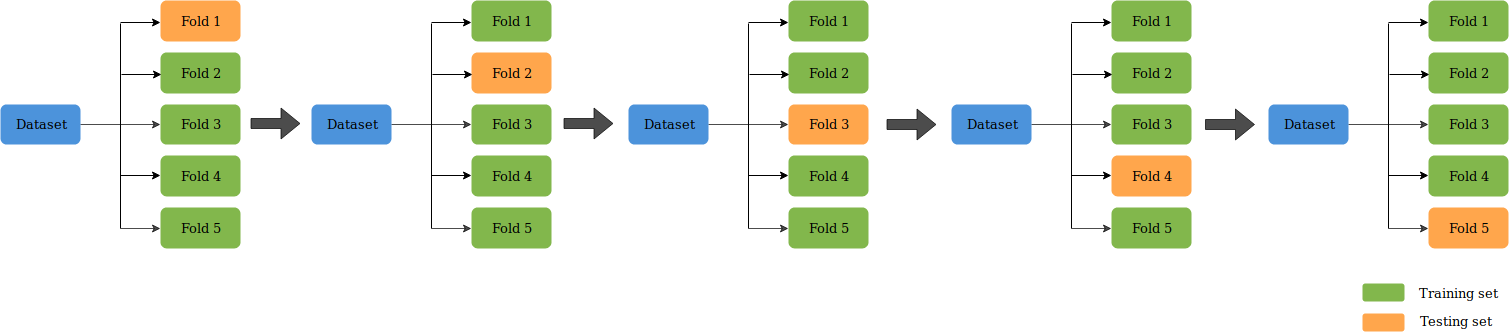
***from sklearn.model\_selection import train\_test\_split***

***x\_train, x\_test, y\_train, y\_test= train\_test\_split(X, Y, test\_size=0.3, random\_state=1234)***

**K-fold Cross Validation**

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.



In this project value of k is 4.

**Train Decision Tree Model with Gini and Entropy Criterion-**

* K-Fold (4-Fold) Cross validation to attain high accuracy
* Predict data for every fold

Picking average of scores

**Model Summary**

* 84.46% (Gini Criterion with K-Fold)
* 84.29% (Entropy Criterion with K-Fold)
* 84.08% (Gini Criterion with train\_test\_split)
* 83.92% (Entropy Criterion with train\_test\_split)