India Housing Finance offers home loans for low-income housing. They have presence across all urban, semi urban and rural areas. When customer applies for home loan, the company validates the customer eligibility for loan. They want to automate the loan eligibility process based on customer details provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History etc.

The training data provided with the problem statement gives approval status of past applications. The following features are included

|  |  |
| --- | --- |
| **Feature** | **Description** |
| Loan\_ID | Unique Loan ID |
| Gender | Male/ Female |
| Married | Applicant married (Y/N) |
| Dependents | Number of dependents |
| Education | Applicant Education (Graduate/Under Graduate) |
| Self\_Employed | Self-employed (Y/N) |
| ApplicantIncome | Applicant income |
| CoapplicantIncome | Applicant income |
| LoanAmount | Loan amount in thousands |
| Loan\_Amount\_Term | Term of loan in months |
| Credit\_History | Credit history meets guidelines |
| Property\_Area | Urban/ Semi Urban/ Rural |
| Loan\_Status | Loan approved (Y/N) - Target (Class) Variable |

Aim of the assignment is to

* Building a Predictive Model
* Evaluate the model.
* Refine the model, as appropriate

The student needs to

1. Select a method for performing the analytic task
2. Carry out descriptive summarization of data and make observations
3. Identify relevant, irrelevant attributes for building model.
4. Perform appropriate data transformations with justifications
5. Generate new features if needed
6. Carry out the chosen analytic task. Show results including intermediate results, as needed
7. Evaluate the solutions
8. Look for refinement opportunities

K-Means clustering

Clustering is the mechanism which is trying to group the similar elements together.

Grouping similar items in market. Students with same marks.

Clustering Types:

1. Hierarchical
   1. Top Down
   2. Bottom Up
2. Partitioning clustering
   1. K-Means
   2. Fuzzy CMeans

Targeted marketing

Random points are something