

# **Alma Mater Studiorum – Università di Bologna**

---

Report of the project about Bayesian Networks on Loan Approval

Fundamentals of Artificial Intelligence and Knowledge  
Representation Module – 3

Professor: Paolo Torroni

---

Ganesh Pavan Kartikeya Bharadwaj Kolluri - [ganeshpavan.kolluri@studio.unibo.it](mailto:ganeshpavan.kolluri@studio.unibo.it)  
Matricola : 0000976948

---

GitHub Repository: <https://github.com/karthikbharadhwajKB/LoanApprovalBayesianNet/>

## 1 Introduction

The problem statement is “A Company wants to automate the loan approval process based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History, and others”. Here target is Loan Status which has two values yes or no. yes means that Loan Approved and no means Loan is not Approved.

Here I am going to build a Bayesian network with Loan Approval dataset. Where it can infer that a person will get Loan Approved or not by providing Evidence.

## 2 Dataset and Preprocessing

Loan Approval dataset consists of Loan\_ID, Gender, Married, Dependents, Education, Self\_Employed, ApplicantIncome, CoapplicantIncome, LoanAmount, Loan\_Amount\_Term, Credit\_History, Property\_Area, Loan\_Status. Here Loan\_Status variable is Target.

I have downloaded the dataset from Kaggle challenge. Here below I am providing source link.

Dataset Source : <https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset/>

Description of Each Variable:

Variable	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female

Variable	Description
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	Coapplicant 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)

Firstly, we have to load the dataset into notebook using pandas library method `pd.read_csv(Loan_approval.csv)`. and we have to drop “Loan\_ID” feature

because it not an informative feature. we are going to check for the missing variables. Here we can observe that there are missing values in Gender, Married, Dependents, Self\_Employed, LoanAmount, Loan\_Amount\_Term, Credit\_History Variables. Here we have to treat the missing values by using fillna() method with mode.

After that I have converted continuous features into equal bins discrete features by take their max and min value and I opted 3 bins and I have used cut() method to create discrete intervals.

For Example:

### **ApplicantIncome Binned into Three bins**

```
min(data['ApplicantIncome'].unique()), max(data['ApplicantIncome'].unique())
```

output: (150, 81000)

Here, we can observe that min value is 150 and max value is 81000.

```
bin_applicantincome = ['150 to 20,250', '20,250 to 40,500', '40,500 to 81000']
```

I have grouped the intervals like 150 to 20,250, 20,250 to 40,500 and 40,500 to 81,000.

```
data['ApplicantIncome'] = pd.cut(data['ApplicantIncome'], 3, labels=
bin_applicantincome)
```

```
data.ApplicantIncome.value_counts()
```

150 to 20,250	607
20,250 to 40,500	5
40,500 to 81000	2

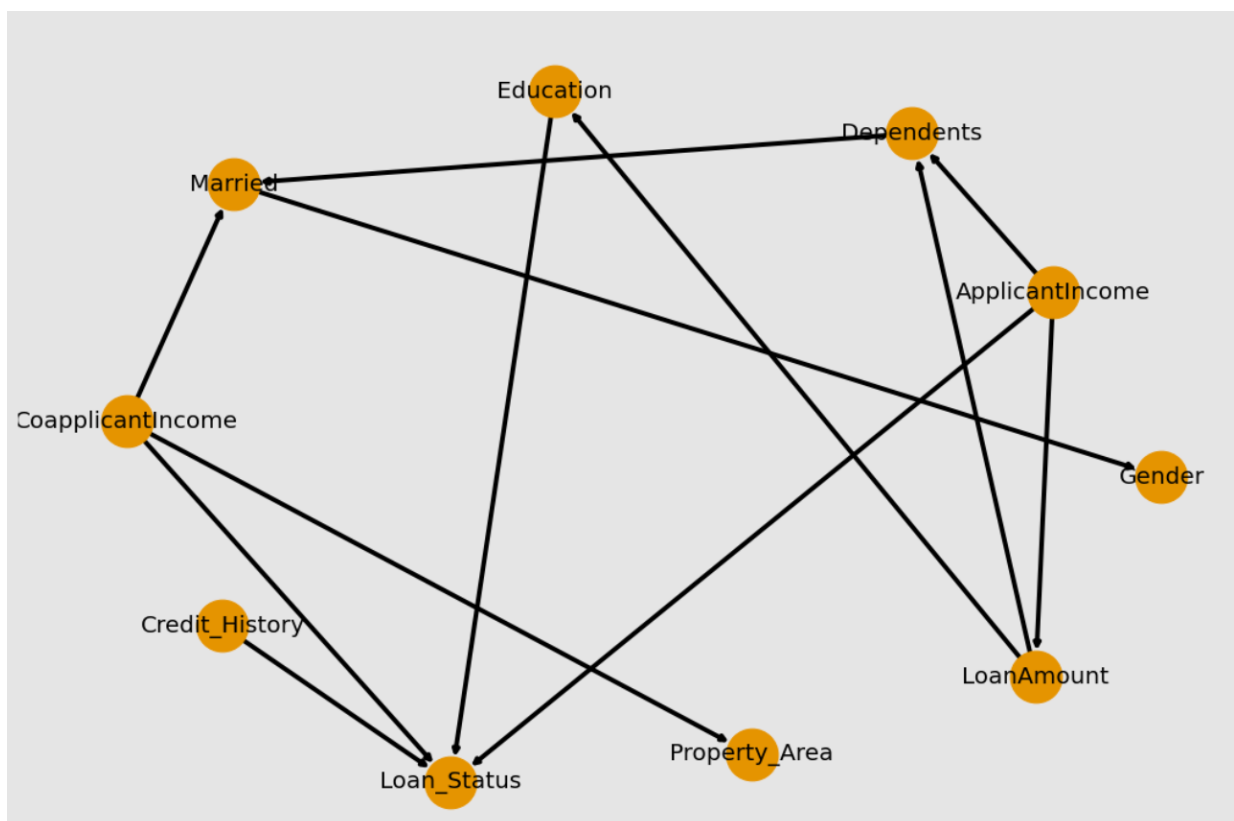
## **3 Bayesian Model**

A Bayesian network, Bayes network, belief network, Bayesian model or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Bayesian networks are mostly used when we want to represent causal relationship between the random variables. Bayesian Networks are parameterized using Conditional Probability Distributions (CPD). Each node in the network is parameterized using where represents the parents of node in the network.

## Building the Bayesian Model

```
model = BayesianModel([('Married', 'Gender'), ('Dependents', 'Married'), ('ApplicantIncome', 'LoanAmount'), ('ApplicantIncome', 'Dependents'), ('ApplicantIncome', 'Loan_Status'), ('CoapplicantIncome', 'Loan_Status'), ('CoapplicantIncome', 'Married'), ('CoapplicantIncome', 'Property_Area'), ('Education', 'Loan_Status'), ('LoanAmount', 'Education'), ('LoanAmount', 'Dependents'), ('Credit_History', 'Loan_Status'), ])
```

## Visualizing the Bayesian Network



## Fitting the model with Loan Approval dataset

fit() method will estimate the CPD for each variables based on a given dataset and it will take estimator. Here I am providing BayesianEstimator as estimator, “BDeu” as prior type and 10 equivalent\_sample\_size.

```
model.fit(data, estimator=BayesianEstimator, prior_type='BDeu', equivalent_sample_size=10)
```

## 4 Conditional Probability Distribution (CPD)

### Conditional probability Distribution of Gender

Married	Married(No)	Married(Yes)
Gender(Female)	0.37844036697247707	0.08497536945812807
Gender(Male)	0.6215596330275229	0.9150246305418719

### Conditional probability Distribution of ApplicantIncome

ApplicantIncome(150 to 20,250)	0.978098
ApplicantIncome(20,250 to 40,500)	0.0133547
ApplicantIncome(40,500 to 81000)	0.00854701

### Conditional probability Distribution of CoapplicantIncome

CoapplicantIncome(0.0 to 10,416.75)	0.982906
CoapplicantIncome(10,416.75 to 20,833.5)	0.00854701
CoapplicantIncome(20,833.5 to 41,667.0)	0.00854701

### Conditional probability Distribution of Credit History

Credit_History(0.0)	0.150641
Credit_History(1.0)	0.849359

## 5 Conditional Independencies

### Independence of Married:

(Married  $\perp$  Education, LoanAmount, Property\_Area, Loan\_Status, ApplicantIncome, Credit\_History | Dependents, CoapplicantIncome)

### Independence of Gender:

(Gender  $\perp$  CoapplicantIncome, Loan\_Status, Dependents, Property\_Area, Education, LoanAmount, Credit\_History, ApplicantIncome | Married)

### Independence of Dependents:

(Dependents  $\perp$  CoapplicantIncome, Loan\_Status, Credit\_History, Property\_Area, Education | ApplicantIncome, LoanAmount)

### Independence of Loan Status:

(Loan\_Status  $\perp$  Married, Gender, Dependents, LoanAmount, Property\_Area | CoapplicantIncome, Education, ApplicantIncome, Credit\_History)

### Independence of Credit History:

(Credit\_History  $\perp$  CoapplicantIncome, Dependents, Property\_Area, Education, Married, LoanAmount, Gender, ApplicantIncome)

We can observe that Credit history is Independent but conditionally independent

.

## 6 Markov Blanket

The Markov Blanket of a Node is its parents, children & parents of its children (excluding itself).

### Markov Blanket for Education:

Markov Blanket of "Education" are Loan\_Status, ApplicantIncome, Credit\_History, CoapplicantIncome, LoanAmount.

**Markov Blanket for LoanAmount:**

Markov Blanket of "LoanAmount" are Dependents, Education, ApplicantIncome.

**Markov Blanket for Credit History:**

Markov Blanket of "Credit\_History" are Education , ApplicantIncome, CoapplicantIncome, Loan\_Status.

**Markov Blanket for ApplicantIncome:**

Markov Blanket of "ApplicantIncome" are Education , ApplicantIncome, CoapplicantIncome, Loan\_Status, Credit\_History, Dependents.

**Markov Blanket for Loan\_Status:**

Markov Blanket of "Loan\_Status" are Education , ApplicantIncome, CoapplicantIncome, Loan\_Status, Credit\_History, Dependents.

## **7 Reasoning Patterns and Active Trails**

**Q1. Is there is any active trail between Education and LoanAmount ?**

There is active trail between Education and LoanAmount. So, Education and LoanAmount are Not Independent.

**Q2. Is there is any active trail between Loan\_Status and Credit\_History ?**

There is active trail between Loan\_Status and Credit\_History. So, Loan\_Status and Credit\_History are Not Independent.

**Q3. Is there is any active trail between Property Area and Gender ?**

There is no active trail between Property\_Area and LoanAmount. So, Property\_Area and LoanAmount are Independent of each other.



## Causal reasoning (Prediction)

### Q1. Do Tony get Loan Approved without any Evidence?

There is 68% of chance that Tony will get Loan Approved.

### Q2. Do Tony get Loan Approved with given Evidence he is "Graduate", and his Credit Score is 1.0 and his income is between 40,500 to 81000 ?

There is 88% of chance that Tony will get Loan Approved.

## Evidential Reasoning (Explaining)

### Q1. Is Tony being a Graduate given that he got Loan Approved ?

There is 79% of chance that Tony is Graduated given evidence that he got Loan Approved.

### Q2. Is Tony is Married given that he got Loan Approved ?

There is 65% of chance that Tony is Married given evidence that he got Loan Approved. May be his wife is also working and earning money.

## 8 Inference

### 8.1 Variable Elimination

Infer = VariableElimination(model)

#### Q1) Loan\_Status given evidence 'Education': 'Graduate', 'ApplicantIncome': '40,500 to 81000' ?

```
print(infer.query(['Loan_Status'], evidence={'Education': 'Graduate', 'ApplicantIncome': '40,500 to 81000'}))
```

```
+-----+-----+
| Loan_Status | phi(Loan_Status) |
+=====+=====+
| Loan_Status(N) | 0.2313 |
+-----+-----+
| Loan_Status(Y) | 0.7687 |
+-----+-----+
```

*Loan\_Status given Education = Graduate and ApplicantIncome = 40,500 to 81000 has 76.87% chance to get Loan Approved.*

**Q2) Loan\_Status given evidence 'Education':'Graduate','ApplicantIncome':'40,500 to 81000', 'CoapplicantIncome':'0.0 to 10,416.75' ?**

```
print(infer.query(['Loan_Status'], evidence={'Education':'Graduate','ApplicantIncome':'40,500 to 81000','CoapplicantIncome':'0.0 to 10,416.75'}))
```

```
+-----+-----+
| Loan_Status | phi(Loan_Status) |
+=====+=====+
| Loan_Status(N) | 0.2266 |
+-----+-----+
| Loan_Status(Y) | 0.7734 |
+-----+-----+
```

*Loan\_Status given Education = Graduate, ApplicantIncome = 40,500 to 81000 and CoapplicantIncome = 0.0 to 10,416.75 has 77.34% chance to get Loan Approved.*

**Q3) Loan\_Status given evidence 'Education':'Graduate','ApplicantIncome':'40,500 to 81000', 'CoapplicantIncome':'0.0 to 10,416.75', 'Credit\_History': 1.0 ?**

```
+-----+-----+
| Loan_Status | phi(Loan_Status) |
+=====+=====+
| Loan_Status(N) | 0.1087 |
+-----+-----+
| Loan_Status(Y) | 0.8913 |
+-----+-----+
```

*Loan\_Status given Education = Graduate, ApplicantIncome = 40,500 to 81000, CoapplicantIncome = 0.0 to 10,416.75 and Credit\_History = 1.0 has 89.13% chance to get Loan Approved.*

## Conclusion

Loan\_Status given Education = Graduate, ApplicantIncome = 40,500 to 81000, CoapplicantIncome = 0.0 to 10,416.75 and Credit\_History = 1.0 has 89.13% chance to get Loan Approved.

There is nearly 90% of chance to get Loan Approved with these Evidence.

A person who graduated in master's degree and earning 40,500 to 81000 euros and his wife who has income of 0 to 10,416.75 euros and has Credit History of 1.0 has nearly 90% chance to get Loan Approved.

## 8 Sampling

There are two Sampling techniques I have used in this Notebook they are Rejection sampling and Likelihood weighting sampling that are based on inference by stochastic simulation drawing N samples from a sampling distribution and compute an approximate posterior probability.

### 9.1 Rejection Sampling

The rejection sampling method is to reject those samples that disagree with the given evidence of the query and compute the probability as a mean of the samples.

#### Calculating the Probabilities from Samples

$$\frac{\text{Number of times Query and Evidence variables occurs}}{\text{Number of times Evidence occurs}}$$

Q1) Query -> Loan\_Status: Y

Evidence -> Education: Graduate, ApplicantIncome = 40,500 to 81000

There is 77% of chance getting Loan Approved given Evidence Education = "Graduate", ApplicantIncome = "40,500 to 81000".

Q2) Query --> Loan\_Status: Y

Evidence --> Education: Graduate, ApplicantIncome = 40,500 to 81000, Credit\_History = 1.0

There is 87.5% of chance getting Loan Approved with given Evidence Education = "Graduate", ApplicantIncome = "40,500 to 81000", Credit\_History = 1.0.

### 9.2 Likelihood Weighted Sampling

The likelihood weighting method adopts the approach to use the evidence to weight the samples in the calculation of the probability.

#### Calculating probabilities from weighted samples

$$\frac{\sum \text{Weights of samples that have Query and Evidence}}{\sum \text{Weights of Sample that have Evidence}}$$

Q1) Query -> Loan\_Status: Y

Evidence -> Education: Graduate, ApplicantIncome = 40,500 to 81000

There is 76% of chance getting Loan Approved with given Evidence Education = "Graduate", ApplicantIncome = "40,500 to 81000".

Q2) Query --> Loan\_Status: Y

Evidence --> Education: Graduate, ApplicantIncome = 40,500 to 81000, Credit\_History = 1.0

There is 85% of chance getting Loan Approved with given Evidence Education = "Graduate", ApplicantIncome = "40,500 to 81000", Credit\_History = 1.0.