LOAN REPAYMENT DECISION SYSTEM

PROJECT TO PREDICT IF LOAN WILL BE REPAID USING BAYESIAN NETWORKS

MEGAMIND

CS 514

Applied Artificial Intelligence

Project 3

INDEX

ABSTRACT:

Loan Repayment Decision system is a Bayesian Network designed using NETICA wherein information of different type of loan applicants and types of loan is asserted by the user of the Bank, and finally concludes whether applicant will be to repay loan or not.Accurately predicting whether a loan will be repaid (credit" scoring) is an important task for any bank. High accuracy benefits both the banks and the loan applicants.

FEATURES:

1. The created is designed for a generic credit scoring scenario using 20 nodes.
2. Nodes within the net are based upon the intuition.
3. The resultant network had 12 evidence variables which do not depend on any external nodes and have their prior probabilities feed, feeding into 6 inner nodes, which, in turn, feed into the outcome node, LoanRepaid.
4. The inner nodes serve two purposes: they group data into qualitative conceptual categories, making interpretation of the BayesianNetwork easier, and they reduce the size of the CPTs in the network (via "divorcing"). However, even when using these inner nodes, the CPT was still too large to fill in manually.
5. CPT i.e conditional probability table for each independent variable were used by Netica to generate the CPTs for dependent variables.
6. The node, IndividualCreditworthiness, used a weighted average of its parents.
7. Note: While some of the probabilities are as close to reality as possible, many have been formulated with hypothesizing and following the general trend defined by banks.
8. Most important variables like IndividualCreditWorthiness and FinancialStability decides the posterior probabilities of every other factors.

USAGE MANUAL:

- Download and place the "Megamind.dne" file in any directory and open it through NETICA application. Compile the project and after the network is initialized, modify and set the values accordingly and see how the network adapts to the change.

SAMPLE OUTPUT:
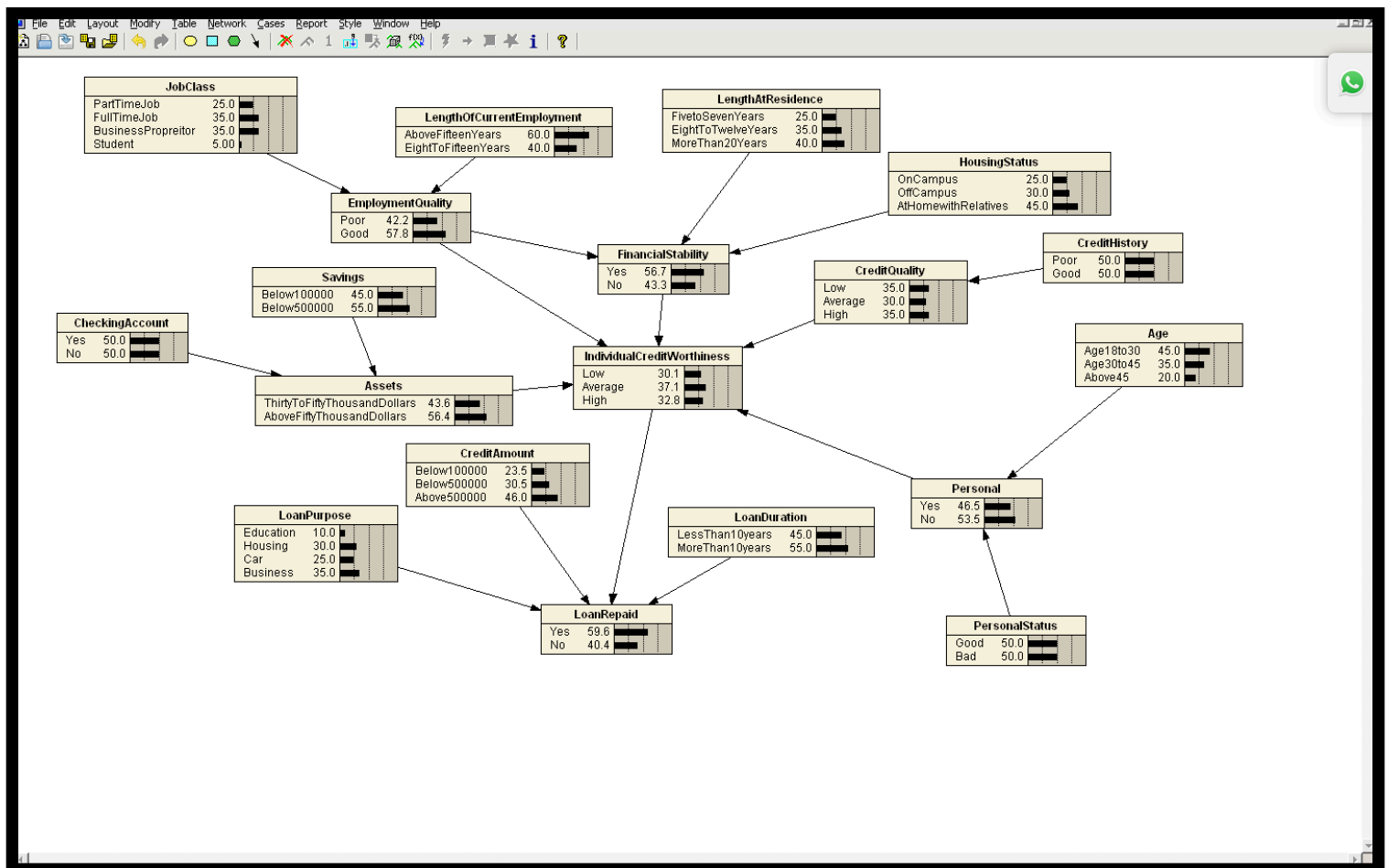
SAMPLE TEST CASES TO TRY:(Follow sequence for each test case as mentioned)

TEST CASE 1:("Select" means Set probability to 100%)

1. **Select, JobClass -> Student**

2. **Employment Quality ->Good**

3. **IndividualCreditWorthiness ->High.**

Output: You will see that the probability of LoanRepaid getting increased if you select IndividualCreditWorthiness ->Low , Probability of LoanRepaid decreases back.

TEST CASE 2:

1. **Select, JobClass -> Student**

2. **Employment Quality ->Good**

3. **IndividualCreditWorthiness ->High.**

4. **Credit Amount -> Below100000**
   Observation: LoanRepaid probability shoots above 70% since everything is good and credit amount i.e loan amount taken is low so higher chances of loan getting repaid by a student. As you go down in the table i.e of Credit Amount's range increases, the probability of loan getting repaid decreases.

5. **Further if , LoanDuration -> LessThan10Years**
   Observation: LoanRepaid probability shoots above 80% since everything is good and credit amount i.e loan amount taken is low and duration to repay is less than 10 years so higher chances of loan getting repaid by a student.(Since majorly loan taken by people are paid within 10 years, if given less duration). Similarly, if duration to repay the loan increases, LoanRepaid probability decreases to 60% since it is observed by the bank that people don't repay the loan if given more than sufficient time to repay.