PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORITHM THROUGH WEKA

DECISION TREE(trees.J48):

cross validation 10:

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: german_credit

Instances: 1000

Attributes: 21

checking_status

duration

credit_history

purpose

credit_amount

savings_status

employment

installment_commitment

personal_status

other_parties

residence_since

property_magnitude

age

other_payment_plans

housing

existing_credits

job

num_dependents

own_telephone

foreign_worker

class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

----

checking_status = <0

| foreign_worker = yes

| | duration <= 11

| | | existing_credits <= 1

| | | | property_magnitude = real estate: good (8.0/1.0)

| | | | property_magnitude = life insurance

| | | | | own_telephone = none: bad (2.0)

| | | | | own_telephone = yes: good (4.0)

| | | | property_magnitude = car: good (2.0/1.0)

| | | | property_magnitude = no known property: bad (3.0)

| | | existing_credits > 1: good (14.0)

| | duration > 11

| | | job = unemp/unskilled non res: bad (5.0/1.0)

| | | job = unskilled resident

| | | | purpose = new car

| | | | | own_telephone = none: bad (10.0/2.0)

| | | | | own_telephone = yes: good (2.0)

| | | | purpose = used car: bad (1.0)

| | | | purpose = furniture/equipment

| | | | | employment = unemployed: good (0.0)

| | | | | employment = <1: bad (3.0)

| | | | | employment = 1<=X<4: good (4.0)

| | | | | employment = 4<=X<7: good (1.0)

| | | | | employment = >=7: good (2.0)

| | | | purpose = radio/tv

| | | | | existing_credits <= 1: bad (10.0/3.0)

| | | | | existing_credits > 1: good (2.0)

| | | | purpose = domestic appliance: bad (1.0)

| | | | purpose = repairs: bad (1.0)

| | | | purpose = education: bad (1.0)

| | | | purpose = vacation: bad (0.0)

| | | | purpose = retraining: good (1.0)

| | | | purpose = business: good (3.0)

| | | | purpose = other: good (1.0)

| | | job = skilled

| | | | other_parties = none

| | | | | duration <= 30

| | | | | | savings_status = <100

| | | | | | | credit_history = no credits/all paid: bad (8.0/1.0)

| | | | | | | credit_history = all paid: bad (6.0)

| | | | | | | credit_history = existing paid

| | | | | | | | own_telephone = none

| | | | | | | | | existing_credits <= 1

| | | | | | | | | | property_magnitude = real estate

| | | | | | | | | | | age <= 26: bad (5.0)

| | | | | | | | | | | age > 26: good (2.0)

| | | | | | | | | property_magnitude = life insurance: bad (7.0/2.0)

| | | | | | | | | | property_magnitude = car

| | | | | | | | | | | credit_amount <= 1386: bad (3.0)

| | | | | | | | | | | credit_amount > 1386: good (11.0/1.0)

| | | | | | | | | | property_magnitude = no known property: good (2.0)

| | | | | | | | | existing_credits > 1: bad (3.0)

| | | | | | | | own_telephone = yes: bad (5.0)

| | | | | | | credit_history = delayed previously: bad (4.0)

| | | | | | | credit_history = critical/other existing credit: good (14.0/4.0)

| | | | | | savings_status = 100<=X<500

| | | | | | | credit_history = no credits/all paid: good (0.0)

| | | | | | | credit_history = all paid: good (1.0)

| | | | | | | credit_history = existing paid: bad (3.0)

| | | | | | | credit_history = delayed previously: good (0.0)

| | | | | | | credit_history = critical/other existing credit: good (2.0)

| | | | | | savings_status = 500<=X<1000: good (4.0/1.0)

| | | | | | savings_status = >=1000: good (4.0)

| | | | | | savings_status = no known savings

| | | | | | | existing_credits <= 1

| | | | | | | | own_telephone = none: bad (9.0/1.0)

| | | | | | | | own_telephone = yes: good (4.0/1.0)

| | | | | | | existing_credits > 1: good (2.0)

| | | | | duration > 30: bad (30.0/3.0)

| | | | other_parties = co applicant: bad (7.0/1.0)

| | | | other_parties = guarantor: good (12.0/3.0)

| | | job = high qualif/self emp/mgmt: good (30.0/8.0)

| foreign_worker = no: good (15.0/2.0)

checking_status = 0<=X<200

| credit_amount <= 9857

| | savings_status = <100

| | | other_parties = none

| | | | duration <= 42

| | | | | personal_status = male div/sep: bad (8.0/2.0)

| | | | | personal_status = female div/dep/mar

| | | | | | purpose = new car: bad (5.0/1.0)

| | | | | | purpose = used car: bad (1.0)

| | | | | | purpose = furniture/equipment

| | | | | | | duration <= 10: bad (3.0)

| | | | | | | duration > 10

| | | | | | | | duration <= 21: good (6.0/1.0)

| | | | | | | | duration > 21: bad (2.0)

| | | | | | purpose = radio/tv: good (8.0/2.0)

| | | | | | purpose = domestic appliance: good (0.0)

| | | | | | purpose = repairs: good (1.0)

| | | | | | purpose = education: good (4.0/2.0)

| | | | | | purpose = vacation: good (0.0)

| | | | | | purpose = retraining: good (0.0)

| | | | | | purpose = business

| | | | | | | residence_since <= 2: good (3.0)

| | | | | | | residence_since > 2: bad (2.0)

| | | | | | purpose = other: good (0.0)

| | | | | personal_status = male single: good (52.0/15.0)

| | | | | personal_status = male mar/wid

| | | | | | duration <= 10: good (6.0)

| | | | | | duration > 10: bad (10.0/3.0)

| | | | | personal_status = female single: good (0.0)

| | | | duration > 42: bad (7.0)

| | | other_parties = co applicant: good (2.0)

| | | other_parties = guarantor

| | | | purpose = new car: bad (2.0)

| | | | purpose = used car: good (0.0)

| | | | purpose = furniture/equipment: good (0.0)

| | | | purpose = radio/tv: good (18.0/1.0)

| | | | purpose = domestic appliance: good (0.0)

| | | | purpose = repairs: good (0.0)

| | | | purpose = education: good (0.0)

| | | | purpose = vacation: good (0.0)

| | | | purpose = retraining: good (0.0)

| | | | purpose = business: good (0.0)

| | | | purpose = other: good (0.0)

| | savings_status = 100<=X<500

| | | purpose = new car: bad (15.0/5.0)

| | | purpose = used car: good (3.0)

| | | purpose = furniture/equipment: bad (4.0/1.0)

| | | purpose = radio/tv: bad (8.0/2.0)

| | | purpose = domestic appliance: good (0.0)

| | | purpose = repairs: good (2.0)

| | | purpose = education: good (0.0)

| | | purpose = vacation: good (0.0)

| | | purpose = retraining: good (0.0)

| | | purpose = business

| | | | housing = rent

| | | | | existing_credits <= 1: good (2.0)

| | | | | existing_credits > 1: bad (2.0)

| | | | housing = own: good (6.0)

| | | | housing = for free: bad (1.0)

| | | purpose = other: good (1.0)

| | savings_status = 500<=X<1000: good (11.0/3.0)

| | savings_status = >=1000: good (13.0/3.0)

| | savings_status = no known savings: good (41.0/5.0)

| credit_amount > 9857: bad (20.0/3.0)

checking_status = >=200: good (63.0/14.0)

checking_status = no checking: good (394.0/46.0)

Number of Leaves : 103

Size of the tree : 140

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 705 70.5 %

Incorrectly Classified Instances 295 29.5 %

Kappa statistic 0.2467

Mean absolute error 0.3467

Root mean squared error 0.4796

Relative absolute error 82.5233 %

Root relative squared error 104.6565 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.84	0.61	0.763	0.84	0.799	0.639	good
	0.39	0.16	0.511	0.39	0.442	0.639	bad
Weighted Avg.	0.705	0.475	0.687	0.705	0.692	0.639	

=== Confusion Matrix ===

a	b	<-- classified as
588	112	a = good
183	117	b = bad

DECISION TREE(trees.J48):

cross validation 5:

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: german_credit

Instances: 1000

Attributes: 21

checking_status

duration

credit_history

purpose

credit_amount

savings_status

employment

installment_commitment

personal_status

other_parties

residence_since

property_magnitude

age

other_payment_plans

housing

existing_credits

job

num_dependents

own_telephone

foreign_worker

class

Test mode:5-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

----

checking_status = <0

| foreign_worker = yes

| | duration <= 11

| | | existing_credits <= 1

| | | | property_magnitude = real estate: good (8.0/1.0)

| | | | property_magnitude = life insurance

| | | | | own_telephone = none: bad (2.0)

| | | | | own_telephone = yes: good (4.0)

| | | | property_magnitude = car: good (2.0/1.0)

| | | | property_magnitude = no known property: bad (3.0)

| | | existing_credits > 1: good (14.0)

| | duration > 11

| | | job = unemp/unskilled non res: bad (5.0/1.0)

| | | job = unskilled resident

| | | | purpose = new car

| | | | | own_telephone = none: bad (10.0/2.0)

| | | | | own_telephone = yes: good (2.0)

| | | | purpose = used car: bad (1.0)

| | | | purpose = furniture/equipment

| | | | | employment = unemployed: good (0.0)

| | | | | employment = <1: bad (3.0)

| | | | | employment = 1<=X<4: good (4.0)

| | | | | employment = 4<=X<7: good (1.0)

| | | | | employment = >=7: good (2.0)

| | | | purpose = radio/tv

| | | | | existing_credits <= 1: bad (10.0/3.0)

| | | | | existing_credits > 1: good (2.0)

| | | | purpose = domestic appliance: bad (1.0)

| | | | purpose = repairs: bad (1.0)

| | | | purpose = education: bad (1.0)

| | | | purpose = vacation: bad (0.0)

| | | | purpose = retraining: good (1.0)

| | | | purpose = business: good (3.0)

| | | | purpose = other: good (1.0)

| | | job = skilled

| | | | other_parties = none

| | | | | duration <= 30

| | | | | | savings_status = <100

| | | | | | | credit_history = no credits/all paid: bad (8.0/1.0)

| | | | | | | credit_history = all paid: bad (6.0)

| | | | | | | credit_history = existing paid

| | | | | | | | own_telephone = none

| | | | | | | | | existing_credits <= 1

| | | | | | | | | | property_magnitude = real estate

| | | | | | | | | | | age <= 26: bad (5.0)

| | | | | | | | | | | age > 26: good (2.0)

| | | | | | | | | | property_magnitude = life insurance: bad (7.0/2.0)

| | | | | | | | | | property_magnitude = car

| | | | | | | | | | | credit_amount <= 1386: bad (3.0)

| | | | | | | | | | | credit_amount > 1386: good (11.0/1.0)

| | | | | | | | | | property_magnitude = no known property: good (2.0)

| | | | | | | | | existing_credits > 1: bad (3.0)

| | | | | | | | own_telephone = yes: bad (5.0)

| | | | | | | credit_history = delayed previously: bad (4.0)

| | | | | | | credit_history = critical/other existing credit: good (14.0/4.0)

| | | | | | savings_status = 100<=X<500

| | | | | | | credit_history = no credits/all paid: good (0.0)

| | | | | | | credit_history = all paid: good (1.0)

| | | | | | | credit_history = existing paid: bad (3.0)

| | | | | | | credit_history = delayed previously: good (0.0)

| | | | | | | credit_history = critical/other existing credit: good (2.0)

| | | | | | savings_status = 500<=X<1000: good (4.0/1.0)

| | | | | | savings_status = >=1000: good (4.0)

| | | | | | savings_status = no known savings

| | | | | | | existing_credits <= 1

| | | | | | | | own_telephone = none: bad (9.0/1.0)

| | | | | | | | own_telephone = yes: good (4.0/1.0)

| | | | | | | existing_credits > 1: good (2.0)

| | | | | duration > 30: bad (30.0/3.0)

| | | | other_parties = co applicant: bad (7.0/1.0)

| | | | other_parties = guarantor: good (12.0/3.0)

| | | job = high qualif/self emp/mgmt: good (30.0/8.0)

| foreign_worker = no: good (15.0/2.0)

checking_status = 0<=X<200

| credit_amount <= 9857

| | savings_status = <100

| | | other_parties = none

| | | | duration <= 42

| | | | | personal_status = male div/sep: bad (8.0/2.0)

| | | | | personal_status = female div/dep/mar

| | | | | | purpose = new car: bad (5.0/1.0)

| | | | | | purpose = used car: bad (1.0)

| | | | | | purpose = furniture/equipment

| | | | | | | duration <= 10: bad (3.0)

| | | | | | | duration > 10

| | | | | | | | duration <= 21: good (6.0/1.0)

| | | | | | | | duration > 21: bad (2.0)

| | | | | | purpose = radio/tv: good (8.0/2.0)

| | | | | | purpose = domestic appliance: good (0.0)

| | | | | | purpose = repairs: good (1.0)

| | | | | | purpose = education: good (4.0/2.0)

| | | | | | purpose = vacation: good (0.0)

| | | | | | purpose = retraining: good (0.0)

| | | | | | purpose = business

| | | | | | | residence_since <= 2: good (3.0)

| | | | | | | residence_since > 2: bad (2.0)

| | | | | | purpose = other: good (0.0)

| | | | | personal_status = male single: good (52.0/15.0)

| | | | | personal_status = male mar/wid

| | | | | | duration <= 10: good (6.0)

| | | | | | duration > 10: bad (10.0/3.0)

| | | | | personal_status = female single: good (0.0)

| | | | duration > 42: bad (7.0)

| | | other_parties = co applicant: good (2.0)

| | | other_parties = guarantor

| | | | purpose = new car: bad (2.0)

| | | | purpose = used car: good (0.0)

| | | | purpose = furniture/equipment: good (0.0)

| | | | purpose = radio/tv: good (18.0/1.0)

| | | | purpose = domestic appliance: good (0.0)

| | | | purpose = repairs: good (0.0)

| | | | purpose = education: good (0.0)

| | | | purpose = vacation: good (0.0)

| | | | purpose = retraining: good (0.0)

| | | | purpose = business: good (0.0)

| | | | purpose = other: good (0.0)

| | savings_status = 100<=X<500

| | | purpose = new car: bad (15.0/5.0)

| | | purpose = used car: good (3.0)

| | | purpose = furniture/equipment: bad (4.0/1.0)

| | | purpose = radio/tv: bad (8.0/2.0)

| | | purpose = domestic appliance: good (0.0)

| | | purpose = repairs: good (2.0)

| | | purpose = education: good (0.0)

| | | purpose = vacation: good (0.0)

| | | purpose = retraining: good (0.0)

| | | purpose = business

| | | | housing = rent

| | | | | existing_credits <= 1: good (2.0)

| | | | | existing_credits > 1: bad (2.0)

| | | | housing = own: good (6.0)

| | | | housing = for free: bad (1.0)

| | | purpose = other: good (1.0)

| | savings_status = 500<=X<1000: good (11.0/3.0)

| | savings_status = >=1000: good (13.0/3.0)

| | savings_status = no known savings: good (41.0/5.0)

| credit_amount > 9857: bad (20.0/3.0)

checking_status = >=200: good (63.0/14.0)

checking_status = no checking: good (394.0/46.0)

Number of Leaves : 103

Size of the tree : 140

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	733	73.3 %
Incorrectly Classified Instances	267	26.7 %
Kappa statistic	0.3264	

Mean absolute error 0.3293

Root mean squared error 0.4579

Relative absolute error 78.3705 %

Root relative squared error 99.914 %

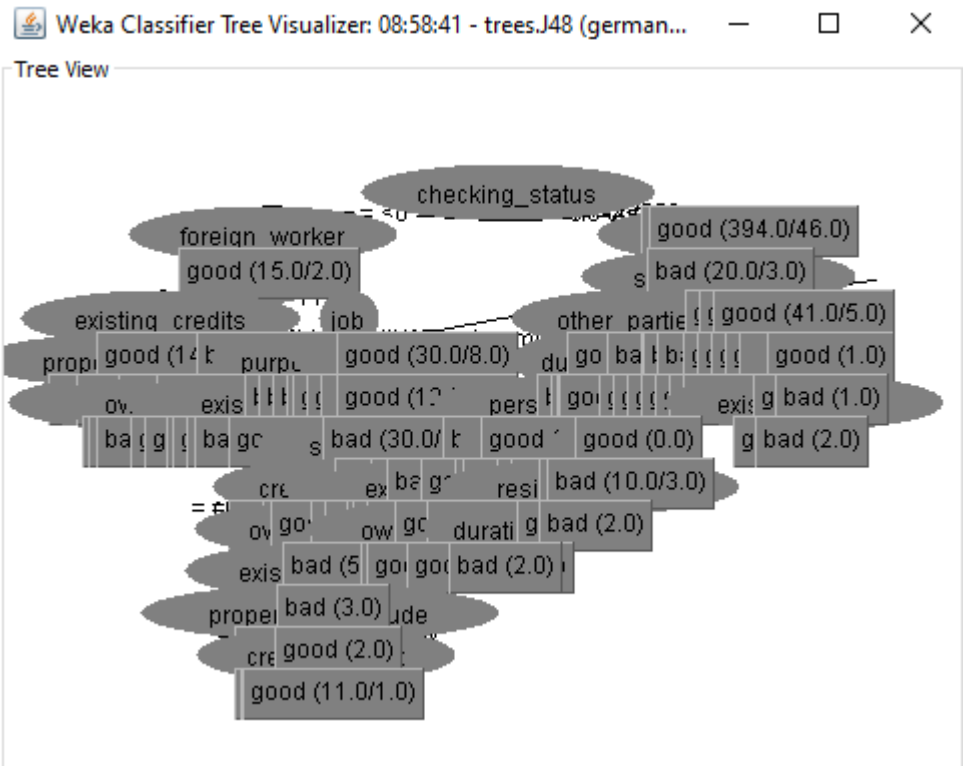
Total Number of Instances 1000

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.851	0.543	0.785	0.851	0.817	0.685	good
	0.457	0.149	0.568	0.457	0.506	0.685	bad
Weighted Avg.	0.733	0.425	0.72	0.733	0.724	0.685	

=== Confusion Matrix ===

a	b	<-- classified as
596	104	a = good
163	137	b = bad



FUNCTION(function.logistic):

=== Run information ===

Scheme:weka.classifiers.functions.Logistic -R 1.0E-8 -M -1

Relation: german_credit

Instances: 1000

Attributes: 21

checking_status

duration

credit_history

purpose

credit_amount

savings_status

employment

installment_commitment

personal_status

other_parties

residence_since

property_magnitude

age

other_payment_plans

housing

existing_credits

job

num_dependents

own_telephone

foreign_worker

class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Variable	Class good

==========

checking_status=<0 -0.778

checking_status=0<=X<200 -0.4032

checking_status=>=200 0.1877

checking_status=no checking 0.9338

duration -0.0279

credit_history=no credits/all paid -0.8129

credit_history=all paid -0.9562

credit_history=existing paid -0.2268

credit_history=delayed previously 0.0403

credit_history=critical/other existing credit 0.6229

purpose=new car -0.692

purpose=used car 0.9744

purpose=furniture/equipment 0.0996

purpose=radio/tv 0.1996

purpose=domestic appliance -0.1692

purpose=repairs -0.4756

purpose=education -0.7283

purpose=vacation 0

purpose=retraining 1.3674

purpose=business 0.0481

purpose=other 0.7968

credit_amount	-0.0001
savings_status=<100	-0.4402
savings_status=100<=X<500	-0.0825
savings_status=500<=X<1000	-0.0641
savings_status=>=1000	0.8989
savings_status=no known savings	0.5065
employment=unemployed	-0.2934
employment=<1	-0.2265
employment=1<=X<4	-0.1106
employment=4<=X<7	0.5376
employment=>=7	-0.0168
installment_commitment	-0.3301
personal_status=male div/sep	-0.4923
personal_status=female div/dep/mar	-0.2168
personal_status=male single	0.3238
personal_status=male mar/wid	-0.1252
personal_status=female single	0
other_parties=none	-0.1798
other_parties=co applicant	-0.6158
other_parties=guarantor	0.7988
residence_since	-0.0048
property_magnitude=real estate	0.2572
property_magnitude=life insurance	-0.0242

property_magnitude=car 0.0627

property_magnitude=no known property -0.4732

age 0.0145

other_payment_plans=bank -0.3273

other_payment_plans=stores -0.2041

other_payment_plans=none 0.3191

housing=rent -0.3498

housing=own 0.0939

housing=for free 0.3341

existing_credits -0.2721

job=unemp/unskilled non res 0.5095

job=unskilled resident -0.0265

job=skilled -0.0451

job=high qualif/self emp/mgmt 0.0301

num_dependents -0.2647

own_telephone 0.3

foreign_worker 1.3922

Intercept 3.1983

Odds Ratios...

Class

Variable good

==========

checking_status=<0 0.4593

checking_status=0<=X<200 0.6682

checking_status=>=200 1.2064

checking_status=no checking 2.5443

duration 0.9725

credit_history=no credits/all paid 0.4436

credit_history=all paid 0.3843

credit_history=existing paid 0.7971

credit_history=delayed previously 1.0411

credit_history=critical/other existing credit 1.8643

purpose=new car 0.5006

purpose=used car 2.6497

purpose=furniture/equipment 1.1047

purpose=radio/tv 1.2209

purpose=domestic appliance 0.8443

purpose=repairs 0.6215

purpose=education 0.4827

purpose=vacation 1

purpose=retraining 3.9251

purpose=business 1.0493

purpose=other 2.2184

credit_amount	0.9999
savings_status=<100	0.6439
savings_status=100<=X<500	0.9208
savings_status=500<=X<1000	0.9379
savings_status=>=1000	2.4569
savings_status=no known savings	1.6594
employment=unemployed	0.7457
employment=<1	0.7973
employment=1<=X<4	0.8953
employment=4<=X<7	1.7119
employment=>=7	0.9834
installment_commitment	0.7189
personal_status=male div/sep	0.6112
personal_status=female div/dep/mar	0.8051
personal_status=male single	1.3824
personal_status=male mar/wid	0.8823
personal_status=female single	1
other_parties=none	0.8354
other_parties=co applicant	0.5402
other_parties=guarantor	2.2229
residence_since	0.9952
property_magnitude=real estate	1.2933
property_magnitude=life insurance	0.9761

property_magnitude=car 1.0647

property_magnitude=no known property 0.623

age 1.0146

other_payment_plans=bank 0.7209

other_payment_plans=stores 0.8154

other_payment_plans=none 1.3758

housing=rent 0.7049

housing=own 1.0984

housing=for free 1.3967

existing_credits 0.7618

job=unemp/unskilled non res 1.6645

job=unskilled resident 0.9738

job=skilled 0.9559

job=high qualif/self emp/mgmt 1.0306

num_dependents 0.7675

own_telephone 1.3499

foreign_worker 4.0237

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 752 75.2 %

Incorrectly Classified Instances 248 24.8 %

Kappa statistic 0.375

Mean absolute error 0.3098

Root mean squared error 0.4087

Relative absolute error 73.727 %

Root relative squared error 89.1751 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.864	0.51	0.798	0.864	0.83	0.785	good
	0.49	0.136	0.607	0.49	0.542	0.785	bad
Weighted Avg.	0.752	0.398	0.741	0.752	0.744	0.785	

=== Confusion Matrix ===

a	b	<-- classified as
605	95	a = good
153	147	b = bad