PREDICTION OF CATEGORICAL DATA USING SMO ALGORITHM

DECISION TREE (trees.J48)

=== 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 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

SMO ALGORITHM (functions.SMO)

=== Run information ===

Scheme:weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -C 250007 -E 1.0"

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) ===

SMO

Kernel used:

Linear Kernel: K(x,y) = <x,y>

Classifier for classes: good, bad

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.6805 * (normalized) checking_status=<0

+ 0.3347 * (normalized) checking_status=0<=X<200

+ -0.4616 * (normalized) checking_status=>=200

+ -0.5537 * (normalized) checking_status=no checking

+ 1.6987 * (normalized) duration

+ 0.5398 * (normalized) credit_history=no credits/all paid

+ 0.6015 * (normalized) credit_history=all paid

+ -0.109 * (normalized) credit_history=existing paid

+ -0.3182 * (normalized) credit_history=delayed previously

+ -0.7141 * (normalized) credit_history=critical/other existing credit

+ 0.5673 * (normalized) purpose=new car

+ -0.5615 * (normalized) purpose=used car

+ -0.1464 * (normalized) purpose=furniture/equipment

+ -0.0798 * (normalized) purpose=radio/tv

+ 0.5456 * (normalized) purpose=domestic appliance

+ 0 * (normalized) purpose=repairs

+ 0.4441 * (normalized) purpose=education

+ -0.3951 * (normalized) purpose=retraining

+ -0.0823 * (normalized) purpose=business

+ -0.2919 * (normalized) purpose=other

+ 1.1473 * (normalized) credit_amount

+ 0.4056 * (normalized) savings_status=<100

+ 0.115 * (normalized) savings_status=100<=X<500

+ 0.1378 * (normalized) savings_status=500<=X<1000

+ -0.3775 * (normalized) savings_status=>=1000

+ -0.2809 * (normalized) savings_status=no known savings

+ 0.2887 * (normalized) employment=unemployed

+ 0.1663 * (normalized) employment=<1

+ 0.0021 * (normalized) employment=1<=X<4

+ -0.3348 * (normalized) employment=4<=X<7

+ -0.1222 * (normalized) employment=>=7

+ 0.6503 * (normalized) installment_commitment

+ 0.3335 * (normalized) personal_status=male div/sep

+ 0.1177 * (normalized) personal_status=female div/dep/mar

+ -0.3697 * (normalized) personal_status=male single

+ -0.0815 * (normalized) personal_status=male mar/wid

+ 0.0514 * (normalized) other_parties=none

+ 0.5697 * (normalized) other_parties=co applicant

+ -0.6211 * (normalized) other_parties=guarantor

+ -0.0001 * (normalized) residence_since

+ -0.2247 * (normalized) property_magnitude=real estate

+ -0.0544 * (normalized) property_magnitude=life insurance

+ -0.0795 * (normalized) property_magnitude=car

+ 0.3586 * (normalized) property_magnitude=no known property

+ -0.4191 * (normalized) age

+ 0.0697 * (normalized) other_payment_plans=bank

+ 0.159 * (normalized) other_payment_plans=stores

+ -0.2287 * (normalized) other_payment_plans=none

+ 0.3271 * (normalized) housing=rent

+ -0.0702 * (normalized) housing=own

+ -0.257 * (normalized) housing=for free

+ 0.4503 * (normalized) existing_credits

+ -0.2026 * (normalized) job=unemp/unskilled non res

+ 0.1501 * (normalized) job=unskilled resident

+ 0.1027 * (normalized) job=skilled

+ -0.0502 * (normalized) job=high qualif/self emp/mgmt

+ 0.0198 * (normalized) num_dependents

+ -0.1394 * (normalized) own_telephone

+ -0.9888 * (normalized) foreign_worker

- 1.5398

Number of kernel evaluations: 436644 (90.558% cached)

Time taken to build model: 0.22 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	**751**	**75.1 %**
Incorrectly Classified Instances	**249**	**24.9 %**
Kappa statistic	**0.3654**	
Mean absolute error	**0.249**	

Root mean squared error 0.499

Relative absolute error 59.2607 %

Root relative squared error 108.8905 %

Total Number of Instances 1000

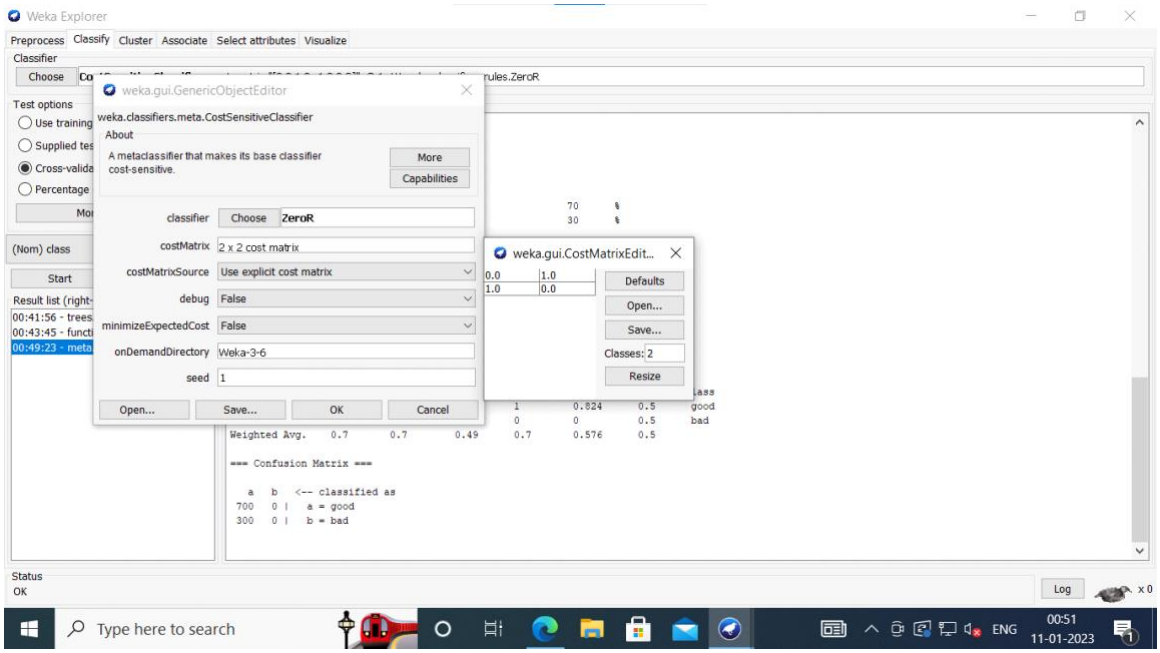
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.871	0.53	0.793	0.871	0.83	0.671	good
	0.47	0.129	0.61	0.47	0.531	0.671	bad
Weighted Avg.	0.751	0.41	0.738	0.751	0.741	0.671	

=== Confusion Matrix ===

a	b	<-- classified as
610	90	a = good
159	141	b = bad

COST SENSITIVE EVALUATION (meta.CostSensitiveClassifier)



=== Run information ===

Scheme:weka.classifiers.meta.CostSensitiveClassifier -cost-matrix "[0.0 1.0; 1.0 0.0]" -S 1 -W weka.classifiers.rules.ZeroR

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

Evaluation cost matrix:

0 1

1 0

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

CostSensitiveClassifier using reweighted training instances

weka.classifiers.rules.ZeroR

Classifier Model

ZeroR predicts class value: good

Cost Matrix

0 1

1 0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	**700**	**70 %**
Incorrectly Classified Instances	**300**	**30 %**
Kappa statistic	**0**	
Total Cost	**300**	
Average Cost	**0.3**	
Mean absolute error	**0.4202**	
Root mean squared error	**0.4583**	
Relative absolute error	**100 %**	
Root relative squared error	**100 %**	
Total Number of Instances	**1000**	

=== Detailed Accuracy By Class ===

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	1	**1**	**0.7**	**1**	**0.824**	**0.5**	**good**
	0	**0**	**0**	**0**	**0**	**0.5**	**bad**
Weighted Avg.	**0.7**	**0.7**	**0.49**	**0.7**	**0.576**	**0.5**	

=== Confusion Matrix ===

a	**b**	**<-- classified as**
700	**0**	**| a = good**
300	**0**	**| b = bad**