Monthly Session

Self-Learning Decision Models

RuleLearner.com

Presented on Sep 4, 2020

Jacob Feldman, PhD
OpenRules, Inc., CTO

RuleLearner.com

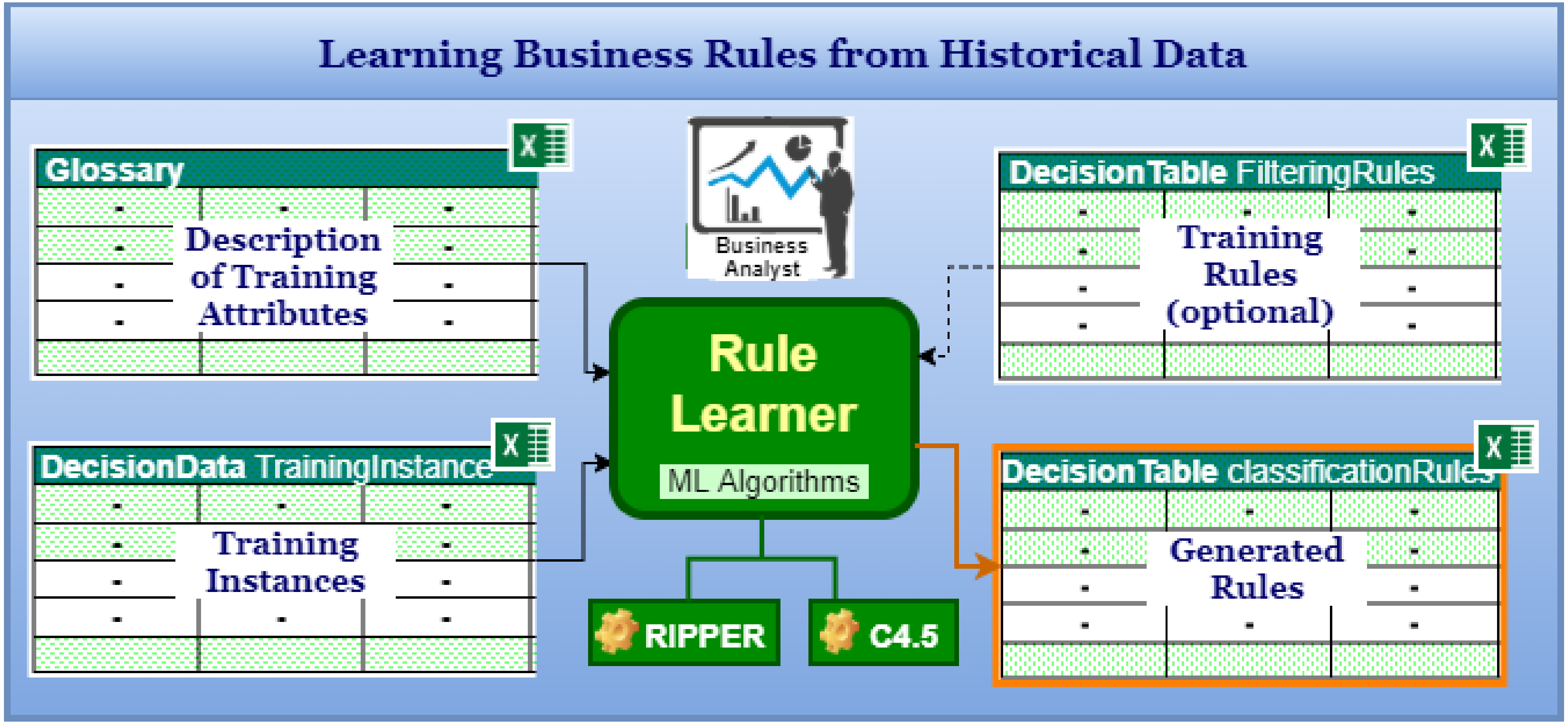
- Machine Learning offers powerful algorithms for the extraction of patterns from large collections of historical data to present them in the form of readable classification rules
- Rule Learner is an *open source tool* that naturally integrates **Machine Learning** (**ML**) and **Business Rules (BR)** techniques by incorporating ML algorithms into rules-based Decision Models

RuleLearner.com

- Rule Learner is oriented to ***business analysts*** who want to *apply Machine Learning to their historical data and get the business rules* without necessity to learn complex data formats, new interfaces, or programming
- How it works?

RuleLearner.com

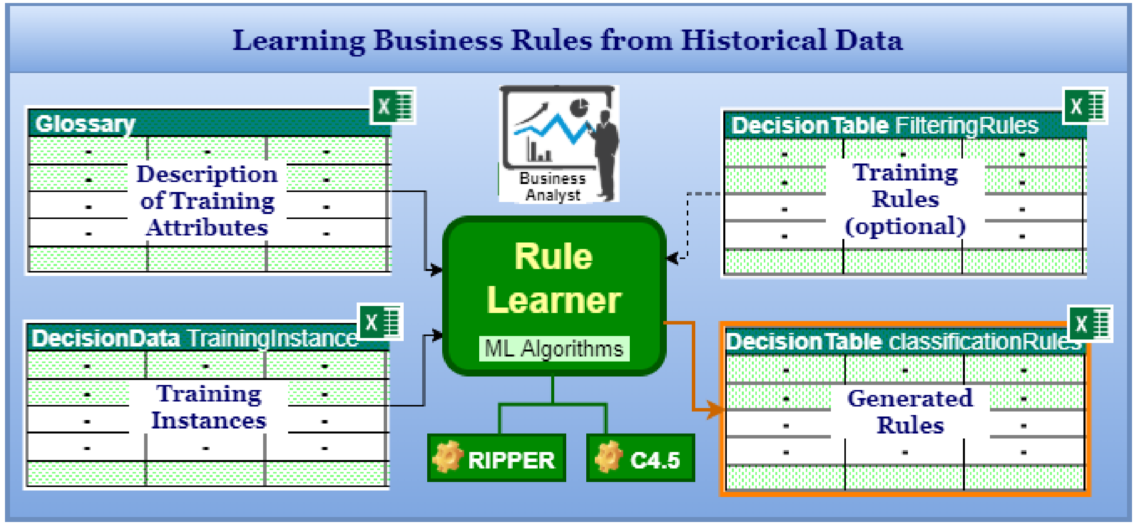
- How it works


Learning Business Rules from Historical Data
Glossary
Description of Training Attributes
DecisionData TrainingInstance
Training Instances
Business Analyst
Rule Learner
ML Algorithms
RIPPER
C4.5
DecisionTable FilteringRules
Training Rules (optional)
DecisionTable classificationRules
Generated Rules

© 2020 OpenRules, Inc.

RuleLearner.com

- Business analyst does the following:
  1. Creates 2 Excel tables:
     - Training Instances
     - Glossary that describes used attributes
  2. Select an ML Algorithm
     - RIPPER
     - C4.5
     - Other
  3. Click on “learn.bat”
  4. Business rules and quality metrics are generated in
     - Human-readable and Rule Engine executable formats

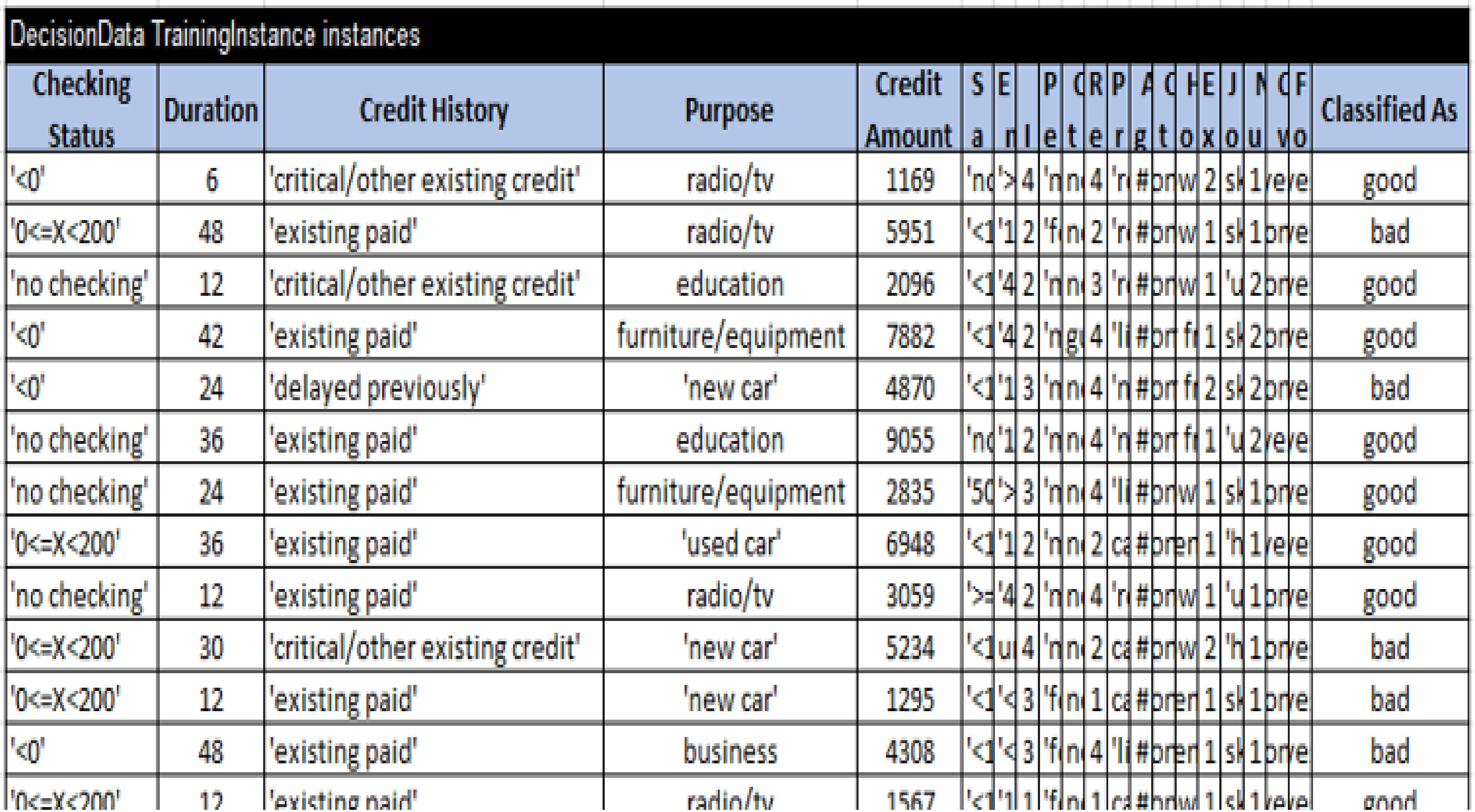

Learning Business Rules from Historical Data
Glossary
Description of Training Attributes
DecisionData TrainingInstances
Training Instances
Business Analyst
Rule Learner
ML Algorithms
RIPPER
C4.5
DecisionTable FilteringRules
Training Rules (optional)
DecisionTable classificationRules
Generated Rules

RuleLearner.com – Example “Credits Classification Rules”

- We have a credit data set which contains 1,000 records about different debtors with such characteristics as Checking Status, Duration, Credit History, Purpose, Credit Amount, Saving Status, Employment, and many others
- Each record is already classified as "good“ or "bad“
- We need to find rules capable to classify new debtors

RuleLearner.com – Example “Credits Classification Rules”

- Training Instances – total 1,000 instances

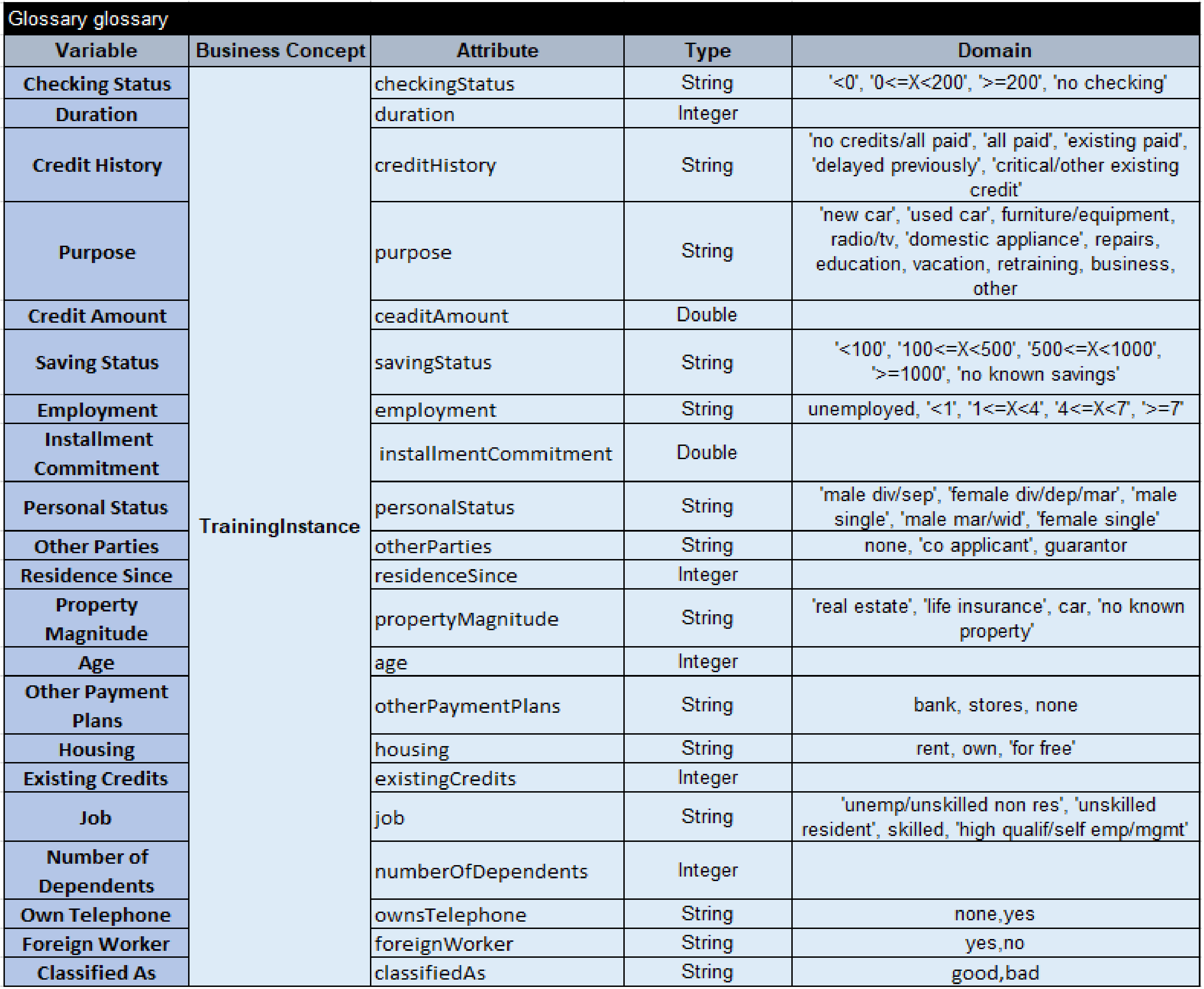


DecisionData TrainingInstance instances

Checking Status	Duration	Credit History	Purpose	Credit Amount	S a	E n	l	P e	C t	R e	P r	A g	C t	H o	E x	J o	N u	C v	F o	Classified As
'<0'	6	'critical/other existing credit'	radio/tv	1169	'nc	'>	4	'n	n	4	'r	#	on	w	2	sk	1	ye	ye	good
'0<=X<200'	48	'existing paid'	radio/tv	5951	'<1	'1	2	'f	n	2	'r	#	on	w	1	sk	1	on	ye	bad
'no checking'	12	'critical/other existing credit'	education	2096	'<1	'4	2	'n	n	3	'r	#	on	w	1	'u	2	on	ye	good
'<0'	42	'existing paid'	furniture/equipment	7882	'<1	'4	2	'n	gu	4	'li	#	on	fr	1	sk	2	on	ye	good
'<0'	24	'delayed previously'	'new car'	4870	'<1	'1	3	'n	n	4	'n	#	on	fr	2	sk	2	on	ye	bad
'no checking'	36	'existing paid'	education	9055	'nc	'1	2	'n	n	4	'n	#	on	fr	1	'u	2	ye	ye	good
'no checking'	24	'existing paid'	furniture/equipment	2835	'50	'>	3	'n	n	4	'li	#	on	w	1	sk	1	on	ye	good
'0<=X<200'	36	'existing paid'	'used car'	6948	'<1	'1	2	'n	n	2	ca	#	or	en	1	'h	1	ye	ye	good
'no checking'	12	'existing paid'	radio/tv	3059	'>=	'4	2	'n	n	4	'r	#	on	w	1	'u	1	on	ye	good
'0<=X<200'	30	'critical/other existing credit'	'new car'	5234	'<1	u	4	'n	n	2	ca	#	on	w	2	'h	1	on	ye	bad
'0<=X<200'	12	'existing paid'	'new car'	1295	'<1	'<	3	'f	n	1	ca	#	or	en	1	sk	1	on	ye	bad
'<0'	48	'existing paid'	business	4308	'<1	'<	3	'f	n	4	'li	#	or	en	1	sk	1	on	ye	bad
'0<=X<200'	12	'existing paid'	radio/tv	1567	'<1	'1	1	'f	n	1	ca	#	on	w	1	sk	1	ye	ye	good

RuleLearner.com

- Glossary

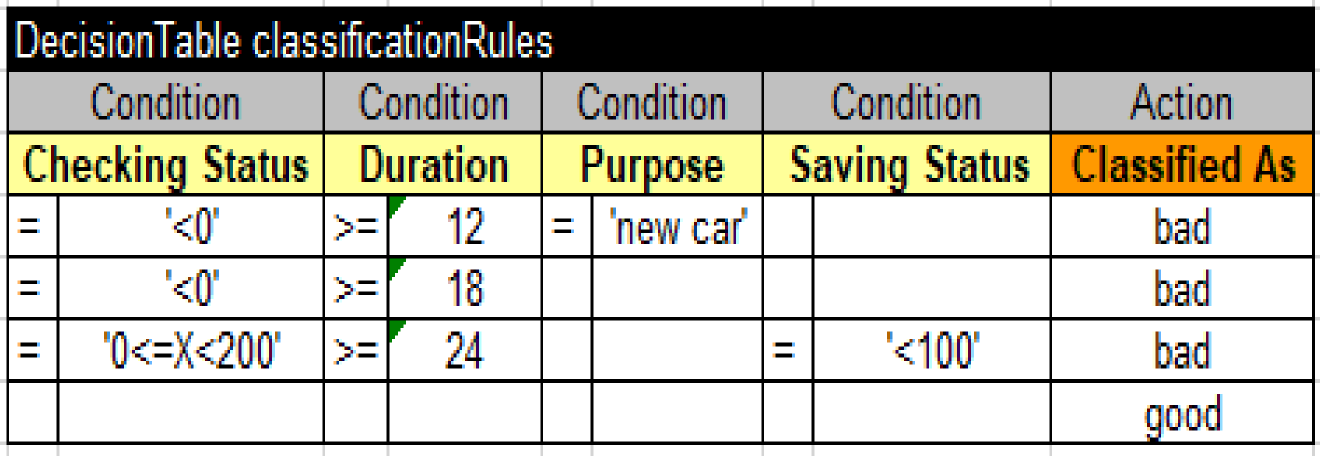


Glossary glossary

Variable	Business Concept	Attribute	Type	Domain
Checking Status	TrainingInstance	checkingStatus	String	'<0', '0<=X<200', '>=200', 'no checking'
Duration		duration	Integer	
Credit History		creditHistory	String	'no credits/all paid', 'all paid', 'existing paid', 'delayed previously', 'critical/other existing credit'
Purpose		purpose	String	'new car', 'used car', furniture/equipment, radio/tv, 'domestic appliance', repairs, education, vacation, retraining, business, other
Credit Amount		ceaditAmount	Double	
Saving Status		savingStatus	String	'<100', '100<=X<500', '500<=X<1000', '>=1000', 'no known savings'
Employment		employment	String	unemployed, '<1', '1<=X<4', '4<=X<7', '>=7'
Installment Commitment		installmentCommitment	Double	
Personal Status		personalStatus	String	'male div/sep', 'female div/dep/mar', 'male single', 'male mar/wid', 'female single'
Other Parties		otherParties	String	none, 'co applicant', guarantor
Residence Since		residenceSince	Integer	
Property Magnitude		propertyMagnitude	String	'real estate', 'life insurance', car, 'no known property'
Age		age	Integer	
Other Payment Plans		otherPaymentPlans	String	bank, stores, none
Housing		housing	String	rent, own, 'for free'
Existing Credits		existingCredits	Integer	
Job		job	String	'unemp/unskilled non res', 'unskilled resident', skilled, 'high qualif/self emp/mgmt'
Number of Dependents		numberOfDependents	Integer	
Own Telephone		ownsTelephone	String	none,yes
Foreign Worker		foreignWorker	String	yes,no
Classified As		classifiedAs	String	good,bad

RuleLearner.com – Example “Credits Classification Rules”

- Generated Rules (using RIPPER)



DecisionTable classificationRules

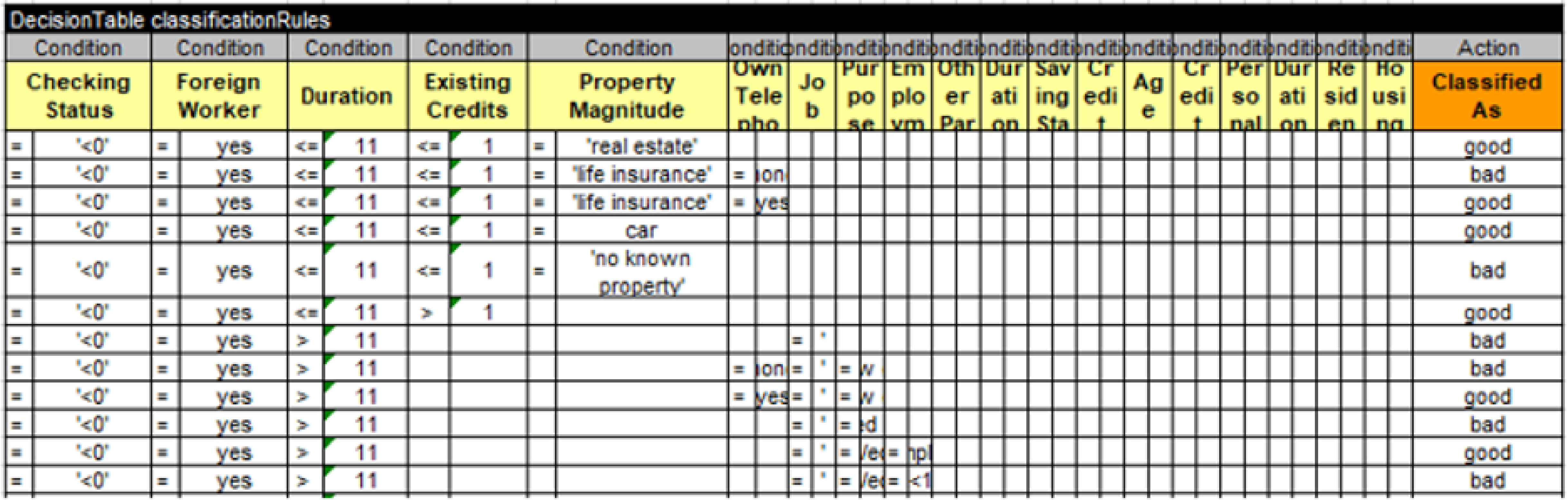
Condition		Condition		Condition		Condition		Action
Checking Status		Duration		Purpose		Saving Status		Classified As
=	'<0'	>=	12	=	'new car'			bad
=	'<0'	>=	18					bad
=	'0<=X<200'	>=	24			=	'<100'	bad
								good

- These 4 rules correctly classify **756** out of 1,000 instances

- Considering cross-validation, it gives 72.7% success rate for new instances

RuleLearner.com – Example “Credits Classification Rules”

- Generated Rules (using C4.5)



DecisionTable classificationRules

Condition	Condition	Condition	Condition	Condition	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	onditi	Action
Checking Status	Foreign Worker	Duration	Existing Credits	Property Magnitude	Own Tele pho	Jo b	Pur po se	Em plo ym	Oth er Par	Dur ati on	Sav ing Sta	Cr edi t	Ag e	Cr edi t	Per so nal	Dur ati on	Re sid en	Ho usi ng	Classified As
= '<0'	= yes	<= 11	<= 1	= 'real estate'															good
= '<0'	= yes	<= 11	<= 1	= 'life insurance'	= ion														bad
= '<0'	= yes	<= 11	<= 1	= 'life insurance'	= yes														good
= '<0'	= yes	<= 11	<= 1	= car															good
= '<0'	= yes	<= 11	<= 1	= 'no known property'															bad
= '<0'	= yes	<= 11	> 1																good
= '<0'	= yes	> 11				= '													bad
= '<0'	= yes	> 11			= ion	= '	= w												bad
= '<0'	= yes	> 11			= yes	= '	= w												good
= '<0'	= yes	> 11				= '	= d												bad
= '<0'	= yes	> 11				= '	= /e	= npl											good
= '<0'	= yes	> 11				= '	= /e	= <1											bad

- These 103 rules correctly classify **855** out of 1,000 instances

- Considering cross-validation, it gives 70.5% success rate for new instances

RuleLearner.com – Quality of the Generated Rules

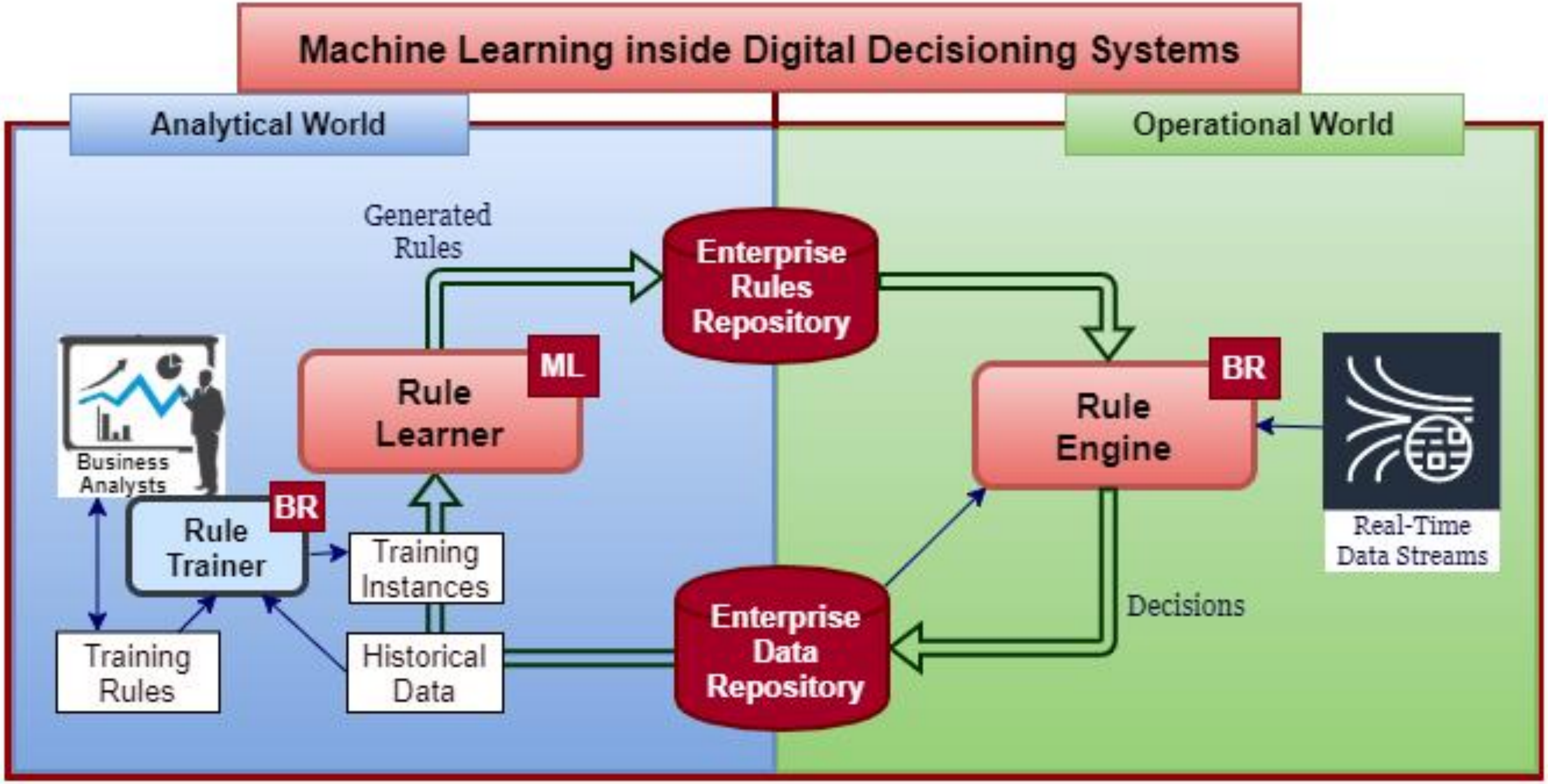
- Should we even try to generate *perfect* rules that guarantee to give the correct classification on all instances in the training set?
- The answer is “No":

  "*You would rather generate 'sensible' rules that avoid over-fitting the training set and thereby stand a better chance of performing well on new instances*" WEKA's book

RuleLearner.com – Improving Rules Quality

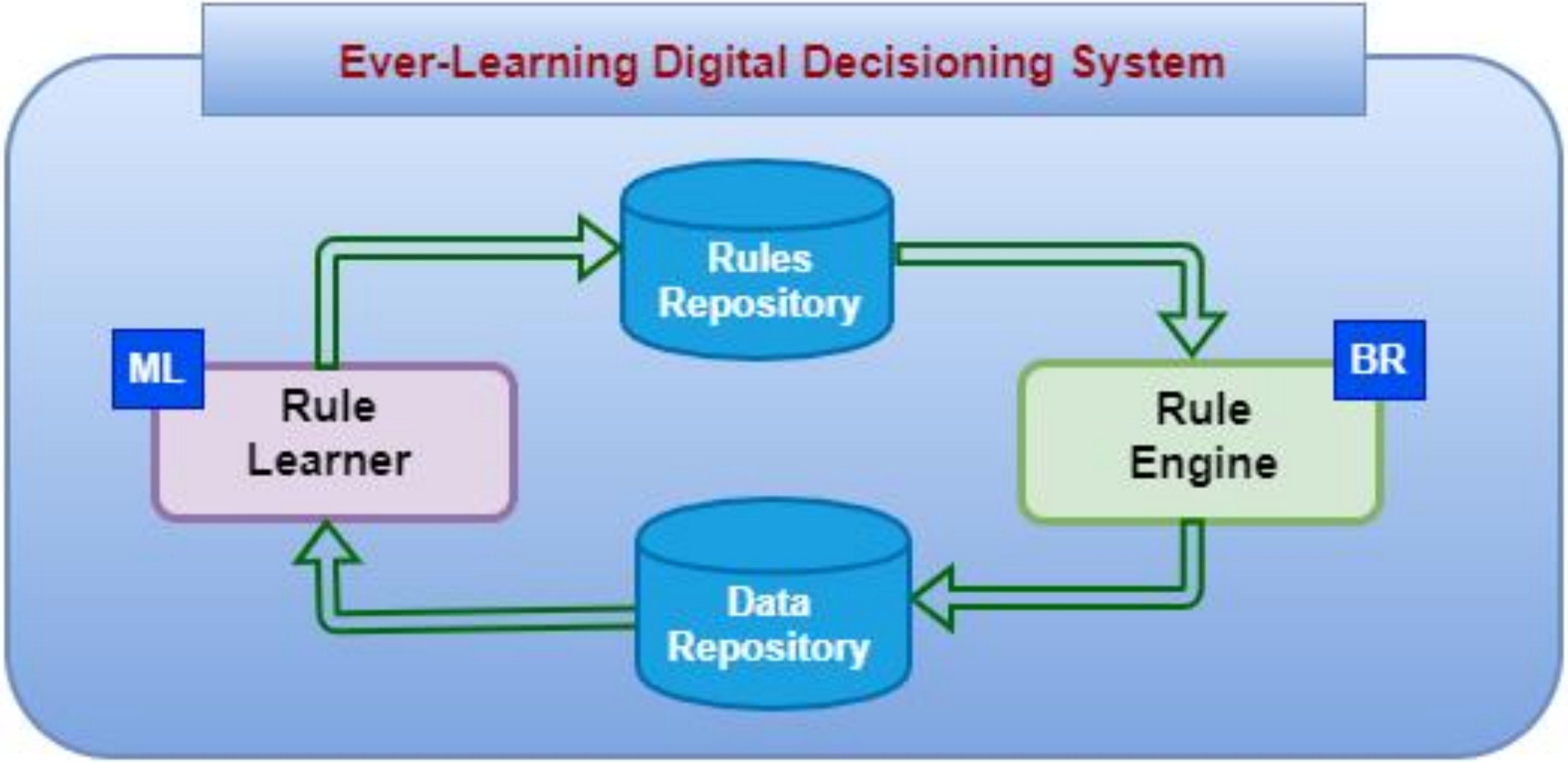
- **Involve Domain Knowledge**
- **Using Rule Trainer** (before generation)
  - Domain experts create Business Rules for filtering training instances
  - BR+ML+BR
- **Adjusting Generated Rules** (after generation)
  - Generated rules in Excel are easy to understand and adjust

- How it works


Machine Learning inside Digital Decisioning Systems
Analytical World
Operational World
Generated Rules
Enterprise Rules Repository
ML
Rule Learner
BR
Rule Engine
Business Analysts
BR
Rule Trainer
Training Instances
Real-Time Data Streams
Training Rules
Historical Data
Enterprise Data Repository
Decisions

RuleLearner.com

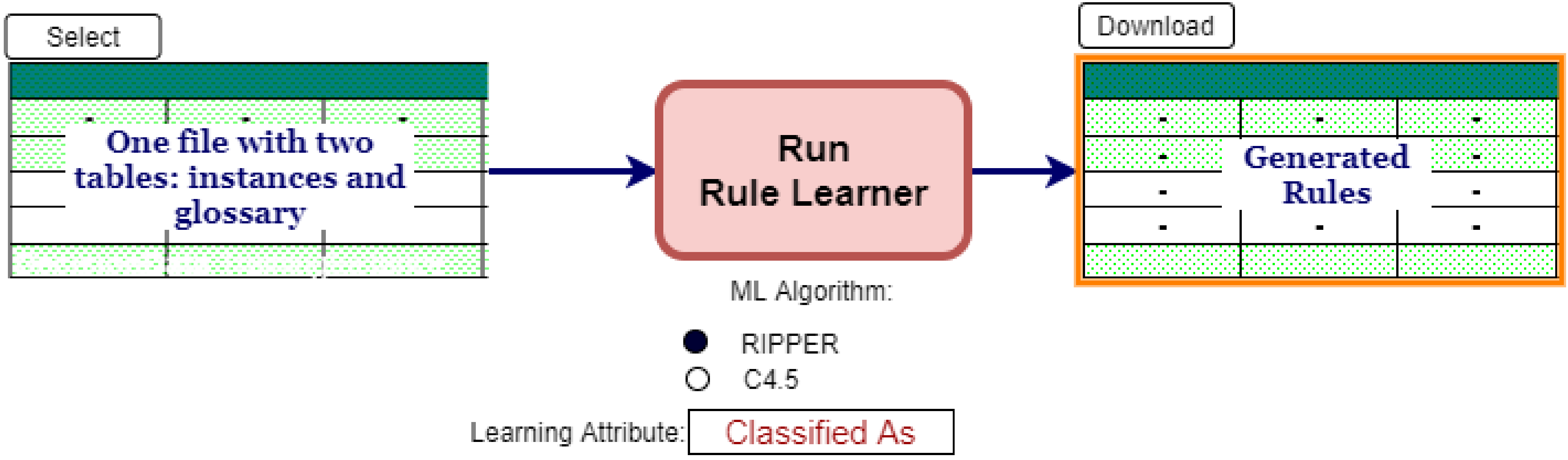
- Rule Learner supports **Self-Learning Decision Models** by implementing this ever-learning loop:



© 2020 OpenRules, Inc.

RuleLearner.com

- **SaaS Rule Learner** in AWS Marketplace
- No Installation required:
  - Drag & drop an Excel file with training data
  - Get back generated rules with metrics



Select

One file with two tables: instances and glossary

Run Rule Learner

Download

Generated Rules

ML Algorithm:

● RIPPER
○ C4.5

Learning Attribute: Classified As

- **Implementation:**
  - Open Source (LGPL) with a simple Java API
  - Uses underlying ML algorithms available from open-source WEKA (proven records)
  - More ML implementations/algorithms are on the way
  - Works with OpenRules Decision Manager
  - Easy to integrate with other Decisioning Products

Questions?

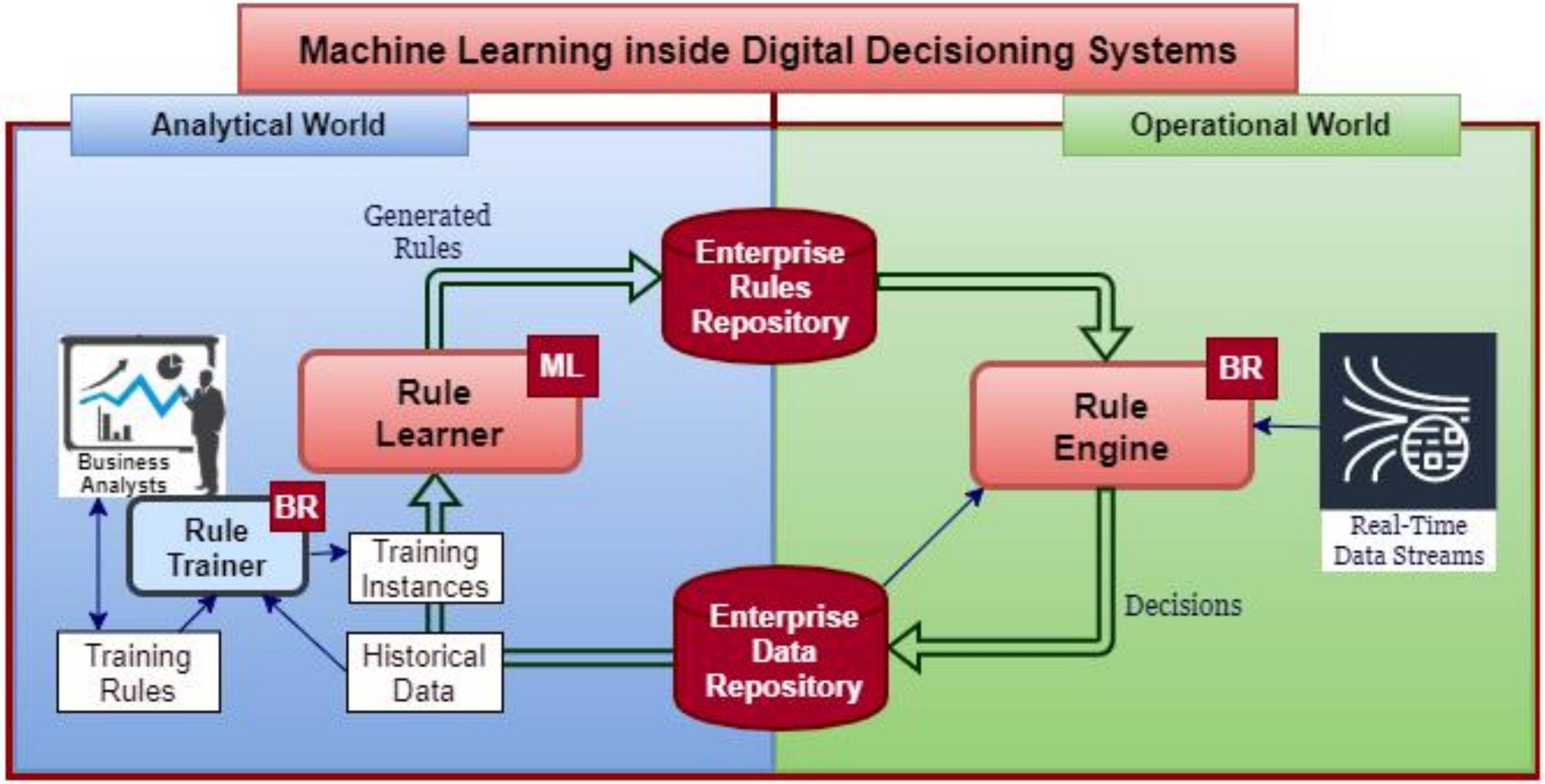
Jacob Feldman, PhD

OpenRules, Inc.

www.OpenRules.com

jacobfeldman@openrules.com

RuleLearner.com



Machine Learning inside Digital Decisioning Systems
Analytical World
Operational World
Generated Rules
Enterprise Rules Repository
Rule Learner
ML
Rule Engine
BR
Business Analysts
Rule Trainer
BR
Training Instances
Real-Time Data Streams
Decisions
Training Rules
Historical Data
Enterprise Data Repository