



Survey Analysis & Churn Risk Detection

Case Study (non-technical): Identifying Intend-to-Churn Customers from Post-Customer Contact Surveyed Verbatim

October 20, 2011

Han Lai

Director, Operational Excellence & Voice of Customer

BUSINESS CHALLENGE

Given large number of customer surveyed verbatim, how can customers' intend-to-churn (ITC) sentiment be quickly and accurately detected

... without having to read 000s customer verbatim feedback?

EXAMPLES OF INTEND-TO-CHURN (ITC) SENTIMENT

" I am going to cancel my account, I only was allowing \$30 to be taken out of my account and \$58.05 was taken out, if that is you term and conditions policy I will never use your service again and I will tell others and have told others not to use your service. "

" Terrible service, no protection for Sellers. Paypal should take responsibility for accounts which have been hacked into by ensuring more sophisticated security. Very disappointed and will never use Paypal again. Explanation from the Paypal Agent was not relevant nor helpful. "

EXAMPLES OF NON OR INCONCLUSIVE INTEND-TO-CHURN SENTIMENT

“Because I had a dispute going on for weeks now. I still havnt got my money back. I received the wrong goods and was asked 2 return the goods recorded delivery which wud have left me even further out of pocket. I tried 2 call on saturday n was left holding for 15 minutes with still no answer so I hung up and as a consequence have had 2 cancel my claim still leavin me out of pocket. Im out raged at the lack of support from paypal! “

“ The website is difficult to navigate when you want to set up an account, link it to your website and provide your bank account information. It took several phone calls to understand how to do this and i am technilogically literate. My account was not on file so I had to go through a whole lot of digital questions before I could get to an agent.”

***BUT FIRST, A QUICK
INTRODUCTION ON PAYPAL***

50 billion connected devices globally in 2020



PayPal will enable payments across devices and channels

PAYPAL BUSINESS VISION

Delivering the future of money today ...

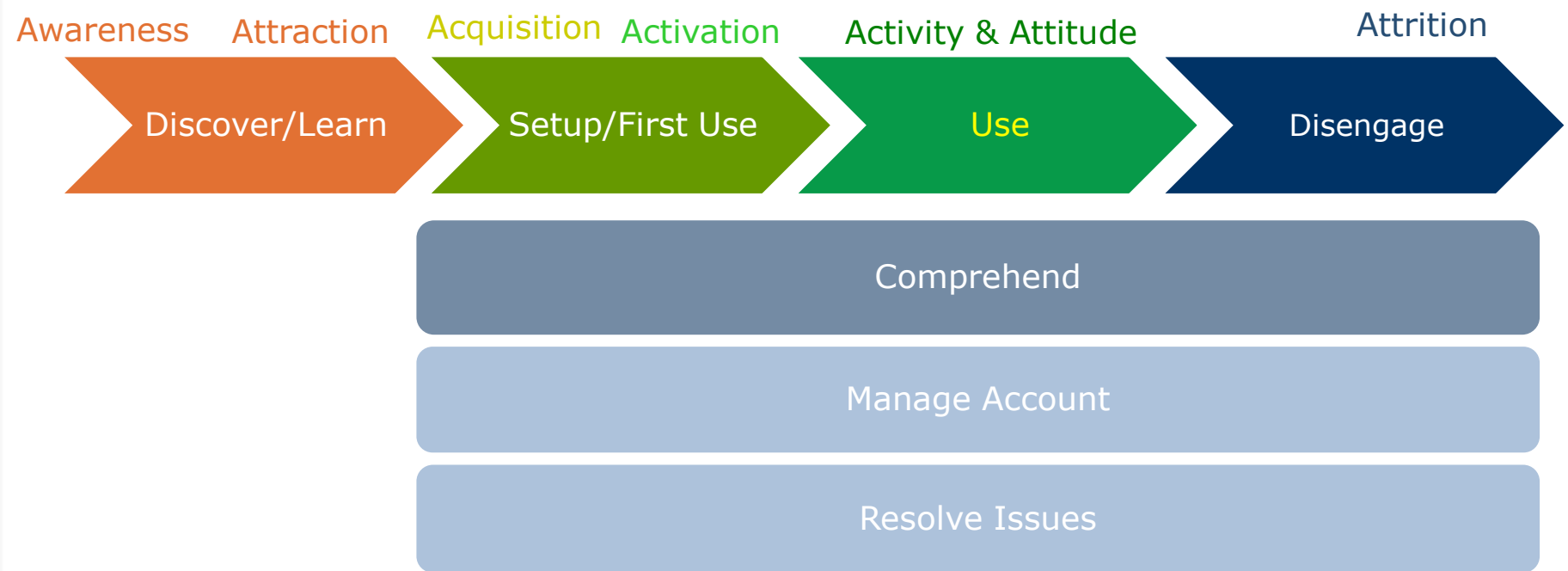
An essential part of our customer's financial
and business lives, enabling secure
commerce anytime, anywhere, any way

A FEW NUMBERS ON PAYPAL

Stats	2011*
Active consumers	~100M
Revenue per quarter	~\$1B
Net promoter score (NPS)	50-55%
# of Employees	~10,000
# of post-CS interaction verbatim comments (English)	~89,000 in Aug

* projection

CONSUMER (BUYER) LIFECYCLE

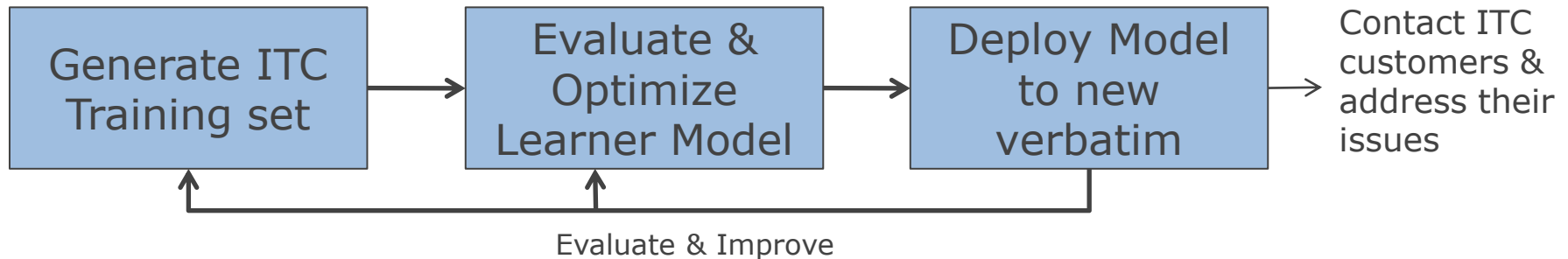


Customer Intent-to-Churn can take place anywhere throughout the lifecycle

TEXT MINING & ANALYTICS AT PAYPAL

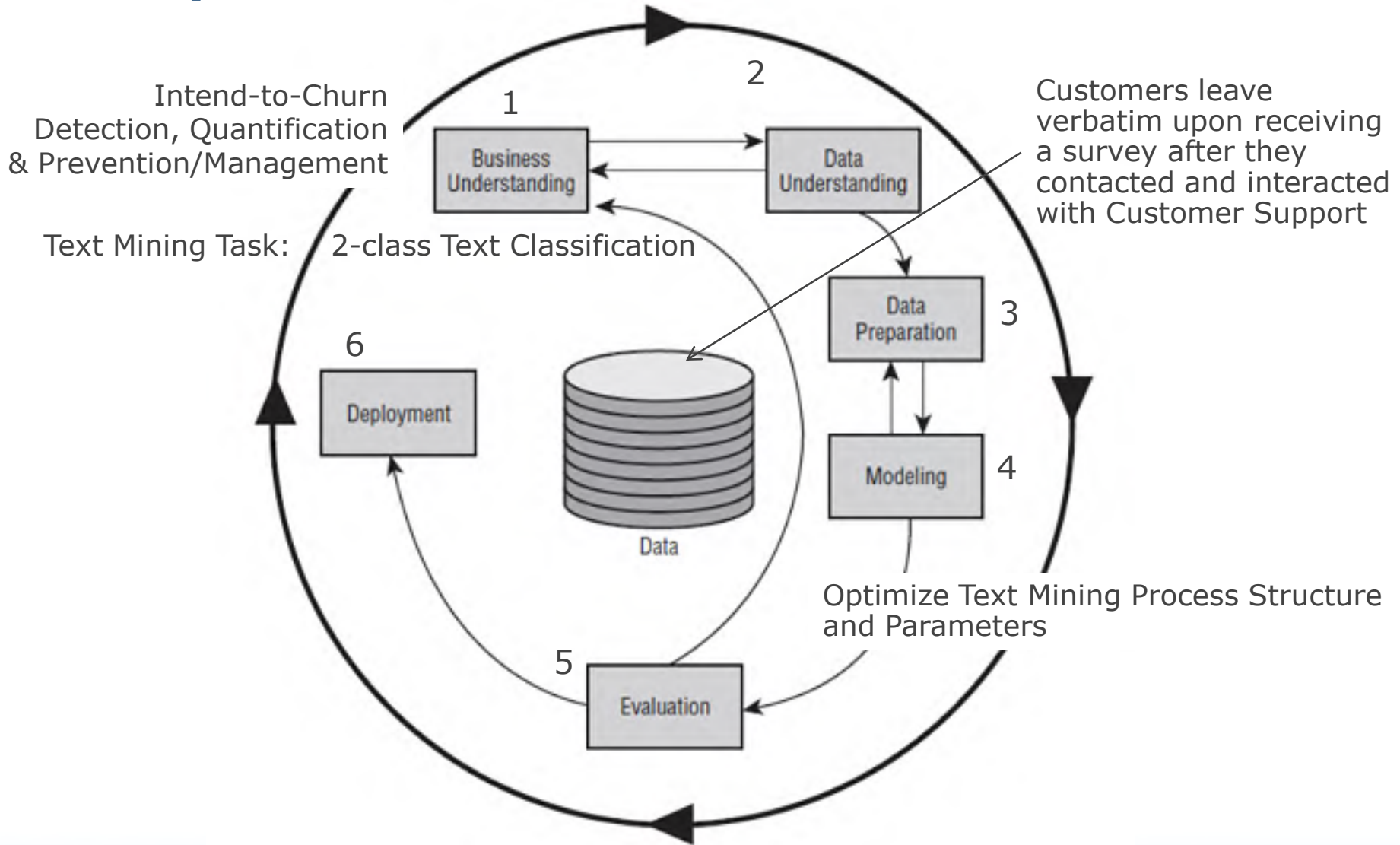
- We are at an early stage of exploring using text mining and analytics
- Used the generated training set for a supervised learning model to classify customer verbatim into at-risk and non-at-risks (so that actions can be taken to prevent churn and learn from their feedback) (tool used: RapidMiner)

CASE STUDY



1. Combined human-classified verbatim and rule/query-based text search to *quickly* generate ~1000 'at-risk' ITC and ~1000 non-ITC verbatim training examples (tool used: Tovek)
2. Used the generated training set for a supervised learning model to classify customer verbatim into at-risk and non-at-risks (so that actions can be taken to prevent churn and learn from their feedback) (tool used: RapidMiner)

WE FOLLOWED THE GENERAL TEXT/DATA MINING APPROACH



TEXT PREPROCESSING & CLASSIFICATION

Document Vector Representation of Verbatim Texts

Text Preprocessing

- **To-Lower-Case-Conversion**
- **Tokenization**
- Stop Word Removal (optional)
- **Term Length Filtering** (optional)
- Term Frequency Filtering (optional)
- Word Stemming (optional)
- **Term N-Grams** (Words plus Word Pairs) (optional)
- Character N-Grams (partial words like 3-grams or 4-grams) (optional)
- ...

Statistical Machine Learning Approaches:

- **Support Vector Machines** (SVM, LibSVM)
- Fast Large Margin (= Special Linear LibSVM Learner*)
- Naive Bayes

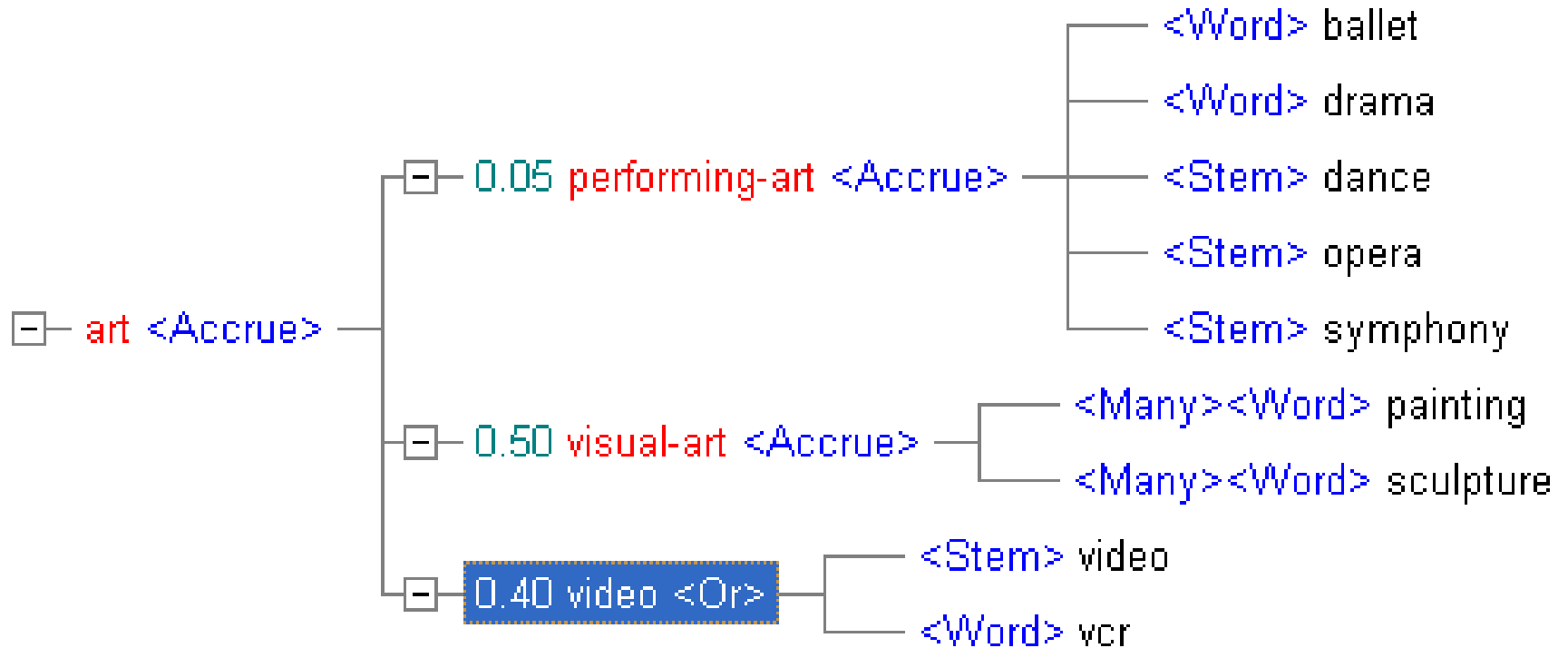
Optimization: find best combination of

- preprocessing options and
- modelling technique and
- parameter settings

* Based on the linear support vector learning scheme proposed by Fan, Chang, Hsieh, Wang, and Lin. Although the result is similar to those delivered by classical SVM or logistic regression implementations, this linear classifier is able to work on data set with millions of examples and attributes

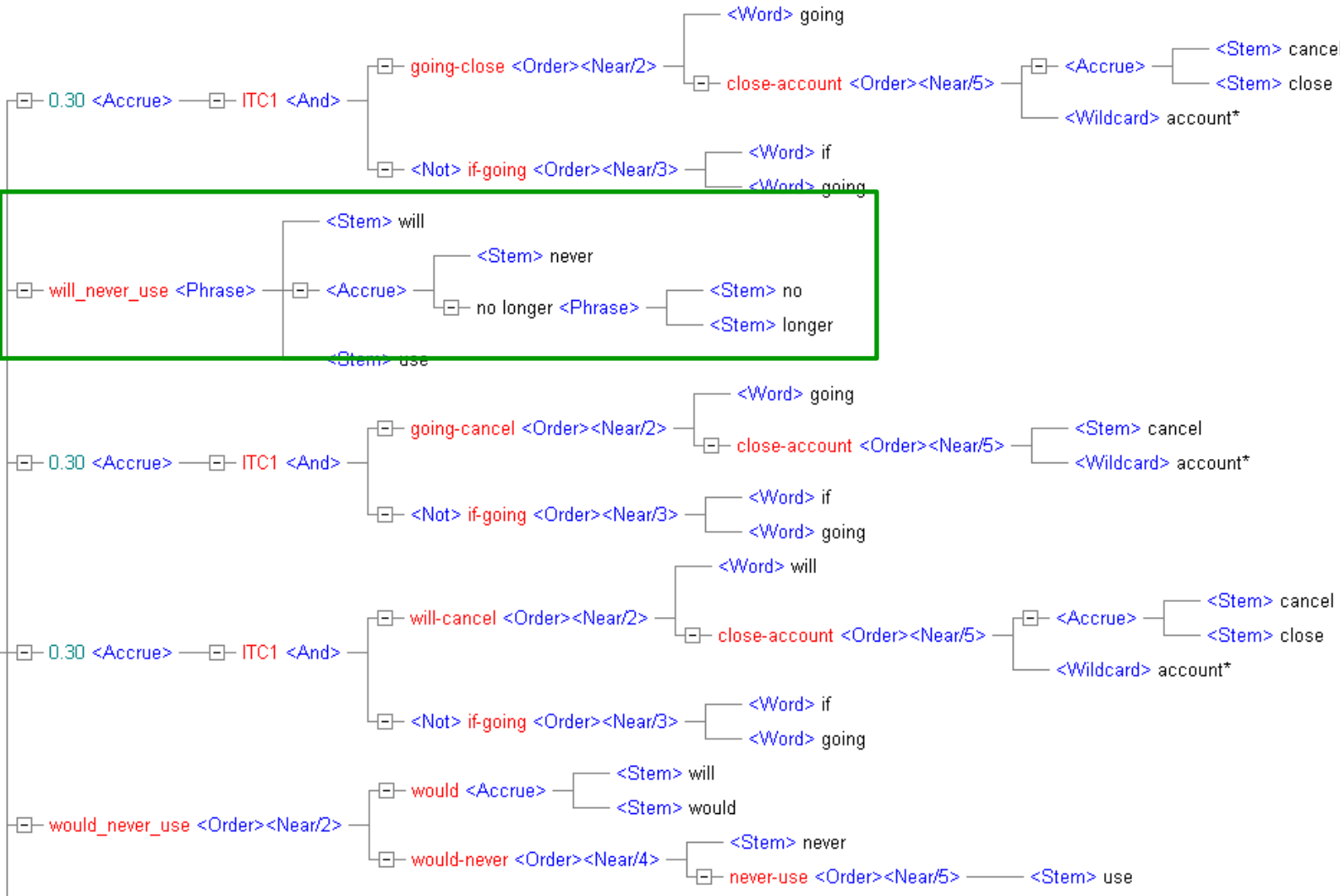
GENERATE TRAINING DATASET BASED ON TEXT FILTERS (TOOL: TOVEK) (SIMPLE EXAMPLE)

art.otl



RULE-BASED TEXT QUERY TO FIND MATCHING INTEND-TO-CHURN VERBATIM FOR TRAINING

Topic Tree for Intend-to-Churn (ITC)



GENERATED ~1000 INTEND-TO-CHURN TRAINING EXAMPLESET OUT OF ~2000 ITC/NON-ITC VERBATIM

Untitled - Tavok Agent

File Edit View Query Tools Help

Brian1
 Brian2
 Brian3
 Detractor
 Detractor Consumer 2
 Files
 Index Consumer Detractor
 ITC
 ITC Test
 small journal entries1
 small journal entries2

<Accrue>{ <Phrase>{ <Stem> not, <Stem> to, <Stem> use, <Stem> PayPal, <Stem> again }, <Accrue>{ <And>{ <Order> <Near/2>{ <Word> going, <Order> <Near/5>{ <Accrue>{ <Stem> cancel, <Stem> close }, <Wildcard> account* } }, <Not> <Order> <Near/3>{ <Word> if, <Word> going } } }, <Phrase>{ <Stem> will, <Accrue>{ <Stem> never, <Phrase>{ <Stem> no, <Stem> longer } }, <Stem> use }, <Accrue>{ <And>{ <Order> <Near/2>{ <Word> going, <Order> <Near/5>{ <Stem> cancel, <Wildcard> account* } }, <Not> <Order> <Near/3>{ <Word> if, <Word> going } } }, <Accrue>{ <And>{ <Order> <Near/2>{ <Word> will, <Order> <Near/5>{ <Accrue>{ <Stem> cancel, <Stem> close }, <Wildcard> account* } }, <Not> <Order> <Near/3>{ <Word> E, <Word> going } } }, <Order> <Near/2>{ <Accrue>{ <Stem> will, <Stem> would }, <Order> <Near/4>{ <Stem> never, <Order> <Near/5>{ <Stem> use } } }, <P
 <Near/4>{ <Accrue>{ <Stem> c
 <Stem> no, <Stem> longer, <St
 account }, <Phrase>{ <Stem> si
 <Near/2>{ <Stem> use, <Order
 <Stem> no, <Stem> need, <Stem> to, <Stem> use, <Stem> use, <Stem> use, <Stem> never, <Stem> using, <Stem> using, <Stem> use }, <Phrase>{ <Stem> no, <Stem> longer, <Stem> interested, <Stem> in, <Stem> using })
 AND not (Paypal close, EBAY CLOSED, reopen <order> <near/2> account, was never able, can close, can cancel, either cancel, either close, without closing, without cancelling)

Query Required = Yes
 Max Number of Documents = 50
 Minimal Score = 0
 Query Type = Single Query

Score	ITC	Document
0.98	ITC	C:\Documents and Se
0.98	ITC	C:\Documents and Se
0.97	ITC	C:\Documents and Se
0.97	ITC	C:\Documents and Se
0.97	ITC	C:\Documents and Se
0.97	ITC	C:\Documents and Se
0.96	ITC	C:\Documents and Se
0.95	ITC	C:\Documents and Se

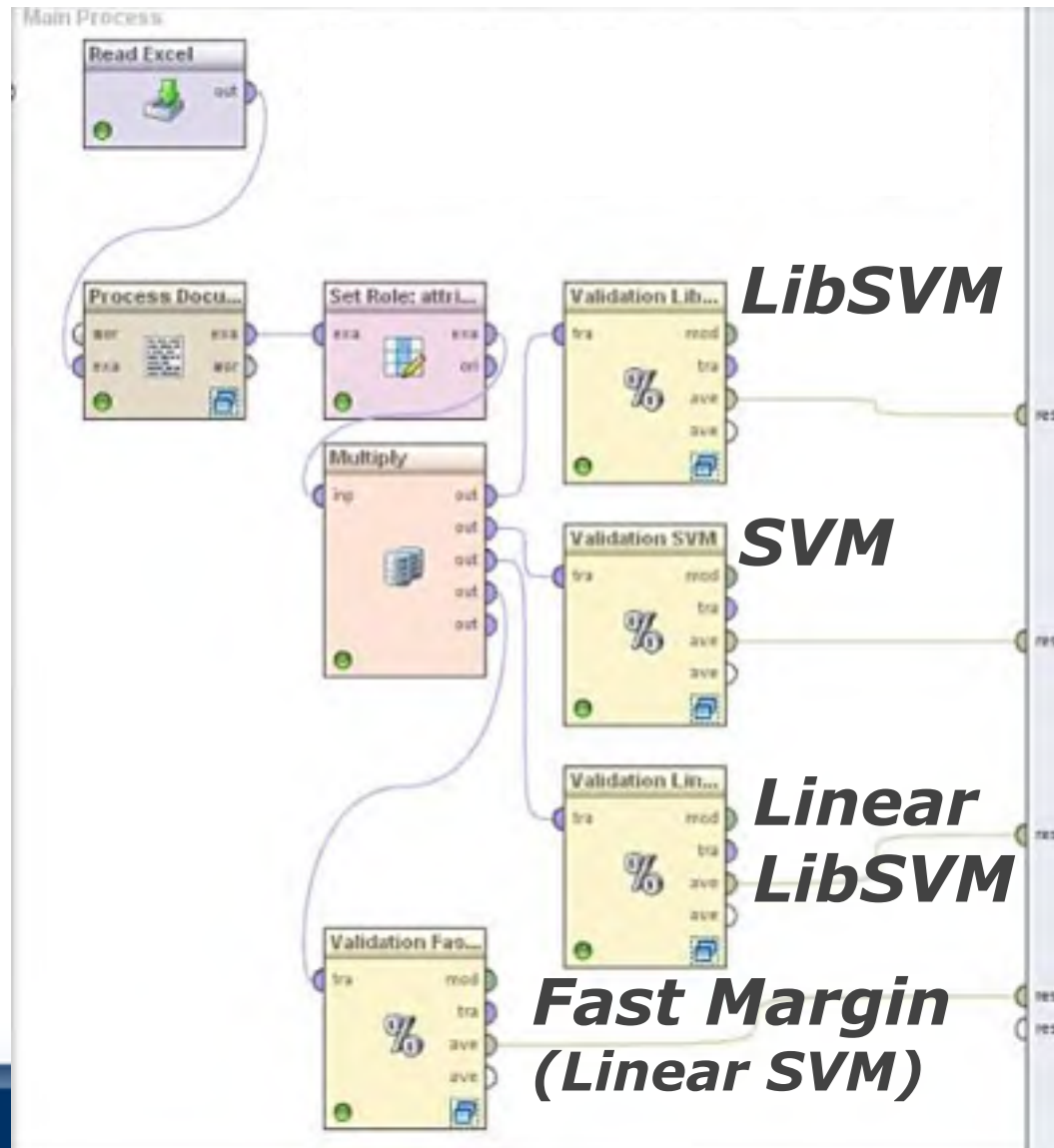
Generated labeled Training exampleset exported to Excel

ITC, "I am going to cancel my account. I only was allowing \$30 to be taken out of my account and \$58.05 was taken out, if that is you term and conditions policy I will never use your service again and I will tell others and have told others not to use your service. When you went into my account and with drew the extra \$28.05 you over drew my account. So thank you for making my purchase of \$28.05 cost \$58.05 with my overdraft fees, and thank you for not using my default card. Once you have moved my money from my pay pal account back to my bank account you will be hearing from me again to close my account and delete all my info. As for the people who helped me they were no help they only help they gave me was to tell that I was shit out of luck and you can take money out of my account and there is nothing that I could do about it because that is your policy. As of now the only thing I want done is my \$30 overdraft fee paid. I am sure you do not care about what I have to say just like your customer service reps, so I sure I will not here back from anyone about this."

Ready

50 1 030 2 056

INITIAL RAPIDMINER TRAINING EVALUATION PROCESS: CROSS-VALIDATION BASED COMPARISON OF MODELLING TECHNIQUES



INITIAL EVALUATION PROCESS: CROSS-VALIDATION RESULTS FOR DIFFERENT MODEL TECHNIQUES (CONFUSION MATRIX)

Multiclass Classification Performance Annotations

Table View Plot View

LibSVM

accuracy: 78.9273% +/- 3.6723% (mikro: 78.9292%)

	true NON ITC	true ITC
pred. NON ITC	773	9
pred. ITC	235	141
class recall	76.6865%	94.0000%

Table View Plot View

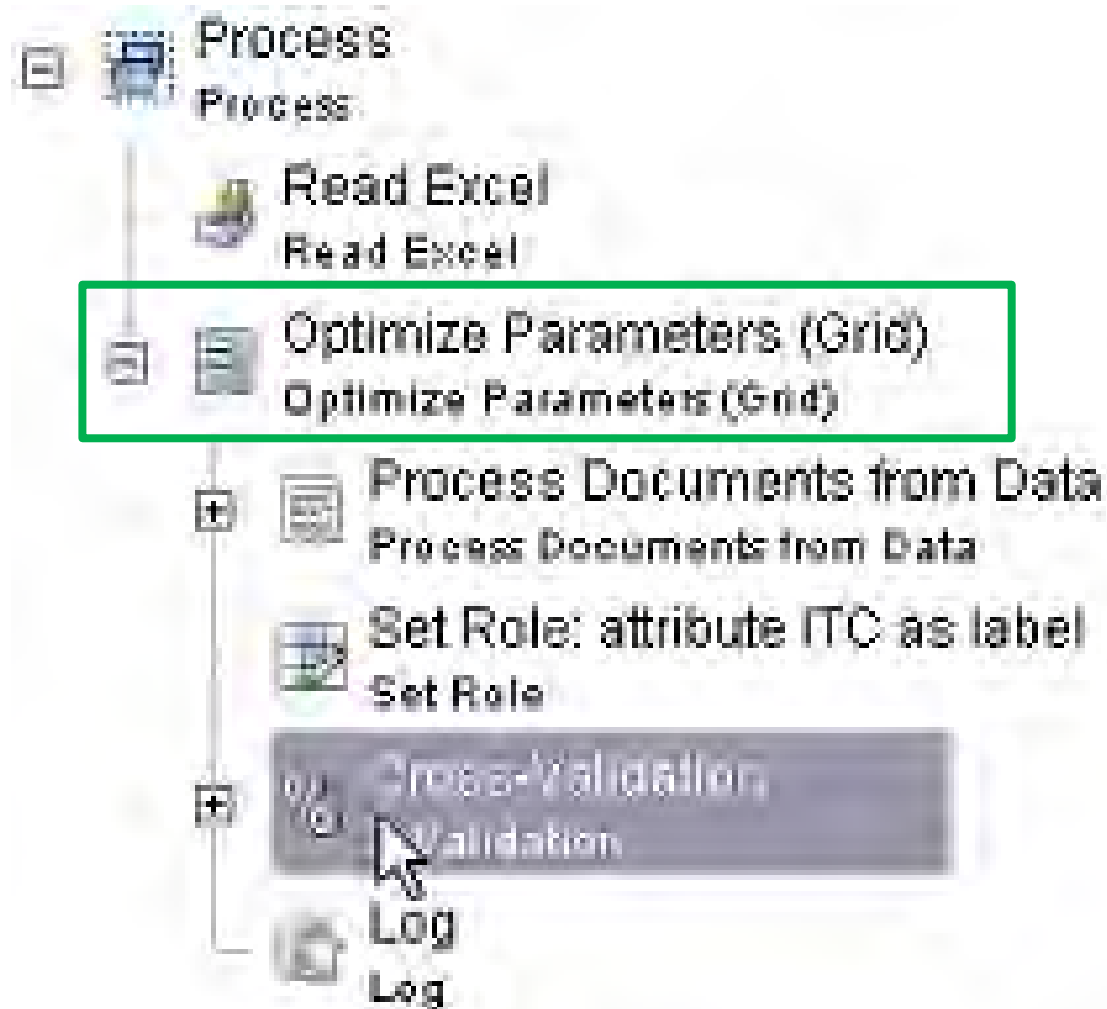
Fast Large Margin (Linear LibSVM)

accuracy: 72.1087% +/- 3.0211% (mikro: 72.1071%)

	true NON ITC	true ITC
pred. NON ITC	696	11
pred. ITC	312	139
class recall	69.0476%	92.6667%

***There is room for accuracy improvement:
Optimization Process: Automatically Find
the Process with the Highest Accuracy***

Optimization Process: Automatically Find the Process with the Highest Accuracy



Optimization Process: specifying parameter values to be optimized and logged



Edit Parameter List **log**

List of key value pairs where the key is the column name and the value specifies the process value to log.

column name		value
Minimum_Token_Length	Text Preprocessing options	Filter Tokens (by Length) parameter min_chars
Remove_Stopwords		Select Subprocess: Remove Stopwords parameter select_which
Word_Stemming		Select Subprocess: Word Stemming parameter select_which
N-Grams		Select Subprocess: N-Grams parameter select_which
Learner	Select Subprocess: Select Learner parameter select_which	
LibSVM_C	Modeling techniques	SVM (LibSVM) parameter C
SVM_C		SVM (with cost) parameter C
Fast_Large_Margin_C		Fast Large Margin (SVM) parameter C
Accuracy	Performance metrics & other measurements	Classification Performance value accuracy
CPU_Execution_Time		Classification Performance value cpu-execution-time
Precision		Classification Performance value weighted_mean_precision
Recall		Classification Performance value weighted_mean_recall

Buttons: Add Entry, Remove Entry, Ok, Cancel

RESULTS FOR DIFFERENT PROCESS DESIGNS AND PARAMETERS: BEST PROCESS DESIGN ACHIEVED ~95-96% ACCURACY

P_219_Optimize_Text_Classification_LibSVM_vs_SVM_vs_Fast_Large_Margin_Cross-Validation_Cost-Sensitive* - RapidMiner@DBZM

File Edit Process Tools View Help

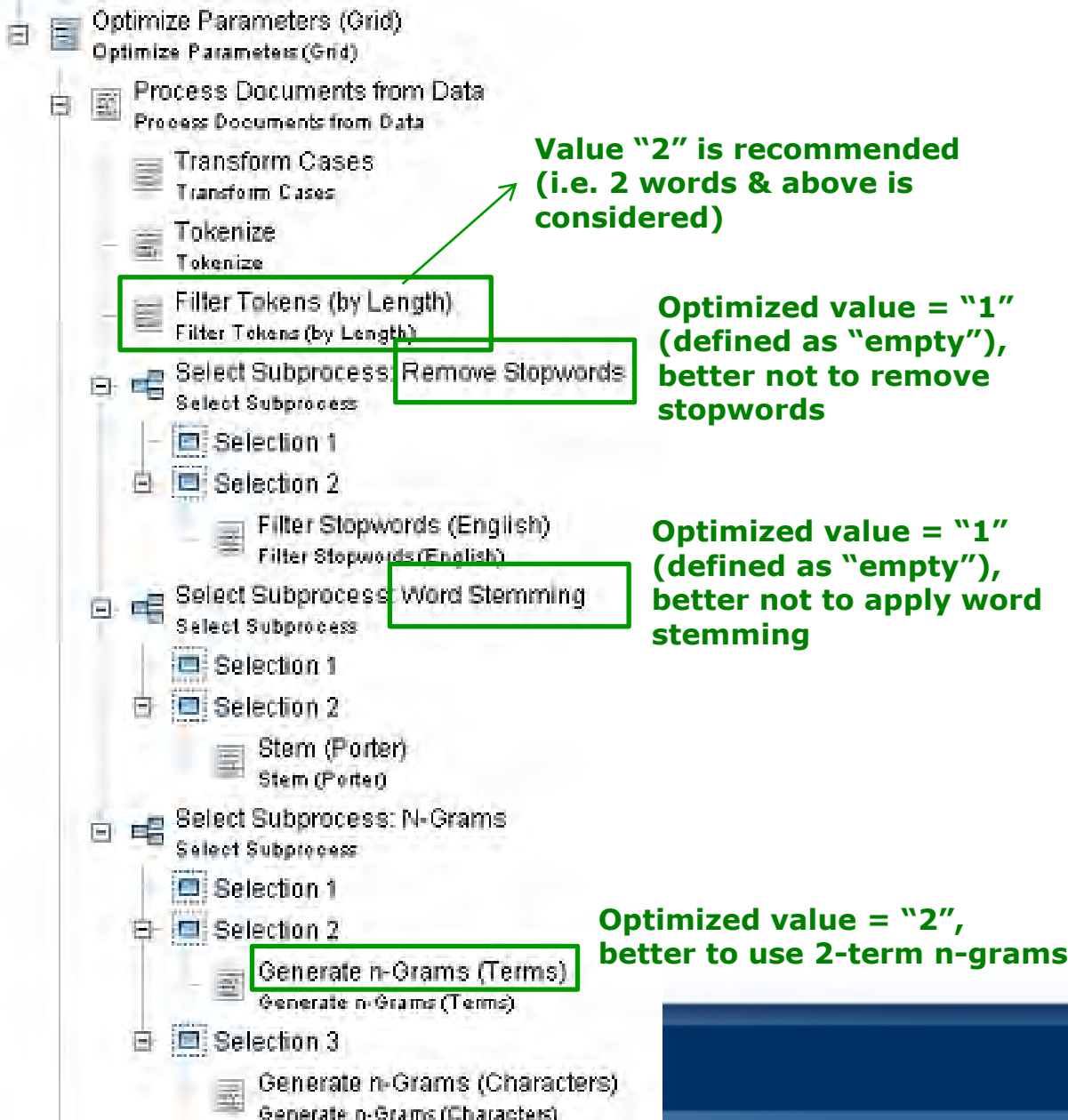
Result Overview Log

Table View Plot View **Text preprocessing** **Learner** **Parameters**

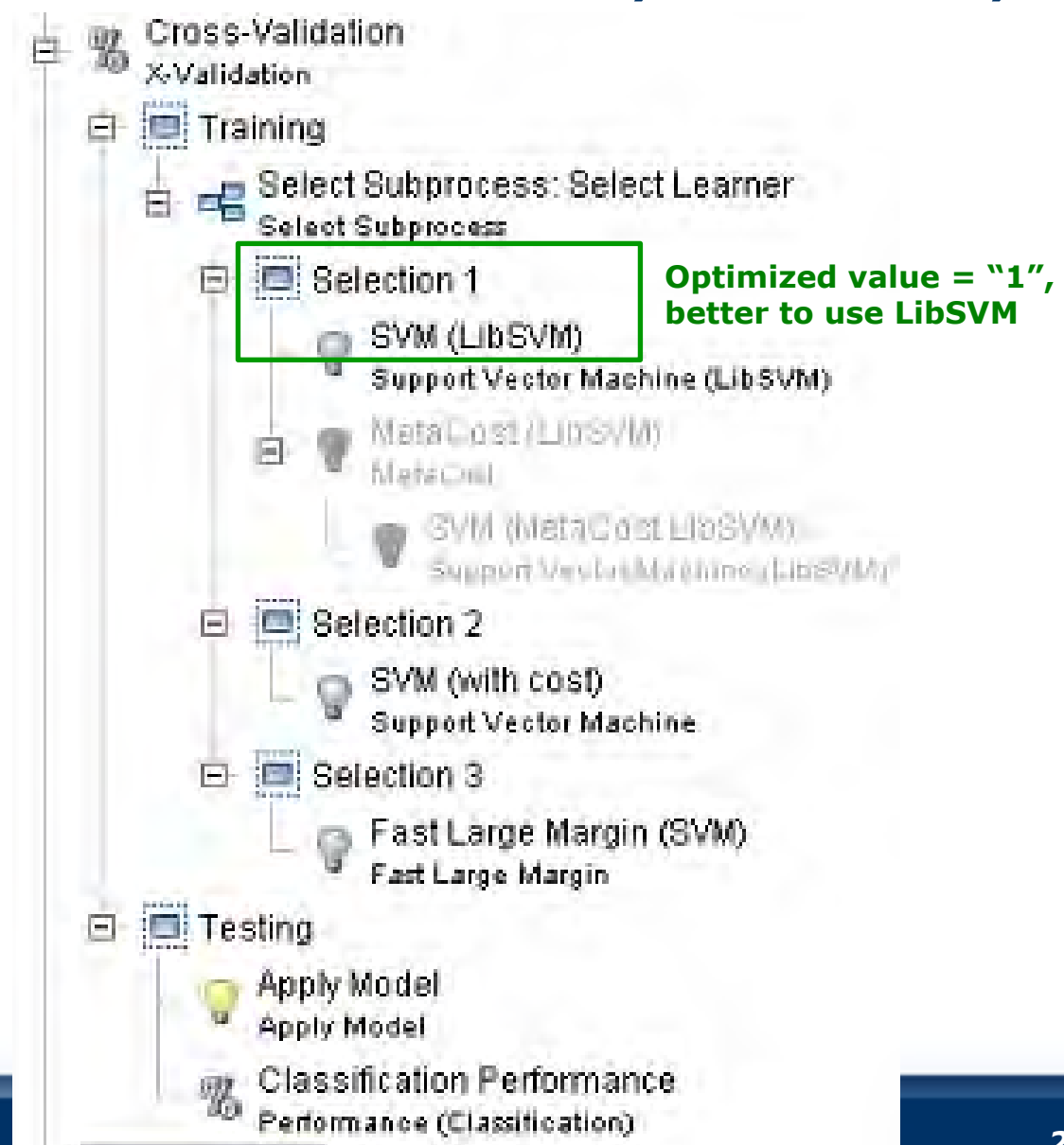
Log (127 rows, 12 columns)

Minimum_Token_Length	Remove_Stopwords	Word_Stemming	N-Gram_Method	Learner	LibSVM_C	SVM_C	Fast_Large_Margin_C	Accuracy
2	1	1	2	1	0	0	1	0.96585
2	1	1	2	1	0	0	1	0.95610
1	1	2	2	1	0	0	1	0.95122
3	1	1	2	1	0	0	1	0.94634
3	1	2	2	1	0	0	1	0.94146
1	1	2	2	1	0	0	1	0.93659
1	1	1	2	1	0	0	1	0.93659
3	1	1	2	1	0	0	1	0.93659
3	1	2	2	1	0	0	1	0.93171
1	1	1	2	1	0	0	1	0.92683
2	1	1	1	1	0	0	1	0.92195
2	1	2	2	1	0	0	1	0.92195
2	2	2	2	1	0	0	1	0.92195

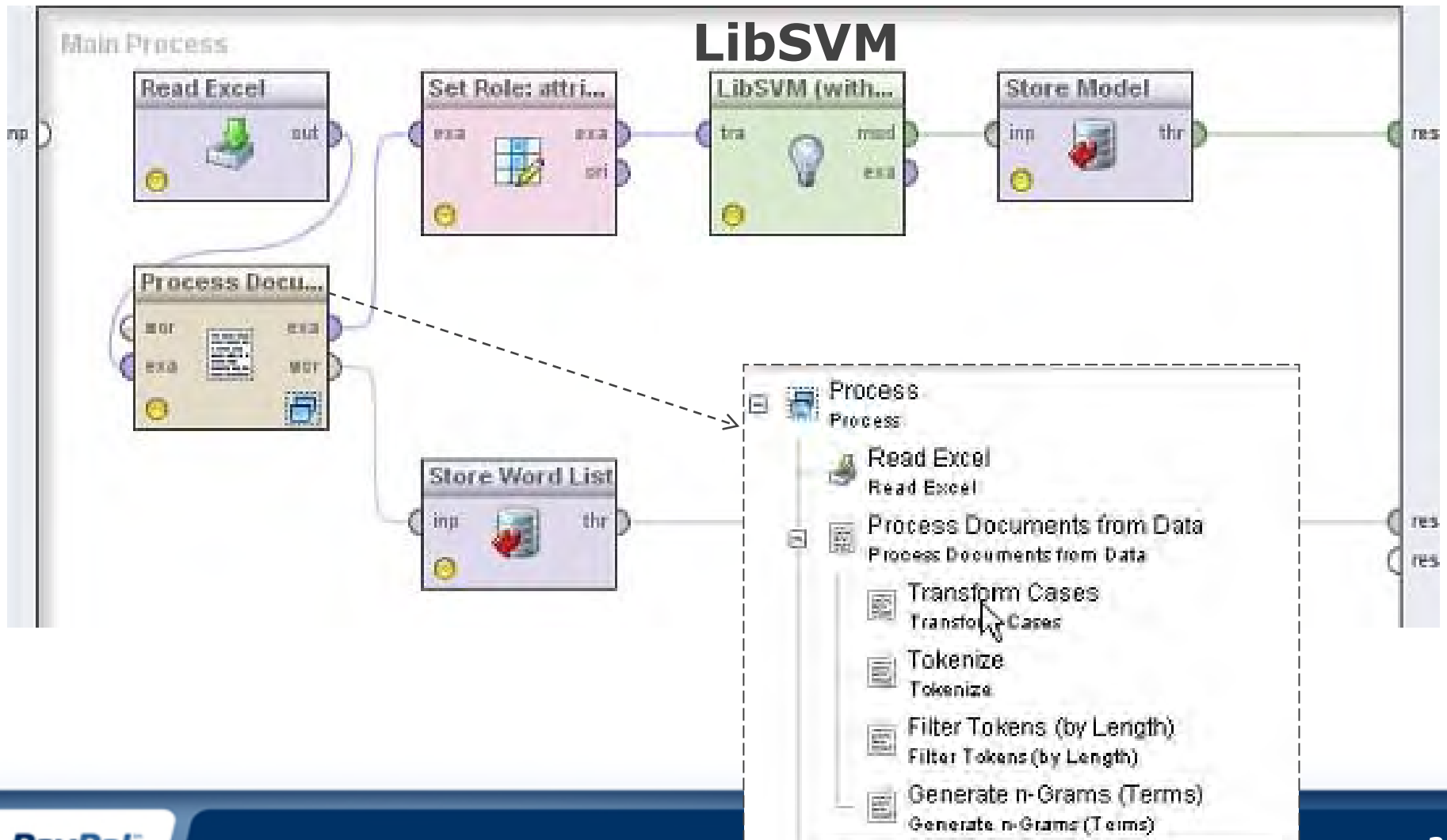
OPTIMIZATION PROCESS: TEXT PREPROCESSING STEPS/OPTIONS



RAPIDMINER OPTIMIZATION PROCESS: MODELLING OPTIONS INCL. SVM, LIBSVM, LINEAR LIBSVM

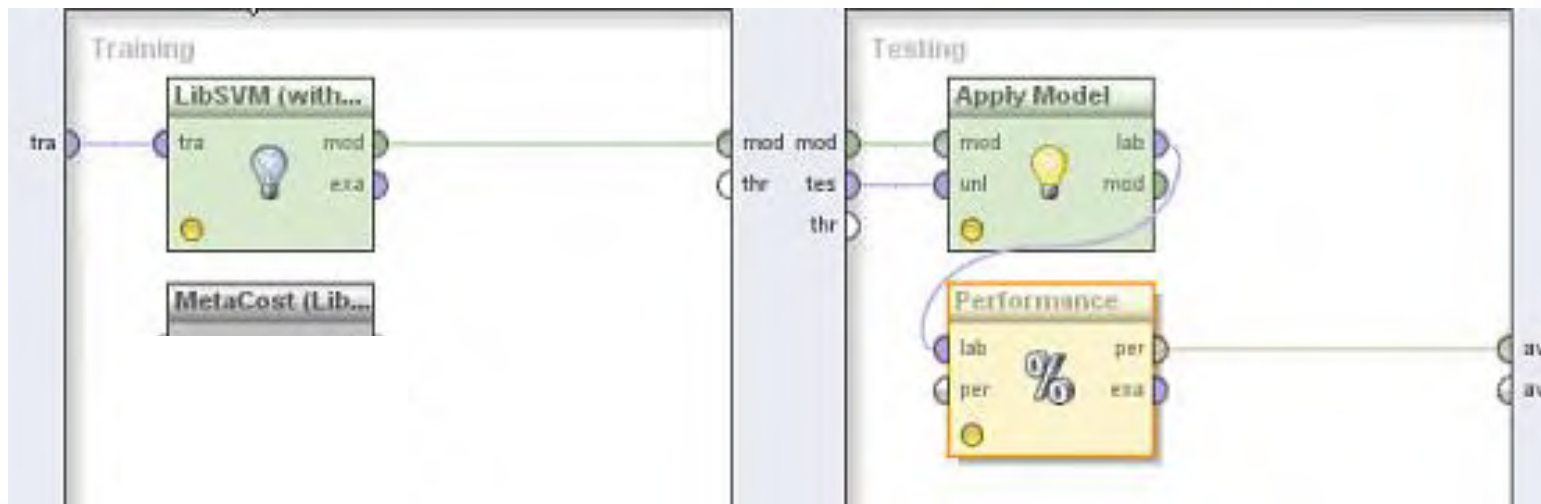


RAPIDMINER TRAINING OF THE MODEL WITH OPTIMIZED TEXT PREPROCESSING, MODELLING TECHNIQUE, AND PARAMETERS

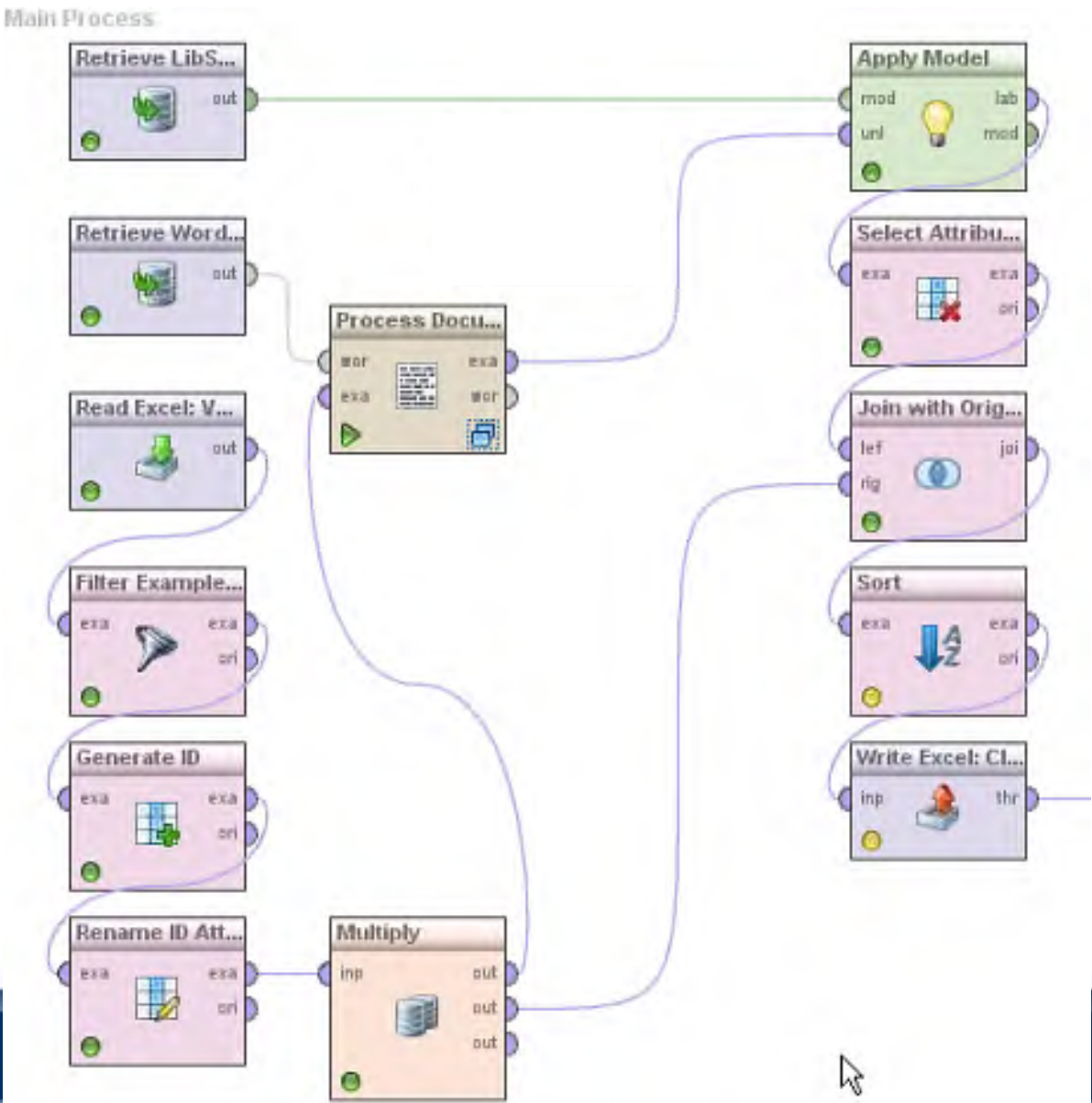


DEPLOY MODEL WITH OPTIMIZED TEXT PREPROCESSING, MODEL TECHNIQUE & PARAMETERS

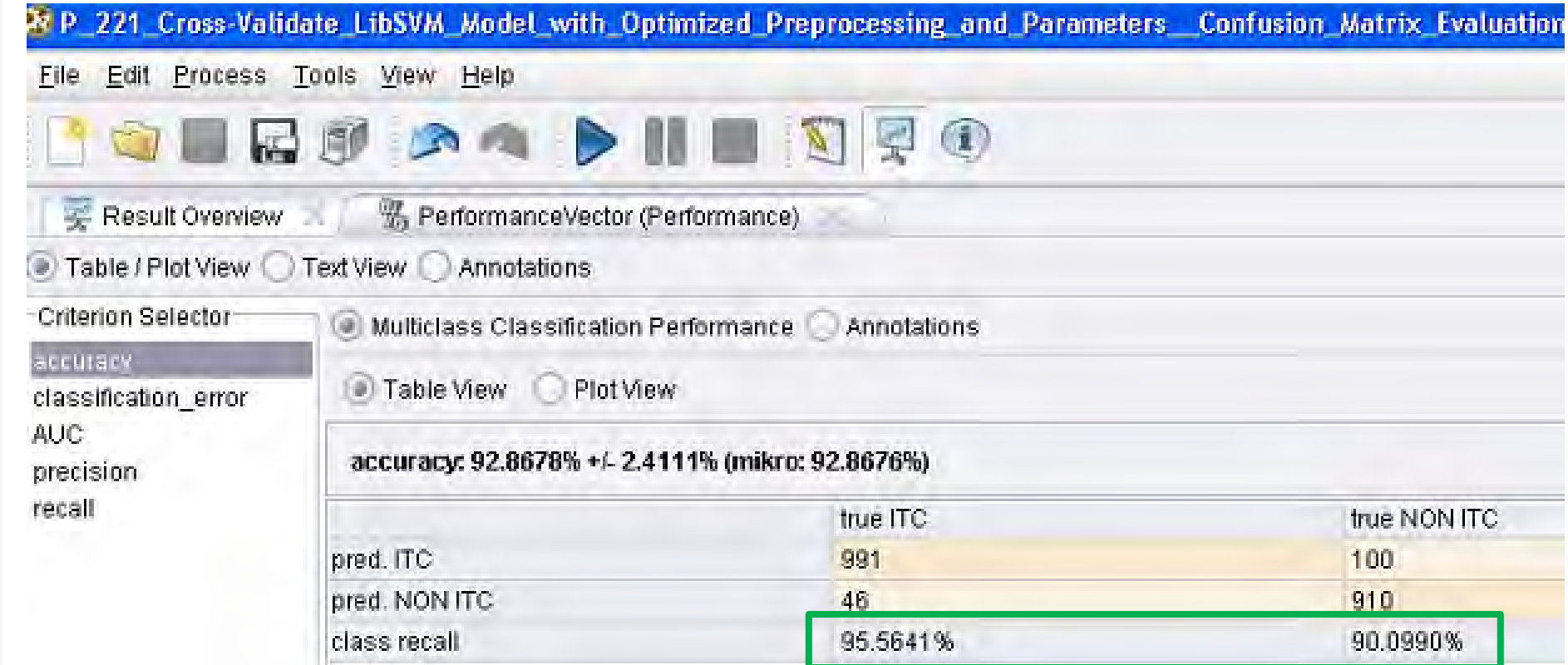
Main Process



Deployment of the Optimized Model: Retrieve „Optimized“ Model, then Classify, Score, Rank and write New ITC Cases



IMPROVED PERFORMANCE OF MODEL WITH OPTIMIZED TEXT PREPROCESSING, MODEL TECHNIQUE & PARAMETERS



RESULTS OF DEPLOYED MODEL WITH OPTIMIZED PARATMETERS ON NEW VERBATIM – ITC CLASSIFIED, SCORED AND RANKED

P_222_Deploy_LibSVM_Model_with_Optimized_Preprocessing_and_Parameters_on_New_Data_and_Store_Scored_T

File Edit Process Tools View Help

Result Overview ExampleSet (Sort by Decreasing Confidence in ITC Class Membership)

Meta Data View Data View Plot View Annotations

ExampleSet (857 examples, 4 special attributes, 1 regular attribute)

Role	Name	Type	Statistics
confidence_ITC	confidence(ITC)	real	avg = 0.44036 +/- 0.12095 [0.13673 ; 0.83464]
id	Verbatim_Text_ID	integer	avg = 429 +/- 247.53889 [1.00000 ; 857.00000]
confidence_NON ITC	confidence(NON ITC)	real	avg = 0.55964 +/- 0.12095 [0.16536 ; 0.86327]
prediction	prediction(ITC)	binominal	mode = NON ITC (599), least = ITC (258) [ITC (258), NON ITC (599)]
regular	Verbatims	text	mode = Because a mistake was made and 1 dont use paypal often and

MANUALLY-VERIFIED ITC ACCURACY WITH MODERATELY SUCCESSFUL RESULTS

P_222_Deploy_LibSVM_Model_with_Optimized_Preprocessing_and_Parameters_on_New_Data_and_Store_Scored_Texts_to_Excel_Sheet - RapidMiner@DB7M

File Edit Process Tools View Help

Result Overview ExampleSet (Sort by Decreasing Confidence in ITC Class Membership)

Meta Data View Data View Plot View Annotations

ExampleSet (857 examples, 4 special attributes, 1 regular attribute) View Filter (857)

Row No.	Verbatim_Text_ID	prediction(ITC)	confidence(ITC)	confidence(NON ITC)	Verbatims
1	41	ITC	0.16536	0.16536	because a mistake was made and it could not be corrected in any way. I find this unac
2	435	ITC	0.77050	0.22950	would tell my friends to NEVER use PayPal! I think that the way you guys handle dis
3	721	ITC	0.75645	0.24355	he Because a mistake was made and it could not be corrected in any way. I find this
4	424	ITC	0.75261	0.24739	W service rep was rude. I will be cancelling my paypal account as soon as I can.
5	852	ITC	0.75144	0.24856	our s
6	218	ITC	0.74741	0.25259	am ve
7	426	ITC	0.73111	0.26889	WILL
8	32	ITC	0.73108	0.26891	s i ha
9	681	ITC	0.73108	0.26892	embla
10	680	ITC	0.73108	0.26892	embla
11	387	ITC	0.73105	0.26895	think the whole thing is a fraud, I have paid my bill and will never use again, I thought
12	627	ITC	0.73104	0.26896	paypal, on at least two different occasions, look money out of the wrong accounts, lea
13	19	ITC	0.73104	0.26896	fter my first experience with ebay and paypal I unless it is a life or death situation I wil
14	358	ITC	0.73102	0.26898	purchased products under bill me later using my husbands accountthe product was.
15	370	ITC	0.72623	0.27377	set up a simple account and I was not aware of the terms and conditions on a nonpr
16	230	ITC	0.71942	0.28058	I cant recommend a service that demands unnecessary personal information. I am rt
17	425	ITC	0.71364	0.28636	I will NEVER recommend Paypal to anyone. Paypal has cost me over \$320 in my che
18	613	ITC	0.71004	0.28996	PayPal locked me out of my account for years. hi your representatives were unable to

Chosen Threshold (conf. => 0.72)

13 out of 15 of classified detractor verbatim are verified Intend-to-Churn (ITC)

Example ITC verbatim predicted by the optimized learner model

OUTLOOK ON DEPLOYMENT OF THE OPTIMIZED MODEL:

CONTACT HIGHEST RANKED ITC CASES ONLY (E.G. CONFIDENCE SCORE => 0.7)

- Typical application: budget to contact the 100 most highly ranked ITC candidates out of the ~200 cases classified as ITC
- Improvement: higher precision in ITC prediction leading to higher customer retention (given the same limited budget)

SUMMARY: TEXT MINING APPROACH

1. Generate ~2000 training data with ITC and non-ITC verbatim texts (labeled semi-manually)
2. Cross-Validate initial text mining process on training data (70-80% average accuracy)
3. Optimize text mining process: find best text processing steps, best modelling technique, and best parameters automatically (128 combinations tested out of more than 5000 => LibSVM on words + word pairs, term filtering => ~96% accuracy)
4. Cross-Validate: compute confusion matrix and performance metrics (accuracy, precision, recall)
5. Train: generate model for automated classification with best process structure and parameters
6. Deploy: apply model trained with best parameters to new verbatim texts, automatically classify into ITC or non-ITC, and rank by confidence score
7. Verify: manual evaluation yields ~85%% accuracy on ~850 new unclassified verbatim data with ≥ 0.72 confidence scores
8. Future: use automatically optimized threshold on the confidence score to maximize benefit

Tools used: Tovek tool for step 1 and RapidMiner for steps 2-6.

THANK YOU FOR YOUR ATTENTION

HLAI@PAYPAL.COM

APPENDIX

Exported Classified New Potential ITC Cases Ranked by Confidence Scores

Microsoft Excel - Deployment_Data_ITC_vs_Non-ITC_auto_classified_by_learned_and_optimized_model_20111012.xls

File Edit Insert Format Tools Data Window Help

IBM Cognos 8

Verbatim

A	B	C	D	E	F
Verbatims	confidence(ITC)	Verbatim_Text ID	confidence(NON ITC)	prediction(ITC)	
2. Because a mistake was made and it could not be corrected in any way. I find this unacceptable. The cu	.8	41.0		2 ITC	
3. I would tell my friends to NEVER use PayPal! I think that the way you guys handle disputes is ridiculous	.8	435.0		2 ITC	
4. The security key still cannot be used & you have not let me know what's happening or when I will be	.8	721.0		2 ITC	
5. I WILL NEVER EVER BUY ANYTHING USING PAYPAL AGAIN . I RECENTLY BOUGHT SOMETHING .	.8	424.0		2 ITC	
6. Your sole interest is getting customers to accept your credit card or to provide you with a checking acco	.8	952.0		2 ITC	
7. I am very upset with the confusion between paypal and smart connect. I spoke to customer service and v	.7	218.0		3 ITC	
8. I WILL NEVER USE PAY PAL - EVER! I DID NOT CHOOSE TO USE PAY PAL THIS TIME BECAUSE	.7	426.0		3 ITC	
9. As I have mentioned it is my first time using this online purchasing and my luck was so bad I would neve	.7	32.0		3 ITC	
10. Terrible service, no protection for Sellers. Paypal should take responsibility for accounts which have bee	.7				
11. terrible service no support total lies when say dont want to use PayPal but use credit card direct as Payf	.7				
12. I think the whole thing is a fraud, I have paid my bill and will never use again, I thought you were chargi	.7				
13. Paypal, on at least two different occasions, took money out of the wrong accounts, leading to errors in re	.7				
14. After my first experience with ebay and paypal I unless it is a life or death situation I will never use eith	.7				
15. I purchased products under bill me later using my husbands account the product was not in stock yet i w	.7				
16. I set up a simple account and I was not aware of the terms and conditions on a nonprofit and it was my r	.7				
17. I cant recommend a service that demands unnecessary personal information. I am referring to the policy	.7				
18. I will NEVER recommend Paypal to anyone. Paypal has cost me over \$320 in my checking account and	.7				
19. PayPal locked me out of my account for years b/c your representatives were unable to verify my identity	.7				
20. the idea of pay pal is good and the personnel that I have dealt with are courteous and professional. howe	.7				
21. I would not recommend paypal, I had fraud on my account & getting it resolved & my card repl	.7				
22. I have perfect feedback I do same day shipping and my product is always what it is described and now p	.7	323.0		3 ITC	
23. I sign up and when it was approved I received your ok with in minutes I got a message from you that I ha	.7	371.0		3 ITC	
24. Because I am disgusted with the way I have been treated, unfortunately for me, I had to change my email	.7	47.0		3 ITC	
25. Because I had a dispute going on for weeks now. I still havnt got my money back. I received the wrong g	.7	54.0		3 ITC	
26. I was calling about an item I had purchased & returned because it was not as described. I paid \$23	.7	404.0		3 ITC	
27. Your system is illogical, not user-friendly, and absolutely idiotic. Believe me, in this age of rampant hack	.7	655.0		3 ITC	
28. I somehow got Paypal Smart Connect I didnt even sign up for it and I cannot see off the balance on it bu	.7	774.0		3 ITC	

Manually verified ~85% ITC accuracy deployed on new verbatim data (for confidence score of >= 0.72)