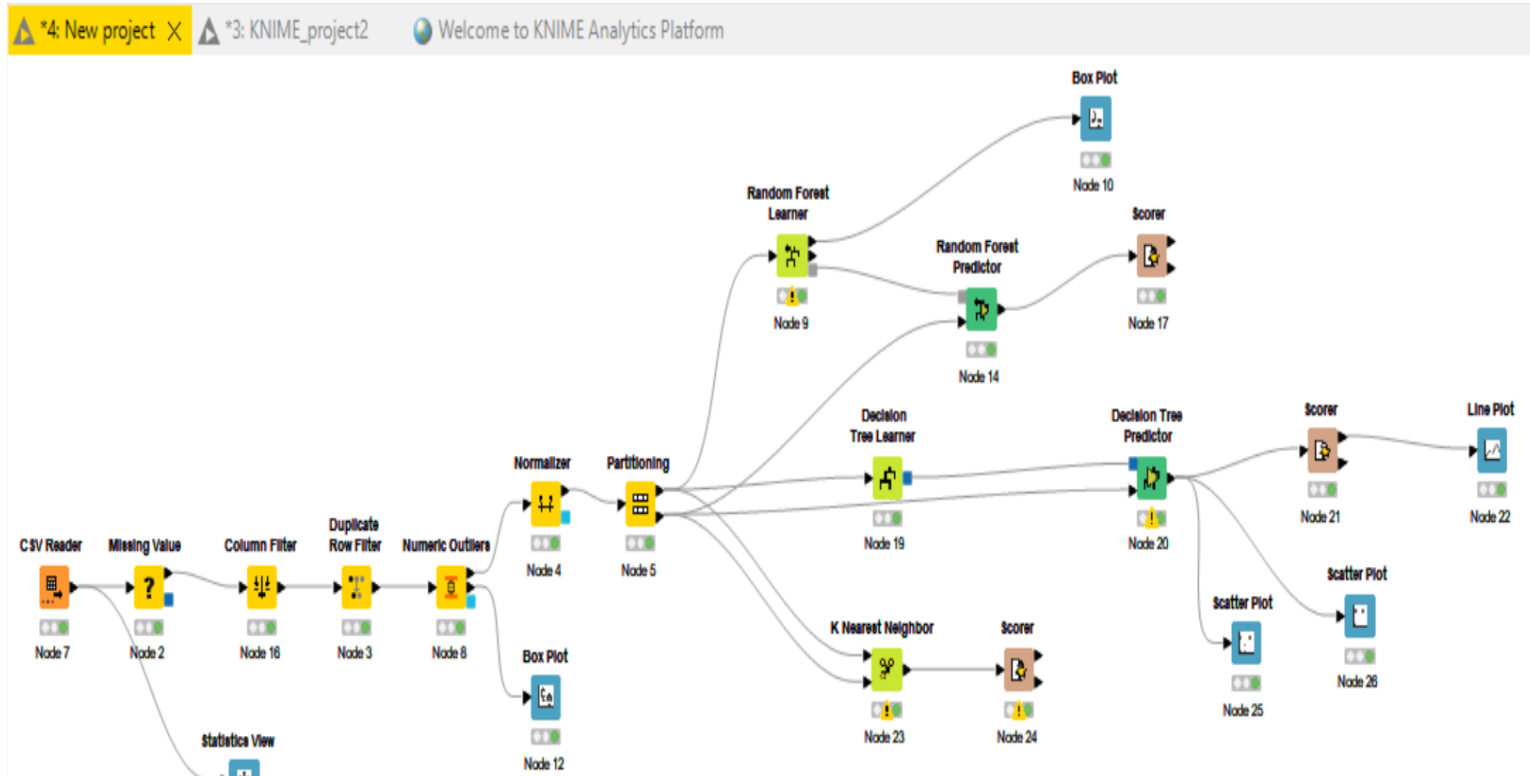


PHISHING URL KNIME PROJECT

WORKFLOW



STATISTIC VIEW

Statistics

Rows: 11-20 of 56 | Columns: 14 < >

Search 🔍

Name	Type	# Mi...	# Uni...	Mini...	Maxi...	25% ...	50% ...	75% ...	Me	🔍
URLCharProb	Number (double)	0	227421	0.001	0.091	0.051	0.058	0.063	0.056	
TLDLength	Number (integer)	0	12	2	13	2	3	3	2.764	
NoOfSubDomain	Number (integer)	0	10	0	10	1	1	1	1.165	
HasObfuscation	Number (integer)	0	2	0	1	0	0	0	0.002	
NoOfObfuscatedChar	Number (integer)	0	20	0	447	0	0	0	0.025	
ObfuscationRatio	Number (double)	0	146	0	0.348	0	0	0	0	
NoOfLettersInURL	Number (integer)	0	421	0	5,191	10	14	20	19.429	
LetterRatioInURL	Number (double)	0	709	0	0.926	0.435	0.519	0.594	0.516	

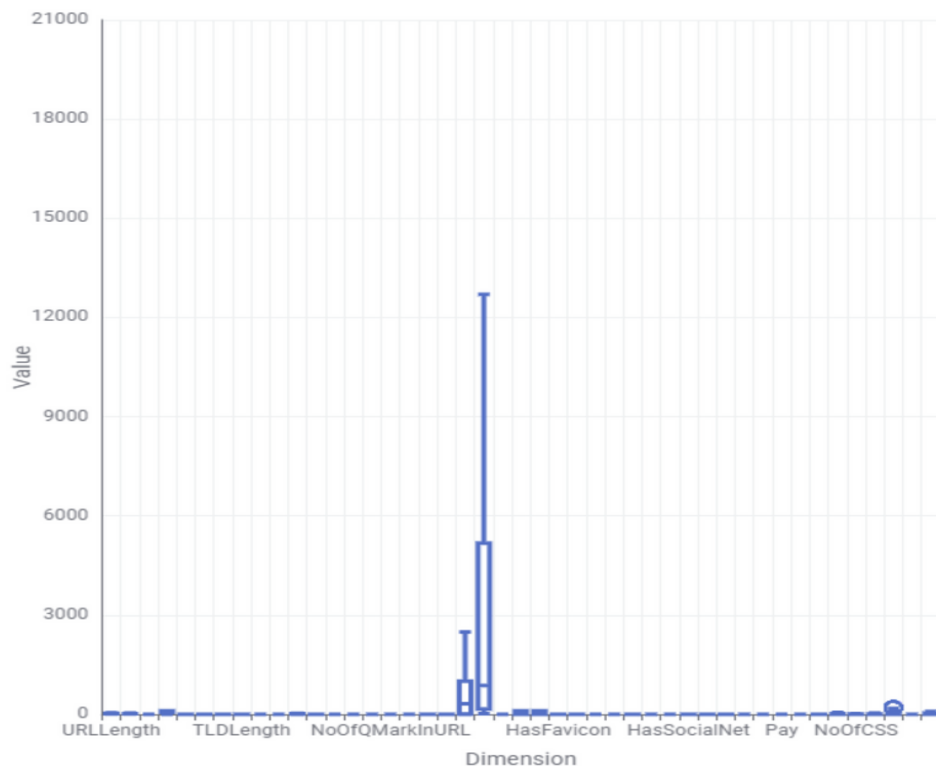
OUTLIER DETECTION AND REMOVAL

File Edit Hilite Navigation View

Table "default" - Rows: 235795 Spec - Columns: 50 Properties Flow Variables

Row ID	S URL	D URLLen...	S Domain	D Domain...	D IsDoma...	S TLD	D URLSimi...	D CharCo...	D TLDLegi...	D URLCh...
Row0	https://www.southbankmosaics.com	31	www.southbankmosaics.com	24	0	com	100	1	0.523	0.062
Row1	https://www.uni-mainz.de	23	www.uni-mainz.de	16	0	de	100	0.667	0.033	0.05
Row2	https://www.voicefmradio.co.uk	29	www.voicefmradio.co.uk	22	0	uk	100	0.867	0.029	0.064
Row3	https://www.sfmjournal.com	26	www.sfmjournal.com	19	0	com	100	1	0.523	0.058
Row4	https://www.rewildingargentina.org	33	www.rewildingargentina.org	26	0	org	100	1	0.08	0.059
Row5	https://www.globalreporting.org	30	www.globalreporting.org	23	0	org	100	1	0.08	0.061
Row6	https://www.saffronart.com	25	www.saffronart.com	18	0	com	100	1	0.523	0.064
Row7	https://www.nerdscandy.com	25	www.nerdscandy.com	18	0	com	100	1	0.523	0.06
Row8	https://www.hyderabadonline.in	29	www.hyderabadonline.in	22	0	in	100	1	0.005	0.057
Row9	https://www.aap.org	18	www.aap.org	11	0	org	100	1	0.08	0.07
Row10	https://www.religionenlibertad.com	33	www.religionenlibertad.com	26	0	com	100	1	0.523	0.064
Row11	http://www.teramill.com	22	www.teramill.com	16	0	com	82.645	1	0.523	0.067
Row12	https://www.socialpolicy.org	27	www.socialpolicy.org	20	0	org	100	1	0.08	0.064
Row13	https://www.aoh61.com	20	www.aoh61.com	13	0	com	100	1	0.523	0.055
Row14	https://www.bulgariaski.com	26	www.bulgariaski.com	19	0	com	100	1	0.523	0.056
Row15	https://www.brightika.com	24	www.brightika.com	17	0	com	100	1	0.523	0.054
Row16	https://www.motley.ie	20	www.motley.ie	13	0	ie	100	1	0.002	0.061
Row17	https://www.funzine.hu	21	www.funzine.hu	14	0	hu	100	1	0.002	0.043
Row18	https://www.dixon.com	21	www.dixon.com	14	0	com	100	1	0.523	0.054
Row19	https://www.ooty.ind.in	22	www.ooty.ind.in	15	0	in	100	0.625	0.005	0.061
Row20	http://www.f0519141.xsph.ru	26	www.f0519141.xsph.ru	20	0	ru	55.285	0.923	0.018	?
Row21	http://www.shprakserf.gq	24	www.shprakserf.gq	17	0	gq	56.551	1	0	0.043
Row22	https://www.cryptocompare.com	28	www.cryptocompare.com	21	0	com	100	1	0.523	0.067
Row23	https://www.diariodealmeria.es	29	www.diariodealmeria.es	22	0	es	100	1	0.006	0.068
Row24	https://www.town.minamichita.lg.jp	33	www.town.minamichita.lg.jp	26	0	jp	100	0.632	0.023	0.055
Row25	https://www.bwresearch.com	25	www.bwresearch.com	18	0	com	100	1	0.523	0.062

Box Plot



NORMALIZATION

Table "default" - Rows: 235795 Spec - Columns: 50 Properties Flow Variables												
Row ID	D NoOfO...	D NoOfLe...	D LetterR...	D NoOfD...	D DegrR...	D NoOfEq...	D NoOfQ...	D NoOfA...	D NoOfOt...	D Spacial...	D LineOf...	D Largest...
Row0	0	0.514	0.603	0	0	0	0	0	0.167	0.213	0.176	0.473
Row1	0	0.257	0.302	0	0	0	0	0	0.333	0.58	0.195	0.473
Row2	0	0.429	0.502	0	0	0	0	0	0.333	0.46	0.147	0.033
Row3	0	0.371	0.475	0	0	0	0	0	0.167	0.253	?	?
Row4	0	0.571	0.642	0	0	0	0	0	0.167	0.2	?	?
Row5	0	0.486	0.581	0	0	0	0	0	0.167	0.22	0.382	0.036
Row6	0	0.343	0.443	0	0	0	0	0	0.167	0.267	0.323	0.049
Row7	0	0.343	0.443	0	0	0	0	0	0.167	0.267	0.162	0.019
Row8	0	0.457	0.557	0	0	0	0	0	0.167	0.227	0.749	0.651
Row9	0	0.143	0.123	0	0	0	0	0	0.167	0.373	0.862	0.023
Row10	0	0.571	0.642	0	0	0	0	0	0.167	0.2	0.826	0.028
Row11	0	0.286	0.403	0	0	0	0	0	0.167	0.3	0	0.003
Row12	0	0.4	0.505	0	0	0	0	0	0.167	0.247	0.329	0.051
Row13	0	0.143	0.079	?	?	0	0	0	0.167	0.333	?	?
Row14	0	0.371	0.475	0	0	0	0	0	0.167	0.253	0.834	0.078
Row15	0	0.314	0.408	0	0	0	0	0	0.167	0.28	?	0.027
Row16	0	0.2	0.237	0	0	0	0	0	0.167	0.333	0.897	0.405
Row17	0	0.229	0.286	0	0	0	0	0	0.167	0.32	0.498	0.095
Row18	0	0.229	0.286	0	0	0	0	0	0.167	0.32	?	0.698
Row19	0	0.229	0.259	0	0	0	0	0	0.333	0.607	0.227	0.092
Row20	0	0.171	0.049	?	?	0	0	0	0.333	0.513	0.002	0.001
Row21	0	0.343	0.475	0	0	0	0	0	0.167	0.28	0.006	0.01
Row22	0	0.429	0.532	0	0	0	0	0	0.167	0.24	0.762	0.899
Row23	0	0.457	0.557	0	0	0	0	0	0.167	0.227	?	0.022
Row24	0	0.514	0.546	0	0	0	0	0	0.5	0.607	0.128	0.075

RANDOM FOREST CLASSIFIER MODEL

Table "Tree Ensemble Column Statistic" - Rows: 45 Spec - Columns: 6 Properties Flow Variables						
Row ID	#splits ...	#splits ...	#splits ...	#candi...	#candi...	#candi...
URLLength	4	7	20	17	27	50
DomainLength	1	2	11	20	26	47
IsDomainIP	0	0	0	18	26	50
URLSimilarityI...	5	27	46	5	27	52
CharContinua...	0	5	4	9	25	52
TLDLegitimate...	0	0	3	20	32	53
URLCharProb	0	2	10	11	25	65
TLDLength	0	1	2	16	23	60
HasObfuscation	0	0	0	11	28	58
NoOfObfusca...	0	0	0	17	27	56
NoOfLettersIn...	2	13	11	14	29	51
LetterRatioIn...	0	5	10	12	21	56
NoOfDegrIn...	0	0	0	15	29	48
DegrRatioInURL	0	0	0	10	26	54
NoOfEqualsIn...	0	0	0	8	34	48
NoOfQMarkIn...	0	0	0	12	22	53
NoOfAmpersa...	0	0	0	13	21	52
NoOfOtherSp...	1	11	28	11	25	60
SpacialCharR...	1	1	12	20	28	51
LineOfCode	9	12	25	10	22	47
LargestLineLe...	2	8	14	10	32	51
HasTitle	0	0	0	9	23	43
DomainTitleM...	3	1	5	16	22	47
URLTitleMatc...	0	4	5	23	28	52
HasFavicon	1	0	4	12	29	49
IsResponsive	1	2	0	15	30	61
NoOfURLRedi...	0	0	0	13	29	53
NoOfSelfRedi...	0	0	0	11	19	49
HasDescription	6	9	12	15	35	46
NoOfFrame	0	2	8	6	27	62
HasExternalF...	0	0	0	18	27	45

Columns: 55	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound	Value 0	Value 1
!ioOfEqualsInURL	Number (do...	17					0	1	?	?
!ioOfQMarkInURL	Number (do...	18					0	1	?	?
!ioOfAmpersandInURL	Number (do...	19					0	1	?	?
!ioOfOtherSpecialCharsInURL	Number (do...	20					0	1	?	?
!spacialCharRatioInURL	Number (do...	21					0	1	?	?
!lineOfCode	Number (do...	22					0	1	?	?
!argestLineLength	Number (do...	23					0	1	?	?
!hasTitle	Number (do...	24					0	1	?	?
!title	String	25					?	?	?	?
!omainTitleMatchScore	Number (do...	26					0	1	?	?
!JRLTitleMatchScore	Number (do...	27					0	1	?	?
!hasFavicon	Number (do...	28					0	1	?	?
!isResponsive	Number (do...	29					0	1	?	?
!ioOfURLRedirect	Number (do...	30					0	1	?	?
!ioOfSelfRedirect	Number (do...	31					0	1	?	?
!hasDescription	Number (do...	32					0	1	?	?
!ioOfIFrame	Number (do...	33					0	1	?	?
!hasExternalFormSubmit	Number (do...	34					0	1	?	?
!hasSocialNet	Number (do...	35					0	1	?	?
!hasSubmitButton	Number (do...	36					0	1	?	?
!hasHiddenFields	Number (do...	37					0	1	?	?
!hasPasswordField	Number (do...	38					0	1	?	?
!rank	Number (do...	39					0	1	?	?
!pay	Number (do...	40					0	1	?	?
!Crypto	Number (do...	41					0	1	?	?
!hasCopyrightInfo	Number (do...	42					0	1	?	?
!ioOfImage	Number (do...	43					0	1	?	?
!ioOfCSS	Number (do...	44					0	1	?	?
!ioOfJS	Number (do...	45					0	1	?	?
!ioOfSelfRef	Number (do...	46					0	1	?	?
!ioOfEmptyRef	Number (do...	47					0	1	?	?
!ioOfExternalRef	Number (do...	48					0	1	?	?
!abel	String	49					?	?	yes	no
! (label=yes)	Number (do...	50					0	1	?	?
! (label=no)	Number (do...	51					0	1	?	?
!abel (Out-of-bag)	String	52					?	?	yes	no
!abel (Out-of-bag) (Confidence)	Number (do...	53					0.512	1	?	?

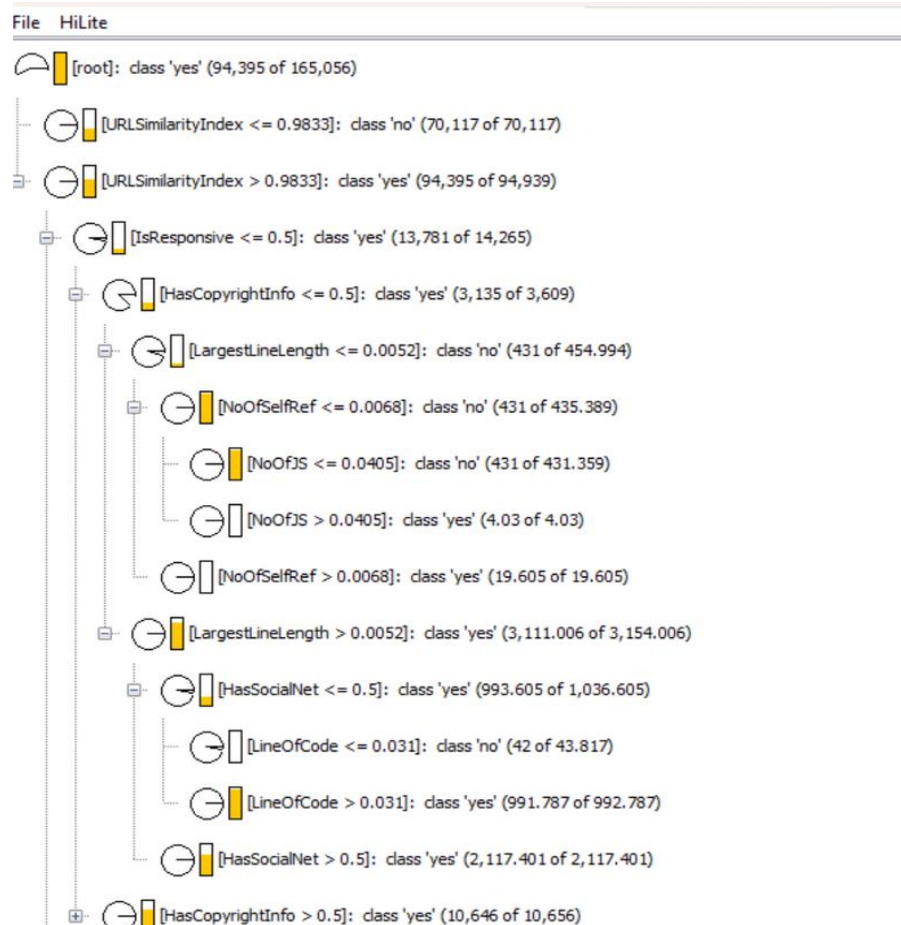
EVALUATION METRIC

Table "default" - Rows: 3 Spec - Columns: 11 Properties Flow Variables											
Row ID	<input type="checkbox"/> TruePo...	<input type="checkbox"/> FalsePo...	<input type="checkbox"/> TrueNe...	<input type="checkbox"/> FalseN...	<input type="checkbox"/> Recall	<input type="checkbox"/> Precision	<input type="checkbox"/> Sensitivity	<input type="checkbox"/> Specificity	<input type="checkbox"/> F-meas...	<input type="checkbox"/> Accuracy	<input type="checkbox"/> Cohen'...
yes	40455	7	30277	0	1	1	1	1	1	?	?
no	30277	0	40455	7	1	1	1	1	1	?	?
Overall	?	?	?	?	?	?	?	?	?	1	1

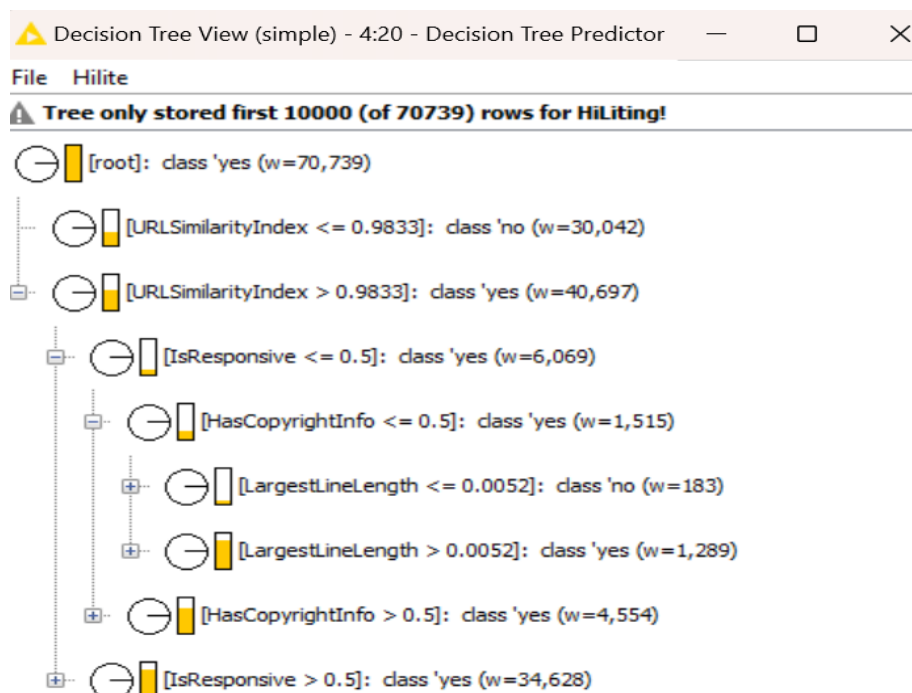
CONFUSION METRIC

Table "spec_name" - Rows: 2 Spec - Columns: 2 Properties Flow Variables		
Row ID	<input type="checkbox"/> yes	<input type="checkbox"/> no
yes	40455	0
no	7	30277

DECISION TREE MODEL LEARNER



DECISION TREE PREDICTOR



EVALUATION OF DECISION TREE

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 11 Properties Flow Variables

Row ID	<input type="checkbox"/> TruePo...	<input type="checkbox"/> FalsePo...	<input type="checkbox"/> TrueNe...	<input type="checkbox"/> FalseN...	<input type="checkbox"/> Recall	<input type="checkbox"/> Precision	<input type="checkbox"/> Sensitivity	<input type="checkbox"/> Specificity	<input type="checkbox"/> F-meas...	<input type="checkbox"/> Accuracy	<input type="checkbox"/> Cohen'...
yes	40448	13	30271	7	1	1	1	1	1	?	?
no	30271	7	40448	13	1	1	1	1	1	?	?
Overall	?	?	?	?	?	?	?	?	?	1	0.999

Confusion Matrix - 4:21 - Scorer

File Hilite

label \ Pre...	yes	no
yes	40448	7
no	13	30271

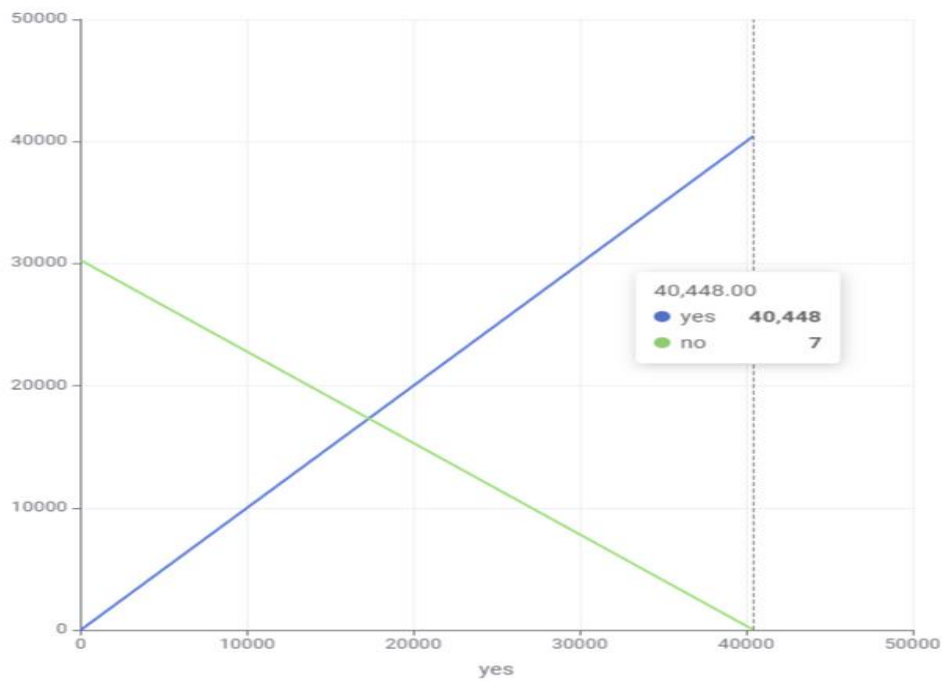
Correct classified: 70,719

Wrong classified: 20

Accuracy: 99.972%

Error: 0.028%

Line Plot



K-Means Model

Classified Data

NoOfURLRedirect	Number (do...	30					0	1	?	?
NoOfSelfRedirect	Number (do...	31					0	1	?	?
HasDescription	Number (do...	32					0	1	?	?
NoOfIFrame	Number (do...	33					0	1	?	?
HasExternalFormSubmit	Number (do...	34					0	1	?	?
HasSocialNet	Number (do...	35					0	1	?	?
HasSubmitButton	Number (do...	36					0	1	?	?
HasHiddenFields	Number (do...	37					0	1	?	?
HasPasswordField	Number (do...	38					0	1	?	?
Bank	Number (do...	39					0	1	?	?
Pay	Number (do...	40					0	1	?	?
Crypto	Number (do...	41					0	1	?	?
HasCopyrightInfo	Number (do...	42					0	1	?	?
NoOfImage	Number (do...	43					0	1	?	?
NoOfCSS	Number (do...	44					0	1	?	?
NoOfJS	Number (do...	45					0	1	?	?
NoOfSelfRef	Number (do...	46					0	1	?	?
NoOfEmptyRef	Number (do...	47					0	1	?	?
NoOfExternalRef	Number (do...	48					0	1	?	?
label	String	49					?	?	yes	no
Class [kNN]	String	50					?	?	yes	no

EVALUATION METRIC

