Table of Contents

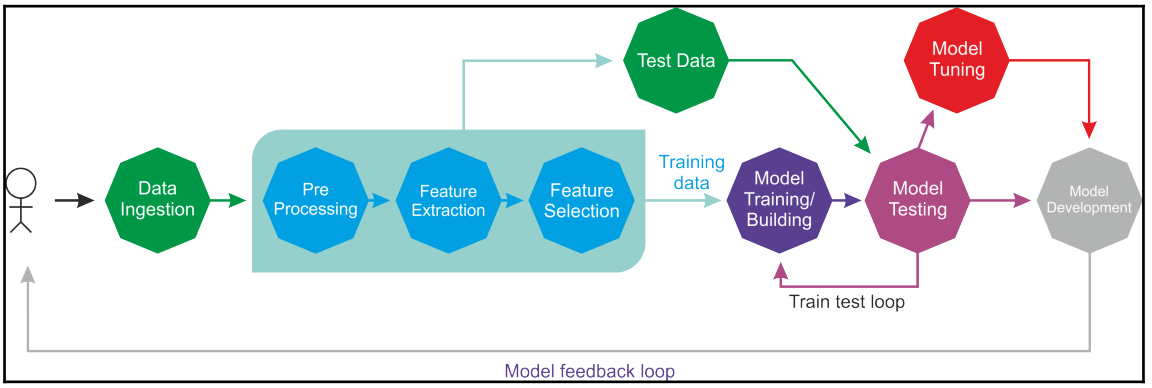
Chapter 1: Graphic bundle 1

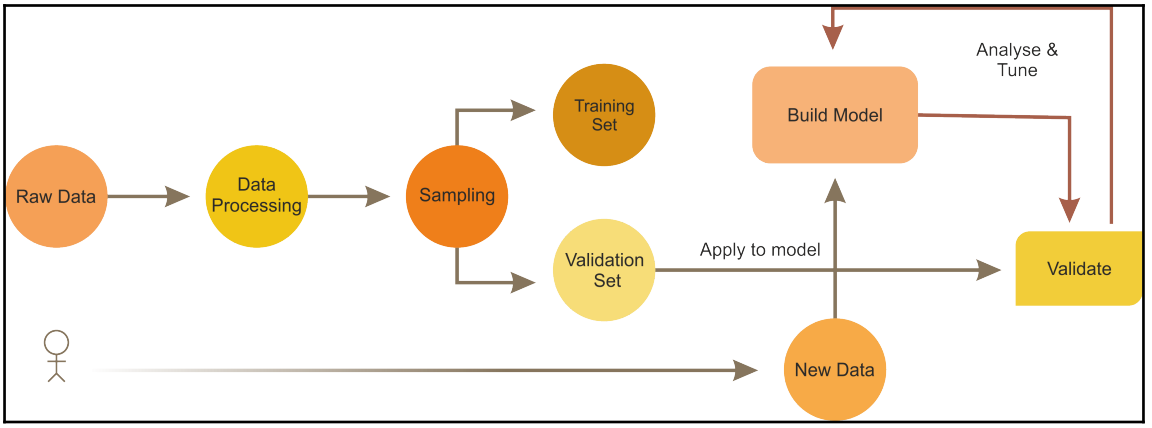
- **Chapter 1: Analyzing Insurance Severity Claims** 1
- **Chapter 2: Analyzing and Predicting Telecommunication Churn** 10
- **Chapter 3: High Frequency Bitcoin Price Prediction from Historical and Live Data** 21
- **Chapter 4: Population-Scale Clustering and Ethnicity Prediction** 26
- **Chapter 5: Topic Modeling - A Better Insight into Large-Scale Texts** 36
- **Chapter 6: Developing Model-based Movie Recommendation Engines** 41
- **Chapter 7: Options Trading Using Q-learning and Scala Play Framework** 47
- **Chapter 8: Clients Subscription Assessment for Bank Telemarketing using Deep Neural Networks** 55
- **Chapter 9: Fraud Analytics Using Autoencoders and Anomaly Detection** 72
- **Chapter 10: Human Activity Recognition using Recurrent Neural Networks** 89
- **Chapter 11: Image Classification using Convolutional Neural Networks** 97

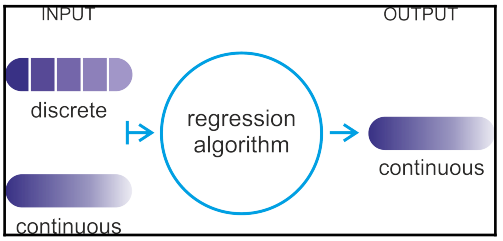
Index 103

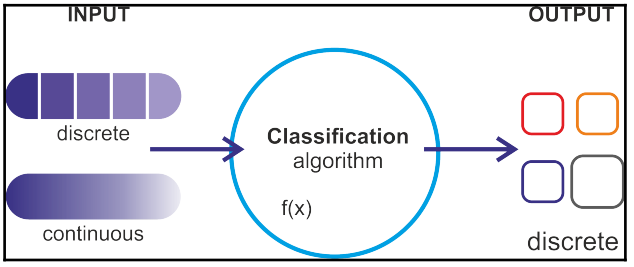
1
Graphic bundle

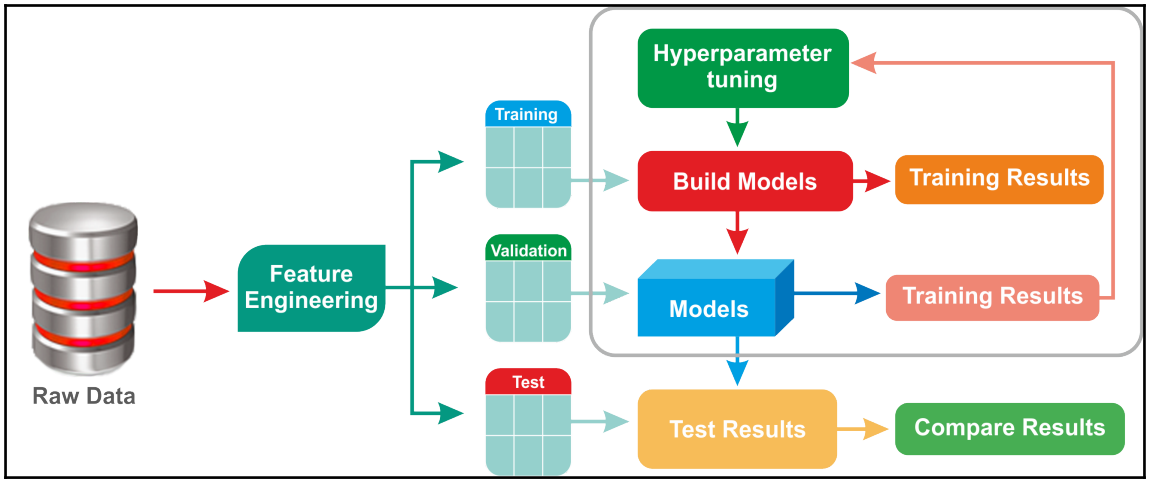
Chapter 1: Analyzing Insurance Severity Claims

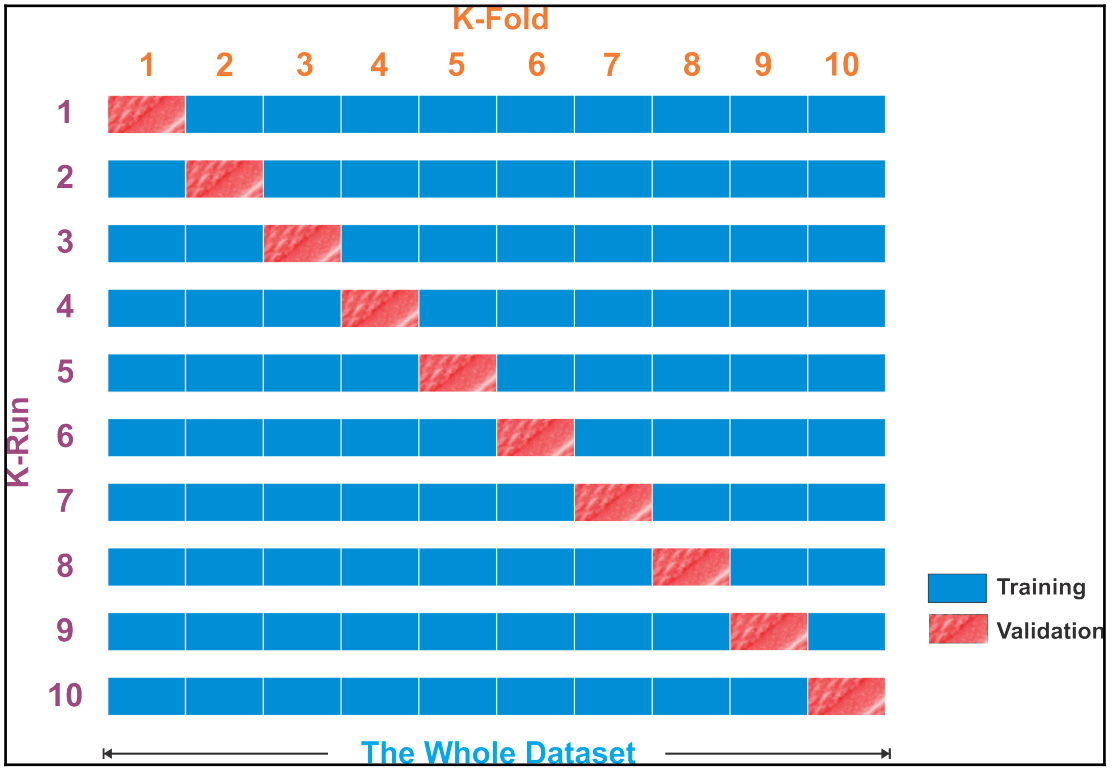



Raw Data
Data Processing
Sampling
Training Set
Validation Set
Build Model
Analyse & Tune
Apply to model
Validate
New Data


INPUT
OUTPUT
discrete
continuous
regression algorithm
continuous


INPUT
OUTPUT
discrete
continuous
Classification algorithm
f(x)
discrete


Hyperparameter tuning
Training
Build Models
Training Results
Validation
Feature Engineering
Models
Training Results
Test
Test Results
Compare Results
Raw Data

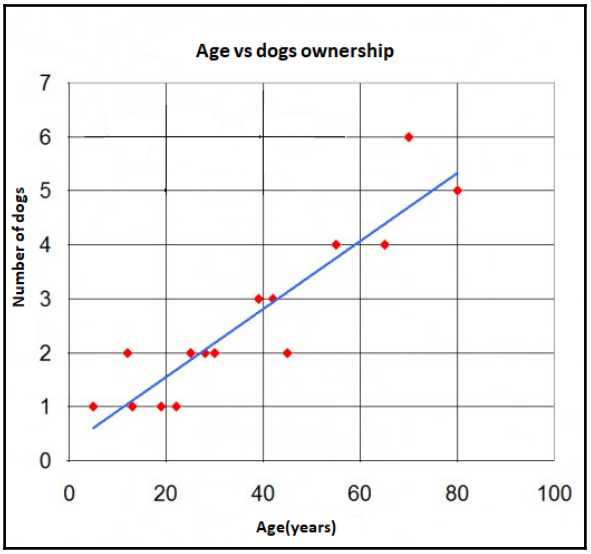

K-Fold
1 2 3 4 5 6 7 8 9 10
K-Run
1
2
3
4
5
6
7
8
9
10
Training
Validation
The Whole Dataset

id	cat1	cat2	cat3	cont1	cont2	cont3	loss
1	A	B	A	0.7263	0.245921	0.187583	2213.18
2	A	B	A	0.330514	0.737068	0.592681	1283.6
5	A	B	A	0.261841	0.358319	0.484196	3005.09
10	B	B	A	0.321594	0.555782	0.527991	939.85
11	A	B	A	0.273204	0.15999	0.527991	2763.85
13	A	B	A	0.54667	0.681761	0.634224	5142.87
14	A	A	A	0.471447	0.737068	0.61366	1132.22
20	A	B	A	0.826591	0.488789	0.26357	3585.75
23	A	B	B	0.330514	0.555782	0.440642	10280.2
24	A	B	A	0.7263	0.358319	0.356819	6184.59
25	A	B	A	0.496063	0.358319	0.65431	6396.85
33	A	B	A	0.520698	0.422197	0.634224	5965.73
34	B	A	A	0.321594	0.555782	0.527991	1193.05
41	B	A	A	0.351358	0.555782	0.440642	1071.77
47	A	A	A	0.894333	0.299102	0.094942	585.18
48	A	A	A	0.472892	0.681761	0.484196	1395.45
49	A	B	B	0.424162	0.737068	0.80643	6609.32
51	A	A	A	0.834747	0.488789	0.246911	2658.7
52	A	A	B	0.488816	0.422197	0.592681	4167.32
55	A	A	A	0.391956	0.15999	0.26357	3797.89

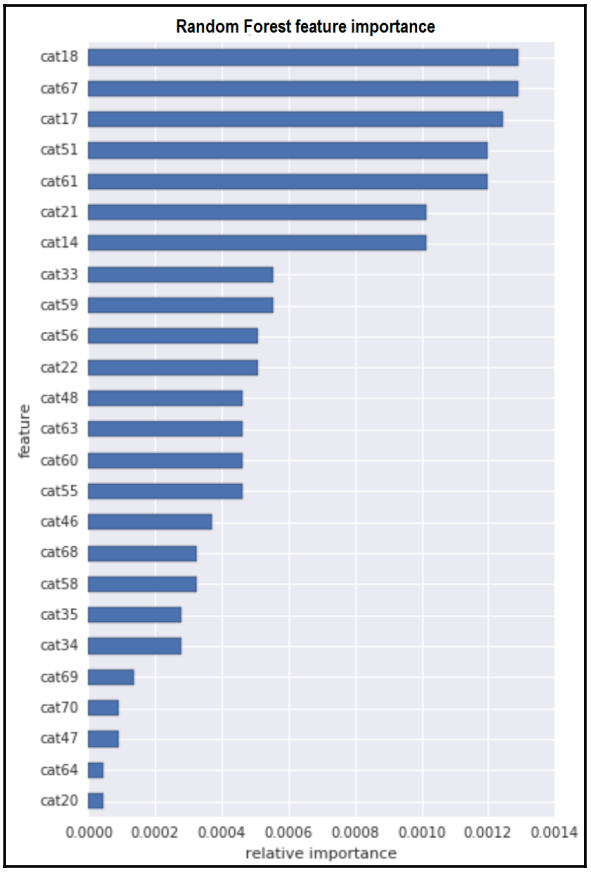
only showing top 20 rows

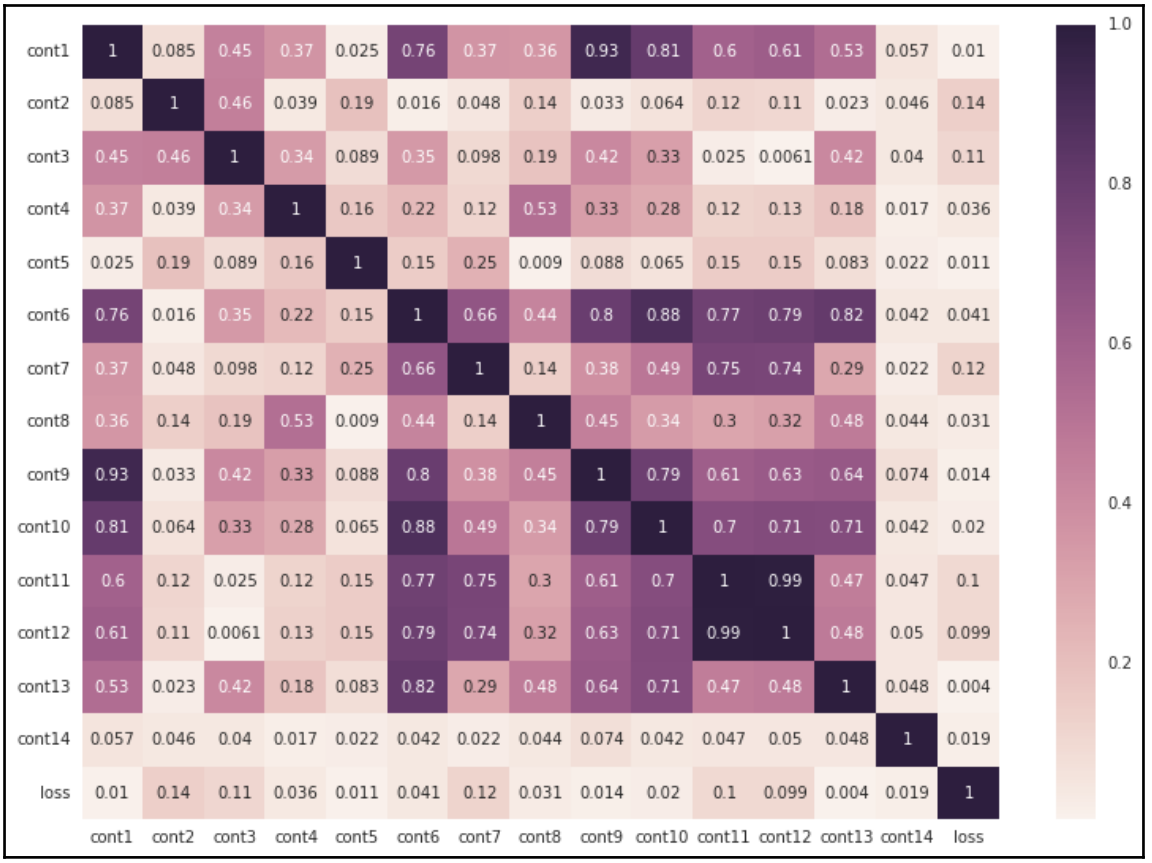
cat109	cat110	cat112	cat113	cat116
BU	BC	AS	S	LB
BI	CQ	AV	BM	DP
AB	DK	C	AF	GK
BI	CS	N	AE	DJ
H	C	Y	BM	CK
BI	CS	AS	AE	DJ
BI	DK	J	AF	DJ
BI	EB	AH	Y	LO
BI	BC	K	AX	IE
BU	DW	U	S	LY
BI	AM	AS	H	GS
BI	AI	E	AX	HK
BI	CS	E	AE	DJ
AB	EG	AK	K	DC
BI	EG	AH	L	MP
BI	CQ	AI	BM	DS
BI	CL	AE	A	DJ
BU	BC	A	S	LE
BI	CS	AV	J	HQ
BI	BS	L	BM	HJ

only showing top 20 rows



Loss	Task	Formula	Description
Squared Error	Regression	$\sum_{i=1}^{N}(y_i - F(x_i))^2$	Also called L2 loss. Default loss for regression tasks
Absolute Error	Regression	$\sum_{i=1}^{N}\|y_i - F(x_i)\|$	Also called L1 loss. Can be more robust to outliers than Squared Error.


Random Forest feature importance
cat18
cat67
cat17
cat51
cat61
cat21
cat14
cat33
cat59
cat56
cat22
cat48
cat63
cat60
cat55
cat46
cat68
cat58
cat35
cat34
cat69
cat70
cat47
cat64
cat20
feature
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
relative importance

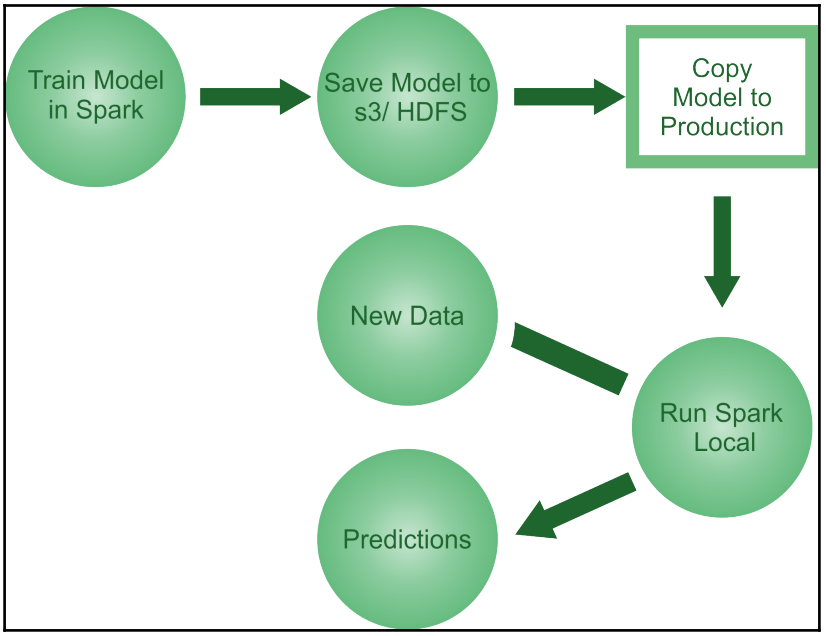

cont1 1 0.085 0.45 0.37 0.025 0.76 0.37 0.36 0.93 0.81 0.6 0.61 0.53 0.057 0.01
cont2 0.085 1 0.46 0.039 0.19 0.016 0.048 0.14 0.033 0.064 0.12 0.11 0.023 0.046 0.14
cont3 0.45 0.46 1 0.34 0.089 0.35 0.098 0.19 0.42 0.33 0.025 0.0061 0.42 0.04 0.11
cont4 0.37 0.039 0.34 1 0.16 0.22 0.12 0.53 0.33 0.28 0.12 0.13 0.18 0.017 0.036
cont5 0.025 0.19 0.089 0.16 1 0.15 0.25 0.009 0.088 0.065 0.15 0.15 0.083 0.022 0.011
cont6 0.76 0.016 0.35 0.22 0.15 1 0.66 0.44 0.8 0.88 0.77 0.79 0.82 0.042 0.041
cont7 0.37 0.048 0.098 0.12 0.25 0.66 1 0.14 0.38 0.49 0.75 0.74 0.29 0.022 0.12
cont8 0.36 0.14 0.19 0.53 0.009 0.44 0.14 1 0.45 0.34 0.3 0.32 0.48 0.044 0.031
cont9 0.93 0.033 0.42 0.33 0.088 0.8 0.38 0.45 1 0.79 0.61 0.63 0.64 0.074 0.014
cont10 0.81 0.064 0.33 0.28 0.065 0.88 0.49 0.34 0.79 1 0.7 0.71 0.71 0.042 0.02
cont11 0.6 0.12 0.025 0.12 0.15 0.77 0.75 0.3 0.61 0.7 1 0.99 0.47 0.047 0.1
cont12 0.61 0.11 0.0061 0.13 0.15 0.79 0.74 0.32 0.63 0.71 0.99 1 0.48 0.05 0.099
cont13 0.53 0.023 0.42 0.18 0.083 0.82 0.29 0.48 0.64 0.71 0.47 0.48 1 0.048 0.004
cont14 0.057 0.046 0.04 0.017 0.022 0.042 0.022 0.044 0.074 0.042 0.047 0.05 0.048 1 0.019
loss 0.01 0.14 0.11 0.036 0.011 0.041 0.12 0.031 0.014 0.02 0.1 0.099 0.004 0.019 1
cont1 cont2 cont3 cont4 cont5 cont6 cont7 cont8 cont9 cont10 cont11 cont12 cont13 cont14 loss
1.0
0.8
0.6
0.4
0.2

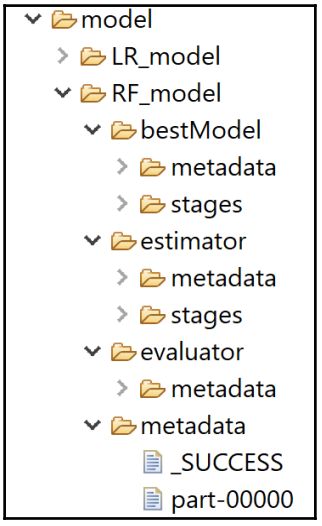
Training set			
Metrics	**Linear regression**	**Gradient Boosted Tree**	**Random Forest**
MSE	4460667.3666198505	2711134.460296872	1340574.3409399686
RMSE	2112.0292059107164	1646.5522950385973	1157.8317412042081
R^2	-0.1514435541595276	0.4979619968485668	0.7642745310548124
MAE	1356.9375609756164	1126.582534126603	809.5917285994619
Explained variance	8336528.638733305	8336528.638733303	8337897.224852404
Validation set			
Metrics	**Linear regression**	**Gradient Boosted Tree**	**Random Forest**
MSE	4839128.978963534	4796065.983773314	4312608.024875177
RMSE	2199.802031766389	2189.9922337244293	2076.6819749001475
R^2	-0.24922962724089603	0.13708582379658474	0.1369507149716651
MAE	1356.419484419514	1289.9808960385383	1273.0714382935894
Explained variance	8724661.329105612	8724866.468978886	8737233.110450774

id	loss
4	1205.818863
6	2124.032479
9	11466.34892
12	4671.905613
15	-29.56590485
17	2312.270982
21	2381.246045
28	199.0961342
32	2526.540265
43	5046.292227
46	3768.996872
50	1368.325947
54	1971.648909
62	2891.711648
70	2945.419087
71	5565.365742
75	2342.345313
77	3035.87975
81	3512.796204

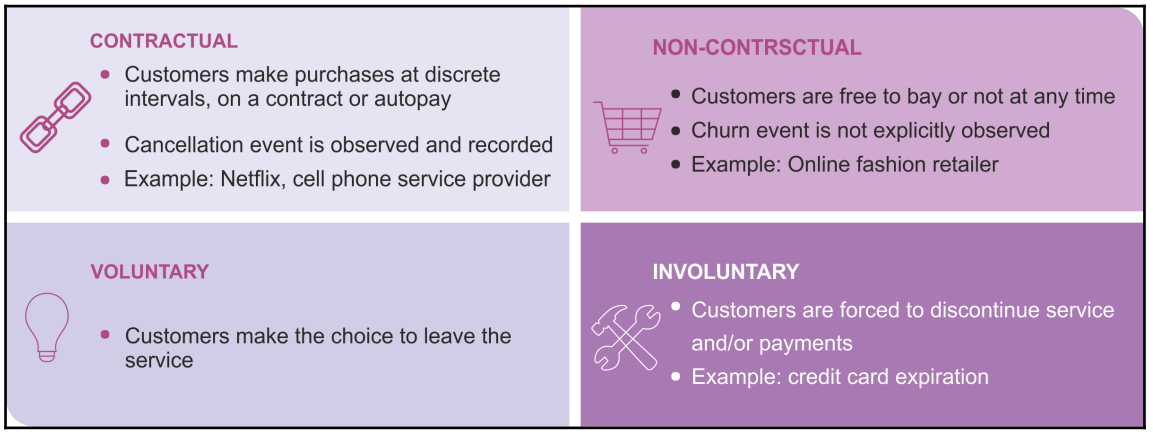
id	loss
4	2118.753
6	2425.418
9	8967.434
12	5772.11
15	907.1143
17	2106.422
21	2134.453
28	1159.63
32	2465.344
43	4356.98
46	2730.829
50	971.0264
54	1818.144
62	1705.926
70	1903.267
71	6196.935
75	2546.685
77	2462.602
81	2926.829

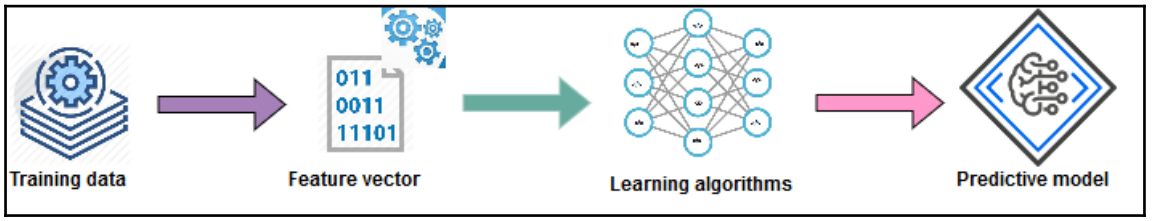
id	loss
4	2182.815
6	1935.753
9	11069.49
12	6639.439
15	988.8615
17	2061.888
21	1982.437
28	1977.683
32	3138.645
43	4143.904
46	4423.25
50	1075.983
54	1552.547
62	1759.637
70	2543.034
71	6566.392
75	2426.998
77	3329.786
81	3685.368

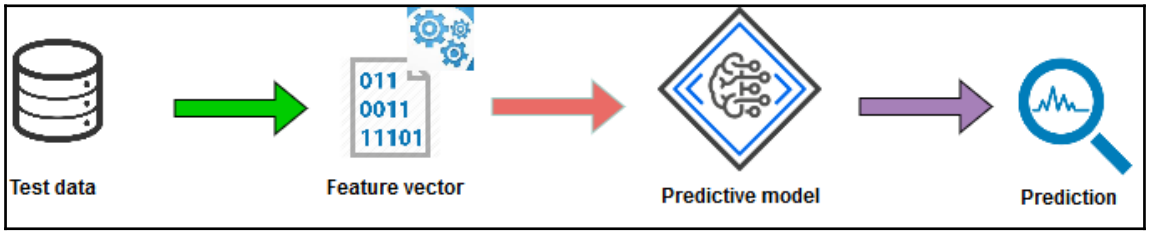

Train Model in Spark
Save Model to s3/ HDFS
Copy Model to Production
New Data
Run Spark Local
Predictions

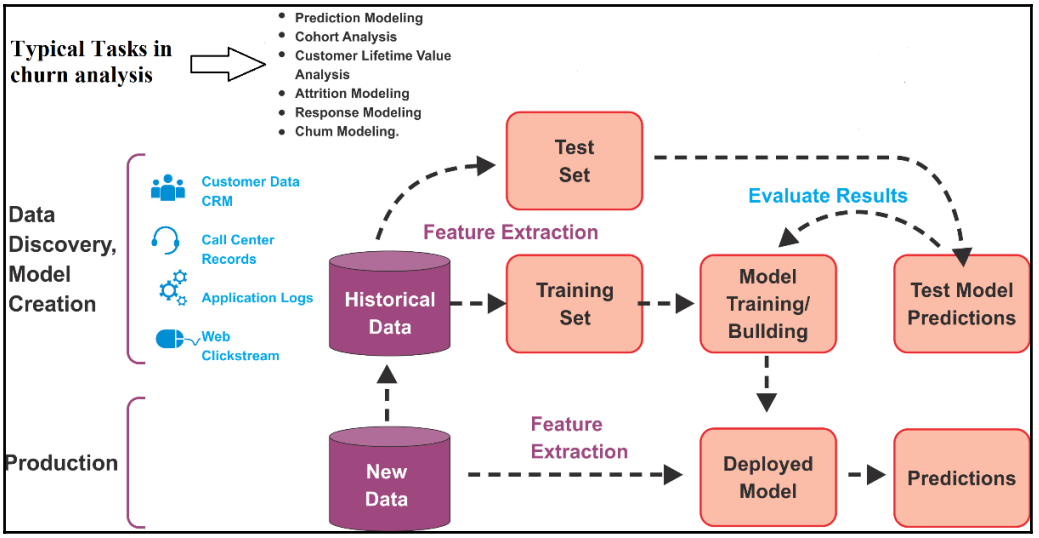

model
LR_model
RF_model
bestModel
metadata
stages
estimator
metadata
stages
evaluator
metadata
metadata
_SUCCESS
part-00000

Chapter 2: Analyzing and Predicting Telecommunication Churn







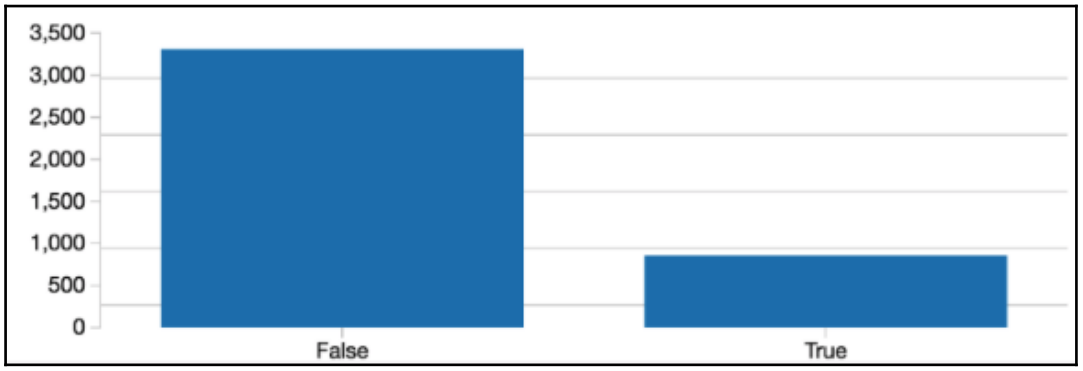



```
root
 |-- state_code: string (nullable = true)
 |-- account_length: integer (nullable = true)
 |-- area_code: string (nullable = true)
 |-- international_plan: string (nullable = true)
 |-- voice_mail_plan: string (nullable = true)
 |-- num_voice_mail: double (nullable = true)
 |-- total_day_mins: double (nullable = true)
 |-- total_day_calls: double (nullable = true)
 |-- total_day_charge: double (nullable = true)
 |-- total_evening_mins: double (nullable = true)
 |-- total_evening_calls: double (nullable = true)
 |-- total_evening_charge: double (nullable = true)
 |-- total_night_mins: double (nullable = true)
 |-- total_night_calls: double (nullable = true)
 |-- total_night_charge: double (nullable = true)
 |-- total_international_mins: double (nullable = true)
 |-- total_international_calls: double (nullable = true)
 |-- total_international_charge: double (nullable = true)
 |-- total_international_num_calls: double (nullable = true)
 |-- churn: string (nullable = true)
```

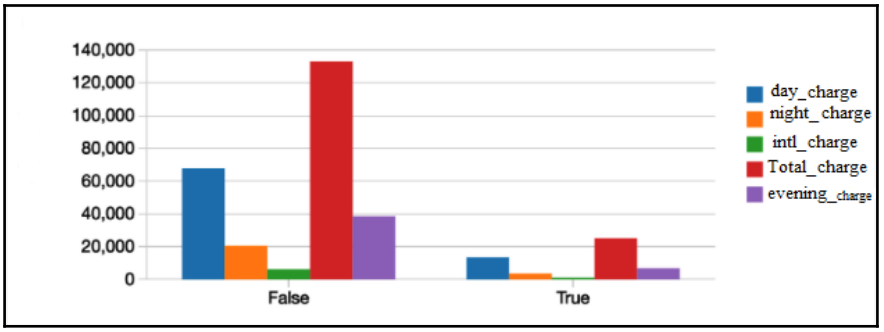
state	len	acode	intlplan	vplan	numvmail	tdmins	tdcalls	tdcharge	temins	tecalls	techarge	tnmins	tncalls	tncharge	timins	ticalls	ticharge	numcs	churn
KS	128	415	No	Yes	25.0	265.1	110.0	45.07	197.4	99.0	16.78	244.7	91.0	11.01	10.0	3.0	2.7	1.0	False
OH	107	415	No	Yes	26.0	161.6	123.0	27.47	195.5	103.0	16.62	254.4	103.0	11.45	13.7	3.0	3.7	1.0	False
NJ	137	415	No	No	0.0	243.4	114.0	41.38	121.2	110.0	10.3	162.6	104.0	7.32	12.2	5.0	3.29	0.0	False
OH	84	408	Yes	No	0.0	299.4	71.0	50.9	61.9	88.0	5.26	196.9	89.0	8.86	6.6	7.0	1.78	2.0	False
OK	75	415	Yes	No	0.0	166.7	113.0	28.34	148.3	122.0	12.61	186.9	121.0	8.41	10.1	3.0	2.73	3.0	False
AL	118	510	Yes	No	0.0	223.4	98.0	37.98	220.6	101.0	18.75	203.9	118.0	9.18	6.3	6.0	1.7	0.0	False
MA	121	510	No	Yes	24.0	218.2	88.0	37.09	348.5	108.0	29.62	212.6	118.0	9.57	7.5	7.0	2.03	3.0	False
MO	147	415	Yes	No	0.0	157.0	79.0	26.69	103.1	94.0	8.76	211.8	96.0	9.53	7.1	6.0	1.92	0.0	False
WV	141	415	Yes	Yes	37.0	258.6	84.0	43.96	222.0	111.0	18.87	326.4	97.0	14.69	11.2	5.0	3.02	0.0	False
RI	74	415	No	No	0.0	187.7	127.0	31.91	163.4	148.0	13.89	196.0	94.0	8.82	9.1	5.0	2.46	0.0	False
IA	168	408	No	No	0.0	128.8	96.0	21.9	104.9	71.0	8.92	141.1	128.0	6.35	11.2	2.0	3.02	1.0	False
MT	95	510	No	No	0.0	156.6	88.0	26.62	247.6	75.0	21.05	192.3	115.0	8.65	12.3	5.0	3.32	3.0	False
IA	62	415	No	No	0.0	120.7	70.0	20.52	307.2	76.0	26.11	203.0	99.0	9.14	13.1	6.0	3.54	4.0	False
ID	85	408	No	Yes	27.0	196.4	139.0	33.39	280.9	90.0	23.88	89.3	75.0	4.02	13.8	4.0	3.73	1.0	False
VT	93	510	No	No	0.0	190.7	114.0	32.42	218.2	111.0	18.55	129.6	121.0	5.83	8.1	3.0	2.19	3.0	False
VA	76	510	No	Yes	33.0	189.7	66.0	32.25	212.8	65.0	18.09	165.7	108.0	7.46	10.0	5.0	2.7	1.0	False
TX	73	415	No	No	0.0	224.4	90.0	38.15	159.5	88.0	13.56	192.8	74.0	8.68	13.0	2.0	3.51	1.0	False
FL	147	415	No	No	0.0	155.1	117.0	26.37	239.7	93.0	20.37	208.8	133.0	9.4	10.6	4.0	2.86	0.0	False
CO	77	408	No	No	0.0	62.4	89.0	10.61	169.9	121.0	14.44	209.6	64.0	9.43	5.7	6.0	1.54	5.0	True
AZ	130	415	No	No	0.0	183.0	112.0	31.11	72.9	99.0	6.2	181.8	78.0	8.18	9.5	19.0	2.57	0.0	False

only showing top 20 rows

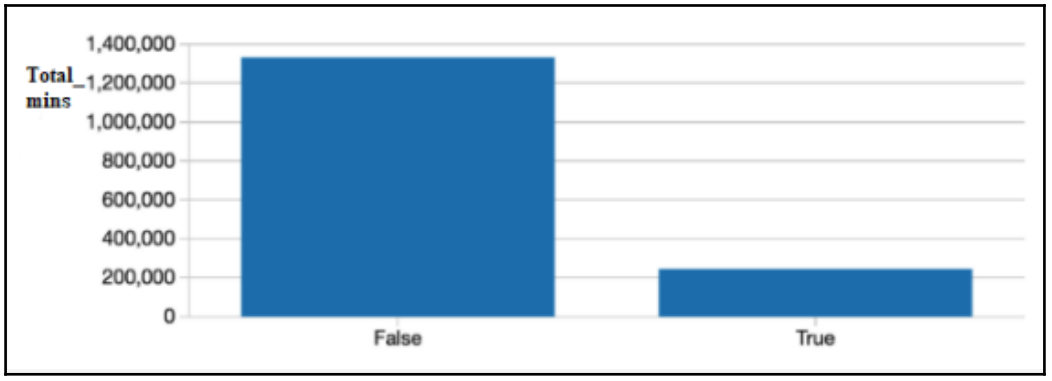
summary	state_code	account_length	area_code	international_plan	voice_mail_plan
count	2666	2666	2666	2666	2666
mean	null	100.62040510127532	437.43885971492875	null	null
stddev	null	39.56397365334985	42.521018019427174	null	null
min	AK	1	408	No	No
max	WY	243	510	Yes	Yes

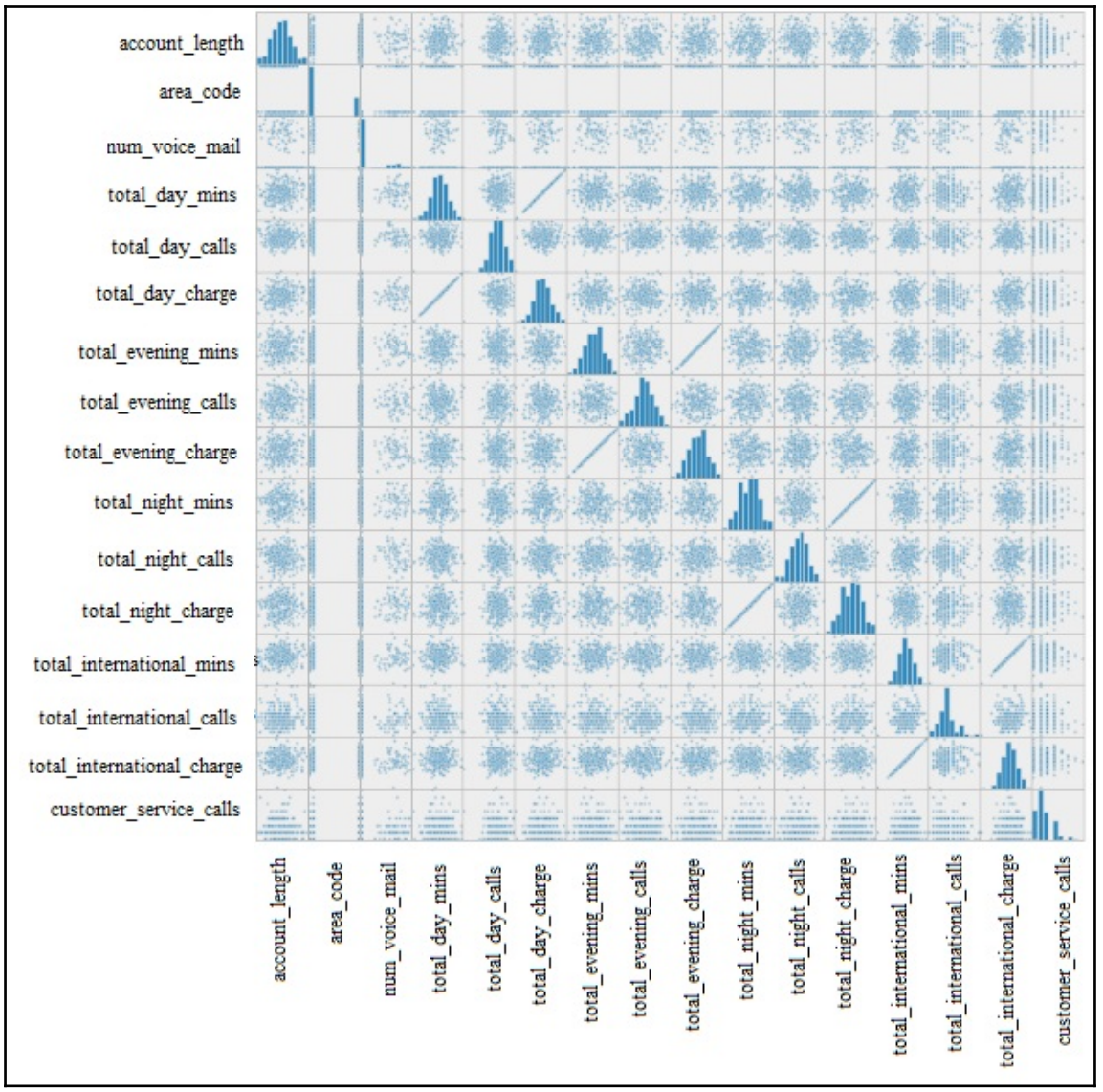


churn	TDC	TEC	TNC	TIC	Total_charge
False	67812.10999999997	38504.60000000002	20549.78	6236.499999999996	133102.99
True	13533.960000000005	6905.569999999997	3584.69	1133.63	25157.850000000002



churn	Total_minutes
False	1331626.900000002
True	244708.89999999994




account_length
area_code
num_voice_mail
total_day_mins
total_day_calls
total_day_charge
total_evening_mins
total_evening_calls
total_evening_charge
total_night_mins
total_night_calls
total_night_charge
total_international_mins
total_international_calls
total_international_charge
customer_service_calls

```
+--------------+------------------+--------------+--------------+---------------------------+-----+
|account_length|international_plan|num_voice_mail|total_day_calls|total_international_num_calls|churn|
+--------------+------------------+--------------+--------------+---------------------------+-----+
|            84|               Yes|           0.0|          71.0|                        2.0|False|
|            74|                No|           0.0|         127.0|                        0.0|False|
|            95|                No|           0.0|          88.0|                        3.0|False|
|            62|                No|           0.0|          70.0|                        4.0|False|
|           147|                No|           0.0|         117.0|                        0.0|False|
|            77|                No|           0.0|          89.0|                        5.0| True|
|           130|                No|           0.0|         112.0|                        0.0|False|
|           142|                No|           0.0|          95.0|                        2.0|False|
|            12|                No|           0.0|         118.0|                        1.0| True|
|            57|                No|          25.0|          94.0|                        0.0|False|
+--------------+------------------+--------------+--------------+---------------------------+-----+
only showing top 10 rows
```

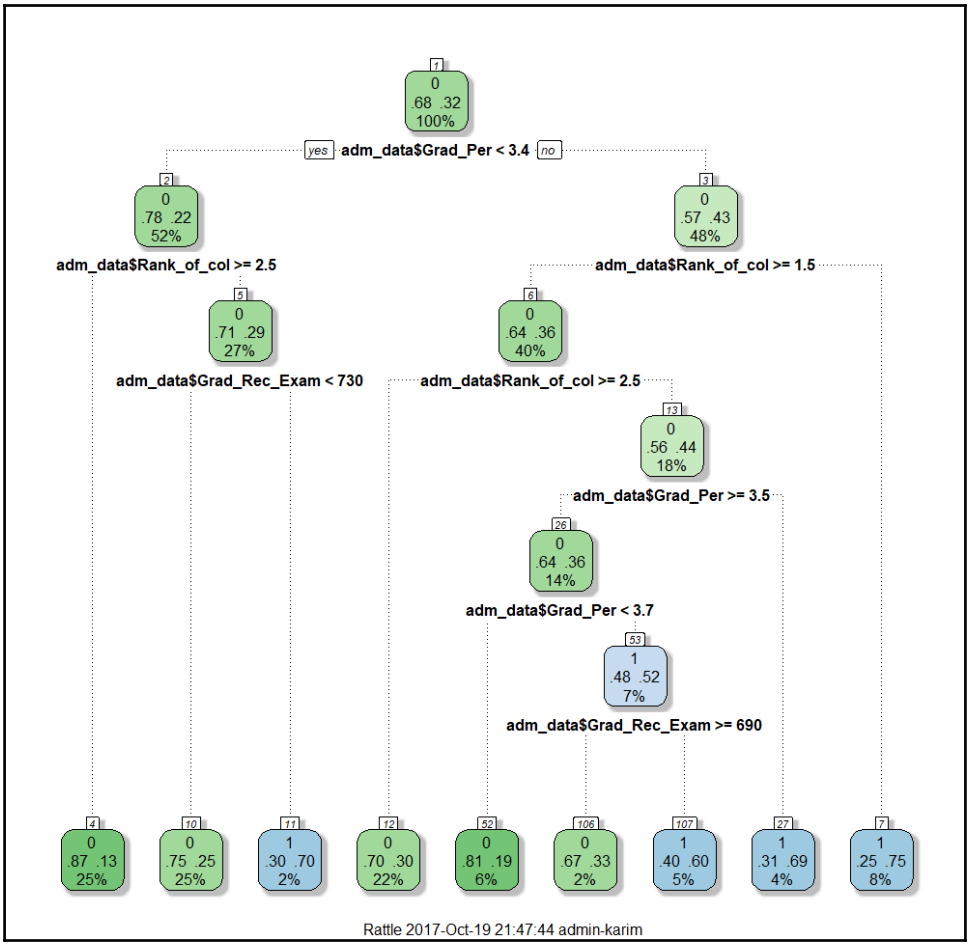
```
+-----+---------------+--------------------+
|label|Predicted_label|         probability|
+-----+---------------+--------------------+
|  0.0|            0.0|[0.50128534704370...|
|  1.0|            0.0|[0.50128534704370...|
|  1.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
|  1.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
|  0.0|            0.0|[0.50128534704370...|
+-----+---------------+--------------------+
only showing top 10 rows
```

```
Markers  Properties  Servers  Data Source Explorer  Progress  JUnit  Console
<terminated> ChurnPredictionLR$ [Scala Application] C:\Program Files\Java\jdk1.8.0_131\bin\javaw.exe (Jan 1,
18/01/01 16:57:27 WARN NativeCodeLoader: Unable to load native-hadoop library for
18/01/01 16:57:30 WARN BLAS: Failed to load implementation from: com.github.fommil
18/01/01 16:57:30 WARN BLAS: Failed to load implementation from: com.github.fommil
18/01/01 16:57:30 WARN SparkSession$Builder: Using an existing SparkSession; some
18/01/01 16:57:32 WARN Utils: Truncated the string representation of a plan since
Total Count: 667
Correct: 534
Wrong: 133
Ratio wrong: 0.19940029985007496
Ratio correct: 0.800599700149925
Ratio true positive: 0.7136431784107946
Ratio false positive: 0.14392803598200898
Ratio true negative: 0.08695652173913043
Ratio false negative: 0.05547226386806597
```

```
+-----+--------------------+--------------------+----------+
|label|            features|       rawPrediction|prediction|
+-----+--------------------+--------------------+----------+
|  0.0|[117.0,0.0,0.0,18...|[0.08973008965728...|       0.0|
|  1.0|[65.0,0.0,0.0,129...|[-0.2242381550714...|       1.0|
|  1.0|[161.0,0.0,0.0,33...|[-1.4711802407566...|       1.0|
|  0.0|[111.0,0.0,0.0,11...|[1.32527571170096...|       0.0|
|  0.0|[49.0,0.0,0.0,119...|[0.85288284318901...|       0.0|
|  0.0|[36.0,0.0,30.0,14...|[1.62981681295705...|       0.0|
|  0.0|[65.0,0.0,0.0,211...|[-0.2107700769615...|       1.0|
|  1.0|[119.0,0.0,0.0,15...|[-0.7029272317993...|       1.0|
|  0.0|[10.0,0.0,0.0,186...|[-0.1759587872812...|       1.0|
|  0.0|[68.0,0.0,0.0,148...|[0.15925905739758...|       0.0|
+-----+--------------------+--------------------+----------+
only showing top 10 rows
```

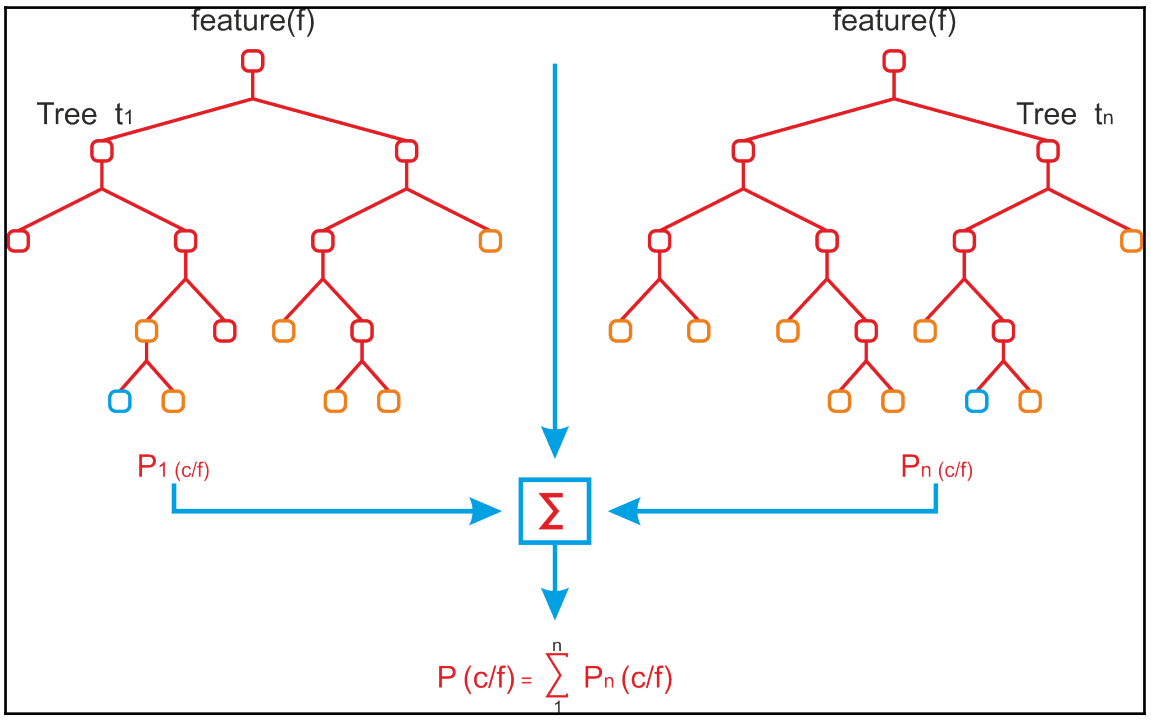
Markers Properties Servers Data Source Explorer Progress JUnit

```
<terminated> ChurnPredictionSVM$ [Scala Application] C:\Program Files\Java\jdk1.8.0_1
18/01/01 17:09:07 WARN NativeCodeLoader: Unable to load native-had
18/01/01 17:09:11 WARN BLAS: Failed to load implementation from: c
18/01/01 17:09:11 WARN BLAS: Failed to load implementation from: c
18/01/01 17:09:11 WARN SparkSession$Builder: Using an existing Spa
18/01/01 17:09:14 WARN Utils: Truncated the string representation
Total Count: 667
Correct: 515
Wrong: 152
Ratio wrong: 0.22788605697151423
Ratio correct: 0.7721139430284858
Ratio true positive: 0.6686656671664168
Ratio false positive: 0.1889055472263868
Ratio true negative: 0.10344827586206896
Ratio false negative: 0.038980509745127435
```

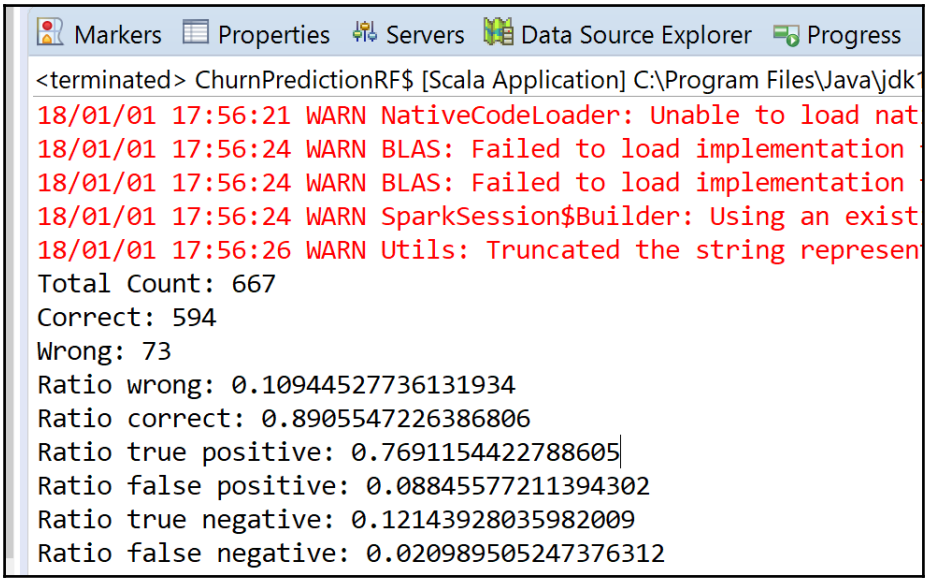


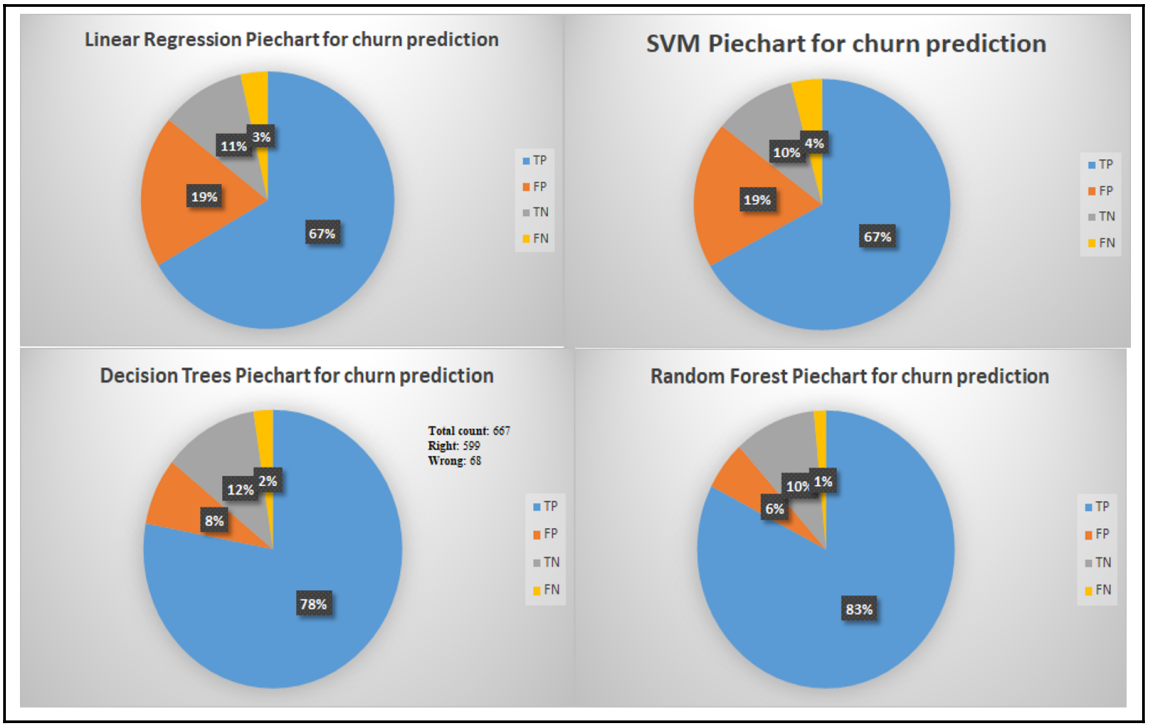

```
+-----+---------------+--------------------+
|label|Predicted_label|         probability|
+-----+---------------+--------------------+
|  0.0|            0.0|           [1.0,0.0]|
|  1.0|            1.0|[0.09523809523809...|
|  1.0|            1.0|           [0.0,1.0]|
|  0.0|            0.0|[0.86545454545454...|
|  0.0|            0.0|[0.86545454545454...|
|  0.0|            0.0|[0.86545454545454...|
|  0.0|            0.0|[0.80263157894736...|
|  1.0|            1.0|[0.09523809523809...|
|  0.0|            0.0|[0.86545454545454...|
|  0.0|            0.0|[0.86545454545454...|
+-----+---------------+--------------------+
```

```
Markers  Properties  Servers  Data Source Explorer  Prog
<terminated> ChurnPredictionDT$ [Scala Application] C:\Program Files\Jav
18/01/01 17:40:19 WARN NativeCodeLoader: Unable to load
18/01/01 17:40:22 WARN BLAS: Failed to load implementat
18/01/01 17:40:22 WARN BLAS: Failed to load implementat
18/01/01 17:40:22 WARN SparkSession$Builder: Using an e
18/01/01 17:40:25 WARN Utils: Truncated the string repr
Total Count: 667
Correct: 599
Wrong: 68
Ratio wrong: 0.10194902548725637
Ratio correct: 0.8980509745127436
Ratio true positive: 0.7796101949025487
Ratio false positive: 0.07796101949025487
Ratio true negative: 0.1184407796101949
Ratio false negative: 0.0239880059970015
```



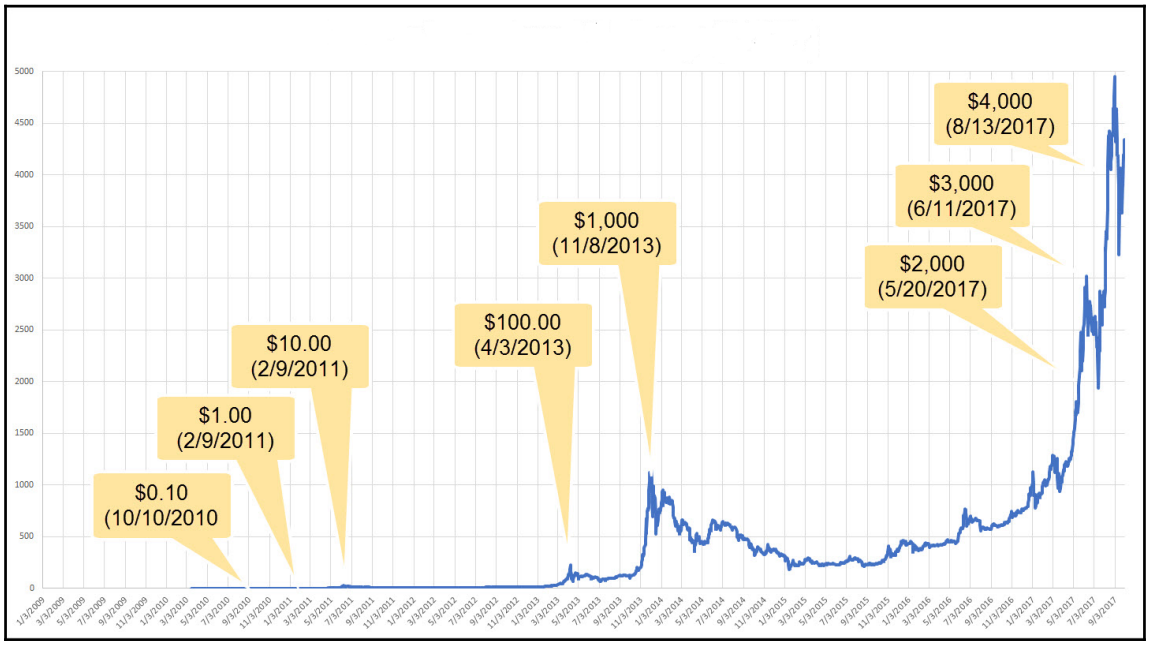
```
+-----+---------------+--------------------+
|label|Predicted_label|         probability|
+-----+---------------+--------------------+
|  0.0|            1.0|[0.47967610185334...|
|  1.0|            1.0|[0.38531389528766...|
|  1.0|            1.0|[0.06850345623033...|
|  0.0|            0.0|[0.88942965019863...|
|  0.0|            0.0|[0.79162145165495...|
|  0.0|            0.0|[0.92841581163596...|
|  0.0|            1.0|[0.41083015977062...|
|  1.0|            1.0|[0.21531047960566...|
|  0.0|            1.0|[0.49987532729656...|
|  0.0|            1.0|[0.49313495545749...|
+-----+---------------+--------------------+
only showing top 10 rows
```

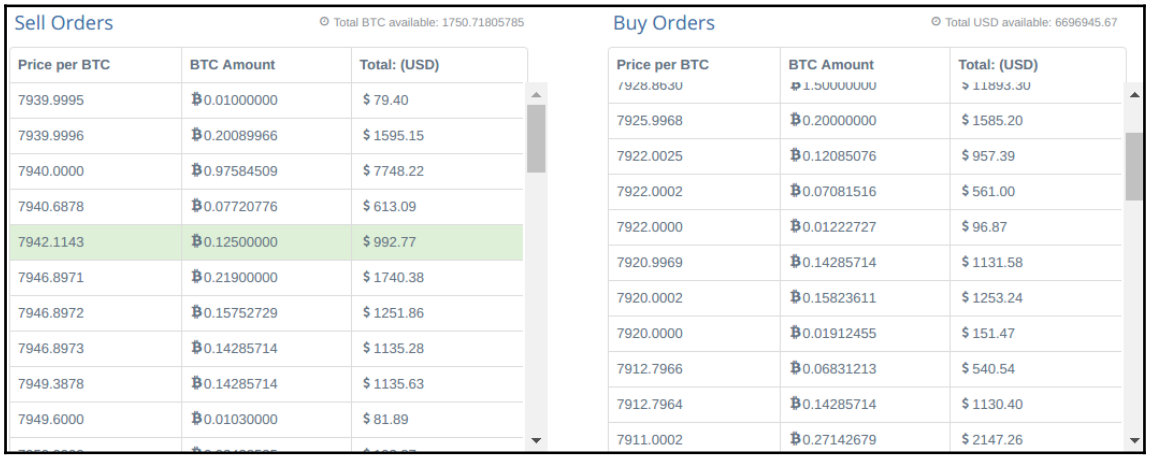





```
Markers  Properties  Servers  Data Source Explorer  Progress  JUnit  Console
<terminated> RandomForestModelReuse$ [Scala Application] C:\Program Files\Java\jdk1.8.0_131\bin\javaw.e
18/01/01 21:07:44 WARN NativeCodeLoader: Unable to load native-hadoop library for
18/01/01 21:07:58 WARN SparkSession$Builder: Using an existing SparkSession; some
18/01/01 21:07:59 WARN Utils: Truncated the string representation of a plan since
Accuracy: 0.8747423629002576
Area under the precision-recall curve: 0.7260962563830867
Area under the receiver operating characteristic (ROC) curve: 0.8747423629002576
Total Count: 667
Correct: 594
Wrong: 73
Ratio wrong: 0.10944527736131934
Ratio correct: 0.8905547226386806
Ratio true positive: 0.7691154422788605
Ratio false positive: 0.08845577211394302
Ratio true negative: 0.12143928035982009
Ratio false negative: 0.020989505247376312
```

Chapter 3: High Frequency Bitcoin Price Prediction from Historical and Live Data



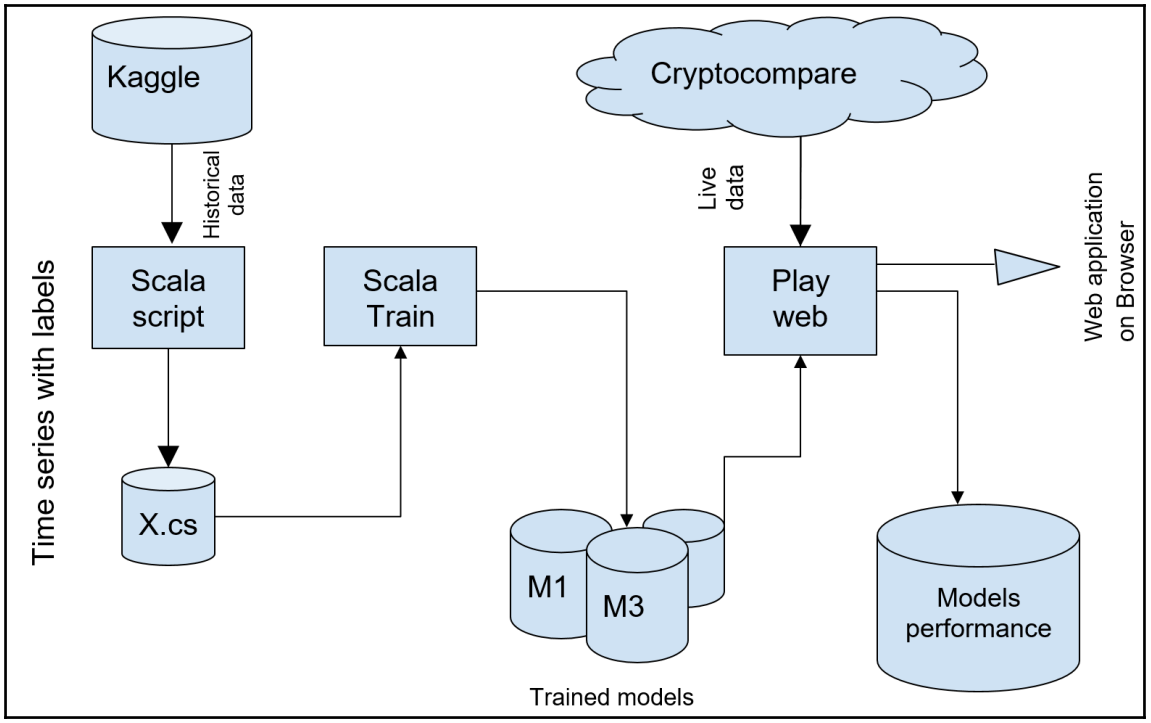


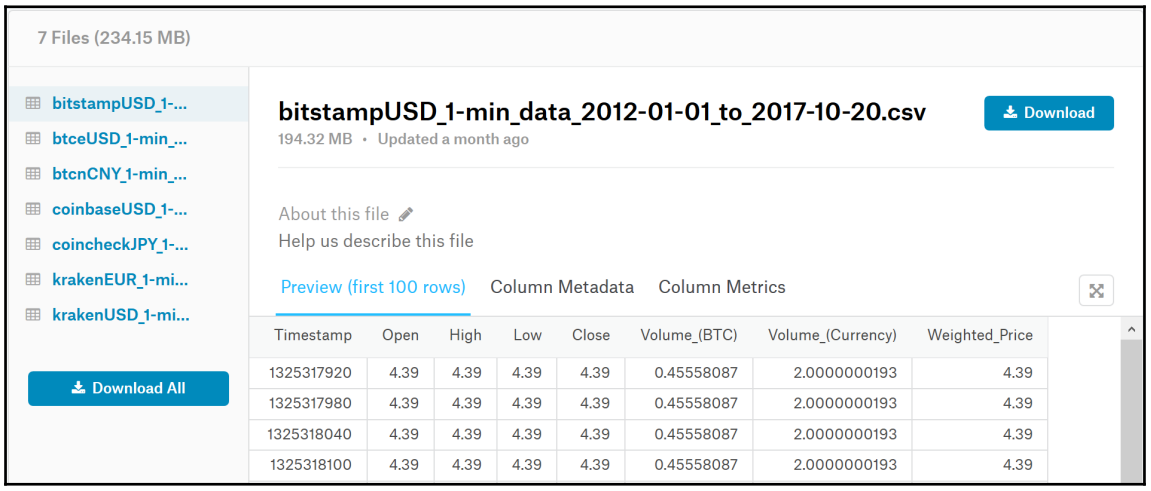
Sell Orders — Total BTC available: 1750.71805785

Price per BTC	BTC Amount	Total: (USD)
7939.9995	Ƀ0.01000000	$ 79.40
7939.9996	Ƀ0.20089966	$ 1595.15
7940.0000	Ƀ0.97584509	$ 7748.22
7940.6878	Ƀ0.07720776	$ 613.09
7942.1143	Ƀ0.12500000	$ 992.77
7946.8971	Ƀ0.21900000	$ 1740.38
7946.8972	Ƀ0.15752729	$ 1251.86
7946.8973	Ƀ0.14285714	$ 1135.28
7949.3878	Ƀ0.14285714	$ 1135.63
7949.6000	Ƀ0.01030000	$ 81.89

Buy Orders — Total USD available: 6696945.67

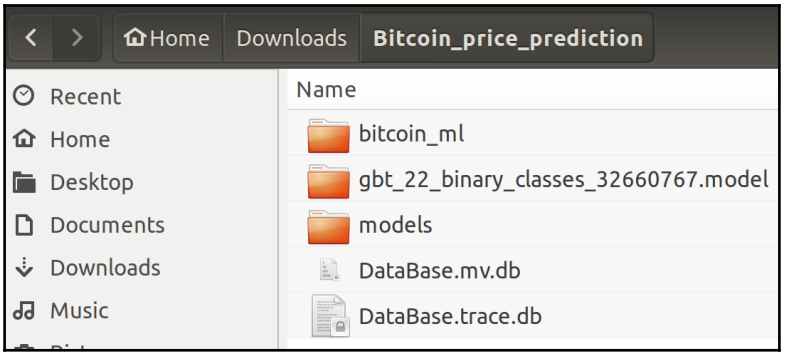
Price per BTC	BTC Amount	Total: (USD)
7928.8630	Ƀ1.50000000	$ 11893.30
7925.9968	Ƀ0.20000000	$ 1585.20
7922.0025	Ƀ0.12085076	$ 957.39
7922.0002	Ƀ0.07081516	$ 561.00
7922.0000	Ƀ0.01222727	$ 96.87
7920.9969	Ƀ0.14285714	$ 1131.58
7920.0002	Ƀ0.15823611	$ 1253.24
7920.0000	Ƀ0.01912455	$ 151.47
7912.7966	Ƀ0.06831213	$ 540.54
7912.7964	Ƀ0.14285714	$ 1130.40
7911.0002	Ƀ0.27142679	$ 2147.26

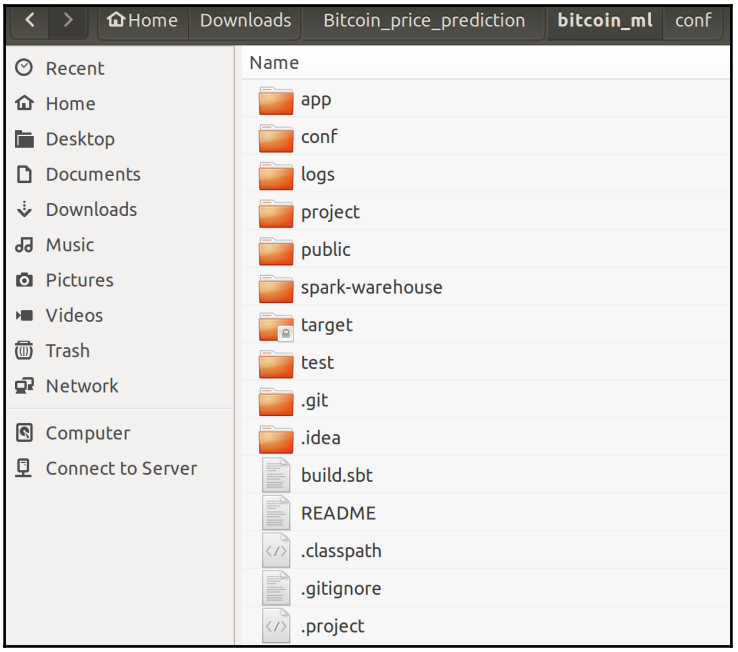

Kaggle
Historical data
Cryptocompare
Live data
Web application on Browser
Time series with labels
Scala script
Scala Train
Play web
X.cs
M1
M3
Models performance
Trained models


7 Files (234.15 MB)
bitstampUSD_1-...
btceUSD_1-min_...
btcnCNY_1-min_...
coinbaseUSD_1-...
coincheckJPY_1-...
krakenEUR_1-mi...
krakenUSD_1-mi...
Download All
bitstampUSD_1-min_data_2012-01-01_to_2017-10-20.csv
194.32 MB • Updated a month ago
Download
About this file
Help us describe this file
Preview (first 100 rows)
Column Metadata
Column Metrics
Timestamp Open High Low Close Volume_(BTC) Volume_(Currency) Weighted_Price
1325317920 4.39 4.39 4.39 4.39 0.45558087 2.0000000193 4.39
1325317980 4.39 4.39 4.39 4.39 0.45558087 2.0000000193 4.39
1325318040 4.39 4.39 4.39 4.39 0.45558087 2.0000000193 4.39
1325318100 4.39 4.39 4.39 4.39 0.45558087 2.0000000193 4.39

```
+----------+----+----+----+-----+-----------+----------------+-------------+
| Timestamp|Open|High| Low|Close|Volume_(BTC)|Volume_(Currency)|Weighted_Price|
+----------+----+----+----+-----+-----------+----------------+-------------+
|1325317920|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325317980|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318040|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318100|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318160|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318220|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318280|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318340|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318400|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
|1325318460|4.39|4.39|4.39| 4.39|  0.45558087|     2.0000000193|         4.39|
+----------+----+----+----+-----+-----------+----------------+-------------+
only showing top 10 rows
```

Parameter	Type	Required	Default	Info
fsym	string	true		From Symbol
tsym	string	true		To Symbols
e	string	true	CCCAGG	Name of exchange
extraParams	string	false	NotAvailable	Name of your application
sign	bool	false	false	If set to true, the server will sign the requests.
tryConversion	bool	false	true	If set to false, it will try to get values without using any conversion at all
aggregate	int	false	1	
limit	int	false	1440	Max 2000
toTs	timestamp	false		


Home
Downloads
Bitcoin_price_prediction
Recent
Home
Desktop
Documents
Downloads
Music
Name
bitcoin_ml
gbt_22_binary_classes_32660767.model
models
DataBase.mv.db
DataBase.trace.db


Home
Downloads
Bitcoin_price_prediction
bitcoin_ml
conf
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
Name
app
conf
logs
project
public
spark-warehouse
target
test
.git
.idea
build.sbt
README
.classpath
.gitignore
.project

```
abel for timestamp: 1516836240Details: probability distribution: [0.6036457093603714,0.39635429063962857]
 label: 0.0 rawPrediction: [0.21033948366056338,-0.21033948366056338]
[info] application - we should not buy -> price will not increase
[info] application - SchedulerActor job runs at 1516836380451
[info] application - __cv__gbt_22_binary_classes_95522293.model has predicted 1 label, actual price delta
 was 0.1999999999989086
[info] application - __cv__gbt_22_binary_classes_94940481.model has predicted 1 label, actual price delta
 was 0.1999999999989086
[info] application - gbt_22_binary_classes_32660767.model has predicted 0 label, actual price delta was 0
.1999999999989086
[info] application - Predicting next price change: gbt_22_binary_classes_32660767.model has predicted 1 l
abel for timestamp: 1516836300Details: probability distribution: [0.46438621191794804,0.535613788082052]
label: 1.0 rawPrediction: [-0.07134839872345475,0.07134839872345475]
[info] application - we should buy -> price will go up
[info] application - SchedulerActor job runs at 1516836410452
[info] application - __cv__gbt_22_binary_classes_95522293.model has predicted 1 label, actual price delta
 was 0.1999999999989086
[info] application - __cv__gbt_22_binary_classes_94940481.model has predicted 1 label, actual price delta
 was 0.1999999999989086
[info] application - gbt_22_binary_classes_32660767.model has predicted 0 label, actual price delta was 0
.1999999999989086
[info] application - Predicting next price change: gbt_22_binary_classes_32660767.model has predicted 1 l
abel for timestamp: 1516836300Details: probability distribution: [0.46438621191794804,0.535613788082052]
label: 1.0 rawPrediction: [-0.07134839872345475,0.07134839872345475]
```

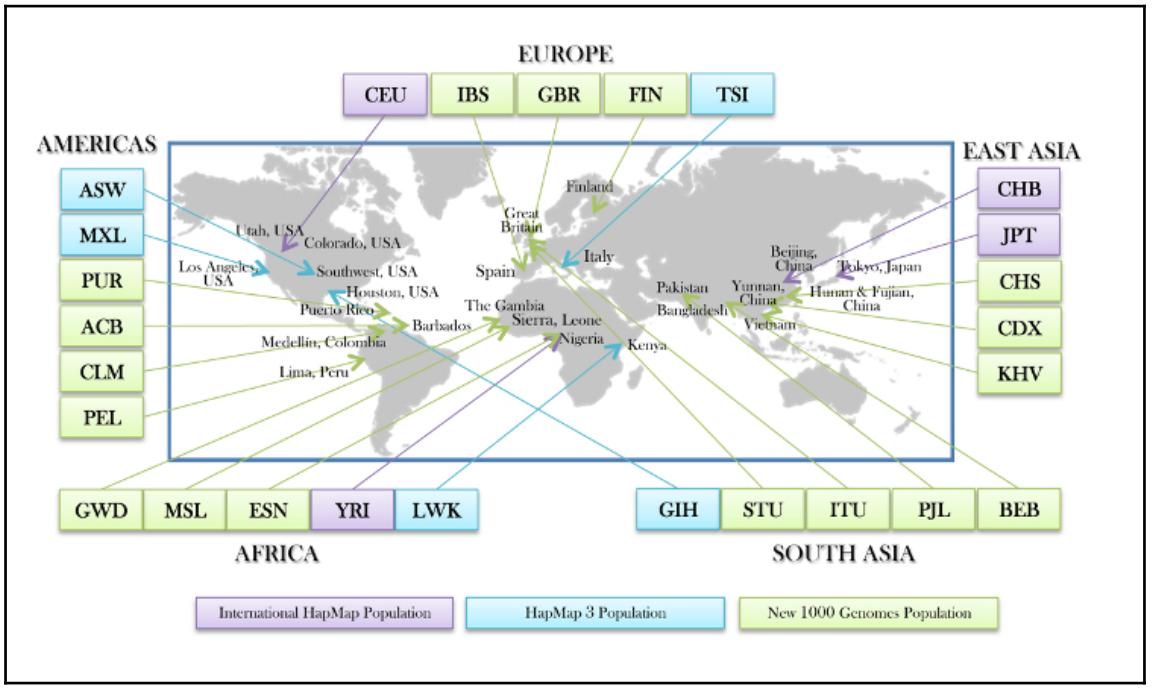
Signal

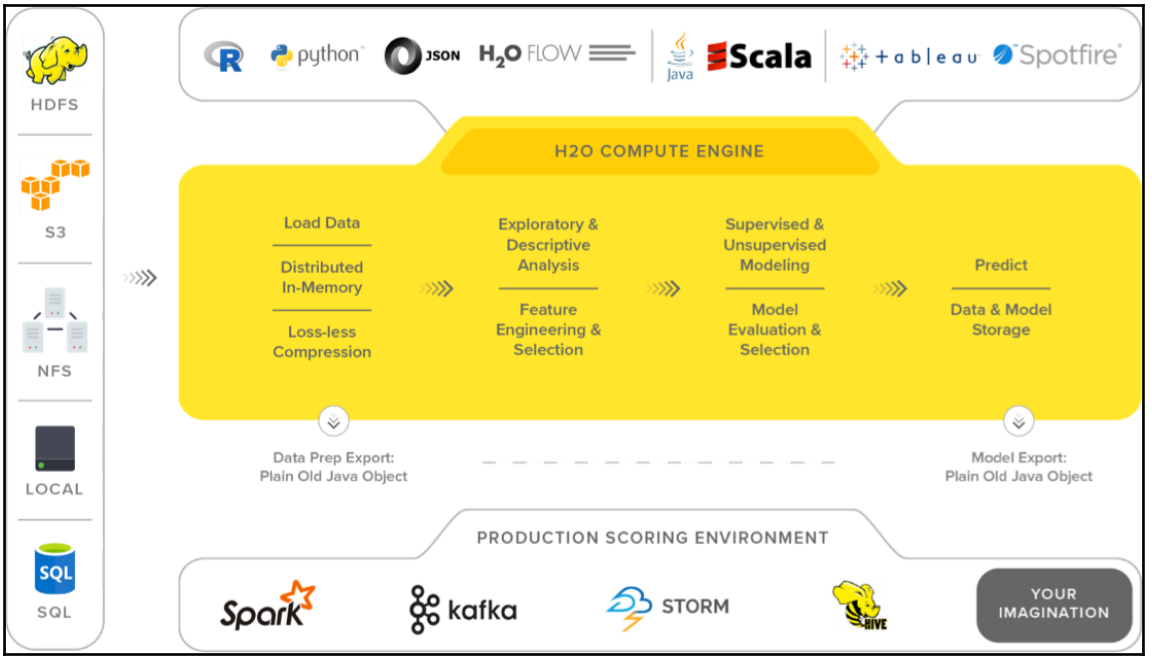
Raw prediction

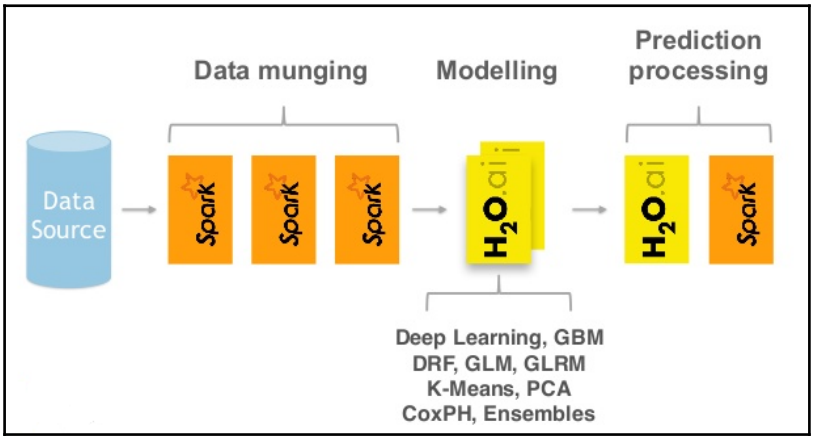
Bitcoin Price Prediction | Inbox - jason.jboss.smit | localhost:9000

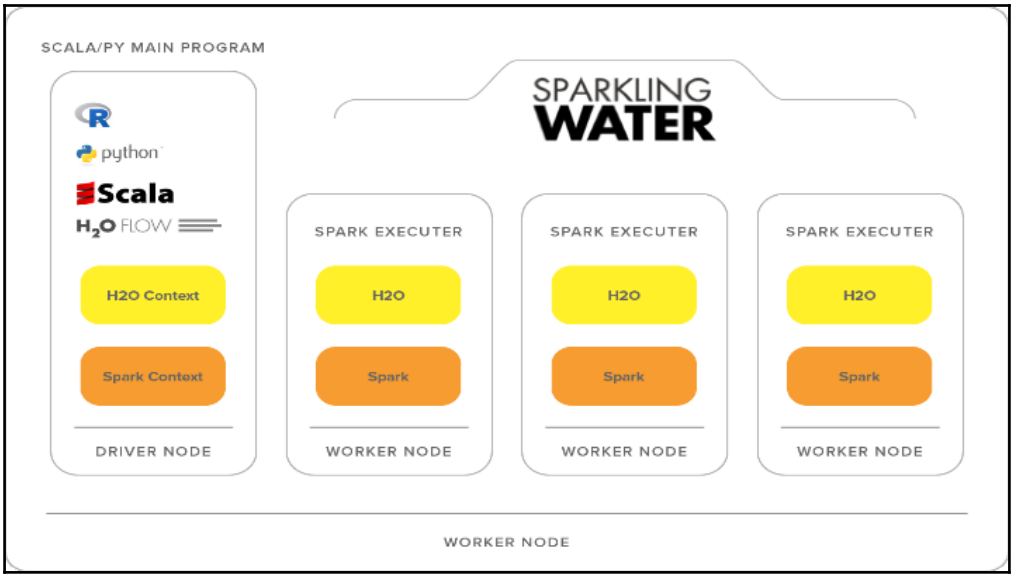
Model Version	True Positive Rate	False Positive Rate	True Negative Rate	False Negative Rate	Total Predictions
gbt_22_binary_classes_32660767.model	24	16	199	124	363
__cv__gbt_22_binary_classes_95522293.model	78	125	90	70	363
__cv__gbt_22_binary_classes_94940481.model	78	125	90	70	363

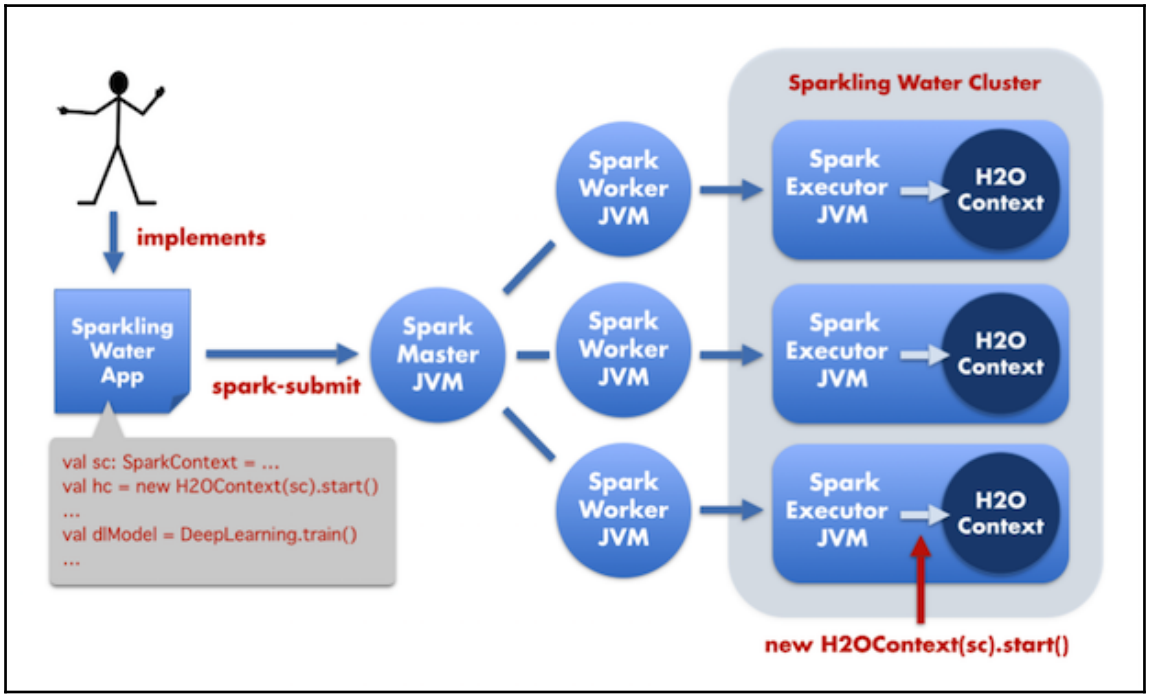
Chapter 4: Population-Scale Clustering and Ethnicity Prediction

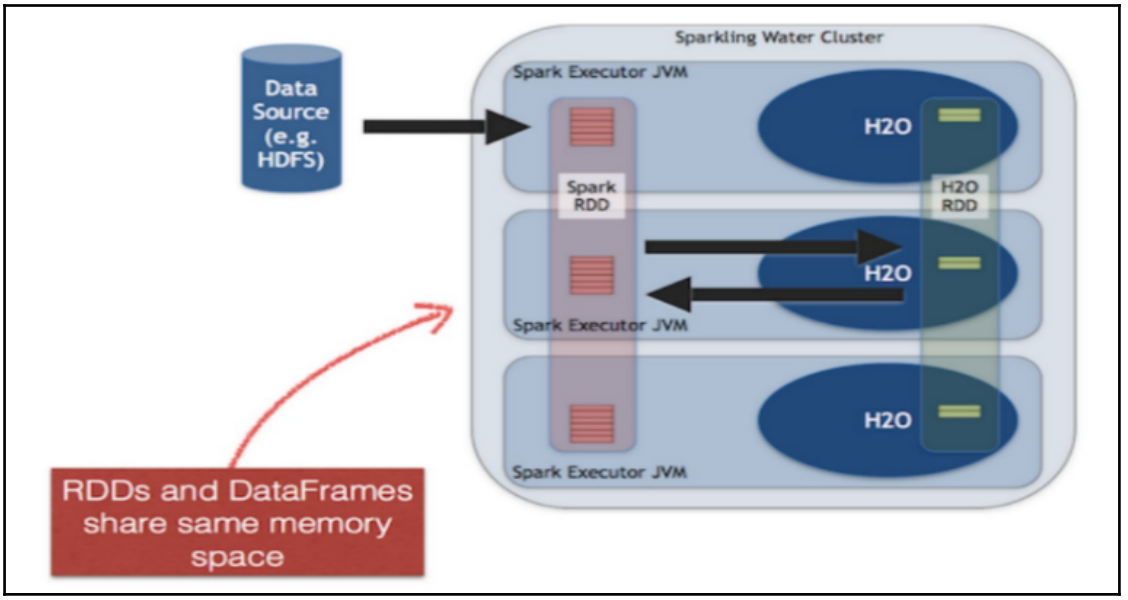


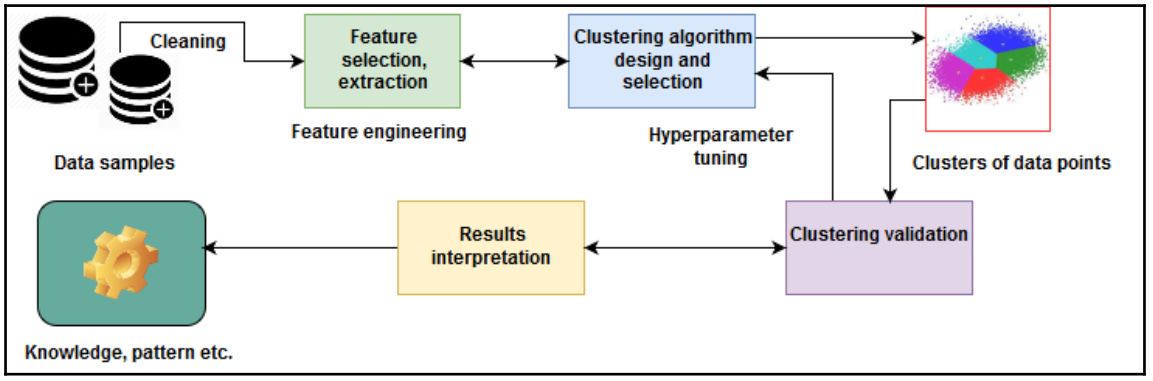

HDFS
S3
NFS
LOCAL
SQL
python
JSON
H2O FLOW
Java
Scala
+ableau
Spotfire
H2O COMPUTE ENGINE
Load Data
Distributed In-Memory
Loss-less Compression
Exploratory & Descriptive Analysis
Feature Engineering & Selection
Supervised & Unsupervised Modeling
Model Evaluation & Selection
Predict
Data & Model Storage
Data Prep Export: Plain Old Java Object
Model Export: Plain Old Java Object
PRODUCTION SCORING ENVIRONMENT
Spark
kafka
STORM
YOUR IMAGINATION

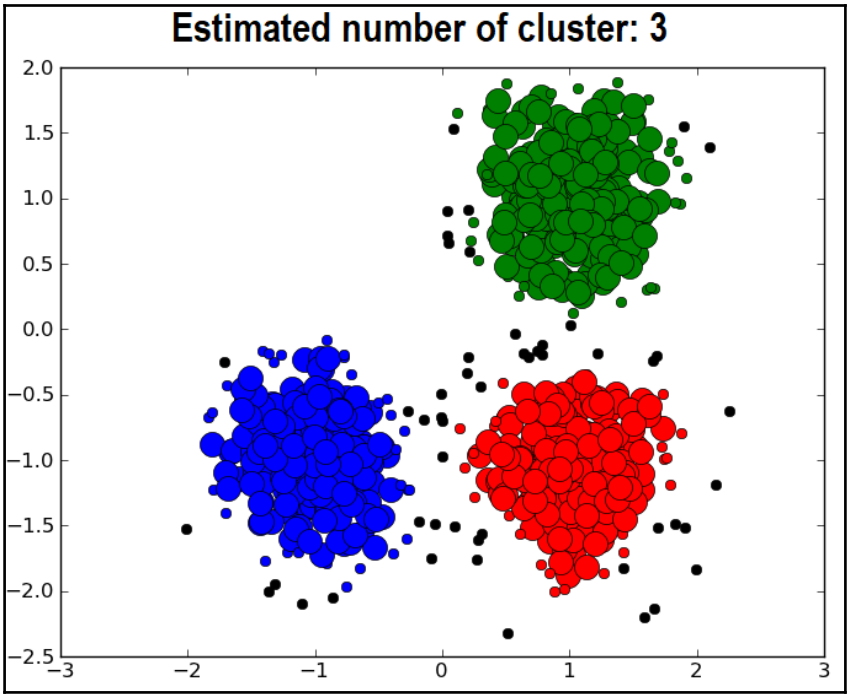

Data munging
Modelling
Prediction processing
Data Source
Spark
Spark
Spark
H2O.ai
H2O.ai
Spark
Deep Learning, GBM
DRF, GLM, GLRM
K-Means, PCA
CoxPH, Ensembles

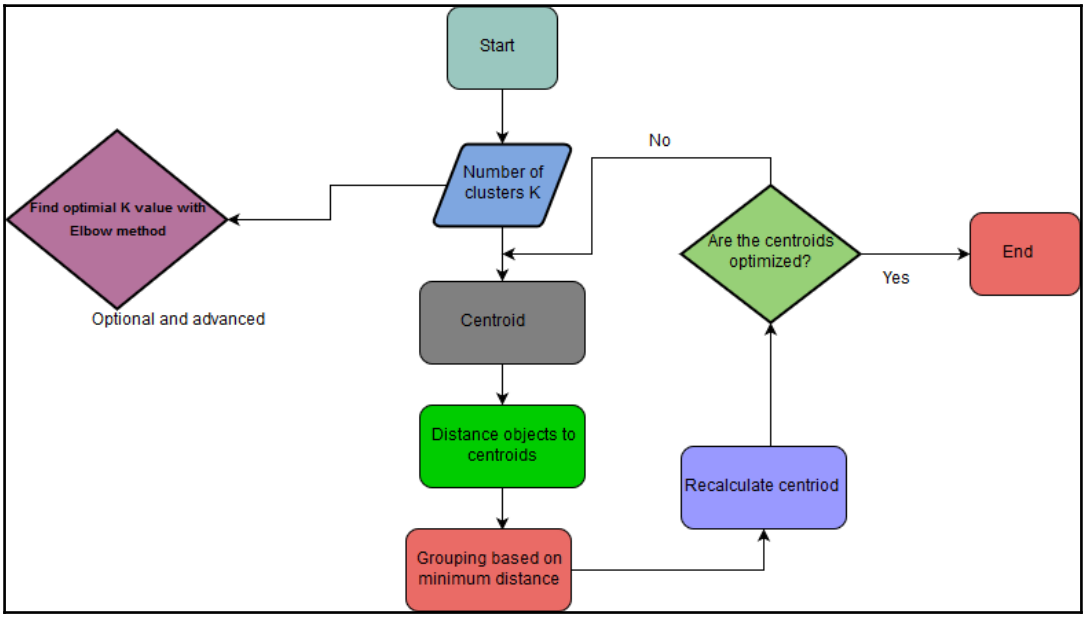

SCALA/PY MAIN PROGRAM
python
Scala
H2O FLOW
H2O Context
Spark Context
DRIVER NODE
SPARKLING WATER
SPARK EXECUTER
H2O
Spark
WORKER NODE
SPARK EXECUTER
H2O
Spark
WORKER NODE
SPARK EXECUTER
H2O
Spark
WORKER NODE
WORKER NODE

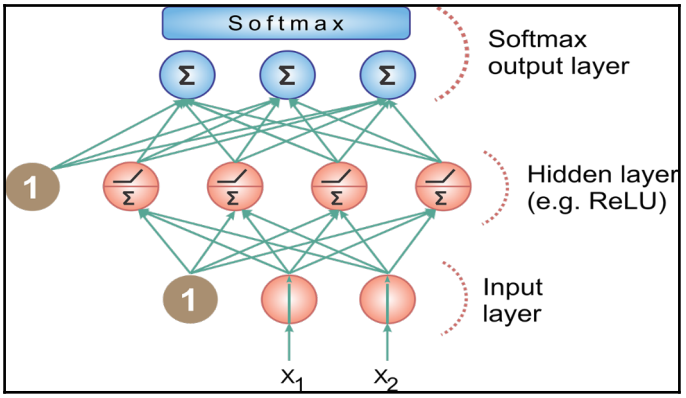

implements
Sparkling Water App
spark-submit
Spark Master JVM
Spark Worker JVM
Spark Worker JVM
Spark Worker JVM
Sparkling Water Cluster
Spark Executor JVM
H2O Context
Spark Executor JVM
H2O Context
Spark Executor JVM
H2O Context
new H2OContext(sc).start()
val sc: SparkContext = ...
val hc = new H2OContext(sc).start()
...
val dlModel = DeepLearning.train()
...

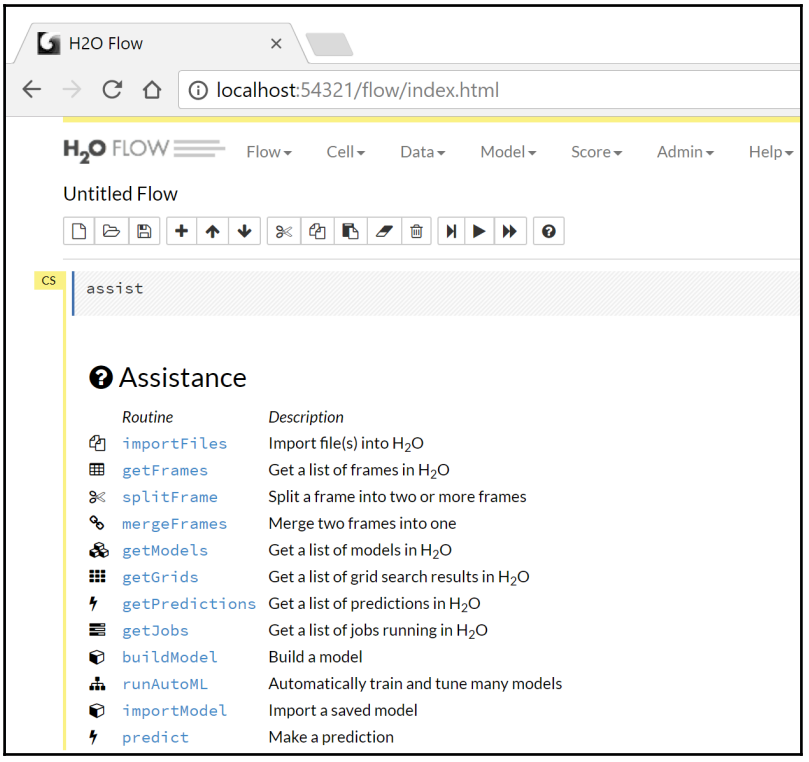

Sparkling Water Cluster
Data Source (e.g. HDFS)
Spark Executor JVM
H2O
Spark RDD
H2O RDD
H2O
Spark Executor JVM
H2O
Spark Executor JVM
RDDs and DataFrames share same memory space


Cleaning
Feature selection, extraction
Clustering algorithm design and selection
Data samples
Feature engineering
Hyperparameter tuning
Clusters of data points
Results interpretation
Clustering validation
Knowledge, pattern etc.

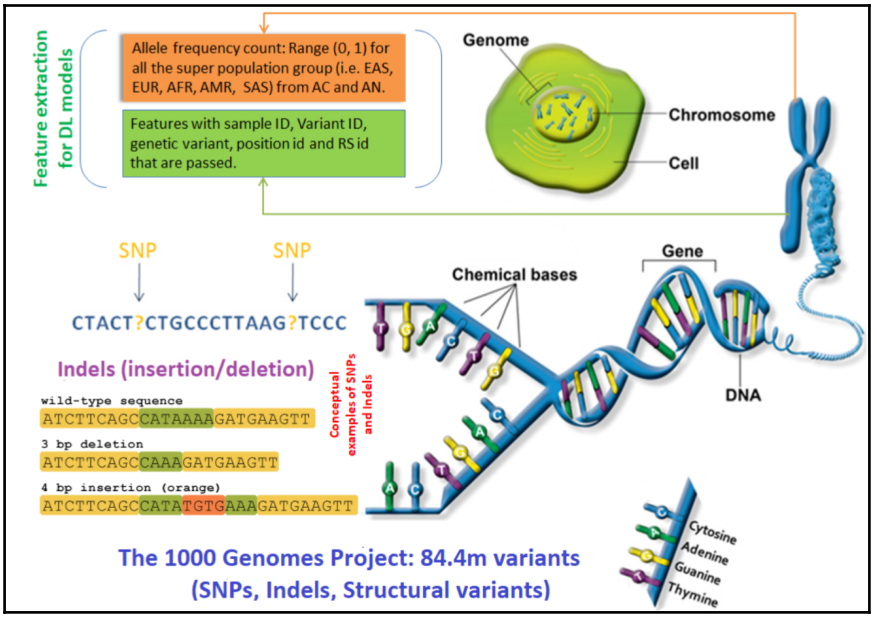

Estimated number of cluster: 3
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3
−2
−1
0
1
2
3

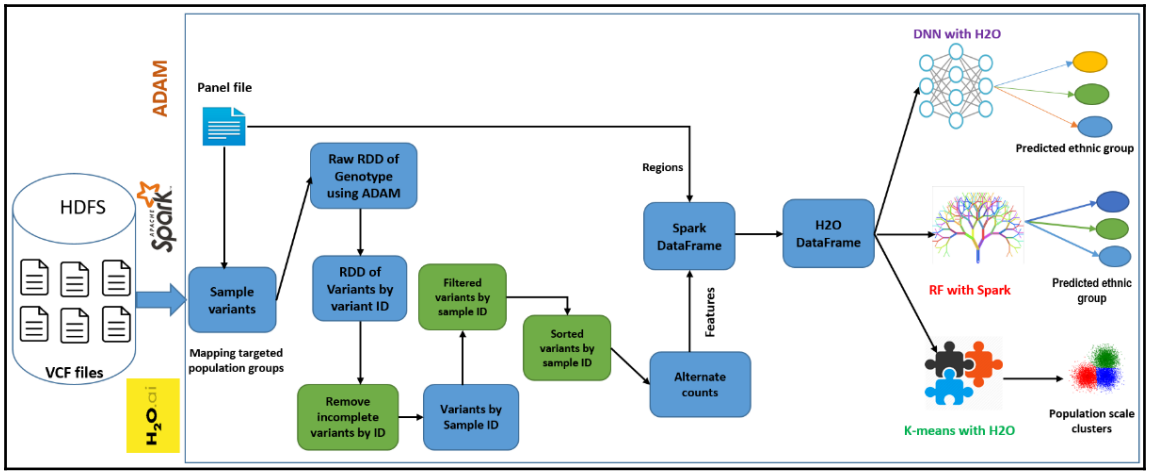

Start
Number of clusters K
Find optimial K value with Elbow method
Optional and advanced
No
Are the centroids optimized?
Yes
End
Centroid
Distance objects to centroids
Grouping based on minimum distance
Recalculate centriod





Sample ID	Pop Group	Ethnicity	Super pop. group	Gender
HG00096	GBR	British in England and Scotland	EUR	male
HG00171	FIN	Finnish in Finland	EUR	female
HG00472	CHS	Southern Han Chinese	EAS	male
HG00551	PUR	Puerto Ricans from Puerto Rico	AMR	female


Feature extraction for DL models
Allele frequency count: Range (0, 1) for all the super population group (i.e. EAS, EUR, AFR, AMR, SAS) from AC and AN.
Features with sample ID, Variant ID, genetic variant, position id and RS id that are passed.
Genome
Chromosome
Cell
SNP
SNP
CTACT?CTGCCCTTAAG?TCCC
Indels (insertion/deletion)
wild-type sequence
ATCTTCAGCCATAAAAGATGAAGTT
3 bp deletion
ATCTTCAGCCAAAGATGAAGTT
4 bp insertion (orange)
ATCTTCAGCCATATGTGAAAGATGAAGTT
Conceptual examples of SNPs and Indels
Chemical bases
Gene
DNA
Cytosine
Adenine
Guanine
Thymine
The 1000 Genomes Project: 84.4m variants
(SNPs, Indels, Structural variants)


ADAM
Panel file
HDFS
VCF files
Sample variants
Mapping targeted population groups
Raw RDD of Genotype using ADAM
RDD of Variants by variant ID
Remove incomplete variants by ID
Filtered variants by sample ID
Variants by Sample ID
Sorted variants by sample ID
Regions
Features
Spark DataFrame
Alternate counts
H2O DataFrame
DNN with H2O
Predicted ethnic group
RF with Spark
Predicted ethnic group
K-means with H2O
Population scale clusters

Region	-2099651974	-2086354790	-2051555302	-2033308294	-2031516326	-1948345126
GBR	1	1	0	0	0	0
CHB	0	0	0	0	0	0
CHB	0	0	0	0	0	0
ASW	0	0	0	0	0	0
ASW	0	0	0	0	0	0
CHB	0	0	0	0	0	0
ASW	0	0	0	0	0	1
GBR	0	0	0	0	0	0
ASW	0	0	0	0	0	0
ASW	0	0	0	0	0	1

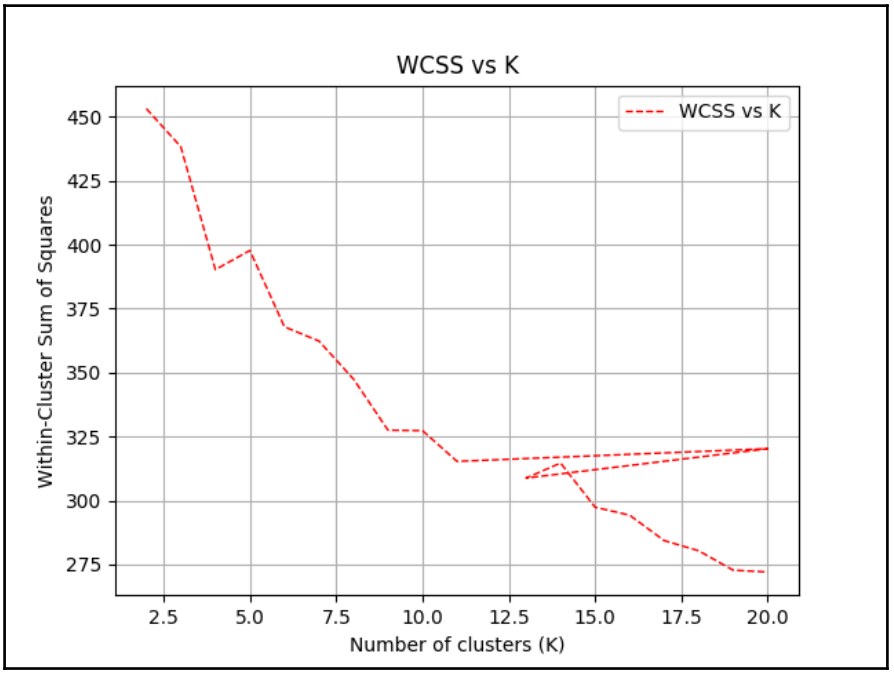
only showing top 10 rows

-2099651974	-2086354790	-2051555302	-2033308294	-2031516326	-1948345126	-1933334022	-1900872614
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	1	1	0

features
(59,[0,1],[1.0,1.0])
(59,[35,51],[1.0,...
(59,[39,42],[1.0,...
(59,[11,16,18,28,...
(59,[9,16],[1.0,1...
(59,[28,35],[1.0,...
(59,[5,9,10,16,28...
(59,[17],[1.0])
(59,[6,9,13,21,44...
(59,[5,6,18,19,21...
(59,[31],[1.0])
(59,[6,11,51],[1....
(59,[13,38,39],[1...
(59,[11,17,28,35,...
(59,[35],[1.0])
(59,[12],[1.0])
(59,[2],[1.0])
(59,[6,10,38,43,4...
(59,[0,1,29,34,51...
(59,[21,26],[1.0,...

only showing top 20 rows

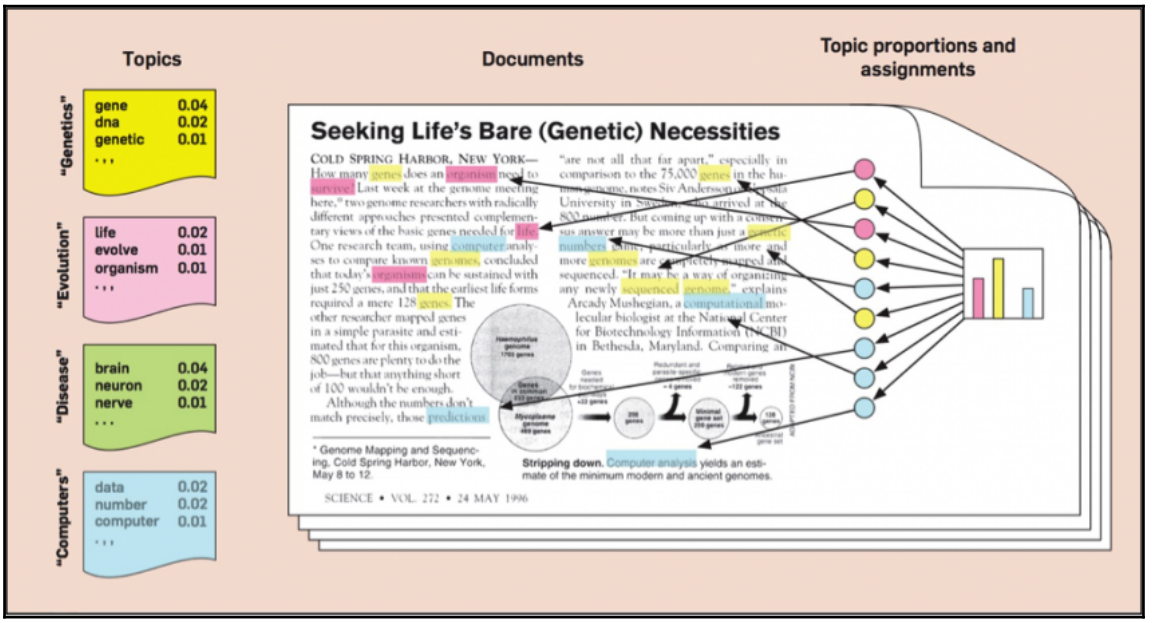
```
+--------------------+
|            features|
+--------------------+
|[-0.0322320102812...|
|[0.04745047735089...|
|[-0.0332140726994...|
|[-0.1110921441100...|
|[-0.0375778701056...|
|[-0.0340456681672...|
|[-0.1083958692025...|
|[-0.0177665347420...|
|[0.05380817444571...|
|[-0.1014376410370...|
|[-0.0178548747978...|
|[0.08328541553353...|
|[0.02980402336806...|
|[-0.1038847392578...|
|[-0.0155973063597...|
|[0.01140843693081...|
|[-0.0158289900287...|
|[0.03367559270881...|
|[-0.0024716378274...|
|[-0.0346824191993...|
+--------------------+
only showing top 20 rows
```

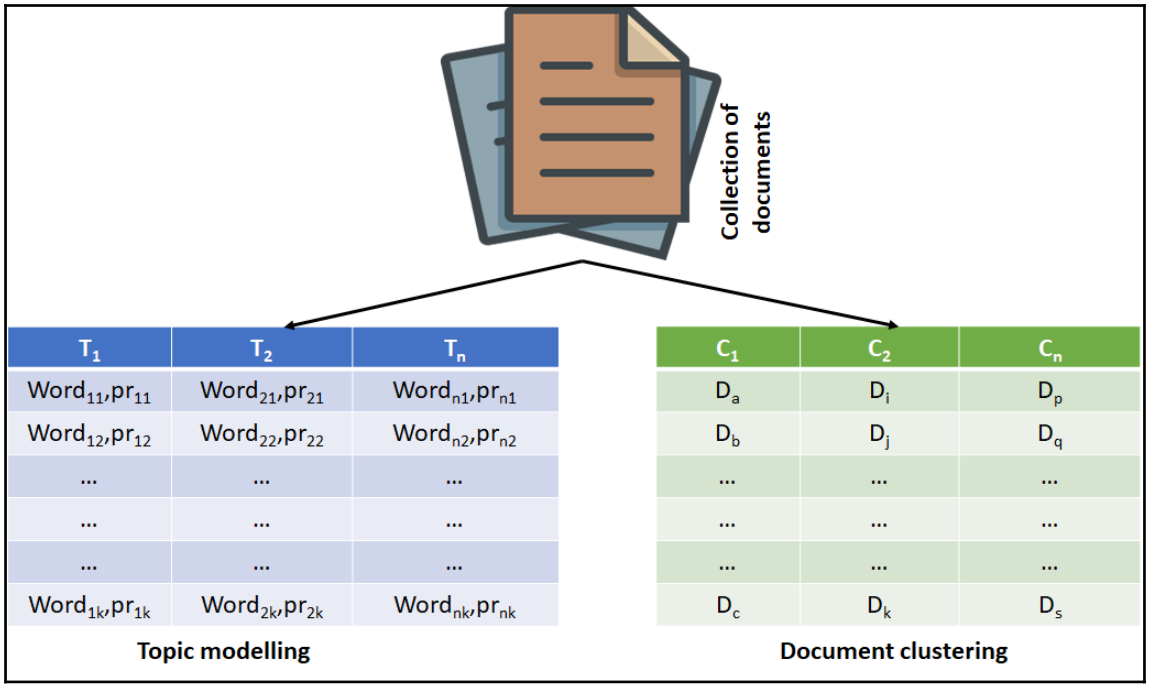


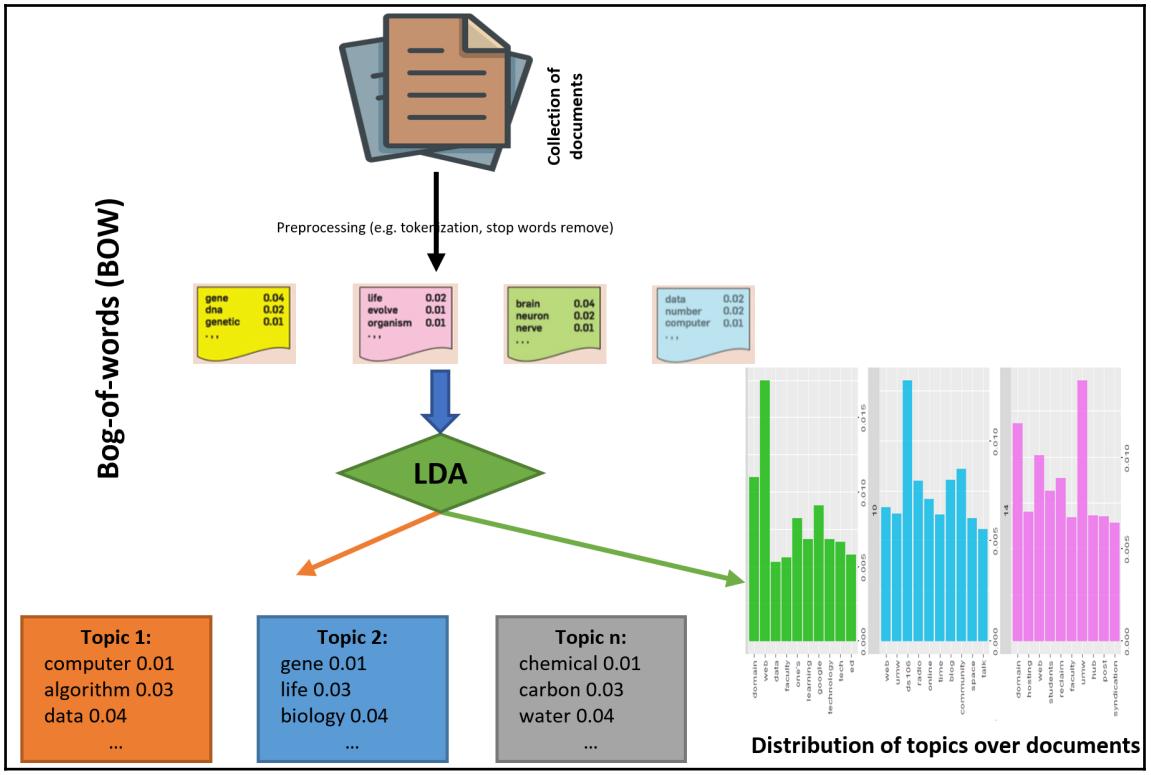
```
14436  FJ-1-13   INFO:    Variable Relative Importance Scaled Importance Percentage
14436  FJ-1-13   INFO:  1158554348            1.000000          1.000000   0.048158
14436  FJ-1-13   INFO:   382875466            0.875208          0.875208   0.042148
14436  FJ-1-13   INFO: -2090607845            0.731845          0.731845   0.035244
14436  FJ-1-13   INFO:  -819576211            0.672748          0.672748   0.032398
14436  FJ-1-13   INFO: -1260378822            0.638622          0.638622   0.030755
14436  FJ-1-13   INFO:  2007209848            0.610795          0.610795   0.029414
14436  FJ-1-13   INFO: -1335940486            0.599973          0.599973   0.028893
14436  FJ-1-13   INFO: -1322225030            0.594180          0.594180   0.028614
14436  FJ-1-13   INFO:   450084336            0.586172          0.586172   0.028229
14436  FJ-1-13   INFO:  -255958118            0.578301          0.578301   0.027850
14436  FJ-1-13   INFO: ---
14436  FJ-1-13   INFO: -1900872614            0.211800          0.211800   0.010200
14436  FJ-1-13   INFO: -1867424774            0.201258          0.201258   0.009692
14436  FJ-1-13   INFO:  1354464014            0.195077          0.195077   0.009394
14436  FJ-1-13   INFO:   -41485958            0.194468          0.194468   0.009365
14436  FJ-1-13   INFO:  -626609286            0.185900          0.185900   0.008952
14436  FJ-1-13   INFO:  -512787196            0.175550          0.175550   0.008454
14436  FJ-1-13   INFO: -1042530870            0.171109          0.171109   0.008240
14436  FJ-1-13   INFO:    82236026            0.168544          0.168544   0.008117
14436  FJ-1-13   INFO:  1956396058            0.156438          0.156438   0.007534
14436  FJ-1-13   INFO:  1607716763            0.146275          0.146275   0.007044
```

```
+--------------------+-----+--------------------+--------------------+----------+
|            features|label|       rawPrediction|         probability|prediction|
+--------------------+-----+--------------------+--------------------+----------+
|(53,[1,3,4,14,21,...|  1.0|[0.38297872340425...|[0.02553191489361...|       2.0|
|(53,[2,8,25,26,27...|  4.0|[0.36666666666666...|[0.02444444444444...|       4.0|
|(53,[3,14,44],[1....|  1.0|[3.90762287854543...|[0.26050819190302...|       2.0|
|(53,[13,34,41],[1...|  4.0|[1.68452289397281...|[0.11230152626485...|       4.0|
|     (53,[15],[1.0])|  0.0|[5.12742506561441...|[0.34182833770762...|       0.0|
|(53,[15,31],[1.0,...|  0.0|[9.67183173350930...|[0.64478878223395...|       0.0|
|          (53,[],[])|  1.0|[5.51940674120168...|[0.36796044941344...|       0.0|
|          (53,[],[])|  2.0|[5.51940674120168...|[0.36796044941344...|       0.0|
|(53,[2,3,14,21,35...|  2.0|[3.92063117699910...|[0.26137541179994...|       2.0|
|(53,[3,17,18,39,4...|  1.0|[0.8,11.333333333...|[0.05333333333333...|       1.0|
+--------------------+-----+--------------------+--------------------+----------+
only showing top 10 rows
```

Chapter 5: Topic Modeling - A Better Insight into Large-Scale Texts




Collection of documents
T1 T2 Tn
Word11,pr11 Word21,pr21 Wordn1,prn1
Word12,pr12 Word22,pr22 Wordn2,prn2
Word1k,pr1k Word2k,pr2k Wordnk,prnk
Topic modelling
C1 C2 Cn
Da Di Dp
Db Dj Dq
Dc Dk Ds
Document clustering


Collection of documents
Preprocessing (e.g. tokenization, stop words remove)
Bog-of-words (BOW)
gene 0.04
dna 0.02
genetic 0.01
life 0.02
evolve 0.01
organism 0.01
brain 0.04
neuron 0.02
nerve 0.01
data 0.02
number 0.02
computer 0.01
LDA
Topic 1:
computer 0.01
algorithm 0.03
data 0.04
...
Topic 2:
gene 0.01
life 0.03
biology 0.04
...
Topic n:
chemical 0.01
carbon 0.03
water 0.04
...
Distribution of topics over documents

```
     Topic: 0
Terms | Index | Weight
----------------------
space    10665   0.046582
just     10667   0.034397
posted   10637   0.016093
love     10661   0.015652
photo    10639   0.013296
cosmic   10635   0.013212
angry    10656   0.012860
like     10666   0.012629
life     10640   0.012107
time     10662   0.011634
-------------------
Sum:= 0.188459750219041
```

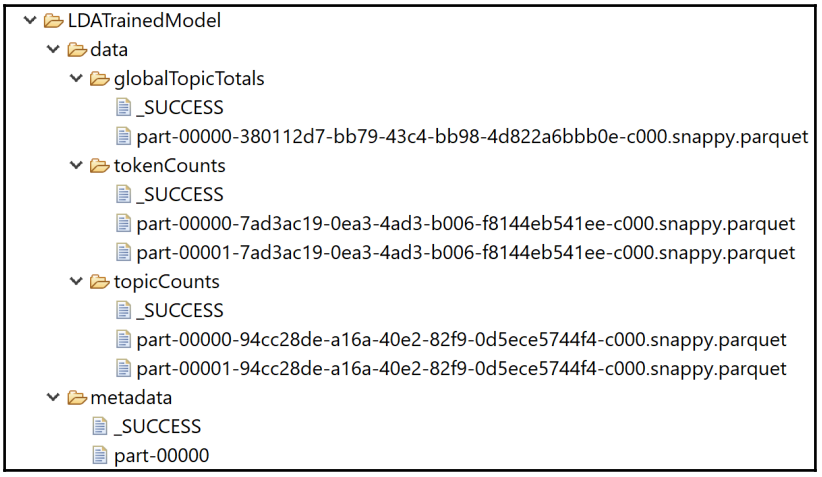
```
      Topic: 1
Terms | Index | Weight
----------------------
smile    10668   0.129227
just     10667   0.024922
good     10663   0.022404
hope     10645   0.017981
going    10655   0.015764
thanks   10648   0.014945
time     10662   0.014941
like     10666   0.014827
think    10659   0.014438
work     10649   0.012702
--------------------
Sum:= 0.28215004471848354
```

```
      Topic: 2
Terms | Index | Weight
-----------------------
grin     10664   0.078958
yang     10628   0.029173
kita     10574   0.017318
disgust          10618   0.016325
udah     10544   0.014584
science          10590   0.012792
space    10665   0.011765
nggak    10501   0.011290
kalo     10476   0.010203
angry    10656   0.009313
--------------------
Sum:= 0.21172148557919923
```

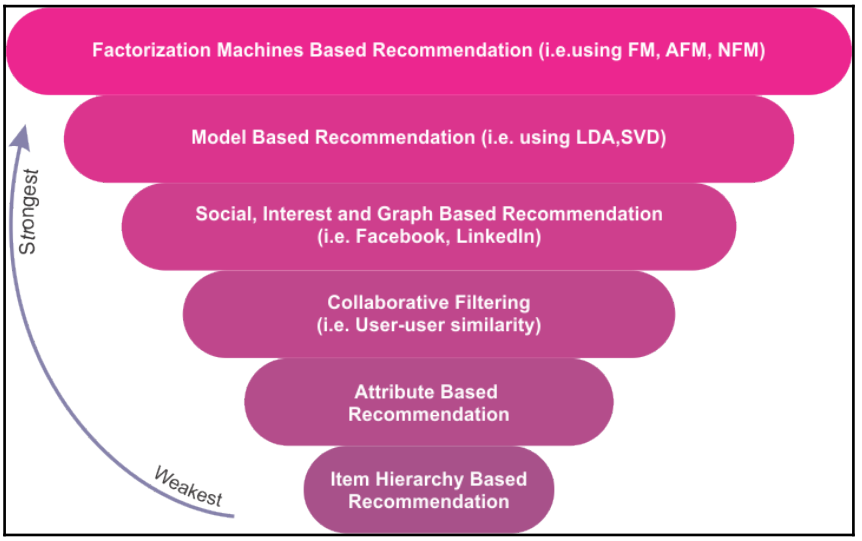
```
      Topic: 3
Terms | Index | Weight
-----------------------
like     10666   0.030890
just     10667   0.020093
know     10660   0.016473
good     10663   0.013343
that     10651   0.012687
people   10658   0.012137
right    10654   0.012097
think    10659   0.011395
love     10661   0.010943
does     10646   0.009002
--------------------
Sum:= 0.14905966677477597
```

```
+--------------------+
|                docs|
+--------------------+
|20 000 2000010 th...|
|fifty two  still ...|
|please this   loo...|
|gable this       ...|
|Sheridan   world ...|
|  wiretap with Ca...|
|                 ...|
|find  plea Lamb h...|
|  find  10th  aes...|
|Disney  alad10 tx...|
| Alcott  Gutenber...|
|   this  hand dom...|
|please this   loo...|
|  HORATIO breaker...|
| HORATIO  ragged ...|
| HORATIO upward s...|
| RESEARCH Electro...|
|tradition Richard...|
|able   across thi...|
+--------------------+
```

ID	Year	Short abstract
1	1978	1H NMR of valine tRNA modified bases. Evidence for multiple conformations.
2	1978	ABBREVIATIONS IN MEDICINE
3	1978	3'End labelling of RNA with 32P suitable for rapid gel sequencing.
4	1978	ABBREVIATIONS IN MEDICINE
5	1978	Absence of plant uptake and translocation of polybrominated biphenyls (PBBs).
6	1978	Abstracts/Sommaires/Zusammenfassungen



Chapter 6: Developing Model-based Movie Recommendation Engines



	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4			5	1		
B	5	5	4				
C				2	4	5	
D		3					3

```
Die Hard (1988) | Fearless (1993) | -0.7674 | -0.4338 | 0.8622 | 0.0549
Die Hard (1988) | How to Make an American Quilt (1995) | -0.5106 | -0.3919 | 0.9056 | 0.1284
Die Hard (1988) | Rich Man's Wife, The (1996) | -0.8807 | -0.3626 | 0.9109 | 0.0303
Die Hard (1988) | Flubber (1997) | -0.5881 | -0.3619 | 0.8630 | 0.0632
Die Hard (1988) | His Girl Friday (1940) | -0.5411 | -0.3608 | 0.9478 | 0.0778
Die Hard (1988) | Believers, The (1987) | -0.8416 | -0.3466 | 0.8789 | 0.0302
Die Hard (1988) | Crow: City of Angels, The (1996) | -0.4753 | -0.3268 | 0.8096 | 0.0913
Die Hard (1988) | Barb Wire (1996) | -0.4983 | -0.3265 | 0.8632 | 0.0805
Die Hard (1988) | Ninotchka (1939) | -0.6143 | -0.3218 | 0.8956 | 0.0464
Die Hard (1988) | Goofy Movie, A (1995) | -0.8528 | -0.3198 | 0.8540 | 0.0253
```

```
Postino, Il (1994) | Volcano (1997) | -0.4861 | -0.3926 | 0.8369 | 0.1419
Postino, Il (1994) | Inventing the Abbotts (1997) | -0.7423 | -0.3888 | 0.9075 | 0.0636
Postino, Il (1994) | Safe (1995) | -0.7462 | -0.3535 | 0.9568 | 0.0549
Postino, Il (1994) | Until the End of the World (Bis ans Ende der Welt) (1991) | -0.8343 | -0.3435 | 0.8963 | 0.0400
Postino, Il (1994) | Cinderella (1950) | -0.4649 | -0.3392 | 0.9236 | 0.1071
Postino, Il (1994) | Man Without a Face, The (1993) | -0.6124 | -0.3340 | 0.8783 | 0.0571
Postino, Il (1994) | Mad City (1997) | -0.7482 | -0.3325 | 0.9370 | 0.0419
Postino, Il (1994) | First Knight (1995) | -0.5258 | -0.3236 | 0.8544 | 0.0714
Postino, Il (1994) | Santa Clause, The (1994) | -0.6433 | -0.3217 | 0.8539 | 0.0433
Postino, Il (1994) | Breakdown (1997) | -0.4763 | -0.3175 | 0.9216 | 0.0948
```

```
Star Wars (1977) | Fathers' Day (1997) | -0.6625 | -0.4417 | 0.9074 | 0.0397
Star Wars (1977) | Jason's Lyric (1994) | -0.9661 | -0.3978 | 0.8110 | 0.0141
Star Wars (1977) | Lightning Jack (1994) | -0.7906 | -0.3953 | 0.9361 | 0.0202
Star Wars (1977) | Marked for Death (1990) | -0.5922 | -0.3807 | 0.8729 | 0.0361
Star Wars (1977) | Mixed Nuts (1994) | -0.6219 | -0.3731 | 0.8806 | 0.0303
Star Wars (1977) | Poison Ivy II (1995) | -0.7443 | -0.3722 | 0.7169 | 0.0201
Star Wars (1977) | In the Realm of the Senses (Ai no corrida) (1976) | -0.8090 | -0.3596 | 0.8108 | 0.0162
Star Wars (1977) | What Happened Was... (1994) | -0.9045 | -0.3392 | 0.8781 | 0.0121
Star Wars (1977) | Female Perversions (1996) | -0.8039 | -0.3310 | 0.8670 | 0.0141
Star Wars (1977) | Celtic Pride (1996) | -0.6062 | -0.3175 | 0.8998 | 0.0220
```

```
+------+-------+------+----------+
|userId|movieId|rating|timestamp |
+------+-------+------+----------+
|1     |16     |4.0   |1217897793|
|1     |24     |1.5   |1217895807|
|1     |32     |4.0   |1217896246|
|1     |47     |4.0   |1217896556|
|1     |50     |4.0   |1217896523|
|1     |110    |4.0   |1217896150|
|1     |150    |3.0   |1217895940|
|1     |161    |4.0   |1217897864|
|1     |165    |3.0   |1217897135|
|1     |204    |0.5   |1217895786|
|1     |223    |4.0   |1217897795|
|1     |256    |0.5   |1217895764|
|1     |260    |4.5   |1217895864|
|1     |261    |1.5   |1217895750|
|1     |277    |0.5   |1217895772|
|1     |296    |4.0   |1217896125|
|1     |318    |4.0   |1217895860|
|1     |349    |4.5   |1217897058|
|1     |356    |3.0   |1217896231|
|1     |377    |2.5   |1217896373|
+------+-------+------+----------+
only showing top 20 rows
```

movieId	title	genres
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action|Crime|Thriller
7	Sabrina (1995)	Comedy|Romance
8	Tom and Huck (1995)	Adventure|Children
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action|Adventure|Thriller
11	American President, The (1995)	Comedy|Drama|Romance
12	Dracula: Dead and Loving It (1995)	Comedy|Horror
13	Balto (1995)	Adventure|Animation|Children
14	Nixon (1995)	Drama
15	Cutthroat Island (1995)	Action|Adventure|Romance
16	Casino (1995)	Crime|Drama
17	Sense and Sensibility (1995)	Drama|Romance
18	Four Rooms (1995)	Comedy
19	Ace Ventura: When Nature Calls (1995)	Comedy
20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller

only showing top 20 rows

title	maxr	minr	cntu
Pulp Fiction (1994)	5.0	0.5	325
Forrest Gump (1994)	5.0	0.5	311
Shawshank Redemption, The (1994)	5.0	0.5	308
Jurassic Park (1993)	5.0	1.0	294
Silence of the Lambs, The (1991)	5.0	0.5	290
Star Wars: Episode IV - A New Hope (1977)	5.0	0.5	273
Matrix, The (1999)	5.0	0.5	261
Terminator 2: Judgment Day (1991)	5.0	0.5	253
Schindler's List (1993)	5.0	0.5	248
Braveheart (1995)	5.0	0.5	248
Fugitive, The (1993)	5.0	1.0	244
Toy Story (1995)	5.0	1.0	232
Usual Suspects, The (1995)	5.0	1.0	228
Star Wars: Episode V - The Empire Strikes Back (1980)	5.0	0.5	228
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)	5.0	1.0	224
Star Wars: Episode VI - Return of the Jedi (1983)	5.0	0.5	222
Batman (1989)	5.0	0.5	217
American Beauty (1999)	5.0	1.0	216
Back to the Future (1985)	5.0	1.5	213
Godfather, The (1972)	5.0	1.0	210

only showing top 20 rows

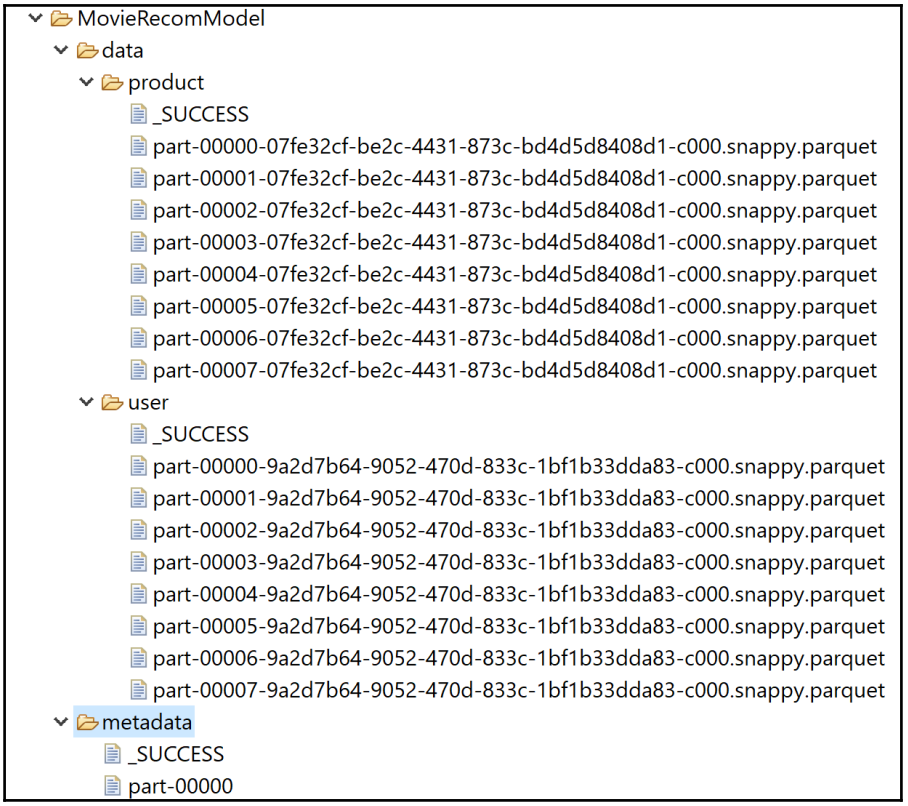
```
+------+----+
|userId|ct  |
+------+----+
|668   |5678|
|575   |2837|
|458   |2086|
|232   |1421|
|310   |1287|
|475   |1249|
|128   |1231|
|224   |1182|
|607   |1176|
|63    |1107|
+------+----+
```

```
+------+-------+------+-----------------------------+
|userId|movieId|rating|title                        |
+------+-------+------+-----------------------------+
|668   |6      |5.0   |Heat (1995)                  |
|668   |858    |5.0   |Godfather, The (1972)        |
|668   |898    |5.0   |Philadelphia Story, The (1940)|
|668   |908    |5.0   |North by Northwest (1959)    |
|668   |910    |5.0   |Some Like It Hot (1959)      |
|668   |912    |5.0   |Casablanca (1942)            |
|668   |913    |5.0   |Maltese Falcon, The (1941)   |
|668   |914    |5.0   |My Fair Lady (1964)          |
|668   |919    |5.0   |Wizard of Oz, The (1939)     |
|668   |1183   |5.0   |English Patient, The (1996)  |
|668   |1204   |5.0   |Lawrence of Arabia (1962)    |
|668   |1217   |5.0   |Ran (1985)                   |
|668   |1219   |5.0   |Psycho (1960)                |
|668   |1256   |5.0   |Duck Soup (1933)             |
|668   |1283   |5.0   |High Noon (1952)             |
|668   |1938   |5.0   |Lost Weekend, The (1945)     |
|668   |1945   |5.0   |On the Waterfront (1954)     |
|668   |1947   |5.0   |West Side Story (1961)       |
|668   |2644   |5.0   |Dracula (1931)               |
|668   |2728   |5.0   |Spartacus (1960)             |
+------+-------+------+-----------------------------+
only showing top 20 rows
```

```
Rating:(UserID, MovieID, Rating)
--------------------------------
Rating(668,5304,4.8525842164777435)
Rating(668,25961,4.8525842164777435)
Rating(668,101862,4.8525842164777435)
Rating(668,80969,4.779325934293423)
Rating(668,93040,4.7528736838886)
Rating(668,25795,4.676957397667861)
--------------------------------
```

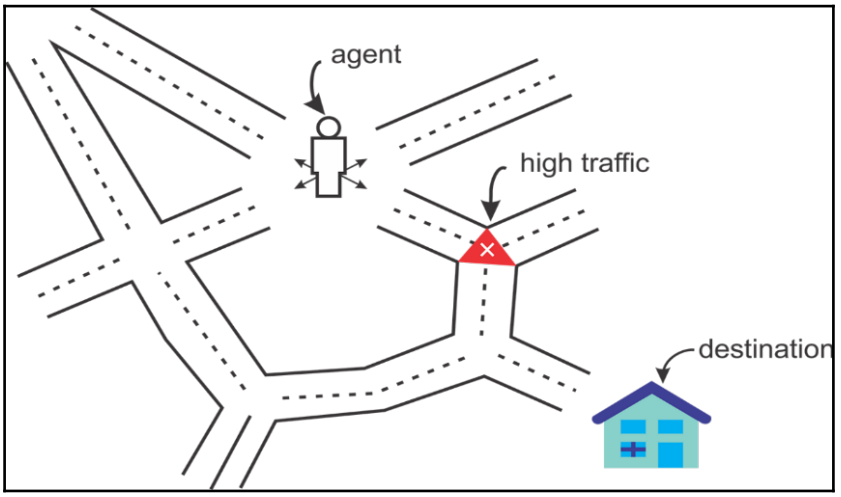
```
(Prediction, Rating)
(3.848087516442212,3.5)
(4.647813269020743,5.0)
(3.578002886107389,4.0)
(3.681217214985231,3.0)
(2.844685318141285,3.0)
```

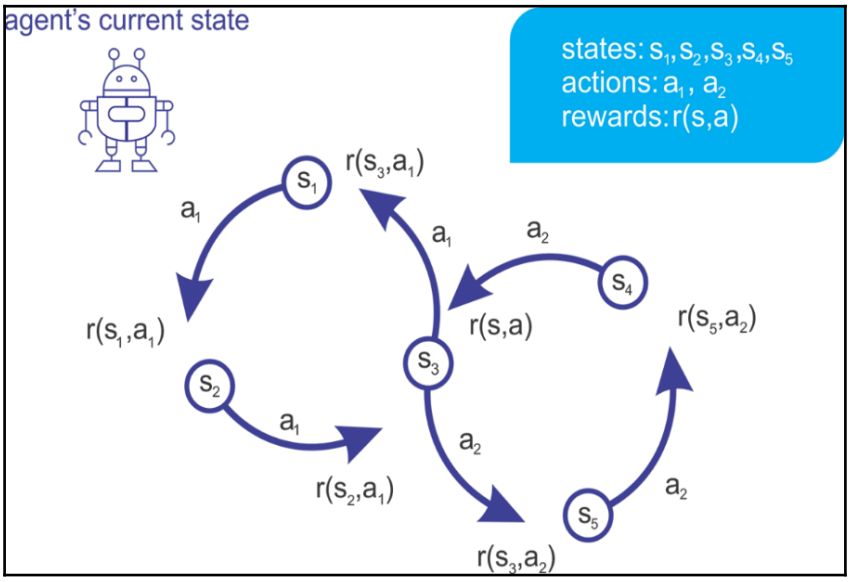
```
Recommendations: (MovieId => Rating)
----------------------------------
(5304,4.8525842164777435)
(25961,4.8525842164777435)
(101862,4.8525842164777435)
(80969,4.779325934293423)
(93040,4.7528736838886)
(25795,4.676957397667861)
----------------------------------
```

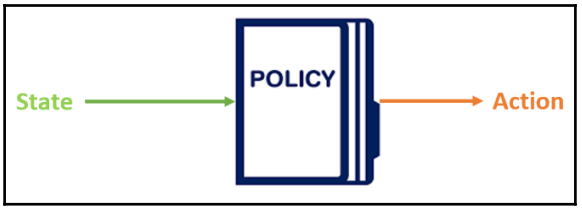


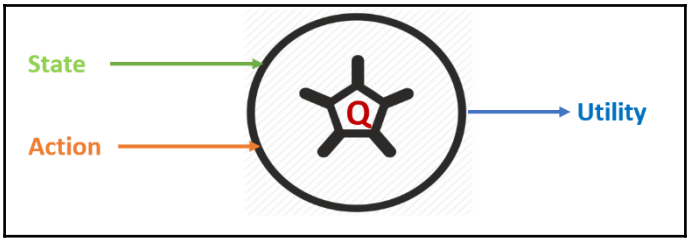
```
Rating:(UserID, MovieID, Rating)
----------------------------------
Rating(458,670,4.687132475422945)
Rating(458,214,4.62504244476065)
Rating(458,3467,4.603944438428056)
Rating(458,858,4.590518214337291)
Rating(458,2745,4.57253593653456)
Rating(458,7158,4.5580179461102235)
Rating(458,1546,4.5526318073335865)
Rating(458,1797,4.551033119658062)
Rating(458,6301,4.550922059713256)
Rating(458,1361,4.522913373522734)
----------------------------------
(Prediction, Rating)
(3.848087516442212,3.5)
(4.647813269020743,5.0)
(3.578002886107389,4.0)
(3.681217214985231,3.0)
(2.844685318141285,3.0)

Test RMSE: = 0.9019872589764074

Recommendations: (MovieId => Rating)
----------------------------------
(670,4.687132475422945)
(214,4.62504244476065)
(3467,4.603944438428056)
(858,4.590518214337291)
(2745,4.57253593653456)
(7158,4.5580179461102235)
(1546,4.5526318073335865)
(1797,4.551033119658062)
(6301,4.550922059713256)
(1361,4.522913373522734)
----------------------------------
```

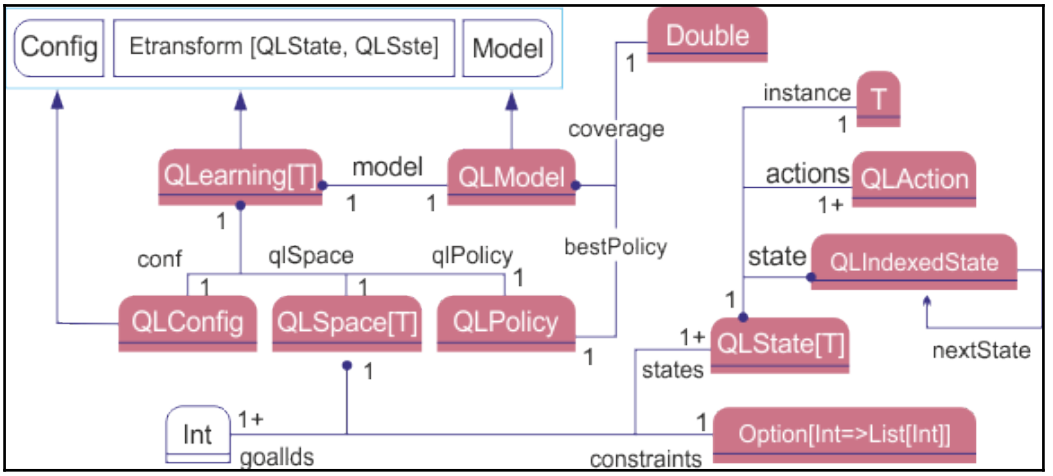
Chapter 7: Options Trading Using Q-learning and Scala Play Framework

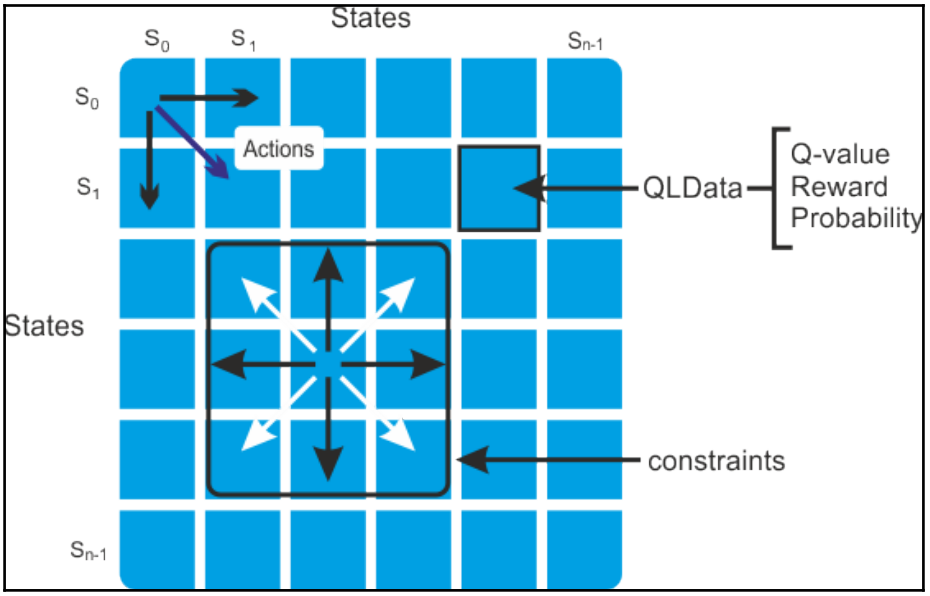


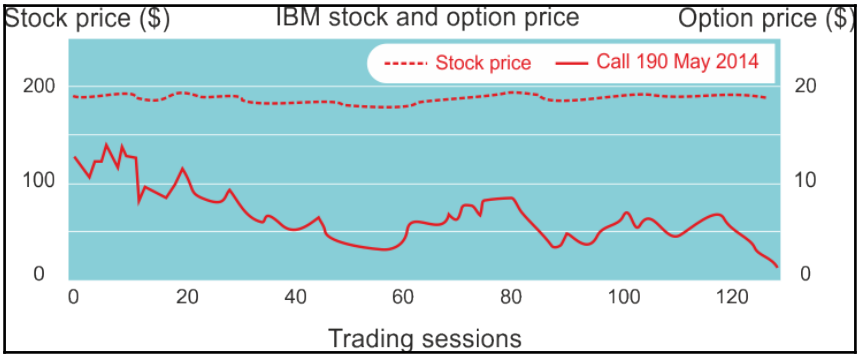


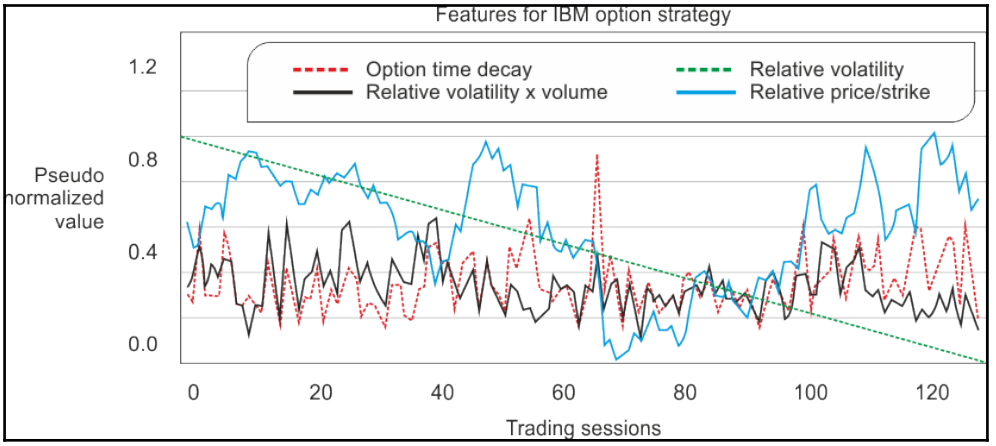

State
POLICY
Action


State
Action
Q
Utility

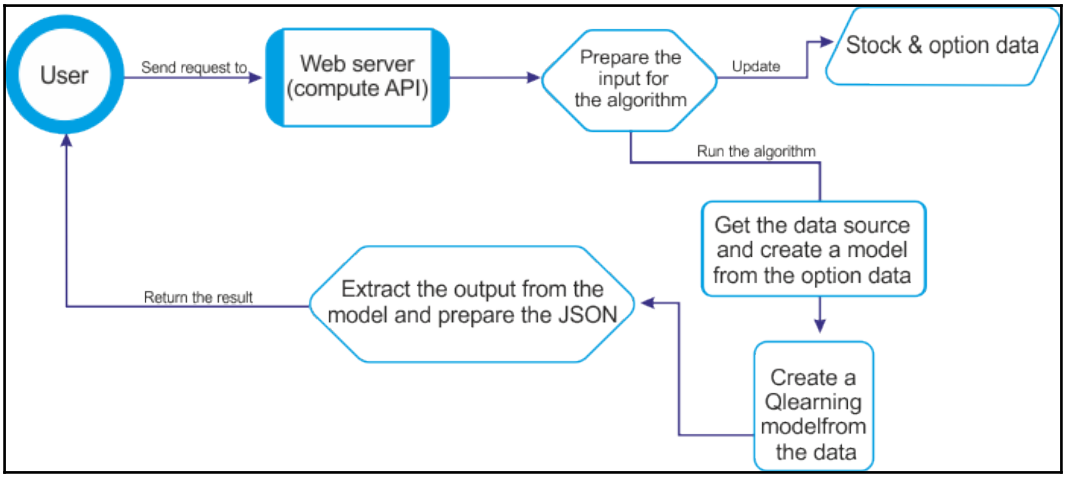

Config
Etransform [QLState, QLSste]
Model
Double
1
coverage
instance
T
1
QLearning[T]
model
QLModel
1
1
1
actions
QLAction
1+
conf
qlSpace
qlPolicy
bestPolicy
state
QLIndexedState
1
1
1
1
QLConfig
QLSpace[T]
QLPolicy
1+
QLState[T]
nextState
1
1
states
Int
1+
goalIds
1
Option[Int=>List[Int]]
constraints

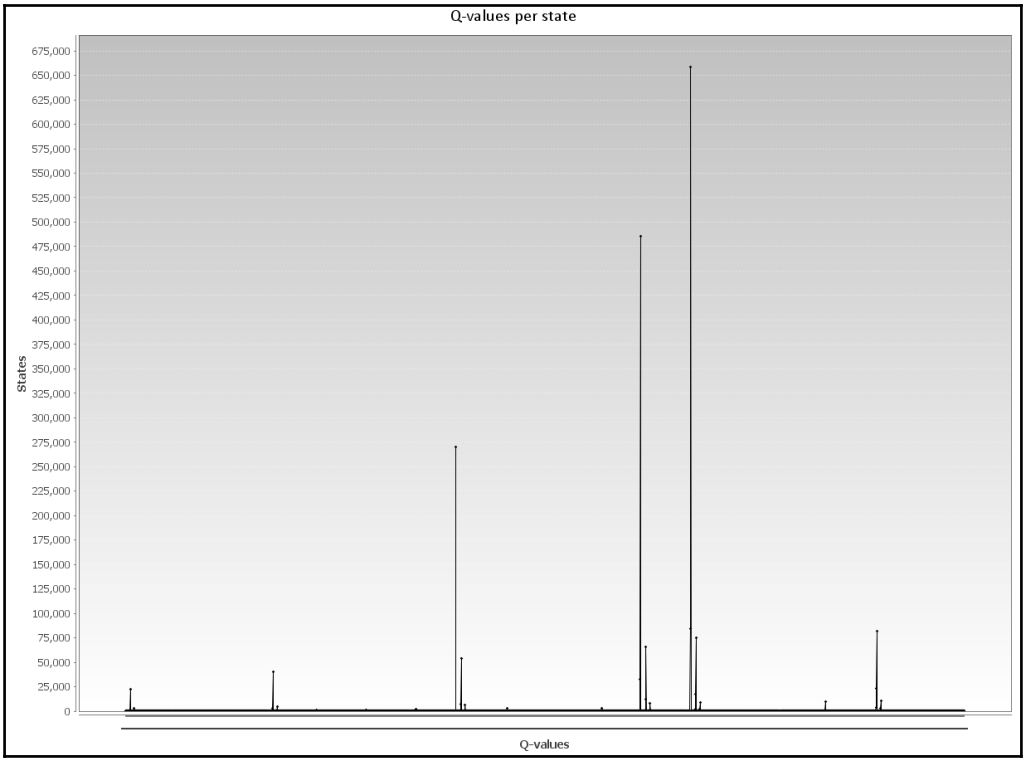

States
S_0
S_1
S_{n-1}
S_0
S_1
Actions
QLData
Q-value
Reward
Probability
States
constraints
S_{n-1}

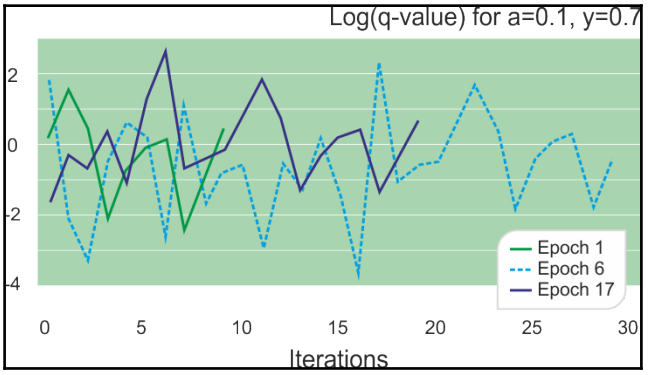

Stock price ($)
IBM stock and option price
Option price ($)
Stock price
Call 190 May 2014
200
100
0
20
10
0
0
20
40
60
80
100
120
Trading sessions

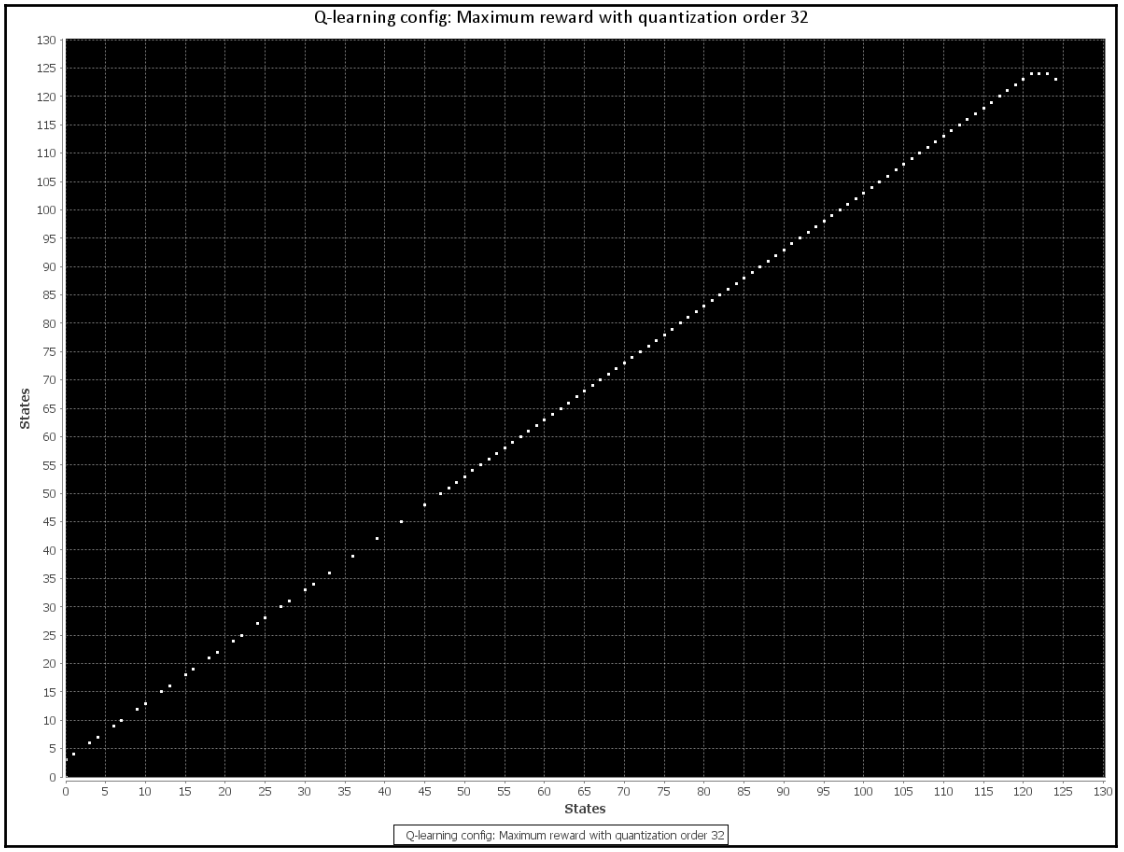


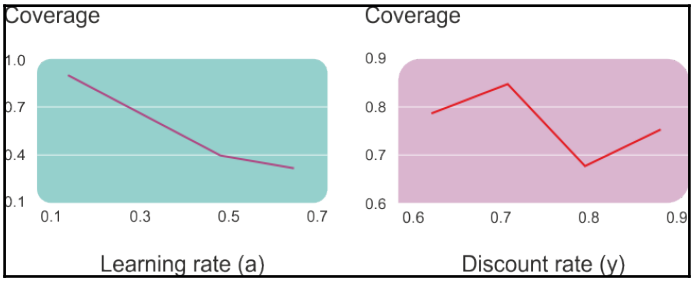
Date	Open	High	Low	Close	Volume	Adj Close
2014-10-03	188.11	189.37	187.56	188.67	3071500	188.67
2014-10-02	187.66	187.78	186.24	186.91	2283100	186.91
2014-10-01	189.91	190.40	186.79	187.17	3705400	187.17
2014-09-30	189.64	190.85	189.15	189.83	2870300	189.83
2014-09-29	188.51	189.96	188.12	189.64	2336300	189.64
2014-09-26	188.93	190.33	188.61	190.06	2493900	190.06
2014-09-25	192.05	192.50	188.97	189.01	4151400	189.01
2014-09-24	191.00	192.45	189.88	192.31	3082600	192.31
2014-09-23	192.75	193.07	191.52	191.62	3301800	191.62

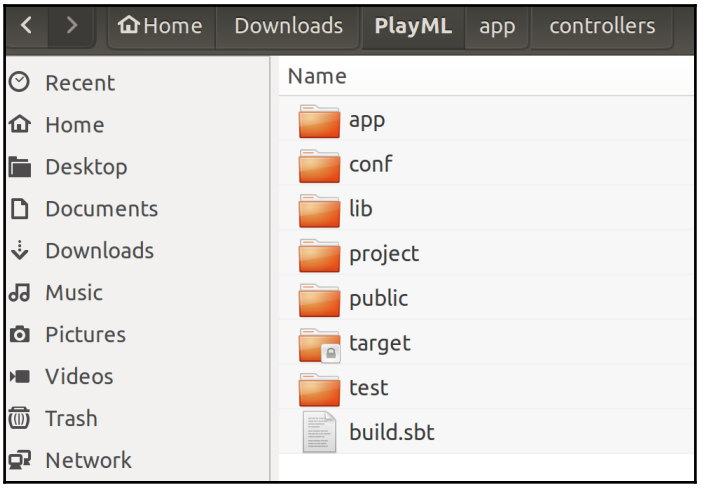


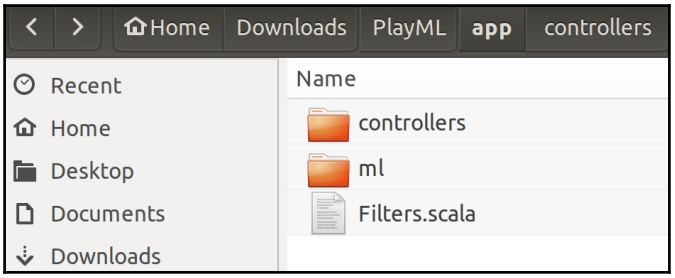

Q-values per state
675,000
650,000
625,000
600,000
575,000
550,000
525,000
500,000
475,000
450,000
425,000
400,000
375,000
350,000
325,000
300,000
275,000
250,000
225,000
200,000
175,000
150,000
125,000
100,000
75,000
50,000
25,000
0
States
Q-values

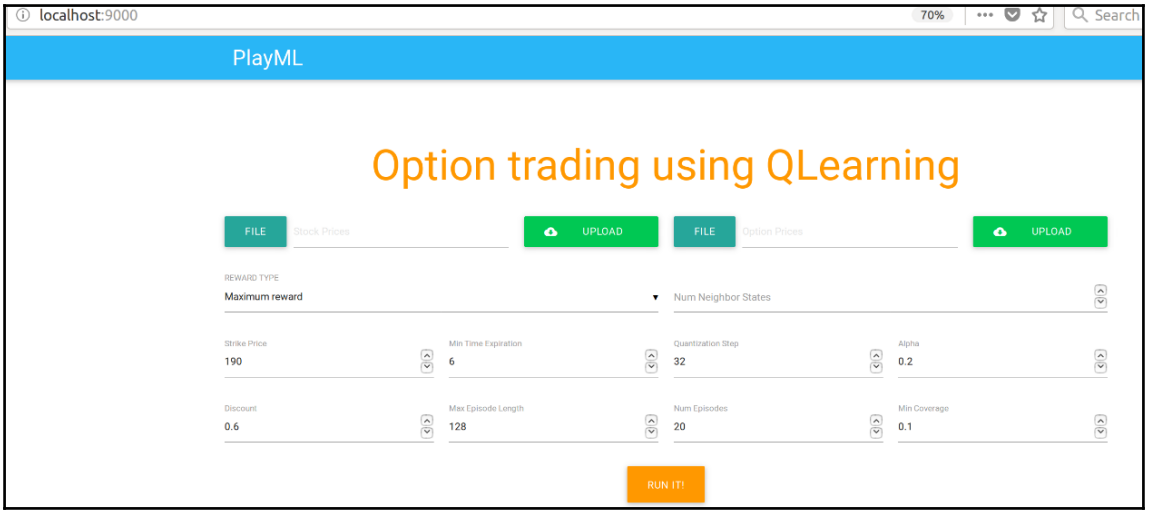

Log(q-value) for a=0.1, y=0.7
2
0
-2
-4
0
5
10
15
20
25
30
Iterations
Epoch 1
Epoch 6
Epoch 17

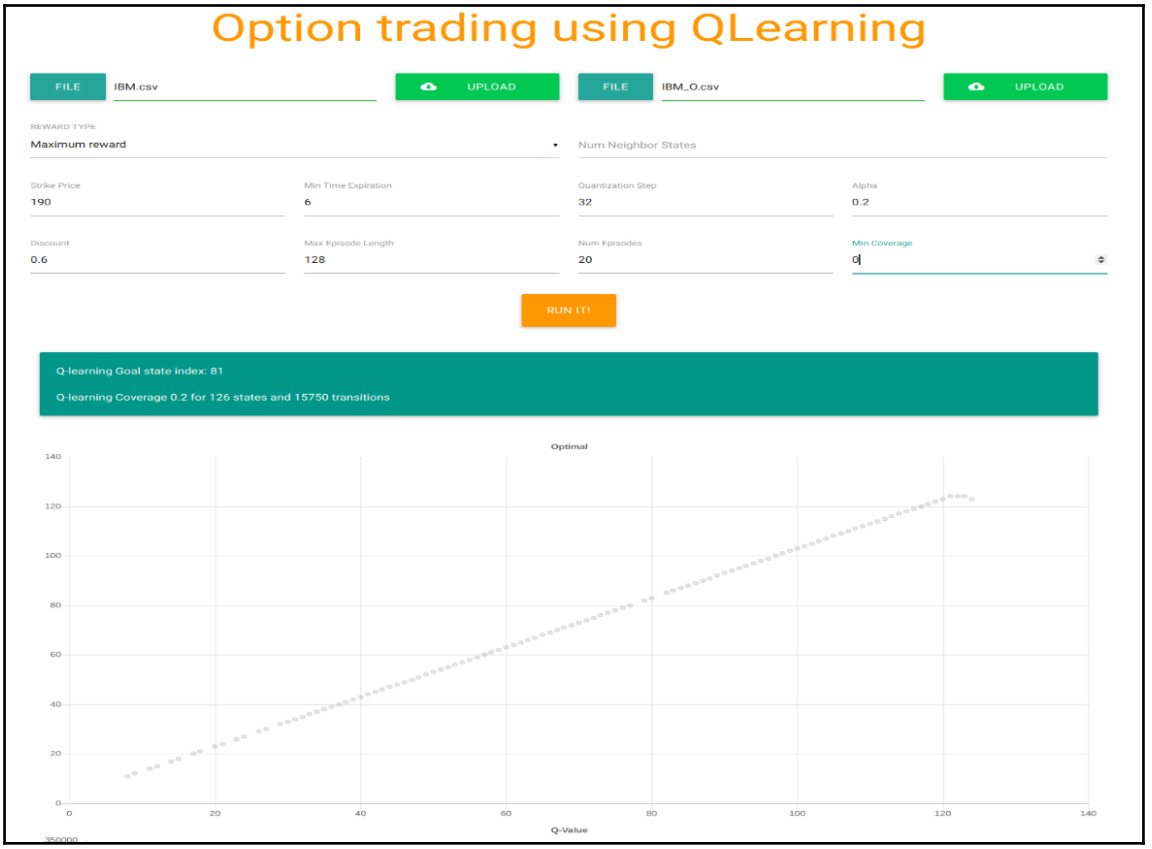

Q-learning config: Maximum reward with quantization order 32
States
States
Q-learning config: Maximum reward with quantization order 32


Coverage
Learning rate (a)
Coverage
Discount rate (y)


Home
Downloads
PlayML
app
controllers
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Name
app
conf
lib
project
public
target
test
build.sbt


Home
Downloads
PlayML
app
controllers
Recent
Home
Desktop
Documents
Downloads
Name
controllers
ml
Filters.scala

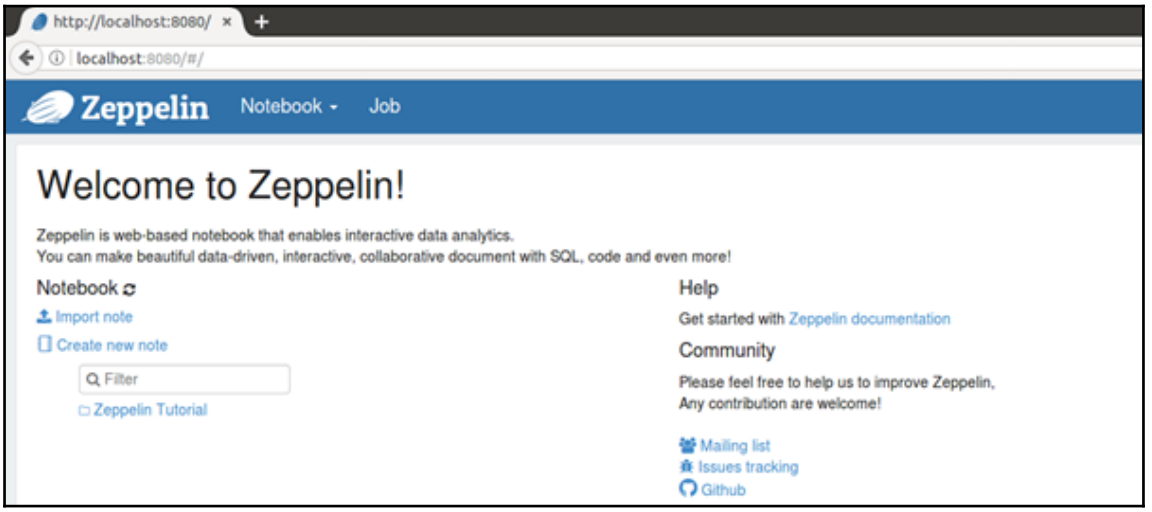

localhost:9000
70%
Search
PlayML
Option trading using QLearning
FILE
Stock Prices
UPLOAD
FILE
Option Prices
UPLOAD
REWARD TYPE
Maximum reward
Num Neighbor States
Strike Price
190
Min Time Expiration
6
Quantization Step
32
Alpha
0.2
Discount
0.6
Max Episode Length
128
Num Episodes
20
Min Coverage
0.1
RUN IT!

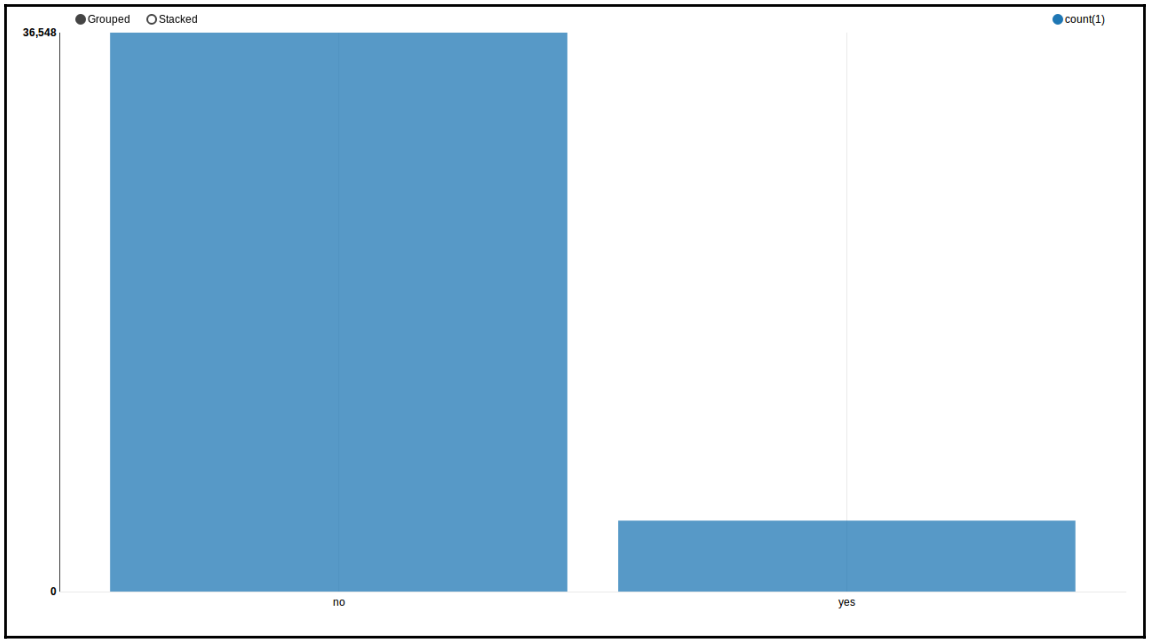

Option trading using QLearning
FILE
IBM.csv
UPLOAD
FILE
IBM_O.csv
UPLOAD
REWARD TYPE
Maximum reward
Num Neighbor States
Strike Price
190
Min Time Expiration
6
Quantization Step
32
Alpha
0.2
Discount
0.6
Max Episode Length
128
Num Episodes
20
Min Coverage
0
RUN IT!
Q-learning Goal state index: 81
Q-learning Coverage 0.2 for 126 states and 15750 transitions
Optimal
Q-Value

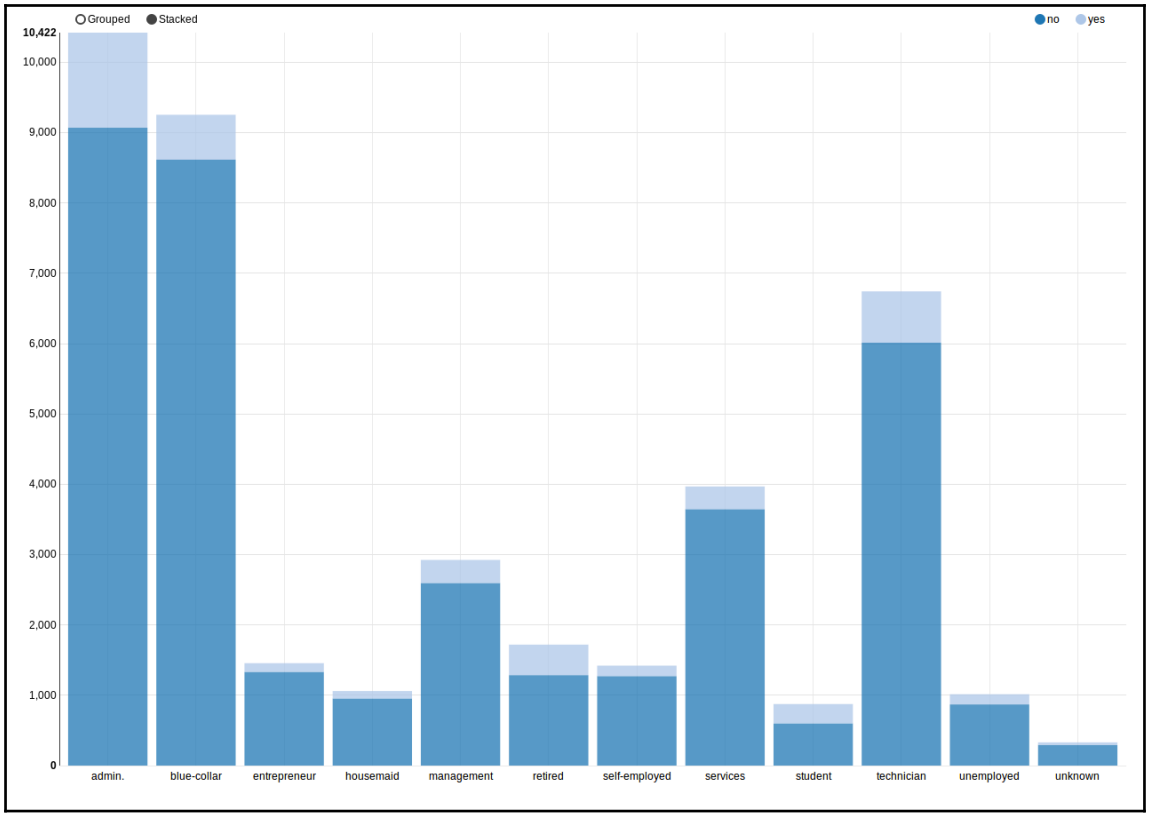
POST
http://localhost:9000/api/compute
Params
Send
Body
Cookies
Headers (3)
Tests
Status: 200 OK
Time: 1343 ms
Pretty
Raw
Preview
JSON
{
"STRIKE_PRICE": 333,
"MIN_TIME_EXPIRATION": 6,
"QUANTIZATION_STEP": 32,
"ALPHA": 0.2,
"DISCOUNT": 0.6,
"MAX_EPISODE_LEN": 128,
"NUM_EPISODES": 20,
"MIN_COVERAGE": 0.1,
"NUM_NEIGHBOR_STATES": 3,
"REWARD_TYPE": "Maximum reward",
"GOAL_STATE_INDEX": 113,
"COVERAGE": 0.2,
"COVERAGE_STATES": 126,
"COVERAGE_TRANSITIONS": 15750,
"Q_VALUE": [...],
"OPTIMAL": [...]
}

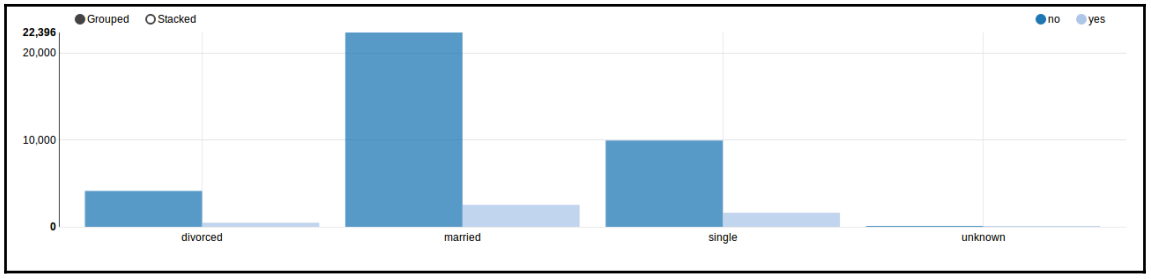
Chapter 8: Clients Subscription Assessment for Bank Telemarketing using Deep Neural Networks

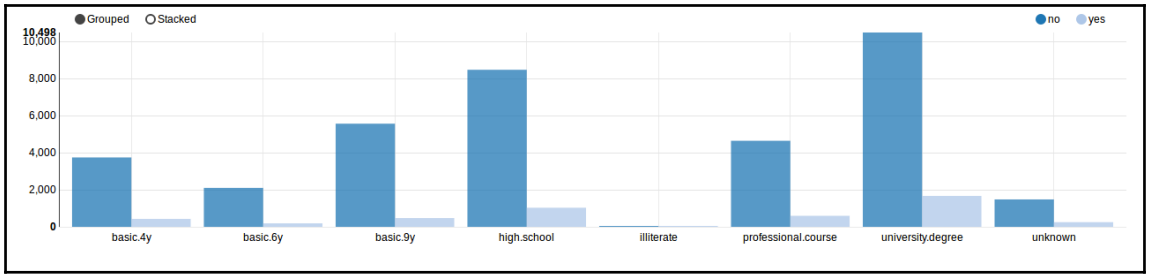
```
asif@ubuntu:~/Zeppelin$ bin/zeppelin-daemon.sh start
Log dir doesn't exist, create /home/asif/Zeppelin/logs
Pid dir doesn't exist, create /home/asif/Zeppelin/run
Zeppelin start                                             [  OK  ]
asif@ubuntu:~/Zeppelin$
```

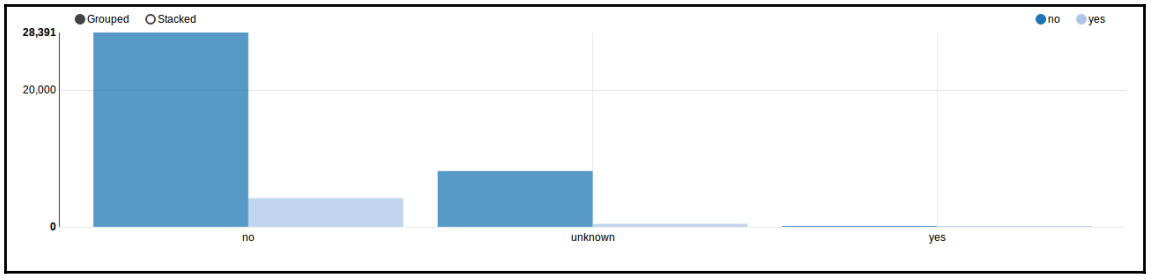


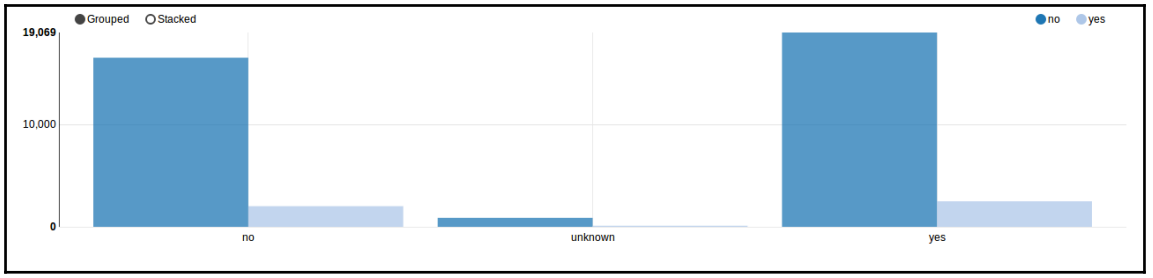

Grouped
Stacked
count(1)
36,548
0
no
yes

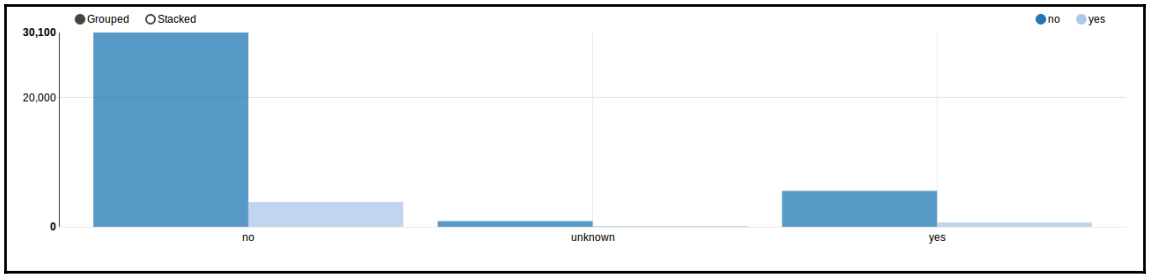

Grouped
Stacked
no
yes
10,422
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
admin.
blue-collar
entrepreneur
housemaid
management
retired
self-employed
services
student
technician
unemployed
unknown

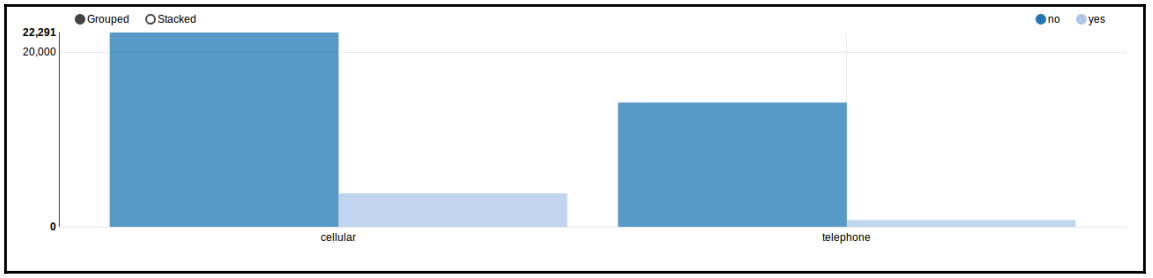

Grouped
Stacked
no
yes
22,396
20,000
10,000
0
divorced
married
single
unknown

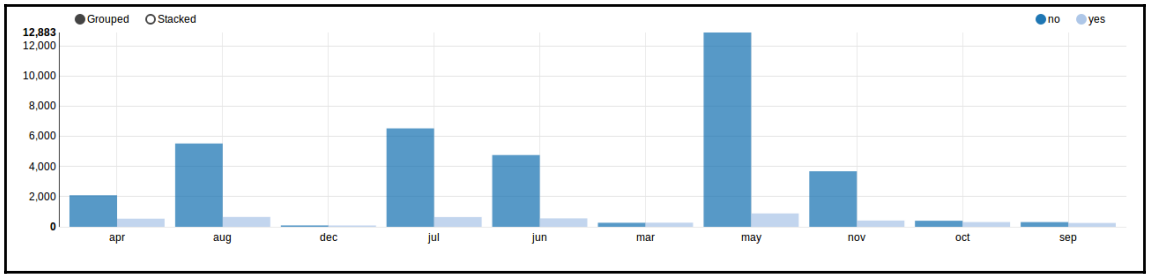

Grouped
Stacked
no
yes
10,498
10,000
8,000
6,000
4,000
2,000
0
basic.4y
basic.6y
basic.9y
high.school
illiterate
professional.course
university.degree
unknown

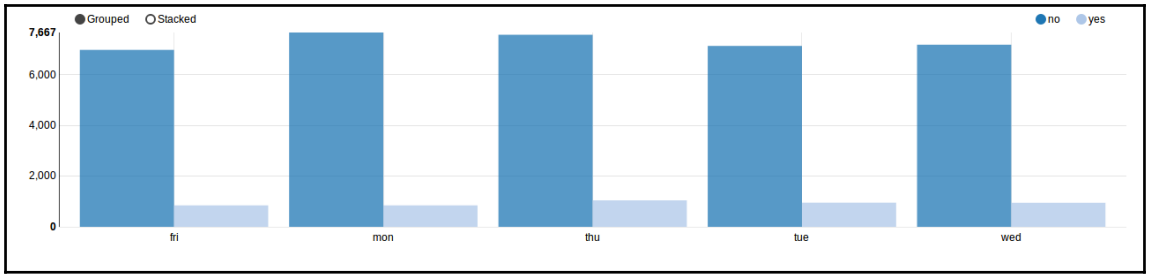

Grouped
Stacked
no
yes
28,391
20,000
0
no
unknown
yes

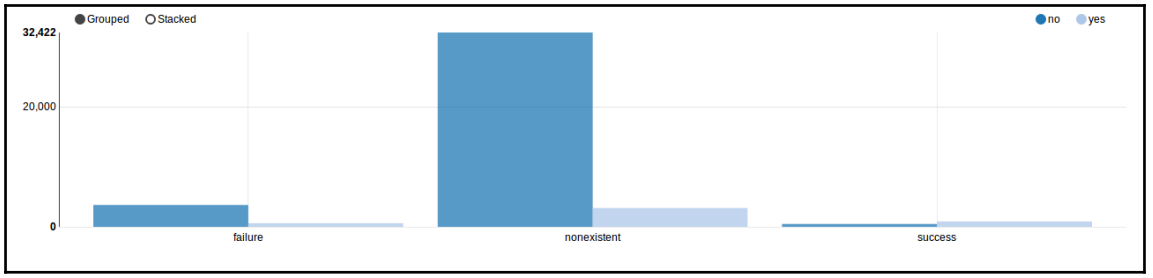

Grouped
Stacked
no
yes
19,069
10,000
0
no
unknown
yes

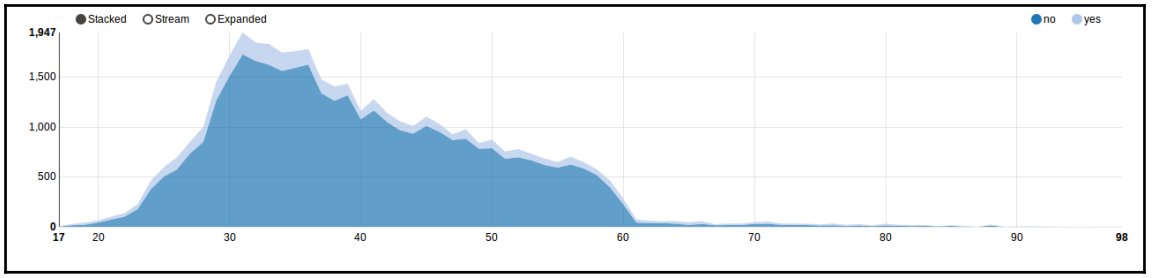

Grouped
Stacked
no
yes
30,100
20,000
0
no
unknown
yes

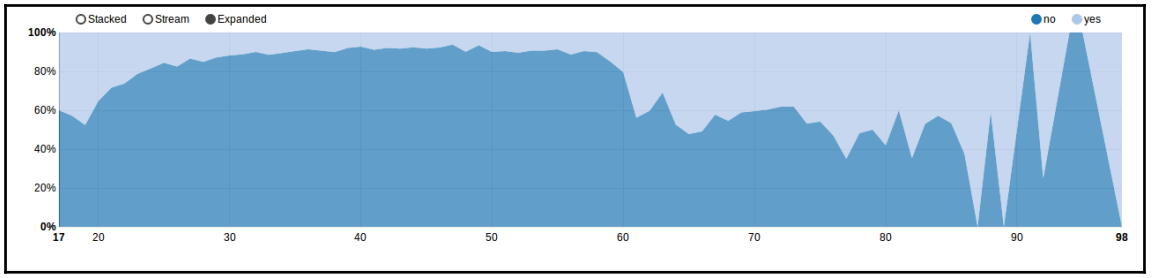

Grouped
Stacked
no
yes
22,291
20,000
0
cellular
telephone

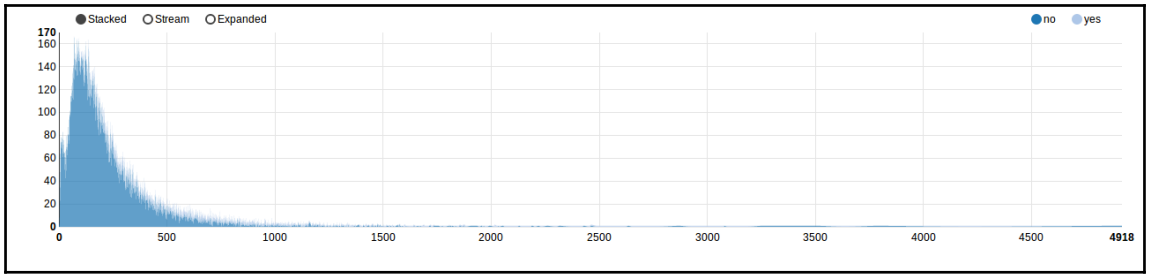

Grouped
Stacked
no
yes
12,883
12,000
10,000
8,000
6,000
4,000
2,000
0
apr
aug
dec
jul
jun
mar
may
nov
oct
sep

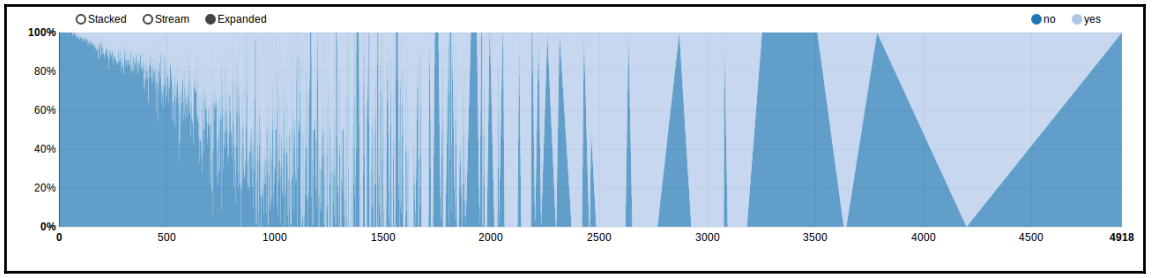

Grouped
Stacked
no
yes
7,667
6,000
4,000
2,000
0
fri
mon
thu
tue
wed

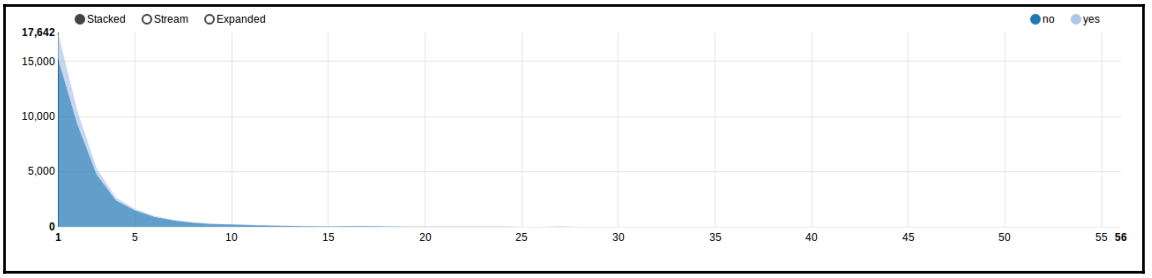

Grouped
Stacked
no
yes
32,422
20,000
0
failure
nonexistent
success

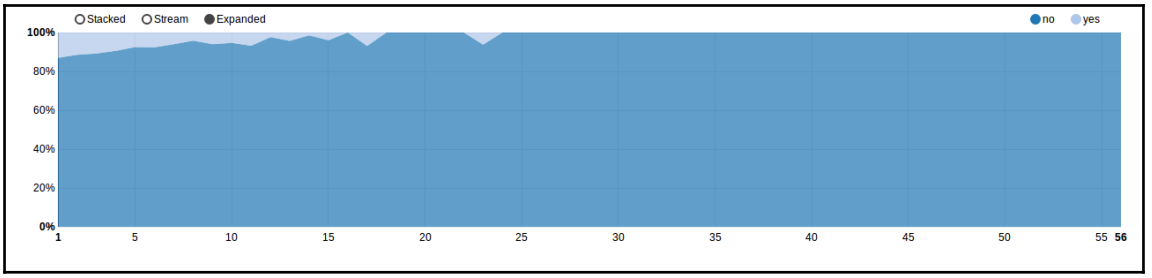

Stacked
Stream
Expanded
no
yes
1,947
1,500
1,000
500
0
17
20
30
40
50
60
70
80
90
98

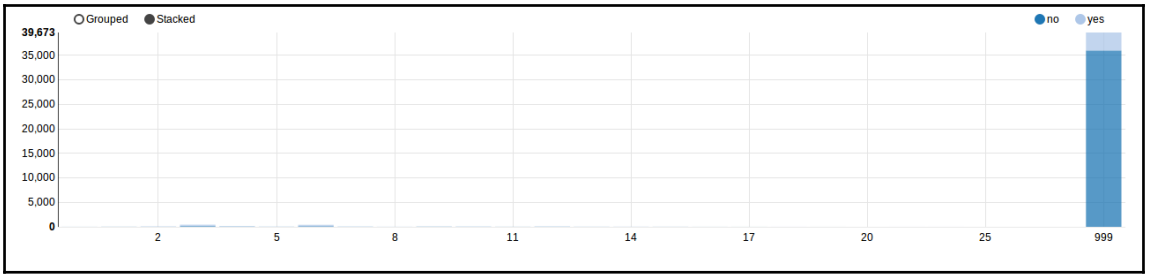

Stacked
Stream
Expanded
no
yes
100%
80%
60%
40%
20%
0%
17
20
30
40
50
60
70
80
90
98

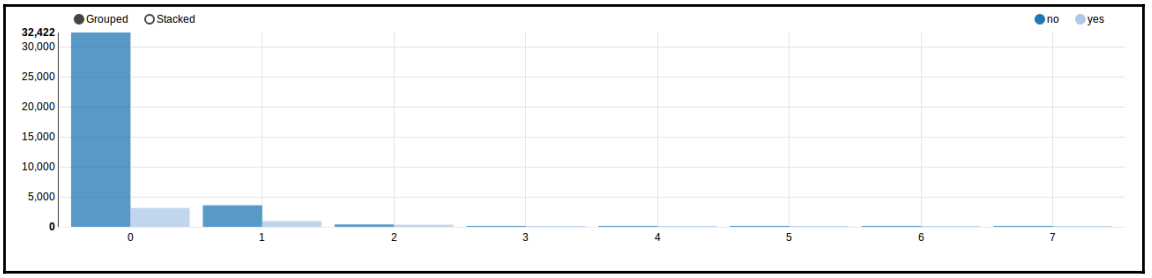

Stacked
Stream
Expanded
no
yes
170
160
140
120
100
80
60
40
20
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
4918

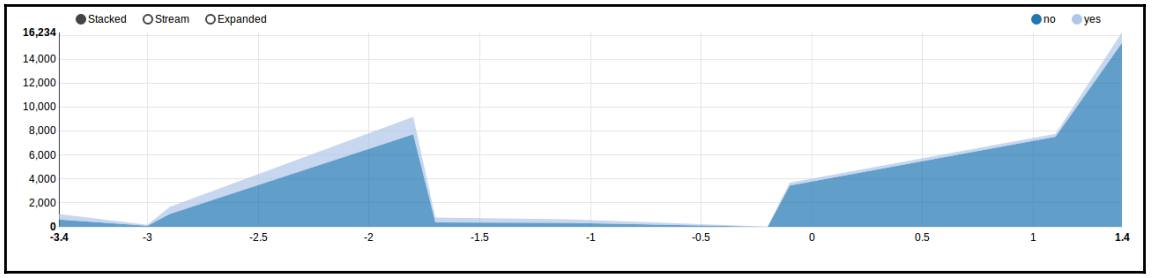

Stacked
Stream
Expanded
no
yes
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
4500
4918

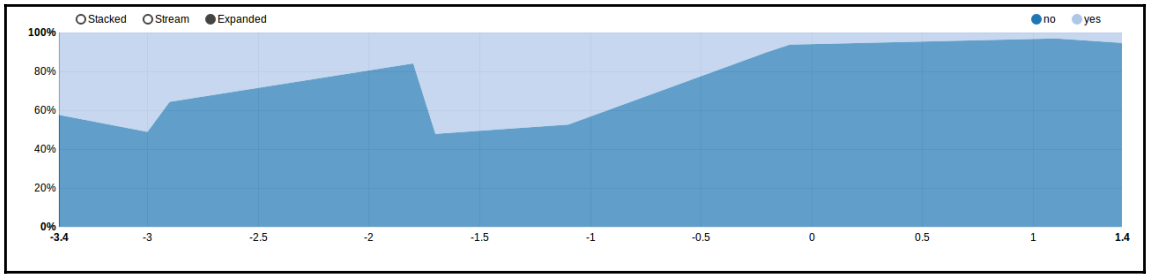

Stacked
Stream
Expanded
no
yes
17,642
15,000
10,000
5,000
0
1
5
10
15
20
25
30
35
40
45
50
55
56

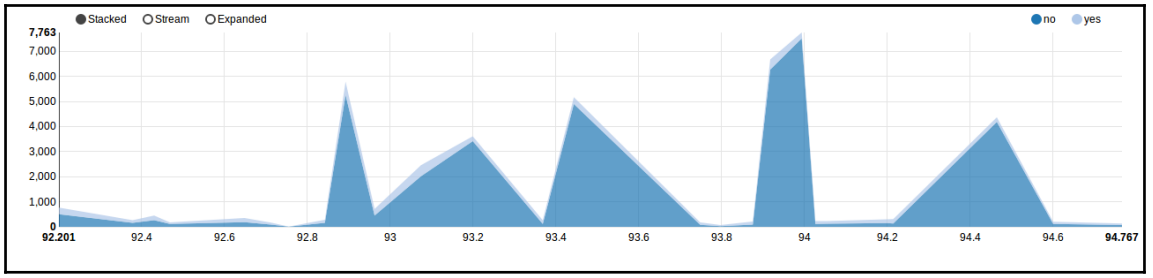

Stacked
Stream
Expanded
no
yes
100%
80%
60%
40%
20%
0%
1
5
10
15
20
25
30
35
40
45
50
55
56

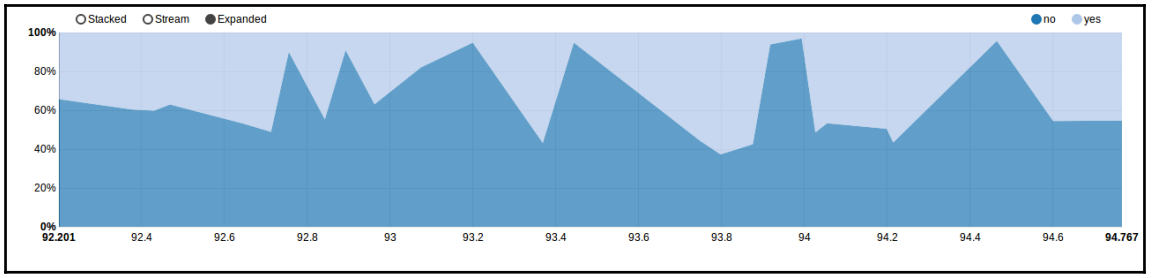

Grouped
Stacked
no
yes
39,673
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
2
5
8
11
14
17
20
25
999

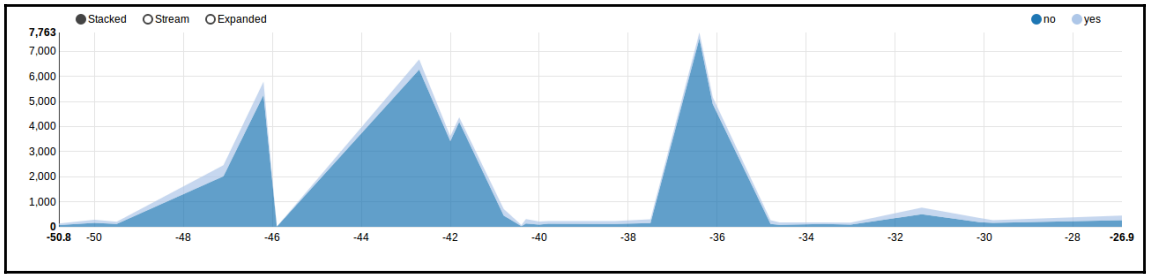

Grouped
Stacked
no
yes
32,422
30,000
25,000
20,000
15,000
10,000
5,000
0
0
1
2
3
4
5
6
7

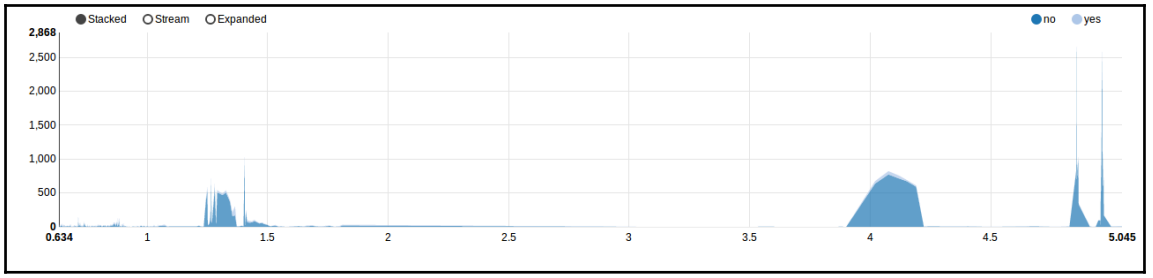

Stacked
Stream
Expanded
no
yes
16,234
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
-3.4
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.4

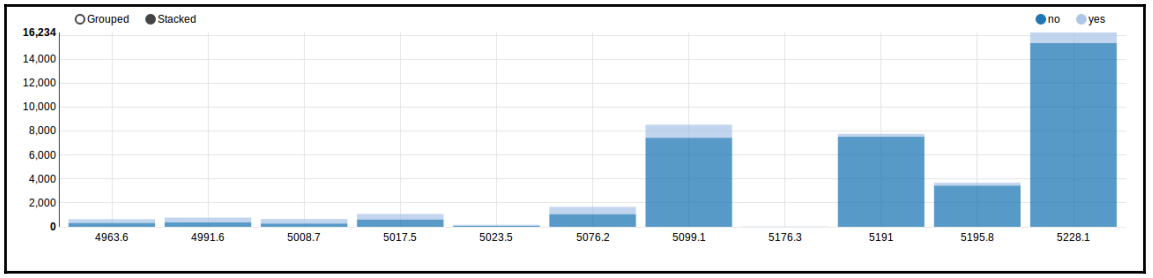

Stacked
Stream
Expanded
no
yes
100%
80%
60%
40%
20%
0%
-3.4
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.4


Stacked
Stream
Expanded
no
yes
7,763
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
92.201
92.4
92.6
92.8
93
93.2
93.4
93.6
93.8
94
94.2
94.4
94.6
94.767

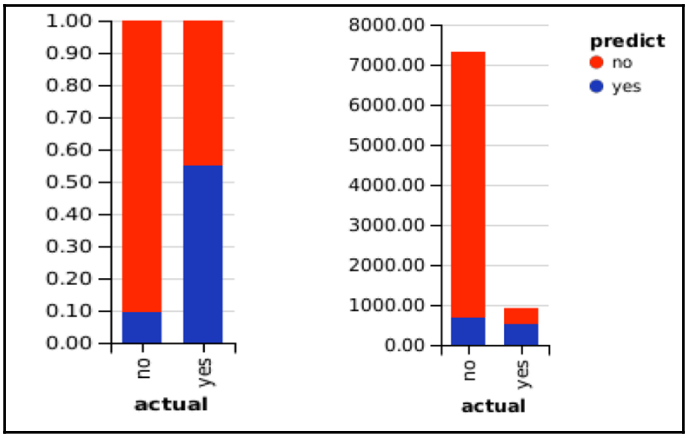

Stacked
Stream
Expanded
no
yes
100%
80%
60%
40%
20%
0%
92.201
92.4
92.6
92.8
93
93.2
93.4
93.6
93.8
94
94.2
94.4
94.6
94.767



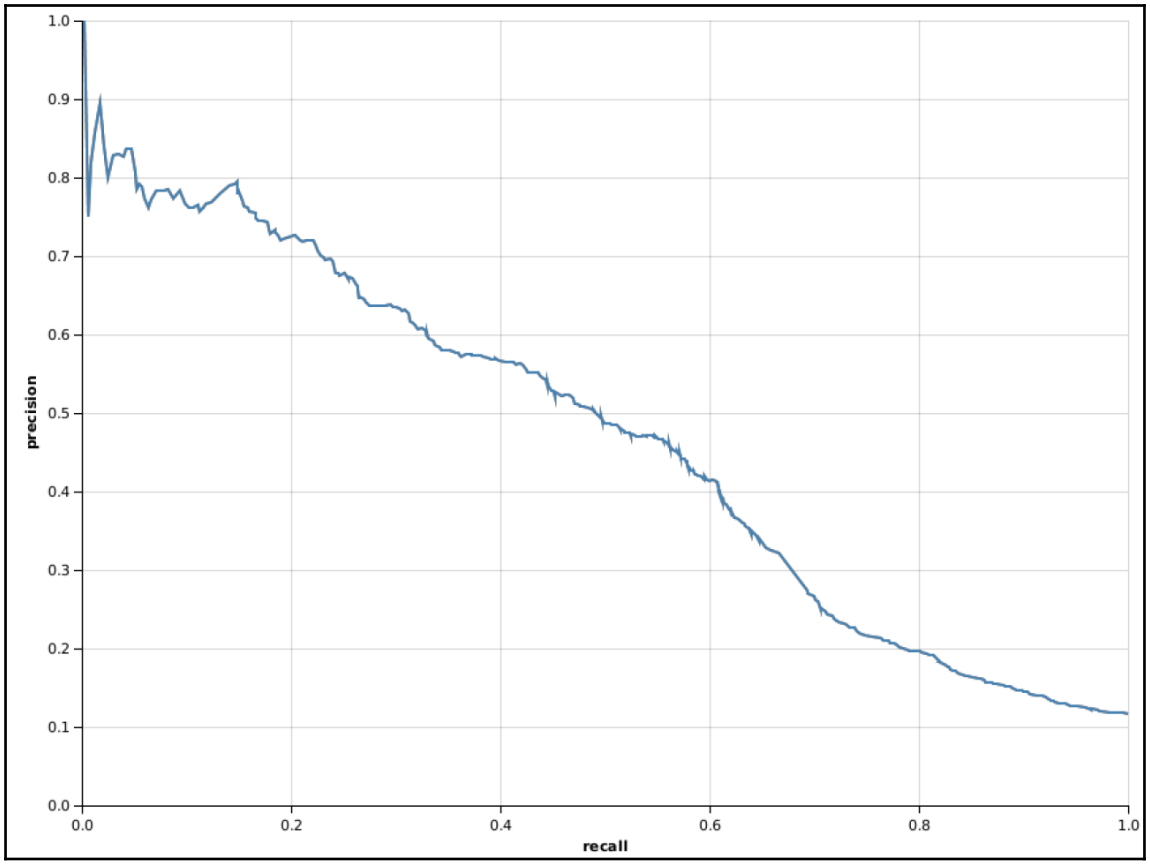


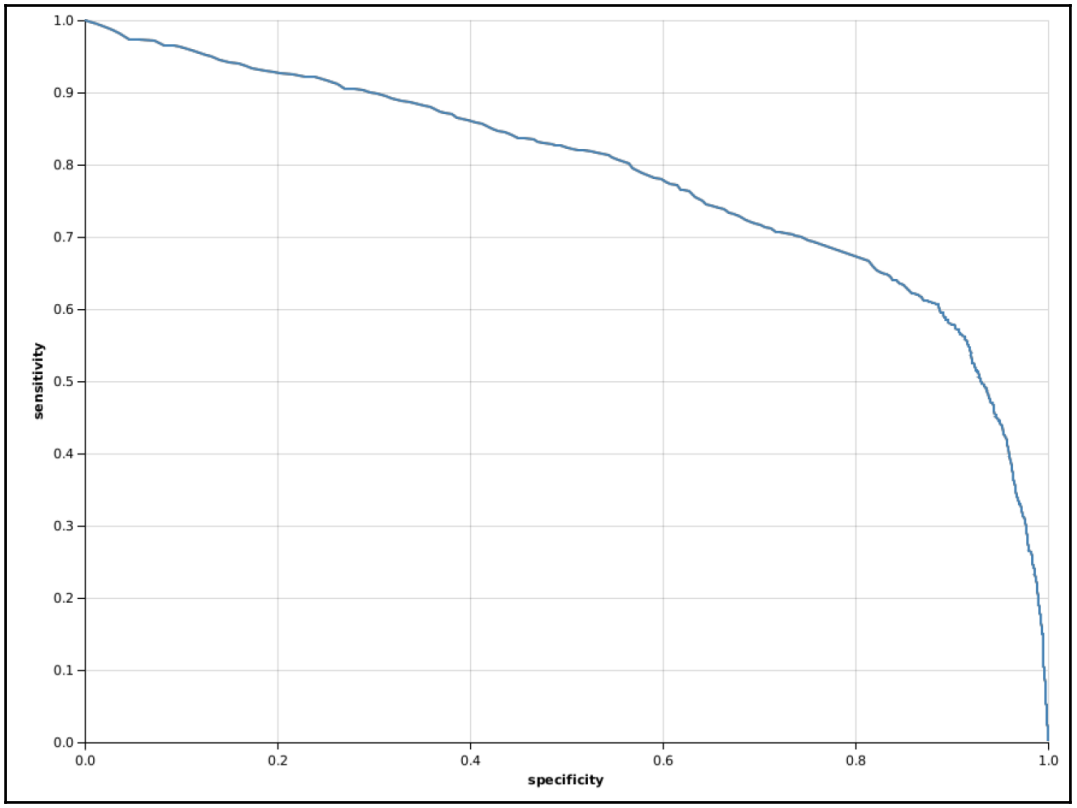


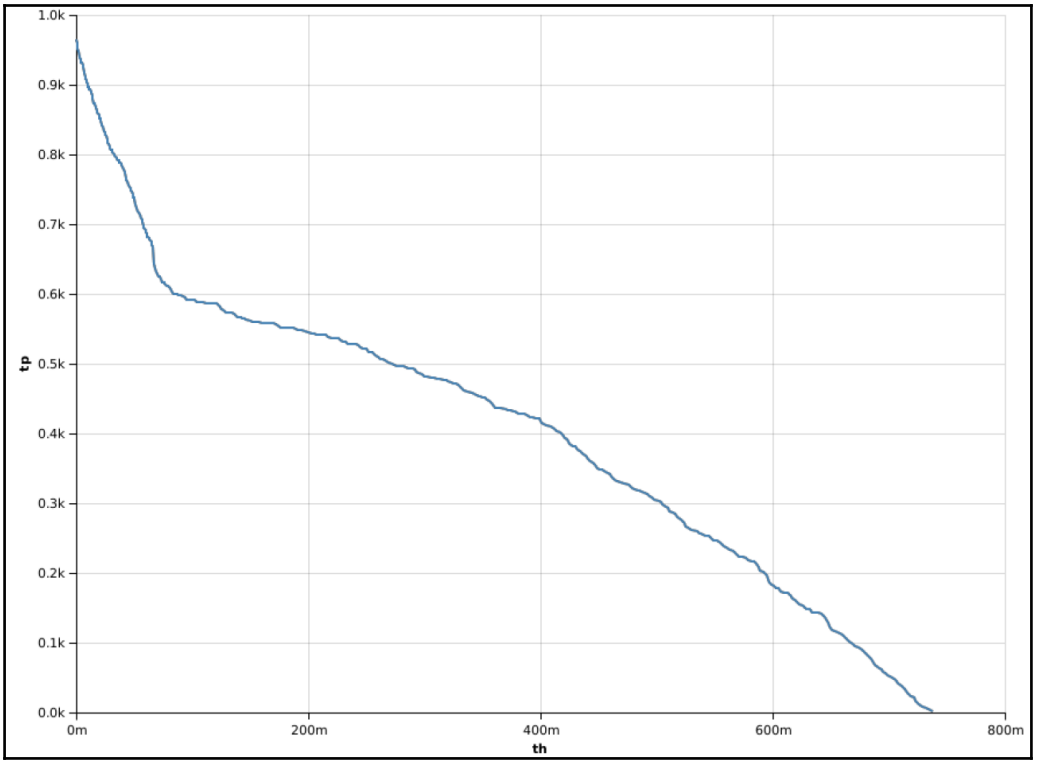

```
+------+-------+-----+
|actual|predict|count|
+------+-------+-----+
|    no|     no| 6654|
|    no|    yes|  678|
|   yes|    yes|  509|
|   yes|     no|  398|
+------+-------+-----+
```

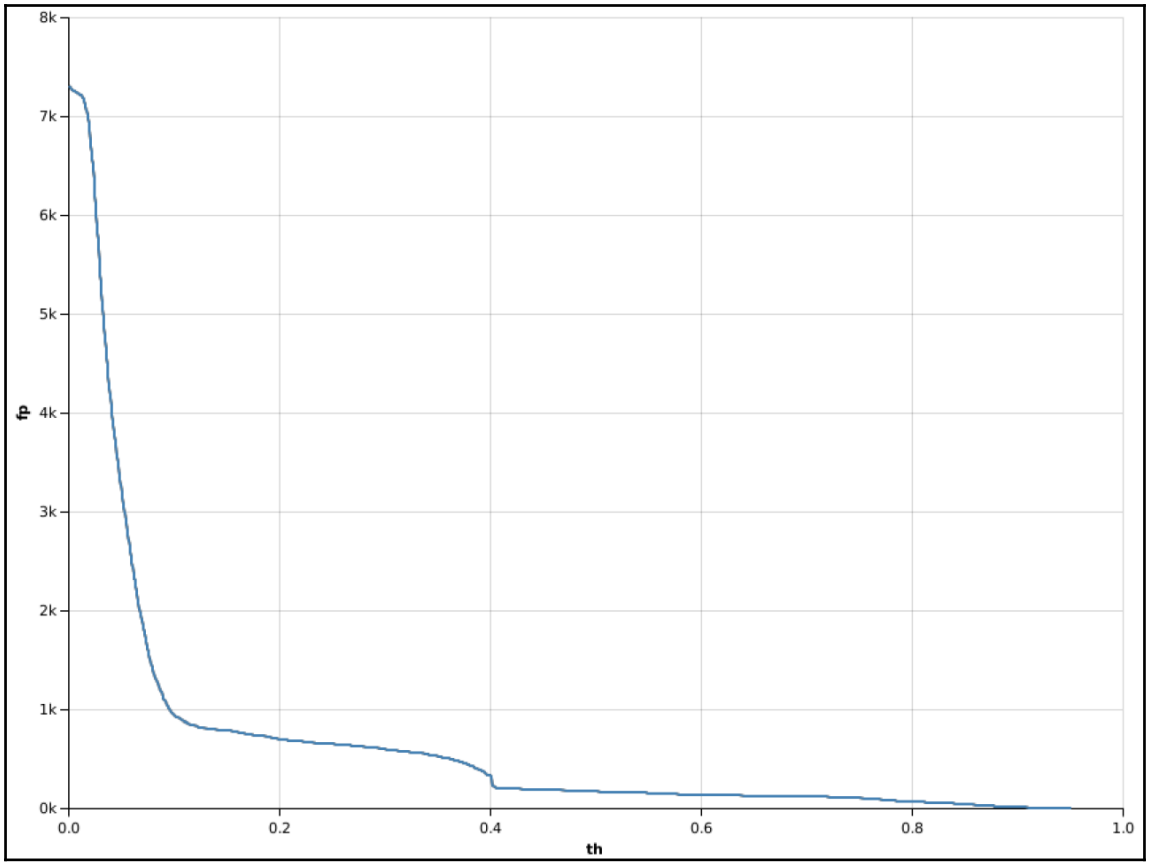


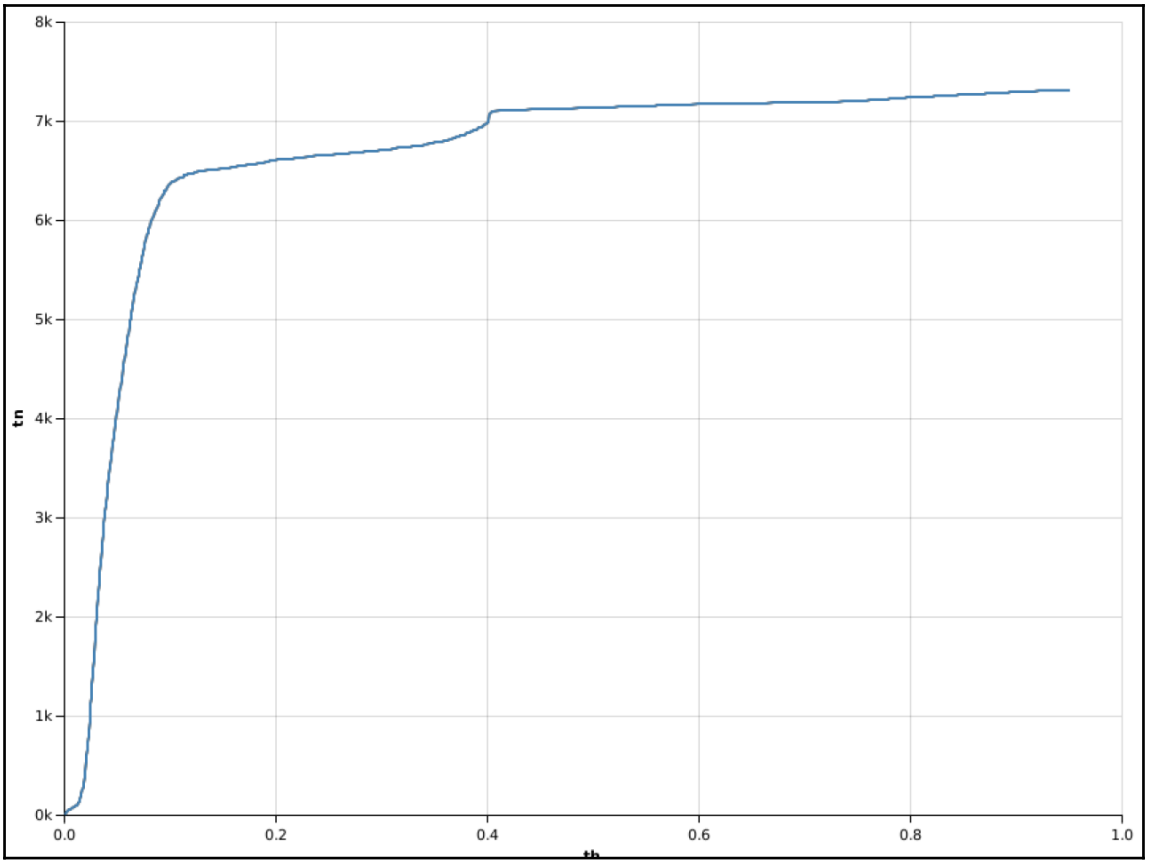

```
Model Metrics Type: Binomial
 Description: N/A
 model id: dlModel.hex
 frame id: test.hex
 MSE: 0.08054043
 RMSE: 0.28379646
 AUC: 0.7632699
 logloss: 0.28570446
 mean_per_class_error: 0.27243903
 default threshold: 0.22040662169456482
 CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class):
          no   yes   Error           Rate
    no  6621   669  0.0918    669 / 7,290
   yes   430   519  0.4531      430 / 949
Totals  7051  1188  0.1334  1,099 / 8,239
```

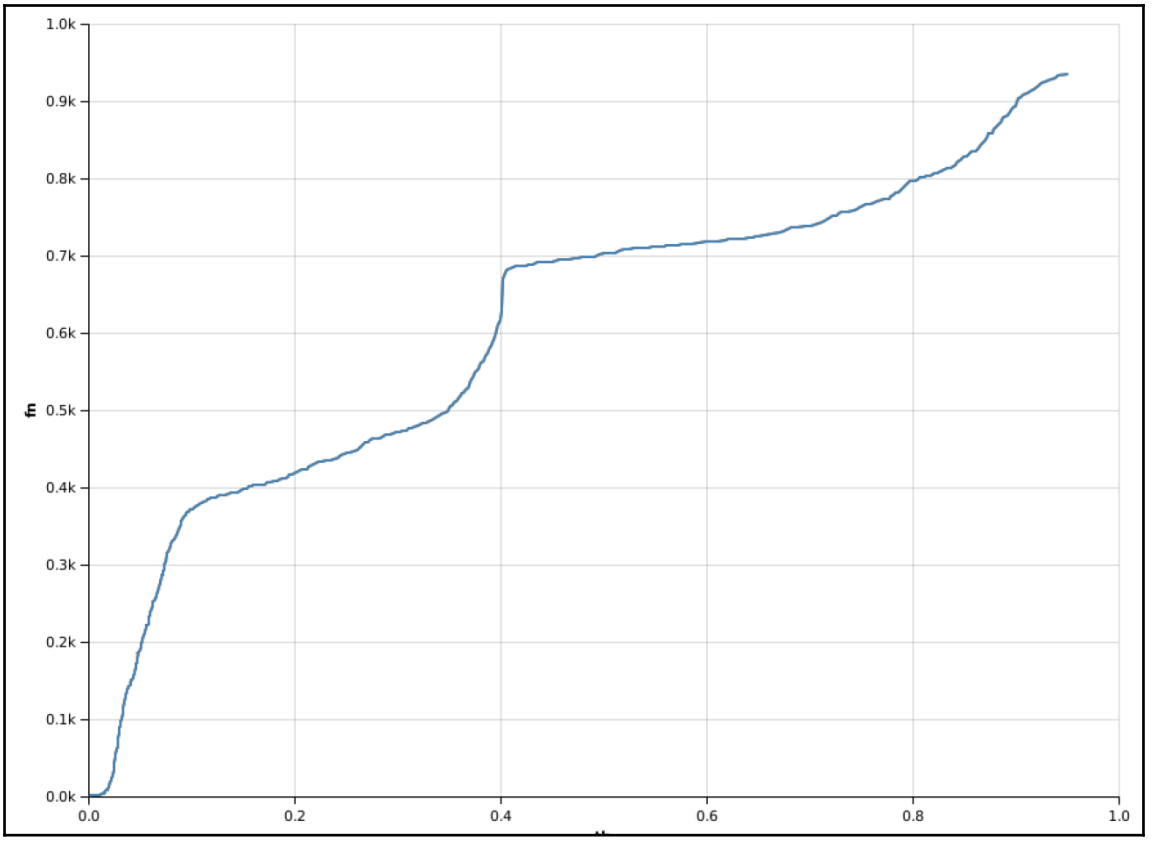

precision
recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
specificity


1.0k
0.9k
0.8k
0.7k
0.6k
0.5k
0.4k
0.3k
0.2k
0.1k
0.0k
tp
0m
200m
400m
600m
800m
th

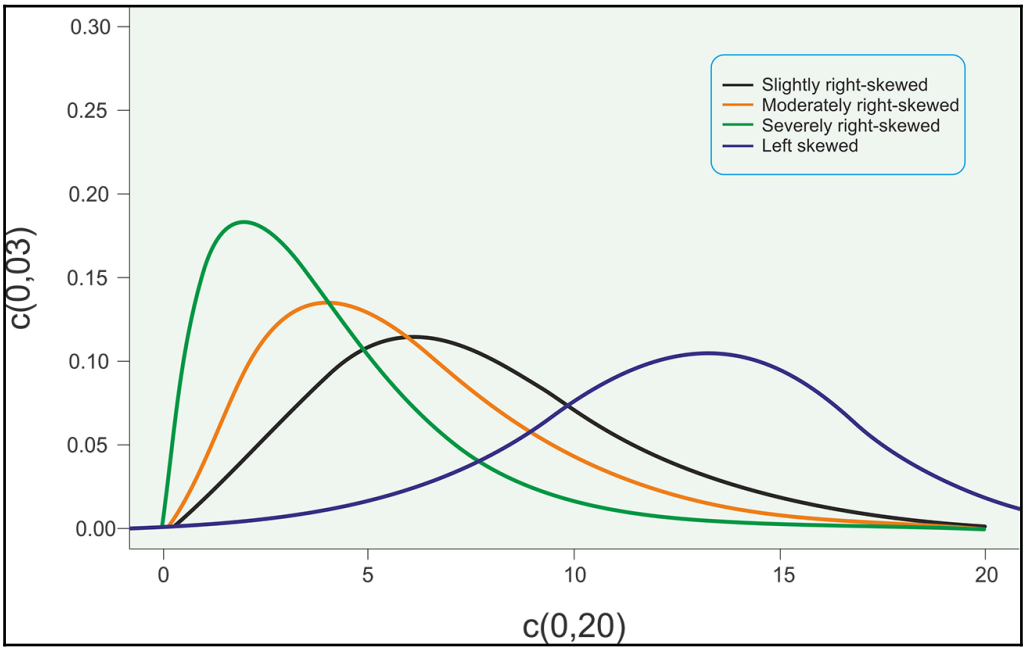

8k
7k
6k
5k
4k
3k
2k
1k
0k
fp
0.0
0.2
0.4
0.6
0.8
1.0
th

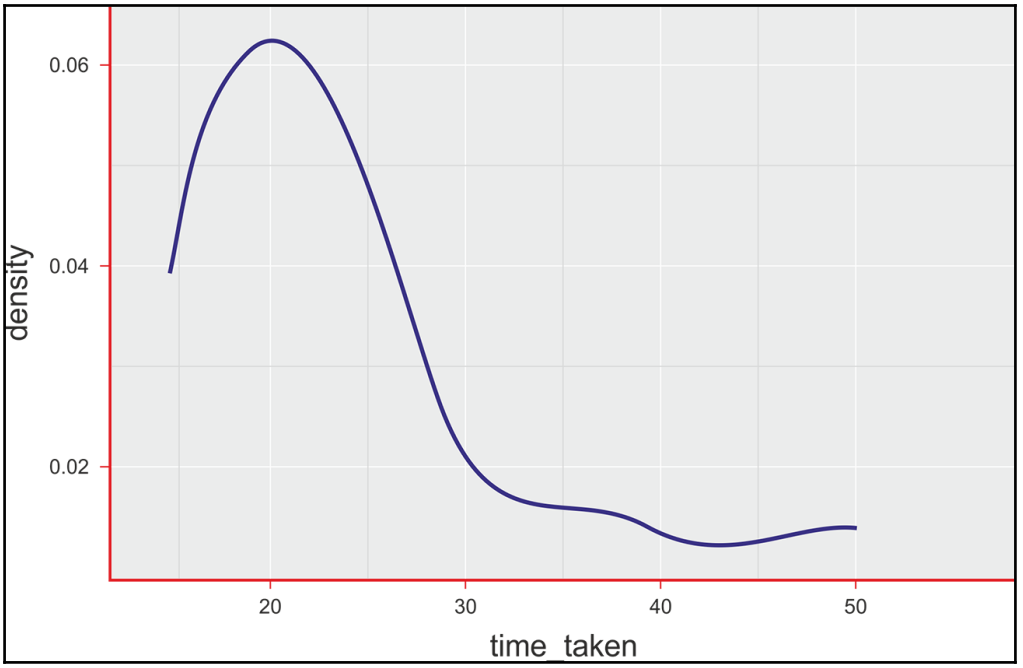

8k
7k
6k
5k
4k
3k
2k
1k
0k
tn
0.0
0.2
0.4
0.6
0.8
1.0

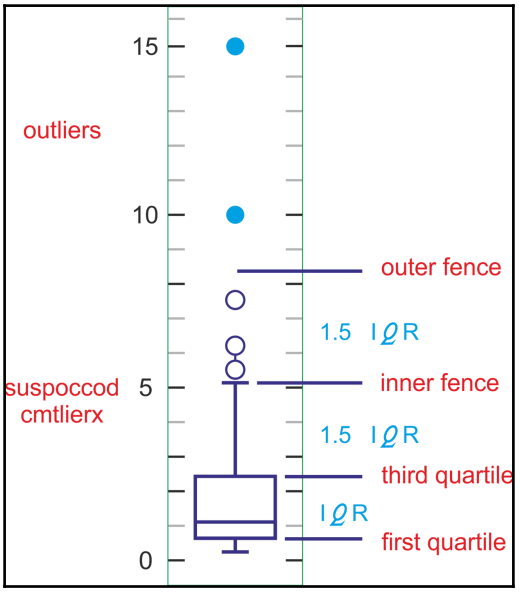

1.0k
0.9k
0.8k
0.7k
0.6k
0.5k
0.4k
0.3k
0.2k
0.1k
0.0k
fn
0.0
0.2
0.4
0.6
0.8
1.0

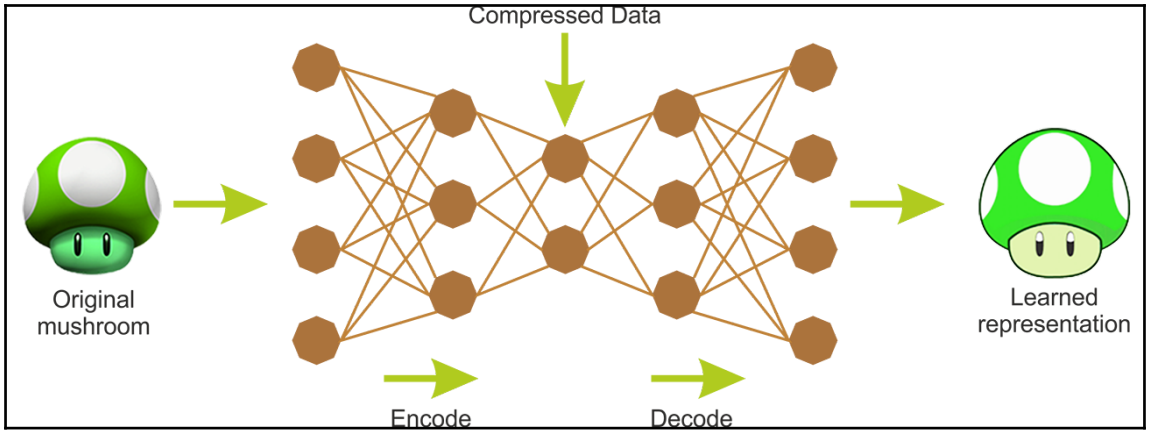
```
INFO: Variable Importances:
INFO:                 Variable Relative Importance Scaled Importance Percentage
INFO:                month.mar            1.000000          1.000000   0.057811
INFO:                month.dec            0.742578          0.742578   0.042930
INFO:    education.illiterate            0.727627          0.727627   0.042065
INFO:             default.yes            0.711488          0.711488   0.041132
INFO:         marital.unknown            0.681409          0.681409   0.039393
INFO:                month.sep            0.642754          0.642754   0.037159
INFO:              job.unknown            0.637595          0.637595   0.036860
INFO:                month.oct            0.619677          0.619677   0.035824
INFO:              job.student            0.536696          0.536696   0.031027
INFO:           job.housemaid             0.520993          0.520993   0.030119
INFO: ---
INFO:          job.missing(NA)            0.000000          0.000000   0.000000
INFO:        month.missing(NA)            0.000000          0.000000   0.000000
INFO:    education.missing(NA)            0.000000          0.000000   0.000000
INFO: day_of_week.missing(NA)             0.000000          0.000000   0.000000
INFO:      marital.missing(NA)            0.000000          0.000000   0.000000
INFO:         loan.missing(NA)            0.000000          0.000000   0.000000
INFO:      default.missing(NA)            0.000000          0.000000   0.000000
INFO:      housing.missing(NA)            0.000000          0.000000   0.000000
INFO:     poutcome.missing(NA)            0.000000          0.000000   0.000000
INFO:      contact.missing(NA)            0.000000          0.000000   0.000000
```

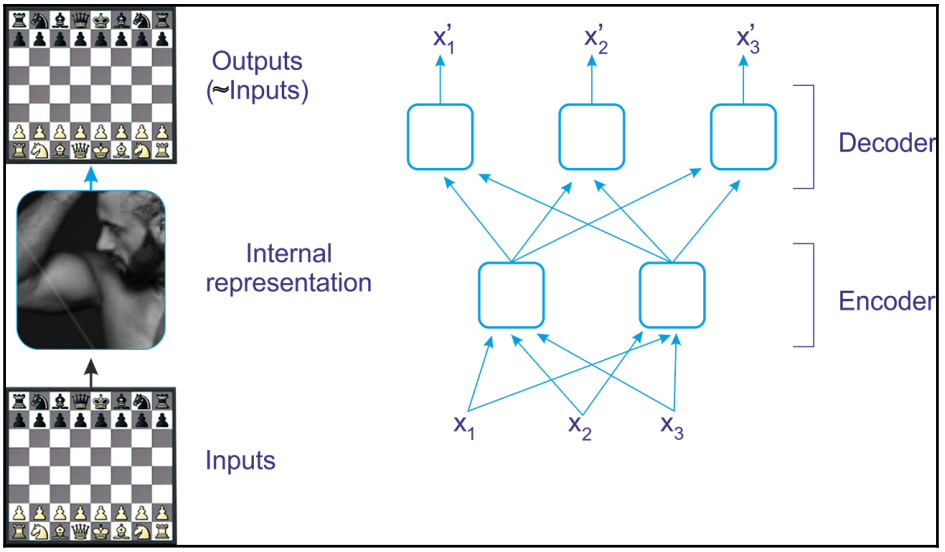
Chapter 9: Fraud Analytics Using Autoencoders and Anomaly Detection




0.06
0.04
0.02
density
20
30
40
50
time_taken

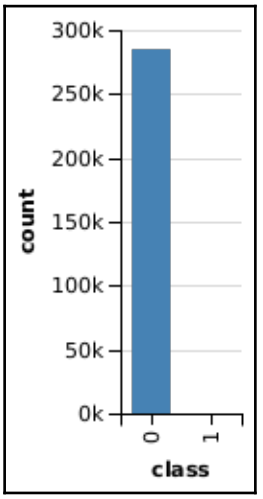

15
10
5
0
outliers
suspoccod
cmtlierx
outer fence
1.5 IQR
inner fence
1.5 IQR
third quartile
IQR
first quartile

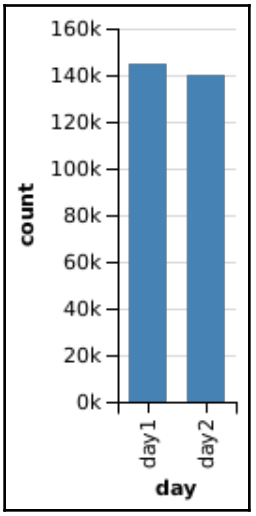

Compressed Data
Original mushroom
Learned representation
Encode
Decode

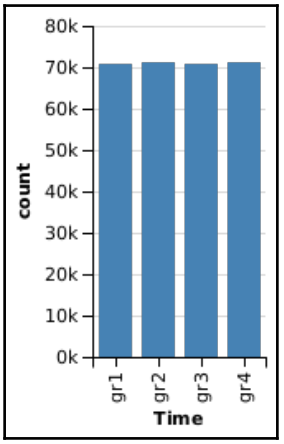

Outputs (≈Inputs)
Internal representation
Inputs
x'_1
x'_2
x'_3
x_1
x_2
x_3
Decoder
Encoder

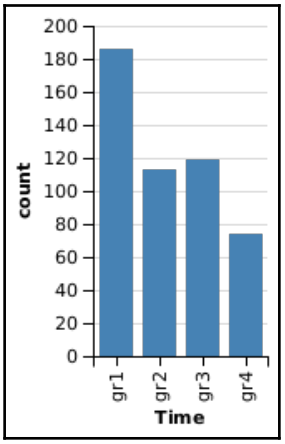
Time	V1	v2	V26	V27	Amount	Class
0	-1.3598071336738	-0.0727811733098497	-0.189114843888824	0.133558376740387	149.62	0
0	1.19185711131486	0.26615071205963	0.125894532368176	-0.00898309914322813	2.69	0
1	-1.35835406159823	-1.34016307473609	-0.139096571514147	-0.0553527940384261	378.66	0
1	-0.966271711572087	-0.185226008082898	-0.221928844458407	0.0627228487293033	123.5	0
2	-1.15823309349523	0.877736754848451	0.502292224181569	0.219422229513348	69.99	0
2	-0.425965884412454	0.960523044882985	0.105914779097957	0.253844224739337	3.67	0
4	1.22965763450793	0.141003507049326	-0.257236845917139	0.0345074297438413	4.99	0
7	-0.644269442348146	1.41796354547385	-0.0516342969262494	-1.20692108094258	40.8	0
7	-0.89428608220282	0.286157196276544	-0.384157307702294	0.0117473564581996	93.2	0
9	-0.33826175242575	1.11959337641566	0.0941988339514961	0.246219304619926	3.68	0

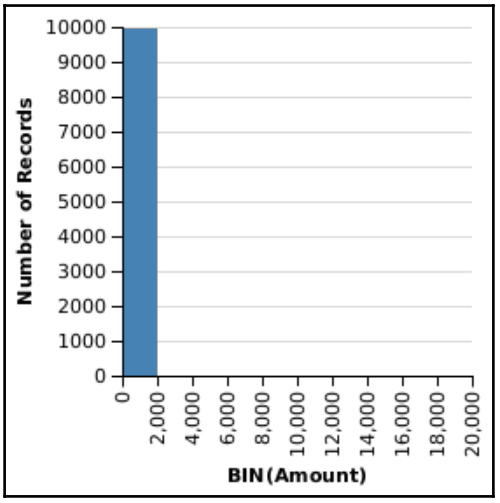
only showing top 10 rows

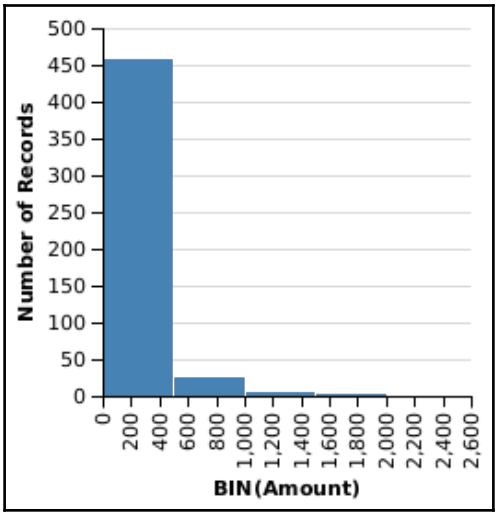



160k
140k
120k
100k
80k
60k
40k
20k
0k
count
day1
day2
day

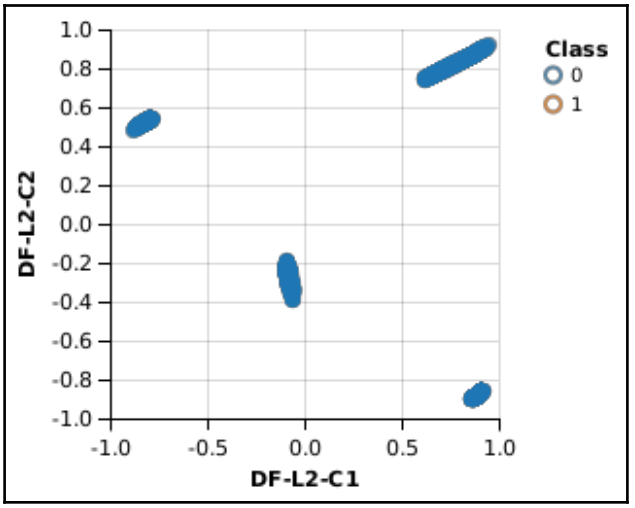

80k
70k
60k
50k
40k
30k
20k
10k
0k
count
gr1
gr2
gr3
gr4
Time


200
180
160
140
120
100
80
60
40
20
0
count
gr1
gr2
gr3
gr4
Time


10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Number of Records
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
BIN(Amount)

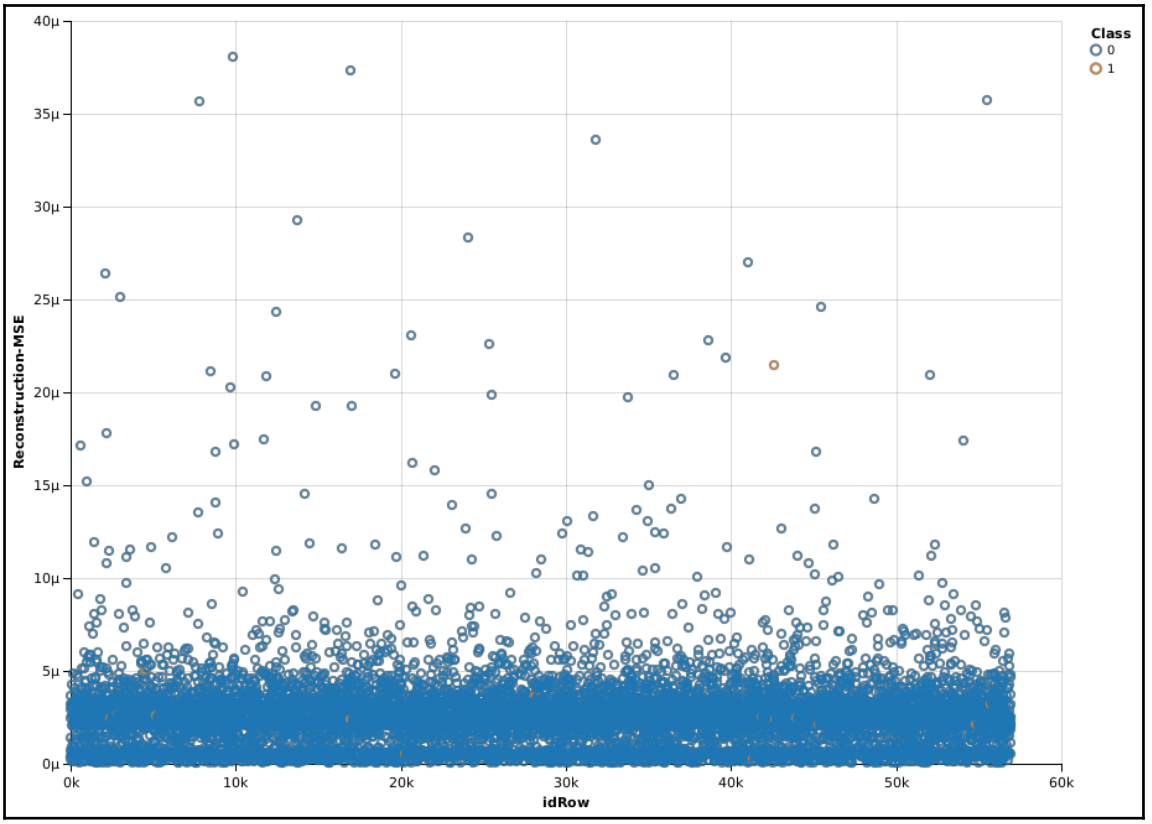


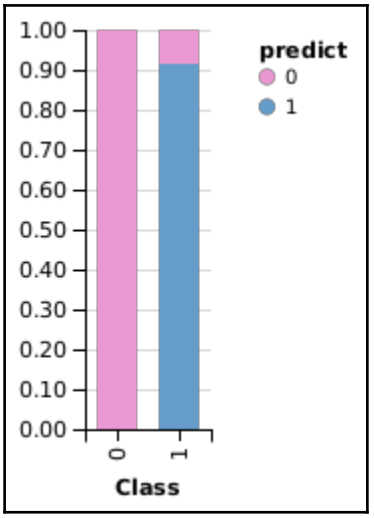

```
Model Metrics Type: AutoEncoder
 Description: N/A
 model id: DeepLearning_model_1510425525966_1
 frame id: train_unsupervised.temporary.sample.8.78%
 MSE: 0.0016747947
 RMSE: 0.040924255
Variable Importances:
       Variable Relative Importance Scaled Importance Percentage
       Time.gr3            1.000000          1.000000   0.266028
       Time.gr1            0.834574          0.834574   0.222020
       Time.gr2            0.458785          0.458785   0.122050
            V13            0.380227          0.380227   0.101151
            V12            0.146439          0.146439   0.038957
       Time.gr4            0.098328          0.098328   0.026158
            V14            0.095269          0.095269   0.025344
             V9            0.068873          0.068873   0.018322
            V17            0.062106          0.062106   0.016522
            V11            0.051645          0.051645   0.013739
             V3            0.016222          0.016222   0.004316
            V22            0.016090          0.016090   0.004280
             V6            0.015982          0.015982   0.004252
            V19            0.015074          0.015074   0.004010
            V28            0.014994          0.014994   0.003989
            V18            0.014616          0.014616   0.003888
            V25            0.010810          0.010810   0.002876
             V1            0.010639          0.010639   0.002830
            V16            0.007121          0.007121   0.001894
Time.missing(NA)           0.000000          0.000000   0.000000
Status of Neuron Layers (auto-encoder, gaussian distribution, Quadratic loss, 776 weights/biases, 16.2 KB, 2,392,362
 Layer Units  Type Dropout       L1       L2 Mean Rate Rate RMS Momentum Mean Weight Weight RMS Mean Bias Bias RMS
     1    34 Input  0.00 %
     2    10  Tanh       0 0.000000 0.000000  0.549217 0.327317 0.000000   -0.040651   0.613107 -0.123076 0.560158
     3     2  Tanh       0 0.000000 0.000000  0.031482 0.060843 0.000000   -0.042979   0.266792  0.124719 0.119967
     4    10  Tanh       0 0.000000 0.000000  0.111525 0.151304 0.000000    0.287284   1.108348  0.023247 0.515582
     5    34  Tanh         0.000000 0.000000  0.252175 0.361352 0.000000    0.004365   0.602132  0.197952 0.758697
Scoring History:
             Timestamp   Duration  Training Speed   Epochs Iterations        Samples Training RMSE Training MSE
 2017-11-11 19:39:19  0.719 sec 0.00000 obs/sec  0.00000          0       0.000000       0.18766      0.03522
 2017-11-11 19:39:21  3.465 sec   42350 obs/sec  1.00000          1  113922.000000       0.04617      0.00213
 ...
 2017-11-11 19:40:16 57.972 sec   42616 obs/sec 21.00000         21 2392362.000000       0.04092      0.00167
```

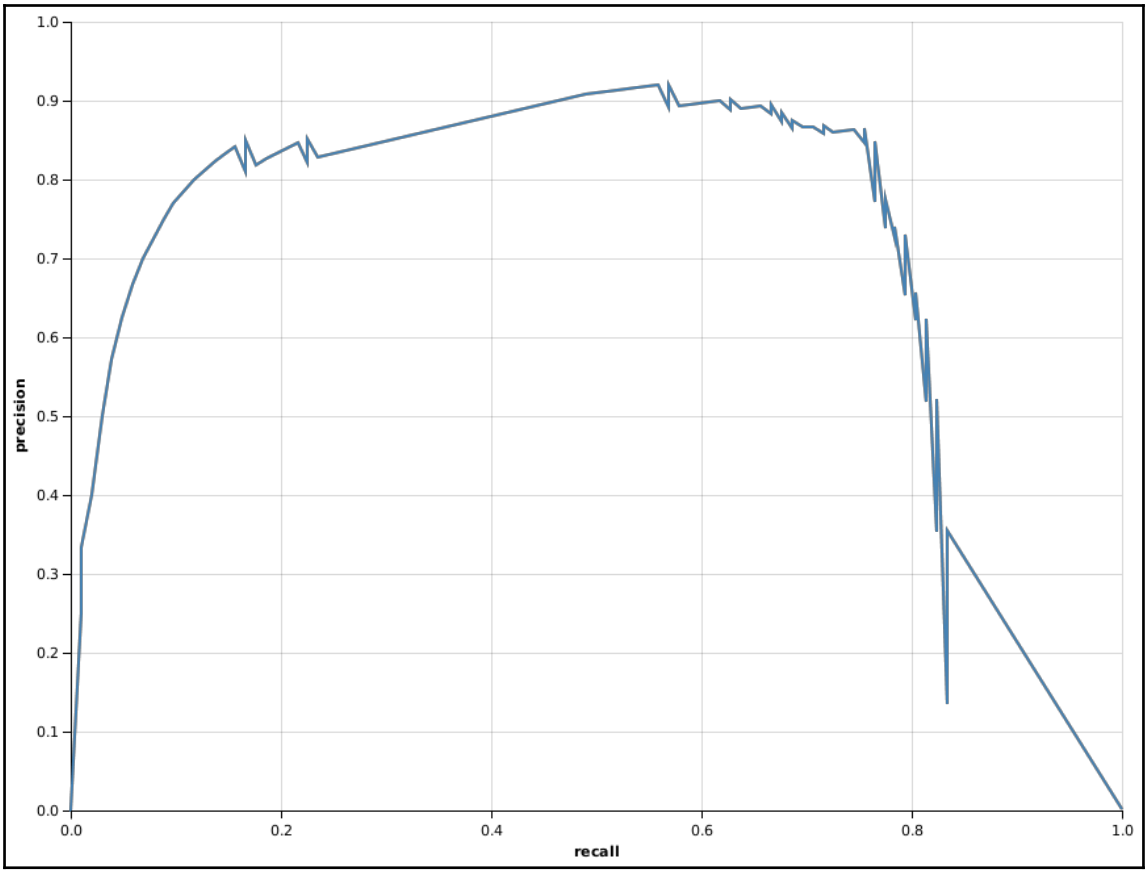


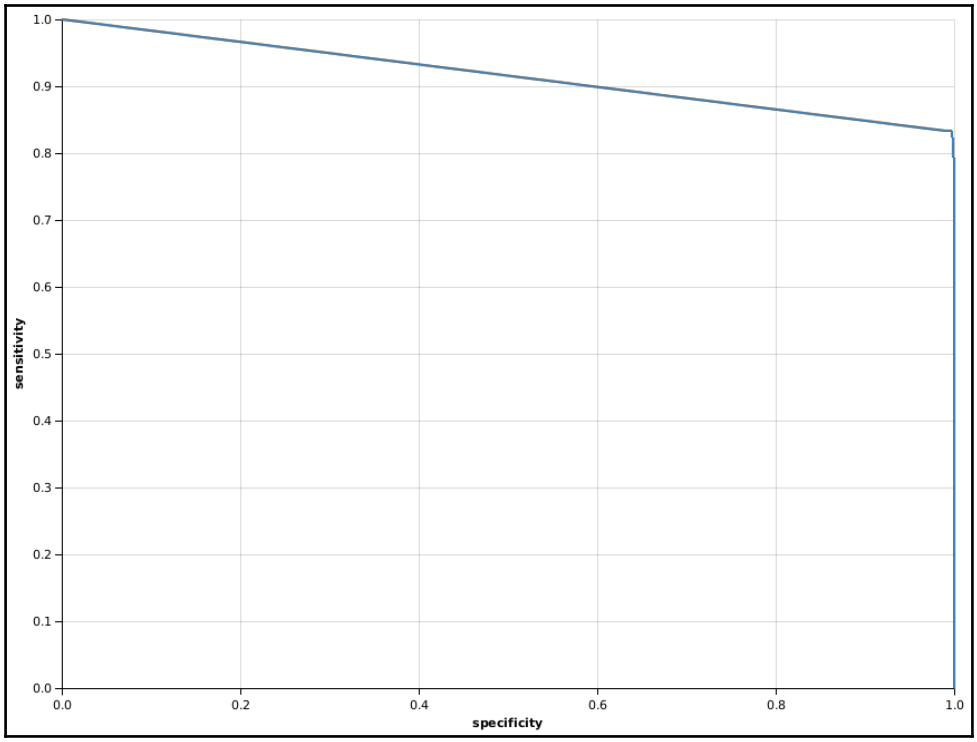
Reconstruction-MSE	Class	idRow
7.956054440127814E-7	0	0
6.106976382315064E-7	0	1
6.558716037639693E-6	0	2
8.753354094541301E-7	0	3
8.759964479355671E-7	0	4
5.720411000239647E-7	0	5
1.015871534556162E-6	0	6
1.110346446902160...	1	7
8.251552481892629E-7	0	8
4.216736950810848E-6	0	9
3.548204047963272...	0	10
3.300497481928 25E-7	0	11
4.504725850972922E-6	0	12
5.81244841042056E-7	0	13
8.988460213316904E-7	0	14
4.308887907806179E-6	0	15
7.33283846864895E-7	0	16
2.571894655334004...	0	17
2.341862714880748...	0	18
6.336200448585121E-7	0	19

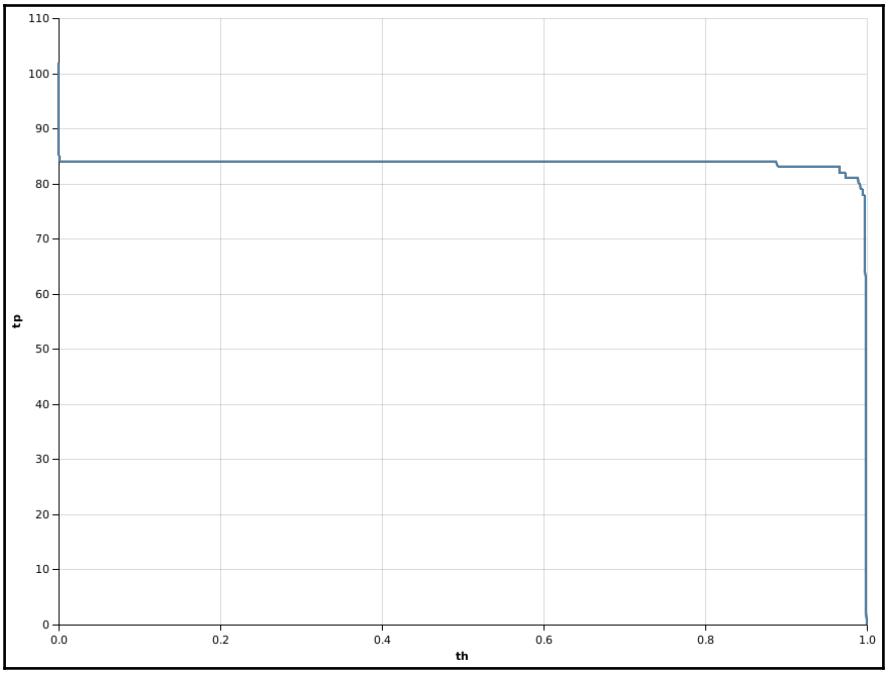
only showing top 20 rows

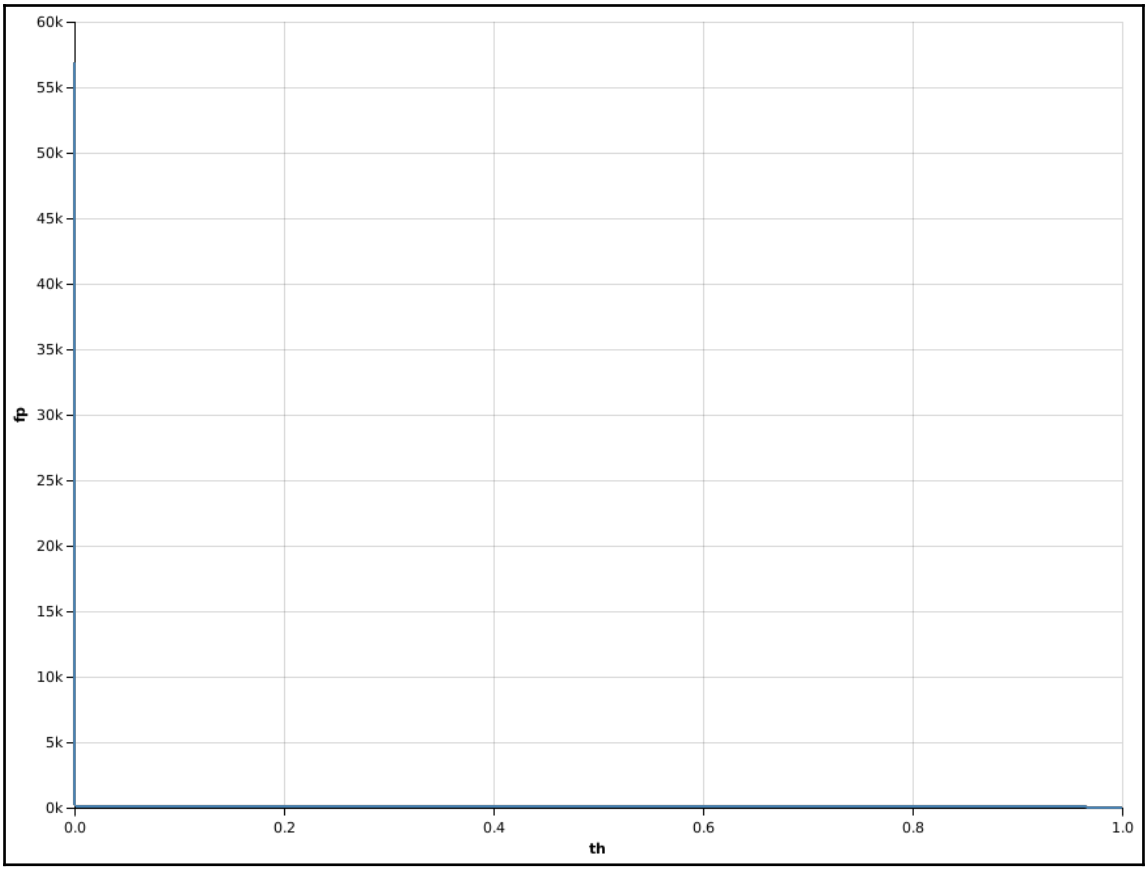

Class
0
1
Reconstruction-MSE
40µ
35µ
30µ
25µ
20µ
15µ
10µ
5µ
0µ
0k
10k
20k
30k
40k
50k
60k
idRow

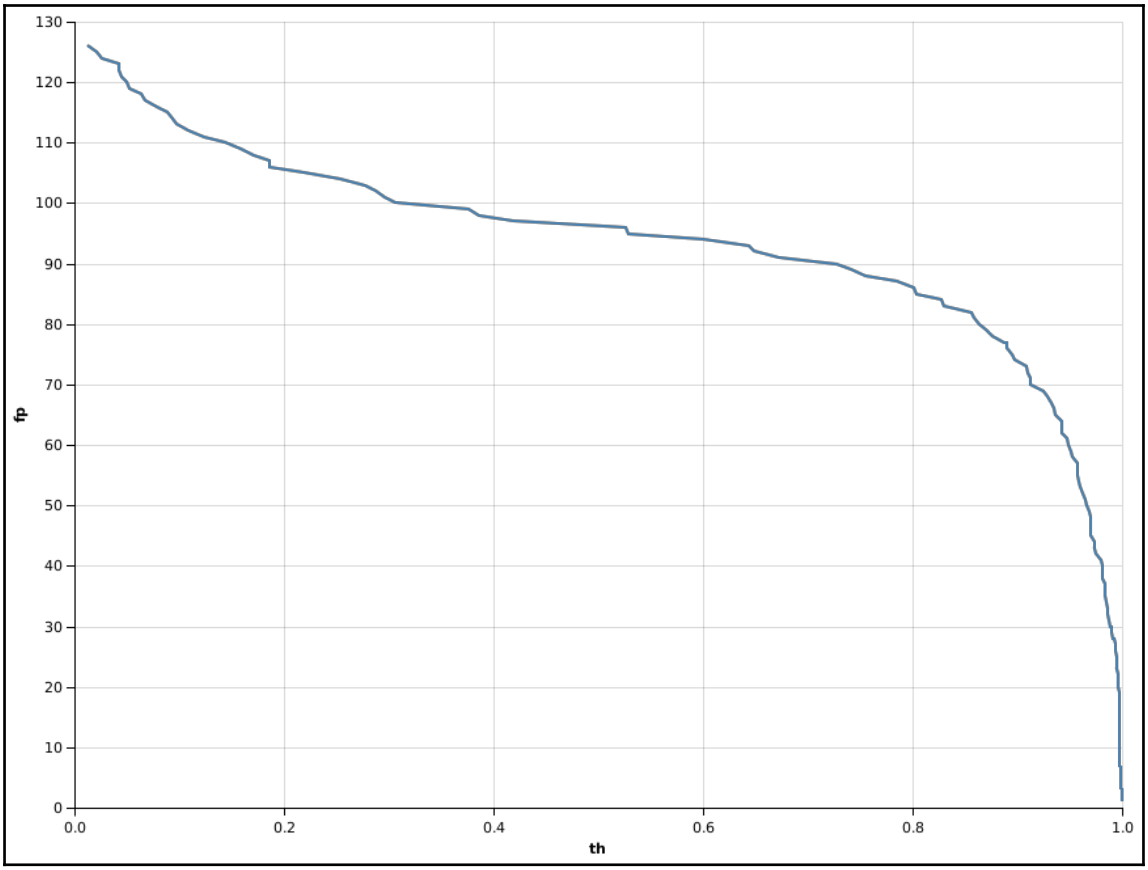

predict
0
1
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
1
Class

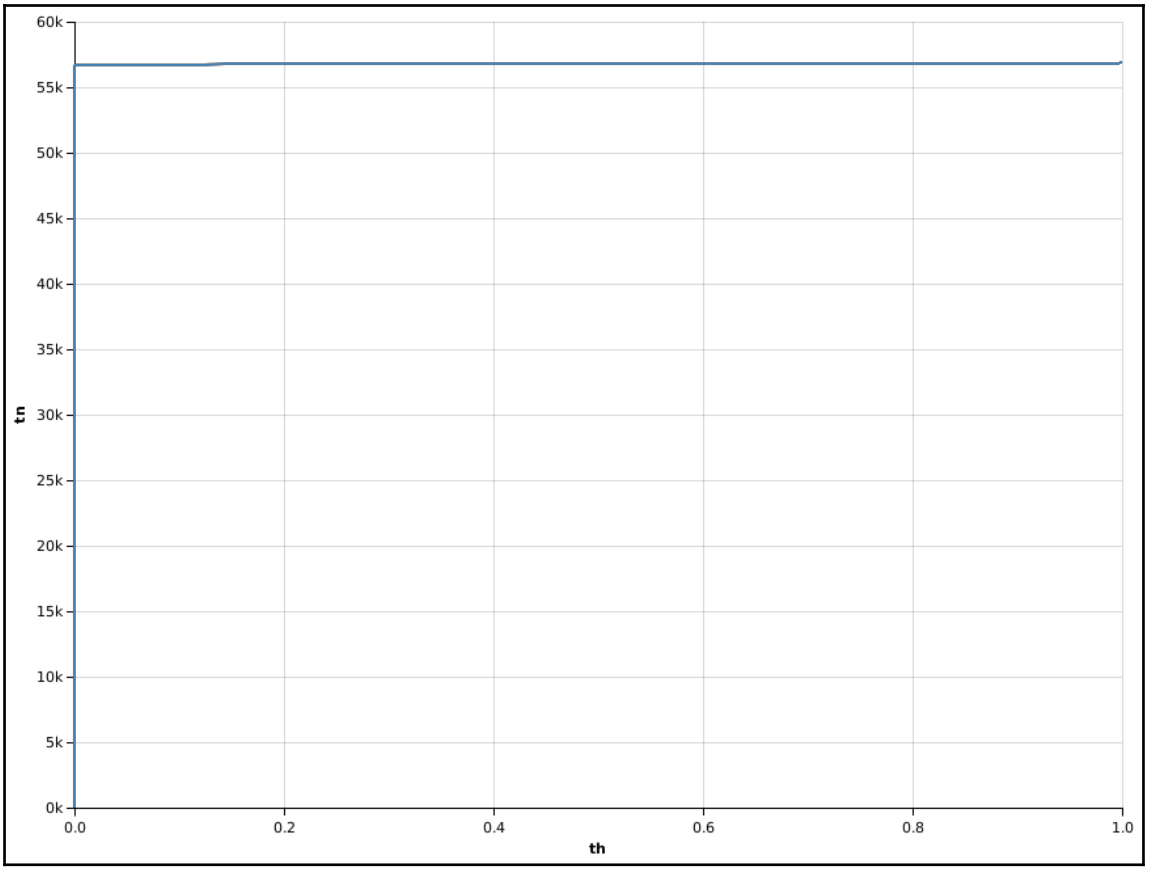

precision
recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0

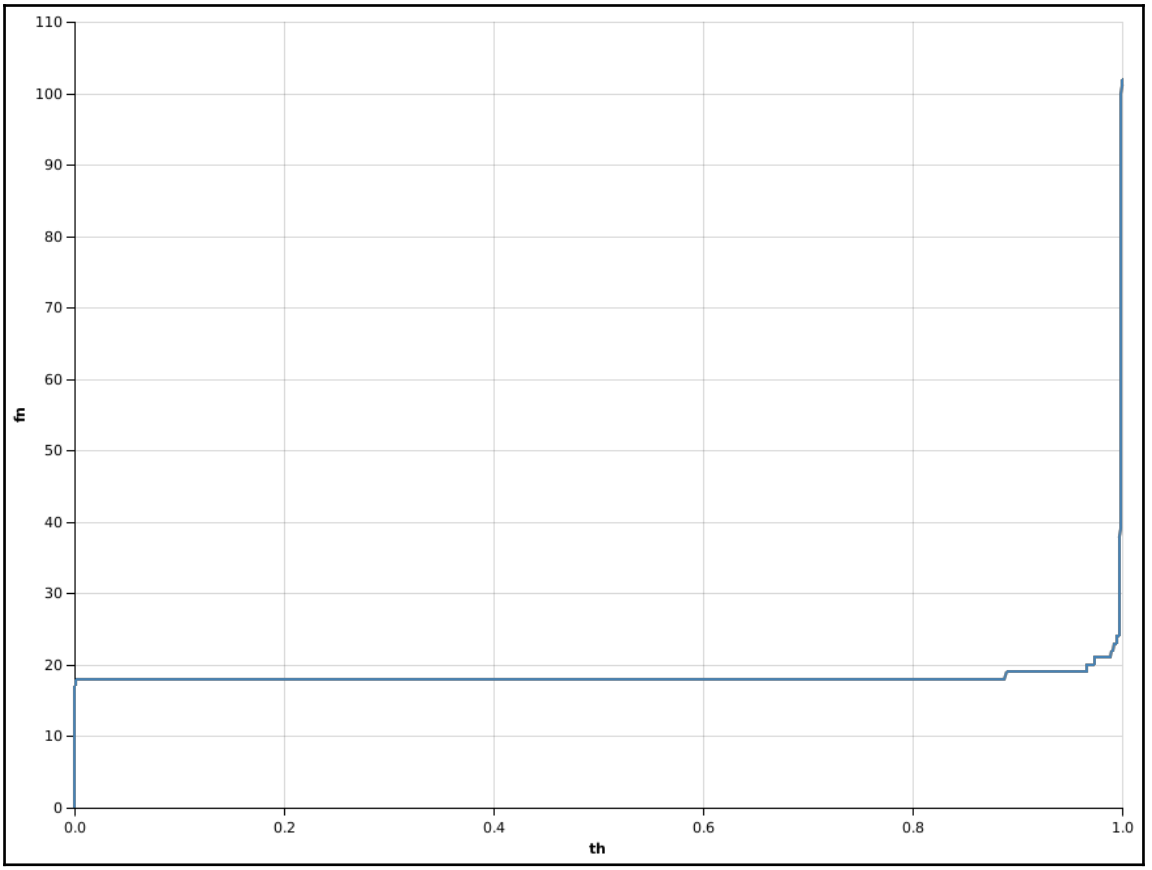

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
specificity


110
100
90
80
70
60
50
40
30
20
10
0
tp
0.0
0.2
0.4
0.6
0.8
1.0
th


60k
55k
50k
45k
40k
35k
30k
25k
20k
15k
10k
5k
0k
fp
0.0
0.2
0.4
0.6
0.8
1.0
th

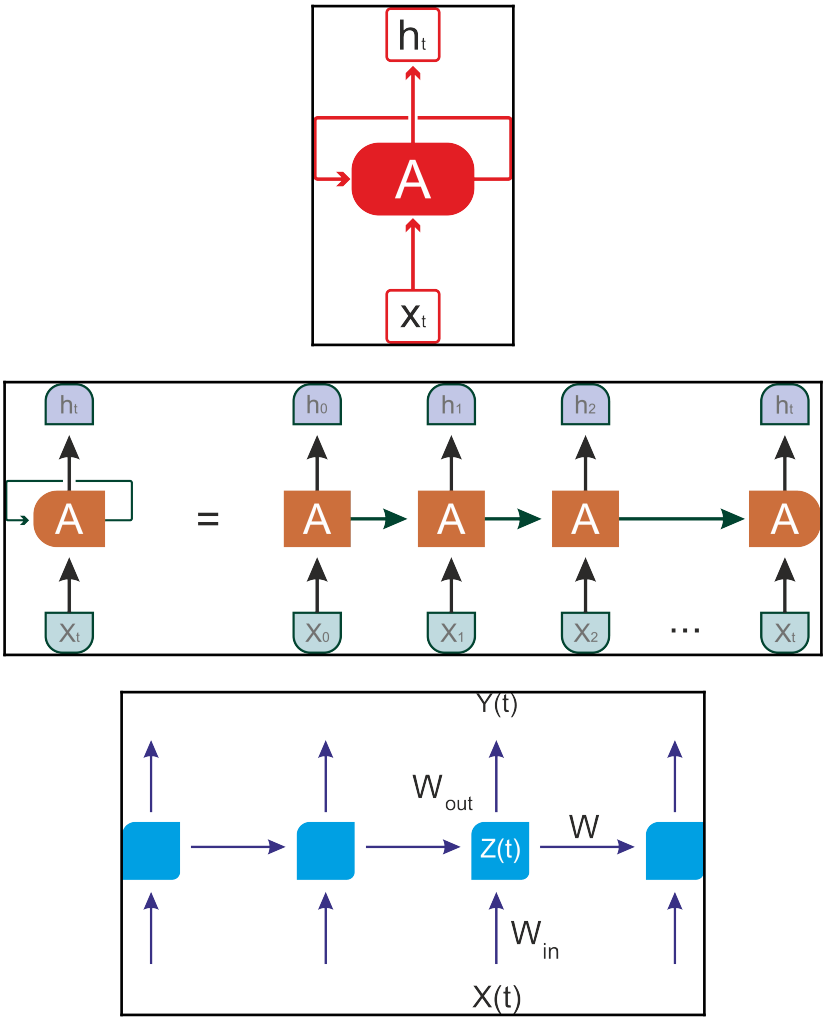

130
120
110
100
90
80
70
60
50
40
30
20
10
0
fp
0.0
0.2
0.4
0.6
0.8
1.0
th

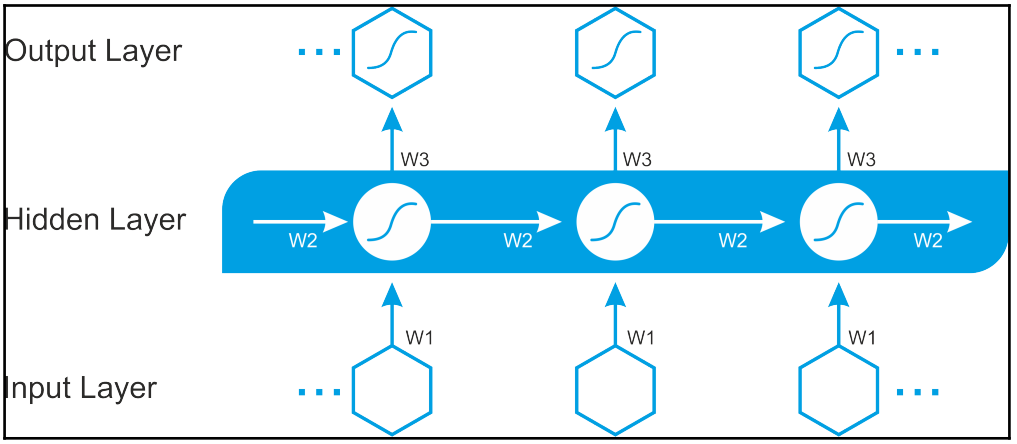

60k
55k
50k
45k
40k
35k
30k
25k
20k
15k
10k
5k
0k
tn
0.0
0.2
0.4
0.6
0.8
1.0
th

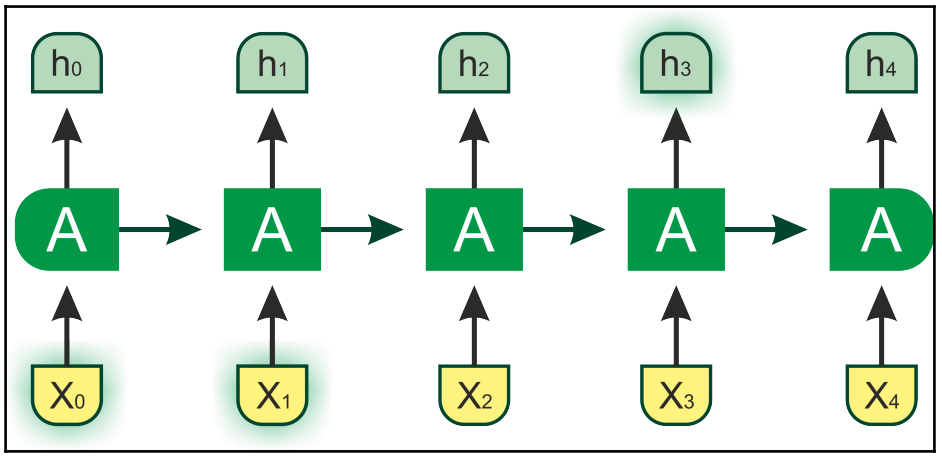

110
100
90
80
70
60
50
40
30
20
10
0
fn
0.0
0.2
0.4
0.6
0.8
1.0
th

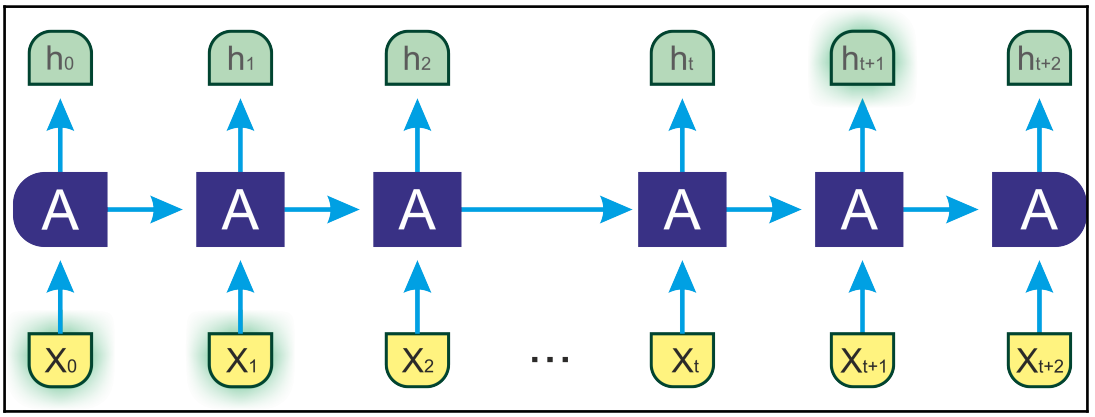
```
INFO: Model Metrics Type: Binomial
INFO:  Description: Metrics reported on temporary training frame with 9975 samples
INFO:  model id: DeepLearning_model_1510449558864_3
INFO:  frame id: null
INFO:  MSE: 0.013227416
INFO:  RMSE: 0.11501051
INFO:  AUC: 0.99984205
INFO:  logloss: 0.04371999
INFO:  mean_per_class_error: 8.1103E-4
INFO:  default threshold: 0.06491753458976746
INFO:  CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class):
INFO:            0     1   Error       Rate
INFO:      0  4924     8  0.0016  8 / 4,932
INFO:      1     0  5043  0.0000  0 / 5,043
INFO: Totals  4924  5051  0.0008  8 / 9,975
INFO: Gains/Lift Table (Avg response rate: 50.56 %):
INFO: Variable Importances:
INFO:         Variable Relative Importance Scaled Importance Percentage
INFO:              V10            1.000000          1.000000   0.054188
INFO:               V2            0.816018          0.816018   0.044218
INFO:              V27            0.729896          0.729896   0.039551
INFO:               V4            0.726038          0.726038   0.039342
INFO:              V12            0.717349          0.717349   0.038872
INFO:              V11            0.709694          0.709694   0.038457
INFO:               V9            0.658070          0.658070   0.035659
INFO:              V22            0.632086          0.632086   0.034251
INFO:               V1            0.624529          0.624529   0.033842
INFO:               V7            0.619922          0.619922   0.033592
INFO: ---
INFO:         Time.gr1            0.451975          0.451975   0.024492
INFO:              V14            0.450914          0.450914   0.024434
INFO:              V15            0.424829          0.424829   0.023021
INFO:              V13            0.397348          0.397348   0.021531
INFO:               V8            0.391477          0.391477   0.021213
INFO:               V6            0.383357          0.383357   0.020773
INFO:              V17            0.352461          0.352461   0.019099
INFO:              V21            0.339572          0.339572   0.018401
INFO:         Time.gr2            0.337747          0.337747   0.018302
INFO: Time.missing(NA)            0.000000          0.000000   0.000000
```

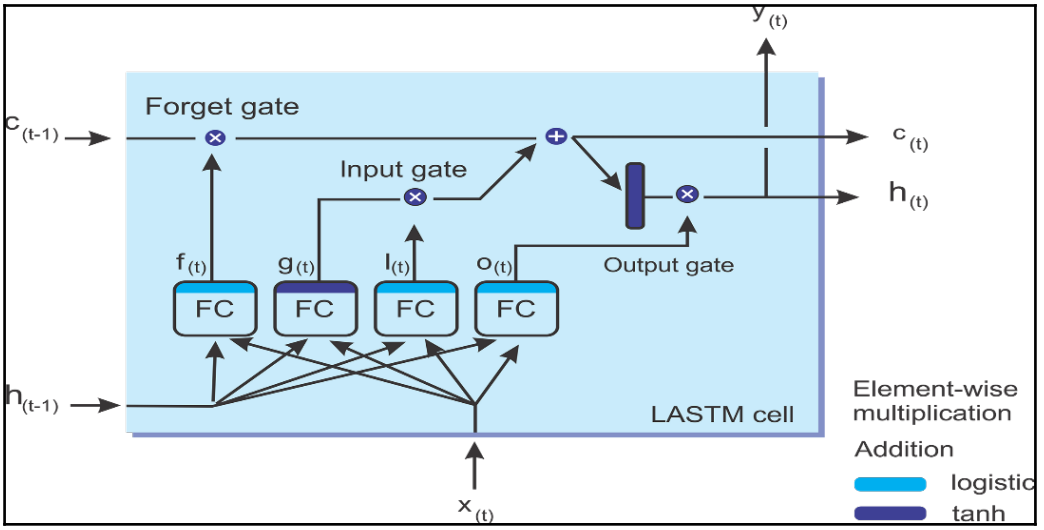
Chapter 10: Human Activity Recognition using Recurrent Neural Networks




Output Layer
Hidden Layer
Input Layer
W1
W2
W3


h0
h1
h2
h3
h4
A
X0
X1
X2
X3
X4


h0
h1
h2
ht
ht+1
ht+2
A
X0
X1
X2
Xt
Xt+1
Xt+2



$$\mathbf{i}_{(t)} = \sigma\left(\mathbf{W}_{xi}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hi}{}^{T} \cdot \mathbf{h}_{(t-1)} + \mathbf{b}_{i}\right)$$

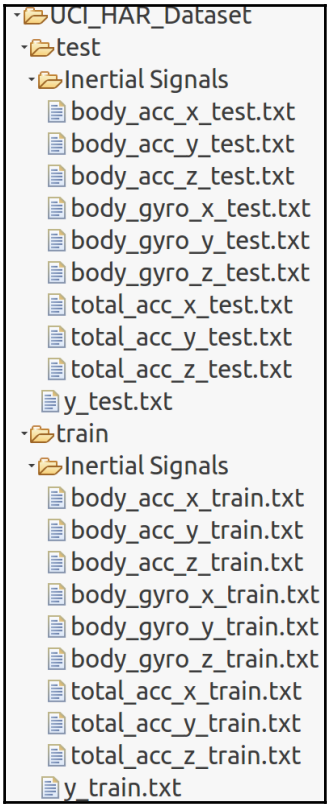
$$\mathbf{f}_{(t)} = \sigma\left(\mathbf{W}_{xf}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hf}{}^{T} \cdot \mathbf{h}_{(t-1)} + \mathbf{b}_{f}\right)$$

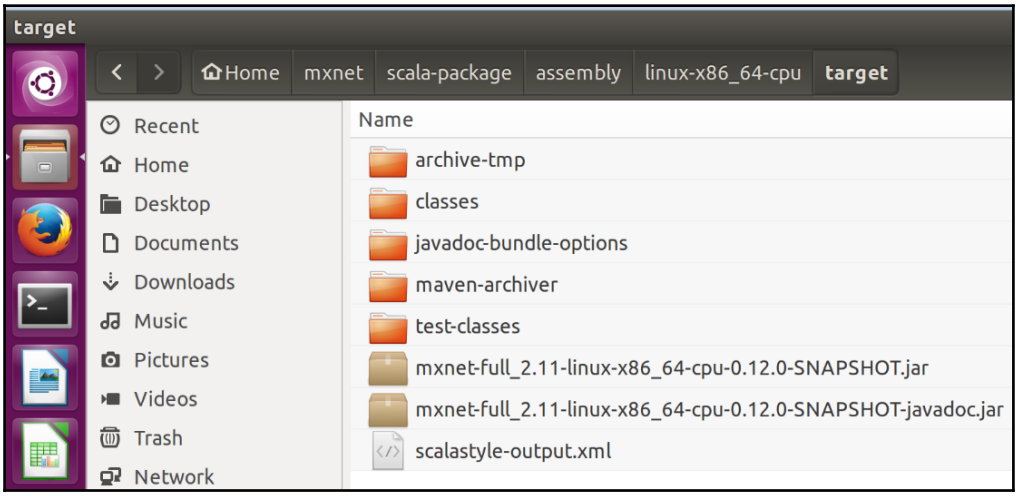
$$\mathbf{o}_{(t)} = \sigma\left(\mathbf{W}_{xo}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{ho}{}^{T} \cdot \mathbf{h}_{(t-1)} + \mathbf{b}_{o}\right)$$

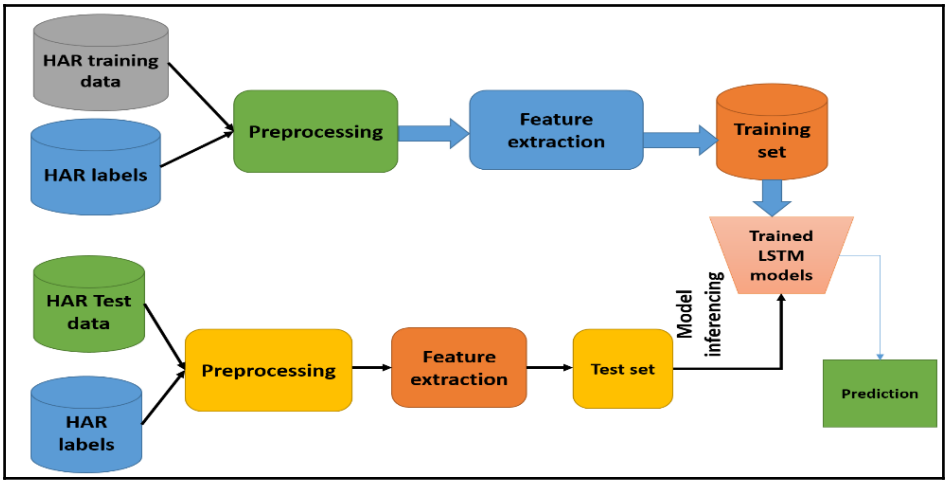
$$\mathbf{g}_{(t)} = \tanh\left(\mathbf{W}_{xg}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hg}{}^{T} \cdot \mathbf{h}_{(t-1)} + \mathbf{b}_{g}\right)$$

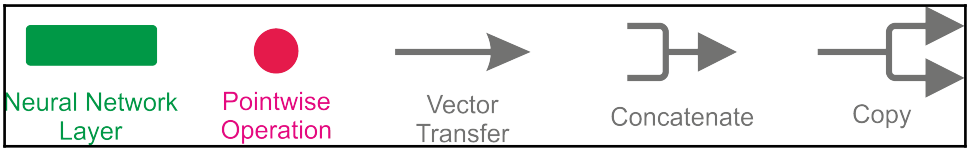
$$\mathbf{c}_{(t)} = \mathbf{f}_{(t)} \otimes \mathbf{c}_{(t-1)} + \mathbf{i}_{(t)} \otimes \mathbf{g}_{(t)}$$

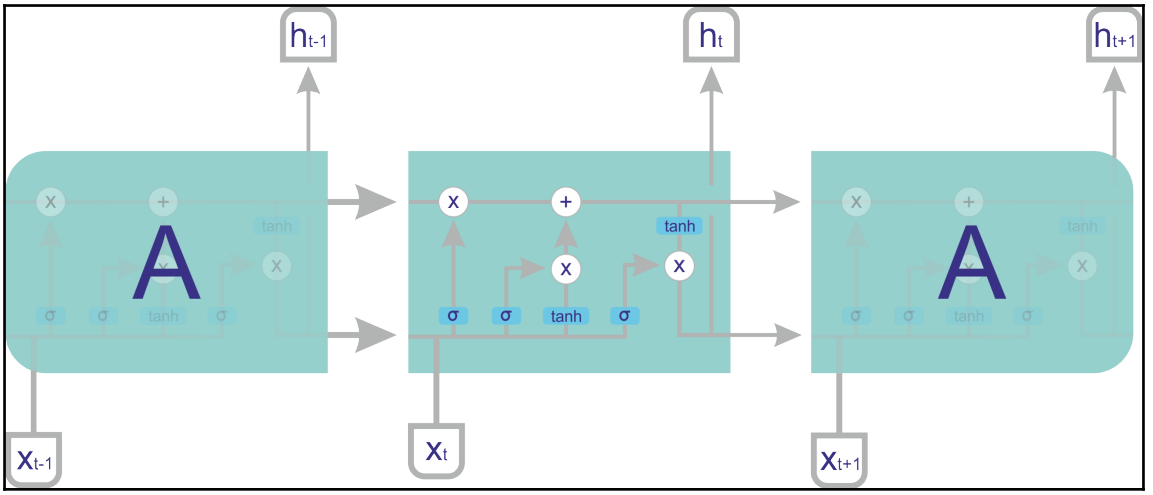
$$\mathbf{y}_{(t)} = \mathbf{h}_{(t)} = \mathbf{o}_{(t)} \otimes \tanh\left(\mathbf{c}_{(t)}\right)$$

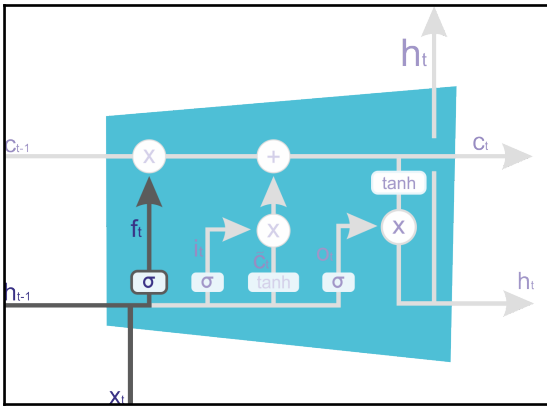

UCI_HAR_Dataset
test
Inertial Signals
body_acc_x_test.txt
body_acc_y_test.txt
body_acc_z_test.txt
body_gyro_x_test.txt
body_gyro_y_test.txt
body_gyro_z_test.txt
total_acc_x_test.txt
total_acc_y_test.txt
total_acc_z_test.txt
y_test.txt
train
Inertial Signals
body_acc_x_train.txt
body_acc_y_train.txt
body_acc_z_train.txt
body_gyro_x_train.txt
body_gyro_y_train.txt
body_gyro_z_train.txt
total_acc_x_train.txt
total_acc_y_train.txt
total_acc_z_train.txt
y_train.txt


target
Home
mxnet
scala-package
assembly
linux-x86_64-cpu
target
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Name
archive-tmp
classes
javadoc-bundle-options
maven-archiver
test-classes
mxnet-full_2.11-linux-x86_64-cpu-0.12.0-SNAPSHOT.jar
mxnet-full_2.11-linux-x86_64-cpu-0.12.0-SNAPSHOT-javadoc.jar
scalastyle-output.xml

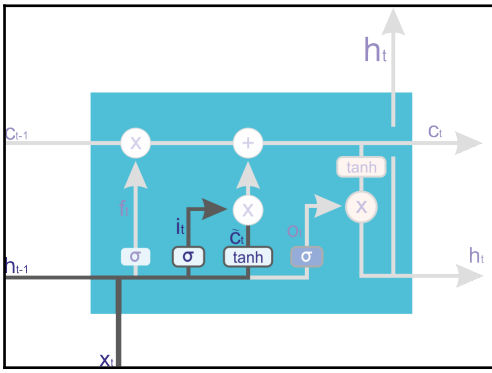

HAR training data
HAR labels
Preprocessing
Feature extraction
Training set
Trained LSTM models
HAR Test data
HAR labels
Preprocessing
Feature extraction
Test set
Model inferencing
Prediction


Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy


h_{t-1}
h_t
h_{t+1}
A
A
tanh
σ
x_{t-1}
x_t
x_{t+1}

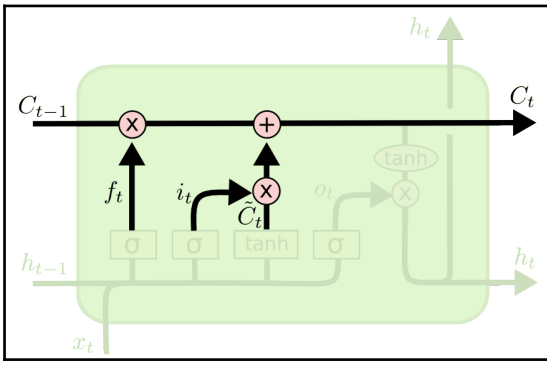


$$f_t = \sigma\left(W_f \cdot [h_t - 1, x_t] + b_f\right)$$

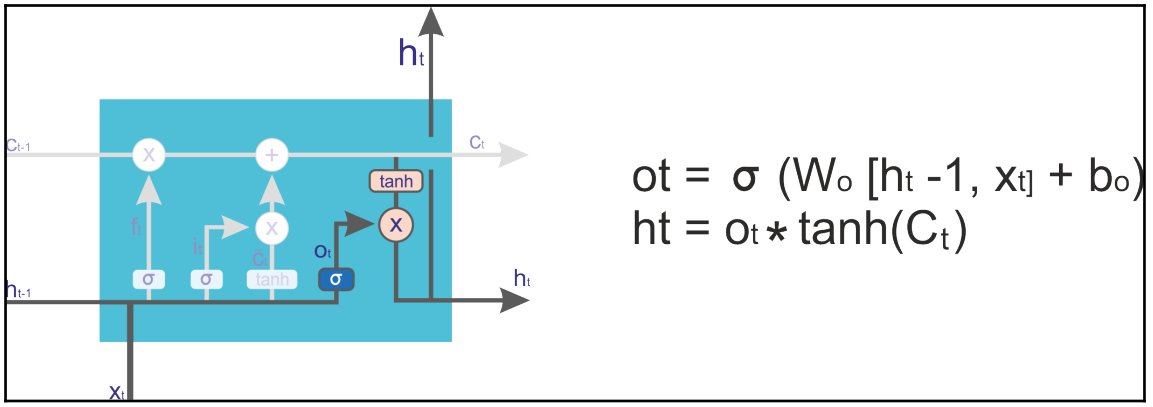


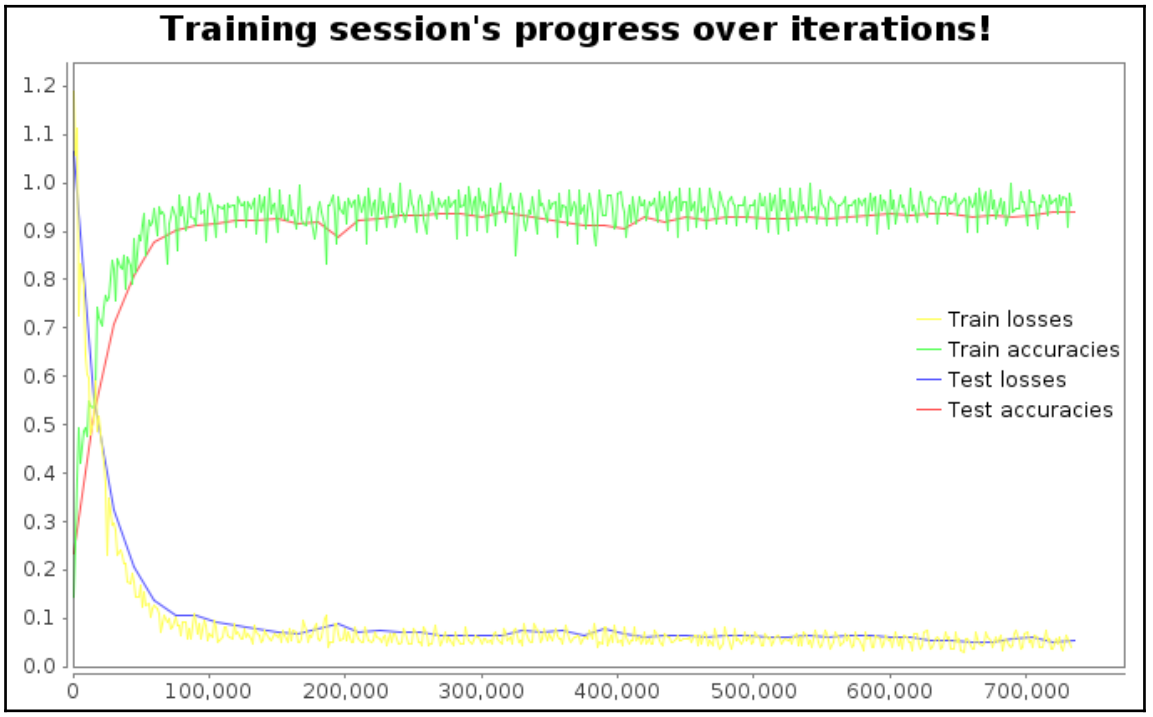
$$i_t = \sigma\left(W_i \cdot [h_t - 1, x_t] + b_i\right)$$

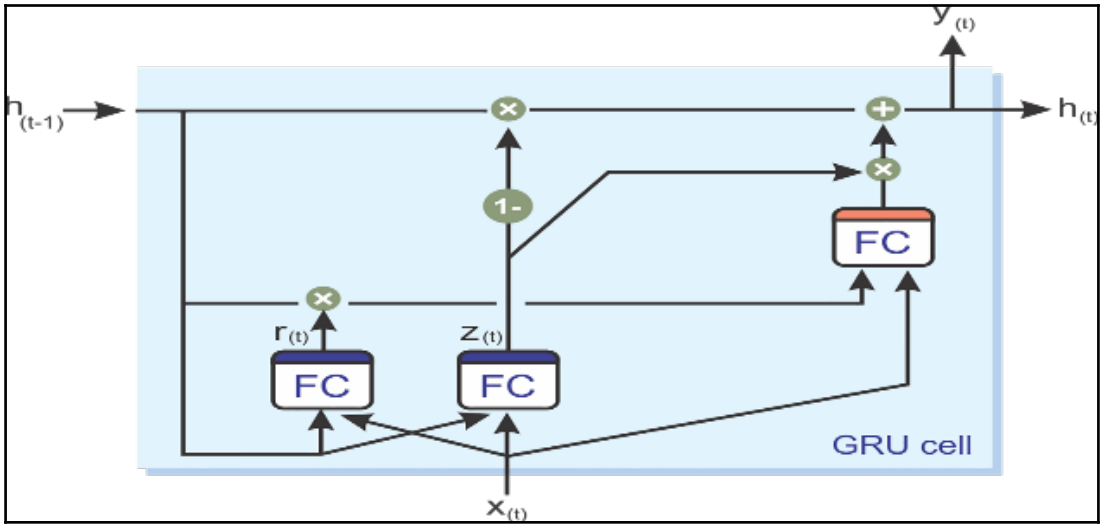
$$\tilde{C}_t = \tanh\left(W_c \cdot [h_t - 1, x_t] + b_c\right)$$



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$


ht
Ct-1
Ct
tanh
ft
it
Ct
ot
ht-1
ht
σ
σ
tanh
σ
xt
ot = σ (Wo [ht -1, xt] + bo)
ht = ot * tanh(Ct)


Training session's progress over iterations!
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100,000
200,000
300,000
400,000
500,000
600,000
700,000
Train losses
Train accuracies
Test losses
Test accuracies



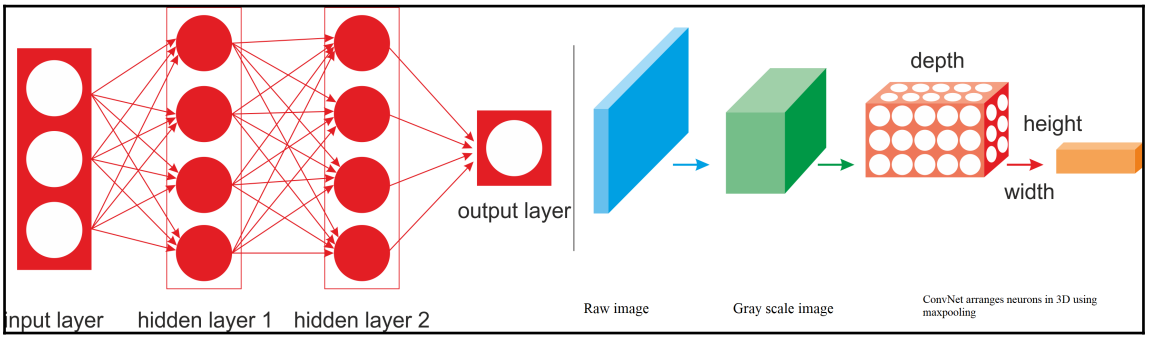
$$\mathbf{z}_{(t)} = \sigma\left(\mathbf{W}_{xz}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hz}{}^{T} \cdot \mathbf{h}_{(t-1)}\right)$$

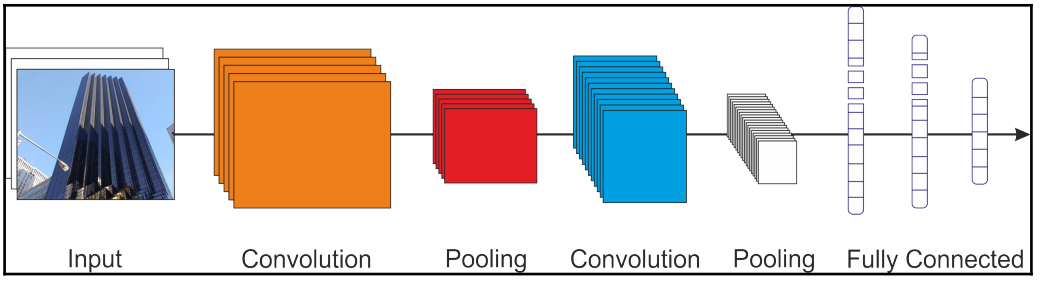
$$\mathbf{r}_{(t)} = \sigma\left(\mathbf{W}_{xr}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hr}{}^{T} \cdot \mathbf{h}_{(t-1)}\right)$$

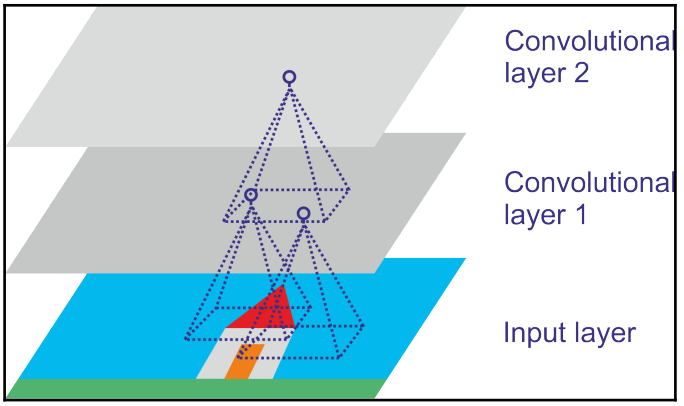
$$\mathbf{g}_{(t)} = \tanh\left(\mathbf{W}_{xg}{}^{T} \cdot \mathbf{x}_{(t)} + \mathbf{W}_{hg}{}^{T} \cdot \left(\mathbf{r}_{(t)} \otimes \mathbf{h}_{(t-1)}\right)\right)$$

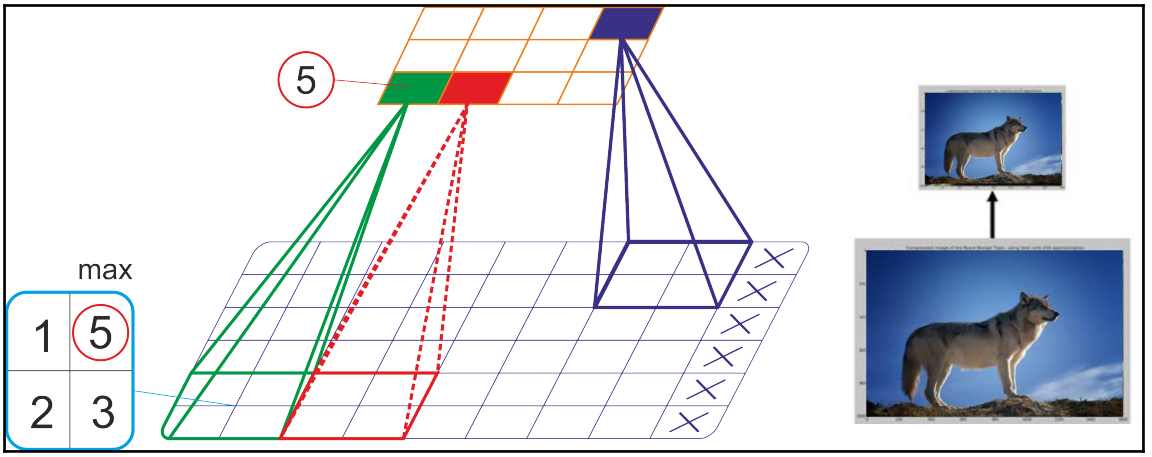
$$\mathbf{h}_{(t)} = \left(1 - \mathbf{z}_{(t)}\right) \otimes \mathbf{h}_{(t-1)} + \mathbf{z}_{(t)} \otimes \mathbf{g}_{(t)}$$

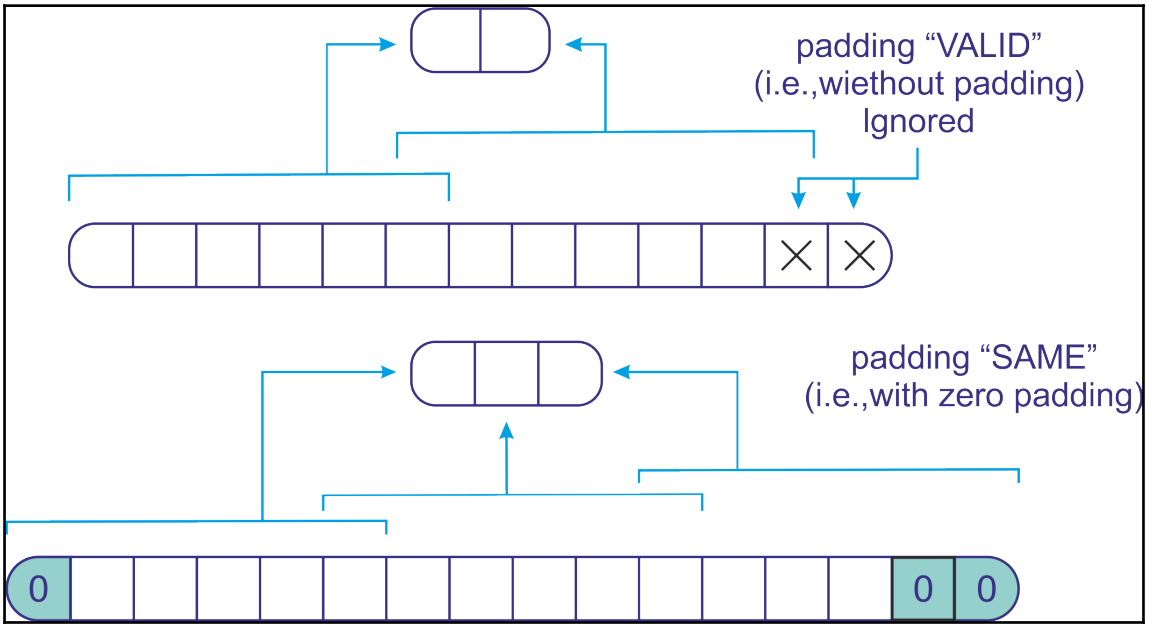
Chapter 11: Image Classification using Convolutional Neural Networks



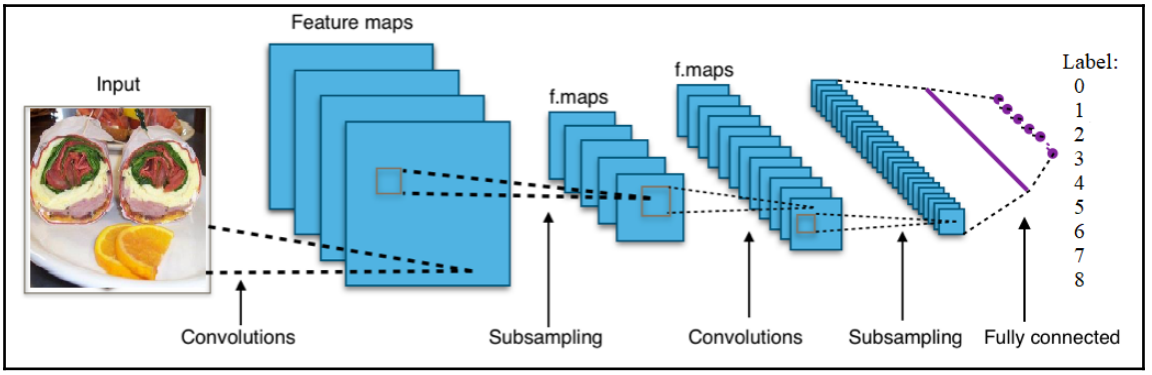


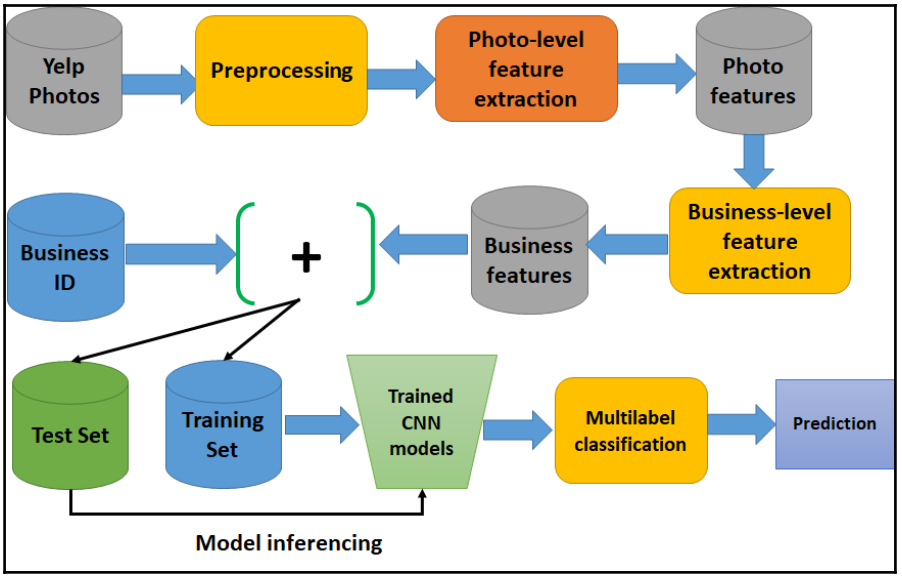


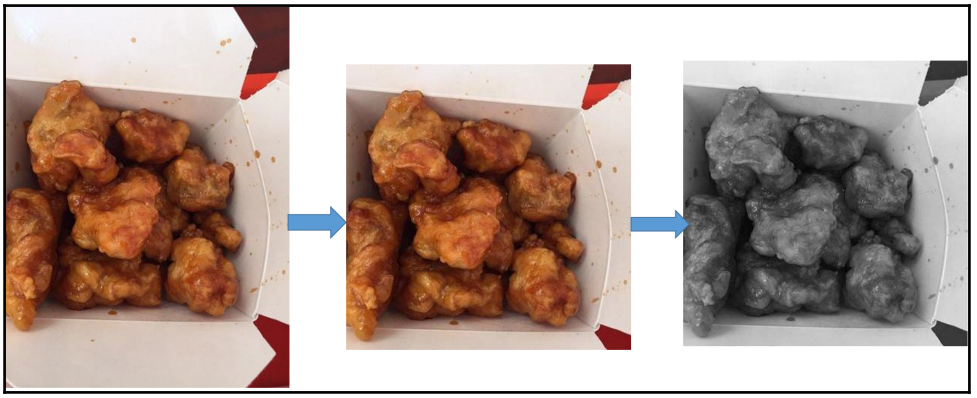

5
max
1
5
2
3

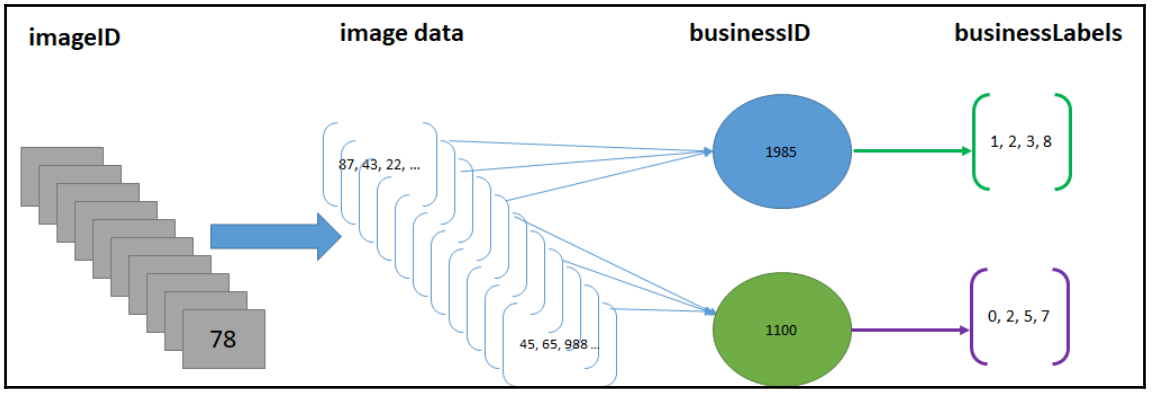

padding “VALID”
(i.e.,wiethout padding)
Ignored
padding “SAME”
(i.e.,with zero padding)
0
0
0

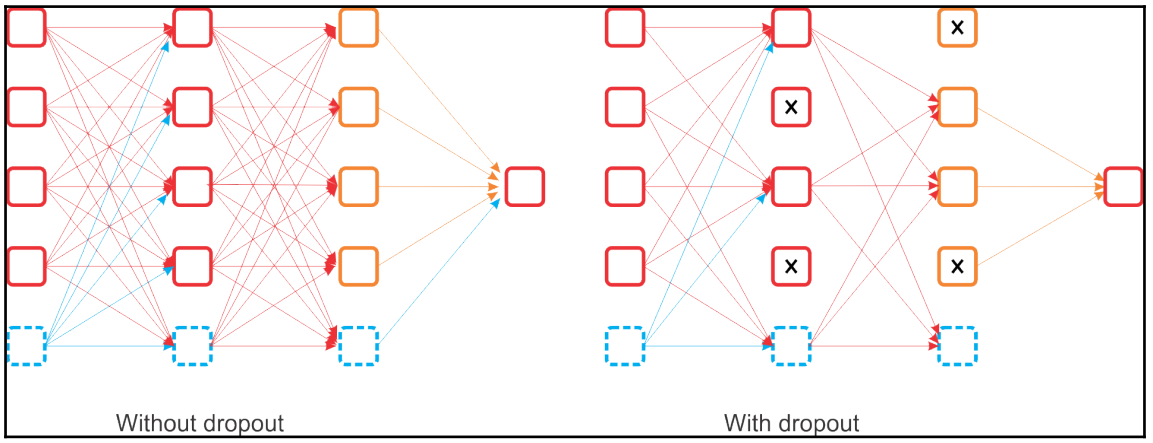
More business info
Takes Reservations No
Delivery No
Take-out Yes
Accepts Credit Cards Yes
Good For Breakfast
Parking Street
Bike Parking Yes
Wheelchair Accessible No
Good for kids No
Good for Groups No
Attire Casual
Ambience Casual


Feature maps
Input
f.maps
f.maps
Label:
0
1
2
3
4
5
6
7
8
Convolutions
Subsampling
Convolutions
Subsampling
Fully connected


Yelp Photos
Preprocessing
Photo-level feature extraction
Photo features
Business ID
+
Business features
Business-level feature extraction
Test Set
Training Set
Trained CNN models
Multilabel classification
Prediction
Model inferencing




imageID
image data
businessID
businessLabels
78
87, 43, 22, ...
45, 65, 988 ...
1985
1100
1, 2, 3, 8
0, 2, 5, 7


Without dropout
With dropout

Index