



UNIVERSITÀ DEGLI STUDI DI TRENTO

Department of Engineering and Information Science

Master in Computer Science

RESEARCH PROJECT IN MULTIMEDIA DATA SECURITY

CLASSIFICATION OF SHARING APPLICATIONS

December 3, 2019

Supervisors:
Prof. Giulia Boato
PHD. Quoc Tin Phan

Student:
Kritjan Gjika

Academic Year 2018/2019

Contents

Bibliografia	2
--------------	---

1 Single Scenario Classification, KFold Validation

Starting with fitting randomly the classifiers, there are some statistics of the data used for the first test:

	count train	count test
messenger	249	100
telegram	244	106
whatsapp	243	107
original	243	107

1.1 Logistic regression results:

Confusion matrix with number of sample and with normalization:

	messenger	telegram	whatsapp	original
messenger	100	0	0	0
telegram	0	106	0	0
whatsapp	0	0	103	4
original	0	0	0	107

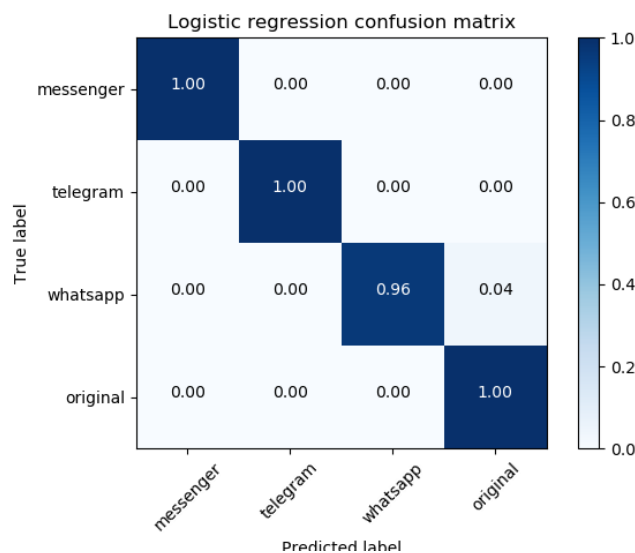


Figure 1.1: logistic regression

Result of the KFold validation with 10 bins:

1.0000	0.9898	0.9898	1.0000	1.0000	0.9796	0.9898	1.0000	0.9898	1.0000
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

The mean is : 0.993878

1.2 Linear Support Vector Machine results:

Confusion matrix with number of sample and with normalization:

	messenger	telegram	whatsapp	original
messenger	100	0	0	0
telegram	0	106	0	0
whatsapp	0	0	103	4
original	0	0	0	107

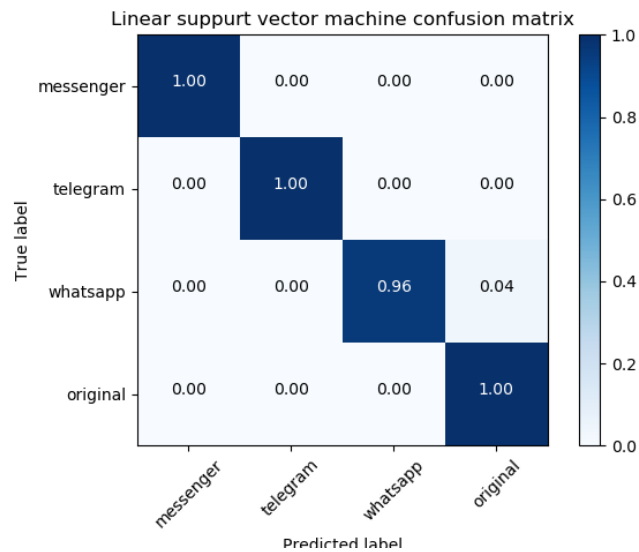


Figure 1.2: linear SVM

Result of the KFold validation with 10 bins:

1.0000	0.9898	1.0000	1.0000	1.0000	0.9898	0.9898	1.0000	0.9898	1.0000
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

The mean is : 0.995918

1.3 Random forest results:

Confusion matrix with number of sample and with normalization:

	messenger	telegram	whatsapp	original
messenger	100	0	0	0
telegram	0	106	0	0
whatsapp	0	0	103	4
original	0	0	0	107

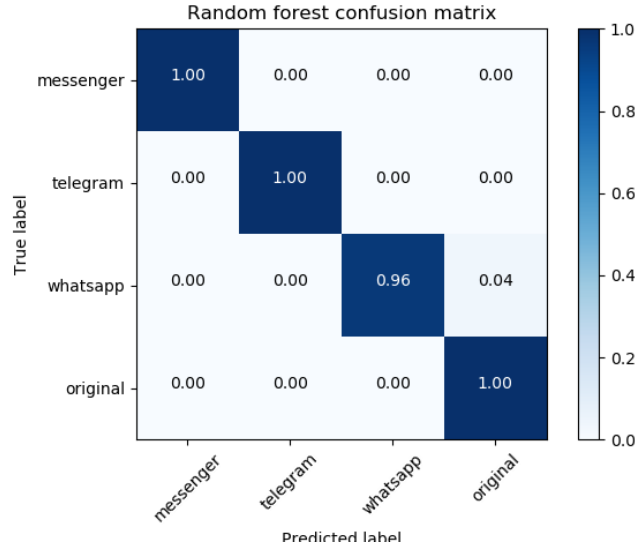


Figure 1.3: random forest

Result of the KFold validation with 10 bins:

1.0000	0.9898	1.0000	1.0000	1.0000	0.9796	0.9898	1.0000	0.9898	1.0000
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

The mean is : 0.994898

2 Single Scenario Classification, Circularly Validation

Here was used the same dataset as before but the training used a 0.3 of the dataset, and it is shifted circularly to cover all the dataset. Here is the table of all steps calculated

step	logistic	linear SVM	random fo.
0	0.9831894364316239	0.9791077839737224	0.987923728495032
1	0.9853026547006402	0.9783844889010689	0.9861107634603096
2	0.9853026547006402	0.9775285159568015	0.9890580733784273
3	0.9853026547006402	0.9783844889010689	0.987165432614862
4	0.9871718848829855	0.9783844889010689	0.9909805452235539
5	0.981755684822845	0.9908606639439437	0.989920141798295
6	0.981755684822845	0.9908606639439437	0.989920141798295
7	0.981755684822845	0.9908606639439437	0.989920141798295
8	0.981755684822845	0.9908606639439437	0.9888913704941559
9	0.9797774593441748	0.9846882267380576	0.9826969666799417
10	0.9807773613945863	0.9857299945019243	0.9827095131017022
11	0.9807773613945863	0.9846882267380576	0.9817251354156551
12	0.9817853927299603	0.9846882267380576	0.9817685950413223
13	0.9817853927299603	0.9857299945019243	0.9826969666799417
14	0.9817853927299603	0.9857299945019243	0.9826773061122597
15	0.9828271604938271	0.9857299945019243	0.9847545798486659
16	0.9828271604938271	0.9857299945019243	0.9798383818675511

17	0.9809303350970018	0.9809303350970018	0.9769817171132961
18	0.9809303350970018	0.9809303350970018	0.9769817171132961
19	0.9809303350970018	0.9809303350970018	0.9769817171132961
20	0.9809303350970018	0.9809303350970018	0.9769817171132961
21	0.9809303350970018	0.9809303350970018	0.97795683313976
22	0.9779061415486145	0.9799140749344002	0.97795683313976
23	0.9779061415486145	0.978906043599026	0.9713928225908935
24	0.9779061415486145	0.978906043599026	0.9713928225908935
25	0.9779061415486145	0.9779067519576579	0.9744170161392807
26	0.9769068499072464	0.9779067519576579	0.9744170161392807
27	0.9820489340633134	0.9849476193424314	0.9820489340633134
28	0.9820489340633134	0.9849476193424314	0.9812130271108159
29	0.9820489340633134	0.9808709993428069	0.9832038974293551
30	0.9820489340633134	0.9799805623885003	0.981280651795602
31	0.9820489340633134	0.9799805623885003	0.9832038974293551
32	0.9810570633546724	0.9799144970334153	0.9729849311036389
33	0.9627812747783722	0.9583378598311828	0.9542004398105941
34	0.953283825521887	0.9489212587825018	0.953283825521887
35	0.953283825521887	0.9521822193250765	0.9332107165025093
36	0.9521336702615147	0.9510715431596557	0.9342712270274949
37	0.9553966439474894	0.9586655168859602	0.941807112194959
38	0.9550624888401142	0.9501411675781424	0.941807112194959
39	0.9508728056918877	0.957897835375157	0.9426760297719203
40	0.949001628304907	0.9569165446997437	0.9444336881405264
41	0.949001628304907	0.9570023590744005	0.9426760297719203
42	0.9518168607439983	0.9570023590744005	0.9444336881405264
43	0.9628462719278571	0.9690702601489536	0.9670237271116444
44	0.9641308448001574	0.9690415311211439	0.9680824885108373
45	0.9611971340022054	0.9540513512651186	0.9724089271961905
46	0.9636358185061074	0.9539595767948307	0.9713033424446343
47	0.9634237045681469	0.9549009600982763	0.9691486092937285
48	0.9630416021422119	0.9457796737082027	0.9670237271116444
49	0.9590950950950952	0.9645468263150108	0.9651904231493449
50	0.9590950950950952	0.9645468263150108	0.9631028529724224
51	0.9590950950950952	0.9645468263150108	0.9631028529724224
52	0.9590950950950952	0.9645468263150108	0.961972924871707
53	0.9704082737708469	0.9750147841513898	0.9734561922907445
54	0.9749984999849999	0.9794007490636705	0.9778754788737738
55	0.9749984999849999	0.9794007490636705	0.9800443458980044
56	0.9749984999849999	0.9794007490636705	0.975720745722868
57	0.9765285344002652	0.9802631578947368	0.9779480414953162
58	0.9775769230769231	0.9792468977321352	0.9790280857354028
59	0.9775769230769231	0.9802631578947368	0.9780137313157126
60	0.9756841360427018	0.978544776119403	0.980188679245283
61	0.9757692307692307	0.978544776119403	0.980188679245283
62	0.9746787205323791	0.9794007490636705	0.980188679245283
63	0.9959839357429718	0.9989837398373984	0.9969594483394227
64	0.9899551291586098	0.9959595462890113	0.9858384733026766
65	0.9814065924422243	0.989179818268771	0.9871656157809847
66	0.9851206921404461	0.989179818268771	0.9881081682594579
67	0.983242407923623	0.9801158153090965	0.9890580733784273

68	0.9841779184557378	0.9791077839737224	0.9870876215497557
69	0.9841779184557378	0.9791077839737224	0.9869750656167979

Average of all steps:

logistic r.	linear SVM	random f.
0.9742098967923488	0.9754432064396276	0.9743308314991495

Confusion matrix estimated on overall tests:

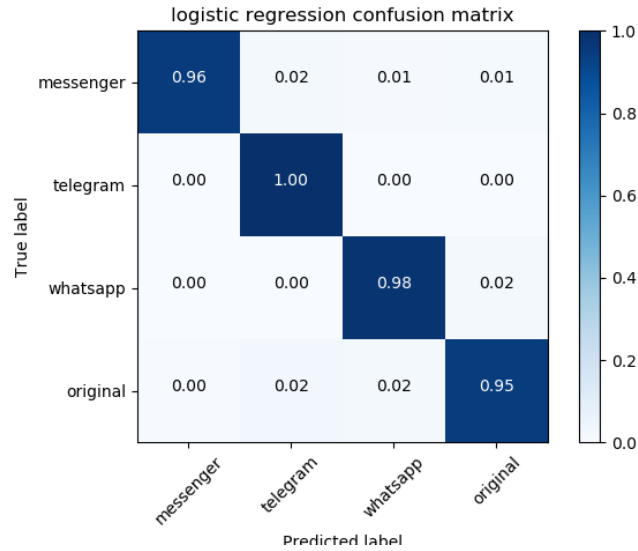


Figure 2.1: logistic regression

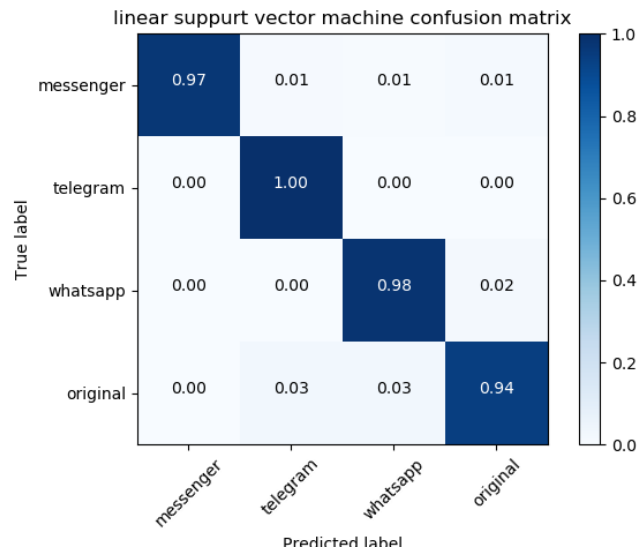


Figure 2.2: linear SVM

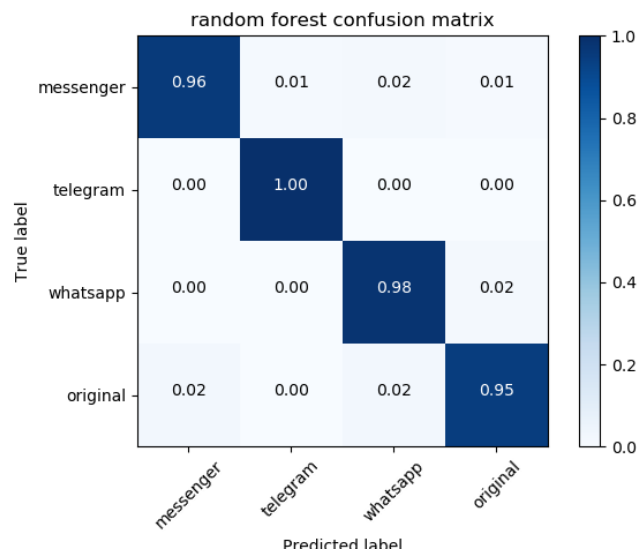


Figure 2.3: random forest

Bibliography