

Progetto Modelli Probabilistici per le Decisioni - Smarthouse

Matteo Angelo Costantini - 795125

Alessandro Longhi - 794235

Obiettivi

- ▶ 1. Definizione della struttura dati
 - ▶ Misurazioni derivanti da sensori, con relativi time stamp
 - ▶ Attività svolte
- ▶ 2. Modello HMM
 - ▶ Definizione della struttura HMM per inferire l'attività date le osservazioni derivanti dai sensori
 - ▶ Stima dei parametri
- ▶ 3. Previsione dell'attività svolta
 - ▶ Inferire l'attività dell'utente tramite i dati rilevati dai sensori
- ▶ 4. Analisi dei dati
 - ▶ Stimare le capacità predittive del modello rispetto alla ground truth

Dataset Iniziali

- ▶ Due dataset relativi a due abitazioni separate (A e B)
 - ▶ 14 giorni per il dataset A, 21 giorni per il dataset B
- ▶ Per ognuna due dataset in formato testuale:
 - ▶ Attività rilevate:
 - ▶ start_time
 - ▶ end_time
 - ▶ activity
 - ▶ Sensori attivi:
 - ▶ start_time
 - ▶ end_time
 - ▶ location
 - ▶ type
 - ▶ place
- ▶ 12 sensori
- ▶ 10 attività

Dataset Iniziali

► Attività

Start time	End time	Activity
2011-11-28 02:27:59	2011-11-28 10:18:11	Sleeping
2011-11-28 10:21:24	2011-11-28 10:23:36	Toileting
2011-11-28 10:25:44	2011-11-28 10:33:00	Showering
2011-11-28 10:34:23	2011-11-28 10:43:00	Breakfast
2011-11-28 10:49:48	2011-11-28 10:51:13	Grooming
2011-11-28 10:51:41	2011-11-28 13:05:07	Spare_Time/TV
2011-11-28 13:06:04	2011-11-28 13:06:31	Toileting
2011-11-28 13:09:31	2011-11-28 13:29:09	Leaving
2011-11-28 13:38:40	2011-11-28 14:21:40	Spare_Time/TV
2011-11-28 14:22:38	2011-11-28 14:27:07	Toileting
2011-11-28 14:27:11	2011-11-28 15:04:00	Lunch

► Sensori

Start time	End time	Location	Type	Place
2011-11-28 02:27:59	2011-11-28 10:18:11	Bed	Pressure	Bedroom
2011-11-28 10:21:24	2011-11-28 10:21:31	Cabinet	Magnetic	Bathroom
2011-11-28 10:21:44	2011-11-28 10:23:31	Basin	PIR	Bathroom
2011-11-28 10:23:02	2011-11-28 10:23:36	Toilet	Flush	Bathroom
2011-11-28 10:25:44	2011-11-28 10:32:06	Shower	PIR	Bathroom
2011-11-28 10:34:23	2011-11-28 10:34:41	Fridge	Magnetic	Kitchen
2011-11-28 10:34:44	2011-11-28 10:37:17	Cupboard	Magnetic	Kitchen

Preprocessing

- ▶ Correzione manuale di alcuni timestamp
- ▶ Eliminazione manuale di caratteri superflui
- ▶ Eliminazione delle righe inconsistenti (finiscono prima di iniziare)
- ▶ Conversione dei valori categorici in interi (es. attività, sensori)
- ▶ Divisione di attività e rilevazioni in timeslice
- ▶ Associazione dei sensori attivi durante ogni attività (bit vector)
 - ▶ Modellata assenza di attività
- ▶ Aggiunto periodo della giornata (divisa in quattro periodi)

Dataset Finali

- ▶ Uno per ogni abitazione
- ▶ Struttura:
 - ▶ timestamp: ora di inizio del timeslice
 - ▶ activity: attività corrente
 - ▶ sensors: bit vector dei sensori (es. 001010010110)
 - ▶ period: periodo della giornata
- ▶ Dataset A: 0: Breakfast, 1: Grooming, 2: Leaving, 3: Lunch, 4: Showering, 5: Sleeping, 6: Snack, 7: Spare_Time/TV, 8: Toileting, 9: No activity
- ▶ Dataset B: 0: Breakfast, 1: Dinner, 2: Grooming, 3: Leaving, 4: Lunch, 5: Showering, 6: Sleeping, 7: Snack, 8: Spare_Time/TV, 9: Toileting, 10: No activity

timestamp	activity	sensors	period
1323128160	7	000000001000	3
1323128220	7	000000001000	3
1323128280	7	000000001000	3
1323128340	8	000000000001	3
1323128400	8	100000000001	3
1323128460	9	000000000000	3
1323128520	5	010000000000	3
1323128580	5	010000000000	3
1323128640	5	010000000000	3
1323128700	5	010000000000	3
1323128760	5	010000000000	3

Modello - Hidden Markov Model

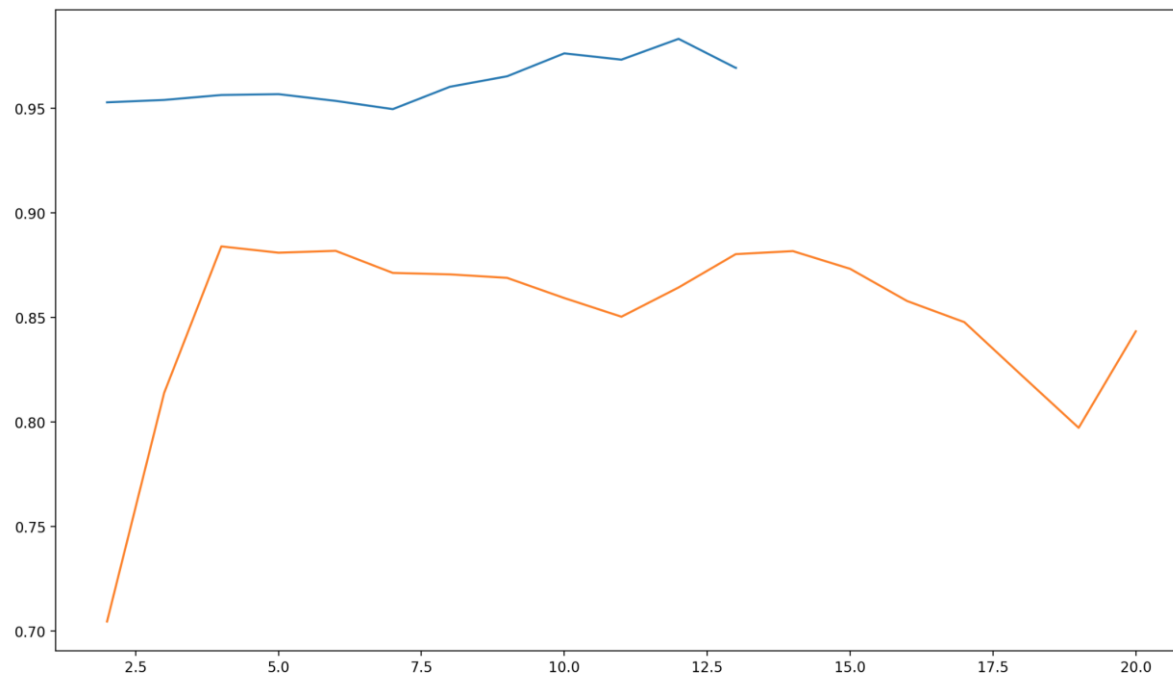
- ▶ Discretizzazione del tempo basata su timeslice di 60s
- ▶ Stato nascosto: attività svolta
- ▶ Osservazione: sensori attivi durante il timeslice
 - ▶ Facoltativamente il periodo della giornata che non è stato considerato a causa delle prestazioni minori
 - ▶ Possibilità di considerare le stanze relative ai sensori o la posizione dei sensori. Non considerato anche in questo caso a causa delle prestazioni.
 - ▶ Osservazioni più generiche

Parametri del modello

- ▶ Probabilità a priori:
 - ▶ Percentuale di occorrenza di ogni stato nel training set
- ▶ Probabilità di transizione:
 - ▶ Ogni stato del training set confrontato col successivo e calcolate le probabilità di transizione
 - ▶ Numero di volte in cui lo stato i è seguito dallo stato j rapportato al numero di occorrenze dello stato i
 - ▶ $P(i|i)$ particolarmente alta per alcuni stati a causa della suddivisione in timeslice di soli 60 secondi
- ▶ Probabilità di emissione:
 - ▶ Probabilità di osservare una configurazione di sensori per una certa attività
 - ▶ Calcolo analogo alla matrice di transizione

Scelta del Training set

- ▶ Divisione in giorni e non casuale o in percentuale causa della dipendenza temporale
- ▶ Dimensione del training set scelta: 5 giorni
- ▶ Accuracy in funzione del training set (in giorni):

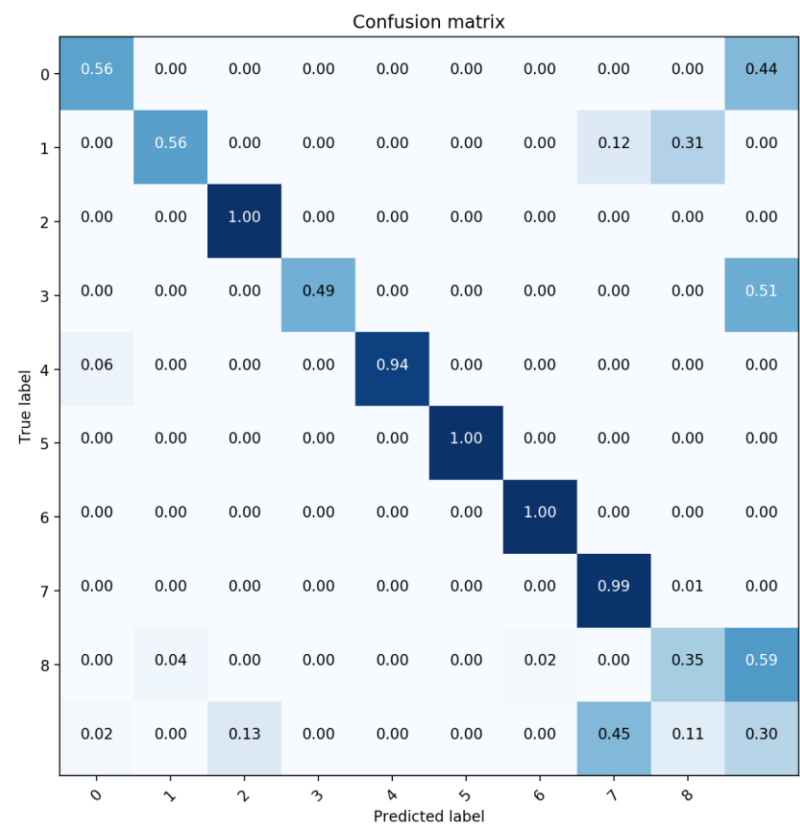


Predizioni

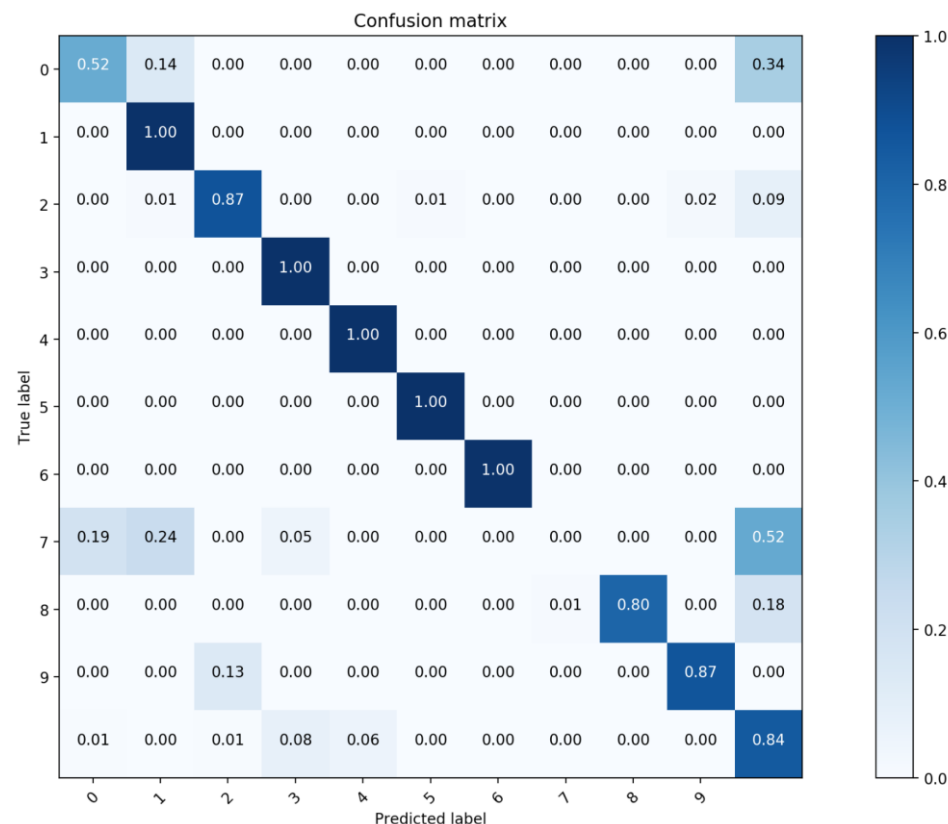
- ▶ Predizione a breve termine:
 - ▶ Training set: 5 giorni
 - ▶ Test set: 3 giorni consecutivi
- ▶ Predizione a lungo termine:
 - ▶ Training set: 5 giorni
 - ▶ Test set: 9 giorni nel caso del dataset A, 16 per il dataset B
- ▶ Predizione su sequenza casuale:
 - ▶ Training set: 5 giorni
 - ▶ Testset: 3000/20000 stati
 - ▶ Generata dalle distribuzioni di probabilità
 - ▶ Poco imprevedibile

Risultati - Previsione 3 giorni

► Accuracy dataset A: 94.9%



► Accuracy dataset B: 92.7%



Risultati - Previsione 3 giorni

► Accuracy dataset A: 94.9%

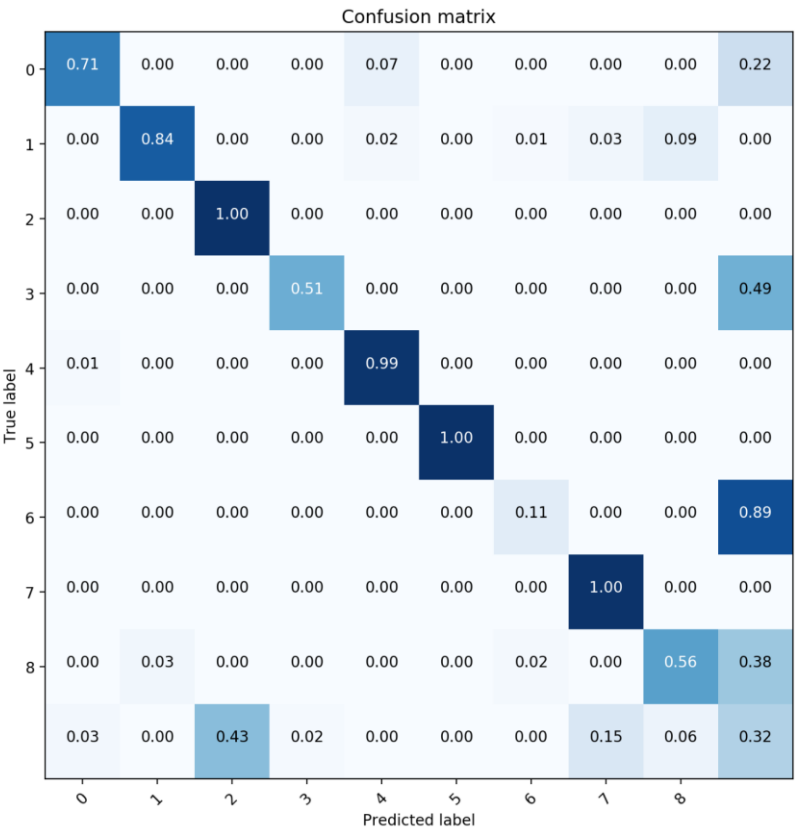
	precision	recall	f1-score	support
0	0.79	0.56	0.65	27
1	0.75	0.56	0.64	16
2	0.95	1.00	0.98	471
3	1.00	0.49	0.66	45
4	1.00	0.94	0.97	18
5	1.00	1.00	1.00	1783
6	0.25	1.00	0.40	1
7	0.95	0.99	0.97	1726
8	0.34	0.35	0.35	54
9	0.44	0.30	0.35	179
avg/total	0.94	0.95	0.94	4320

► Accuracy dataset B: 92.7%

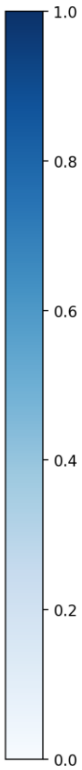
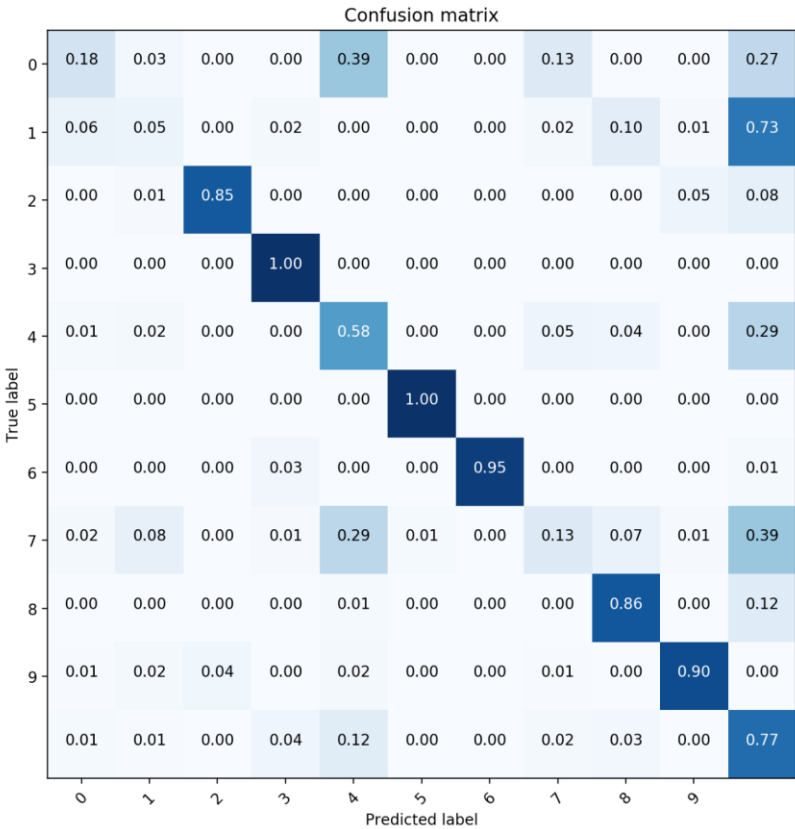
	precision	recall	f1-score	support
0	0.68	0.52	0.59	29
1	0.18	1.00	0.30	3
2	0.89	0.87	0.88	94
3	0.97	1.00	0.98	1219
4	0.60	1.00	0.75	45
5	0.94	1.00	0.97	15
6	1.00	1.00	1.00	1452
7	0.00	0.00	0.00	21
8	1.00	0.80	0.89	899
9	0.81	0.87	0.84	15
10	0.69	0.84	0.76	528
avg/total	0.94	0.93	0.93	4320

Risultati - Previsione 9/16 giorni

► Accuracy dataset A: 95.7%



► Accuracy dataset B: 89.1%



Risultati - Previsione 9/16 giorni

► Accuracy dataset A: 94.9%

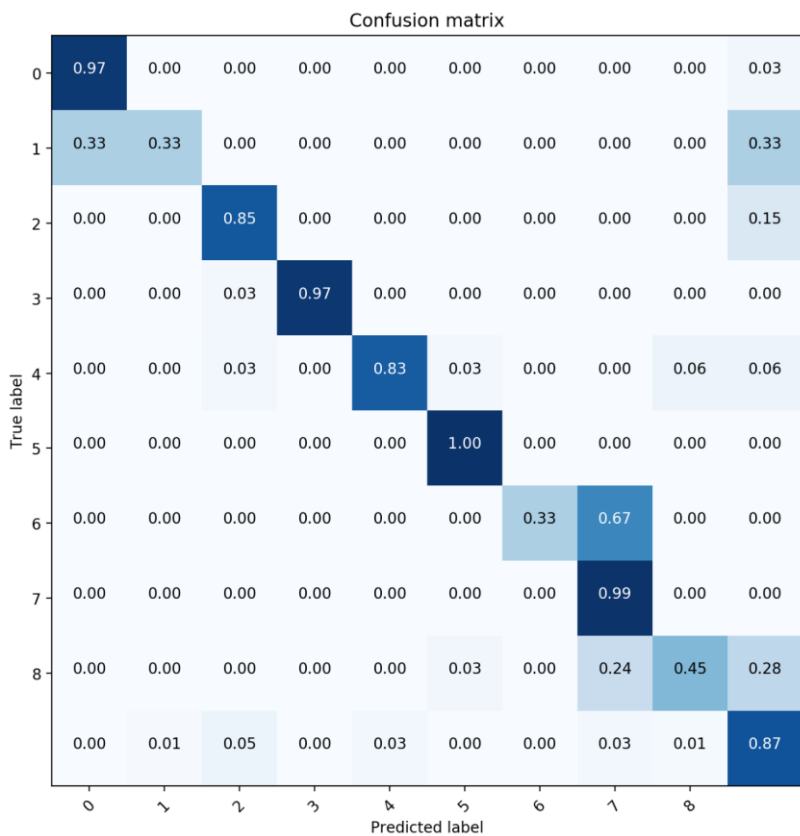
► Accuracy dataset B: 92.7%

	precision	recall	f1-score	support
0	0.80	0.71	0.75	85
1	0.94	0.84	0.88	87
2	0.87	1.00	0.93	1474
3	0.92	0.51	0.65	174
4	0.90	0.99	0.94	72
5	1.00	1.00	1.00	5170
6	0.14	0.11	0.12	9
7	0.98	1.00	0.99	5263
8	0.50	0.56	0.53	94
9	0.53	0.32	0.40	532
avg/total	0.95	0.96	0.95	12960

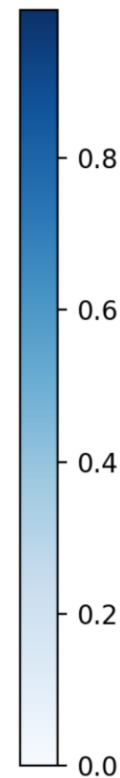
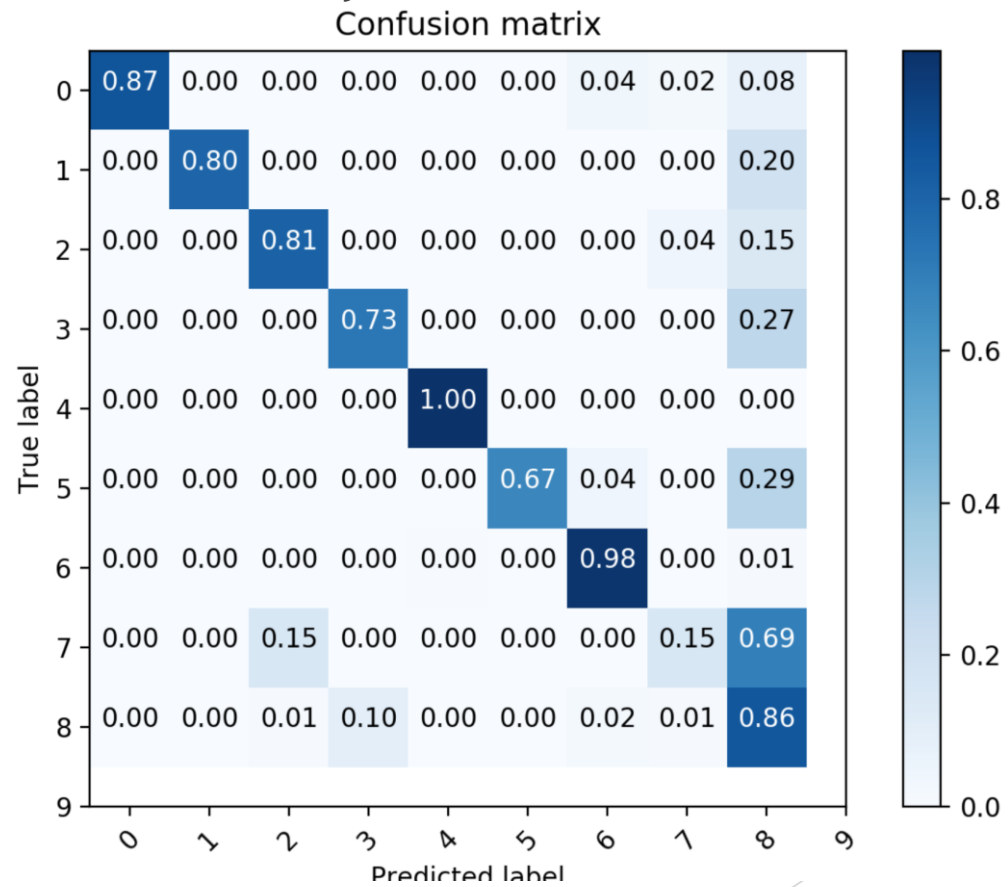
	precision	recall	f1-score	support
0	0.48	0.18	0.27	234
1	0.07	0.05	0.06	96
2	0.93	0.85	0.89	320
3	0.92	1.00	0.96	4417
4	0.19	0.58	0.28	211
5	0.93	1.00	0.96	63
6	1.00	0.95	0.97	8353
7	0.18	0.13	0.15	214
8	0.98	0.86	0.92	6514
9	0.77	0.90	0.83	129
10	0.61	0.77	0.68	2489
avg/total	0.91	0.89	0.90	23040

Risultati - Previsione casuale (3000 stati)

► Accuracy dataset A: 96.8%



► Accuracy dataset B: 91.4%



Risultati - Previsione casuale (3000 stati)

► Accuracy dataset A: 94.9%

► Accuracy dataset B: 91.4%

	precision	recall	f1-score	support
0	0.97	0.97	0.97	30
1	0.20	0.33	0.25	3
2	0.89	0.85	0.87	137
3	0.97	0.97	0.97	116
4	0.83	0.83	0.83	36
5	1.00	1.00	1.00	640
6	0.33	0.33	0.33	3
7	0.99	0.99	0.99	1790
8	0.62	0.45	0.52	29
9	0.83	0.87	0.85	216
avg/total	0.97	0.97	0.97	3000

	precision	recall	f1-score	support
0	1.00	0.87	0.93	52
1	0.80	0.80	0.80	5
2	0.76	0.81	0.79	48
3	0.82	0.73	0.77	393
6	1.00	1.00	1.00	1055
7	1.00	0.67	0.80	48
8	0.98	0.98	0.98	765
9	0.15	0.15	0.15	13
10	0.78	0.86	0.82	621
avg/total	0.92	0.91	0.91	3000

Conclusioni

- ▶ Il modello risulta particolarmente accurato, ma sembra mostrare dell'overfitting sui dati.
- ▶ Alcuni stati sono previsti molto meglio di altri, in particolare quelli per cui si hanno più osservazioni. Questo è dovuto al fatto che queste attività hanno una durata media maggiore rispetto alle altre attività e risulta quindi molto più facile prevedere lo stato successivo (ovvero, lo stato non cambia).
- ▶ Forte bias per lo stato di nessuna attività