1 Data

In the following, some details about the dataset.

The description of the main columns in the Table 1.

2 Embedding

Some basic information of our final dataset to remark is certainly the number of triples 1 558 700, that depends on the number of single instances reported in Tables 2, 3, 4.

3 KG prediction data

Table 5 reports all value for, all the 9 selected clusters of health services, among the initial 13 clusters.

Field-Name	Type	Explanation
booked-date	date	date of the booked appointment
reservation-date	date	date of the first contact with the booking staff, is also the date of the insertion of the record in the database.
last-reservation-change-date	date	last update of the reservation, which usually matches the reservation-date
encrypted-nin-id	str	encrypted national insurance number. Used as a non-decriptable numeric ID of the patient, special value -1 represents undefined patients
gender	long	1 for male, 2 for female, -1 undefined (for anonymous patients)
patient-age	long	age of the patient at the time of the issue of the referral, special value -1 is for anonymous patients
nuts-istat-code	long	Italian nuts code, it identifies local administrative areas, it can be linked to postal codes
booking-agent-id	long	anonimized numeric ID of the booking agent
medical-branch-id	str	alphanumeric string that contains a code of the medical branch
health-service-id	str	alphanumeric ID of the health service
practitioner-id	long	anonimized numerical identifier of the prescriber, special value -1 is for undefined practitioners
referral-centre-id	str	identification of the unit that delivers the appointment. It is the concatenation of the *dispenser* and the *ambulatory*, separated by the hyphen
appointment-encoding	str	concatenation of local-health-department-id and the appointment codification
booking-type	str	Numeric value: 1 is for direct check-in (e.g. from a hospital during a recovery), 0 is for the conventional booking reservation
referral-id	long	anonimized numeric ID of the referral
referral-date	date	the date of the issue of the referral from the practitioner to the patient
priority-code	str	the code of the priority of the referral
exemption-code	str	the code of the exemption with respect to the patient and the health service
number-of-health-services	long	number of needed health-service-ids, rarely greater than 1
status	str	status of the appointment: P: booked; A: canceled
local-health-department-id	long	sub-regional local health department. In our data set there are three departments, identified with A, B and C

Table 1: Data set features. The table contains the description of the main columns of the local health department booking center

@has-relations	Number of instances
@has-booked-date	149198
@has-reservation-date	122744
@has-number-of-health-services	122744
@has-appointment-encoding	122744
@has-booking-type	122744
@has-referral-id	31679
@has-referral-modified-id	31679
@has-patient-age	31679
@has-referral-date	30630
@has-exemption-code	21545
@has-priority-code	10958
@has-nuts-istat-code	3584
@has-local-health-department-id	2907
@has-practitioner-id	1991
@has-referral-centre-id	1837
@has-encrypted-nin-id	1747
@has-gender	1747
@has-refined-medical-branch-id	1337
@has-health-service-id	1316
@has-booking-agent-id	1070
@has-health-service-description	1023
@has-refined-health-service-id	1023
@has-official-branch-description	131
@has-medical-branch-id	131
@has-branch-description	131

Table 2: Number of instances in the @has type relations,, those a re the ones that express a property of either an entity or a relation (i.e. hyper-relation)

relation-entity links	Number of instances
reserved-health-service	122744
updating-agent	122744
booked-referral	122744
booking-agent	122744
referring-centre	84673
referrer	31679
prescribed-health-service	31679
referred-medical-branch	31679
referred-patient	31679
cure-provider	9537
cured-patient	9537
provided-health-service	9471
health-service-provider	9471

Table 3: Number of instances of relation kind.

entities and meta-relations	Number of instances
reservation	122744
referral	31679
health-care	9537
provision	9471
practitioner	1991
appointment-provider	1837
patient	1747
booking-staff	1070
medical-branch	1023

Table 4: Number of instances of entity kind.

lccccccccc									
Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Random Forest	Neural Network	Naive Bayes	QDA	XGBoost	Cluster
10	0.59 ± 0.024	**0.67 ± 0.027**	0.58 ± 0.025	**0.67 ± 0.03**	0.62 ± 0.024	0.64 ± 0.028	0.65 ± 0.027	0.63 ± 0.02	0.65 ± 0.022
5	0.51 ± 0.056	0.58 ± 0.065	0.54 ± 0.079	0.6 ± 0.093	0.57 ± 0.071	0.56 ± 0.075	**0.64 ± 0.082**	0.52 ± 0.098	0.5 ± 0.079
4	0.54 ± 0.037	0.58 ± 0.045	0.55 ± 0.042	**0.65 ± 0.045**	0.58 ± 0.046	0.6 ± 0.051	0.63 ± 0.035	0.51 ± 0.051	0.57 ± 0.041
2	0.51 ± 0.029	0.54 ± 0.028	0.52 ± 0.036	0.56 ± 0.037	0.53 ± 0.05	**0.59 ± 0.031**	0.56 ± 0.036	0.49 ± 0.048	0.57 ± 0.031
1	0.54 ± 0.031	0.6 ± 0.044	0.6 ± 0.034	0.59 ± 0.041	0.6 ± 0.037	**0.62 ± 0.044**	0.61 ± 0.036	0.5 ± 0.055	**0.62 ± 0.033**
7	0.59 ± 0.022	**0.67 ± 0.021**	0.57 ± 0.021	0.66 ± 0.025	0.62 ± 0.021	0.62 ± 0.024	0.65 ± 0.022	0.58 ± 0.022	0.62 ± 0.021
6	0.57 ± 0.031	0.69 ± 0.025	0.59 ± 0.028	0.71 ± 0.028	0.69 ± 0.027	0.67 ± 0.022	**0.72 ± 0.024**	0.54 ± 0.026	0.71 ± 0.026
11	0.52 ± 0.04	0.65 ± 0.049	0.62 ± 0.038	0.68 ± 0.068	0.68 ± 0.046	0.69 ± 0.055	**0.73 ± 0.052**	0.52 ± 0.044	0.69 ± 0.052
13	0.59 ± 0.02	**0.65 ± 0.026**	0.57 ± 0.025	0.64 ± 0.041	0.62 ± 0.027	0.64 ± 0.028	0.64 ± 0.025	0.62 ± 0.022	0.64 ± 0.021
MEAN	0.55 ± 0.034	0.63 ± 0.052	0.57 ± 0.031	0.64 ± 0.048	0.61 ± 0.051	0.63 ± 0.04	**0.65 ± 0.052**	0.55 ± 0.052	0.62 ± 0.065

Table 5: AUC values of every classification techniques, repeated 30 times, for every selected cluster written in the form of (mean ± std). In the last row the means values of every classification techniques. Bolded values are the best values for every row.