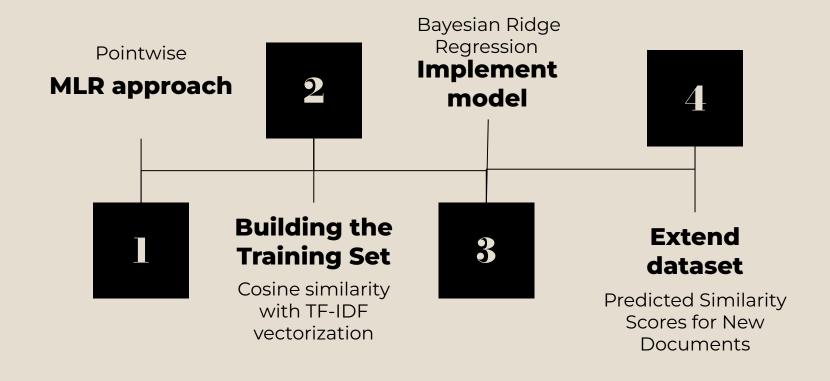
Lab ML Ranking Assignment

Irune Monreal Iraceburu Sergio Marín Sánchez Ander Ros Ollo

Table of contents



Pointwise MLR approach

f(query, document) — relevance score

Order results by relevance score

Building the Training Set





Cosine similarity



Building the Training Set

Dataset example with the similarity ranking for the query "glucose in blood"

loinc_num	c 😭 long_common_name 💠	⇔ component ÷	😭 system 💠	<pre> property</pre>	$\overline{123}$ similarity $ ightharpoonup$
10331-7	Rh [Type] in Blood	Rh	Bld	Туре	0.785288
1959-6	Bicarbonate [Moles/volume] in Blood	Bicarbonate	Bld	SCnc	0.657838
20565-8	Carbon dioxide, total [Moles/volume] in B	Carbon dioxide	Bld	SCnc	0.565994
890-4	Blood group antibody screen [Presence] in	Blood group antibody screen	Ser/Plas	ACnc	0.465409
2143-6	Cortisol [Mass/volume] in Serum or Plasma	Cortisol	Ser/Plas	MCnc	0.388614
2075-0	Chloride [Moles/volume] in Serum or Plasma	Chloride	Ser/Plas	SCnc	0.388614
1988-5	C reactive protein [Mass/volume] in Serum	C reactive protein	Ser/Plas	MCnc	0.371565
1975-2	Bilirubin.total [Mass/volume] in Serum or	Bilirubin	Ser/Plas	MCnc	0.362224
18998-5	Trimethoprim+Sulfamethoxazole [Susceptibi	Trimethoprim+Sulfamethoxazole	Isolate	Susc	0.000000
18906-8	Ciprofloxacin [Susceptibility]	Ciprofloxacin	Isolate	Susc	0.000000

Encode the dataset for the training phase



Tf-Idf Vectorizer →

sum(tfidf_array) / len(tfidf_array)

loinc_num	‡	123 long_common_name	\$	123 component	\$	½ system \$	123 property ÷
50407-6			0.009406	Θ.	. 009933	0.020833	0.045455
53125-1			0.008370	Θ.	. 009779	0.029463	0.045455
14423-8			0.008696	0.	. 005464	0.028092	0.045455
30403-0			0.007179	0.	. 005464	0.027718	0.045455
95226-7			0.010545	0.	. 009504	0.036084	0.045455
50672-5			0.009893	0.	. 009362	0.020833	0.045455
62239-9			0.009689	0.	. 005464	0.028043	0.045455
74354-2			0.007207	0.	.008681	0.020833	0.045455
76674-1			0.008745	Θ.	. 008324	0.020833	0.045455
56770-1			0.010048	Θ.	. 008406	0.020833	0.045455

Implement model

Bayesian Ridge

Example Results for

"Glucose in Blood"

Model Error: 0.21

loinc_num	y_test	y_pred
49926-9	0.47	0.59
15076-3	0.62	0.77
74774-1	0.58	0.36

Extend dataset (from LOINC)

loinc_num	long_com mon_name	component	system	property	similarity_ glucose_in _blood	similarity_ bilirubin_in _plasma	similarity _White_bl ood_cells _count
12345-6	Glucose in Blood	Glucose	Blood	Mass	0.63	-0.17	0.20
12346-7	Bilirubin in Plasma	Bilirubin	Plasma	Mass	0.63	-0.17	0.05
12347-8	White Blood Cells Count	Blood Cells	Blood	Count	0.64	-0.16	0.26

THANK YOU!