Ch03.08.Q. CT Intracranial Mass Effect – Natural Language Processing

Since the brain is enclosed by the skull, swelling due to bleeding, infection, or inflammation is dangerous. This is called intracranial mass effect (IME). According to the Mortality Probability Model (Ch. 6, Box 6.2), IME is an important predictor of mortality in ICU patients.

IME is diagnosed by CT scan of the brain, but the results of the CT appear in the electronic health record (EHR) as a free-text report dictated by the radiologist; there is no <yes/maybe/no> field dedicated to IME. Enter Natural Language Processing (NLP), which may be used to extract <yes/maybe/no> results from free text in the EHR.

In the CTIME Study [REF], trained reviewers assigned <yes/no> labels for IME to 1557 head CT reports on ICU-bound emergency department patients. An NLP algorithm called "**term frequency–inverse document frequency**" (tf-idf) was trained on 1202 and tested on 355 reports. The algorithm assigns a score between 0 and 1 to each report. The authors created 3 score ranges: yes, > 0.5; maybe, 0.05 to 0.5; no, < 0.05. The test-set (validation) performance of the algorithm is summarized in the following table:

NLP Category	Score Range	Average Score	CT IME+ by trained reviewer		CT IME-- by trained reviewer			NLP Yes/Maybe/No Category	
			N	%	N	%	LR	Total N	%
Yes	>0.50	0.886	85	78.7%	8	3.3%	24.20	93	26.2%
Maybe	0.05 - 0.50	0.204	19	17.6%	44	17.9%		63	17.7%
No	<0.05	0.008	4	3.7%	195	78.9%	0.05	199	56.1%
	Total		108	100.0%	247	100.0%		355	100.0%

    a. In the test set of CT reports, what proportion were labelled by the (human) reviewers as IME+?

    b. What is the LR for the "Maybe" result category?

    c. What proportion of reports categorized by the NLP algorithm as "Maybe" were IME+? We are asking for P(IME+|"Maybe").

We are willing to accept the NLP categorization for IME when it is either "Yes" or "No", but we must manually review all reports labelled "Maybe".

d. What proportion of reports require manual review? We are asking for P("Maybe").
17.7%.

e. Does this depend (significantly) on the prevalence of IME+ in the sample?

The table above is what we call an LR table. Here is the corresponding ROC table, partially completed.

	IME+	Sensitivity	IME-	1 - Specificity
>0.50			8	3.2%
>=0.05			52	21.1%
>=0	108	100.0%	247	100.0%

f. Finish completing the ROC table.

g. How many line segments in the resulting ROC curve?

There are 3 segments in the resulting ROC curve.

h. (Extra Credit) What is the AUROC corresponding to the table in (e)?