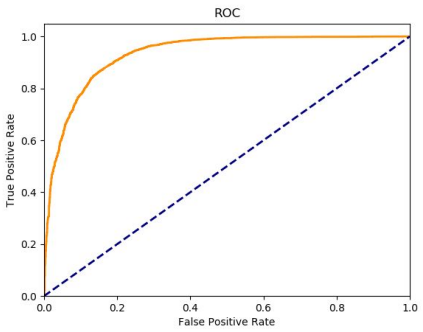
Identifying Non-Clinical Patient Messages Using Naive Bayes

Philip Ingram, MS*, Rom Srinivasan, MS*, Patrick Grennan, John Schrom, MPH
One Medical, San Francisco, CA

Introduction: Physician time and attention is an increasingly scarce resource. Primary care physicians spend nearly two hours in the electronic health record for every hour of patient care, with inbox management alone accounting for 85 minutes of their work day[1]. Further, physician availability via email has been shown to increase patient email volume by over 300%[2]. This creates a substantial burden on physicians, forcing them to choose between being accessible to patients, completing their administrative responsibilities, and maintaining their own time for personal wellbeing. This preliminary study attempts to assess the feasibility of identifying and rerouting non-clinical messages to administrative staff ("admins") in order to decrease the burden on physicians.

Methods: A subset of secure messages sent to physicians was collected between October 2017 and January 2018 from patients in a national primary care clinic system. The employee type (physician or admin) of the last employee to handle the message was recorded. Messages were pre-processed to remove stop words, and tf-idf was calculated. A 10% test set was held out for evaluation. A multinomial naive bayes classifier was trained with scikit-learn v0.19.1[3]. Hyperparameter tuning was conducted using grid search.

Results: A total of 69,316 messages were included: 41,630 (60%) were resolved by physicians and 27,686 (40%) by admins. Grid search revealed optimal performance when features with document frequencies over 0.125 were eliminated, and when using an additive smoothing parameter of 0.1. The final model had equal precision and recall of 0.87, and an AUC of 0.93.



Most Predictive Features			
Admin		**Physician**	
membership	new	prescription	work
insurance	received	week	come
cancel	com	think	results
message	office	time	help
email	year	blood	day
appointment	best	today	wondering
let	account	pain	help

Conclusions: A naive bayes classifier can be effective for distinguishing between clinical and non-clinical patient emails. This is a promising machine learning approach for reducing physicians' administrative burden.

References

1. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *Ann Fam Med.* 2017; 15(5): 419-426.
2. Wolcott V, Agarwal R, Nelson DA. Is Provider Secure Messaging Associated With Patient Messaging Behavior? Evidence From the US Army. Eysenbach G, ed. *J Med Internet Res*. 2017;19(4):e103.
3. Pedregosa F, Varoquaux G, Gramfort A, Michel V, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.

*These authors contributed equally to this work.