

Identifying Non-Clinical Patient Messages Using Naive Bayes

Philip Ingram, MS*, Rom Srinivasan, MS*, Patrick Grennan, John Schrom, MPH

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 day¹. 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.

Natural language processing is increasingly being used in medicine and public health; however, the focus tends to be on identifying unstructured data elements (e.g., race, gender, social determinants of health), identifying diseases, conditions, or events that are not otherwise structured (e.g., adverse drug events), or selecting optimal therapies for complex conditions. Proportionately little research has gone into minimizing the administrative burden clinicians face from the increased access patients have to their clinicians.

This preliminary study attempts to assess the feasibility of identifying and rerouting non-clinical messages to administrative staff in order to decrease the burden on physicians.

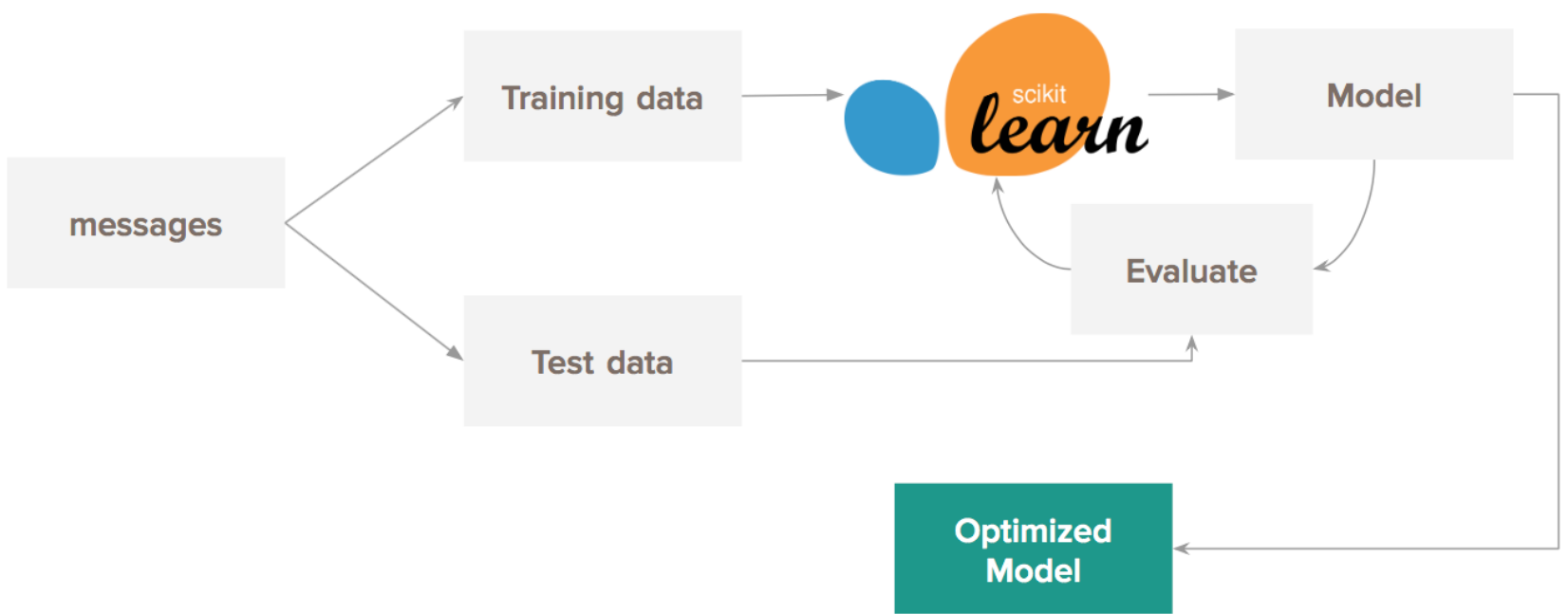
Methods

Data Collection & Processing

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.

Model Selection & Tuning

A 10% test set was held out for evaluation. Multiple models were trained, including Naive Bayes (NB), stochastic gradient descent (SGD), support vector machines with linear (SVM-linear) and radial basis function (SVM-RBF) kernels, and random forest (RF). Models were trained using scikit-learn v0.19.1³. Hyperparameter tuning was conducted using grid search.

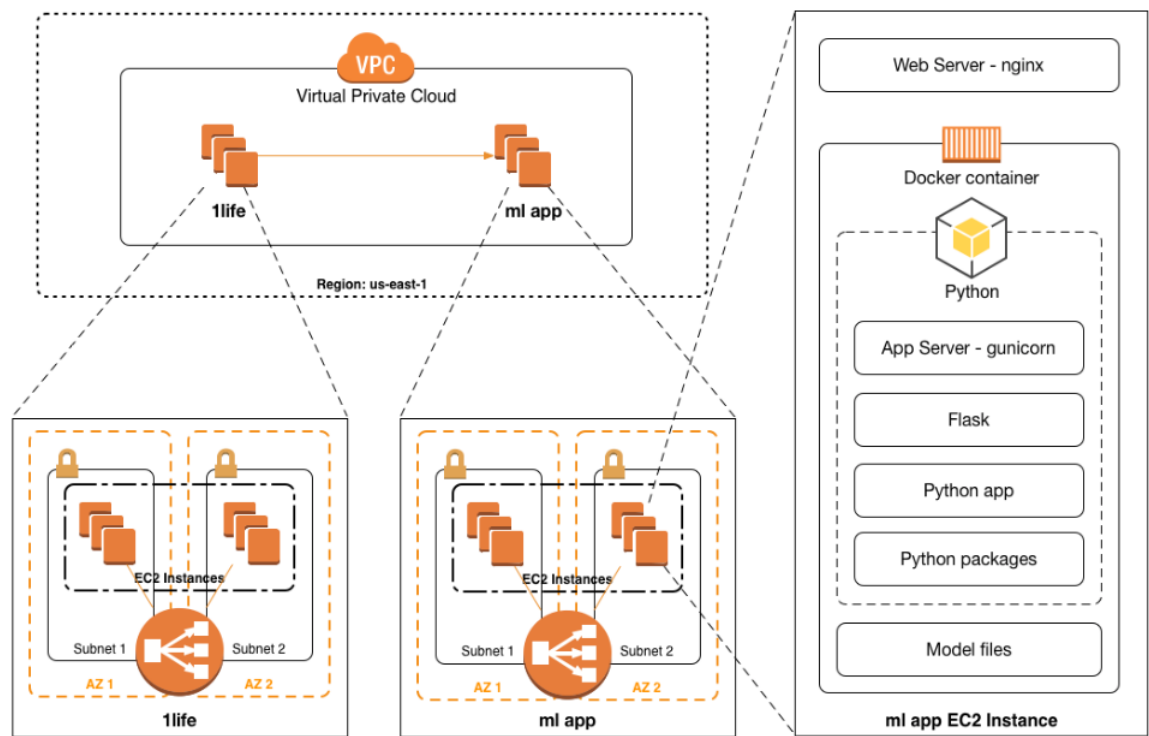


Evaluation

Each model was evaluated on the test set for accuracy, precision, recall, and ROC/AUC. The final model was selected on the basis of AUC and ease of implementation and deployment (i.e., size and calculation speed of the model).

Implementation

The selected model was deployed as a micro-service with Flask (python-based web server) and scikit-learn. Incoming messages were flighted through the micro-service and routed based on the model's result.

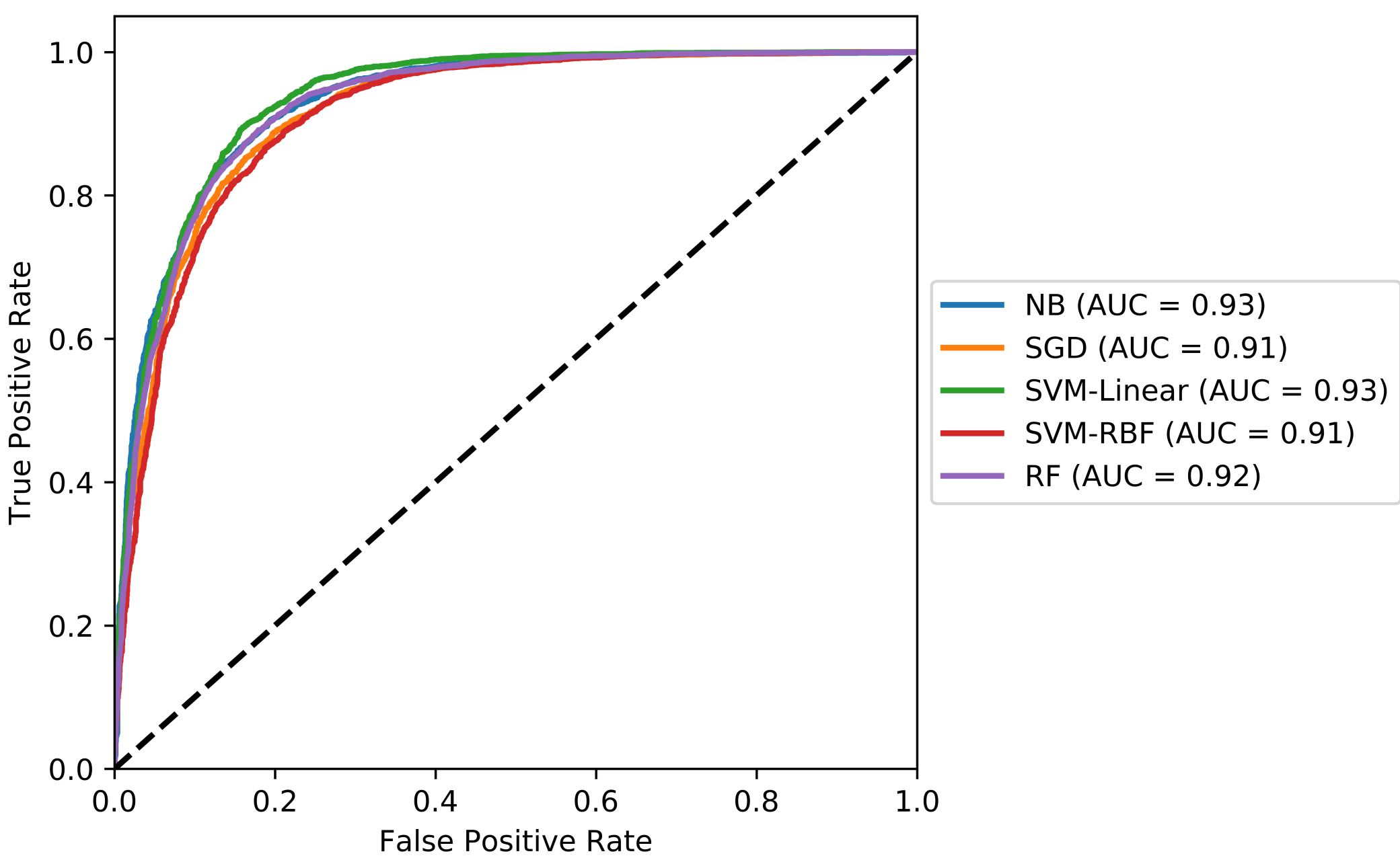


Results

Model Evaluation

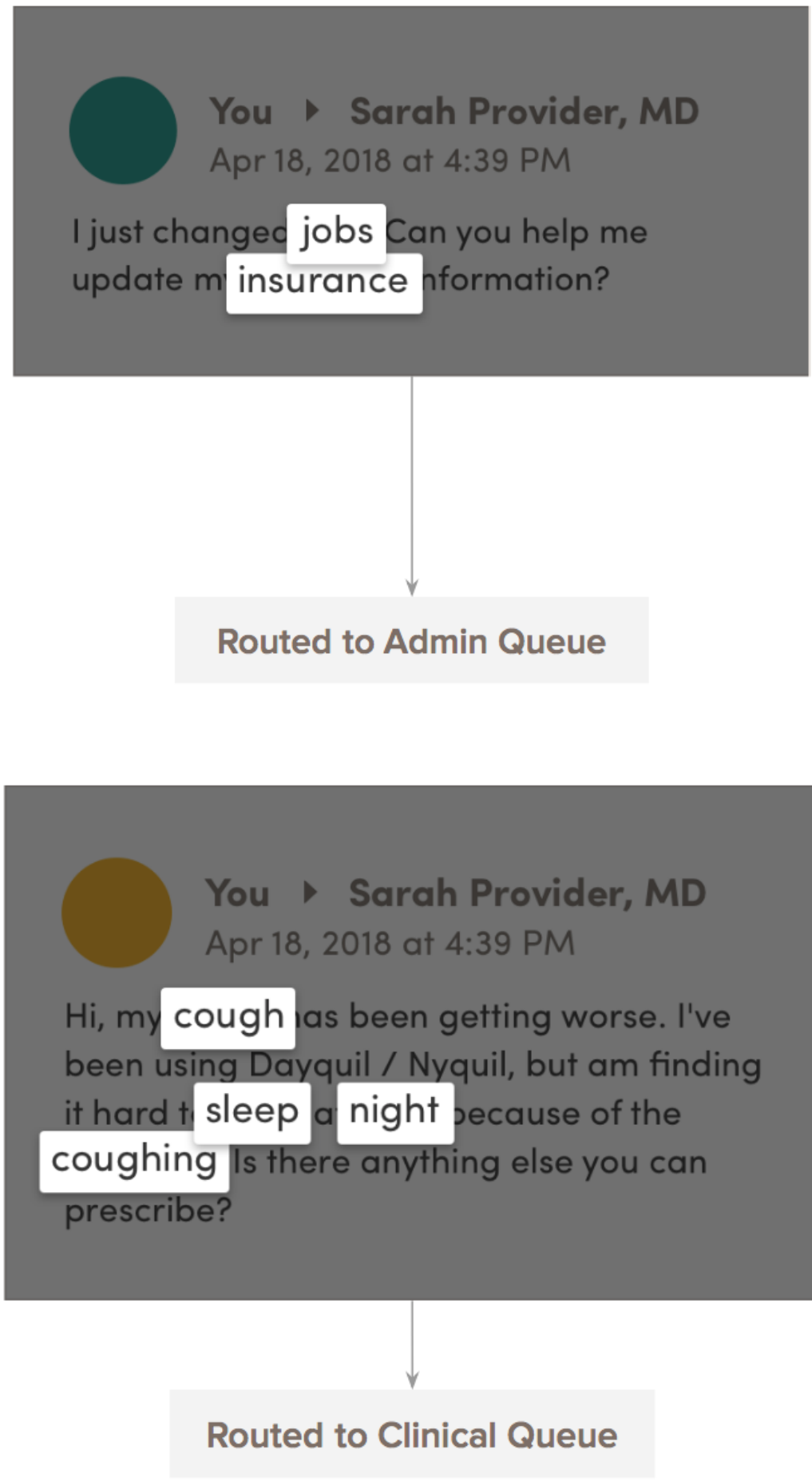
A total of 69,316 messages were included: 41,630 (60%) were resolved by physicians and 27,686 (40%) by admins. For Naive Bayes, grid search revealed optimal performance when features with document frequencies over 0.125 were eliminated, and when using an additive smoothing parameter of 0.1. The final model had equal precision and recall of 0.87, and an AUC of 0.93. In implementation, this model reroutes 9% of incoming messages to non-clinical staff for resolution.

	Accuracy	Precision	Recall
NB	0.8635	0.86	0.86
SGD	0.8465	0.86	0.85
SVM-Linear	0.8747	0.88	0.87
SVM-RBF	0.5999	0.36	0.60
RF	0.8638	0.87	0.86



Model Interpretation

Individual feature importances were identified for the NB model. Empirically, words related to administrative functions (e.g., scheduling, insurance) were predictive of non-clinical messages, while words related to questions (e.g., wondering, think), and clinical concerns (e.g., blood, pain, prescription) were predictive of clinical messages.



Clinical	Non-Clinical
prescription	membership
week	insurance
wanted	cancel
think	message
time	email
blood	appointment
today	let
work	new
pain	received
results	com
come	office
help	year
day	best
wondering	account
test	fee

Conclusions

Incoming patient emails can be effectively distinguished as clinical and non-clinical, and routed accordingly. While every model tested performed well, a Naive Bayes classifier was performant and efficient for production implementation. 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.