

Advancing Holistic Review During Residency Application: Using Natural Language Processing to Applicants' Experiences To Predict an Interview Invitation

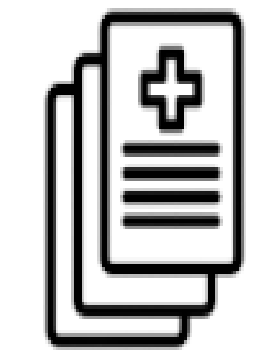
Arun Umesh Mahtani, MD, MS^{1,2}, Ilan Reinstein, MS¹, Marina Marin, MS¹, Jesse Burk-Rafel, MD, MRes¹

Email: arun.mahtani@nyu.edu

¹ NYU Grossman School of Medicine

² Richmond University Medical Center

DESIGN



6500 ERAS applications from 2017-2019



Extracted experiences from each AAMC ID

spaCy

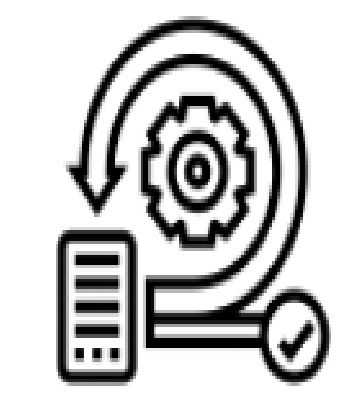
Experiences preprocessed using spaCy's en_core_web_lg pipeline

TFIDF

Word importance via TFIDF



Filtered words were split into train and test set

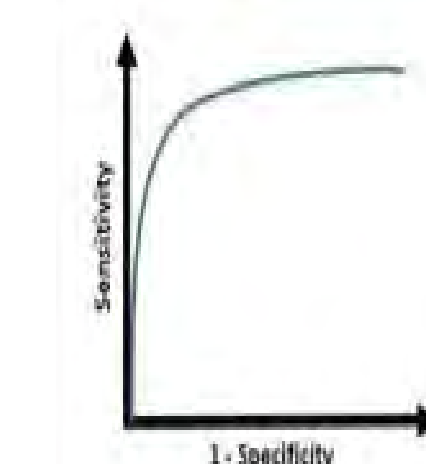


Logistic regression model using NLP & NLP + structured data



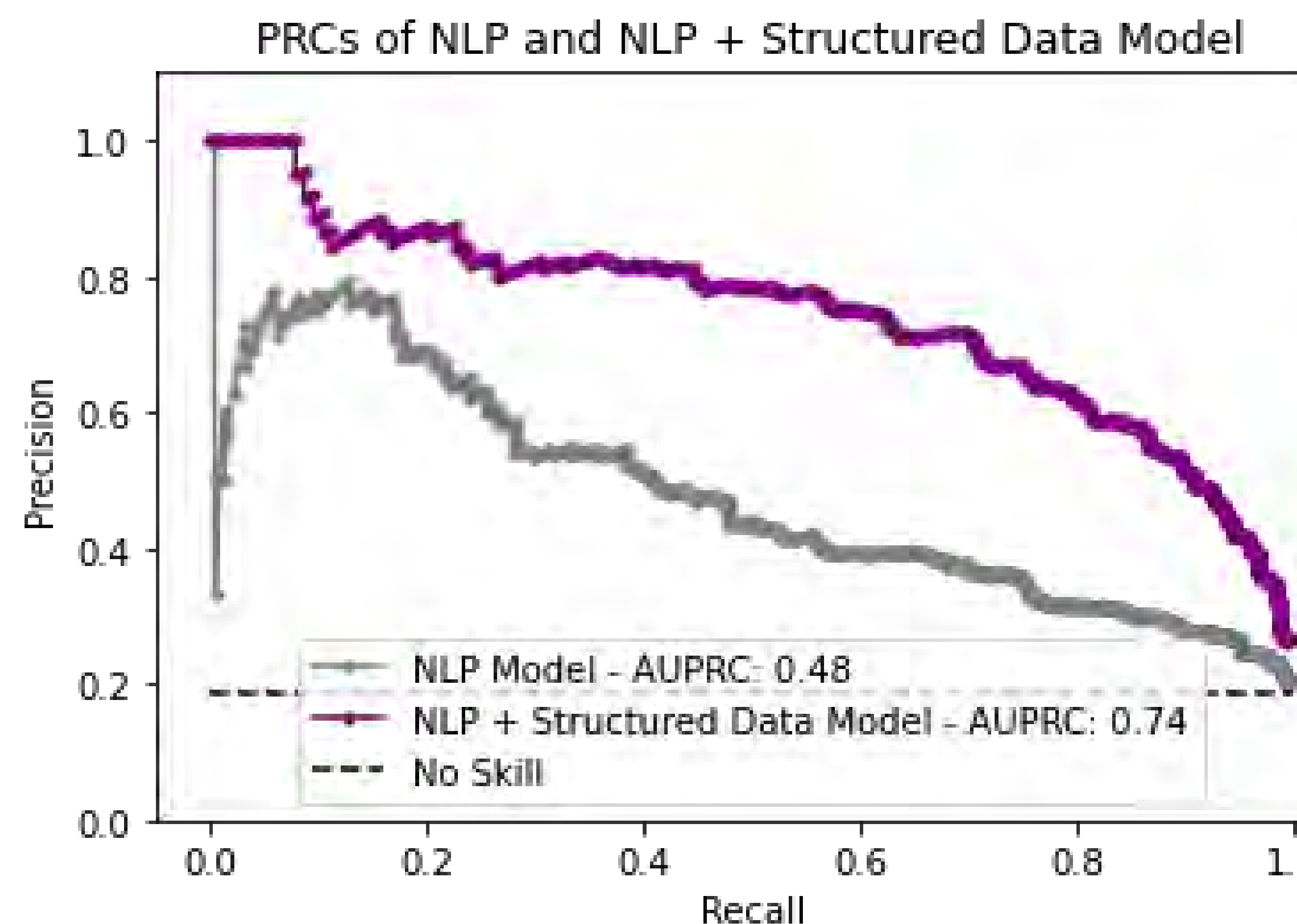
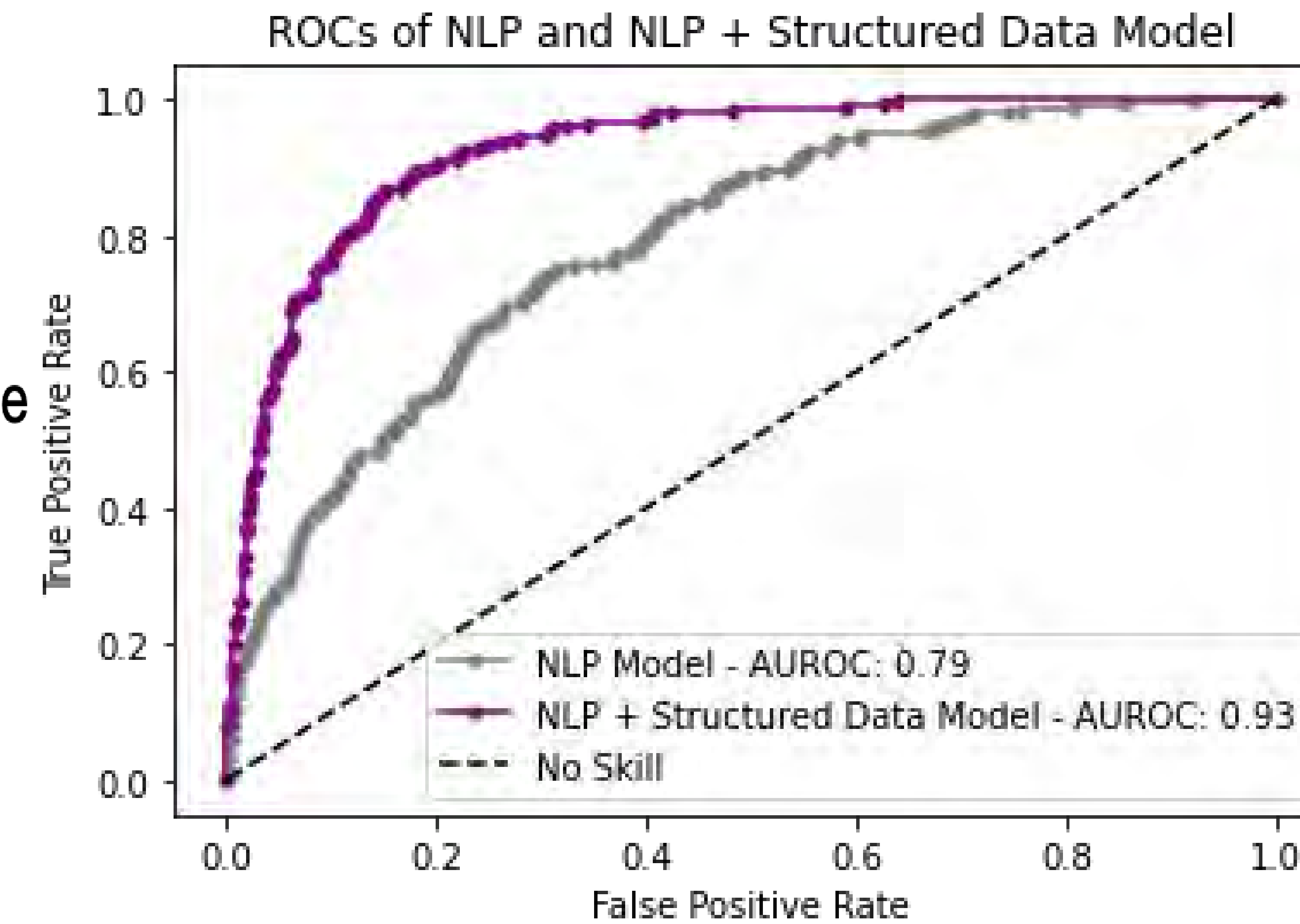
Features selected with nonzero coefficients.

Each applicant was assigned a predicted probability score



Model performance assessed

RESULTS



CONCLUSIONS

- NLP model showed moderate predictive strength
- Phrases indicating active leadership roles, research, or work involving social justice and health disparities were associated with an interview invitation
- The NLP score representing an individual's narrative entries was an important predictor, however it did not improve overall predictive performance compared to structured data alone