



Clinical Trials Prediction

By Jeffery Rosario



Problem Statement

Is the clinical trial successful?

~40 - 80 million

Business Value

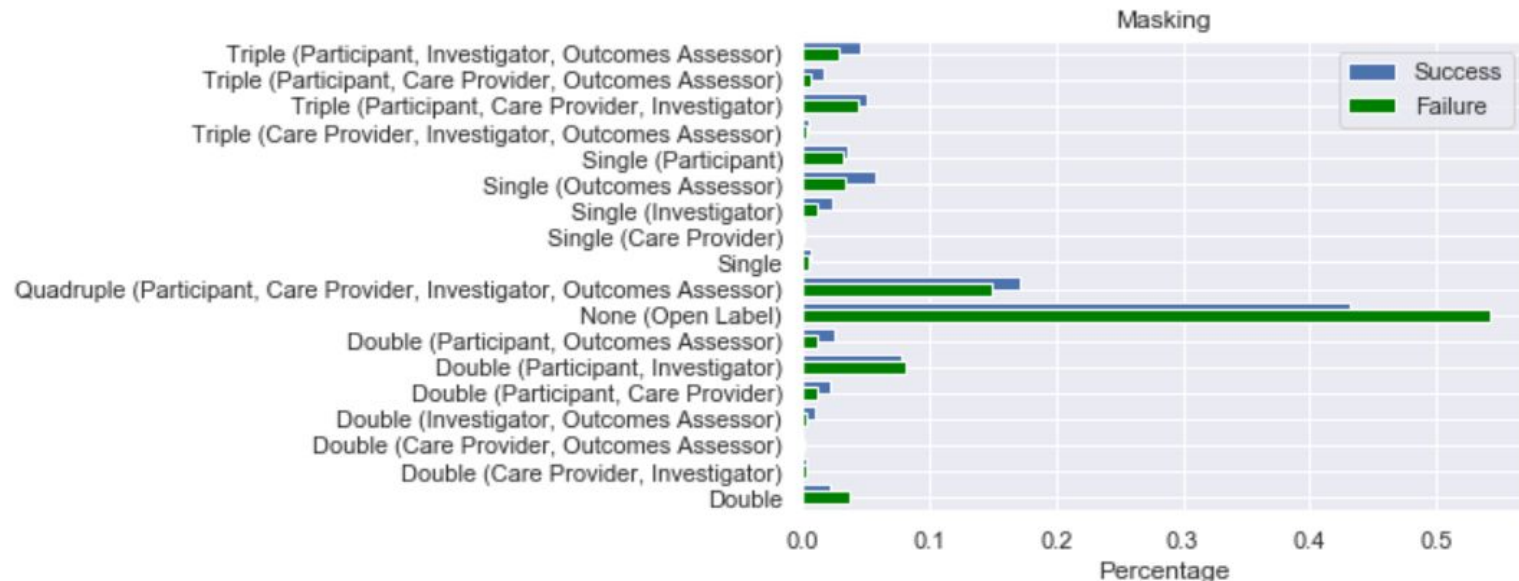
- Tool for healthcare researchers to determine whether a clinical trial will succeed
- Help determine changes needed for a successful predicted outcome



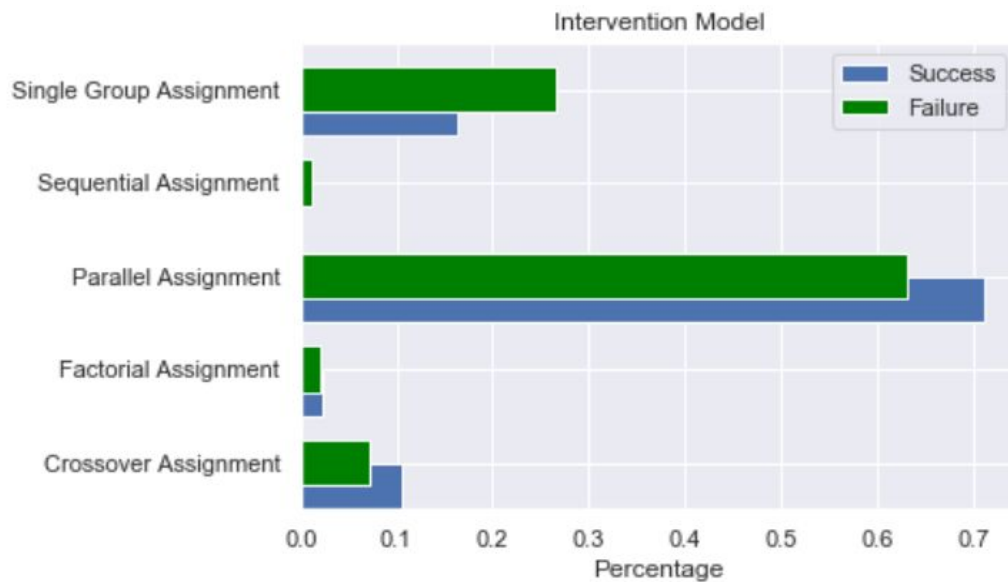
Methodology

- 5395 Clinical Trials - clinicaltrials.gov
 - minimum/maximum age, intervention type, intervention name, etc.
- ~30000 PMID reference papers
- Logistic Regression
- Cross Validation

Masking



Intervention Model



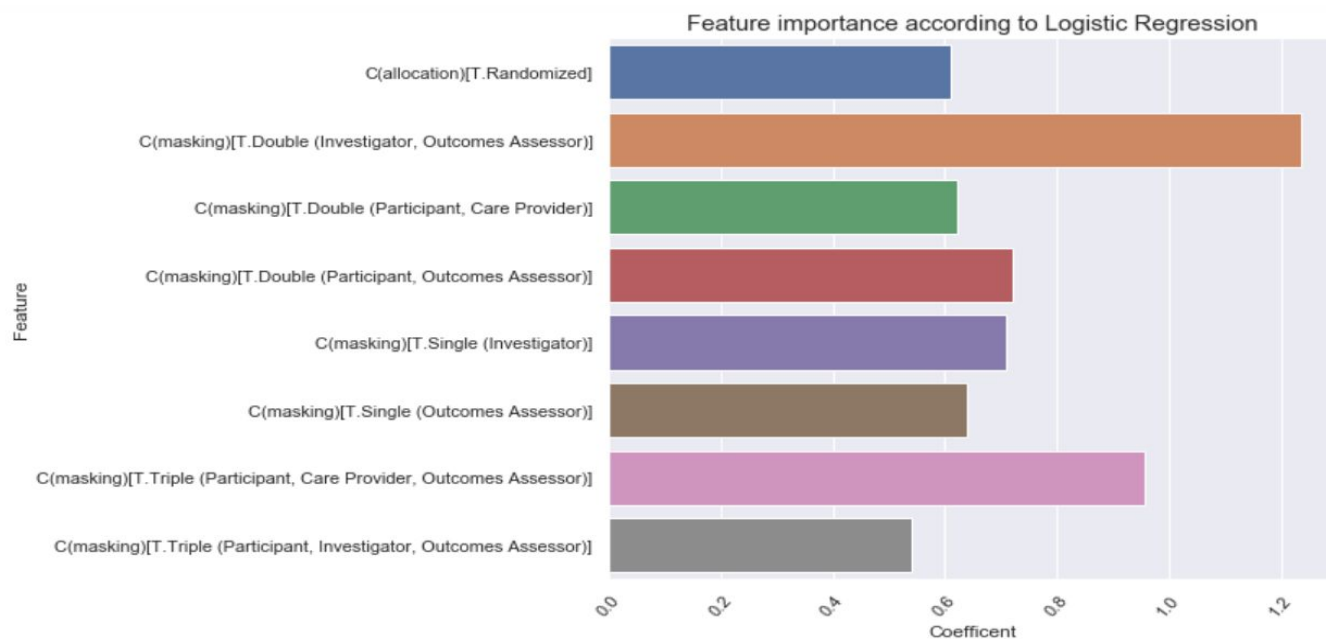




Success vs failure

	minimum_age	maximum_age	num_of_papers	source_len	condition_len	textblock_len	intervention_type_len	inter_name_len	pmid_len
success									
False	17.975237	31.844301	8.972524	27.152985	17.372456	2808.502714	11.411805	33.401628	41668.910448
True	19.487988	33.034535	10.154655	26.780781	17.602102	2529.741742	10.235736	33.695195	41469.408408

Feature Importance





Classification Report

	Precision	Recall	F1-Score	Support
Success - w/out Text	0.55	0.25	0.34	334
Success	0.66(+0.11)	0.36(+0.11)	0.47(+0.13)	334
Failure - w/out Text	0.85	0.95	0.90	1447
Failure	0.87(+0.02)	0.96(+0.01)	0.91(+0.01)	1447



Findings and Recommendations

- Including text vectorizers modestly improved overall model accuracy and precision
- Machine learning models can predict clinical trials outcomes
- Focus on experiment design:
 - masking(double/triple)
 - Allocation(randomized)



Future Work

- Bigger Dataset, institutional access to data
- Search a bigger breadth of hyperparameters for machine learning models



Thank You!



Appendix

Data Imbalance

- Imbalance of target variable
- 1:4 ratio
- Account for imbalance using SMOTE

