Emerge MA: Surveying Opportunities for Women to Run and Win Elections

Sarah Ferry, Neela Kaushik, Alex O'Connor

Abstract

- Emerge Massachusetts is a nonprofit organization whose mission is to get more female Democrats elected into public office
- Project goal is to isolate the factors that make female Democratic political candidates in Massachusetts most likely to win a seat in office
- Focus on analyzing districts demographic composition and voting behavior to help Emerge MA strategically prioritize candidate recruitment

Data Sources & Collection

VoteBuilder:

- Provides summary statistics for voter data in Republican districts. Specifically voter participation in the past four elections, race, sex and party

PD43+:

- Provides election results since 2009 in all districts in MA, including current and past incumbents name, sex, party, vote count, and vote percent

OCPF:

- Provides political contribution data including donation amount and donor district since 2009, specifically contributions that are over $500

US Census data:

- Provides demographic information at all district levels for each year since 2010. Use key features of interest: political affiliation, gender, race and income

Approach

Historical Classification Model:

- Used logistic regression, KNN, support vector machines, and a decision tree classifier to determine which would give us the best accuracy

Present Classification Model:

- Logistic regression model with recursive feature elimination and leave-one-out selection to extract the most statistically significant features

Results

- Accuracy scores indicate the percent of districts for which the model correctly identifies a female democratic winner

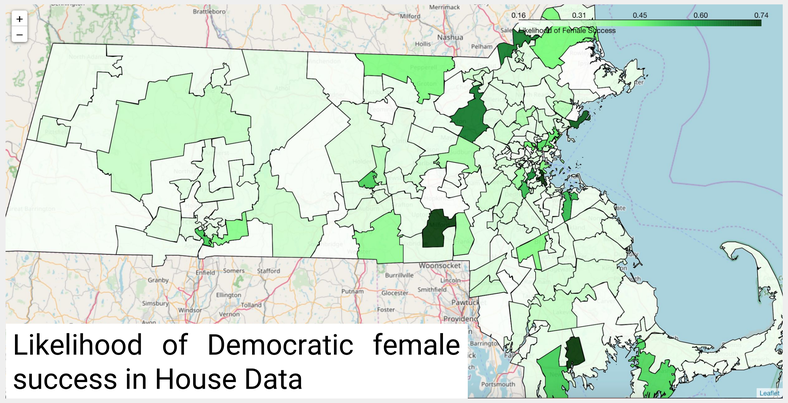
Historical Classification Model:

- SVM classifier demonstrates best performance when senate and house districts are trained and tested on their respective datasets

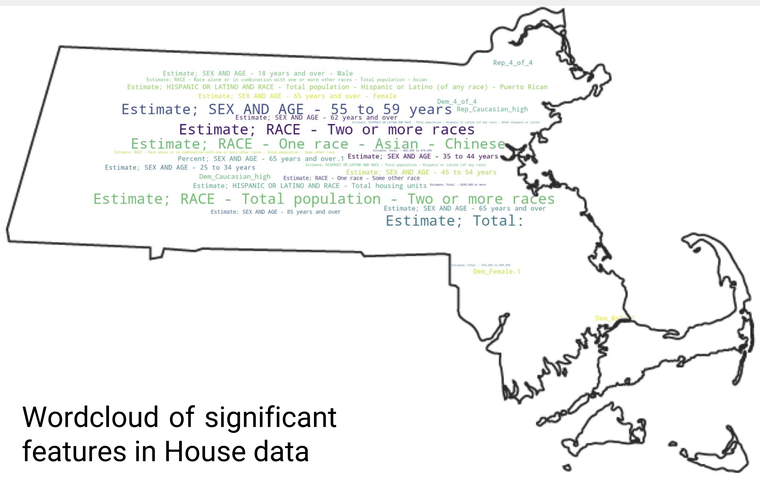
Present Classification Model:

- Goodness-of-fit test determined the significance (where $p < .05$) of all features in the models and output features

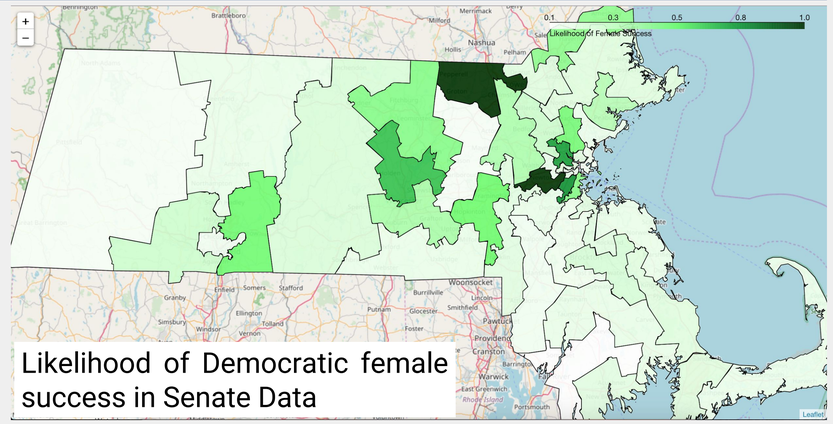
			Test on Senate		Test on House	
Model	Training Data	Method	Fem.	Fem. Dem.	Fem.	Fem. Dem.
Historical	Senate Districts	SVM	0.7442	0.7674	N/A	N/A
	House Districts	SVM	N/A	N/A	0.7048	0.8253
Present	Senate Districts	LR	N/A	0.7778	N/A	N/A
	House Districts	LR	N/A	N/A	N/A	0.9333



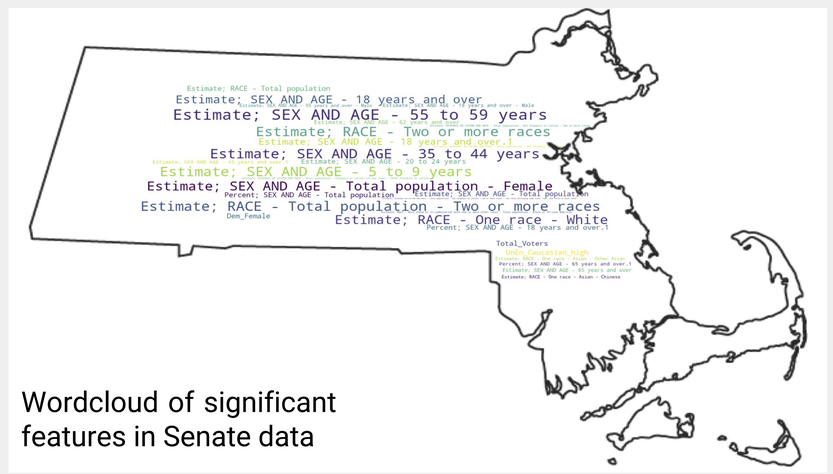
Likelihood of Democratic female success in House Data



Wordcloud of significant features in House data



Likelihood of Democratic female success in Senate Data



Wordcloud of significant features in Senate data

Conclusion

- Models predict probable success for female Democratic candidates to a reasonable level of accuracy in all districts
- Historical model is better for predicting success according to trends from previous years as well as current data
- Present model can identify districts with features that may contribute to a female Democratic candidate's success
- Future work and improvement involves more data collection for years ranging further back than 2009

Github

https://github.com/ferrys/cs-506-emerge-MA