

---

---

# Looking for Gendered Differences in Instructor Evaluation Language

— Samira Bazuzi, Sylvan Tsai, and —  
Annie Shoup

---

---

# Problem

- **Instructors receive feedback** on their course content and teaching methods **from students** who fill out subject evaluation forms
- Studies have indicated a gender bias in course evaluations: **male and female professors** seem to be **evaluated differently**
- Can we **predict the gender of the professor** being evaluated **using learned features** from **student text responses**?

# Data

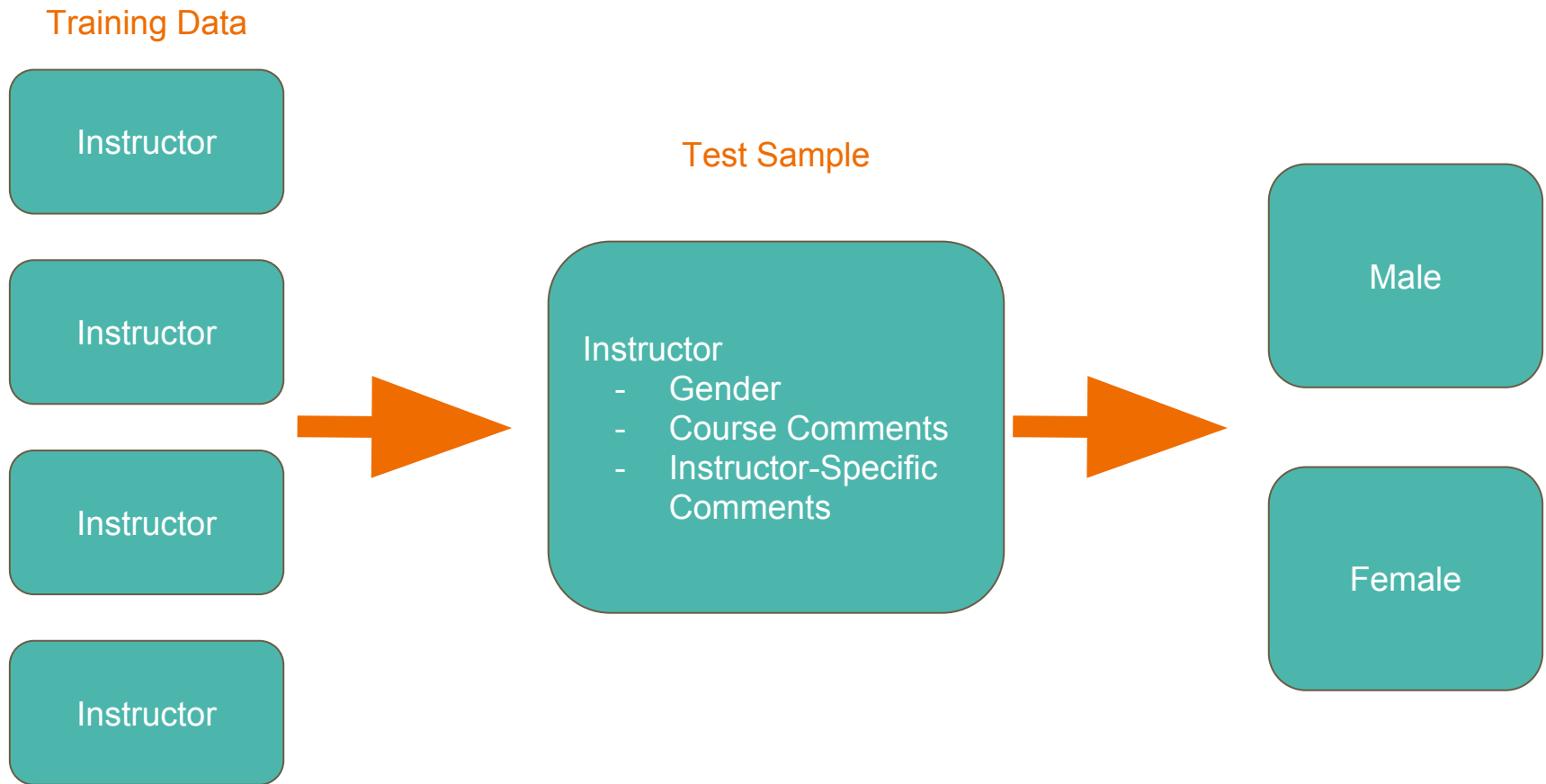
## Alpha Chi Omega Sorority (AXO)

- HASS class evaluations (2006-2014), written by Alpha Chi Omega members
- 152 instructor evaluations (88 males, 64 females)
- Average of 1.7 reviews per course, with a standard deviation of 1.011, a maximum of 6 reviews, and a minimum of 1.

## HKN Underground Guide

- HKN Underground Guide student evaluations from 2006 for 6 classes
- 29 instructor evaluations (26 males, 3 females)
- An average of 32.5 reviews per a course, with a standard deviation of 49, a maximum of 239 reviews, and a minimum of 5.

# Input/Output



# Baselines

	AXO	HKN Underground Guide
Classifier that guesses randomly according to the distribution of gender in the training set	0.4950	0.7922
Consistent assignment of the majority gender in the training set	0.4750	0.8571

# Results

	AXO	HKN Underground Guide
SVM	0.5937	0.8295
Logistic Regression	0.5573	0.8682
KNN	0.5413	0.8372
Topic Modeling Features w/SVM	0.5776	0.8295
Topic Modeling Features w/Logistic Regression	0.5764	0.8287
Topic Modeling Features w/KNN	0.5244	0.8295

# Heavily Weighted Features - AX0

Male	Female
Knowledgeable	Super
Awesome	Amazing
Cool	Gender
Fun	Energetic
Flexible	Helpful
Good	Simple
Easy	Nice
Weird	Enjoyable
Hard	Friendly

# Heavily Weighted Features - HKN

Male	Female
Clear	Suggestions
Good	Prompt
Helpful	Informed
Interesting	Learning
Available	Replying
Liked	Prepared
Understand	Responded
Friendly	Comfortable
Worth	Feedback



# Conclusion

We were not able to effectively classify the gender of the instructor based upon language in the comments at a significantly better rate than our baselines. We saw only small improvements over the baselines for some tests.

- We believe that all-female authorship and public reviewer attribution of the AXO comments may have contributed to the lack of discernible bias.
- We were not allowed access to very much data for our HKN Underground Guide dataset, and very few female instructors were present in the limited sample, a feature which is exacerbated by the gender imbalance in EECS faculty in general.

# Future Work

- Is there a difference in results depending on gender of the reviewer and intended audiences?
- Would results vary for evaluations of different departments?
- Larger dataset, more diversity in reviewers/instructors
- Comparisons of attributed and anonymized results