Thesis Defense

Computer Science Master's Program

"Predicting Newsworthiness of California Bill Discussions"

By **Venkata Kedarnath Chaturvedula**

Abstract:

Among various California bill hearings, it is crucial to identify which hearings are most likely to receive media coverage, as this highlights the importance of the bill and the hearing, while also underscoring its societal impact. Digital Democracy, established at Cal Poly's Institute for Advanced Technology and Public Policy, provides users access to a distinctive dataset and information about state legislative committee hearings, hearing transcripts, assets, bill information, etc.

This thesis presents a machine learning framework for predicting whether California legislative bills are likely to receive media coverage. Drawing on a dataset of legislative "tip sheets" provided by Digital Democracy and media coverage records from CalMatters, we developed a semantic similarity-based mapping engine to identify relevant bill-article pairs, achieving 93% precision through manual validation. This approach expanded our dataset from 424 manually verified pairs to 1,824 samples (3.3× increase), enabling more robust model training.

We evaluated various machine learning algorithms. Our best-performing model, XGBoost, achieved 89.91% accuracy, 90.81% precision, and 89.90% F1-score. Feature importance analysis revealed that temporal features dominate predictions, accounting for 62.3% of total importance.

The results demonstrate that machine learning models can effectively predict media attention for legislative bills, providing a practical foundation for media resource allocation, legislative transparency, and predictive analysis in political journalism and public affairs.

Date: Monday, December 8th, 2025

Time: 2:00 PM – 4:00 PM
ZOOM: https://calpoly.zoom.us/my/foaad
Committee: Dr. Khosmood, Dr. Kurfess, and Dr. Bisheff