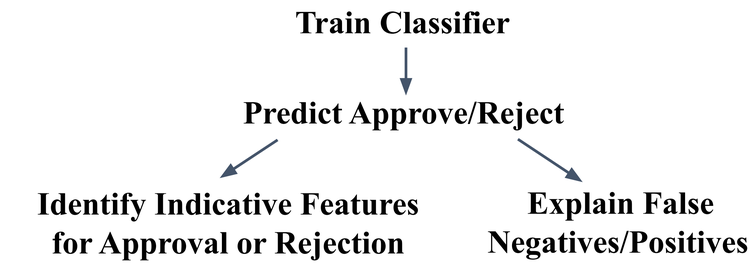
Identifying Discrepancies in the California Coastal Commission Permitting Process Using Machine Learning

Winton Yee, *winton@stanford.edu* **Samuel Kwong**, *samkwong@stanford.edu*

STANFORD UNIVERSITY
THE BILL LANE CENTER FOR THE AMERICAN WEST

Abstract

The California coastline, over 800 miles of land, requires governing in order to preserve its splendor. One factor that shapes the environmental outcomes of the coastline is the permitting process of the California Coastal Commission. Our goal was to detect discrepancies in the process by which the CCC approves permits governing the California coastline. To aid our research in analyzing decades of data from archived meeting agendas, we used machine learning to create natural language processing classifiers to predict permit approvals, and, ultimately, identify what makes an application suitable for approval or rejection.



However, even small discrepancies in how the CCC makes its decisions can affect the landscape of the coastline for many years to come. Thus, it is important to understand how these discrepancies may arise.

Data

We used a dataset from the California Coastal Act and a preprocessed dataset that includes individual permit applications and which sections from the California Coastal Act were mentioned in the application.

ARTICLE 5
LAND RESOURCES

Section	
30240	Environmentally sensitive habitat areas; adjacent developments
30241	Prime agricultural land; maintenance in agricultural production
30241.5	Agricultural lands; determination of viability of uses; economic feasibility evaluation
30242	Lands suitable for agricultural use; conversion
30243	Productivity of soils and timberlands; conversions
30244	Archaeological or paleontological resources

ARTICLE 6
DEVELOPMENT

Section	
30250	Location, existing developed areas
30251	Scenic and visual qualities
30252	Maintenance and enhancement of public areas
30253	Minimization of adverse impacts
30254	Public works facilities
30254.5	Terms or conditions on sewage treatment plant development; prohibition
30255	Priority of coastal-dependent developments

Fig 1. California Coastal Act

Approach

Our linear algorithm takes in a set of permits features, and returns a prediction of acceptance or rejection. We employed three different feature sets: the words in the permits, the sections of the California Coastal Act referenced by each permit, and a "clustering" score.

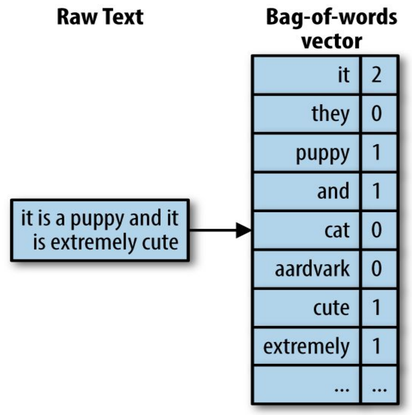


Fig 2. Bag of words

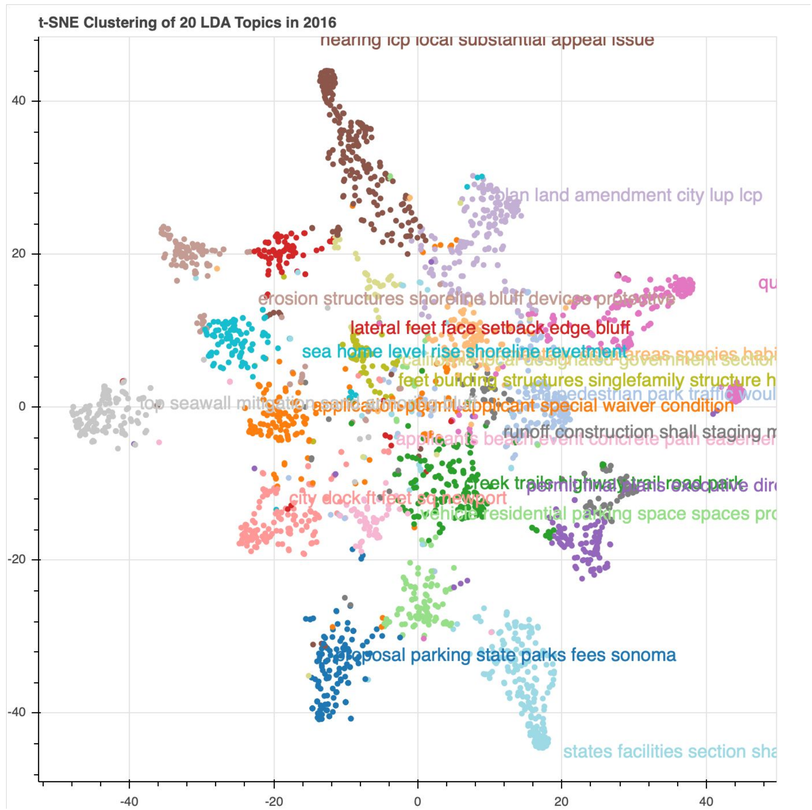


Fig 3. Clustering of topics

We also designed a more complex neural network, which took in words as features, to predict the outcomes of permits. Below are the accuracies of each of the different models:

Results & Analysis

Model	Accuracy
Neural network	84.2%
Linear, words	82.8%
Linear, sections	74.3%
Baseline	71.8%
Linear, clusters	70.7%

Every permit references specific **sections** within the California Coastal Act that specify where/what the permit is requesting to modify.

- We observed variation in permit approval across different **regions**, and found that certain sections indicating where the permit was submitted correlated strongly with *acceptance* or *rejection.*
- We also found that sections related to **natural improvement** were strongly correlated with *acceptance*, while a section related to **industrial expansion** was strongly correlated with *rejection*.

In predicting permits, we were able to see the *false positives* of our model (permits we predicted would be *accepted* but were actually *rejected*). These *false positives* often had to do with **development on natural land**, suggesting the CCC is especially sensitive to this topic.

References

Hui, Iris. "Shaping the Coast with Permits: Making the State Regulatory Permitting Process Transparent with Text Mining." *Coastal Management*, 45(3), 2017