



# Initial Steps Towards Automating Legal Document Editing

## Summary

**Objective:** Determine feasibility of using machine learning to improve the costly legal document drafting process

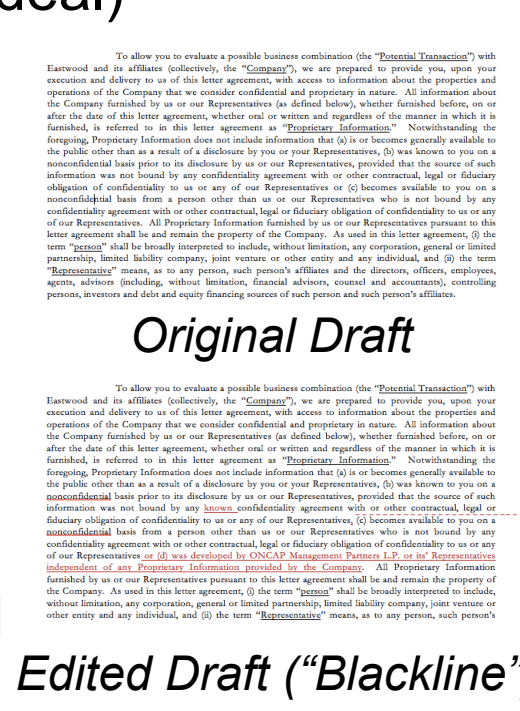
**Research Conducted:** Apply supervised learning models to predict initial edits to Non-Disclosure Agreements (“NDAs”)

**Key Results:** Classification results are promising yet also highlight the challenges from relying on unbalanced and sparse data from legal document text

## Background on NDAs

NDAs are a type of private contract that determines how the participants can use sensitive information

- 1 Business situation requires an NDA (e.g. a business development deal)
- 2 Company A writes a first draft of the NDA
- 3 Company B responds with initial edits
- 4 Edits are provided by both firms until a final NDA is agreed upon
- 5 The final NDA is signed by both firms



How Companies Use NDAs

## Data Set

Data set contained 105 NDA documents from a single investment firm from Summer 2010

- Sentences: ~4,000
- Total Tokens: ~202,000
- Unique Tokens: ~2,400

## Data Processing

Original and initial edited drafts of each NDA were compared to determine what changes were made and label the tokens

- Original tokens (“O”): 87%
- Inserted tokens (“I”): 7%
- Deleted tokens (“D”): 6%

Tokens were created with the Punkt’s sentence tokenizer, a regular expressions word tokenizer, and Porter’s stemmer

### Original Text:

‘...will have any legal effect. Within ten days, after being so requested...’

### Text with HTML-like Tags Representing Edits:

‘...will have any legal effect. ~~Within ten days,~~ Promptly after being so requested...’

### Labeled Tokens:

[...(`will', 'O'), (`have', 'O'), (`ani', 'O'), (`legal', 'O'), (`effect', 'O')] [(‘within', 'D'), (`ten', 'D') (`days', 'D'), (`promptli', 'I'), (`after', 'O'), (`being', 'O'), (`so', 'O'), (`requested', 'O'),...]

Data Processing Steps

## Methodology

Three models that are commonly used to tag text sequences in NLP were evaluated on their ability to correctly label tokens from the edited NDAs

Model	High-level Approach	Discriminative vs. Generative	Independence Assumptions	Additional Features
<b>Hidden Markov Model</b>	MLE for transition and output probabilities	Generative	Markov Property	No (only considers states)
<b>Structured Averaged Perceptron</b>	Online learning for weights and a bias	Discriminative	No	Includes prior and next words, suffixes, prefixes, etc.
<b>Conditional Random Field</b>	MLE for output conditional on feature function probabilities	Discriminative	No	Reputedly takes context into account, very flexible

Emphasis was placed on each model’s performance on Inserted and Deleted tokens because these labels represent the editing process and appear infrequently

## Classification Results For Training On All Sentences

	Hidden Markov Model				Structured Perceptron				Conditional Random Field			
	Overall Accuracy: 90.8%				Overall Accuracy: 91.9%				Overall Accuracy: 94.6%			
	Edited Token Accuracy: 37.5%				Edited Token Accuracy: 34.9%				Edited Token Accuracy: 65.5%			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Original	0.93	0.98	0.95	18,767	0.92	0.99	0.96	18,767	0.96	0.98	0.97	18,767
Inserted	0.85	0.49	0.62	1,458	0.89	0.46	0.61	1,458	0.94	0.78	0.85	1,458
Deleted	0.34	0.20	0.25	943	0.64	0.17	0.27	943	0.61	0.46	0.52	943
Total	0.89	0.91	0.90	21,168	0.91	0.92	0.90	21,168	0.94	0.95	0.94	21,168

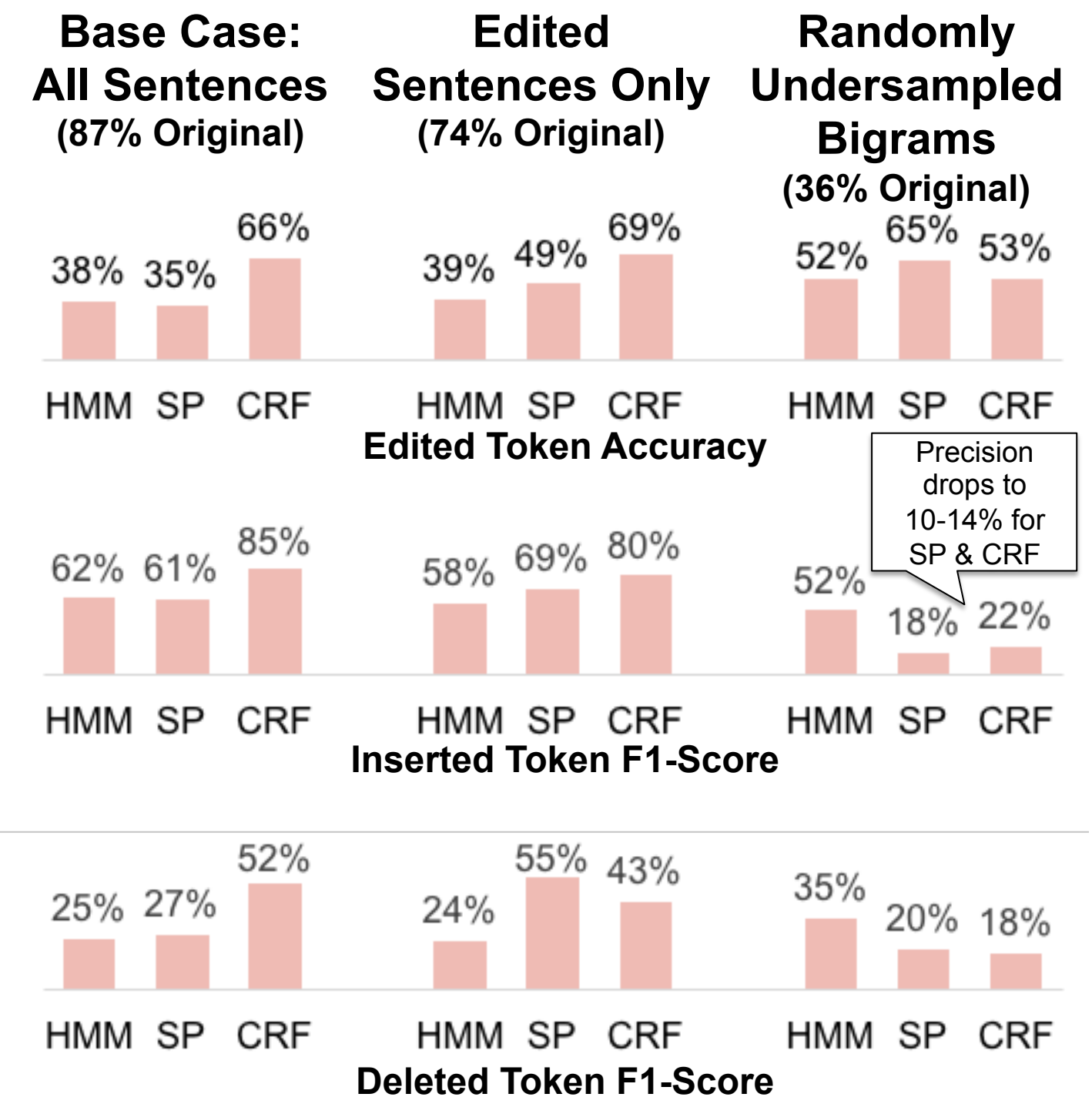
Actual		Predicted			Actual		Predicted			Actual		Predicted		
		Original	Inserted	Deleted			Original	Inserted	Deleted			Original	Inserted	Deleted
Original	18,312	106	349	Original	18,621	70	76	Original	18,447	58	262			
Inserted	720	711	27	Inserted	768	674	16	Inserted	299	1,142	17			
Deleted	730	23	190	Deleted	761	17	165	Deleted	491	21	431			

Note: Red shading indicates the most frequently predicted label for each type of actual token. Adjustments are not made to precision and recall calculations for boundary errors.

## Lauren Blake (lblake@stanford.edu)

### Adjustments for Unbalanced Data

Different modifications to the training data set were considered to mitigate the bias towards “Original” labels



## Models in Actions

**Hidden Markov Model:**  
~~Within ten days~~ Promptly after being so requested by the Company or [NAME] ~~[NAME]~~ [NAME] in writing except to the extent you are advised by legal counsel that complying with such request would be prohibited by law or regulatory authority, you will return or destroy at your cost all Evaluation ~~Material at the option~~ Material.

**Structured Averaged Perceptron:**  
~~Within ten days~~ Promptly after being so requested by the Company or [NAME] ~~[NAME]~~ [NAME] in writing except to the extent you are advised by legal counsel that complying with such request would be prohibited by law or regulatory authority, you will return or destroy at your cost all Evaluation ~~Material at the option~~ Material.

**Conditional Random Field:**  
~~Within ten days~~ Promptly after being so requested by the Company or [NAME] ~~[NAME]~~ [NAME] in writing except to the extent you are advised by legal counsel that complying with such request would be prohibited by law or regulatory authority, you will return or destroy at your cost all Evaluation ~~Material at the option~~ Material.

Example Paragraph Tagged By Each Model (Green = Correct, Red = Incorrect)