



# Automatic Classification of Legal Documents

Hailey (Han Bit) Yoon

hbyoon@stanford.edu

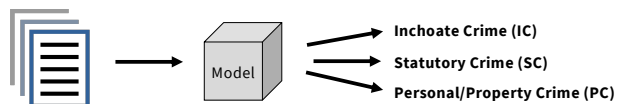


## Motivation

- There is a great need for lawyers for low-income people who cannot afford legal representation due to the continually rising price of legal service.
- The United States Attorneys Office (USAO)'s matrix of reasonable hourly rates for attorneys shows that amount has nearly doubled since 2003; the market rate in 2003-2004 is \$180 to \$380 [1], but the price in 2018-19 is \$307 to \$613 [2].
- I believe that we can make legal access more affordable by utilizing artificial intelligence (AI). AI technology can assist attorneys to minimize time spent on monotonous tasks and reduce hours billed.

## Problem Definition

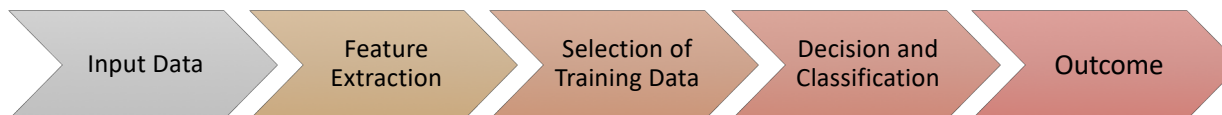
- The primary objective of this project is to automatically classify legal cases into three major crime types.
- **Input:** a set of legal documents
- **Output:** a list of matching crime types (inchoate crime, statutory crime, or personal/property crime)



## Challenges

- Since a court case often has 50+ pages each, parsing a set of legal documents and reshaping it to the format required effort and time.
- The current model classifies a document into one specific category, but there are legal cases with multiple criminal charges.

## Approaches



- Baseline – Logistic Regression:  $Loss_{logistic}(x, y, w) = \log(1 + e^{-1(w \cdot \phi(x))y})$
- Oracle – Human Judgement: The oracle consists of manual prediction of unprocessed documents.
- Feature Extraction: A legal document  $\rightarrow$  {Offender, Victim, Damage, Intent, ...}
- Decision and Classification:
  1. Naïve Bayes (NB):  $Posterior\ Probability = \frac{(Likelihood \cdot Class\ Prior\ Probability)}{Predictor\ Prior\ Probability}$
  2. k-Nearest Neighbors (kNN)
  3. Support Vector Machines (SVM)

## Results

- Analysis
  - In Figure 1, kNN shows the highest accuracy (74.33%) out of all three “decision and classification” strategies.
  - Figure 2 illustrates that the model successfully categorized most of the statutory crimes (72%) and personal/property crimes (84%). However, inchoate crime showed significantly low accuracy (51%).
  - Legal cases are easily misclassified to another category when it is only attempted and not committed.

- Future Work
  - To further increase the accuracy of inchoate crime, future training sets should include more non-violent personal/property crimes such as fraud, tax crimes, and gambling.

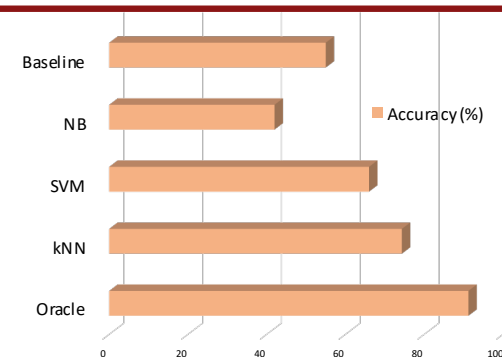


Figure 1. Testing Results

<u>KNN MATRIX</u>	IC	SC	PC	Accuracy (%)
IC	9	5	1	51
SC	3	36	4	72
PC	2	2	47	84

Figure 2. KNN Matrix

References: 1. Justice.gov. (n.d.). *LAFFEY MATRIX -- 2003-2012*. [online] Available at: [https://www.justice.gov/sites/default/files/usao-dc/legacy/2011/07/06/civil\\_Laffey\\_Matrix\\_2003-2012.pdf](https://www.justice.gov/sites/default/files/usao-dc/legacy/2011/07/06/civil_Laffey_Matrix_2003-2012.pdf) [Accessed 20 May 2019]. 2. Justice.gov. (n.d.). *USAO ATTORNEY'S FEES MATRIX -- 2015-2019*. [online] Available at: <https://www.justice.gov/usao-dc/file/796471/download> [Accessed 20 May 2019].