

Dato para actualizarse: Artificial Intelligence and Law (Springer) es una buena fuente de artículos en este tema. El tema abordado se le conoce como Legal Judgment Prediction (LJP).

Predicting Indian Supreme Court Judgments

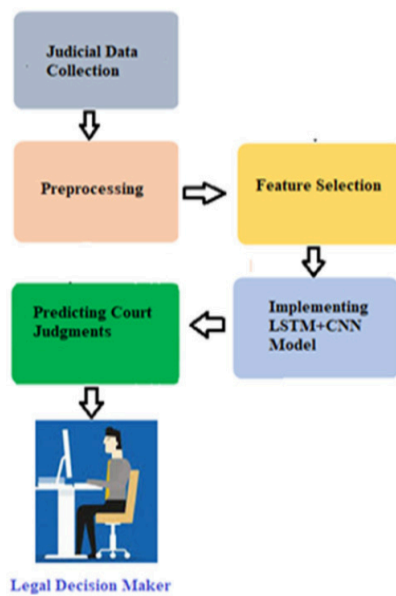
- **Fecha Publicación:** 23 May 2022
- **Revista:** Oxford Academic
- **Autor:** Sugam K Sharma
- **Puntaje y/o Métricas Usadas:** 76% accuracy (F1-score)
- [Enlace](#)
- **Tema:** Modelo Machine Learning entrenado en 3072 casos de la corte suprema de India. También se implementa un módulo operacional para que usuarios lo alimenten con un documento con el caso que desean predecir.
- **Conclusiones:**

Attributes beyond Attitudes: Personality Traits on the US Supreme Court

- **Fecha Publicación:** 21 Oct 2022
- **Revista:** Cambridge University Press
- **Autor:** Matthew E. K. Hall
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:** SCIPES
- **Tema:** Se asume que el comportamiento de jueces se basa en su personalidad individual y en sus ideologías como *Liberal* y *Conservador*. Con esto se refieren a sus **Actitudes Ideológicas**. Pero el paper va más allá de ese paradigma e indaga en los **Atributos** de los jueces, con esto se refiere a sus rasgos de personalidad que pueden basarse en el famoso test **Big Five** como *Apertura a la experiencia*, *Responsabilidad*, *Extroversión*, *Amabilidad*, *Neuroticismo*. Los datos usados fueron de la US Supreme Court desde 1946 a 2015 y su metodología fue usar Machine Learning.
- **Conclusiones:** El enfoque en Atributos aporta una capa explicativa adicional al margen de la Ideología.

Efficient Prediction of Court Judgments Using an LSTM+CNN Neural Network Model with an Optimal Feature Set

- **Fecha Publicación:** 22 Feb 2022
- **Revista:** MDPI
- **Autor:** Daniyal Alghazzawi
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:** 92.05% accuracy, 93% precision, 94% recall, 93% F1-score.
- **Tema:** Propone usar un modelo de redes neuronales híbrido: LSTM con CNN. Priorizando features con mejores scores.



También destacan la interfaz de usuario para el ingreso de casos para forecasting.

Court Judgment Prediction Tool				
Case Entry Screen		Training of Model (LSTM+CNN)		Predict Court Judgment
Case Entry Screen				
Petitioner	125	Law Type	1	
Case Origion	123	Respondent	405	
Lower Court Disagrement	0	Decision Direction Dissent	0	
Lower Court Decision	1	Opinion	3	
Issue	10500	Decision Direction	2	
Save Court Case Particulars				

- **Conclusiones:** Poca data disponible en este dominio (casos judiciales), solo una técnica estadística utilizada para extracción de features (RFE), usan embeddings en vez de red pre-entrenada, no se hizo una eficiente reducción de ruido. La eficacia del modelo varía demasiado en el tiempo, posibles factores: Percepción pública, desacuerdos entre poderes, cambios en los jueces y sus opiniones, cambios en normas y prácticas procesales.

A Study of Legal Judgment Prediction Based on Deep Learning Multi-Fusion Models—Data from China

- **Fecha Publicación:** 9 Sep 2024
- **Revista:** Sage Journals
- **Autor:** Yu Wen and Ping Ti
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:**

Table 3. Comparison of the Results of Different Neural Network Models.		
Task name	Penalty forecast	Law forecast
Models	Micro-Averaged F1	Jaccard
TextCNN	0.5892	0.7264
TextRNN	0.6031	0.7135
Wide&TextCNN	0.6378	0.7536
TextDenseNet	0.7086	0.8175

- **Tema:** Usa el dataset BDCI2017 que contenía:
 - 452 casos legales.
 - 40 000 ejemplos para entrenamiento
 - 10 000 ejemplos para prueba.
 - Todos en formato `.txt`.
 - Los datos de entrenamiento se dividieron en entrenamiento y validación (ratio 9:1). Este dataset viene de la competencia **CCF Big Data and Computing Intelligence Competition**, una de las principales en China (existe otro dataset que podría ser mejor CAIL2018) Se hicieron experimentos y aplicaron algoritmos de deep learning para mejorar predicción. El objetivo del estudio era analizar descripciones de casos legales y predecir: Normas legales aplicables y Multas/penas (sanciones legales). Utilizando modelos de **deep learning** combinados (Multi-Fusion Models), con el fin de mejorar la eficiencia y equidad en el sistema judicial. Los modelos evaluados fueron:
 - TextCNN
 - TextRNN
 - Wide & TextCNN
 - TextDenseNet
- **Conclusiones:** De estos, **TextDenseNet** obtuvo el mejor desempeño tanto en precisión como en las medidas empleadas (F1 micro-promediado y coeficiente de Jaccard)

An LLMs-based neuro-symbolic legal judgment prediction framework for civil cases

- **Fecha Publicación:** 8 Feb 2025
- **Revista:** Artificial Intelligence and Law (Springer)
- **Autor:** Bin Wei, Yaoyao Yu, Leilei Gan, Fei Wu
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:**
- **Tema:** Se centra en la falta de interpretabilidad que tienen los métodos de aprendizaje tradicionales con redes neuronales. Proponen para esto un framework neuro simbólico basado en LLMs. Este framework combina conocimiento legal (e.g., legal rules), represented through first-order logic rules, con deep neural networks (DNNs), se usa discrepancy loss para minimizar diferencias de la predicción entre las 2 componentes. Además desarrollan un Chain-of-Thought prompt que usa LLMs para extraer elementos fácticos de los casos legales. Estos elementos fácticos actúan como variables locales dentro de las reglas, apoyando el modelo lógico y mejorando la interpretabilidad. El dataset usado es sobre casos de préstamo y es privado.
- **Conclusiones:** Esta aproximación al problema no solo mejora la predicción y la interpretabilidad de ella.

A general approach for predicting the behavior of the Supreme Court of the United States

- **Fecha Publicación:** 12 Apr 2017
- **Revista:** Plos One
- **Autor:** Daniel Martin Katz
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:** 70.2% accuracy (caso general) 71.9% (voto por juez)
- **Tema:**

Modelo diseñado para predecir el comportamiento de la Corte Suprema de Estados Unidos con muestras generalizadas fuera de contexto. Se usa un clasificador Time-evolving random forest. Esto mejora los resultados sobre el dataset utilizado. Dataset consiste en 240.000 votos jurídicos (decisión de cada juez) y 28.000 casos (decisión el caso completo) desde 1816 hasta 2015.

Se entrena usando datos antes de la decisión del fallo. (Esto evita leakage).

Para comprobar si la diferencia entre su modelo y los baselines es **estadísticamente significativa**, aplican dos tipos de pruebas: paramétricas y no paramétricas.

Conclusiones: El modelo no solo predice bien en un período anual concreto de la Corte Suprema, sino que se puede aplicar a todos los períodos, pasados y futuros.

Predicting potentially abusive clauses in Chilean terms of services with natural language processing

- **Fecha Publicación:** 9 May 2025
- **Revista:** Artificial Intelligence and Law (Springer)
- **Autor:** Christoffer Löffler, Andrea Martínez Freile & Tomás Rey Pizarro
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:**
Detección: Macro-F1 score 79% a 89%. Micro-F1 score hasta 96%.
Clasificación: Macro-f1 score 60% a 70%. Micro-F1 score 64% a 80%
- **Tema:** En el contexto de una creciente preocupación por la inclusión de cláusulas abusivas en los términos y condiciones, que ha proliferado debido al aumento de servicios online, nace esta investigación con el objetivo de detectar y clasificar cláusulas abusivas en la "letra chica" de estos términos y condiciones. Se utilizaron 4 categorías y 20 clases como esquema de anotación para evaluar 50 "Términos y Condiciones de Servicio" online en Chile. Se utiliza un modelo basado en Transformers.
- **Conclusiones:** Gran variabilidad en el performance para diferentes tareas y modelos.

Artículo 1

- **Fecha Publicación:**
- **Revista:**
- **Autor:**
- [Enlace](#)
- **Puntaje y/o Métricas Usadas:**
- **Tema:**
- **Conclusiones:**