ASEC 2025 Conference
The 6th International Electronic Conference on Applied Sciences
09–11 December 2025 | Online
MDPI

Automatic Classification of Legal Cases: A Comparative Study of AutoGluon and the Gemini Family

José Maria Drumond de Vasconcelos Dias Filho[1], Carlos Mauricio Serodio Figueiredo[1]
1 PPGEEL – Universidade do Estado do Amazonas (UEA), Avenida Darcy Vargas, Parque Dez de Novembro 1.200 – Manaus – AM – Brasil

INTRODUCTION & AIM

The growing volume of legal proceedings in Brazil highlights the need for automating repetitive tasks in the Judiciary. A critical step is the classification of the "Procedural Class" in initial petitions, according to the Unified Procedural Tables (TPUs) of the National Council of Justice (CNJ). This step directly impacts case distribution and procedural flow. This study investigates the application of Artificial Intelligence (AI) to automatically classify petitions, comparing AutoGluon (traditional AutoML framework) with the Large Language Models (LLMs) from the Gemini family (2.0 Flash and 2.5 Pro). The objective is to evaluate the practical feasibility of using AI in a zero-shot regime, without specific training, to support lawyers and court staff in procedural triage.

METHOD

The dataset consists of authentic procedural documents from the Amazonas State Court of Justice (TJAM). It is composed of the full text of initial petitions and contains 9 distinct Procedural Classes, with a maximum of 3,000 cases per class, totaling a corpus of 19,614 documents. Baseline AutoGluon: Experiments were conducted using the complete set of petitions, except for the 900 documents reserved for direct comparative evaluation with the LLMs. Gemini Models: To assess the zero-shot learning capability of the LLMs, a random sample of 100 examples per class was extracted, totaling 900 documents. This subset served exclusively as the test set for the LLMs, which were not trained or fine-tuned on any TJAM data. The experimental design was developed to compare two fundamentally different approaches: A machine learning approach, based on task-specific supervised training, an LLM approach, leveraging general-purpose knowledge through prompts designed for the classification task. The first approach served as the baseline for comparison with the generative models of the second approach. Both approaches provide quantitative contributions to the task of legal document classification.

RESULTS & DISCUSSION

When evaluated on the same test set as the LLMs, AutoGluon achieved an overall accuracy of 92%, with a macro-F1 of 0.92.

Classe	Precisão	Recall	F1-score	Suporte
(7) Procedimento Comum Cível	0.92	0.93	0.93	100
(40) Monitória	0.98	1.00	0.99	100
(436) Juizado Especial Cível	0.92	0.96	0.94	100
(1268) Lei Maria da Penha	0.99	0.97	0.98	100
(1417) Regularização de Registro Civil	0.62	0.81	0.70	75
(1682) Retificação de Registro Civil	0.86	0.70	0.74	125
(1707) Reintegração / Manutenção de Posse	1.00	0.95	0.97	100
(12154) Execução de Título Extrajudicial	0.98	0.97	0.97	100
(14695) Juizado Esp. da Fazenda Pública	0.98	0.98	0.98	100
Acurácia			0.92	900
Macro avg	0.92	0.92	0.92	900
Weighted avg	0.92	0.92	0.92	900

In Table 2, the results obtained with Gemini 2.5 Pro (06-05) are presented. With an accuracy of 87.4% and a macro-F1 of 0.85, the model showed good performance across different classes.

Classe	Precisão	Recall	F1-score	Suporte
(7) Procedimento Comum Cível	0.83	0.65	0.73	100
(40) Monitória	0.99	0.97	0.98	100
(436) Juizado Especial Cível	0.79	0.99	0.88	100
(1268) Lei Maria da Penha	1.00	0.98	0.99	100
(1417) Regularização de Registro Civil	0.89	0.21	0.34	75
(1682) Retificação de Registro Civil	0.68	0.98	0.80	125
(1707) Reintegração / Manutenção de Posse	1.00	0.92	0.96	100
(12154) Execução de Título Extrajudicial	1.00	0.97	0.98	100
(14695) Juizado Esp. da Fazenda Pública	0.90	1.00	0.95	100
Acurácia			0.88	900
Macro avg	0.90	0.85	0.85	900
Weighted avg	0.89	0.87	0.86	900

In Table 3, the results of Gemini 2.0 Flash are shown. With an accuracy of 93.7% and a macro-F1 of 0.93, the model outperformed both the baseline and Gemini Pro. All classes achieved consistent results with F1 above 0.86, except for Regularização de Registro Civil (F1 = 0.79). Despite reaching its best performance with this model, the class still presents separation difficulties due to semantic proximity with Retificação ou Suprimento ou Restauração de Registro Civil.

Classe	Precisão	Recall	F1-score	Suporte
(7) Procedimento Comum Cível	0.88	0.84	0.86	100
(40) Monitória	0.99	0.97	0.98	100
(436) Juizado Especial Cível	0.88	0.95	0.91	100
(1268) Lei Maria da Penha	1.00	1.00	1.00	100
(1417) Regularização de Registro Civil	1.00	0.65	0.79	75
(1682) Retificação de Registro Civil	0.83	1.00	0.91	125
(1707) Reintegração / Manutenção de Posse	1.00	0.96	0.98	100
(12154) Execução de Título Extrajudicial	1.00	0.98	0.99	100
(14695) Juizado Esp. da Fazenda Pública	0.95	0.99	0.97	100
Acurácia			0.94	900
Macro avg	0.95	0.93	0.93	900
Weighted avg	0.94	0.94	0.94	900

CONCLUSION

Both approaches demonstrate practical feasibility for application in the Judiciary. AutoGluon offers high performance when labeled datasets are available, while Gemini LLMs provide rapid deployment capabilities without the need for training. The automation of classification can reduce rework, streamline case processing, and increase reliability in judicial data analysis.

FUTURE WORK / REFERENCES

As future work, we propose expanding the experiments and diversifying the methodological approaches such as Claude 3 and GPT-5, as well as expanding the scope of the classification beyond the “Procedural Class”, advancing to the automatic identification of “Procedural Subjects”.

References

- [Conselho Nacional de Justiça 2007] Conselho Nacional de Justiça (2007). Resolução nº46, de 18 de dezembro de 2007. Resolution of the National Council of Justice. Provides for the standardization of the Unified Procedural Tables.
- [Erickson et al. 2020] Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., and Smola, A. (2020). Autogluon-tabular: Robust and accurate automl for structured data. arXiv preprint arXiv:2003.06505.
- [Georgiev et al. 2024] Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., et al. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv preprint arXiv:2403.05530.