A Method to Predict Specific Types of Violence Against Women in Pernambuco

Arthur Carvalho da S. Xavier[1], Cleyton M. de Oliveira Rodrigues[1]

[1]Polytechnic School of Pernambuco (PPGEC)
University of Pernambuco (UPE) – Recife, Brazil

arthur.csxavier@upe.br, cleyton.rodrigues@upe.br

Abstract. *This study presents the main observations on the use of Machine Learning algorithms to predict specific types of violence against women in the state of Pernambuco. Only in 2024, more than 30.000 cases of gender-based violence were reported, with more than 100 resulting in homicide, in addition to serious harm to women's health. Therefore, the objective of the study is to find the best models for predicting specific types of violence to mitigate future cases. To this end, data from the Ministry of Women from 2015 to 2021 were used, in addition to tests with different Machine Learning models, resulting in models with accuracy and precision above 70% and 75%, respectively. In addition, an interpretability analysis was performed to identify bias in the best model.*

1. Introduction

Violence against women is a critical problem of health and public security problem, also recognized as one of the Sustainable Development Objectives of the United Nations for the 2030 Agenda. In Brazil, the number of gender-based violence reached more than 500,000 in 2024 [Conselho Nacional de Justiça 2024], where 8,464 were just the new entries for femicide cases. In Pernambuco, the data reinforce the importance of the theme, where the number of violence reports raised by 16,3% from 2023 to 2024 [Secretaria de Comunicação Social 2024]. A study by Barros (2021), between 2011 and 2016, presented also that the main characteristics related to homicide cases against women are related to young, black, and single women, occurring in public ways and with the use of fire arms [de Barros et al. 2021], reaffirming the necessity of efficient preventive measures.

As a response to the observed challenges, this study aims to compare the main machine learning algorithms and predictive analysis methods to present the best model configurations, based on data from the Ministry of Women (2015-2021). Using the main models most cited in the context of machine learning and cases of violence against women, the study seeks to carry out efficiency comparisons with 5 diffentent types of algorithms to predict classes of gender-based violence in Pernambuco, these being more traditional models, such as Recurrent and Convolutional Neural Networks, among others, in order to select the best ones for the task. The public database that is going to be used has different attributes that contributes to its strength, including the general groups of violence and the specific types committed against women in the state over 6 years, these being the research's prediction focus. By identifying the best models and configurations, the study aims to promote the necessary inputs for the development of robust prediction systems and, consequently, prevention, contributing to the reduction of gender-based violence cases in the state. Furthermore, after identifying the best model, an interpretability analysis using SHapely Additive exPlanations (SHAP) will be performed as a way to try to explain the output and possible bias in the model, leading to understanding the possible errors and paving the way to better systems in the future.

This paper is structured as follows. Section 2 discusses the different works to which this study is related, highlighting the differences and objectives of each. Section 3 presents the methodology used to carry out the tests. Section 4 outlines the experiments with the different machine learning algorithms. Finally, Section 5 summarizes the study and discusses potential directions for future research.

2. Related Works

In a systematic review conducted by [da Silva 2024], "Application of Machine Learning in the Detection and Prevention of Femicide: A Systematic Review", it was possible to identify the main applications of machine learning models within the context of femicide detection and prevention. The study, which encompassed more than 40 case studies, showed that algorithms such as Random Forest and Decision Tree are the most referenced, given their prominence in predictions and interpretability, respectively. Algorithms based on Gradient Boosting are also widely cited, while algorithms based on neural networks are the least used, indicating a possible lag of unsupervised algorithms in the context in question.

The study of [de Barros et al. 2021], "Factors associated with the homicides of women who are victims of violence", analyzed 32,308 cases of violence against women and 1,162 homicides from 2011 to 2016 taken place in Pernambuco, highlighting a predominance of cases against women above 20 years old, black and single, with the majority living in urban areas, where also occurs the majority of the cases. Also highlighting that most of the homicide cases had previous cases of physical violence. The study detailed with demographic analysis and factors associated with homicides of women helps with the development of prevention actions and supervision in the fields of health and public security.

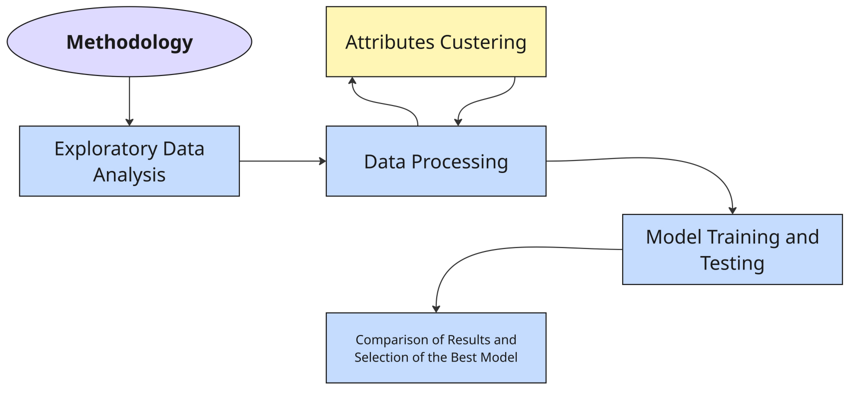
Another study conducted by [da Silva et al. 2024] brought an application of Self-Organizing Maps and Apriori algorithm within the same database of the Ministry of Women (2015-2021) with the aim of finding the most important association patterns and providing the best ideas for the development of targeted interventions and support systems to mitigate violence against women.

All of these studies, in general, deepen our understanding of gender-based violence in Pernambuco and even provide important information on how to work with the existing database and possible paths with different machine learning algorithms. In this sense, a path will be followed to understand, based on the context of the research base, which are the best possible models for predicting and, consequently, preventing cases of violence against women. With this, the study seeks to bring as novelty the prediction not only of the cases of violence against women, but specifically the type of crime that can be committed.

3. Methodology

This study uses a quantitative approach to address the cases of violence against women in Pernambuco, based on public data from the Ministry of Women between 2015 and 2021. The database contains demographic data, case characteristics, geographic locations, and site specifications, among other variables. The applied methodology follows 4 main steps (Figure 1):

Figure 1. Methodology Diagram



First, an Exploratory Data Analysis (EDA) was performed to observe demographic, spatial and temporal characteristics, allowing the definition of predictive variables. Then, the database was organized and filtered, treating null variables and readjusting categories for standardization.

Clustering techniques, such as KMeans, an iterative centroid-based clustering algorithm, with the TF-IDF vectorizer - a statistical measure, frequently used in Natural Language Processing to indicate the importance of a word in a document relative to a collection of documents [Bafna et al. 2016] - were also applied to reduce the dimensionality of a category, requiring adjustments to the parameters. The categorical data were handled by applying Label Encoding before the training with the models, except CatBoost, as it already handles categorical features natively. Next, the data was split in train and test with techniques such as hold-out, with a 70/30 ratio for training and testing, in addition to using cross-validation to compare the model performance that stood out the most.

After processing, the data were used in model training and configuration, adjusting with the most suitable parameters (Tables 1, 2 and 3). The process was defined for each model: Naive Bayes, CatBoost, Artificial Neural Network (ANN), Recurrent Neural Network (RNN), with Long-Short-Term Memory (LSTM) architecture and Convolutional Neural Network (CNN). Finally, tests and results comparison evaluated the efficiency of the models at predicting cases of violence, highlighting those who performed better.

Table 1. CatBoost's Parameters and Hyperparameters

Model	Iterations	Learning Rate	Depth
CatBoost	500	0.1	6

Table 2. ANN and LSTM's Parameters and Hyperparameters

Model	Epochs	Entry Nodes	Ocult Layers	Activ. Func.	Dropout Layers	Optmizer
ANN (y = groups)	100	128	1: 64 nodes	ELU	2: 0.1 and 0.2	Adam
ANN (y = specific types)	50	256	2: 128 and 256 nodes	ELU	3: 0.1	Adam
LSTM	50	32	1: 32 nodes	-	-	AdaMax

Table 3. CNN's Parameters and Hyperparameters

Model	Epochs	Filters and Kernel	Ocult Layers	Activ. Func.	Dropout Layers	Optmizer
CNN (y = groups)	50	64 and 2	2: 128 and 64 nodes	ReLU	2: 0.25	Adam
CNN (y = specific types)	50	32 and 2	1: 64 nodes	ReLU	0	Adam

4. Results

4.1. Exploratory Data Analysis

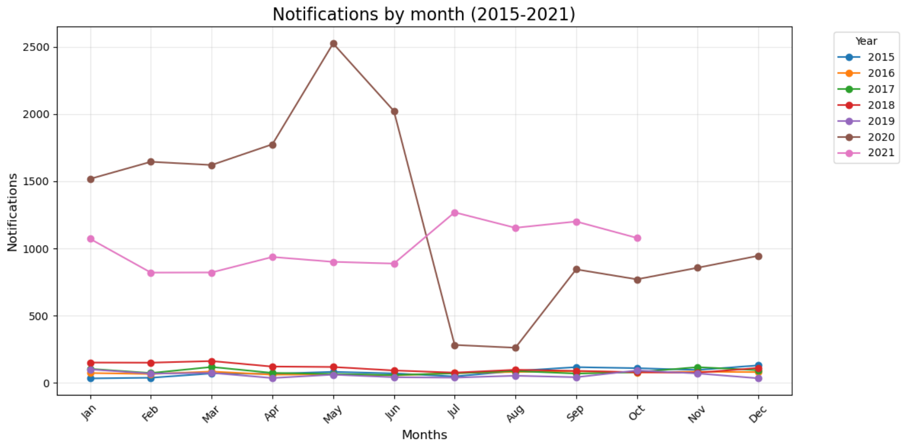
The database analysis includes a significant number of instances and categories of violence against women in Pernambuco. In a first moment, an initial treatment was performed, where the existence of columns with circa 60% of null values was observed, the latter being those removed. After that, predictive variables for model training were selected, described in Table 4.

Table 4. Descriptive table of predictive variables

Column	Description	Data type
Date	Data of the incident or report	Int
Group	General category of violence	String
Violation	Specific type of violence	String
State	State where the incident occur	String
Municipality	Municipality where the incident occur	String
Victim's age group	Approximately victim's age group	String
Victim's gender	Victim's gender	String
Victim's ethnicity	Victim's self declared ethnicity	String
Victim & suspect's relation	Relation between the victim and suspect	String
Suspect's gender	Suspect's gender	String

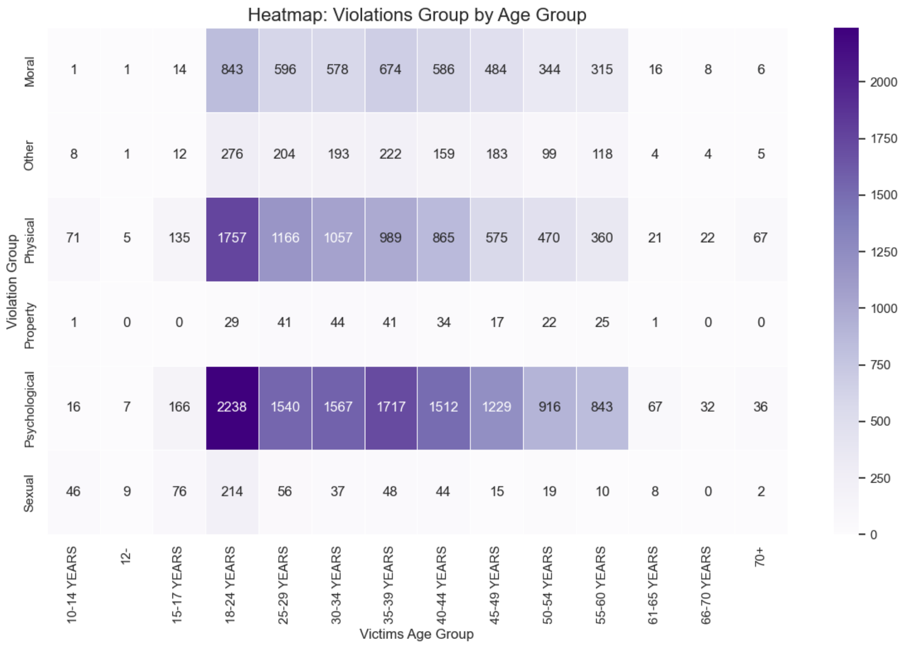
Sequentially, it was possible to generate charts to reinforce the understanding of the cases of femicide and they density over the years (Figure 2), being able to observe a peak of notifications in the initial months of the COVID-19 pandemic, leading one to assume that the database, starting in 2020, passed to have more robust data from the notifications.

Figure 2. Line chart with incidents of domestic violence over the years



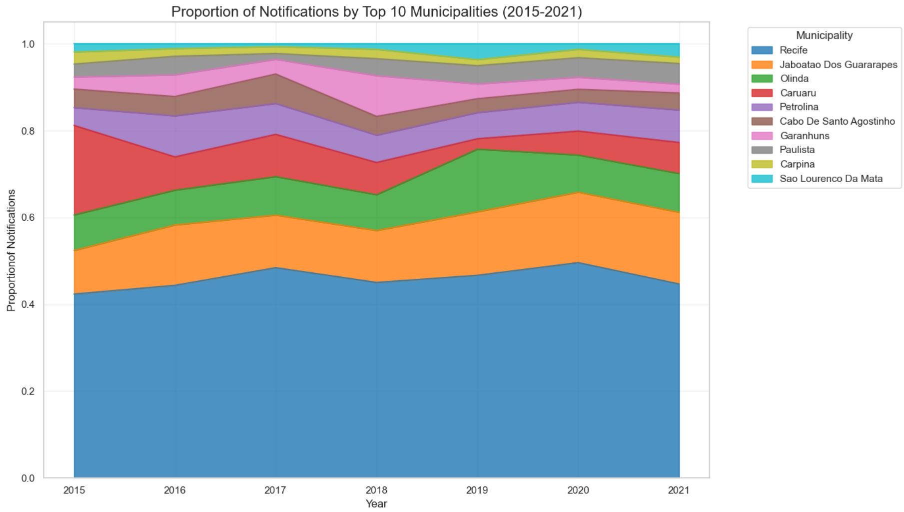
In addition to that, charts were generated that allowed visualization of victim age groups, centered on young women between 18 and 24 years of age, which are the main forms of violence, physical and psychological, respectively (Figure 3).

Figure 3. Heatmap of domestic violence incidents by age



Another important observation is the geographic location where the incidents occurred, presented in Figure 4, showing a strong tendency for most incidents occur in the most populated and urban cities, with emphasis on the state capital, Recife.

Figure 4. Piled area chart with number of domestic violence incidents at top 10 municipalities



4.2. Data Organization and Filtering

After completing the EDA, the instances were filtered and the database was readjusted before training the models. After completing the initial filtering, the database had 43,293 instances (excluding duplicates), but some classifications presented inconsistencies. The field for specifying types of violence presented the greatest inconsistencies, where, in certain instances, the specific type of violence was the same as the macro group of gender violence - e.g. instances in which physical violence was the group and the specific type. After removing these instances, the database had 30,151 instances, a significant and good number for training.

An important characteristic worth mentioning about the database, and that influenced significantly the obtained results, was the almost categorical-only nature of the database, besides the high unbalance of the data (Figure 5) - with types of violence such as Threat and Coercion with more than 4,800 instances, and Crimes Against Life representing less than 800 cases. Regarding instances with null data, it was opted to impute the lacking data considering the most frequent ones, since the option of removing these instances could cause more unbalance to the database.

After some initial training, it was observed that the database needed to be readjusted. Previously, the database was divided into 6 macro groups: physical, psychological, moral, sexual, property violence, and others; and 25 specific types of violence against women. However, in order to predict specific cases of violence, difficulties were encountered in establishing good models with acceptable reliability indices, mainly due to the high dimensionality. Because of this, it was decided to reclassify the 25 categories into 6 groups of violence with similar characteristics, which could or could not be from the same macro group. The reclassification was done through clustering using the KMeans method, with the TF-IDF vectorizer to create groups with syntactic proximity (Figure 6).

The next topics will cover the results obtained from the train and test of different chosen models, presenting the results before and after the reclassification with KMeans method.

Figure 5. Group balancing of the specific types of violence

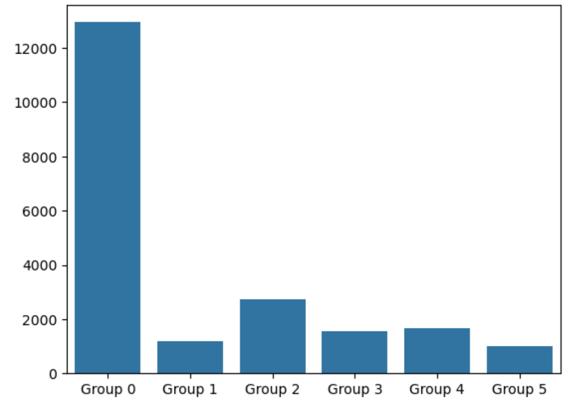
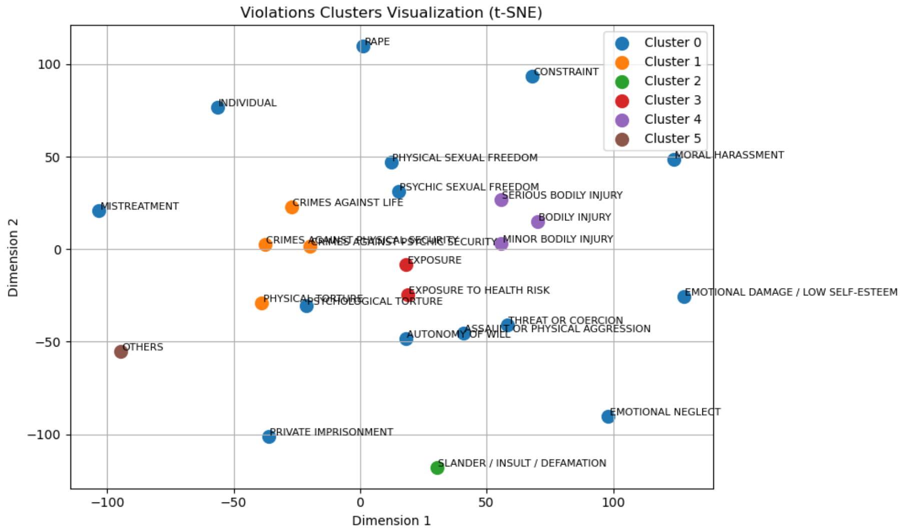


Figure 6. Clustering visualization of the types of violence



4.3. Models Training and Testing

For all model tests, fundamental evaluation metrics were used, such as Precision, F1-Score and Accuracy. For each algorithm, tests were performed to predict macro groups of violence and specific types, which is the final objective of this study. The results of both tests before and after clustering the types of violence will be presented here.

To justify the choice of each model, we opted for CatBoost, based on Gradient Boosting - which uses weaker decision trees with the ensemble method to build a robust model - because it handles categorical data natively [Prokhorenkova et al. 2018]. Naive Bayes was chosen because it is simple, fast and efficient in problems with many categorical variables. Another model tested was a simple Artificial Neural Network (ANN), with the aim of comparing the efficiency of unsupervised models with supervised ones.

The research also sought to work with more traditional machine learning models. Therefore, the fourth model used was a RNN, with LSTM architecture. RNNs are commonly used in predictions where subsequent instances depend on previous ones. However, in cases such as femicide, where the instances are not directly related, the use of this model is unconventional, and this testing is only academic. Finally, a Convolutional

Neural Network (CNN) was tested, which is more suitable for computer vision problems but can be adapted to contexts with categorical data, performing well with data in tabular form [Singh et al. 2021].

Comparing the results of the prediction tests for macro groups and specific types, the Naive Bayes model demonstrated low efficiency in both, especially in specific types, possibly due to the high complexity and dimensionality of the data. CatBoost obtained excellent results for macro groups, indicating its effectiveness and suitability for the parameters used, but had lower performance when predicting the 25 types of violence. The ANN model presented intermediate results for macro groups and weak results in specific types. The RNN (LSTM), had the worst results in both tests, reinforcing its inadequacy to the context. Finally, the CNN model showed high performance in macro groups, close to CatBoost, but had lower performance in specific types, like most models (Tables 5 and 6).

When observing the limited performance in predicting the 25 types of violence in all the models tested, it was assumed that the low performance would be directly related to the very high number of prediction classes, leading to the decision to regroup the specific types into 6 new classes. The sequential tests showed, in general, improvements in the prediction of the types of violence, but a significant reduction in the prediction of the macro groups.

Table 5. Performance Metrics of the Models for Macro Groups Before Reclassification

Model	Accuracy (%)	Precision (%)	F1-Score (%)	Recall (%)
Naive Bayes	35.38	37.64	31.19	35.38
CatBoost	**97.80**	**97.82**	**97.56**	**97.56**
ANN	85.82	68.86	62.44	59.65
RNN (LSTM)	45.46	7.57	10.41	16.66
CNN	97.31	84.10	83.77	83.62

Table 6. Performance Metrics of the Models for Specific Types Before Reclassification

Model	Accuracy (%)	Precision (%)	F1-Score (%)	Recall (%)
Naive Bayes	29.31	43.91	28.56	29.31
CatBoost	**52.65**	**45.98**	**45.66**	**52.65**
ANN	47.20	27.17	23.85	27.59
RNN (LSTM)	17.94	0.69	1.17	3.84
CNN	48.38	31.71	28.70	31.47

The Naive Bayes model was the only one that showed significant improvement in the prediction of macro groups and, mainly, of the reclassified specific types, although still below the ideal levels. The CatBoost model, in turn, showed a decrease in the metrics of the macro groups, observing a substantial improvement in the prediction of the specific types. The ANN model also showed a decrease in performance in the prediction of the macro groups, despite registering higher accuracy in the prediction of the types of violence, accompanied by a significant reduction in precision and F1-score. The RNN (LSTM) model, despite a slight improvement in the prediction of the specific types, maintained a low overall performance, demonstrating, once again, its inadequacy in the context of prediction crimes against women. Finally, the CNN model showed reasonable performance after the reclassification, with results similar to or better than the ANN for both categories (Tables 7 and 8).

Table 7. Performance Metrics of the Models for Macro Groups After Reclassification

Model	Accuracy (%)	Precision (%)	F1-Score (%)	Recall (%)
Naive Bayes	51.20	50.14	47.47	51.20
CatBoost	**76.64**	**78.68**	**75.05**	**76.64**
ANN	71.34	63.90	50.42	48.62
RNN (LSTM)	42.40	21.67	18.76	19.23
CNN	74.59	70.54	57.14	53.29

Table 8. Performance Metrics of the Models for Specific Types After Reclassification

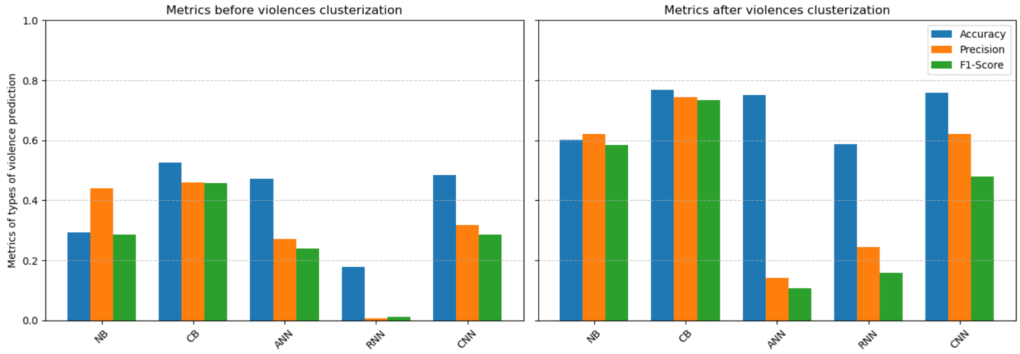
Model	Accuracy (%)	Precision (%)	F1-Score (%)	Recall (%)
Naive Bayes	60.15	62.07	58.45	60.15
CatBoost	**76.84**	**74.40**	**73.43**	**76.84**
ANN	75.14	14.22	10.80	10.41
RNN (LSTM)	58.71	24.48	15.79	17.96
CNN	75.92	62.21	47.84	46.00

4.4. Comparison of Results

Based on the developed analyses, charts were created to facilitate the comparison of accuracy, precision and f1-score results with the 5 different models, within the same database of the Ministry of Women, making it possible to observe their highlights within the tests before and after the reclassification of the specific types of violence (Figure 7).

From this, it became clear that, within the context worked on in the research, the prediction of many classes is still challenging, especially in a context with very unbalanced and scarce data. Because of this, in order to make a more targeted prediction of specific types of violence against women, it is still necessary to group these types of incidents into small classes with similar behaviors. In addition, it is also clear that the most balanced models for the context are those based on Gradient Boosting, indicating that, in order to achieve greater assertiveness in working with this type of prediction, it will be necessary to delve deeper into this architecture. Convolutional Networks also presented satisfactory performances, which may indicate a positive future within the studies.

Figure 7. Performance metrics at predicting types of violence



4.5. CatBoost Deep Dive

As already observed in this study, CatBoost, among the models tested, was the one that presented the best global metrics within the context of the base. In this sense, new tests were carried out that helped to understand the challenges with an unbalanced database and also brought new insights into new possibilities of working with the model in question.

First, it is important to analyze that the global data from the base model does not reflect reality for all classes, taking into account the high imbalance already mentioned. In this sense, it is possible to see that there are classes that have high false negative rates, raising a considerable alert, considering that the context of the prediction is quite sensitive.

To try to solve this challenge, two tests were proposed, one improving the model to work with weights, giving a greater focus to classes with fewer instances, and another using cross-validation – with the Stractified K-Fold variation – with 5 folds as a possible solution.

The test with the weighted model proved to be more appropriate within the context of the application, considering that it managed to improve sensitivity in relation to minority classes. As a result, classes such as Group 1 – which represents a critical group of violence – began to have fewer false negatives, even increasing the false positive rate, and this trade-off was acceptable to avoid serious errors in decision-making (Table 9). On the other hand, the test with cross-validation reinforced the stability of the model with behaviors similar to the base model and with weights, but it still did not bring a clear improvement in balance, failing in sensitivity in minority classes (such as 1, 3 and 4).

4.5.1. Feature importance analysis

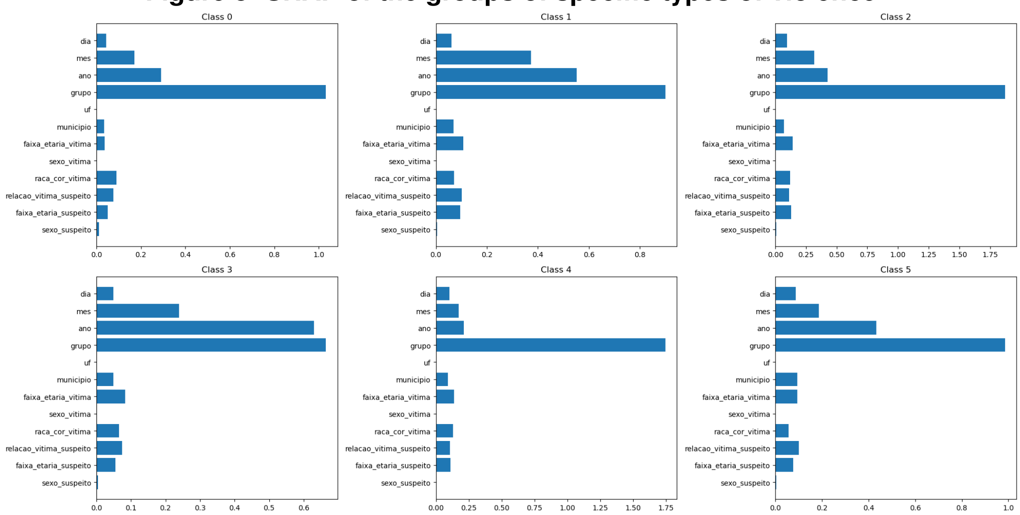
As a final part of the Deep Dive study of the CatBoost model, a brief analysis of the main attributes that influence the predictions was also carried out for each class. In view of

Table 9. Catboost Models Classification Report

Class	Precision (Base→Weighted→C.V.)	Recall (Base→Weighted→C.V.)	F1-Score (Base→Weighted→C.V.)
Group 0	0.78→0.87→0.78	0.92→0.77→0.93	0.84→0.82→0.85
Group 1	0.54→0.26→0.63	0.13→0.61→0.13	0.21→0.36→0.21
Group 2	0.84→0.84→0.84	1.00→1.00→1.00	0.91→0.91→0.92
Group 3	0.47→0.46→0.52	0.27→0.29→0.28	0.34→0.35→0.36
Group 4	0.64→0.42→0.62	0.29→0.46→0.29	0.40→0.44→0.40
Group 5	0.91→0.91→0.89	0.54→0.56→0.53	0.67→0.70→0.66

this, using SHAP, it was possible to observe that in some categories the main determining factor was, in addition to the macro groups of violence, the temporal variables (Figure 8). In this sense, it is important to reflect that this type of characteristic can induce biases to some extent that can mask possible more serious types of violence simply because of the time in which they occurred. Therefore, it is clear that there is a need to look for new discriminating characteristics in the base variables that can help to better interpret the predictions, in addition to looking for more recent data that occurred after 2021.

Figure 8. SHAP of the groups of specific types of violence



5. Final Considerations

This study analyzed cases of violence against women that occurred in Pernambuco from January 2015 to October 2021. By applying different types of machine learning algorithms, it was possible to identify the best configurations and models to predict cases of gender-based violence within the context of the database used, resulting in CatBoost being ranked as the most balanced and assertive model. In this sense, although it was identified that CatBoost is currently the best model, according to the parameters used, during the continuation of the research it will still be necessary to explore other models based on

Gradient Boosting. Even so, the final model found so far reflects that, with the current database, it is necessary to seek new strategies to deal with imbalance, testing techniques such as SMOTE (Synthetic Minority Oversampling Technique) and its variants, in addition to optimized parameter searches for new models.

Furthermore, during executions, it was observed that, even with the reclassification of specific types of violence, the accuracy of the models did not exceed 72%. In this sense, knowing that the clustering performed did not use any other restriction other than the Euclidean distance between the vectorized elements, it may be necessary to understand other ways of regrouping the types of violence, so that they respect the macro groups and their differences, in order to increase the efficiency metrics. However, the study showed that there is a possible path for the construction of more robust prediction models using a public database and with the processing of appropriate data, ensuring the continuity of efforts to achieve the objective of mitigating cases of violence against women. In this logic, it may be possible to build more efficient support systems for security forces that assist in the decision-making process of allocation efforts.

References

Bafna, P., Pramod, D., and Vaidya, A. (2016). Document clustering: Tf-idf approach. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pages 61–66.

Conselho Nacional de Justiça (2024). Painel violência contra a mulher [violence against women dashboard]. Available at: https://justica-em-numeros.cnj.jus.br/painel-violencia-contra-mulher/.

da Silva, B. B. (2024). Application of machine learning in the detection and prevention of femicide - a systematic review. https://doi.org/10.5281/zenodo.10802399.

da Silva, B. B. A., Pereira, R. M., de Oliveira Rodrigues, C. M., and de Araújo Fagundes, R. A. (2024). Analyzing patterns of violence against women in pernambuco using self-organizing maps and apriori algorithm. In *2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6, Bogotá D.C., Colombia.

de Barros, S. C., da Rocha Pimentel, D., de Oliveira, C. M., and do Bonfim, C. V. (2021). Factors associated with the homicides of women who are victims of violence. *Revista Brasileira de Enfermagem*, 21(5).

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). Catboost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pages 6638–6648.

Secretaria de Comunicação Social (2024). Ligue 180 registra aumento de 40 % nos atendimentos em 2024 [ligue 180 registers a 40 % increase in calls in 2024]. Available at: https://www.gov.br/secom/pt-br/assuntos/noticias-regionalizadas/ligue-180-balanco-2024/em-pernambuco-ligue-180-registra-aumento-de-40-6-nos-atendimentos-em-2024.

Singh, K., Mahajan, A., and Mansotra, V. (2021). 1d-cnn based model for classification and analysis of network attacks. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(11).