

Comparing machine learning and deep learning models to moderate inappropriate messages on social media.

Nicolas Bedoya Figueroa
Daniel Escalante Pérez
Marilyn Stephany Joven Fonseca
Eder Leandro Carbonero Baqueron

Abstract—Aquí va el resumen del artículo, una breve descripción del problema, la metodología y los resultados.

I. INTRODUCTION

II. METHODOLOGIES

III. MACHINE LEARNING METHODS

This section describes the classification methods used in this study. Each represents a different approach within supervised learning.

A. Random Forest

1) *Definition:* Random Forest is an ensemble algorithm that builds multiple decision trees during training and outputs the class that is the mode of the individual trees' classes. It is robust against overfitting and usually performs well without complex tuning.

B. Decision Tree

1) *Definition:* A decision tree is a predictive model that represents decisions and their possible consequences. It is easy to interpret and train but can be prone to overfitting if the tree depth is not controlled.

C. Logistic Regression

1) *Definition:* Logistic regression is a statistical model used to predict the probability of a binary class. It is efficient for linear problems and serves as a baseline for comparison with more complex models.

To reference one of the earliest formulations of the model, we draw upon the perspective presented in *The Regression Analysis of Binary Sequences* by D. R. Cox (1958), which introduced a rigorous statistical framework for the analysis of binary outcomes. In this work, both dependent and independent variables are explicitly identified, and the probability of a binary event is modeled using a logistic transformation of the predictors. Dependent variables are represented as binary categorical variables, typically coded as 0 and 1, which enables the analysis of how one or more explanatory variables

influence the likelihood of an event of interest. Through illustrative examples, the author demonstrates how the model can be estimated using maximum likelihood methods, thereby establishing a methodological foundation that continues to be widely applied in fields such as biostatistics, the social sciences, and machine learning [Cox, 1958].

Given the capabilities offered by the model, it has been selected for use in this study. However, prior to its implementation, we will review some previous approaches used in earlier research.

As one of the initial and relevant approaches, the study presented in the article "Offensive Language Detection Using Soft Voting Ensemble Model" [Suhartono et al., 2023] incorporates logistic regression as part of its ensemble strategy. The article emphasizes the use of L2 regularization, which enhances the precision of offensive language classification by mitigating overfitting. Furthermore, the model's simplicity is highlighted as a significant advantage, enabling efficient implementation even with limited computational resources—demonstrating its usefulness and versatility within the set of evaluated classifiers.

The research article titled "Comparison between Machine Learning and Deep Learning Approaches for the Detection of Toxic Comments on Social Networks" [Bonetti et al., 2023] employs logistic regression once again, combining it with two semantic modeling techniques: Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA). These methods will be further discussed in the following paragraphs. Notably, the study reports a prediction accuracy exceeding 91

D. AdaBoost

1) *Definition:* AdaBoost (Adaptive Boosting) combines multiple weak classifiers, such as shallow decision trees, to form a strong classifier. Each new classifier focuses more on the examples misclassified by the previous ones.

E. Naïve Bayes

1) *Definition:* Naïve Bayes is a probabilistic classifier based on Bayes' theorem with a strong (naïve) assumption of independence between features. It is fast, efficient, and works well on text and classification problems with many features.

F. SVM

1) *Definition:* Support Vector Machines (SVM) aim to find the optimal hyperplane that separates classes with the maximum margin. It is particularly effective in high-dimensional spaces and when the number of features exceeds the number of samples.

IV. DEEP LEARNING METHODS

Deep learning is a subset of machine learning that uses artificial neural networks with many layers to automatically learn complex patterns from large amounts of data.

A. Artificial Neural Networks (ANN)

1) *Definition:* Artificial Neural Networks are computational models inspired by the human brain, composed of layers of interconnected nodes (neurons). Each neuron receives inputs, processes them with a weight and bias, and applies an activation function to produce an output.

B. Convolutional Neural Networks (CNN)

1) *Definition:* Convolutional Neural Networks are specialized neural networks primarily used for image processing tasks. They use convolutional layers to automatically extract spatial features from images, followed by pooling and fully connected layers for classification.

C. Recurrent Neural Networks (RNN)

1) *Definition:* Recurrent Neural Networks are designed to recognize patterns in sequences of data by using loops in the network to maintain information across time steps. They are commonly used in natural language processing and time-series prediction.

D. Long Short-Term Memory (LSTM)

1) *Definition:* Long Short-Term Memory networks are a type of RNN that can learn long-term dependencies using a special architecture that controls the flow of information. They are effective in handling the vanishing gradient problem during training.

E. Autoencoders

1) *Definition:* Autoencoders are unsupervised neural networks that learn efficient codings of data. They consist of an encoder that compresses the data and a decoder that reconstructs it. They are used for tasks like dimensionality reduction and anomaly detection.

F. Generative Adversarial Networks (GAN)

1) *Definition:* Generative Adversarial Networks consist of two neural networks—the generator and the discriminator—that compete against each other. The generator creates synthetic data, while the discriminator evaluates them, leading to the generation of highly realistic data.

V. RESULTS

VI. CONCLUSIONS

REFERENCES

- Andrea Bonetti, Marcelino Martínez-Sober, Julio C. Torres, Jose M. Vega, Sebastien Pellerin, and Joan Vila-Francés. Comparison between machine learning and deep learning approaches for the detection of toxic comments on social networks. *Applied Sciences*, 13(10):6038, 2023. doi: 10.3390/app13106038. URL <https://www.mdpi.com/2076-3417/13/10/6038>.
- D. R. Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2):215–242, 1958.
- Derwin Suhartono et al. Offensive language detection using soft voting ensemble model. *Mendel Journal*, 29(1):1–10, 2023.