Universidad Politécnica De Yucatán

Machine Learning

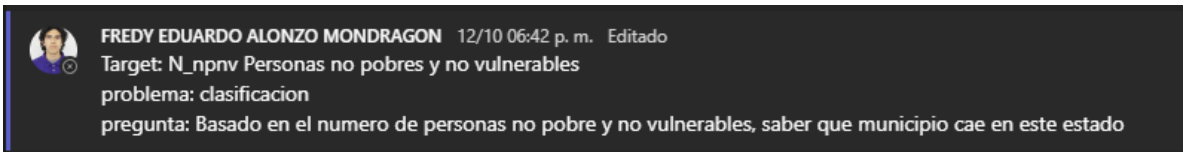
U2 – Implementing a Predictor from Scratch

- Fredy Eduardo Alonzo Mondragon

Professor: Victor Ortiz

Date: October 23rd, 2023

a. **A screenshot of the post in the general channel, where you indicate your selected objective, your selected problem (regression or classification), and the question the predictor will answer.**



b. **A chronological document or journal that describes in detail:**

I. **all the steps taken to get your dataset ready (including the rationale over why to keep or drop features or observations).**

1. Initial Analysis of the Dataset:

Before any manipulation, it is crucial to deeply understand the dataset. This initial analysis involves a detailed exploration of the characteristics, their relevance to the problem at hand, and clear identification of the target variable. This step is essential to establish a solid foundation for subsequent work.

2. Deal with Missing Data:

One of the common challenges in dataset preparation is the presence of null values. It is crucial to identify columns with missing values and make informed decisions on how to address them. Whether removing rows with null values or strategically padding them, these decisions affect the integrity and relevance of the data.

3. Feature Selection:

Not all features contribute equally to the prediction. Careful feature selection is vital to improve model accuracy and avoid overfitting. A correlation analysis is performed to understand the relationships between the characteristics and the target variable. Irrelevant and highly correlated features are removed to maintain the independence and relevance of the variables.

4. Treatment of Categorical Variables:

Categorical variables need to be transformed to be used by machine learning algorithms. One-hot coding is a common technique to avoid misinterpretations of order relationships. This encoding ensures that each category is treated as an independent entity, preserving the integrity of the categorical data.

5. Normalization or Standardization (Optional):

Variation in feature scales can negatively affect some machine learning algorithms. Normalizing or standardizing the data ensures that all characteristics are on a similar scale, avoiding bias in the model due to magnitude differences.

II. All steps taken to train your predictor. Including an explanation on Why to choose such algorithm and an explanation on its characteristics.

1. Dataset Partition:

Dividing the dataset into training and test sets is essential to evaluate the generalization capacity of the model. Using one portion for training allows the model to learn patterns, while the other portion is reserved for evaluating its performance on unseen data. This division provides an objective evaluation of the model.

2. Model Validation and Adjustment:

Choosing a suitable machine learning algorithm is essential and depends on the specific problem. Model parameters are adjusted using techniques such as cross-validation to avoid overfitting. This step ensures that the model fits well with the training data and can generalize correctly to unseen data, ensuring optimal performance in the real world.

V. A link to a github project where you publish your code.

https://github.com/FredyA13/Implementing-a-Predictor-from-scratch

c. A reflection on what were your main challenges during this project, how do you see this could be applied to the field of Robotics.

Throughout this project, numerous challenges arose. Firstly, identifying the dataset we were working with was crucial, as understanding how each feature impacted the model, we were constructing proved to be a complex and educational process. Then came the challenge of cleaning the dataset to prepare it properly for the model. This task involved getting rid of null or nonexistent values, something I hadn't encountered much before. Converting categorical features into numerical ones was also a new task for me. We had to carefully consider which features were essential and what dimensions of the dataset were sufficient for the model we planned to use.

Once the dataset was ready, the next challenge was building the classification model. Although the result wasn't the best, this process was fundamental for learning to handle the data. I am aware that there is much room for improvement in this procedure by applying more appropriate methods.

Regarding the application of machine learning in robotics, a complex yet exciting landscape emerges. Integrating machine learning into robotics is crucial for a wide range of applications, from obstacle avoidance in mobile robots to detecting moving objects, mapping, and localization for precise path following. However, it's not all sunshine and rainbows. Applying machine learning in robotics can be costly and time-consuming, especially when considering the size of simulations and the specific nature of real-world tasks.

Simulations are essential in this domain as they provide a controlled environment to test and refine algorithms before deploying them on real robots. Despite being a close representation of reality, simulations can fail in some cases due to limitations in cost and processing power. It's a delicate balance between accuracy and economic viability.

In summary, machine learning offers tremendous possibilities in the field of robotics, but its application requires careful consideration of the technical and financial challenges involved. Ongoing research and development in this field are essential to overcoming these barriers and enabling significant advancements at the intersection of machine learning and robotics.