



What is AdaBoost?

AdaBoost, short for Adaptive Boosting, is a powerful ensemble learning method that combines multiple weak learners to create a strong classifier. Introduced by Yoav Freund and Robert Schapire in 1996, AdaBoost is particularly effective in improving the performance of weak classifiers by sequentially training them on the same dataset, with each subsequent learner focusing more on the instances that were misclassified by the previous ones. By assigning higher weights to these misclassified instances, AdaBoost ensures that subsequent weak learners pay more attention to them, ultimately leading to a strong, accurate classifier. Yoav Freund and Robert Schapire won 2003 Gödel Prize for this work.^[1] AdaBoost has been widely applied in various domains, including Spam filtering, face recognition, and sentiment analysis, due to its simplicity, effectiveness, and ability to handle complex classification tasks

The Idea behind AdaBoost

Step 1: Assign equal weights to all training data points in the dataset. These weights represent the importance of each sample during the training process.

Step 2: Train one or a set of weak classifiers like decision stump on the training data.

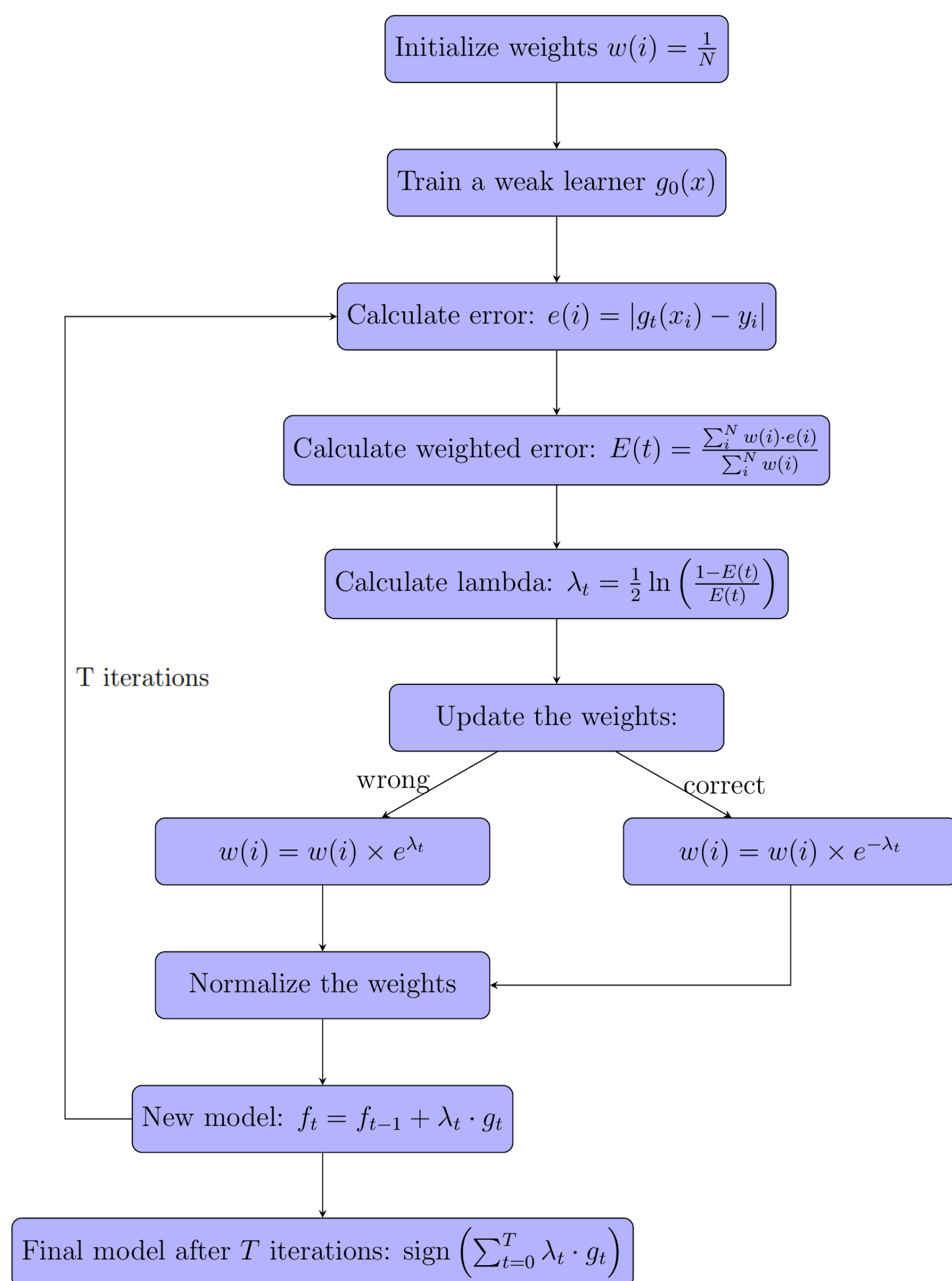
Step 3: Calculate the error made by the weak classifier. Assign weights to classifiers based on their accuracy. Better classifiers receive higher weights.

Step 4: Assign higher weights to misclassified data points and lower weights to correctly classified data points.

Step 5: Repeat step 2 to 4 for a specified number of iterations, refining predictions at each step.

Step 6: Aggregate predictions of all weak classifiers with their respective weights using a weighted majority vote to make the final prediction.^[2]

The Algorithm



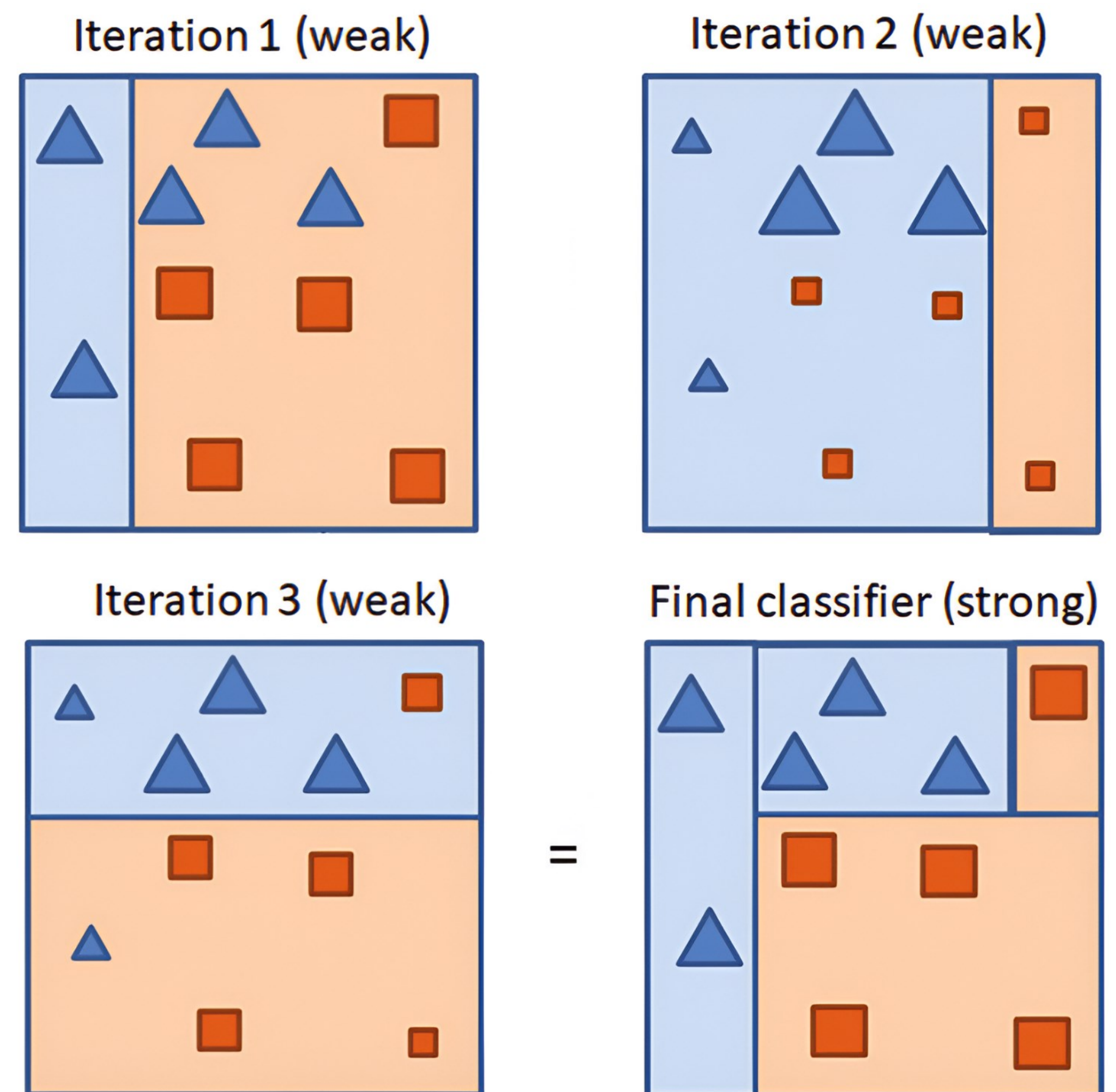
How to Implement AdaBoost in your Code!

AdaBoost can be easily implemented in your code using sklearn.ensemble module.^[4]

```

from sklearn.ensemble import AdaBoostClassifier
from sklearn.tree import DecisionTreeClassifier
base = DecisionTreeClassifier(max_depth=1)
adaboost_classifier = AdaBoostClassifier(base,
    n_estimators=50, random_state=42)
adaboost_classifier.fit(X_train, Y_train)
predictions = adaboost_classifier.predict(X_test)
  
```

Illustrated Example



Advantages and Disadvantages

Advantages	Disadvantages
Boosts accuracy by combining weak models to create a strong model.	Sensitive to outliers and noisy data.
Adapts to challenging examples by focusing on past mistakes.	Training can be computationally expensive for large datasets or many models.
Works with various types of base models for adaptable ensemble creation.	May struggle with imbalanced datasets. Additional techniques need to handle them. ^[2]

Applications using AdaBoost

- Facial recognition^[5]
- Spam filtering^[6]
- Recommendation system^[7]
- Financial Fraud Detection^[8]
- Sentiment analysis^[9]
- Gesture Recognition^[10]

References

1. Wikipedia contributors. (2024, March 6). AdaBoost. Wikipedia. <https://en.wikipedia.org/wiki/AdaBoost>
2. Wizards, D. S. (2023, July 8). Understanding the AdaBoost Algorithm - Data Science Wizards - medium. Medium. <https://medium.com/@datasciencewizards/understanding-the-adaboost-algorithm-2e9344d83d9b>
3. Machine Learning Mastery. (2020, July 9). AdaBoost Math with Example clearly explained Step by Step Machine Learning Ensembles [Video]. YouTube. <https://www.youtube.com/watch?v=peh2l4dePBc>
4. sklearn.ensemble.AdaBoostClassifier. (n.d.). Scikit-learn. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>
5. Corso, J. (2014). AdaBoost for face detection. In EECS 598 Fall 2014 (pp. 1–61). https://web.eecs.umich.edu/~jjcorso/t/598F14/files/lecture_1117_boosting.pdf
6. Using Adaboost and Stochastic gradient descent (SGD) Algorithms with R and Orange Software for Filtering E-mail Spam. (2019, September 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/8974319>
7. AdaBoosting for case-based recommendation system. (2016, October 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/7857591>
8. Credit card fraud detection using AdaBoost and majority voting. (2018). IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/document/8292883>
9. Silva, N. F. F., Hruschka, E. R., University of São Paulo, USP, Federal University of Sao Carlos, & Hruschka, E. R. Jr. (2014). Biocom Usp: Tweet Sentiment Analysis with Adaptive Boosting Ensemble. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 129–134). <https://aclanthology.org/S14-2018.pdf>
10. Head Gesture Recognition System Using Adaboost Algorithm with Obstacle Detection. (2015, November 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/7425580>