Background & Motivation:

Water is an ultimate gift of nature to humanity and every living organism. Most of us have the luxury to lead our daily lives without feeling blunt of water scarcity but there are people in some parts of the world where they would have to walk several miles to have access to water. The situation worsens for them if this access, such as pumps, wells or other water-points, also shuts down unexpectedly.

If a Machine Learning algorithm can predict which water point may fail and identify patterns/reasons for the failure then appropriate measures can be taken by concerned authorities.

Data:

Data presented is collected from few of such regions and it represents the information related to construction, geography, physical characteristics and other aspects of the water-point. The target column is the working condition which could be either functional/non-functional.

Objective:

Your goal is to build a ML model which trains on the given data, is appropriately validated and finally predicts on the test data provided on the grader tool.

Metric: Accuracy

Accuracy is defined as :

(No. of True Positives+ No. of True Negatives)/ (Total No. of Instances)

Other Instructions:

Spend enough time on pre-processing and data understanding. Think of the problem from domain's perspective to make your model smarter.
Your final grader score carries much lower weightage than your overall approach which includes data exploration and model validation. Use your time wisely.

Attribute Information:

1. Id - row id
2. Amount_of _water - Amount of water available in the water pump
3. Gps_height - Altitude of the well

4. Waterpoint name - Name of the waterpoint if there is one
5. Waterpoint_type - The kind of waterpoint
6. Basin_name - Geographic water basin
7. Village - Geographic location
8. Regionname - Geographic location
9. Region_code- Geographic location (coded)
10. Wardname- Geographic location (renamed the original location)
11. Districtcode - Geographic location (coded)
12. Population – No of people around the well
13. Public_meeting – Is there any public meetings were conducted. (If Yes TRUE otherwise False)
14. Organization_funding - organization that provides money for that well construction
15. Organization_surveyed– organization which has done survey to collect the data.
16. Scheme_management - Organization operating the water point
17. Schemename – scheme name under which water point sanctioned.
18. Permit - Is the waterpoint has the permission to use
19. Company_installed - Organization that installed the pump
20. Management - How the waterpoint is managed
21. Management_group – Group which manages the water point.
22. Extraction_type – Way of extracting the water
23. Extraction_type_group - Way of extracting the water
24. Extraction_type_class – class of equipment used to extract the water
25. Payment - What the water costs type of payment
26. Payment_type - What the water costs type of payment
27. Water_quality - The quality of the water
28. Quality_group - The quality of the water
29. Quantity - The quantity of water
30. Quantity_group - The quantity of water
31. Source - The source of the water
32. Source_type - The source of the water
33. Source_class - The source of the water
34. Waterpoint_type_group - The kind of waterpoint.

35. Target: Status- functional/non-functional