**Title : Large Soybean Database**

**ABSTRACT**

This project report entitled to **“A study on soybeans diseases”**. The main objective of the study is to analyze the species, size of soybeans, plantation, temperature, and many more attributes related to soybeans.

The tool used for the study is **Weka 3.8.** Charts and tables are used for better understanding.

Regarding the dataset we used a version of the soybean which I believe to recognize 17 different diseases of soybean.

We have increased the accuracy and decreased the number of attributes.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Contents** | **Page No** |
| **1.** | **Abstract** |  |
| **2.** | **Table of Contents** |  |
| **3.** | **Introduction** |  |
| **4.** | **Experimental Techniques and Tool used** |  |
| **5.** | **Results and Conclusion** |  |
| **6.** | **References** |  |

**Introduction**

**Attribute Information:**

19 Classes

( diaporthe stem canker, charcoal rot, rhizoctonia root rot,

phytophthora rot, brown stem rot, powdery mildew,

downy mildew, brown spot, bacterial blight,

bacterial pustule, purple seed stain, anthracnose,

phyllosticta leaf spot, alternarialeaf spot,

frog eye leaf spot, diaporthe pod & stem blight,

cyst nematode, 2 4 d injury, herbicide injury. )

1. date: april, may, june, july, august, september, october.

2. plant stand: normal, lt normal .

3. precip: lt norm, norm, gt norm .

4. temp: lt norm, norm, gt norm .

5. hail: yes, no .

6. crop hist: diff lst year, same lst yr, same lst two yrs,

same lst sev yrs .

7. area damaged: scattered, low areas, upper areas, whole field .

8. severity: minor, pot severe, severe .

9. seed tmt: none, fungicide, other .

10. germination: 90%, 100%, 80 89%, lt 80% .

11. plant growth: norm, abnorm .

12. leaves: norm, abnorm.

13. leafspots halo: absent, yellow halos , no yellow halos .

14. leafspots marg: w s marg, no w s marg, dna .

15. leafspot size: lt 1/8, gt 1/8,dna .

16. leaf shread: absent, present .

17. leaf malf: absent, present .

18. leaf mild: absent, upper surf, lower surf .

19. stem: norm, abnorm .

20. lodging: yes, no .

21. stem cankers: absent, below soil, above soil, above sec nde .

22. canker lesion: dna, brown, dk brown blk,tan .

23. fruiting bodies: absent, present .

24. external decay: absent, firm and dry, watery .

25. mycelium: absent, present .

26. int discolor: none, brown, black .

27. sclerotia: absent, present .

28. fruit pods: norm ,diseased, few present, dna .

29. fruit spots: absent,colored, brown w/blk specks,distort,dna.

30. seed: norm, abnorm .

31. mold growth: absent, present .

32. seed discolor: absent, present .

33. seed size: norm, lt norm .

34. shriveling: absent, present .

35. roots: norm, rotted, galls cysts .

**Class Distribution:**

1. diaporthe stem canker: 10

2. charcoal rot: 10

3. rhizoctonia root rot: 10

4. phytophthora rot: 40

5. brown stem rot: 20

6. powdery mildew: 10

7. downy mildew: 10

8. brown spot: 40

9. bacterial blight: 10

10. bacterial pustule: 10

11. purple seed stain: 10

12. anthracnose: 20

13. phyllosticta leaf spot: 10

14. alternarialeaf spot: 40

15. frog eye leaf spot: 40

16. diaporthe pod & stem blight: 6

17. cyst nematode: 6

18. 2 4 d injury: 1

19. herbicide injury: 4

**Experimental Techniques and Tool Used**

The tool we used is **WEKA 3.8**.

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre processing, classification, regression, clustering, association rules, and visualization. It is also well suited for developing new machine learning schemes.

Experimental Technique used is **J48 –C 0.25 –M 2**.

J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of the **C4.5** **algorithm**.

What is **J48 decision tree**?

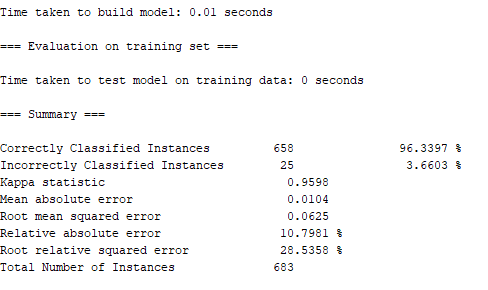
Imagine that you have a dataset with a list of predictors or independent variables and a list of targets or dependent variables. Then, by applying a decision tree like J48 on that dataset would allow you to predict the target variable of a new dataset record.

**Result and Conclusion**

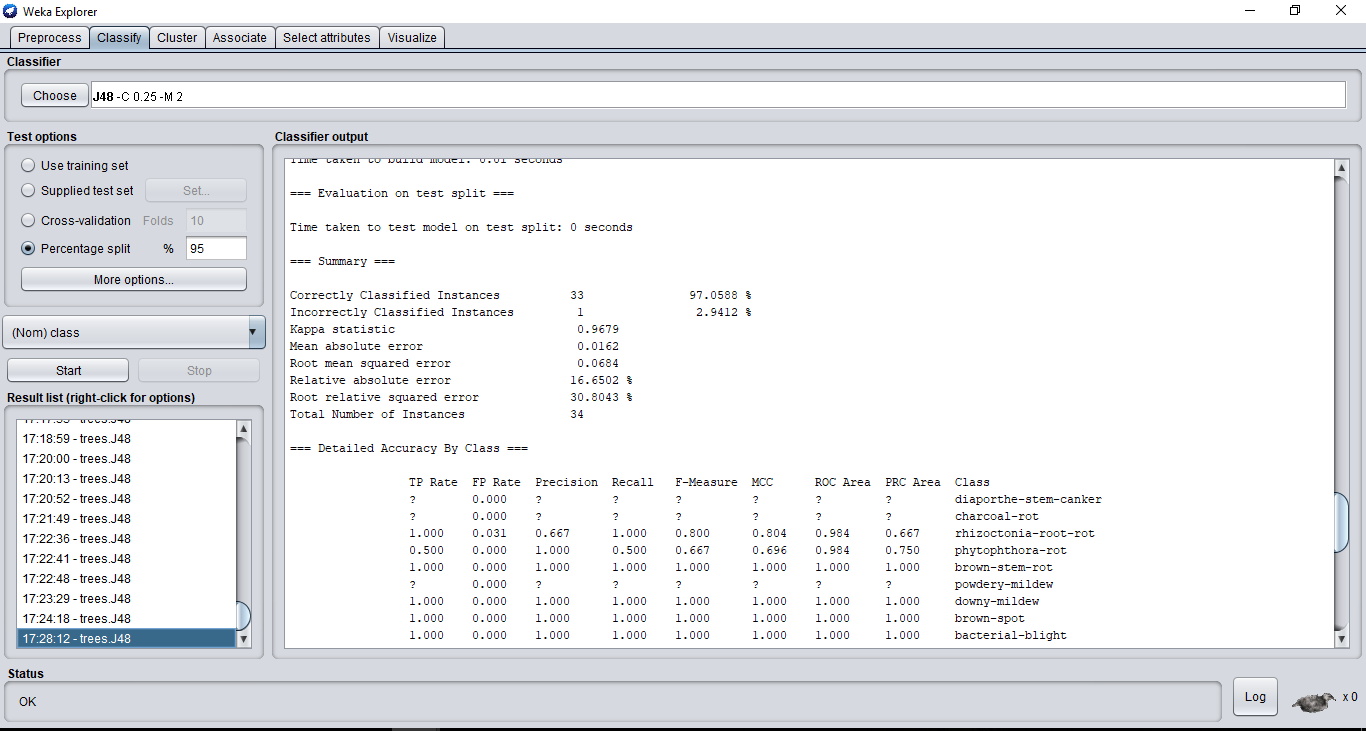
We got some amazing results by changing the training size of datasets and we have also reduced the number of attributes from **35** to **15** most influential ones and increased its accuracy from **96.3397** to **97.0588**, using the information related to soybean diseases.

Followings are the screenshots of the result of our experiments and the previous original result.

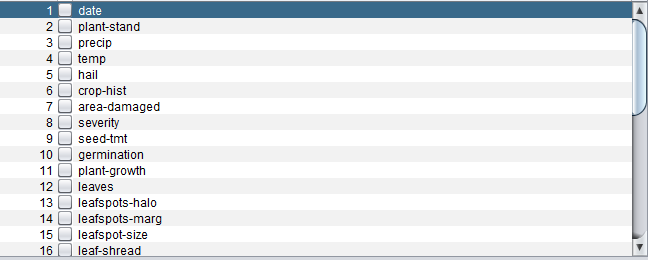
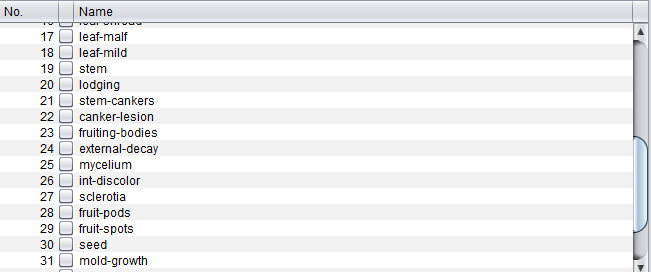
**Previous Results Accuracy**

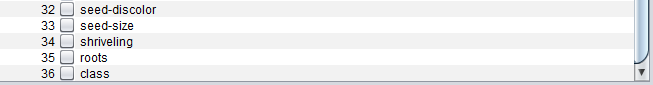
****

**Final Accuracy**

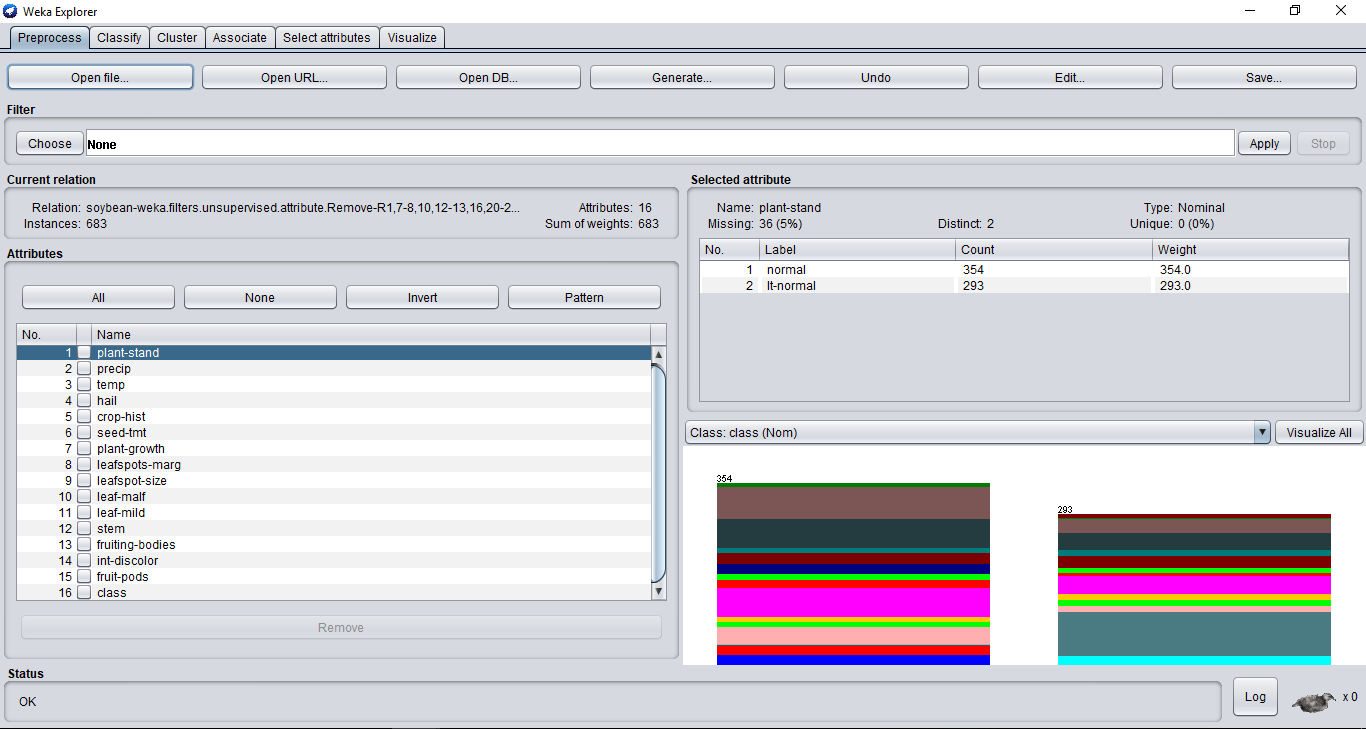
****

**Previous number of Attributes**

****

****

**Final Attributes**

****