IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PLANT DISEASE IDENTIFICATION USING MACHINE LEARNING ALGORITHMS

BORRA PRASANTHI [#1], V.SARALA [#2]

[#1] MCA Student, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

[#2] Assistant Professor, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

Abstract

The Machine Learning (ML) field has gained its momentum in almost any domain of research and just recently has become a reliable tool in the medical domain. Identification of the plant diseases is the key to prevent the losses in the yield and quantity of the agricultural product. The studies of the plant diseases mean the study of visually observed patterns seen on the plant. Health monitoring and disease detection on the plant is very critical for the substantial growth. It is very difficult to identify the diseases on the plant manually and provide the treatment for that appropriate disease. It requires a tremendous amount of work experience and should be expertise in the plant diseases and also requires excessive time for processing. Here in this proposed application we try to find out the disease of the plant based on the inputs which we observe physically on any plant. For any plant there are 5 levels of disease occurrences: Stem Level, Leaves Level, Seed Level,Lessions Level, Plant Level. So any disease on the plant can be either of these five levels. We try to design an medical dictionary in which all the physical inputs are substituted according to any of the level and then try to detect which disease plant is suffered with and it will try to provide cure for that appropriate disease. Our evaluation results on the proposed method using ML approach for identifying diseases on plant able to identify the diseases accurately and try to provide a solution for the end users.

Keywords: Plant Diseases, Machine Learning, Medical Diseases

1. INTRODUCTION

People care deeply about their health and want to be, now more than ever, in charge of their health and healthcare. Life is more hectic than has ever been, the medicine that is practiced today is an Evidence-Based Medicine (hereafter, EBM) in which medical expertise is not only based on years of practice but on the latest discoveries as well. Tools that can help us manage and better keep track of our health such as Google Health and Microsoft Health Vault are reasons and facts that make people more powerful when it comes to healthcare knowledge and management. The traditional healthcare system is also becoming one that embraces the Internet and the electronic world. Electronic Health Records (hereafter, EHR) are becoming the standard in the healthcare domain. Researches and studies show that the potential benefits of having an EHR system are

1. Health information recording and clinical data repositories— immediate access to patient diagnoses, allergies, and lab test results that enable better and time-efficient medical decisions.
2. Medication management—rapid access to information regarding potential adverse drug reactions, immunizations, supplies, etc.
3. Decision support—the ability to capture and use quality medical data for decisions in the workflow of healthcare
4. Obtain treatments that are tailored to specific health needs—rapid access to information that is focused on certain topics.
5. First the symptoms provided by the user are processed by a expert system for identifying the diseases. If the rules required for processing the data by the above are not present in the database, then the system automatically calls the machine learning algorithm technique.

AIM OF THE PROJECT

In order to embrace the views that the EHR system has, we need better, faster, and more reliable access to information. In the medical domain, the richest and most used source of information is Medline, a database of extensive life science published articles. All research discoveries come and enter the repository at high rate, making the process of identifying and disseminating reliable information a very difficult task. The work that we present in this paper is focused on two tasks: automatically identifying sentences published in medical abstracts (Medline) as containing or not information about diseases and treatments, and automatically identifying semantic relations that exist between diseases and treatments, as expressed in these texts. The second task is focused on three semantic relations: Cure, Prevent, and Side Effect

The tasks that are addressed here are the foundation of an information technology framework that identifies and disseminates healthcare information. People want fast access to reliable information and in a manner that is suitable to their habits and workflow. Medical care related

Information is a source of power for both healthcare providers and laypeople. Studies reveal that people are searching the web and read medical related information in order to be informed about their health. show how a new outbreak of the influenza virus can be detected from search engine query data. Our objective for this work is to show what Natural Language Processing (NLP) and Machine Learning (ML) techniques—what representation of information and what classification algorithms—are suitable to use for identifying and classifying relevant medical information in short texts.

We acknowledge the fact that tools capable of identifying reliable information in the medical domain stand as building blocks for a healthcare system that is up-to-date with the latest discoveries. In this research, we focus on diseases and treatment information, and the relation that exists between these two entities. Our interests are in line with the tendency of having a personalized medicine, one in which each patient has its medical care tailored to its needs. It is not enough to read and know only about one study that states that a treatment is beneficial for a certain disease. Healthcare providers need to be up-to-date with all new discoveries about a certain treatment, in order to identify if it might have side effects for certain types of patients.

We envision the potential and value of the findings of our work as guidelines for the performance of a framework that is capable to find relevant information about diseases and treatments in a medical domain repository. The results that we obtained show that it is a realistic scenario to use NLP and ML techniques to build a tool, similar to an RSS feed, capable to identify and disseminate textual information related to diseases and treatments. Therefore, this study is aimed at designing and examining various representation techniques in combination with various learning methods to identify and extract biomedical relations from literature.

2. LITERATURE SURVEY

RELATED WORK

In 2011, an innovative approach was presented[1] to automatically grade the disease on plant leaves. According to that, plant pathologists mainly rely on naked eye prediction and a disease scoring scale to grade the disease. That leads some problems associated with manual grading This manual grading is not only time consuming but also not feasible. Hence an image processing-based approach to automatically grade the disease spread on plant leaves by employing Fuzzy Logic had been proposed. The results are proved to be accurate and satisfactory in contrast with manual diseases are inevitable in plants. The proposed methodology aims to model a promising disease grading system for plant leaves. The system was divided into the following steps:

(1) Image acquisition

(2) Image Pre-processing

(3) Color image segmentation

(4) Calculating AT and

(5) Disease grading by Fuzzy Logic.

In 2014, an survey report was published[2], based on different classification techniques that could be used for plant leaf disease classification. A classification technique deals with classifying each pattern in one of the distinct classes. A classification is a technique where leaf is classified based on its different morphological features. There are so many classification techniques such as k-Nearest Neighbor Classifier, Probabilistic Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis, Artificial neural network, Fuzzy logic. Selecting a classification method is always a difficult task because the quality of result can vary for different input data. Plant leaf disease classifications have wide applications in various fields such as in biological research, in Agriculture etc. The paper provides an overview of different classification techniques used for plant leaf disease classification.

In 2012, an article was published[3] having a detailed description on definition of disease, types of diseases, symptoms and causes of most commonly observed plant diseases One article was published by Michigan University[4] regarding the threats caused due to diseases. Various conditions for disease development had been discussed there. An overview of major disease-causing organisms and the effect of diseases caused by them was given.

Deep neural networks have recently been successfully applied in many diverse domains as examples of end to end learning. Neural networks provide a mapping between an input—such as an image of a diseased plant—to an output—such as a crop disease pair. The nodes in a neural network are mathematical functions that take numerical inputs from the incoming edges, and provide a numerical output as an outgoing edge. Deep neural networks are simply mapping the input layer to the output layer over a series of stacked layers of nodes. The challenge is to create a deep network in such a way that both the structure of the network as well as the functions (nodes) and edge weights correctly map the input to the output. Deep neural networks are trained by tuning the network parameters in such a way that the mapping improves during the training process.

3. EXISTING SYSTEM

In Existing System, there is no concept like identifying the plant diseases based on ML Approach or using data mining algorithms. All the existing methods try to utilize manual method for identifying the plant disease and those disease identification may sometimes accurate and sometimes in efficient.

LIMITATION OF EXISTING SYSTEM

The following are the limitation of the existing system :

1) All the existing methods try to use manual method to identify the disease and provide cure for that disease.
2) A lot of time is consumed in identifying and verifying the type of disease on the plant.
3) Outcome may be accurate or sometimes in accurate.
4) A lot of research knowledge should be their in order to judge the disease on that plant based on the symptoms observed on that plant.
5) There is no system to identify and predict the disease using ML approach in the existing system
6) All the existing methods try to exhaust a lot of time in identifying and provide cure for the infected plant.

4. PROPOSED SYSTEM

In proposed System, we try to implement ML Approach in data mining to identify the plant diseases. The proposed method tries to create a medical dictionary for the end users in order to predict and identify the diseases very accurately on plant.

ADVANTAGES OF THE PROPOSED SYSTEM

The following are the advantages of the proposed system :

1) The proposed system try to use ML approach to identify the disease and provide cure for that disease.
2) It will not exhaust lot of time for identifying and predicting the disease of that infected p76 1 lant.
3) Here we can able to categorize the plant into 5 levels for disease prediction and based on physical appearance of that plant, we can able to choose appropriate part to identify the disease name and provide cure according to that.
4) Here the outcome or end result will always be accurate
5) There is no need to having any previous research knowledge to judge the disease of that plant, the user just need to know how to find the problem on that plant by physical appearance.
6) The proposed system is very accurate in identifying the type of disease in one of the 5 levels

5. SOFTWARE PROJECT MODULES

Implementation is the stage where the theoretical design is converted into programmatically manner. In this stage we will divide the application into a number of modules and then coded for deployment. We have implemented the proposed concept on Java programming language with JEE as the chosen language in order to show the performance this proposed protocol. The proposed application is mainly divided into 5 modules. They are as follows:

1. Load Disease Data Set
2. User Module
3. Machine Learning Technique
4. Identify Disease using Rule Based Expert System
5. Identify Disease using Machine Learning Algorithm

Now let us discuss about each and every module in detail as follows:

5.1 Load Plant Dataset Module

This is a predefined task which is done by the administrator in order to maintain a proper data set to classify the plant diseases. Whenever a new plant disease is invented, then that disease details should be maintained into the database by the administrator. This is nothing but collecting training data set information for the naïve bayes classification algorithm.

5.2 User Module

In this module the user is one who try to verify the diseases on a plant. he try to identify all the infected areas of that plant physically and then try to substitute those symptoms on that plant library. Here those symptoms which he found will be choose as yes and remaining those symptoms which he didn't observe on that plant will be selected as No. So once after choosing all the values the corresponding inputs is send to the ML approach. Here the user try to give test inputs for the Naïve Bayes classification algorithm. These test values should be matched with training data set and then probability of disease is identified by the ML approach.

5.3 Machine Learning Technique Module

In this module, the predefined disease data sets and user inputs are to be learned by the machine (computer). Machine learning is the study of how to make computers learn; the goal is to make computers improve their performance through experience. This ML is mainly used in clustering the diseases based on the type of symptoms. As we all know that all plants may not suffer with same type of disease and same level of complaints. So based on the individual problem ,the ML system need to guide the user to take cure on those conditions.

5.4 Rule Based Expert Module

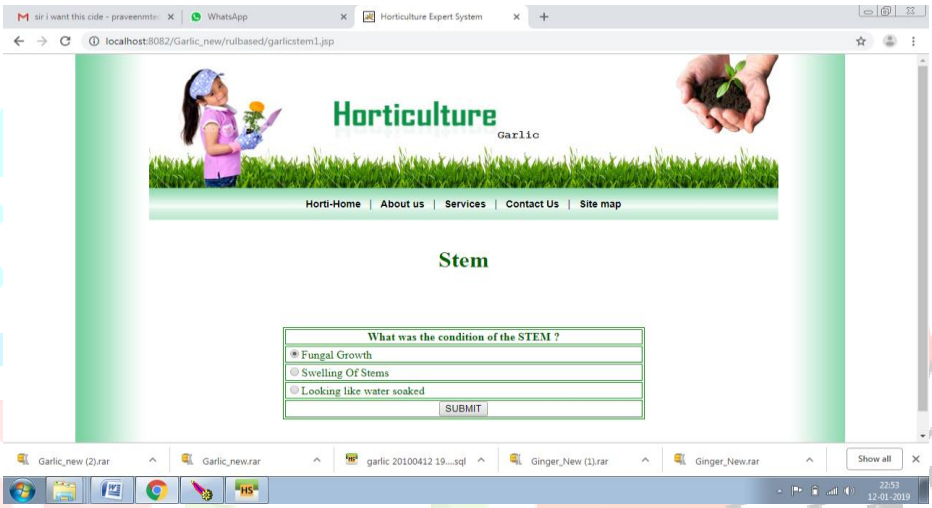
Here the diseases can be identified based on rule based expert system which means user need to input the symptoms based on the one of the 5 levels and after the user choose the level and enters all the symptoms then only the user will get the appropriate disease name based on expert knowledge. Here if the user choose all symptoms as null, then the resultant output will be displayed as no disease for that plant. If the same user try to

choose appropriate inputs then based on that the disease will be predicted and cure will be provided for that user.

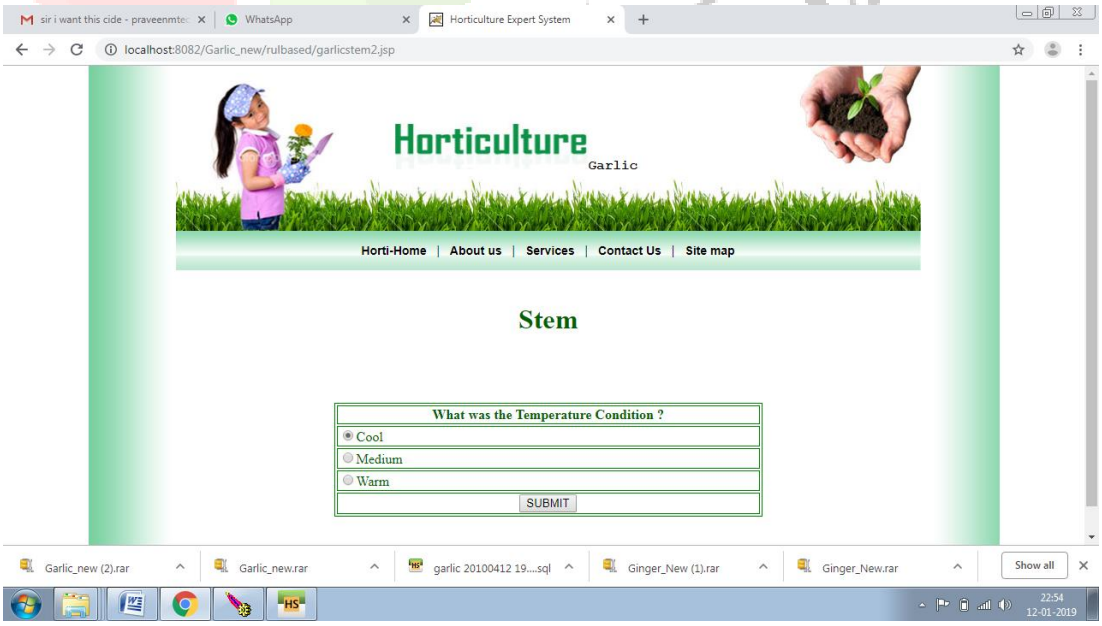
5. 5 Machine Learning Based Module

Here this module clearly tells that there is no need to choose individual category and then input the symptoms. If the user find some common symptoms which can be generally visualized and identified. Those symptoms he try to choose from this common list of attributes and based on those fields ,the ML approach will decide which type of disease the plant suffer from and how much percentage of infection occurred to the plant. We can calculate the percentage of infection on that plant.

6. OUTPUT RESULTS



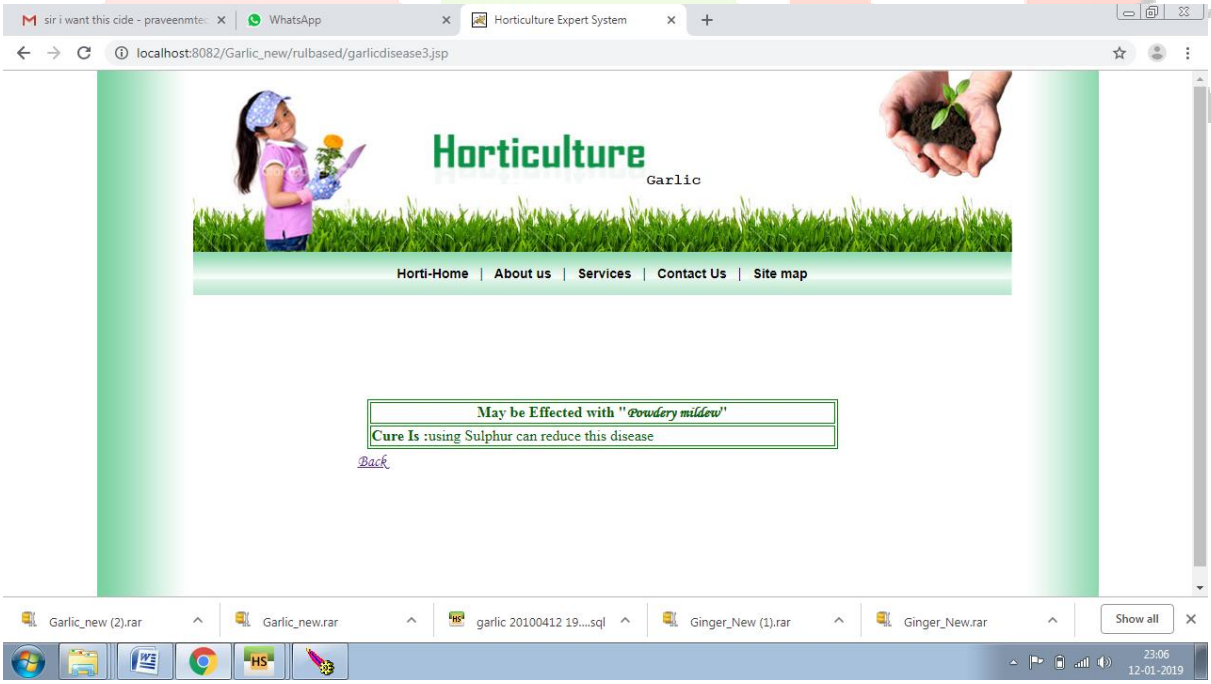
HERE FOR EXAMPLE WE CHOOSE FUNGAL GROWTH ON THE STEM



Then from the above window we can choose what is the temperature condition in that region

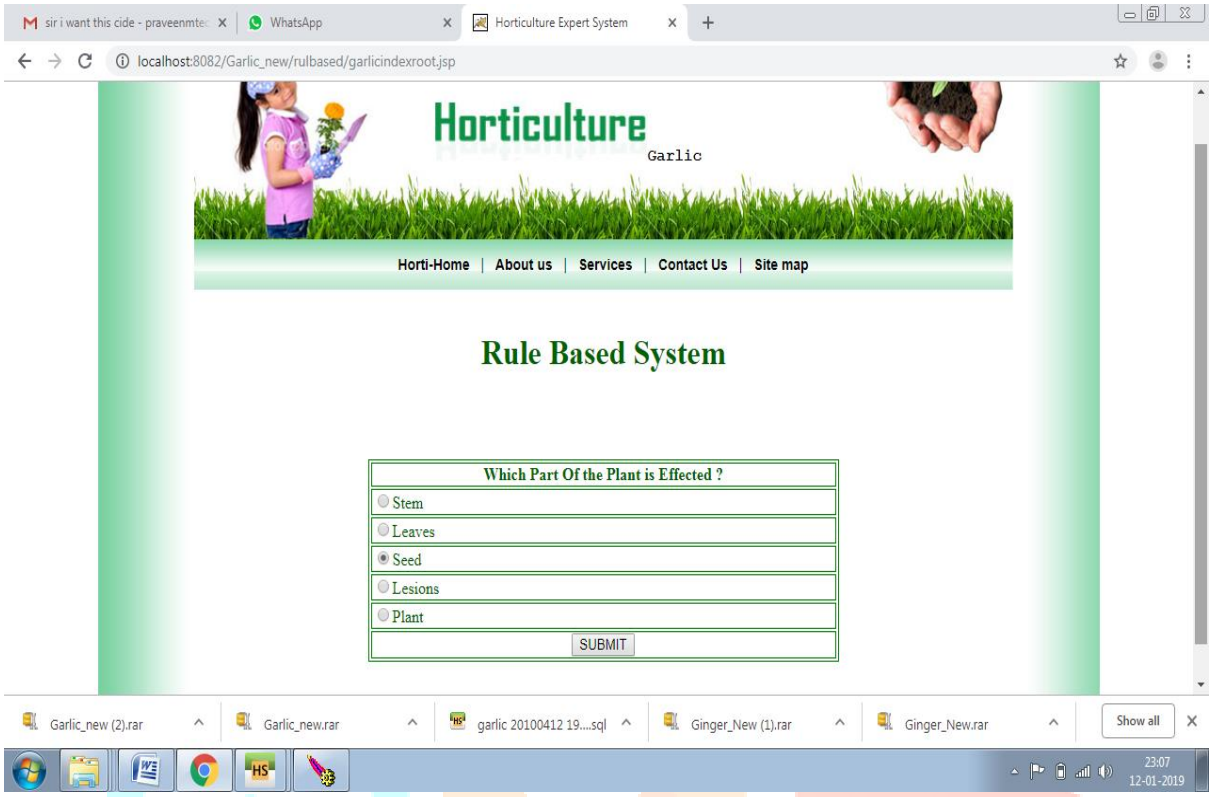
We have choose medium temperature

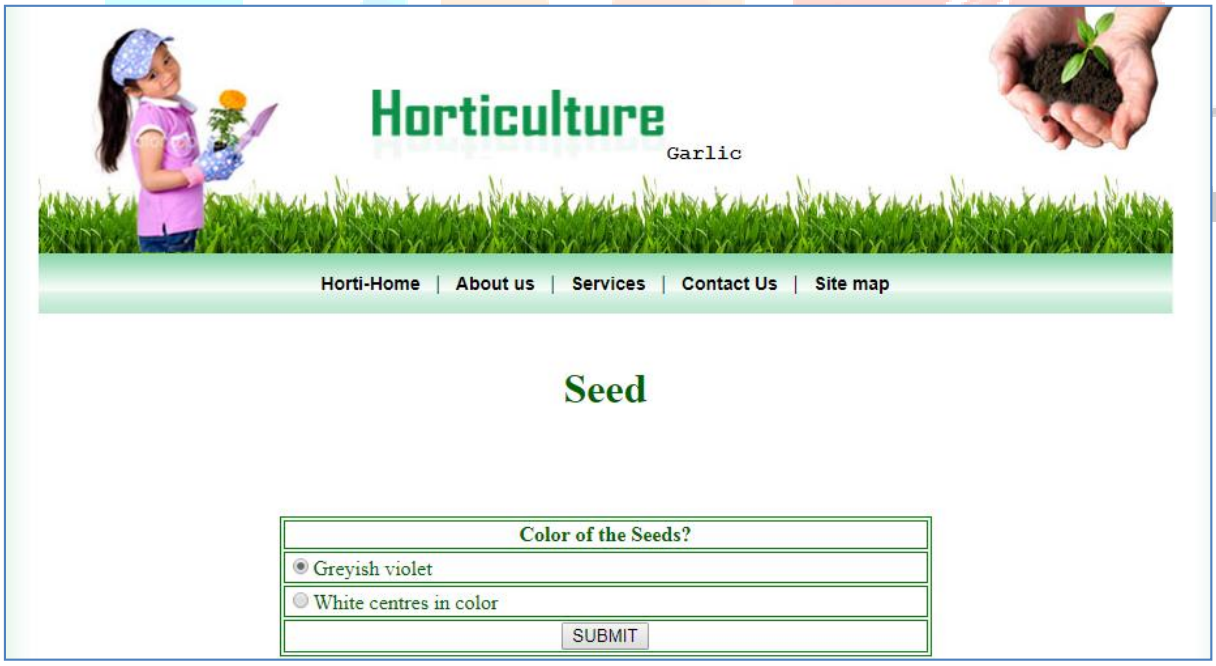
Now the expert system gives the suggestion as below:



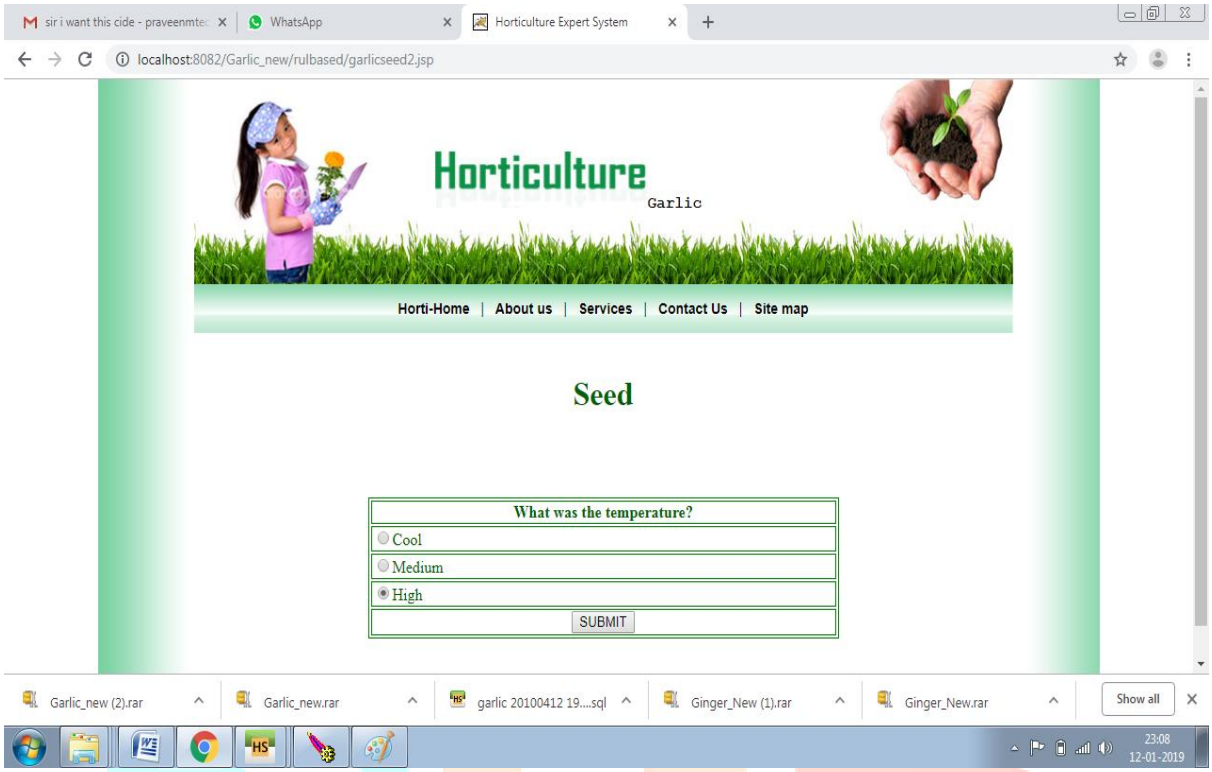
In the same way we can check other attributes of the plant like

SEED LEVEL

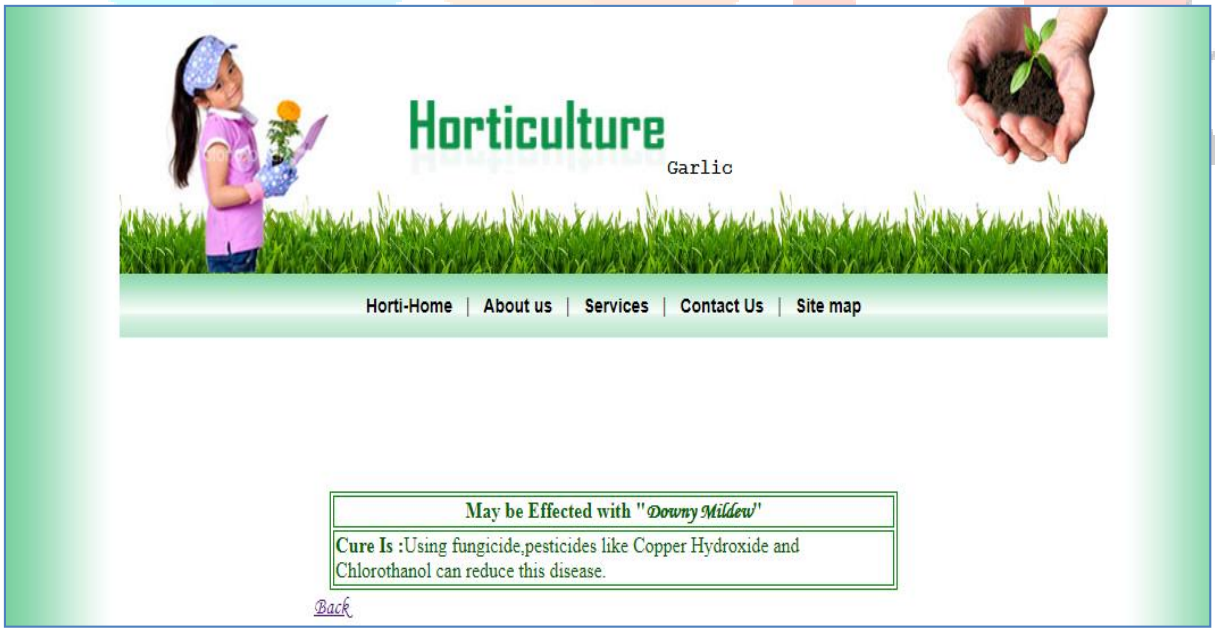

sir i want this cide - praveenmtec
WhatsApp
Horticulture Expert System
localhost:8082/Garlic_new/rulbased/garlicindexroot.jsp
Horticulture
Garlic
Horti-Home | About us | Services | Contact Us | Site map
Rule Based System
Which Part Of the Plant is Effected ?
Stem
Leaves
Seed
Lesions
Plant
SUBMIT
Garlic_new (2).rar
Garlic_new.rar
garlic 20100412 19....sql
Ginger_New (1).rar
Ginger_New.rar
Show all
23:07
12-01-2019


Horticulture
Garlic
Horti-Home | About us | Services | Contact Us | Site map
Seed
Color of the Seeds?
Greyish violet
White centres in color
SUBMIT

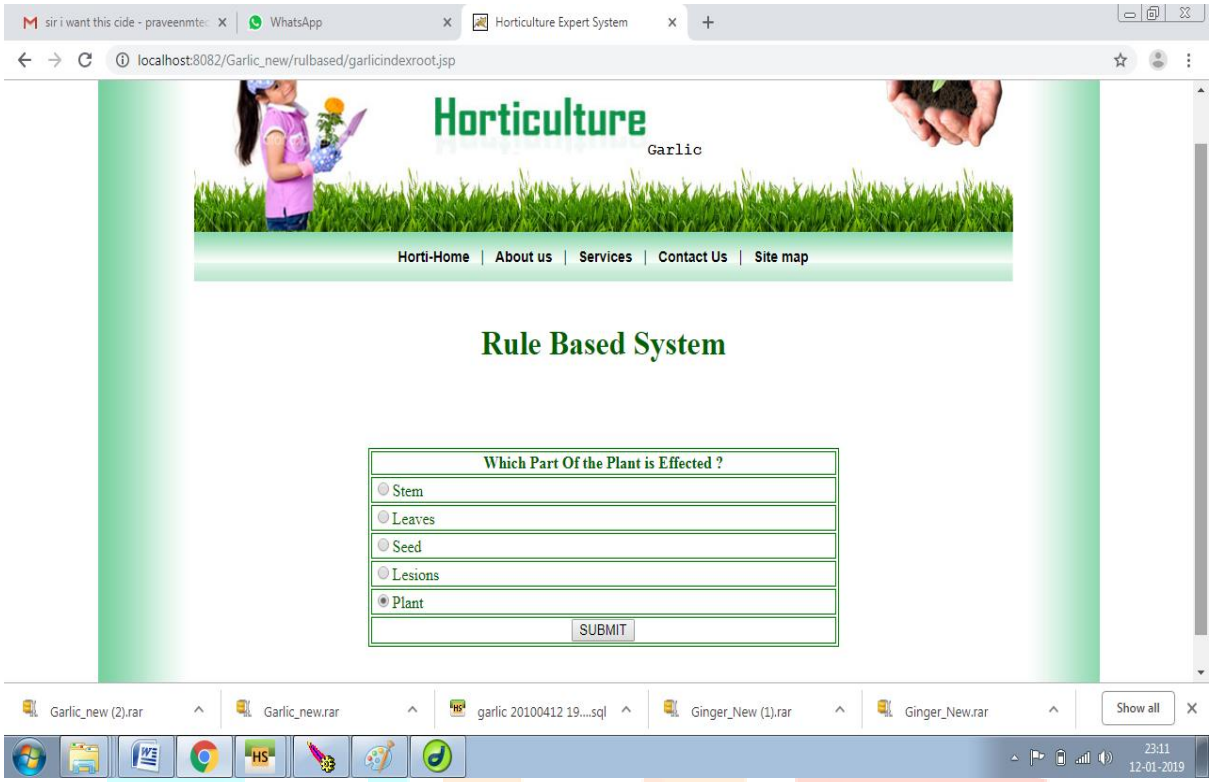
NOW IT WILL ASK THE TEMP



NOW THE EXPERT SYSTEM SUGGEST AS FOLLOWS:



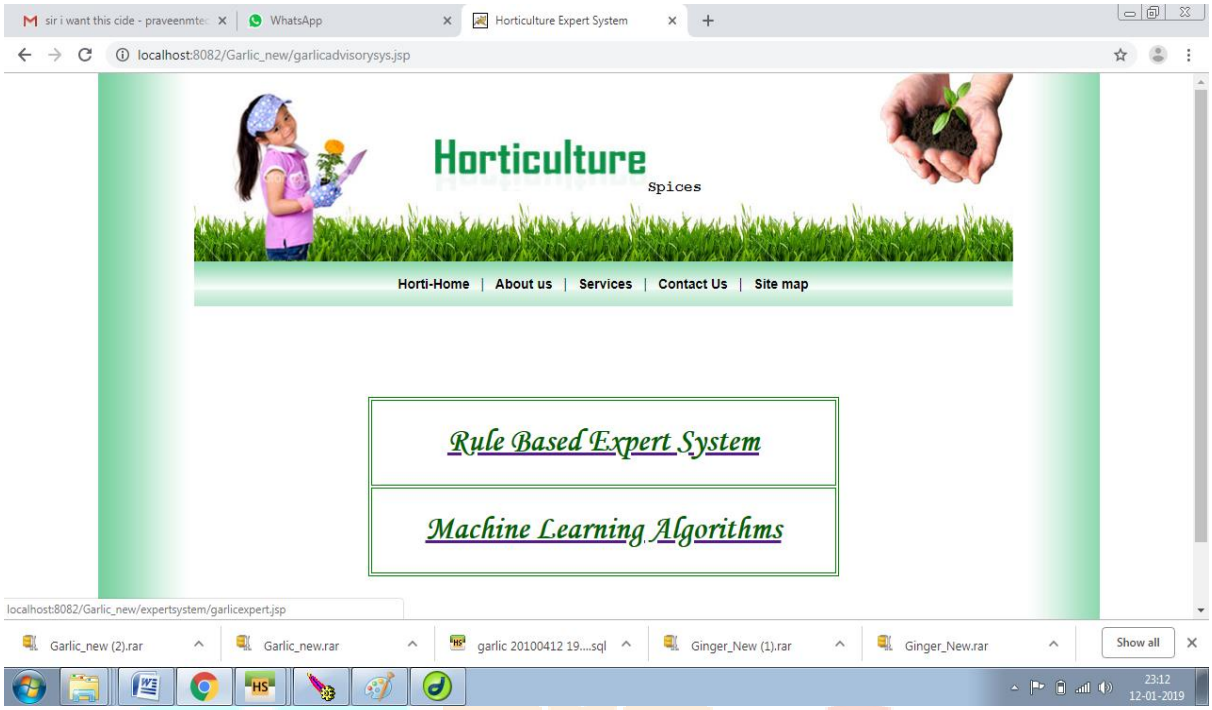
NOW WE CAN CHECK PLANT LEVEL




sir i want this cide - praveenmtec
WhatsApp
Horticulture Expert System
localhost:8082/Garlic_new/rulbased/garlicplant2.jsp
Horticulture
Garlic
Horti-Home | About us | Services | Contact Us | Site map
Plant
What was the Temperature Condition ?
Cool
Medium
Warm
SUBMIT
Garlic_new (2).rar
Garlic_new.rar
garlic 20100412 19....sql
Ginger_New (1).rar
Ginger_New.rar
Show all
23:11
12-01-2019


Horticulture
Garlic
Horti-Home | About us | Services | Contact Us | Site map
May be Effected with "Pink root"
Cure Is :Control of this disease is by using at least a three- to four-year rotation without allium.
Back

NOW WE CAN CHECK THE MACHINE LEARNING BASED APPRAOCH



7. CONCLUSION

The conclusions of our study suggest that domain-specific knowledge improves the results. Probabilistic models are stable and reliable for tasks performed on short texts in the medical domain. The representation techniques influence the results of the ML algorithms, but more informative representations are the ones that consistently obtain the best results.The first task that we tackle in this paper is a task that has applications in information retrieval, information extraction, and text summarization. We identify potential improvements in results when more information is brought in the representation technique for the task of classifying short medical texts. We show that the simple BOW approach, well known to give reliable results on text classification tasks, can be significantly outperformed when adding more complex and structured information from various ontologism.

8. REFERENCES

[1] R. Bunescu and R. Mooney, "A Shortest Path Dependency Kernel for Relation Extraction," Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT/ EMNLP), pp. 724-731, 2005.

[2] R. Bunescu, R. Mooney, Y. Weiss, B. Scho¨ lkopf, and J. Platt, "Subsequence Kernels for Relation Extraction," Advances in Neural Information Processing Systems, vol. 18, pp. 171-178, 2006.

[3] A.M. Cohen and W.R. Hersh, and R.T. Bhupatiraju, "Feature Generation, Feature Selection, Classifiers, and Conceptual Drift for Biomedical Document Triage," Proc. 13th Text Retrieval Conf.(TREC), 2004.

[4] M. Craven, "Learning to Extract Relations from Medline," Proc. Assoc. for the Advancement of Artificial Intelligence, 1999.

[5] I. Donaldson et al., "PreBIND and Textomy: Mining the Biomedical Literature for Protein-Protein Interactions Using a Support Vector Machine," BMC Bioinformatics, vol. 4, 2003.

[6] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky,"GENIES: A Natural Language Processing System for the Extraction of Molecular Pathways from Journal Articles," Bioinformatics, vol. 17, pp. S74-S82, 2001.

[7] O. Frunza and D. Inkpen, "Textual Information in Predicting Functional Properties of the Genes," Proc. Workshop Current Trends in Biomedical Natural Language Processing (BioNLP) in conjunction with Assoc. for Computational Linguistics (ACL '08), 2008.

[8] R. Gaizauskas, G. Demetriou, P.J. Artymiuk, and P. Willett, "Protein Structures and Information Extraction from Biological Texts: The PASTA System," Bioinformatics, vol. 19, no. 1, pp. 135-143, 2003.

[9] C. Giuliano, L. Alberto, and R. Lorenza, "Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature," Proc. 11th Conf. European Chapter of the Assoc. for Computational Linguistics, 2006.

[10] J. Ginsberg, H. Mohebbi Matthew, S.P. Rajan, B. Lynnette, S.S. Mark, and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," Nature, vol. 457, pp. 1012-1014, Feb. 2009.