International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Decision Support System for Industries using Naïve Bayesian Classification with Laplace Smoothing

Y. Ratnavali[1], V. Valli Kumari[2]

[1] M. Tech, Computer Science and Technology, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam

[2] Professor, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam

Abstract*: Industries have a key role in rural development programmes. In most of the countries like India, industries play an important role in employment creation, resource utilization and income generation. Jute industry is mainly labour intensive industry. By establishing a jute industry employability of the village may increase. But establishing an industry in a particular place is a challenging task for entrepreneur. Entrepreneur has to take decision about the place to establish an industry, for this purpose he/she has to visit so many places for studying parameters. These parameters include availability of raw facts, labour, power, fuel, water, area, transportation and climatic conditions. A Decision Support System is required in such conditions, to guide entrepreneurs.*

Keywords: Jute industry, decision support system, Naive Bayesian classification, Laplace smoothing, industry establishment

1. Introduction

Industry is the production of goods or services within an economy. Now a days the main problems in country is income generation. Industrialization may increase the country income and employability. India has many favourable factors for the development of industries. The various favourable factors present in the country for the development of industries are:

- The country is rich in natural resources, such as minerals, forests etc.
- Commercial crops, such as sugar cane, raw jute etc. required for the development of agro-based industries.
- Human resources such as labour required for industries.

The establishment of manufacturing industries [2] depends upon factors like raw materials, water, manpower, area, transportation, market.

1.1 Internet based GIS in mapping industry information

Utilization of GIS and internet has changed the organizing way through the utilization of geographic information accessing process, sharing, and data analysis. Therefore there are many benefits from using GIS and internet online. User makes a request to DSS by selecting mandal and industry and it gives the suitable areas to establish an industry. Each area will be displayed in map with the factors available in that area such as raw material available, manpower, water, land available etc.

1.2 Google map API

Google map is free and online virtual globe map service provided by Google, Google map offer map which can be copied and real images of the earth from satellites all around the earth, moon and also provides travel route system of the selected areas. Google map API is free for commercial use, can be accessed by public and there is no charge for the access. Therefore, the use of Google map API is very useful in developing interactive map application.

2. Problem Statement

Over the years, rural people in India depending on indigenous or local knowledge for starting industry. Such knowledge refers to skill and experience gained through oral tradition and practice over many generations. After gaining knowledge from skilled domain experts, taking decision is very difficult. By manual procedure some parameters may get omitted and there may be a chance of the facts getting suppressed. To avoid the above problems we need decision support system which can integrate the data and give decisions to users based on factors like raw material availability, man-power, transportation facilities, water, market etc.

3. Proposed Method

The proposed system mainly focuses on minimizing the time taken by the domain experts in analyzing the gathered data. A decision support system is developed at mandal level for analyzing the gathered data and providing list of suitable villages for the establishment of jute industry, pointing those villages in Google maps and also decision support system is developed at village level on entrepreneur requisites and providing decision about whether the particular village is suitable or not.

DSS is „an interactive, flexible and adaptable computer-based information system“ and is developed to provide solution for non-structured management problem [6] .

In DSS, data mining technique has been applied on village database and industry database to draw useful results which will form the basis for decision making. Data mining technique like Naïve Bayesian classification with Laplacian smoothing [4]. Village database contains the information about village like population, marginal workers, raw material

available, water availability, waste land, and latitude, longitude values of a village. By using latitude longitude values DSS point out the villages in Google map. Industry database contains the information about the minimum requirements of the industry.

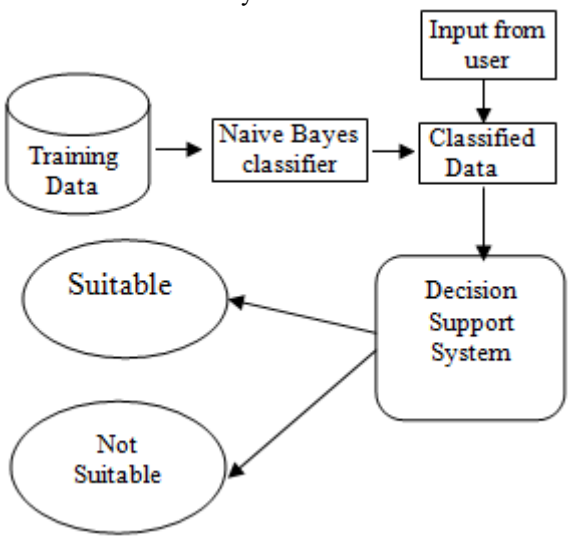

Figure 1: Flow diagram for the Decision support System for establishing an industry

Entrepreneur needs a DSS in decision making process of establishing a jute industry. DSS takes the input from the user and it is trained by machine learning algorithm i.e Naïve Bayesian classification with Laplace smoothing. Every classification algorithm requires training data to assign class label to newly arrived model. DSS consider the input as newly arrived model and applies classification algorithm then assign class label to input data. Based on class label it gives the output to entrepreneur as suitable or not suitable.

3.1 Algorithm

Naive classifier [3] is a term dealing with probabilistic classifier based on applying bayes theorem with strong independence assumptions. It assumes that the presence or absence of particular feature of class is unrelated to the presence or absence of any feature.

The naive bayes algorithm based on conditional probabilities. It uses Bayes theorem [7], a formula that calculates a probability by counting the frequency of values and combination of values in historical data. Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred.

If the classifier encounters a feature value that has not been present in training data, the probability of both the classes becomes zero because of zero probability accuracy may decrease. To improve the accuracy, naïve Bayesian classification algorithm implemented with smoothing method i.e, Laplacian Smoothing [5].

Calculate the probability using Bayes theorem as follows

$$P(C_j|X) = \frac{P(X|C_j) * P(C_j)}{P(X)}$$

$P(C_j/X)$ = probability of instance X being in class C_j.
$P(X/C_j)$ = probability of generating instance X given class C_j.
$P(C_j)$ = probability of occurrence of class C_j.
$P(X)$ = probability of instance X occurring.

Steps:-

1. Each data sample is represented by an n dimensional feature vector, $X=(x_1,x_2,x_3,\ldots x_n)$ depicting n measurements made on the sample from n attributes, respectively $A_1,A_2,A_3,\ldots,A_n$.
2. Suppose that there are m classes, $C_1,C_2,\ldots..,C_n$. Given an unknown data sample, X(i.e. having no class label), the classifier will predict that X belongs to class having highest posterior probability, conditioned on X. That is, the Naïve probability assigns unknown sample X to the class C_i.

If and only if :
$P(C_i/X)>P(C_j/X)$ for all $1<=j<=m$ and $j!=i$.
Thus we maximize $P(C_i/X)$. The class C_i for which $P(C_i/X)$ is maximized is called the maximum posterior hypothesis.
By Bayes theorem,

$$P(C_j|X) = \frac{P(X|C_j) * P(C_j)}{P(X)}$$

3. P(X) is constant for all classes only $P(X/C_j)*P(C_j)$ needed to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e. $P(C_1)=P(C_2)=\ldots..=P(C_n)$, and we should therefore maximize $P(X/C_i)$ otherwise we maximize $P(X/C_i)*P(C_i)$. Note that the class prior probabilities may be estimated by $P(C_i)=S_i/s$ where S_i is number of training samples of class C_i, and s is the total number of training samples.

4. Posterior probability mainly depends on the value of $P(X/C_i)$. With the Naïve Bayes Assumption, we can still end up with zero probability. NB requires each conditional probability to be non-zero, hence, Laplace smoothing method is usually used to avoid zero-probability values:

$$P\left(\frac{X}{C_j}\right) = \frac{1 + N(X, C_j)}{K + N(C_j)}$$

Where
$N(X,C_j)$ is Number of occurrences of attribute X in C_j.
K is the total number of possible values of X.
$N(C_j)$ is the Number of data touples belongs to class C_j.

4. Results and Discussions

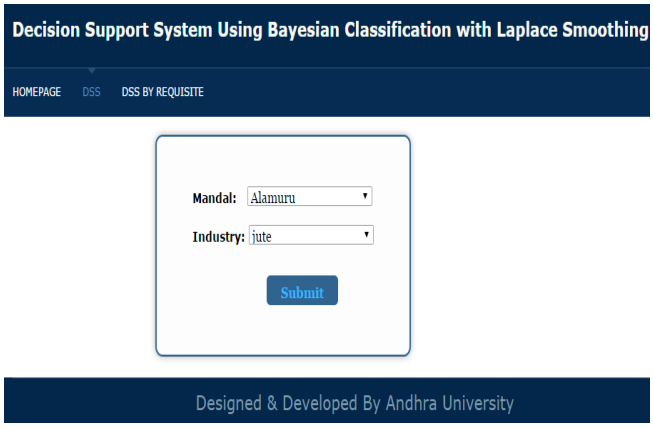

Figure 2: Input for DSS at mandal level

DSS developed for the villages in the East Godavari district. In east Godavari district, there are 59 mandals and 1370 villages. Entrepreneur wants to start a jute industry he/she has to visit all the villages. It takes lot of time as well as cost effective. Entrepreneur gives the mandal and industry as input to DSS.

Figure 3: Output for DSS at mandal Level

By taking input from the entrepreneur, DSS applies Naïve Bayesian classification with Laplace smoothing on input data then it gives the list of villages as output that are suitable to start a jute industry. DSS considers the parameters of all villages in a particular mandal and industry requirements. Based on those parameters DSS gives the list of villages that are suitable to start a jute industry and those villages are pointed in the google maps. Google maps helps the user to visualize the route map, transportation facilities of a particular village. DSS will give all the details regarding a village when the cursor is placed on marker of the map.

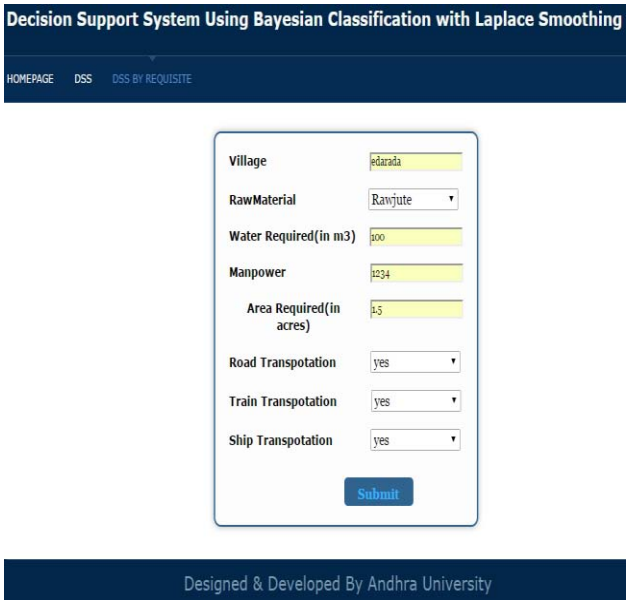

Figure 4: Input for DSS at village level

DSS by Requisite module used by the entrepreneur has some constraints and he/she has to given those constraints to DSS as input based on constraints DSS will give decision about suitability of place to establish an industry.

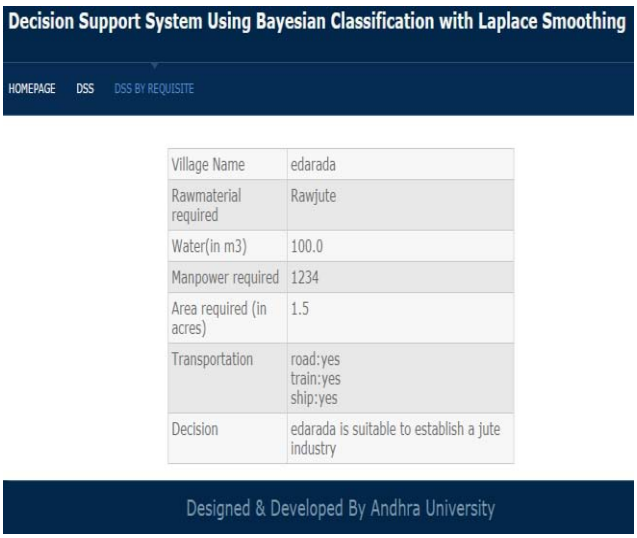

Figure 5 output for DSS at Village Level

Naive Bayesian classifier with Laplace smoothing is applied on input data and it assigns the class label to input data. DSS gives decision about suitability of establishment of jute industry. Based on that decision entrepreneur will take decision.

5. Conclusion

In this paper a Decision Support System for establishing a jute industry is presented. Where the system guides the user (Entrepreneur) to make a decision about establishment of jute industry in villages of East Godavari. This Decision Support System is user friendly as it provides an interactive web interface which takes the user through series of screens that provide the user with supporting decision depending on the user selection like mandal, industry in DSS and village name, raw material, transportation, water, area required in DSS by Requisite.

6. Future scope

The proposed system can be enhanced in the following ways:

- Implementation of Decision Support System for several types of industries.
- It can be scaled up entire state by using Big Data analytics.

References

[1] http://www.indianmirror.com/indianindustries/jute.html
[2] Factors influencing location of jute industry.
[3] http://www.yourarticlelibrary.com/industries/factors-influencing-the-location-of-industries-geographical-and-non-geographical-factors/19695/

[4] D. Ratnam and P. Hima bindhu, "Computer based clinical decision support system for prediction of heart diseases using naive Bayesian classification"
[5] Giudici, P.: "Applied Data Mining: Statistical Methods for Business and Industry", New York: John Wiley, 2003.
[6] Vivek Narayanan1 , Ishan Arora2 , Arjun Bhatia3 "Fast and Accurate Sentiment Classification Using an Enhanced Naïve Bayes Model"
[7] Randall E. Louw, Decision Support System
[8] Kevin P. Murphy, Naive Bayes classifiers.
[9] http://www.yourarticlelibrary.com/industries/production-and-distribution-of-jute-textiles-industry-in-india/19699

Author Profile

Yerubandi Ratnavali received B. Tech. degree in Computer Science and Engineering from Swarnandhra Institute of Engineering and Technology in 2013 and currently pursuing Master of Technology in Computer science and Technology from the Department of Computer Science and Systems Engineering, Andhra University Visakhapatnam.