

# AUTOMATED RESUME EXTRACTION AND CANDIDATE SELECTION SYSTEM

**Sneha Kumari<sup>1</sup>, Punam Giri<sup>2</sup>, Swati Choudhury<sup>3</sup>, S.R. Patil<sup>4</sup>**

<sup>1, 2, 3</sup>Student, <sup>4</sup>Lecturer, Department of Computer Engineering, University of Pune, DYPSOE, Lohegaon, Pune, Maharashtra, India

### Abstract

*Automated Resume Extraction and Candidate Selection System (ARE & CSS) is a product which can be best suited for any organization's recruitment process. The system will be robust enough which will automatically extract the resume content and store it in a structure form within the Data Base. Classification algorithm (Naïve Bayes) will be run on the profiles to identify profile categories or classes. Also the employer can specify his criteria and also decide the importance level.*

**Keywords:** -Naïve Bayes, Clustering, Classification, Data Mining, Preprocessing.

-----\*\*\*-----

## 1. INTRODUCTION

Automated Resume Extraction and Candidate Selection System will be built on Google's Cloud. Large enterprises and head-hunters receive several thousands of resumes from job applicants every day. HRs and Managers go through a hundreds of resumes manually. Resumes or Profiles are unstructured documents and have typically number of different formats (eg: .doc, .pdf, .txt). As a result manually reviewing multiple profiles is a very time consuming processes. How to ensure you have the Appropriate Candidate in the right jobs at the right time. This is a significant problem faced by large companies today in the market.

Now a day's many job portals are available but the basic problem in available system are it required manual efforts for both candidates and Employers. Candidate has to provide complete information in given text filed and employer also needs to apply many filters to select the candidate. Even though Employer has applied many filters he would get thousands of resume even going through it and selecting candidates is very inefficient and time consuming task. Some costly extraction systems are available in the market that also do the search on keyword basis and has many extraction limitations like forcing candidates to fill templates and keep updating the templates as per job profiles.

## 2. LITERATURE SURVEY

### 2.1 Naïve Bayes Classifier

Here we are going to discuss NAÏVE BAYES Classifier which includes:

- Strong independence (naïve) assumptions
- Bayes' theorem

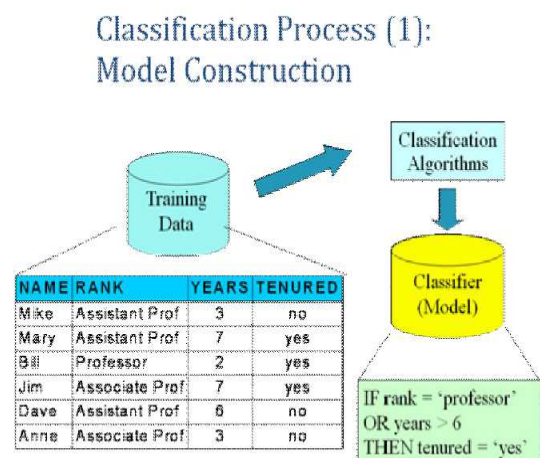
- Prior and Posterior probability
- How naïve bayes classifier is used in classification of resumes

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (or Bayes' rule) with strong independence (naïve) assumptions.

### 2.2 Classification—A Two-Step Process

**Model Construction:** Describing a set of predetermined classes.

Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute.

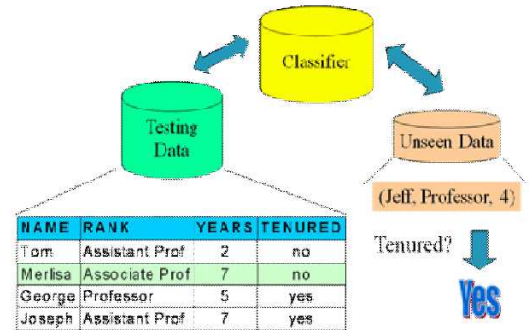


The set of tuples used for model construction: training set.

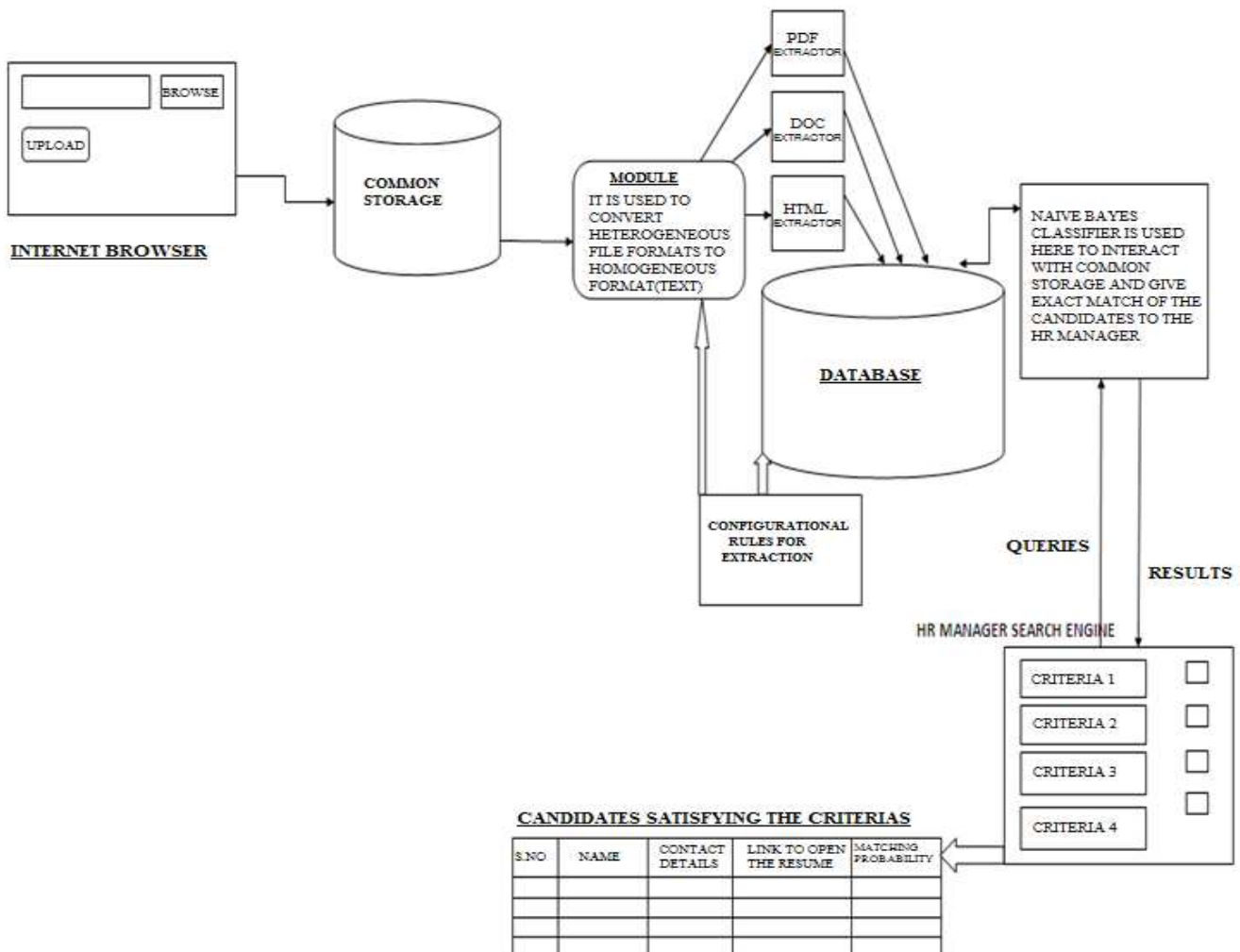
The model is represented as classification rules or mathematical formula.

**Model Usage:** for classifying future or unknown objects. Estimate accuracy of the model. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur.

Classification Process (2): Use the Model in Prediction



### 3. PROPOSED SYSTEM



#### 4. ADVANTAGES

This system provides time efficient and very effective candidate selection mechanism. It is highly customizable as employer can specify their criteria along with importance level. It is easy for user as they just need to upload their resumes on portal. No form filling is required. Automatic Email notification to candidate / employers can be possible.

#### 5. DISADVANTAGES

As these applications need internet it requires GPRS connection every time. Bad weather and poor Internet connection can affect the system. If there is no database connection then the system will not function.

#### 6. FUTURE SCOPE

Automated Resume Extraction and Candidate Selection System basically extracts all the information about the candidate only through his/her resume ,without forcing the candidates to fill any other information about them. After extraction it stores the information in a centralized data base, allowing the HR Managers to search in the data base for their criteria satisfying candidates. There can be future enhancements like

1. The HR can have a video conference with the candidate in order to take his/her interview.
2. The candidates can also appear for online aptitude test for practice
3. The employees can give reviews of the company they are working in.

#### CONCLUSIONS

Here we are providing a unique system which is robust enough to automatically extract the resume content and store it in a structure form within the Data Base. This system will make the task of both candidate and HR Manager easier and faster. This system avoids the hectic form filling procedure of the candidates by directly asking the user to upload only the resume. The HR Manager also just need to fill his/her criteria instead of manually going through all the resumes.

#### ACKNOWLEDGEMENTS

We would like to sincerely thank Mr. S.R Patil, our guide, for his support and encouragement.

#### REFERENCES

- [1].Jongwoo Kim, Daniel X. Le, and George R. “Naïve Bayes Classifier for Extracting Bibliographic Information from Biomedical Online Articles”, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA
- [2]. Ajay S. Patil, B.V. Pawar “Automated Classification of Web Sites using Naive Bayesian Algorithm”, Proceedings of the International MultiConference of Engineers and Computer

Scientists 2012 Vol I,IMECS 2012, March 14-16,2012,Hong Kong

[3]. Md. Faisal Kabir “Enhanced Classification Accuracy on Naive Bayes Data Mining Models”, International Journal of Computer Applications (0975 – 8887) Volume 28– No.3, August 2011

[4]. Mauricio A. Valle , Samuel Varas , Gonzalo A. Ruz “Job performance prediction in a call center using a naive Bayes classifier”, Facultad de Ciencias Económicas y Administrativas, Universidad de Valparaíso, Santiago, Chile,2011

[5].Gla´ucia M.Bressan “Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop ”, UniversidadedeSaõ

Paulo,DepartamentodeEngenhariaEle´trica, 13566-590Saõ Carlos,SP,Brazil 2009

[6]. S.L. Ting, W.H. Ip, Albert H.C. Tsang. “ Is Naïve Bayes a Good Classifier for Document Classification?”, International Journal of Software Engineering and Its Applications Vol. 5, No. 3, July, 2011

[7].YasinUzun“ Keyword Extraction Using Naive Bayes”, Bilkent University, Department of Computer Science, Turkey, yasinu@cs.bilkent.edu.tr

[8].Binal A. Thakkar, Mosin I. Hasan, Mansi A. Desai ”Health Care Decision Support System For Swine Flu Prediction Using Naïve Bayes Classifier”, International Conference on Advances in Recent Technologies in Communication and Computing, india, 2010

[9].[Online]

[http://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](http://en.wikipedia.org/wiki/Naive_Bayes_classifier)