Moocs Video Mining Using Decision Tree J48 and Naive Bayesian Classification Models

EL HARRAK Othman[1], GHADI Abderrahim[2], EL BOUHDIDI Jaber[3]
[1,2]FST, UAE, TANGIER / [3]ENSA, UAE, TETUAN

Abstract— **Nowadays, the internet has become the first source of information for most people, it plays a vital role in the teaching, research and learning process. MOOCs are probably the most important "novelty" in the field of e-learning of the last years, it represents an emerging methodology of online teaching and an important development in open education. MOOCs makes it possible for everyone to access the education over the world, but due to the large resources in the web, it becomes increasingly difficult for a learner to identify a suitable course for him. This task can be tedious because it involves access to each platform, search available courses, select some courses, read carefully each course syllabus, and choose the appropriate content. Web video mining is retrieving the content using data mining techniques from World Wide Web. There are two approaches for web video mining using traditional image processing (signal processing) and metadata based approach. In this work, effective attempts will classify and predict the metadata features of web videos such as category of the Mooc video, length, number of comments, rate, ratings information and view counts of the Mooc videos. In this perspective, the data mining algorithms such as Decision tree J48 and naive Bayesian algorithms will be used as a part of Mooc video mining. The results of Decision tree J48 and naive Bayesian classification models will be analyzed and compared as a step in the process of knowledge discovery from Moocs videos.**

Keywords— Moocs; e-Learning; Web video mining; Data mining; Metadata; Classification techniques; Decision tree J48; Naive Bayesian;

I- INTRODUCTION

Learning is a life-long activity that people are embracing. Along with the evaluation of Internet, Massive open online course (MOOC) has recently been experiencing a booming growth in nowadays, which brings a novel experience of learning to people. The year 2012 was even tagged as “The Year of the MOOC” by The New York Times. MOOC is a model for delivering learning content online to any person who wants to take a course, with no limit on attendance, with the option of free and open registration. MOOCs integrate social networking, accessible online resources, and are facilitated by leading practitioners in the field of study. Most significantly, MOOCs build on the engagement of learners who self-organize their participation according to learning goals, prior knowledge, skills, and common interests.

Data mining is the extraction of hidden predictive information from large database. It is a powerful new technology with great potential. Data mining is a knowledge discovery process in large and complex data sets, referring to extract or “mining” knowledge from big data. Data mining is a multidisciplinary field with many techniques that create a mining model describing the data to be used.

The entire process of applying the techniques to extract the data from the World Wide Web is the Web mining. A Multimedia database system includes a multimedia database management system, which manages and provides support for storing, manipulating and retrieving multimedia data from database. Multimedia gives a lot of information on each entity but not the same information for each one. The difference between Multimedia Mining and Structured data mining is the sequence or time element.

The classification of web videos is an increasingly outstanding area of research, growing with the quantity of videos shared through some social sites such as YouTube, Yahoo Screen etc. In recent years many works have been implemented to discover knowledge from videos using traditional image processing techniques such as video image retrieval and indexing technique; object recognition/detection in videos such as face recognition, vehicle detection… Video object is tracking from one frame to another and many more. A number of algorithms have been developed to apply the mining to videos using image processing approach. However, in recent years the discovered knowledge from web videos has concluded that a less number of research works have been implemented using metadata based approach which is a challenging task nowadays. The reason for this fact is insufficient of metadata for many web videos. This leads to ineffective and poor results in the discovery knowledge filed [3]. Some of the web videos don’t have sufficient metadata, because the uploader of the video ignores to give sufficient metadata while uploading video to the websites.

In this work, effective attempts are made to classify and predict the metadata features of Moocs videos such as length of the web videos, number of comments of the web videos, ratings information and view counts of the web videos using data mining algorithms such as Decision J48 and Naive Bayesian algorithms as a part of web video mining.

The rest of the paper is organized as follows: The section 2 represents related works on the classification of web videos, section 3 represents proposed Moocs videos classification methodology. In section 4, the results of Decision tree J48 and Naive Bayesian classification models are analyzed and compared as a step in the process of knowledge discovery from Moocs videos. Finally, section 5 represents conclusion and future enhancements.

II- RELATED WORKS

In recent years, many works have been implemented to discover knowledge from videos, in this section we present some related previous works, which are implemented to classify web videos using metadata objects.

In [1], Y. Song, Ming Zhao, J. Yagnik, and X. Wu worked on a large scale video taxonomic classification system, which uses the category taxonomic structure in training and in interpreting the classification results. To take advantage of the available text information from videos on the Internet and to compensate for the lack of labeled video training data, a novel scheme is proposed to adapt the web-classifiers to video domain. Video content based features are integrated with text features to gain power in case of degradation of one type of features. Experiments show that the proposed algorithms generate significant performance improvement over the original text classifiers, and over the classifiers trained from using video content based features. The performance has reached a satisfactory level for practical deployment.

In [2], the authors S. P. Algur, P. Bhat and S. Jain described significance of web video descriptive metadata, presented an effective and efficient method for construction and extraction of web video descriptive metadata. The proposed method demonstrated the effectiveness of constructing the descriptive metadata with timeline for a domain specific web video. The papers also suggested the construction of event specific and object specific metadata which are considered to be very effective. Using proposed descriptive metadata model, users may process the video contents effectively and efficiently.

In [8], effective attempts are made to cluster web videos based on metadata objects such as – category, view counts, length, number of comments, and rating information. The clusters are made to form automatically using unsupervised Expectation Maximization (EM) and Density Based (DB) clustering approach. Effective clustering models were built using EM and DB algorithms and applied on large scale web video metadata object dataset. Different clusters were formed according to the web video metadata objects. Each resultant clusters are analyzed in depth and normal distribution of each numerical metadata object within clusters are found, the log likelihood of EM and DB cluster models was also found.

In [9], they classified/partitioned web videos based on their category using web metadata. The web video metadata are extracted and stored in a database for classification. The Random Tree (RT) and J48 classification algorithms are chosen to classify/partition the web videos. The classification results of RT and J48 classification models are compared and RT classification model was found more efficient to classify web videos using metadata. Also, category wise cost/benefit analysis of RT classification result is analyzed and minimum cost/benefit and maximum cost/benefit are found with classification accuracy. Difficulties were arrived during the classification due to insufficient web metadata. Only 79% tuples were classified and 21% tuples are ignored by the Random Tree and J48 classification models due to poor tuple records. Also the J48 classification model has less efficiency on partitioning web video categories based on independent attributes. Because the independent attribute values are numeric.

In [10], they worked on a novel approach to classify the multimedia contents based on internal metadata such as- video bit rate kbps, maximum bit rate kbps, width pixels, and height pixels etc. The proposed work used Decision Tree and Support Vector Machine (SVM) approach for classification process. The web multimedia-video metadata are extracted and stored and pre-processed in a database for classification. In this research supervised learning algorithms are compared to predict the best classifier. An experimental result shows the effectiveness of the proposed method. The Model is also evaluated using precision and recall and F-Score. The Decision Tree (DT) and SVM classification algorithms are chosen to classify the web multimedia-videos. The all class of multimedia data classification results of DT and SVM classification models are compared and found DT classification model is more efficient. Also the SVM classification model has less efficiency web multimedia video based on independent attributes.

In [11], they classified web videos based on their metadata attributes/features such as- length, view counts, rate, ratings, and number of comments as a part of knowledge discovery from web videos. The web video metadata are extracted from standard website and stored in a database for classification. The J48 and naive Bayesian (NB) classification algorithms are chosen to classify/predict the class labels of different attributes chosen. The classification results of J48 and NB classification models are compared and found NB classification model with predictive analysis is more efficient to classify web videos using metadata.

In [4], the authors J. R. Zhang, Y. Song and T. Leung explored an approach which exploits YouTube video co-watch data to improve the performance of a video taxonomic classification system. A graph is built whereby edges are created based on video co-watch relationships and weakly-labeled videos are selected for classifier training through local graph clustering. Evaluation is performed by comparing against classifiers trained using manually labeled web documents and videos. We find that data collected through the proposed approach can be used to train competitive classifiers versus the state of the art, particularly in the absence of expensive manually-labeled data.

In [6], S. Agrawal, N. Agarwal, A. Surekha presented a one class classifier approach to detect the privacy invading harassment and misdemeanor videos having objectionable content on YouTube. They conduct a series of experiments on YouTube and identify 13 discriminatory features including linguistic, popularity and temporal metadata of given video and related videos. The experimental results indicate that identified discriminatory features can be used to exploit the harassment detection on YouTube unto a reasonable accuracy.

In [5], the authors C. Huang, T. Fu and H. Chen proposed a text-based framework for video content classification of online-video sharing websites. Different types of user-generated data (e.g., titles, descriptions, and comments) were used as proxies for online videos, and three types of text features (lexical, syntactic, and content-specific features) were extracted. Three feature-based classification techniques (C4.5, Naïve Bayes, and SVM) were used to classify videos. To evaluate the proposed framework, user-generated data from candidate videos, which were identified by searching user-given keywords on YouTube, were first collected. Then, a subset of the collected data was randomly selected and manually tagged by users as our experiment data. The experimental results showed that the proposed approach was able to classify online-videos based on users' interests with accuracy rates up to 87.2%, and all three types of text features contributed to discriminating videos. SVM outperformed C4.5 and Naïve Bayes in their experiments.

Automatic categorization of videos in a Web-scale unconstrained collection such as YouTube is a challenging task. A key issue is how to build an effective training set in the presence of missing, sparse or noisy labels. In this regard, the authors Z. Wang, M. Zhao, Y. Song, S. Kumar, and B. Li [3], proposed to achieve this by first manually creating a small labeled set and then extending it using additional sources such as related videos, searched videos, and text based web pages.

III- PROPOSED APPROACH

This section represents the detailed method of the approach used. The web metadata of Moocs are extracted using Info Extractor tool [12]. This metadata includes uploader information, category, comments, ratings, length of the Mooc, descriptions about content of the Mooc etc. The metadata features are discretized and transformed to nominal values by 'Equal Width Partitioning' method, and out of the total metadata dataset, 60% are used for training and the remaining 40% are used for testing the classification model built using Decision Tree J48 and Naive Bayesian classification methods. The classification/prediction results of each considered metadata features are analyzed and the efficiency of the proposed method has been demonstrated. The approach model is represented in Fig. 1, and it consists of the following components:

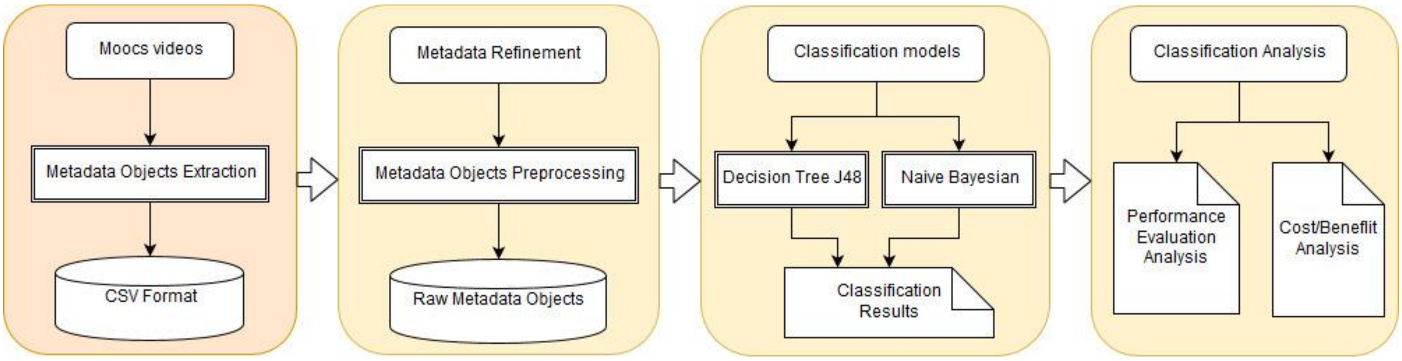

Figure 1 : Approach Model of the Proposed Methodology

1) Moocs Videos Metadata Extraction

In this component, the quality and efficiency of the mining result is directly depended to the richness of the available metadata in web Moocs. An effective method is needed to extract the metadata from web videos. A traditional procedure to extract metadata from web pages is using XML platform. To extract metadata from web videos, different open source tools are available such as Media Info, Video Inspector, InfoExtractor, MooO etc. The Mooc metadata objects are then stored in a disk with CSV file format for experimental purpose.

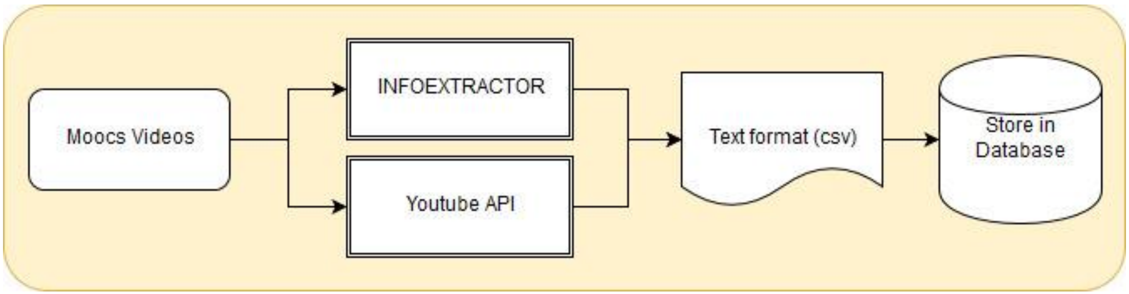

Figure 2 : Extraction of Metadata using InfoExtractor

The extracted web metadata includes information of Mooc URL, website data, user comments, number of views, number of likes and dislikes, rating, tags, date time of uploading, and subscription information etc...

2) Metadata Refinement and Categorization

The input to this component is raw metadata extracted from the Moocs. This raw metadata has to be preprocessed for the refinement such as file format conversion and to identify the unimportant metadata. The WEKA tool is used for file preprocessing and to build classification model. The extracted raw metadata are converted to CSV format from the text format for effective classification. Some Mooc might have less metadata information, whereas some Moocs might have more metadata information. Through observations, it is found that, all Moocs contains minimum metadata information such as- length, view counts, ratings, average ratings and number of comments, author information and URL. Among this minimum metadata information of Moocs, only the numeric and nominal attributes-length, view counts, ratings, average ratings, category and number of comments are considered for classification. To improve the classification and prediction accuracy, we will generate subcategories based on Mooc's title and description metadata, an algorithm was developed to generate categories. The Mooc genres are categories like Education, Science & Technology, How-to & Style, etc. The genre names are pre-defined in the system along with a small set of root words for each genre. The root words act like a description of the genre. For example, Language and Mathematics act as the key characteristics of the Education genre. This allows new genres to be easily defined in the system in terms of the root words as well as to have a fine distinction between the genres.

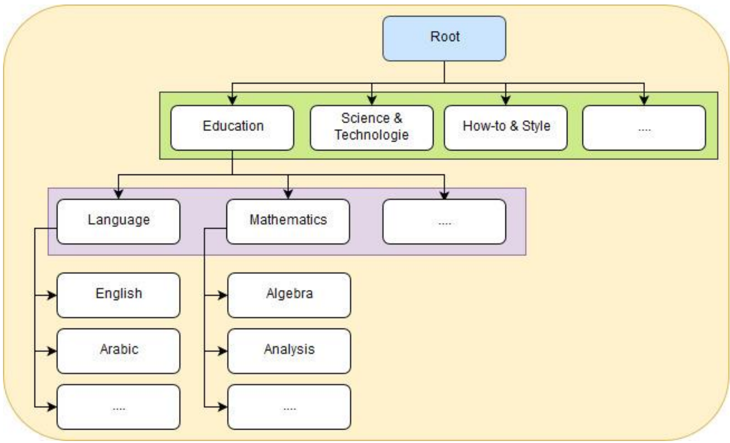

Figure 3 : Metadata Categorization

The different kind of web video metadata objects are extracted using InfoExtractor tool and web video meta-objects are then stored in a disk with CSV file format for experimental purpose. The raw web video metadata objects are then preprocessed to conduct effective proposed experiments. A typical structure of web video metadata object dataset is presented in Table 1. The attribute "Category" is nominal and contains 15 different classes of web videos. Each Category has several sub-categories. The remaining attributes are numeric and represents features of each web videos.

Table 1 : Structure of Moocs Metadata Object Dataset

N°	Category	Sub-Category 1	Sub-Category 2	Length	View	Rate	Rating	Comment
1	Education	Language	Arabic	105	885	3.62	8	22
2	Education	Language	English	77	5081	4.52	10	500
3	Sc. & Technology	C. Science	Programming	599	2559	4.12	55	652
4	Education	Mathematics	Algebra	407	52236	4.64	286	507
5	Education	Mathematics	Analysis	386	10011	4.57	56	86
n		…	…	…	…	…	…	…

The missing values are replaced by mean of each numeric attribute and the missing values of the attribute 'category' are replaced by most repeated nominal values. Since, the metadata attributes length, comments, rate, views, and ratings are extracted in the form of continuous/numeric values and hence, to improve the classification and prediction accuracy, data transformations are needed to transfer from continuous numeric values to nominal values. The considered numeric attributes are then discretized and transformed to nominal values by 'equal width partitioning' method. The Table 2 represents typical structure of transformed web video metadata dataset. The transformed metadata dataset is stored in a database for classification.

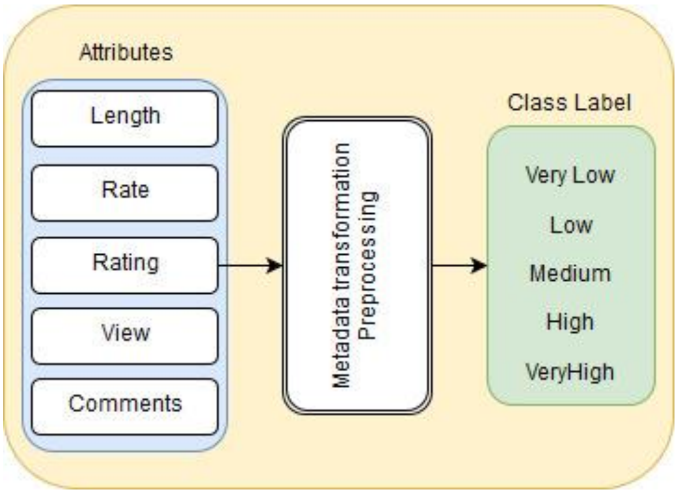

Figure 4 : Metadata Transformation Preprocessing

3) Classification model

We used the Decision Tree J48 and Naive Bayesian classification algorithms to classify/partition the Moocs videos. Decision Tree is a tree structured predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the different attributes; the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes represent the final value (classification) of the dependent variable. The process of the Decision Tree J48 classification algorithm is described as follows:

Building the J48 classification model:

The proposed J48 classification model consists of three major steps such as,

a) Attribute selection measures

b) J48 algorithm

c) Classification rules.

J48 classification algorithm

J48 is bespoke version of C4.5 classification algorithm. The J48 algorithm generates a classification-decision tree for the web video metadata dataset by recursive partitioning the tuples. The decision tree is grown using depth-first strategy. The algorithm considers all the possible tests that can split the metadata data set and selects a test that gives the best information gain. For each continuous attributes of the web video such as 'Length', 'Ratings', 'Comments' etc. Binary tests involving every distinct values of the attribute are considered. In order to gather the information gain of all these binary tests efficiently, the training data set belonging to the node in consideration is sorted and the information gain of the binary partition point based on each distinct values are calculated and sub trees are formed accordingly. This process is repeated for each continuous attributes.

Naive Bayesian algorithm

The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

4) Classification Analysis

In this section, performance evaluation measures such as TP, FP, precision, recall and F-Measure will be calculated to measure classification accuracy and efficiency of Decision Tree J48 and Naive Bayesian classification model. Also the classification accuracy of NB and J48 will be compared. To analyze the cost/benefit of each Mooc category, the result of classification model with highest accuracy will be taken. In the cost/benefit analysis, the minimum cost and maximum cost of each Mooc category with classification accuracy will be represented.

IV- EXPERIMENTATION

To test the efficiency of the classification models constructed using Decision tree J48 and Naive Bayesian, the Moocs videos dataset is extracted from the data mining tool which consists of 20 web Moocs video metadata instances. The performance of the model is measured in terms of number of correctly classified instances, number of incorrectly classified instances, TP rate, FP rate, Precision, Recall, F-Measure, MCC, ROC Area, PRC Area. The multimedia data contains three different classes which represents the 20 attributes. The Figure 5 represents classification result obtained by the Decision Tree J48 and Naïve Bayesian classification models. The performance evaluation of J48 and Naive Bayesian (NB) classification models on the class label prediction of each attributes is discussed in the following subsections.

Table 2 : Classification Result of J48 and Naive Bayesian Classification Models

N°	Attribute	Classification Model	Classified Instances 20		TP	FP	Precision	Recall	F-measure	MCC	ROC Area	PRC Area
			Correctly	Incorrectly								
1	Category	J48	18	2	0,900	0,005	0,867	0,900	0,875	0,877	0,992	0,900
		NB	20	0	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000
2	Sub-Category 1	J48	15	5	0,750	0,044	0,665	0,750	0,684	0,673	0,957	0,725
		NB	19	1	0,950	0,009	0,913	0,950	0,929	0,926	1,000	1,000
3	Sub-Category 2	J48	12	8	0,600	0,047	0,407	0,600	0,473	0,469	0,970	0,630
		NB	18	2	0,900	0,018	0,890	0,900	0,885	0,876	0,997	0,988
4	Length	J48	18	2	0,900	0,082	0,918	0,900	0,900	0,848	0,983	0,959
		NB	19	1	0,950	0,041	0,955	0,950	0,949	0,919	0,988	0,984
5	Views	J48	18	2	0,900	0,042	0,918	0,900	0,899	0,852	0,971	0,924
		NB	18	2	0,900	0,067	0,920	0,900	0,899	0,852	0,982	0,982
6	Rate	J48	18	2	0,900	0,060	0,912	0,900	0,886	0,859	0,972	0,915
		NB	16	4	0,800	0,064	0,856	0,800	0,808	0,739	0,977	0,961

7	Ratings	J48	13	7	0,650	0,200	0,558	0,650	0,591	0,456	0,827	0,617
		NB	18	2	0,900	0,042	0,918	0,900	0,893	0,869	0,983	0,972
8	Comments	J48	15	5	0,750	0,124	0,623	0,750	0,672	0,616	0,923	0,778
		NB	17	3	0,850	0,056	0,858	0,850	0,849	0,799	0,993	0,984

In the prediction of the class labels of the ‘Category’ attribute, 18 metadata instances are correctly classified and 2 instances are incorrectly classified by the J48 classification model, whereas, 20 metadata instances are correctly classified and 0 instances are incorrectly classified by the Naive Bayesian classification model. These statistics shows that, the Naive Bayesian classifier predicted class labels more correctly compared to J48 classification model. However, the classification efficiency of the J48 classification model is 90%, and the Naive Bayesian classification model is 100%.

In the prediction of the class labels of the ‘Length’ attribute, 18 metadata instances are correctly classified and 2 instances are incorrectly classified by the J48 classification model, whereas, 19 metadata instances are correctly classified and 1 instance are incorrectly classified by the Naive Bayesian classification model. It means that the Naive Bayesian classifier predicted class labels more correctly compared to J48 classification model. However, the classification efficiency of the J48 classification model is 90%, and the Naive Bayesian classification model is 95%. This is because, the J48 classification model incorrectly classified all the instances that are belongs to the class labels ‘Medium’ and ‘High’.

In the prediction of the class labels of others attribute, with ‘SubCategory 1’ attribute, 15 instances are correctly classified and 5 instances are incorrectly classified by the J48, whereas, 19 instances are correctly classified and 1 instance are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 75%, and it is 95% for NB. With ‘SubCategory 2’ attribute, 12 instances are correctly classified and 8 instances are incorrectly classified by the J48, on the other hand, 18 instances are correctly classified and 2 instances are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 60%, and it is 90% for NB. With ‘View’ attribute, 18 instances are correctly classified and 2 instances are incorrectly classified by the J48, whereas, 18 instances are correctly classified and 2 instances are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 90%, and it is 90% for NB. With ‘Rate’ attribute, 18 instances are correctly classified and 2 instances are incorrectly classified by the J48, whereas, 16 instances are correctly classified and 4 instances are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 90%, and it is 80% for NB. With ‘Rating’ attribute, 13 instances are correctly classified and 7 instances are incorrectly classified by the J48, whereas, 18 instances are correctly classified and 2 instances are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 65%, and it is 90% for NB. With ‘Comments’ attribute, 15 instances are correctly classified and 5 instances are incorrectly classified by the J48, whereas, 17 instances are correctly classified and 3 instances are incorrectly classified by the Naive Bayesian classification model. The classification efficiency of the J48 is 75%, and it is 85% for NB.

Table 3 : Comparison of classification Accuracy

Classification model	Correctly Classified	Incorrectly classified	Accuracy
J48 classifier	127	33	79,38%
Naive Bayesian	145	15	90,63%

By the comparison of results of J48 classification model and NB classification model, the NB classification model is found with highest accuracy. The experimental result shows that the NB classification method is the most effective way to classify the Moocs web videos based on their category. The classification results comparative analysis of efficiency of J48 and NB classification models are represented for the attributes ‘Category’, ‘SubCategory 1’, ’SubCategory 2’, ‘Length’, ‘Views’, ‘Rate’, ‘Ratings’, and ‘Comments’. The J48 classification model is found with good accuracy for the prediction of class labels of the attribute ‘Rate’. For the remaining attributes the NB classification model is found with highest efficiency as compared to J48 classification model. Hence, by considering the overall experimental results, the NB classification model with predictive analysis is found with highest efficiency for the classification of web video metadata attributes/features. The figure 5 represent the comparative analysis of efficiency of J48 and NB classification model.

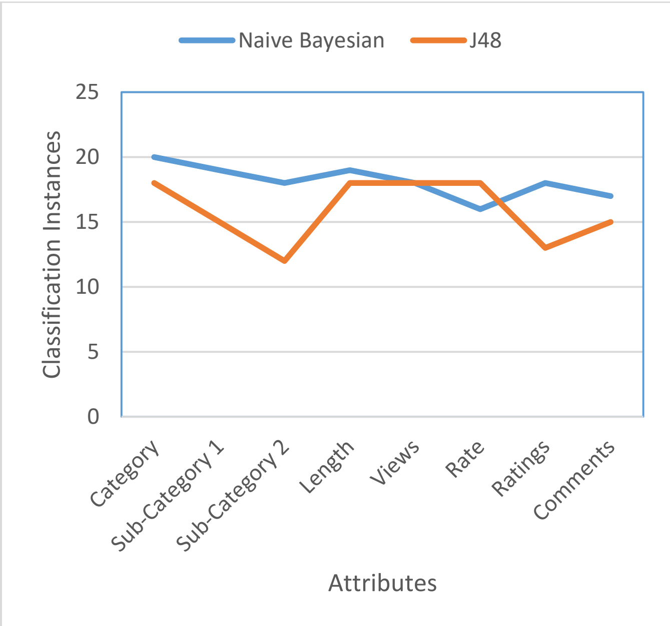

Figure 5 : Comparative Analysis of Efficiency of J48 and NB Classification Models

V- CONCLUSION AND FUTURE WORK

In this work, we reviewed some existing techniques to mine web videos using metadata and we have concluded that there is no ideal framework to extract data from web videos. We have presented approach towards Moocs videos mining using metadata based on J48 and Naive Bayesian classification models to provide learners with better results. We have classified Moocs videos based on their category using web metadata. The Moocs web video metadata are extracted from standard website and stored in a database for classification. The J48 and Naive Bayesian classification algorithms are chosen to classify the class labels of different attributes chosen. The classification results of J48 and Naive Bayeasian classification models are compared and the Naive Bayesian classification model was found more effective to classify Moocs web videos using metadata. The difficulties we've encountered during the classification were due to insufficient web metadata. Only 80% tuples were classified and 20% tuples were ignored by the J48 classification models due to poor tuple records. Also the results demonstrate classification of Moocs videos depends largely on available metadata and accuracy of the classification model. The future work is to improve the classification accuracy of the Naive Bayesain classification model, as well we try to extract low level features from Moocs videos by analyzing the frames, signals, audio etc, along with textual features.

VI- REFERENCE

[1] Yang Song, Ming Zhao, Jay Yagnik, and Xiaoyun Wu, "Taxonomic Classification for Web-based Videos", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, June 2010

[2] Siddu P. Algur, Prashant Bhat, Suraj Jain, "Metadata Construction Model for Web Videos: A Domain Specific Approach", International Journal Of Engineering And Computer Science (IJECS), Volume 3 Issue 12 December 2014, Page No.9742-9748

[3] Zheshen Wang, Ming Zhao, Yang Song, Sanjiv Kumar, and Baoxin Li, "YouTubeCat: Learning to Categorize Wild Web Videos", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, June 2010

[4] John R. Zhang, Yang Song, and Thomas Leung,"Improving Video Classification via YouTube Video Co-Watch Data", SBNMA '11 Proceedings of the 2011 ACM workshop on Social and behavioural networked media access, Scottsdale, Arizona, USA, December 2011

[5] Chunneng Huang, Tianjun Fu and Hsinchun Chen,"Text-based Video Content Classification for Online Video-sharing Sites", Journal of the American Society for Information Science and Technology (JASIST), Volume 61 Issue 5, Pages 891-906, May 2010

[6] Aggarwal N, Agrawal, S. and Sureka, A., "Mining YouTube Metadata for Detecting privacy Invading Harassment and Misdemeanor Videos", Twelfth Annual International Conference on Privacy Security and Trust, vol., no., pp.84,93, 23-24 July 2014.

[7] Shubhangi D. Raverkar, Meghana Nagori "Classification of YouTube Metadata using Shark Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 132 – No.9, December 2015

[8] Siddu P. Algur, Prashant Bhat. "Web Video Object Mining: Expectation Maximization and Density Based Clustering of Web Video Metadata Objects", I.J. Information Engineering and Electronic Business (IJIEEB), 2016, 1, 69-77

[9] Siddu P. Algur, Prashant Bhat. "Metadata Based Classification and Analysis of Large Scale Web Videos", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 3, May-June 2015 ISSN 2278-6856

[10] Siddu P. Algur, Basavaraj A. Goudannavar, Prashant Bhat. "Web Multimedia classification Using DT and SVN Models", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 5, Issue 4, July - August 2016 ISSN 2278-6856

[11] Siddu P. Algur, Prashant Bhat. "Web Video Mining: Metadata Predictive Analysis using Classification Techniques." International Journal of Information Technology and Computer Science (IJITCS), 2016, 2, 69-77

[12] Chirag Shah, Charles File, "Infoextractor – A Tool for Social Media Data Mining", the Journal of Information Technology and Politics (JITP) Volume 9, 2012 - Issue 3.