IDENTIFICATION OF SONG ARTIST AND RECOMMEMDATION OF SONGS USING MACHINE LEARNING

Mrs.M.P.Divya[1], Dr.Kaushik Sekaran[2], Sai Sri.G[3], K.Roja[4] & M.Saineha[5]
[1]Assistant professor, [2]Professor, [3,4,5]B-TECH,
Dept.:-CSE, VIGNAN INSTITUTE OF TECHNOLOGY AND SCIENCE

Received: February 17, 2019 Accepted: March 25, 2019

ABSTRACT: *Textual content classification is an major and original project in supervised computing device studying. The Naïve Bayes Classifier is a fashionable algorithm that can be utilized for this cause. The intention of our research used to be prediction of tune performer utilising Naive Bayes Classification algorithm based completely on lyrics. A dataset that has been created contains lyrics performed by means of Nirvana and Metallica, 207songs in whole. Model analysis measures confirmed superb outcome: precision of zero.93, recollect of zero.95 and F1-measure of zero.94, hence lyrics classification making use of Naive Bayes may also be regarded as positive.*

Key Words: *Naive Bayes, Songs data set.*

1. INTRODUCTION

Text classification is an essential and long-established task in supervised machine finding out. Its application is in e-mail spam detection, sentiment analysis, language detection of written text, classification etc. Many classifiers can be utilized for report classification. A few of them are neural networks, help vector machines, genetic algorithms, Naive Bayes classifier, k-nearest neighbors and Rocchio classifier [1]. The wide variety of tune, particularly on the internet, is growing speedily and its organizing is a difficult assignment. Given the significant measurement of music collections, classification of music should be made robotically. Classification may also be made in line with style, temper, performer, geographical neighborhood, and so forth. To make classification positive, it is easy to depend on audio elements corresponding to tempo, rhythm, timbre, pitch, loudness or lyric aspects akin to word and sentence length, word frequencies, word n-grams, sentence and phrase constitution, error, synonyms, rhyme patterns and so forth. According to [2] most existing work on automated music temper classification is based on audio points (spectral and rhythmic aspects are probably the most popular). Relying on type of classification, combining audio and lyrics information is a normal strategy. In [3] four very certain genres (classical, jazz, metallic and pop) were chosen for audio-situated classification utilising Mel Frequency Cepstral Coefficients. Accuracy in genre prediction when Direct Acyclic Graph help Vector Machines used to be applied different from 67 % to 97 %. When Neural Networks had been used, accuracy varied from 76 % to a 100 % relying on style. Automated identification of tune performers, given a set of piano performances of the equal piece of music is an fascinating study described in [4]. Pianists played two portions by way of Frederick Chopin. Success cost was once excessive: the accuracy was 70 % in 10-type undertaking. Fell and Sporleder in [5] dealt with predicament of discovering out whether or not it is feasible to robotically predict the approximate newsletter time of a song given its lyrics. They selected pop/rock songs and divided them into three periods: 2008 and newer, from 1998 to 2001, and those released before 1988. Results confirmed that songs which might be released two decades and more in the past will also be distinguished rather good, however for newer songs results of classification are rather low. Authors in [6] record that there's no giant difference in outcome of tune mood classification relying on whether or not stemming was used or now not. In [7] authors highlight that stemming and removing of discontinue words may do extra damage than just right when coping with multilingual lyrics. Text authorship identification is a field with long research historical past [8]. The major suggestion behind statistically or computationally supported authorship attribution (which began at the finish of nineteenth century) is that the texts written via special authors can also be special through measuring some textual features [9]. This subject speedily developed with the progress of computer studying classification procedures. The purpose of this study used to be checking out whether or not the Naive Bayes

classifier can effectually predict music performer situated exclusively on lyrics. A dataset which include lyrics of two performers (Nirvana and Metallica) was once created for this cause. Two performers are chosen deliberately to separate problems of classification in line with performer from the main issue of classification in line with the genre of tune due to the fact genres of their song should not some distance away from each other. Nirvana is a rock band, at the same time Metallica is heavy metallic (which is one sub-genre of rock) band. No single writer writes lyrics for one performer, but songs are written having a performer in intellect (and audience of direction), so style and style of songs will have to be almost each other. As a subject of a fact, commonly one song is written by a couple of creator. Within the case of Metallica, many songs are written with the aid of three or 4 authors. As dataset has 127 Metallica's and eighty Nirvana's songs, Naive Bayes Classifier was once used, in view that it is suitable for small datasets [10]. The the rest of this paper is organized as follows. In the next part we in brief describe the ways and measures we used. In section three we describe our experiment and reward the outcome. In part 4 we draw conclusions and point out future instructional materials.

2. RELATED WORK

Existing System

The range of song, specially on the net, is developing quickly and its organizing is a difficult undertaking. The gigantic measurement of track collections, classification of track is very problematic. On the whole in current system track classification and ideas can use by means of the overall cluster approaches. Utilizing classification algorithms in computer learning we will classify the songs using lyrics knowledge. From an affective computing point of view, it's exciting to investigate the connection between a artist and lyrics.

Disadvantages

1. Song recommendation based on static information like film name, actor and so on.
2. No classification work in present options.
3. Consumer mood no longer taking into the concerns.

Proposed System

The purpose of this study was trying out whether or not the Naive Bayes classifier can efficiently predict music performer established completely on lyrics. A dataset consisting of lyrics of two performers (Nirvana and Metallica) was created for this purpose. Two performers are chosen intentionally to separate issues of classification in line with performer from the main issue of classification in keeping with the genre of tune seeing that genres of their music are not a ways far from every different. Naive Bayes is a laptop studying algorithm whose classification efficiency is proved in purposes corresponding to document categorization and electronic mail unsolicited mail filtering. This classifier learns via a file classification algorithm, and is established on a easy usage of the Bayes'rule

Advantages

1. Music suggestion centered on dynamicdata like artist and lyrics of the songs.
2. Utilising computer learning classification work in proposed options.
3. Person temper taking into the issues.

3. IMPLEMENTATION

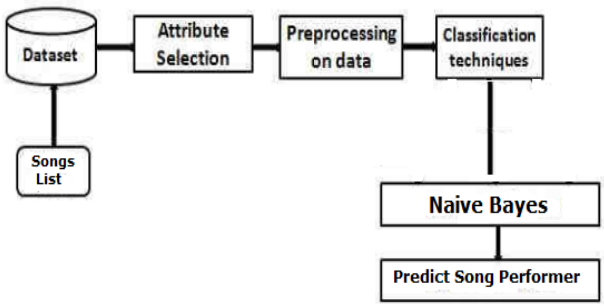

Fig:-1 System Architecture

Admin

Admin is predominant consumer of our utility. Admin will upload the songs dataset containing of tune name, artist title, online play hyperlink and lyrics of the track. Admin will participate in classification of the artist prediction utilisingNaïve Bays algorithm. Admin can verify the reviews of our software results.

User

Person is end person of our utility. User can login with his/her register details. User can search the songs through provide some key phrases of the song. Person can get the record of songs with lyrics and with online link of the music play. Consistent with the tune s/he listening person will get the song suggestions according to Naïve bays algorithm.

Naïve Bayes

Naive Bayes is a computing device finding out algorithm whose classification efficiency is proved in applications equivalent to file categorization and e mail junk mail filtering. This classifier learns by way of a file classification algorithm, and is based on a easy utilization of the Bayes'rule. We use this algorithm for lyrics classification

e ISSN 2348 –1269, Print ISSN 2349-5138
Cosmos Impact Factor 4.236

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

Algorithm used in Project

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

wherein:

- c is a class,
- d is a document,
- P(c) is a class probability,
- P(d) is the probability of a document,
- P(d|c) is conditional probability of the class forthe given document d,
- P(c|d) is conditional probability that document dbelongs to class c.

Computational efficiency in modeling and predictingis an unquestionable advantage over some otherclassification algorithms, which is due to the possibility ofeasy parallelization, especially important for largedatasets.

4. EXPERIMENTAL RESULTS

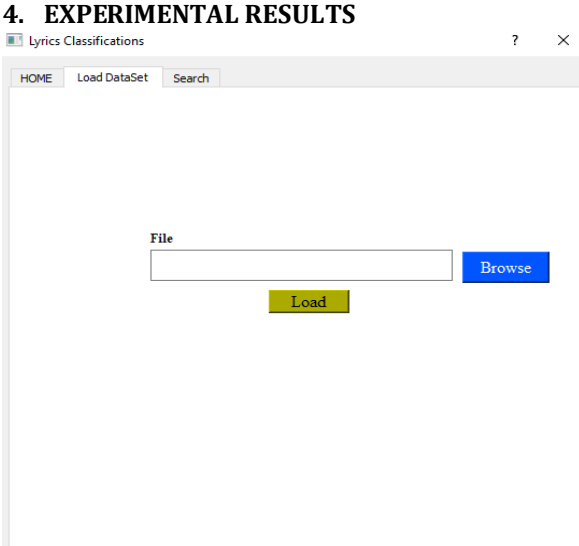

Fig:-2 Data Set Upload

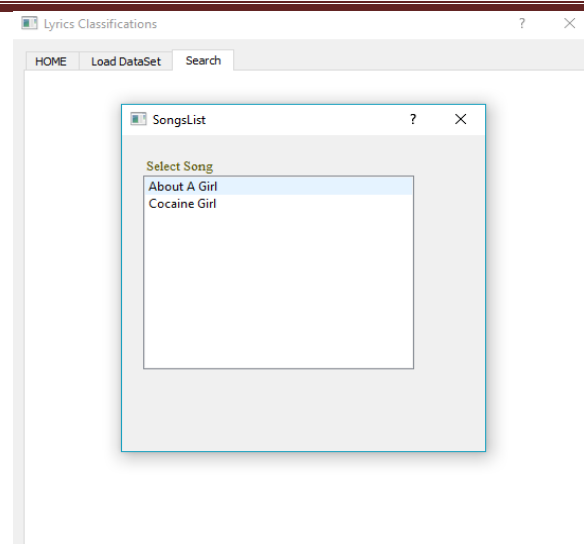

Fig:-3 Search Results

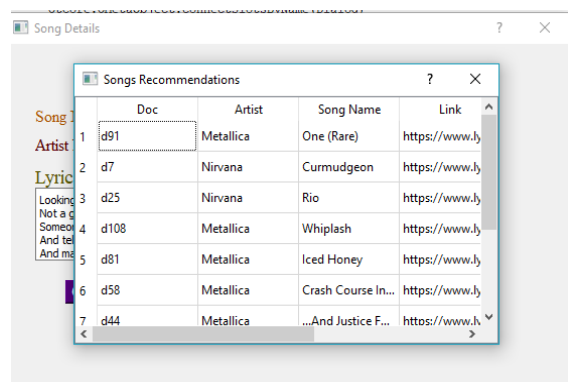

Fig:-4 Result Data Set

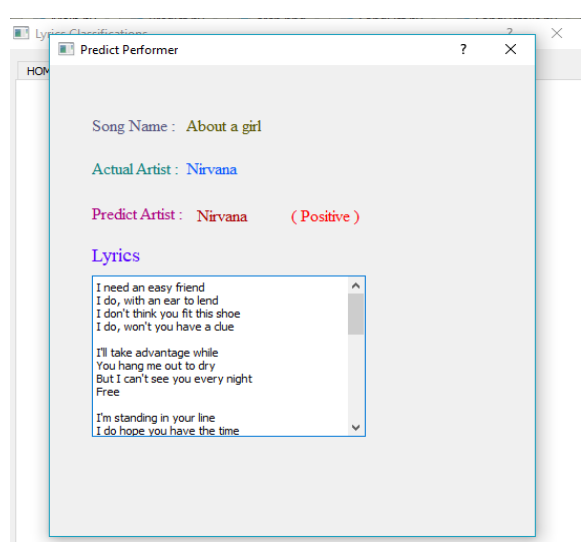

Fig:-5 Naïve Bayes

5. CONCLUSION

Making a dataset was tedious and time-drinking challenge, partly due to the fact that it used to be

created manually, and partly for the reason that of doubt about inserting some songs into a dataset. Specifically, instances such as visitor appearances of different musicians on the album or two versions of the equal song(a studio and a slightly altered live variant) had to behandled with care. Apart from, it used to be no longer always clear whether a song belong to a performer or not – the doubt used to be resolved with a help of Wikipedia's list of songs recorded by using chosen artist. Outcome of a created mannequin are very good. Naive Bayes classifier is a good choice for this task – once once more it proved its capabilities. Given that the dataset was once really small, it was a logical candidate for the mannequin. Outcome showed that Nirvana's and Metallica's songs have textual 'signatures' that may be unique to alarge measure completely on studying textual content. Outcome are more exciting when one takes into consideration the fact that songs for one band are almost always written by using extra authors. In some future study, it could be intriguing to examine how the mannequin behaves in a bigger quantity of classes (artists)and to evaluate outcome bought by Naïve Bayes classifier with outcome received by way of other classifiers, primarily with support vector machines.

REFERENCES

1. A. Khan, B. Baharudin, L. H. Lee & K. Khan, "A review of machine learning algorithms for text-documents classification", Journal of advances in information technology, 1(1), 2010, pp. 420.
2. X. Hu & J. S. Downie, "When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis", 11th International Society for Music Information Retrieval Conference (ISMIR), August 2010, pp. 619-624.
3. M. Haggblade, Y. Hong & K. Kao, "Music genre classification", Department of Computer Science, 2011, http://cs229.stanford.edu/proj2011/HaggbladeHongKaoMusicGenreClassification.pdf
4. E. Stamatatos& G. Widmer, "Music performer recognition using an ensemble of simple classifiers", ECAI, 2002, pp. 335-339.
5. M. Fell & C. Sporleder, "Lyrics-based analysis and classification of music", Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers,2014, pp. 620-631.
6. X. Hu, J. S. Downie& A. F. Ehmann, "Lyric text mining in music mood classification", American music, 183.5, 049, 2009, 2-209.
7. S. Howard, C. N. SillaJr& C. G. Johnson, "Automatic lyricsbased music genre classification in a multilingual setting", Proceedings of the Thirteenth Brazilian Symposium on Computer Music, 2011.
8. N. Homem& J. P. Carvalho, "Authorship identification and author fuzzy "fingerprints"", Fuzzy Information Processing Society (NAFIPS), 2011 Annual Meeting of the North American, IEEE, 2011, pp. 1-6.
9. E. Stamatatos, "A survey of modern authorship attribution methods", Journal of the Association for Information Science and Technology 60.3, 2009, pp. 538-556.
10. B. Lantz, "Machine learning with R", Packt Publishing Ltd, 2015.