**Phân loại lời bài hát sử dụng Naive Bayes**

Dalibor Bužić \*, Jasminka Dobša \*\*

# \* College for Information Technologies, Klaićeva 7, Zagreb, Croatia

# \*\* Faculty of Organization and Informatics, Pavlinska 2, Varaždin, Croatia dalibor.buzic@vsite.hr, jasminka.dobsa@foi.hr

***Abstract* - Phân loại văn bản là một nhiệm vụ chung trong máy học có giám sát. Trong đó Naive Bayes là một thuật toán phổ biến có thể được sử dụng.** **Mục tiêu nghiên cứu là dự đoán người biểu diễn bài hát sử dụng thuật toán phân loại Naive Bayes chỉ dựa trên lời bài hát. Tập dữ liệu đã được tạo bao gồm lời bài hát của 207 bài hát do Nirvana và Metallica. Các biện pháp đánh giá mô hình cho thấy kết quả: độ chính xác 93% (recall of 0.95 and F1-measure of 0.94), do đó phân loại lời bài hát sử dụng Naive Bayes có thể coi như thành công.**

***Keywords* - *Naive Bayes classifier, text classification, machine learning***

# I. TỔNG QUAN VỀ VẤN ĐỀ

Phân loại văn bản là một nhiệm vụ quan trọng và phổ biến trong học máy có giám sát. Một số ứng dụng của nó như: phát hiện email spam, phân tích cám xúc, phát hiện ngôn ngữ của văn bản viết tay, phân lớp, v.v. Một vài thuật toán như: Mạng neural (ANN), support vector machines (SVM), thuật toán di truyền, Phân lớp Naive Bayes, K-Láng giềng (KNN) và phân lớp Rocchio [1].

Số lượng các bài hát, đặc biệt là trên internet, đang phát triển một cách nhanh chóng và việc tổ chức quản lý nó là một việc khá khó. Với một bộ sưu tập nhạc khổng lồ, việc phân loại nhạc nên được thực hiện một cách tự động. Phân loại có thể dựa trên các tiêu chí như: thể loại, cảm xúc, người biểu diễn, địa lý, khu vực, v.v.

Để có thể phân loại thành công, người ta có thể dựa vào một số đặc điểm của âm thanh như là: nhịp độ (tempo), nhịp điệu (rhythm), âm sắc (timbre), cao độ (pitch), độ lớn (loudness) hoặc đặc điểm của lời bài hát như là: từ và độ dài của câu, tần số của từ, biểu đồ n-grams của từ, câu và cấu trúc đoạn, các lỗi, từ đồng nghĩa, từ đồng âm v.v. Theo [2], vấn đề cần làm nhất trong phân loại nhạc tự động theo tâm trang dựa trên đặc điểm của âm thanh (phổ và nhịp điệu của âm thanh là phổ biến nhất).

Tùy vào các kiểu phân lớp, cách kết hợp âm thanh và thông tin từ lời bài hát là một cách phổ biến để phân loại.

Trong [3] có 4 thể lại nhạc phổ biến (cổ điển, jazz, rock và pop) đã đươc chọn để làm nền tảng phân loại âm thanh dự trên Hệ số bề mặt tần số Mel (Mel Frequency Cepstral Coefficients). Độ chính xác khi dùng Direct Acyclic Graph SVM rơi vào khoảng từ 67% tới 97%. Khi dùng Mạng Nơ-ron (Neural Networks), kết quả đạt được vào khoảng 76% đến 100 % tùy theo thể loại.

Tự động nhận dạng các nghệ sĩ biểu diễn, được cung cấp bởi các buổi hòa nhạc dương cầm của cùng một bản nhạc là một nghiên cứu thú vị được mô tả trong [4]. Nghệ sĩ dương cầm đã thể hiện 2 tác phẩm Frederick Chopin. Với kết quả cao với độ chính xác khoảng 70% trong 10-class task.

Fell và Sporleder trong mục [5] đã giải quyết các vấn đề tìm kiếm liệu xem có thể tự động dự đoán thời gian xuất bản gần đúng của một bài hát với lời bài hát. Họ chọn các bài hát pop/rock và chia chúng thành 3 giai đoạn: 2008 và mới hơn, từ 1998 tới 2001, và phát hành trước 1988. Kết quả cho thấy những bài hát được phát hành hơn 20 năm trước có thể phân biệt tương đối tốt, nhưng với các bài hát mới hơn kết quả phân loại tương đối thấp.

Tác giả trong [6] nói rằng không có sự khác biệt trong kết quả của phân loại nhạc theo cảm súc tùy thuộc vào bất cứ stemming được sử dụng hay không. Trong [7] tác giả nhấn mạnh rằng stemming và removing của stop words có thể tạo ra nhiều khó khăn khi xử lý nhạc đa ngôn ngữ.

Xác định quyền tác giả là một trường có lịch sử nghiên cứu trong [8]. Ý tưởng chính đằng sau việc thống kê hoặc phân bổ quyền tác giả được hỗ trợ tính toán (bắt đầu ở cuối thế kỷ thứ XIX) là các văn bản được viết bởi các tác giả khác nhau có thể phân biệt bằng một vài đặc điểm của văn phong [9]. Lĩnh vực này phát triển nhanh chóng với sự phát triển của công nghệ phân lớp của máy học. //Đoạn này nó nói cái gì dịch xong cũng ko biết nó nói cái gì luôn 😊

Mục tiêu của nghiên cứu này để kiểm tra xem phân lớp Naive Bayes có thể dự đoán thành công người biểu diễn dựa trên lời bài hát. Một tập dữ liệu bao gồm lời của hai nghệ sĩ biểu diễn (Nirvana và Metallica) đã được tạo ra cho mục đích này. Hai nghệ sĩ được chọn với mục đích tách vấn đề của phân lớp theo người biểu diễn với vấn đề phân lớp theo thể loại nhạc bới vì thể loại nhạc của họ cần như tương đồng. Nirvana là nhạc rock, và Metallica là nhạc heavy metal (một nhánh nhỏ của rock). Không có tác giả nào viết lời chỉ riêng cho 1 người biểu, nhưng các bài hát được viết có thể được biểu diễn trong tâm trí (của khán giả), vì vậy phong cách và thể loại của bài hát nên giống nhau nhất có thể. Thực tế là, đôi khi một bài hát được viết bởi nhiều tác giả. Trong trường hợp của Metallica, nhiều bài hát được viết bởi ba hoặc bốn tác giả.

Vì tập dữ liệu có 127 bài hát của Metallica và 80 bài hát của Nirvana, phân lớp Naive Bayes đã được sử dụng, vì nó phù hợp với bộ dữ liệu nhỏ [10].

Cấu trúc của bài báo này sẽ như sau. Trong phần tiếp theo chúng tôi mô tả ngắn gọn các phương pháp và các biện pháp chúng tôi đã sử dụng. Trong Phần 3, chúng tôi mô tả thí nghiệm và trình bày kết quả. Trong Phần 4 chúng tôi rút ra kết luận và chỉ ra hướng đi trong tương lai.

# II. CƠ SỞ LÝ THUYẾT

## A. Naive Bayes

Naive Bayes là một thuật toán học máy có hiệu quả phân loại được chứng minh trong các ứng dụng như phân loại tài liệu và lọc thư rác e-mail [11]. Nó học thông qua phân loại tài liệuthuật toán và dựa trên cách sử dụng đơn giản của quy tắc Bayes [12]:

Trong đó:

* *c* là một class,
* *d* là một document,
* *P*(*c*) là xác suất của class,
* *P*(*d*) là xác suất document,
* *P(d|c*) là xác suất điều kiện của class với document d đã cho,
* *P(c|d)* là xác suất điều kiện của mà document d thuộc về class c.

Phân lớp Naive Bayes đặc trưng bởi: [13]

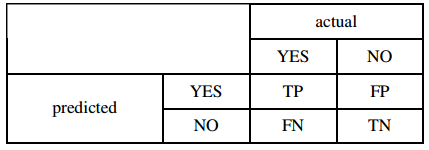
* Hiệu quả tính toán,
* Phương sai thấp,
* Học gia tăng (Incremental learning),
* Dự đoán trực tiếp về xác suất sau,
* Xử lý tốt với nhiễu và khi có giá trị thiếu.

Hiệu quả tính toán trong mô hình hóa và dự đoán là một lợi thế không thể nghi ngờ so với một số thuật toán phân loại, do khả năng song song dễ dàng, đặc biệt quan trọng đối với bộ dữ liệu. Đối với các đặc điểm đã đề cập, nó có giá trị thêm hai điều nữa: khả năng chống lại việc trang bị quá nhiều và khả năng xử lý với số lượng lớn các thuộc tính mà không cần sự lựa chọn của họ [14].

## B. Performance measures

Sau khi tạo một mô hình học máy để đo lường hiệu suất của mô hình để quyết định xem mô hình đạt yêu cầu, cho dù nó có thể được cải thiện hoặc thậm chí bỏ đi. Model nên mắc lỗi càng ít càng tốt - nhưng khái niệm sai lầm có thể được định nghĩa trên các tùy thuộc vào miền vấn đề. Dưới đây là một số biện pháp đánh giá chung nhất. Họ bắt nguồn từ ma trận nhầm lẫn (Table 1) chứa các quyết định của trình phân loại trong các hàng và quyết định thực tế về phân loại trong lớp trong cột. Bốn trường của bảng chứa số dương tính thực (TP), âm tính thực (TN), dương tính giả (FP) và các tài liệu phân loại âm tính giả (FN).

**Table 1. Confusion matrix**



Độ chính xác (precision) được biểu thị bằng tỷ lệ dương các trường hợp được công nhận chính xác là tích cực trên tất cả các trường hợp được phân loại là tích cực và được tính theo công thức:

Thu hồi (recall) được biểu thị bằng tỷ lệ các trường hợp tích cực được công nhận một cách chính xác là tích cực trên thực tế các trường hợp dương tính và được tính theo công thức:

Độ chính xác (accuracy) được biểu thị bằng tỷ lệ của các trường hợp được phân loại trên tất cả các trường hợp và được tính toán theo vào công thức:

Lỗi (error) được biểu thị bằng tỷ lệ không chính xác các trường hợp được phân loại trên tất cả các trường hợp và được tính toán theo công thức:

Hoặc:

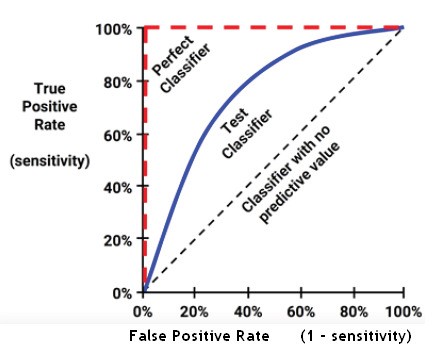
Các biện pháp riêng lẻ không nên được xem xét riêng biệt. Sẽ rất dễ dàng để xây dựng một trình phân loại vô dụng sẽ phân loại tất cả các trường hợp là tích cực, làm cho biện pháp thu hồi trở nên hoàn hảo. Độ chính xác và việc thu hồi là bổ sung cho nhau, vì nó thể hiện khả năng để phát hiện các trường hợp tích cực và khả năng khác để tránh phát hiện không chính xác các trường hợp tiêu cực. Bằng cách tăng một đo lường nó có khả năng giảm một (hoặc tốt nhất là một sẽ giữ nguyên) [14].

Một biện pháp kết hợp độ chính xác và khả năng thu hồi được gọi là thước đo F1 và đại diện cho sóng hài có trọng số của chúng nghĩa là. Nó được tính theo công thức:

Thang đo à một trong những phương pháp được sử dụng phổ biến là các biện pháp số đơn trong truy xuất thông tin, tự nhiên xử lý ngôn ngữ và học máy. Nó có giá trị đề cập rằng biện pháp này có nhiều vấn đề thực tế hơn, một số trong số đó là [15]

* Giống , và , nó cũng chỉ tập trung vào một lớp
* Giống , và độ , nó cũng thiên về một lớp chính,
* Nó không xem xét các trường hợp âm tính thực sự (TN)
* Nó giả định rằng thực tế và tiên lượng phân phối bằng nhau..

Biện pháp đánh giá cuối cùng được đề cập ở đây là đường cong ROC (Đặc tính hoạt động của máy thu - Receiver Operating Characteristic). Nó là một biểu diễn đồ họa (Figure 1) của bộ phân loại nhị phân hiệu suất mà đường cong thể hiện sự thỏa hiệp giữa các trường hợp dương tính thật và dương tính giả.



**Figure 1. ROC curve (source: [10])**

* Đường màu đen trên đường chéo biểu thị một bộ phân loại rằng các trường hợp dương tính thật và dương tính giả phát hiện ở tỷ lệ như nhau, do đó không hữu ích trong phân loại.
* Trong sự tương phản, một bộ phân loại hoàn hảo được đánh dấu bằng đường màu đỏ dự đoán dương tính thật 100% với 0% trường hợp âm tính giả.
* Đường cong của bộ phân loại thực tế gần với màu đỏ là bao nhiêu, nó tốt hơn cho việc phát hiện các trường hợp tích cực.

# III. THỰC NGHIỆM VÀ KẾT QUẢ

Mục tiêu của nghiên cứu này là tìm hiểu xem liệu bộ phân loại có thể xác định chính xác người biểu diễn (Metallica hoặc Nirvana) chỉ bằng lời bài hát. Một tập hợp dữ liệu đã được tạo cho mục đích nghiên cứu và dữ liệu được chuẩn bị. Sau đó, mô hình được đào tạo và đã đánh giá. Bước cuối cùng là cải tiến mô hình.

## A. Thu thập dữ liệu và tiền xử lý

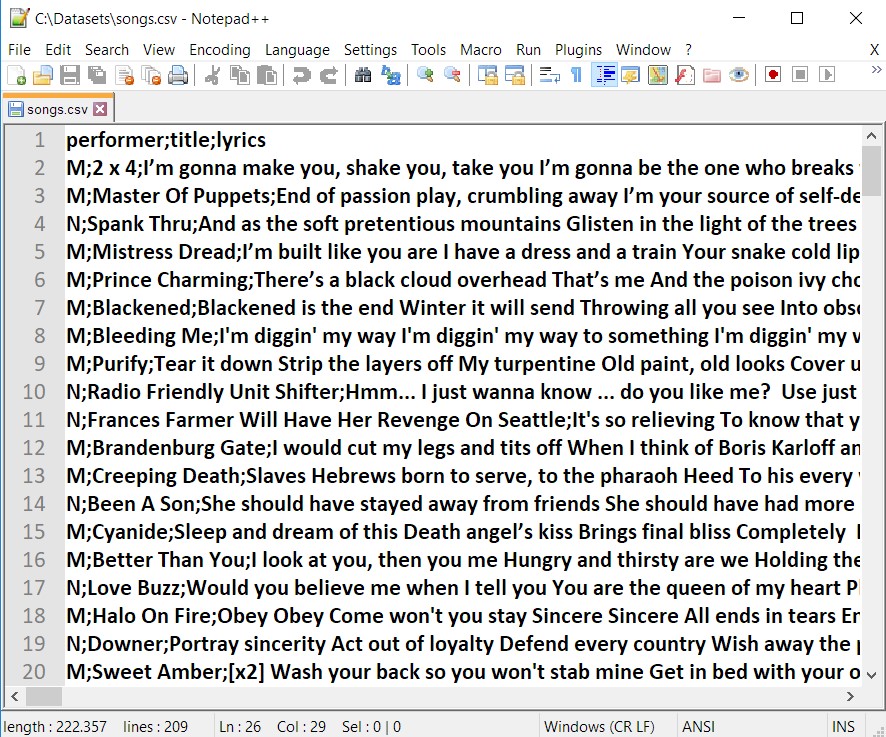
Câu hỏi nghiên cứu trong bài báo này là liệu một phân loại (và ở mức độ nào) chỉ dựa trên lời bài hát có thể nhận ra đó là bài hát của Nirvana hay Metallica.

Bước đầu tiên là tạo tập dữ liệu. Tất cả các bài hát đã lấy từ trang web azlyrics.com. Tập dữ liệu bao gồm ba cột (loại, tiêu đề và lời bài hát) và 207 hàng (127 bài hát của Metallica và 80 bài hát của Nirvana). Ở cột đầu tiên, *type*, chứa thông tin một chữ cái viết tắt bài hát thuộc về ai ('M' cho Metallica hoặc 'N' cho Nirvana). Cột thứ hai, *title*, chứa tên bài hát – nó không được dùng trong nghiên cứu này, nhưng nó rất quan trọng trong việc dễ dàng nhận ra bài hát cụ thể để quản lý và kiểm soát các bản sao nếu có. Cột thứ ba, *lyrics*, chứa lời bài hát.

Trong một số trường hợp nó không hoàn toàn rõ ràng liệu bài hát cụ thể thực sự thuộc về ban nhạc được quan sát (do sự nghiệp âm nhạc chẳng hạn như người đứng đầu ban nhạc trước khi thành lập ban nhạc). Để giải quyết những vấn đề như vậy, chỉ những bài hát được liệt kê trên Wikipedia (https://en.wikipedia.org/wiki/List\_of\_songs\_recorded\_by\_Metallica và https://en.wikipedia.org/wiki/List\_of\_songs\_recorded\_by\_Nirvana) có thể được đưa vào tập dữ liệu. Nó cũng là quan trọng cần nhấn mạnh rằng tập dữ liệu không đại diện cho toàn bộ đĩa hát của hai dải.

Sau khi tạo tập dữ liệu ban đầu, ngẫu nhiên hóa các hàng được làm từ. Vì vào cuối quá trình tạo tập dữ liệu, đã có tỷ lệ bài hát đã biết (61,4% Metallica và 38,6% Nirvana), và thực tế là 2/3 dữ liệu (138) sẽ được sử dụng cho việc học và phần ba còn lại để thử nghiệm, trong số lượng bài hát của cả hai ban nhạc được thiết lập theo tỷ lệ đã được đặt (85 bài hát của Metallica và 53 bài hát của Nirvana).

Cuối cùng, một lần nữa đặt cho đào tạo và đặt cho thử nghiệm được ngẫu nhiên riêng biệt. Điều chỉnh này sau cho phép đơn giản hóa hoạt động đào tạo và kiểm tra trong R. Một phần của tập dữ liệu cuối cùng được thể hiện trong Figure 2.



**Figure 2. Dataset**

Bước cuối cùng trong xử lý trước tập dữ liệu là phổ biến chuyển đổi: thay đổi tất cả các chữ cái thành chữ thường, loại bỏ các từ dừng, số, dấu chấm câu, dấu cách, xuống dòng... Mẫu lời bài hát trước và sau khi các phép biến đổi được thể hiện trên Figure 3.



**Figure 3. Lyrics before and after transformations**

Sau khi xử lý trước, ma trận thuật ngữ tài liệu được tạo ra. Từ điển bao gồm 2,932 thuật ngữ và ma trận kỳ hạn tài liệu thưa thớt là 98%. Danh sách nhiều nhất các thuật ngữ thường xuyên cho cả hai nghệ sĩ được trình bày trong Table 2.

**Table 2. Most frequent terms**

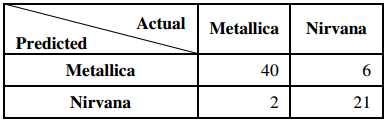
|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Nirvana** | |
| **term** | **freq** | **term** | **freq** |
| see | 157 | like | 85 |
| never | 156 | yeah | 67 |
| just | 152 | know | 57 |
| now | 141 | take | 48 |
| one | 130 | got | 47 |
| come | 128 | feel | 46 |
| take | 128 | one | 46 |
| away | 118 | said | 46 |
| feel | 118 | away | 45 |
| life | 110 | can | 44 |
| will | 106 | way | 41 |
| time | 96 | never | 36 |
| let | 95 | make | 35 |
| death | 87 | get | 34 |
| way | 87 | love | 34 |
| can | 86 | want | 33 |
| die | 85 | just | 32 |
| like | 85 | think | 32 |
| want | 82 | mind | 30 |
| day | 81 | see | 30 |

Ngưỡng lượng từ thường xuyên được đặt thành 8. Từ xuất hiện ít hơn 8 lần đã bị loại trước đó đào tạo người mẫu. Mô hình cho kết quả tốt nhất khi ngưỡng là 8 hoặc 9. Bằng cách tăng hoặc giảm ngưỡng, bộ phân loại đưa ra nhiều quyết định không chính xác hơn.

## B. Kết quả đạt được và kết luận

Kết quả phân loại của Naive Bayes được hiển thị trong ma trận nhầm lẫn (Table3). Có thể nhận thấy rằng trình phân loại đã tiên lượng không chính xác người biểu diễn 8 trong số 69 lần.

**Table 3. Confusion matrix**

****

Các trường hợp quan tâm đặc biệt là những trường hợp mà bộ phân loại đã không đưa ra quyết định chính xác. Nhìn vào xác suất trong những trường hợp như vậy trong (Table 4), đáng chú ý là năm trong số tám trường hợp người phân loại rất tự tin (hơn 95%) trong quyết định của mình.

**Table 4. Probabilities in incorrect decisions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Actual** | **Predicted** | **Probability Metallica** | **Probability Nirvana** |
| Metallica | Nirvana | 0.02909 | 0.97091 |
| Nirvana | Metallica | 0.97369 | 0.02631 |
| Nirvana | Metallica | 0.54434 | 0.45566 |
| Nirvana | Metallica | 0.99939 | 0.00061 |
| Nirvana | Metallica | 0.52747 | 0.47253 |
| Metallica | Nirvana | 0.02356 | 0.97644 |
| Nirvana | Metallica | 0.52038 | 0.47962 |
| Nirvana | Metallica | 0.99992 | 0.00008 |

Để đánh giá đo lường tính toán, trong ma trận nhầm lẫn lớp quan tâm là Metallica và nó đại diện cho một lớp tích cực. Do đó, Nirvana là một lớp tiêu cực.

The accuracy is (40 + 21) / (40 + 21 + 6 + 2) = 61/69 = 0.88406

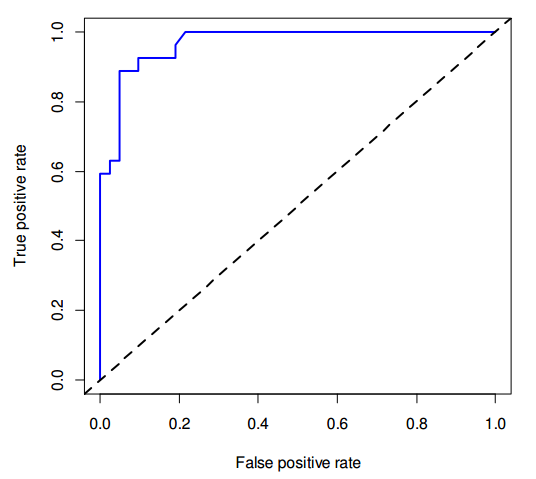
The error is 1 – 0.88406 = 0.11594

The precision is 40 / (40 + 6) = 40/46 = 0.86957

The recall is 40 / (40 + 2) = 40/42 = 0.95238

The F-measure is (2 × 0.86957 × 0.95238) / (0.86957 + 0.95238) = 1.65632 / 1.82195 = 0.90909

Đường cong ROC được thể hiện ở Figure 4. Diện tích dưới đường cong ROC là 0.969, đây là một kết quả tốt.

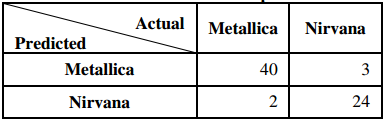


**Figure 4. ROC curve**

## C. Cải tiến Model – Hướng phát triển

Để cải thiện mô hình, phương pháp làm mịn Laplace đã được áp dụng. Mô hình kết quả tốt nhất đã đưa ra khi Laplace giá trị của công cụ ước tính là 0.06. Với điều chỉnh này, bộ phân loại đã nhận ra chính xác ba bài hát nữa của Nirvana (Table 5).

**Table 5. Confusion matrix of improved model**



Cuối cùng, chúng tôi đã kiểm tra hiệu suất của mô hình không có hai biến đổi: loại bỏ các từ dừng và gốc từ. Kết quả mà không làm mịn Laplace và với nó là giống hệt nhau. Table 6 cho thấy sự so sánh của các kết quả.

**Table 6. Comparison of results**

|  |  |  |  |
| --- | --- | --- | --- |
| Measure | Without  Laplace smoothing | Laplace estimator  = 0.06 | Without stemming and  removing stop words |
| precision | 0.86957 | 0.93023 | 0.88636 |
| recall | 0.95238 | 0.95238 | 0.92857 |
| F1-measure | 0.90909 | 0.94117 | 0.90697 |

# IV. KẾT LUẬN

Tạo tập dữ liệu thật tẻ nhạt và tốn thời gian một phần vì nó được tạo theo cách thủ công và một phần vì nghi ngờ về việc chèn một số bài hát vào tập dữ liệu. Cụ thể, các trường hợp như khách xuất hiện của các nhạc sĩ trong album hoặc hai phiên bản của cùng một bài hát (một studio và một phiên bản trực tiếp được thay đổi một chút) phải được xử lý cẩn thận. Bên cạnh đó, nó không phải lúc nào cũng rõ ràng liệu một bài hát có thuộc về một nghệ sĩ biểu diễn hay không - nghi ngờ đã được giải quyết với sự trợ giúp của danh sách các bài hát trên Wikipedia được ghi lại bởi nghệ sĩ được chọn

Kết quả của một mô hình được tạo ra là rất tốt. Naive Bayes phân loại là một lựa chọn tốt cho nhiệm vụ này - một lần nữa nó đã chứng minh khả năng của nó. Vì tập dữ liệu khá nhỏ, nó là một ứng cử viên hợp lý cho mô hình

Kết quả cho thấy bài hát của Nirvana và Metallica có 'chữ ký' văn ​​bản có thể được phân biệt với mức độ lớn chỉ về đọc văn bản. Kết quả nhiều hơn thú vị khi người ta tính đến thực tế là các bài hát cho một ban nhạc thường được viết bởi nhiều tác giả hơn. Trong một số nghiên cứu trong tương lai, sẽ rất thú vị khi xem xét cách mô hình hoạt động trong một số lượng lớn hơn các lớp (nghệ sĩ) và để so sánh kết quả thu được bởi bộ phân loại Naive Bayes với kết quả thu được bởi các bộ phân loại khác, đặc biệt là với SVM.

# TÀI LIỆU THAM KHẢO

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 .

1. E. Stamatatos & G. Widmer, "Music performer recognition using an ensemble of simple classifiers", ECAI, 2002, pp. 335-339.
2. 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.
3. X. Hu, J. S. Downie & A. F. Ehmann, "Lyric text mining in music mood classification", American music, 183.5, 049, 2009, 2-209.
4. S. Howard, C. N. Silla Jr & C. G. Johnson, "Automatic lyricsbased music genre classification in a multilingual setting", Proceedings of the Thirteenth Brazilian Symposium on Computer Music, 2011.
5. 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.
6. E. Stamatatos, "A survey of modern authorship attribution methods", Journal of the Association for Information Science and Technology 60.3, 2009, pp. 538-556.
7. B. Lantz, "Machine learning with R", Packt Publishing Ltd, 2015.
8. K. Ramasubramanian & A. Singh, "Machine Learning Using R", Apress, 2017.
9. F. Peng, "Augmenting Naive Bayes Classifiers with Statistical Language Models", Computer Science Department Faculty

Publication Series, Paper 91, 2003.

1. C. Sammut & G. I. Webb, "Encyclopedia of machine learning and data mining", Springer, 2017.
2. P. Cichosz, "Data mining algorithms: explained using R", John Wiley & Sons, 2015.
3. D. M. Powers, "What the F-measure doesn't measure: Features, Flaws", Fallacies and Fixes. arXiv, 2015.