

MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA
Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică și Microelectronică
Departamentul Informatică și Ingineria Sistemelor

Admis la susținere
Șef departament:
conf. univ. dr. V. Sudacevski

„___” _____ 2023

**Sistem informațional pentru determinarea unui set de
compoziții muzicale în baza clusterizării informației**

Teză de master

Student: Paniș Iulian (_____)

Conducător: Moraru Vasile (_____)

Chișinău – 2023

ADNOTARE

La teza de master cu tema “ Sistem informațional pentru determinarea unui set de compoziții muzicale în baza clusterizării informației” a masterandului gr. MAI-211M Paniș Iulian

Cuvinte cheie: Set de date, Clasificare, Clasterizare, Model, K-means, Cold start, Serviciu Web, API.

Scopul prezentei teze de master constă în crearea unui sistem informațional pentru constituirea sugestiilor muzicale pe baza clusterizării informațiilor, utilizând seturi de date și modele de clasificare.

Obiectivele cercetării includ identificarea și analiza seturilor de date relevante pentru constituirea sugestiilor muzicale, dezvoltarea și implementarea unui model de clusterizare a datelor muzicale, precum și crearea unui serviciu web cu API pentru a permite utilizatorilor să acceseze și să beneficieze de recomandările muzicale.

Aplicația a fost creată utilizând limbajele de programare Python cu bibliotecile aferente necesare pentru prelucrarea și clusterizarea datelor muzicale, precum și limbajul Java cu bibliotecile acestuia pentru crearea API-urilor suitabile. De asemenea, a fost utilizat un algoritm de clusterizare K-means și un filtru de buclă pentru îmbunătățirea precisiei recomandărilor muzicale.

Aplicația creată este utilă pentru orice persoană pasionată de muzică, deoarece oferă sugestii muzicale personalizate, bazate pe preferințele și gusturile individuale. De asemenea, prin intermediul API-ului, aplicația poate fi integrată în alte aplicații sau servicii web pentru a oferi recomandări muzicale personalizate și mai relevante.

ANNOTATION

The master thesis entitled “Information system for determining a set of musical compositions based on information clustering ”

of the master student gr. MAI-201M Paniş Iulian

Keywords: Dataset, Classification, Clusterization, Model, K-means, ColdStart, Web Service, API

The **scope** of this master's thesis is to create an information system for generating music recommendations based on the clustering of information, using datasets and classification models.

The research **objectives** include identifying and analyzing relevant datasets for generating music recommendations, developing and implementing a data clustering model, as well as creating a web service with an API to allow users to access and benefit from music recommendations.

The application was created using Python programming language with the necessary libraries for processing and clustering music data, as well as Java language with its libraries for creating suitable APIs. Additionally, a K-means clustering algorithm and a loop filter were used to improve the accuracy of music recommendations.

The created application is useful for anyone passionate about music, as it offers personalized music suggestions based on individual preferences and tastes. Furthermore, through the API, the application can be integrated into other applications or web services to provide even more personalized and relevant music recommendations.

CUPRINS

INTRODUCERE	8
1 INIȚIEREA ÎN DOMENIU	9
1.1 Domeniul problematic	9
1.2 Scopul și obiectivele	10
1.3 Soluții existente.....	11
1.4 Limitări existente	12
1.5 Implicații etice și personale	13
2 BAZELE TEORETICE ȘI TEHNOLOGICE	15
2.1 Particularitățile recomandărilor muzicale.....	15
2.2 Problema „Cold Start”	17
2.3 Continuarea automată a listei de redare	22
2.4 Analiza și selecția	25
2.5 Clasificarea și Clusterizarea.....	28
3 REALIZAREA ȘI IMPLIMENTAREA PRACTICĂ	34
3.1 Conceptualizarea sistemului informațional	34
3.2 Obținerea și formatarea datelor.....	37
3.3 Clusterizarea și publicarea datelor	40
3.4 Livrarea recomandărilor.....	43
CONCLUZII GENERALE ȘI RECOMANDĂRI	48
BIBLIOGRAFIE	49
ANEXE	52

INTRODUCERE

În era digitală, volume mari de date sunt colectate într-un ritm constant, inclusiv date muzicale. În acest context, dezvoltarea unui sistem informațional pentru analiza și clusterizarea acestor date poate oferi o mai bună înțelegere a compozițiilor muzicale și poate îmbunătăți experiența utilizatorilor cu muzica pe care o ascultă. Acest sistem informațional va fi conceput pentru a oferi utilizatorilor un set de compoziții muzicale relevante, bazate pe preferințele lor personale și pe caracteristicile muzicale ale melodiilor.

Scopul acestei lucrări este să dezvolte un sistem informațional care să determine un set de compoziții muzicale bazat pe clusterizarea informației. Aceasta va implica analiza datelor muzicale prin intermediul unor parametri specifici, precum tonalitatea, tempo-ul, durata și alte caracteristici relevante. În plus, se va explora utilizarea unui furnizor de date muzicale, cum ar fi Spotify, pentru a colecta date și a le analiza în cadrul sistemului.

Sistemul va fi proiectat pentru a fi utilizat ca un web-serviciu, ceea ce va permite utilizatorilor să acceseze și să utilizeze sistemul într-un mod convenabil și eficient. În plus, se va explora utilizarea unei API pentru a permite integrarea sistemului în alte aplicații sau platforme.

Pentru a atinge acest scop, vor fi utilizate metode de analiză și clusterizare a datelor, precum algoritmul k-means, care va fi adaptat și aplicat pentru a clusteriza compozițiile muzicale pe baza caracteristicilor lor comune. În plus, se vor explora metodele de prelucrare a datelor, care vor fi folosite pentru a procesa și analiza volume mari de date muzicale.

Astfel, această teză are ca scop dezvoltarea unui sistem informațional pentru determinarea unui set de compoziții muzicale bazat pe clusterizarea informației. Acest sistem va utiliza metode de analiză și clusterizare a datelor muzicale și va fi proiectat pentru a fi utilizat ca un web-serviciu. Se va explora, de asemenea, utilizarea unui furnizor de date muzicale, și a unei API pentru a permite integrarea sistemului în alte aplicații și platforme.

BIBLIOGRAFIE

1. *How Many Songs Are There In The World?*. Pe: *Audio Captain* [online]. Din 17.10.2022 [citat 02.02.2023]. Disponibil pe Internet: <https://audiocaptain.com/how-many-songs-are-there-in-the-world/>
2. *About Spotify*. Pe: *Spotify AB*. [online] © 2023 [citat 09.02.2023]. Disponibil pe Internet: <https://newsroom.spotify.com/company-info/>
3. *Our strategy and history*. Pe: *Deezer Investors*. [online] © 2022 Deezer [citat 09.02.2023]. Disponibil pe Internet: <https://www.deezer-investors.com/>
4. *The Causes and Effects of “Filter Bubbles” and how to Break Free*. Pe: *Medium*. [online] Din 13.04.2018 [citat 16.02.2023] Disponibil pe Internet: <https://medium.com/@10797952/the-causes-and-effects-of-filter-bubbles-and-how-to-break-free-df6c5cbf919f>
5. Ke Yin Cao. *Improving the Cold Start Problem in Music Recommender Systems*. Publicat Online, 2020 J. Phys.: Conf. Ser. 1651 012067 [citat 16.02.2023] Disponibil pe Internet: <https://iopscience.iop.org/article/10.1088/1742-6596/1651/1/012067/pdf>
6. *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information* Pe: *Pew Research Center*. [online] Din 15.10.2019 [citat 23.02.2023] Disponibil pe Internet: <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>
7. *Privacy by Design*. Pe: *IntersofConsulting* [online] [citat 23.02.2023] Disponibil pe Internet: <https://gdpr-info.eu/issues/privacy-by-design/>
8. SCHÄFER, T., MEHLHORN, C. Can personality traits predict musical style preferences? A meta-analysis [online]. *Personal Individ Differ*, 2017, 116:265–273 [citat 06.03.2023]. Disponibil pe Internet: <https://doi.org/10.1016/j.paid.2017.04.061>
9. KAMINSKAS, M., RICCI, F. Contextual music information retrieval and recommendation: state of the art and challenges [online]. *Comput Sci Rev*, 2012, 6(2):89–119 [citat 06.03.2023]. Disponibil pe Internet: <https://doi.org/10.1016/j.cosrev.2012.04.002>
10. AGGARWAL, Charu. C. *Recommender systems* [online]. Yorktown Heights, NY, USA: Springer International Publishing Switzerland, 2016, 139–166 [citat 07.03.2023].

- ISBN 978-3-319-29659. Disponibil pe Internet:
http://pzs.dstu.dp.ua/DataMining/recom/bibl/laggarwal_c_c_recommender_systems_the_textbook.pdf
11. EGHBAL-ZADEH, H., LEHNER, B., SCHEDL, M., WIDMER, G. I-Vectors for timbre-based music similarity and music artist classification [online]. Malaga, Spain: Conferinta: 16th ISMIR, 2015 [citat 08.03.2023]. Disponibil pe Internet: <http://dx.doi.org/10.13140/RG.2.1.1341.1287>
 12. KNEES, P., SCHEDL, M. Music similarity and retrieval: an introduction to audio- and web-based strategies [online]. Springer Berlin Heidelberg. 2016 [citat 08.03.2023]. Disponibil pe Internet: <https://books.google.it/books?id=MdRhjwEACAAJ>
 13. ELAHI M., Ricci F., Rubens N. Active learning strategies for rating elicitation in collaborative filtering: a system-wide perspective. Publicat Online, 2014 ACM Trans Intell Syst Technol 5(1):13:1–13:33 [citat 09.03.2023]. Disponibil pe Internet: <https://doi.org/10.1145/2542182.2542195>
 14. ELAHI, M., Adaptive active learning in recommender systems [online]. Springer, Berlin, Heidelberg: User Model Adapt Pers, 2011. 414–417 [citat 09.03.2023]. Disponibil pe Internet: https://doi.org/10.1007/978-3-642-22362-4_40
 15. CREMONESI, P., ELAHI, M., GARZOTTO, F. User interface patterns in recommendation-empowered content intensive multimedia applications [online]. Multimed Tools Appl, 2017. 76(4):5275–5309 [citat 09.03.2023]. Disponibil pe Internet: <https://doi.org/10.1007/s11042-016-3946-5>
 16. *Announcing MIDiA's State Of The Streaming Nation 2 Report*. MiDiA, 05.06.2017 [citat 09.03.2023]. Disponibil pe Internet <https://www.midiaresearch.com/blog/announcing-midias-state-of-the-streaming-nation-2-report>
 17. PICHL, M., ZANGERLE, E., SPECHT, G. Towards a context-aware music recommendation approach: what is hidden in the playlist name? [online]. ICDMW, 2015. IEEE, pp 1360–1365 [citat 09.03.2023] Disponibil pe Internet: <https://evazangerle.at/publication/pichl-somera-2015/pichl-somera-2015.pdf>
 18. *Classification vs. Clustering*. Simplilearn, 23.03.2023 [citat 27.03.2023]. Disponibil pe Internet: <https://www.simplilearn.com/tutorials/data-analytics-tutorial/classification-vs-clustering>

19. DABBURA, I. *K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks*, Publicat Online, 17.09.2018 [citat 27.03.2023]. Disponibil pe Internet: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
20. *K-means Clustering Algorithm: Applications, Types, & How Does It Work?* Simplilearn, 16.02.2023 [citat 27.03.2023]. Disponibil pe Internet: <https://www.simplilearn.com/tutorials/machine-learning-tutorial/k-means-clustering-algorithm>
21. *What is an API?*. RedHat [online] 02.06.2022 [citat 03.05.2023]. Disponibil pe Internet: <https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces>
22. *Get Track's Audio Analysis*. Spotify for Developers [online] [citat 03.05.2023]. Disponibil pe Internet: <https://developer.spotify.com/documentation/web-api/reference/get-audio-analysis>
23. T. ALADE. Tutorial: How to determine the optimal number of clusters for k-means clustering [online]. 27.05.2018 [citat 05.05.2023] Disponibil pe Internet: <https://blog.cambridgespark.com/how-to-determine-the-optimal-number-of-clusters-for-k-means-clustering-14f27070048f>
24. *Authorization Code Flow*. Spotify for Developers [online] [citat 05.05.2023]. Disponibil pe Internet: <https://developer.spotify.com/documentation/web-api/tutorials/code-flow>