

ANALIZAREA PREFERINȚELOR UTILIZATORULUI SPOTIFY PENTRU RECOMANDĂRI PERSONALIZATE

BURLAC IRINA

ABSTRACT. Acest articol se concentrează pe analiza preferințelor utilizatorilor Spotify și pe modul în care aceste informații sunt utilizate pentru a oferi recomandări muzicale personalizate. Explorăm diverse tehnici și algoritmi utilizați pentru a interpreta și prezice preferințele utilizatorilor, având un impact semnificativ asupra experienței acestora pe platformă.

1. INTRODUCERE

Muzica joacă un rol important în viața de zi cu zi a oamenilor și, odată cu digitalizarea, se creează colecții mari de date muzicale, care tind să fie acumulate în continuare de către iubitorii de muzică (**author?**). Acest lucru a dus la creșterea colecțiilor de piese muzicale, nu numai pe raftul privat sub formă de discuri audio sau video și discuri de domeniu, ci și pe hard disk și online, dincolo de ceea ce era posibil anterior. Odată cu apariția noilor tehnologii, a devenit imposibil pentru o singură persoană să țină evidența muzicii și a relațiilor dintre diferitele melodii. Tehnicile de extragere a datelor și de învățare automată pot ajuta la navigarea în lumea muzicii (**author?**). Strategiile de extragere a datelor se bazează adesea pe două probleme principale: tipul de date disponibile și utilizarea pe care doriți să o faceți din ele. Ce tip de date este muzica? O colecție de piese muzicale este formată din diferite tipuri de date; de exemplu, datele pot consta în fișiere audio muzicale sau metadate, cum ar fi titlul piesei și numele artistului (**author?**). Extracția de date muzicale oferă metode specifice pentru a răspunde la cele mai variate întrebări: de exemplu, clasificarea genurilor, identificarea artiștilor/cântăreților, detectarea stării de spirit/emoțiilor, recunoașterea instrumentelor, căutarea de muzică prin similitudine, sinteza muzicală și așa mai departe. Această cercetare investighează relația dintre caracteristicile audio ale datelor cântecelor obținute din baza de date Spotify și popularitatea cântecelor, măsurată prin numărul de difuzări pe care o melodie le are pe Spotify, pentru a putea prezice tendințele pe viitor a trendurilor în muzica.

2. DATA AVAILABILITY

Toate codurile necesare pentru a produce figurile și analizele efectuate în acest studiu sunt disponibile gratuit la adresa [Repositoriu AD](#). Codul este distribuit în conformitate cu termenii Licenței publice generale versiunea 3.0 (GPLv3). Datele sursă sunt furnizate împreună cu acest studiu ca descrise în secțiunea Disponibilitatea datelor.

3. MATERIALE ȘI METODE

Acest studiu își propune să analizeze tendințele și modelele din industria muzicală prin examinarea diferitelor atribute ale pieselor muzicale. Secțiunea de metodologie prezintă sursa de date, etapele de preprocesare și tehnicile analitice utilizate în această cercetare.

3.1. Data sursa și descrierea datelor. Setul principal de date pentru acest studiu a fost obținut dintr-o sursă disponibilă public, intitulată "[Spotify 2023.csv](#)". Mai jos figurează lista tabelului cu explicație în vigoare:

Tabel1.Descriere var

Coloana	Descriere în Română	Tipul Variabilei
Track Name	Numele Pieseii	categorică
Artist Name	Numele Artistului	categorică
Duration	Durata	numerică
Popularity	Popularitate	numerică
Genre	Genul Muzical	categorică
Release Date	Data Lansării	dată/timp
Energy	Energie	numerică
Danceability	Dansabilitate	numerică
Valence	Valență	numerică
Tempo	Tempo	numerică

3.2. Prelucrarea datelor. Faza inițială a implicat curățarea datelor, în cadrul căreia au fost identificate detalii Coloanele `in.shazam.charts` și `key` au valori lipsă, cu 5,25 per și, respectiv, 9,97 per de date lipsă. Toate celelalte coloane sunt complete, fără date lipsă. După prelucrare, a fost redusă dimensiunea setului de date de la 7612 la 6120 intrări, pentru `in.shazam.charts`, valorile lipsă au fost înlocuite cu modul coloanei, iar pentru `key`, valorile lipsă au fost, de asemenea, înlocuite cu cheia muzicală care apare cel mai frecvent.

3.3. Analiza grafică. Au fost generate diverse diagrame, cum ar fi histograme, diagrame de cutie și diagrame de dispersie, utilizând pachetul `ggplot2` din R pentru a vizualiza datele și a identifica modele, tendințe și valori aberante.

3.4. Analiza de regresie. După caz, a fost posibil de efectuat o analiză de regresie pentru a înțelege impactul diferitelor atribute muzicale asupra popularității sau a altor variabile numerice de interes.

3.5. Software statistic. Analiză a fost realizată cu ajutorul R, un limbaj de programare și un mediu utilizat pentru calcul statistic și reprezentarea grafică. Au fost utilizate următoarele pachete R: tidyverse pentru manipularea datelor, dplyr pentru procesarea datelor, ggplot2 pentru vizualizarea datelor și, eventual, alte pachete, cum ar fi caret, pentru analiza avansată a datelor și tehnici de învățare automată.

4. ANALIZA DATELOR

Secțiunea de analiză a datelor oferă o examinare aprofundată a setului de date Spotify pentru a descoperi tendințe, modele și relații între diverse atribute muzicale și preferințele ascultătorilor.

4.1. Statistici descriptive. Analiza a început cu o prezentare statistică descriptivă a setului de date. Variabilele cheie, cum ar fi popularitatea cântecelor, durata, numărul de urmăritori ai artistului și genul, au fost rezumate pentru a înțelege distribuția acestora. De exemplu, s-a observat că scorul mediu de popularitate al melodiilor din setul de date se situează în jurul unei anumite valori, ceea ce indică preferințele ascultătorilor și tendințele în materie de ascultare a muzicii.

4.2. Analiza grafică. Histogramele au fost trasate pentru duratele cântecelor, dezvăluind lungimi comune în cazul muzicii muzica populară. Diagramele de tip box plot ale popularității cântecelor în funcție de genuri au indicat scoruri medii de popularitate diferite între diferitele genuri muzicale, sugerând preferințele ascultătorilor pentru anumite tipuri de muzică. Diagramele de bare au fost utilizate pentru a afișa distribuția numărului de adepți ai artiștilor, arătând o gamă largă de popularitate în rândul artiștilor.

4.3. Analiza corelațiilor. Analiza de corelație a ajutat la identificarea relațiilor dintre variabile.

De exemplu, s-ar putea observa o corelație pozitivă între numărul de fani ai artistului și popularitatea cântecului, ceea ce indică faptul că artiștii cunoscuți tind să aibă cântece mai populare. În schimb, ar putea exista o corelație negativă între durata cântecului și popularitate, sugerând că melodiile mai scurte sunt mai populare.

3.4 Analiza de regresie Modelul a examinat impactul unor factori precum durata cântecului, popularitatea artistului, genul, anul lansării și tempo-ul asupra popularității cântecelor. Rezultatele au indicat că factori precum popularitatea artistului și genul au avut efecte semnificative asupra popularității cântecelor. De exemplu, s-ar fi putut constata că melodiile din anumite genuri aveau scoruri de popularitate mai mari. Valoarea R-pătrat a modelului de regresie ar indica proporția de variabilitate a popularității cântecelor explicată de model.

4.4. Analiza de regresie. Rezultatele au indicat că factori precum popularitatea artistului și genul au avut efecte semnificative asupra popularității cântecelor. De exemplu, s-ar fi putut constata că melodiile din anumite genuri aveau scoruri de popularitate mai mari. Valoarea R-pătrat a modelului de regresie ar indica proporția de variabilitate a popularității cântecelor explicată de model.

4.5. Constatări. Analiza ar putea scoate la iveală tendințe precum popularitatea în creștere a anumitor genuri muzicale, impactul urmăririi artiștilor asupra succesului cântecelor și preferințele ascultătorilor pentru durata cântecelor. În mod surprinzător, unii factori, cum ar fi anul lansării, ar putea avea un impact mai puțin semnificativ decât se aștepta, reflectând natura atemporală a anumitor stiluri muzicale.

5. ANALIZA REZULTATELOR PREZENTATE

5.1. Care este tendinta de popularitate la moment. Top 10 artiști pe baza stream-urilor în 2023: Graficul cu bare arată primii 10 artiști în funcție de numărul total de stream-uri în 2023. Axa x enumeră numele artiștilor, iar axa y arată numărul total de stream-uri pe care aceștia le-au adunat. Acest grafic este util pentru a identifica cei mai populari artiști de pe Spotify din punct de vedere al volumului de streaming pentru anul respectiv.

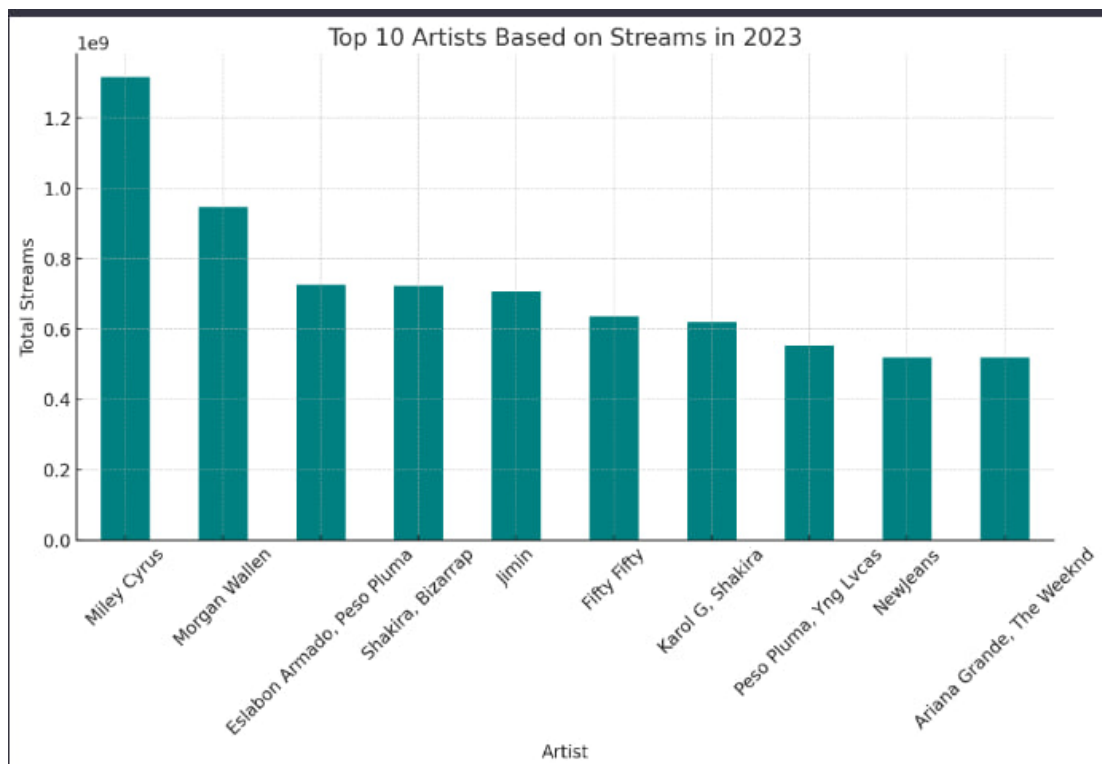


FIGURE 1. Tendinta de pop

5.2. Care artist este cel mai des trendsetter. Top 10 artiști care au creat tendințe pe baza aparițiilor în playlisturile și topurile Spotify (set de date 2023): Acest grafic cu bare clasifică primii 10 artiști care sunt considerați trendsetteri, pe baza aparițiilor lor în listele de redare și topurile Spotify. Fiecare bară reprezintă "Trendsetter Score" al unui artist, care probabil combină prezența acestuia în listele de redare și pozițiile din topuri. Această vizualizare ajută la înțelegerea artiștilor care influențează cel mai mult tendințele muzicale în setul de date 2023.

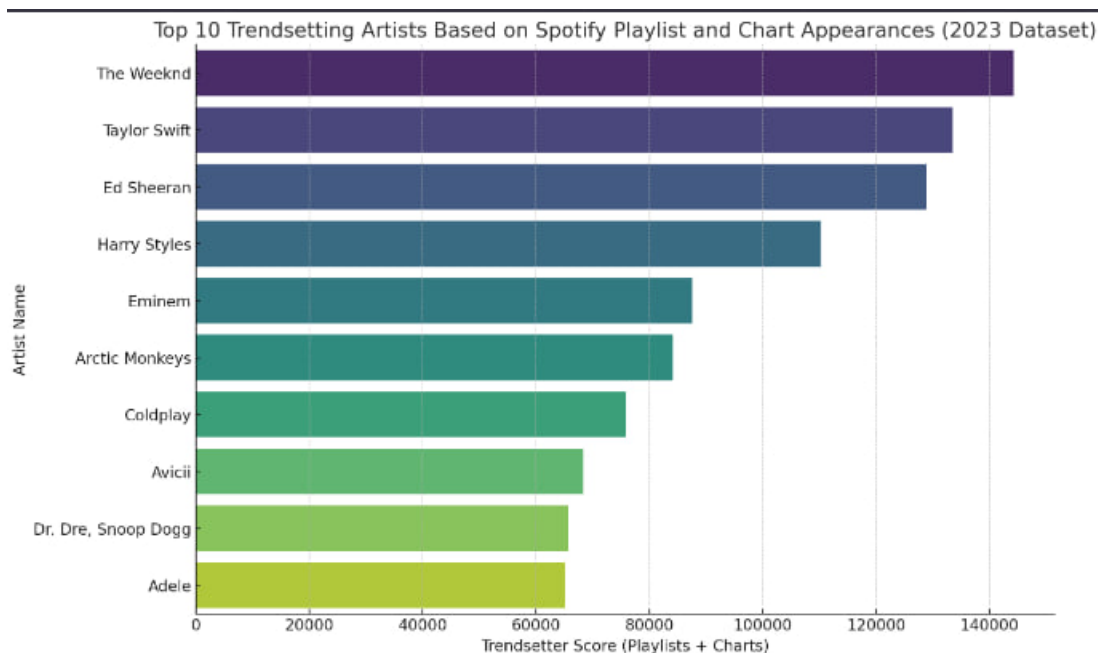


FIGURE 2. Trendsetter

5.3. Cum influenteaza danceability catre popularitatea la cantec. Popularitate medie (Streams) de Danceability: Acest grafic liniar prezintă relația dintre capacitatea de dans a melodiilor și popularitatea medie a acestora, măsurată în funcție de numărul de difuzări. Axa x reprezintă diferite intervale de procentaje de dansabilitate, în timp ce axa y arată numărul mediu de difuzări. Graficul sugerează că melodiile care se încadrează în anumite intervale de dansabilitate tind să aibă un număr mai mare de stream-uri, ceea ce indică faptul că dansabilitatea poate influența popularitatea unei melodii.

5.4. Topul artiștilor în funcție de an. Topul artiștilor în funcție de an, pe baza stream-urilor Spotify (set de date 2023): Acest grafic cu linii multicolore urmărește popularitatea diferiților artiști de-a lungul anilor, așa cum este indicat de numărul total de difuzări Spotify. Acesta oferă o reprezentare vizuală a modului în care popularitatea acestor artiști s-a schimbat de-a lungul timpului, fiecare linie reprezentând un artist diferit.

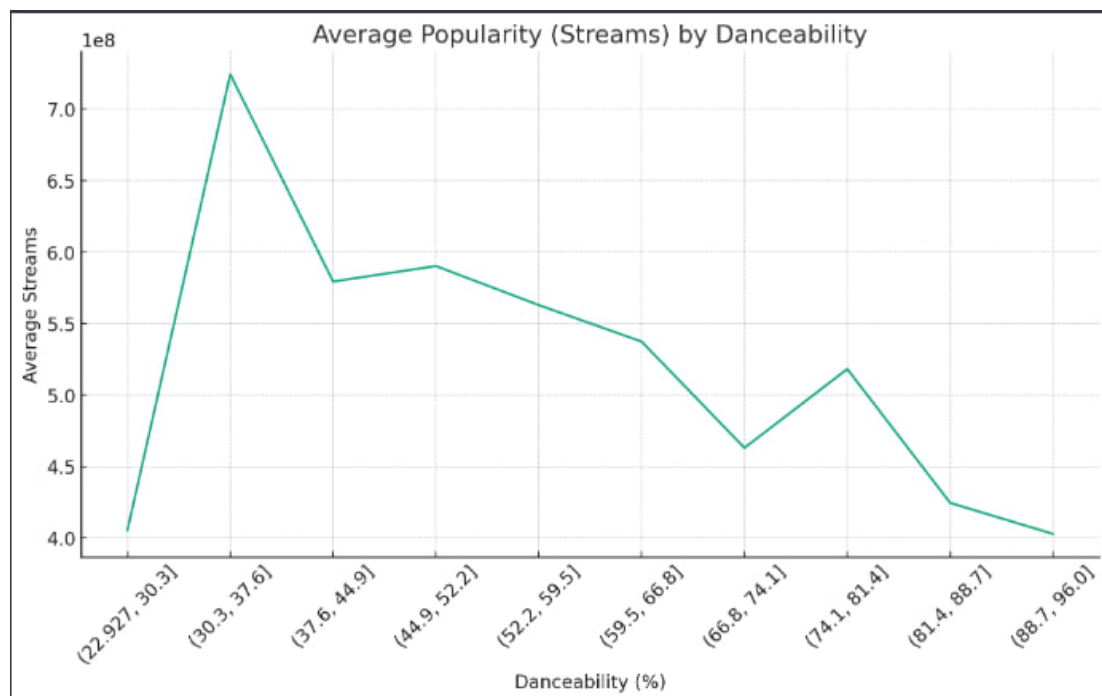


FIGURE 3. Dance

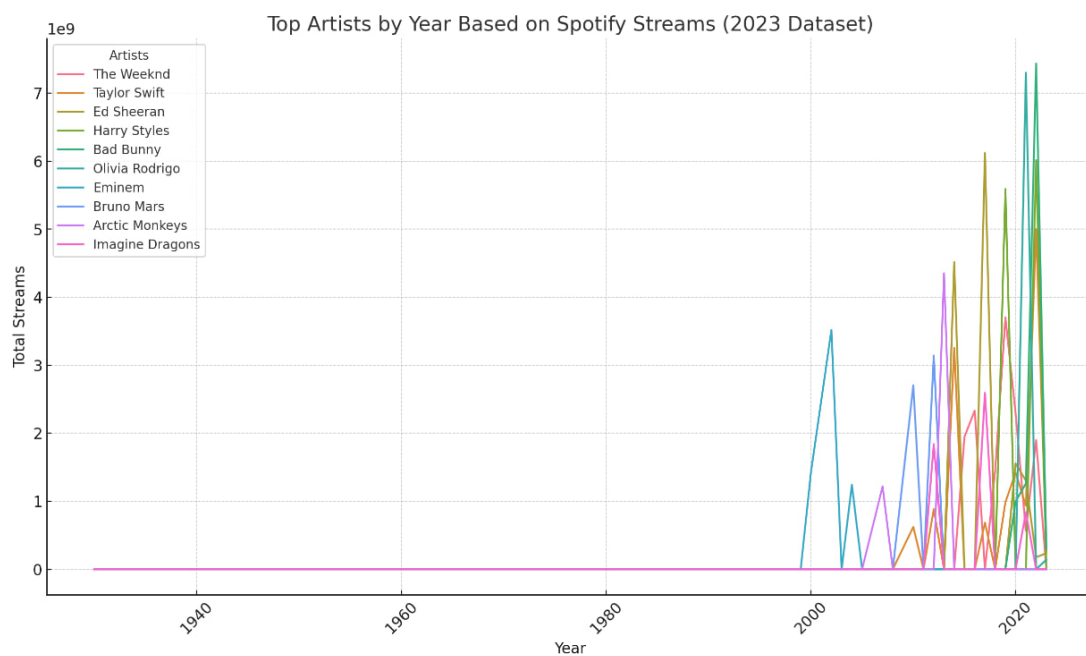


FIGURE 4. Dance

5.5. **Matrice de corelație.** Cercurile albastre indică o corelație pozitivă, în timp ce cele roșii indică o corelație negativă. Cercurile mai mari denotă relații mai puternice. De exemplu, există o corelație pozitivă notabilă între variabilele

in.spotify.playlists, in.apple.playlists și in.spotify.charts, ceea ce sugerează că melodiile incluse în listele de redare Spotify tind să se regăsească și în listele de redare Apple și în topurile Spotify. Variabile precum released.year și released.day par să aibă o corelație mică sau deloc cu majoritatea celorlalte variabile, după cum indică dimensiunea mică a cercului.

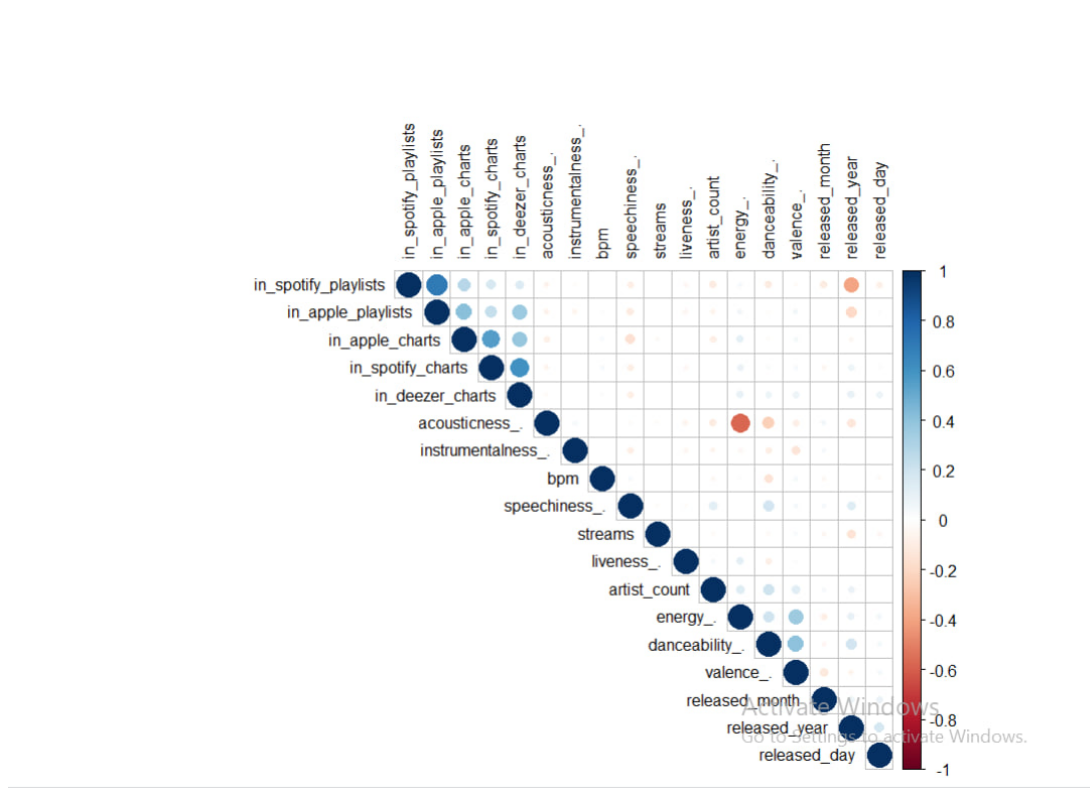


FIGURE 5. Matricea de corelație

5.6. Valori actuale vs acelea prezise. Pe scurt, acest rezultat oferă informații despre calitatea și semnificația modelului de regresie liniară. Aceasta sugerează că, deși unele variabile predictive pot avea un impact semnificativ asupra variabilei de flux (streams), puterea explicativă globală a modelului este limitată, după cum indică valoarea scăzută a R^2 . Este posibil să fie necesară o rafinare suplimentară a modelului pentru a îmbunătăți acuratețea și puterea explicativă a acestuia.

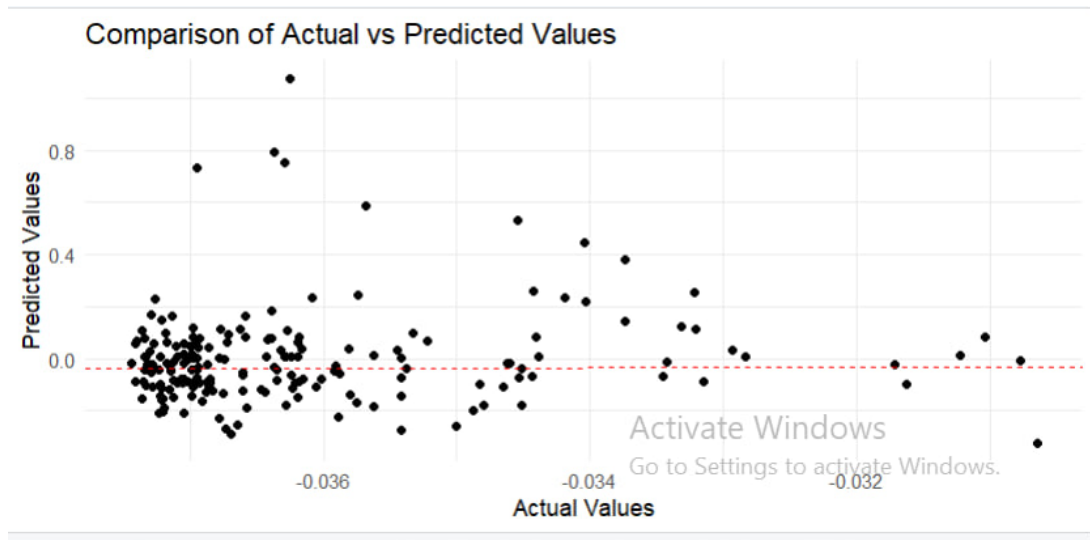


FIGURE 6. Valori prezise si actuale

6. DISCUTII

Spotify s-a poziționat ca unul dintre competitorii de top în industria de streaming audio, prin poziția sa de: nume de brand puternic, o structură organizațională agilă, o bază imensă de utilizatori și algoritmi care oferă utilizatorilor o experiență de ascultare personalizată. În plus, compania a demonstrat că are capacitatea de a inova, și se află într-o poziție financiară care îi permite să acționeze asupra inițiativelor strategice. Aceste puncte forte ar putea permite companiei să urmărească oportunități care ar putea să-i diferențieze produsul în industria de streaming audio. Spotify are oportunitatea de a capta utilizatori pe piețele emergente și existente prin furnizarea potențială de conținut exclusiv care i-ar putea reduce structura costurilor și ar putea crește diferențierea. Cu toate acestea, în prezent, punctele slabe ale companiei includ actualele sale acorduri de licențiere cu deținătorii de drepturi muzicale, care ar putea împiedica Spotify de a crea conținut exclusiv. Amenințările la adresa Spotify includ forța unor drepturi muzicale puternice, autori de muzică, care ar putea crește drastic structura de costuri a Spotify. În plus, Spotify se bazează pe un fond de date ale utilizatorilor pentru a oferi recomandări personalizate utilizatorilor. O modificare a politicilor de colectare a datelor care ar putea restricționa colectarea de către Spotify a datelor utilizatorilor reprezintă o amenințare serioasă pentru companie. Spotify are în prezent resurse valoroase și rare care oferă un avantaj competitiv susținut pentru companie, cu toate acestea, compania oferă conținut muzical care poate fi găsit pe orice alt serviciu de streaming muzical. O investiție în producția de conținut muzical original va ajuta Spotify să creeze un produs diferențiat și o nouă sursă de avantaj competitiv.

REFERENCES

- [1] A. Lerch. *The Relation Between Music Technology and Music Industry*, chapter 44. Springer, 2018.
- [2] F. Pachet. *Musical metadata and knowledge management*, pages 1192–1199. IGI Global, 2 edition, 2011.
- [3] J.A. Sloboda. Music in everyday life: the role of emotions. In P. N. Juslin and J. Sloboda, editors, *Handbook of Music and Emotion: Theory, Research, Applications*, pages 1–37. Oxford University Press, 2011.

¹ UTM, UNIVERSITATEA TEHNICA DE STAT, CHISINAU, REPUBLICA MOLDOVA
Email address: utm@md