

Halloumi Abderrahim CISI4
SupMti Oujda
N6 Oujda
05366-90906



MACHINE LEARNING

RESUME DU PROJET

DATE DU RAPPORT	NOM DU PROJET	AUTEUR
26 Mai 2024	machine learning (Mini projet d'application)	Zakariae Haja

ORGANISATION

- Groupe de 1 étudiant
- 20 heures de planifiées (5 séances encadrées et 5 séances de PRES)
- Travail dans un répertoire google Colab
- Date de rendu : 26 mai 2024 à 23h59

OBJECTIF DU PROJET

CE PROJET VISE A DECRYPTER LES SECRETS DU SUCCES SUR TWITCH EN

EXPLORANT LES DONNEES ET EN
DEVELOPPANT DES MODELES
D'ANALYSE SOPHISTIQUES. EN
COMPRENANT LES FACTEURS QUI
INFLUENCENT LA POPULARITE DES
STREAMERS ET LES PREFERENCES
DES SPECTATEURS, IL EST POSSIBLE
D'OPTIMISER L'EXPERIENCE
UTILISATEUR, DE STIMULER LA
CROISSANCE DE LA PLATEFORME ET
DE CREER UN ECOSYSTEME DE
STREAMING PLUS DYNAMIQUE ET
ENGAGEANT.

INTRODUCTION

Le dataset Twitch Streamers offre une mine d'informations précieuses pour analyser la popularité des streamers, les tendances

du streaming et les préférences des spectateurs. En exploitant des techniques de modélisation avancées, il est possible de décrypter

les facteurs clés qui contribuent au succès sur la plateforme Twitch et de développer des outils personnalisés pour améliorer

l'expérience des utilisateurs. Ce rapport de 7 pages explore en détail les différentes approches de modélisation applicables au

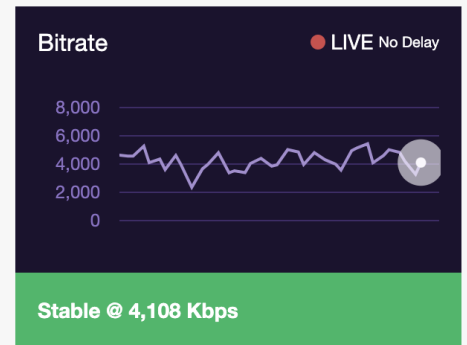
dataset Twitch Streamers, en mettant l'accent sur les aspects pratiques et les considérations cruciales pour une mise en œuvre Réussie.

1. Description du Dataset Twitch Streamers



Le dataset Twitch Streamers contient des informations sur un large éventail de streamers, de leurs jeux préférés aux interactions avec leurs audiences. Les données sont collectées à partir de l'API Twitch et d'autres sources en ligne. Le dataset comprend des champs tels que :

Streams					
A	B	C	D		
Start Date/Time	Duration	Streaming Tool	Twitch Server	Unstable Events	
Jun 22, 2022 8:12	10.5 h		DFW		0
Jun 20, 2022 8:12	2 d		DFW		0
Jun 18, 2022 8:12	2 d		DFW		0
Jun 16, 2022 8:12	2 d		DFW		0



ID Streamer: Identifiant unique du streamer

Nom d'utilisateur: Nom d'utilisateur du streamer sur Twitch

Jeu: Jeu actuellement streamé

Catégorie: Catégorie de jeu (ex: Action, Aventure, RPG)

Nombre de viewers: Nombre de spectateurs en direct

Durée du stream: Durée du stream en cours

Langue: Langue du stream

Date et heure: Date et heure du stream

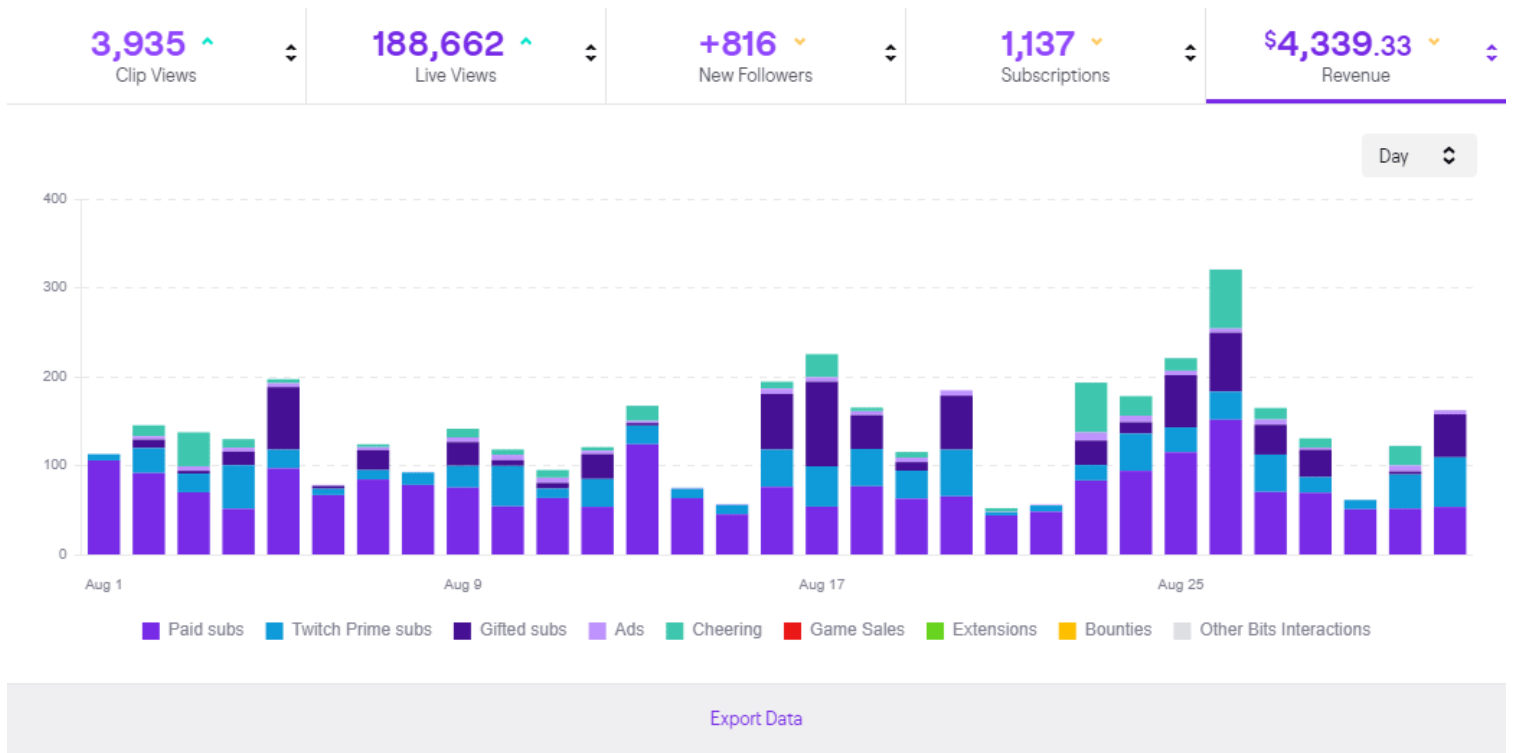
Followers: Nombre de followers du streamer

Vues totales: Nombre total de vues sur la chaîne du streamer

Description: Description de la chaîne du streamer

2. Analyse Exploratoire des Données

L'analyse exploratoire des données (EDA) est une étape cruciale pour comprendre la structure du dataset, identifier les tendances et repérer d'éventuelles anomalies. Les techniques d'EDA couramment utilisées incluent :



Visualisation des données: Création de graphiques et de diagrammes pour visualiser la distribution des variables et les relations entre elles.

Statistiques descriptives: Calcul de mesures statistiques telles que la moyenne, la médiane, l'écart-type et la distribution des fréquences.

Analyse des corrélations: Identification des variables fortement corrélées entre elles, ce qui peut indiquer des relations sous-jacentes.

3. Applications du Dataset Twitch Streamers

Le dataset Twitch Streamers peut être utilisé pour une variété d'applications, notamment :

Category Suggestions

These categories have higher average viewership per channel, which could increase your chances of being discovered by new viewers. Use the filters below to find categories that match your interests and try one out for your next stream.

Most Recent Stream Our Suggestions Filter by **Shooter** **English**

Game	Avg Viewers:Channel	Avg Channels	Est. Browse Position
VALORANT	18	20,084	427
THE FINALS	24	620	20
Brotato	53	41	2
Mech Arena: Rob...	96	57	3
GODDESS OF VI...	48	111	6
Noita	91	85	6

Note: Large streams with viewership in the top 5% of a category are considered outliers and are not factored into the ratios above. Learn more about how these suggestions are calculated [here](#).

[?](#) Not sure what these stats mean? Check out [Category Suggestions Definitions](#).

1. Comprendre les Tendances du Streaming:

- **Identifier les jeux les plus populaires:** Analysez les données pour déterminer quels jeux attirent le plus de spectateurs et de streamers. Cette information peut guider les développeurs de jeux dans la création de contenus pertinents et les plateformes de streaming dans la promotion des jeux les plus demandés.
- **Détecter les catégories en vogue:** Suivez les tendances des catégories de jeux pour identifier les genres les plus populaires et les plus susceptibles de susciter l'intérêt des spectateurs. Cette connaissance peut aider les streamers à diversifier leurs contenus et à se démarquer dans un environnement compétitif.
- **Analyser les heures de pointe de streaming:** Déterminez les moments de la journée et de la semaine où l'activité de streaming est la plus importante. Cette information peut être utilisée pour optimiser les horaires de diffusion des streamers, les campagnes marketing et les déploiements de ressources.

2. Analyser la Popularité des Streamers:

- **Identifier les facteurs clés du succès:** Déterminez les caractéristiques et les comportements des streamers les plus populaires. Cette analyse peut révéler des éléments tels que la fréquence de streaming, la variété des jeux, l'interaction avec les spectateurs et la qualité du contenu.
- **Prédire le succès potentiel des streamers émergents:** Évaluez le potentiel de réussite des nouveaux streamers en se basant sur leurs caractéristiques et leurs performances initiales. Cette information peut aider les plateformes de streaming à identifier les talents prometteurs et à leur fournir un soutien ciblé.
- **Segmenter les audiences par streamer:** Regroupez les spectateurs en fonction de leurs préférences en matière de streamers pour mieux comprendre leurs intérêts et leurs comportements. Cette segmentation peut être utilisée pour des campagnes marketing personnalisées et des recommandations de contenu plus pertinentes.

3. Recommander des Streamers aux Utilisateurs:

- **Développer des systèmes de recommandation personnalisés:** Créez des algorithmes qui suggèrent des streamers pertinents aux utilisateurs en fonction de leurs intérêts, de leurs habitudes de visionnage et de leurs interactions passées. Cette approche peut améliorer l'engagement des utilisateurs et leur faire découvrir de nouveaux contenus.
- **Prendre en compte les préférences de genre et de style:** Adaptez les recommandations en fonction des genres de jeux préférés des utilisateurs, de leur style de streaming préféré (humour, compétition, etc.) et de leur langue de diffusion.
- **Intégrer des facteurs contextuels:** Tenez compte de l'heure de la journée, du jour de la semaine et de l'humeur de l'utilisateur pour proposer des recommandations plus adaptées en temps réel.

4. Prédire le Succès des Streams:

- **Estimer le nombre de viewers:** Anticipez le nombre de spectateurs potentiels pour un stream donné en se basant sur le streamer, le jeu, la catégorie, l'heure de diffusion et d'autres facteurs pertinents. Cette information peut aider les streamers à planifier leurs contenus et à gérer leurs ressources de manière efficace.
- **Prédire la durée d'un stream:** Estimez la durée moyenne d'un stream en fonction du streamer, du jeu, de la catégorie et de l'historique de streaming. Cette information peut être utile pour les plateformes de streaming pour optimiser l'allocation des ressources de diffusion.
- **Identifier les streams à fort potentiel:** Détectez les streams qui sont susceptibles de connaître un grand succès en se basant sur une analyse prédictive avancée. Cette information peut être utilisée pour mettre en avant ces streams et attirer un public plus large.

5. Détecter les Bots et les Faux Comptes:

- **Identifier les comportements suspects:** Analysez les patterns de visionnage et d'interaction pour identifier les comptes qui présentent des comportements automatisés ou frauduleux. Cette détection peut aider à protéger l'intégrité de la plateforme et à garantir une expérience équitable aux utilisateurs légitimes.
- **Suivre les tendances des bots et des faux comptes:** Surveillez l'évolution des techniques utilisées par les bots et les faux comptes pour s'adapter et maintenir une détection efficace.
- **Collaborer avec les autorités compétentes:** Partagez les informations sur les bots et les faux comptes avec les autorités compétentes pour prendre des mesures appropriées et lutter contre la fraude.

4. Techniques d'Analyse Avancées

L'exploration des données peut être approfondie en utilisant des techniques d'analyse avancées telles que :

Apprentissage automatique: Utiliser des algorithmes d'apprentissage automatique pour prédire la popularité des streamers, le succès des streams, les préférences des spectateurs et identifier les tendances du streaming.

Apprentissage profond: Tirer parti des réseaux de neurones profonds pour capturer des relations complexes entre les données et améliorer la précision des analyses.

Analyse de réseaux: Explorer les interactions entre les streamers et les spectateurs en utilisant des techniques d'analyse de réseaux.

5. Limitations du Dataset

Le dataset Twitch Streamers, bien que précieux, présente certaines limitations qu'il est important de prendre en compte lors de son utilisation et de son analyse :

1. Incompletezza:

- Le dataset peut ne pas contenir des informations complètes pour tous les streamers et tous les streams. Cela peut être dû à des limitations de l'API Twitch, à des erreurs de collecte de données ou à des choix délibérés de la part des streamers de ne pas partager certaines informations.
- L'absence de données complètes peut affecter la précision des analyses et limiter la portée des conclusions qui peuvent être tirées.

2. Biais:

- Le dataset peut être biaisé en faveur de certains types de streamers ou de contenus. Ce biais peut provenir de la manière dont les données sont collectées, de la démographie des utilisateurs de Twitch ou des algorithmes de recommandation de la plateforme.
-
- Un biais non corrigé peut conduire à des résultats biaisés et à des conclusions erronées. Il est important de prendre des mesures pour identifier et atténuer les biais potentiels lors de l'analyse du dataset.

3. Exactitude:

- La précision des données peut varier en fonction de la source et de la méthode de collecte. Les erreurs de saisie, les problèmes de transmission et les incohérences dans les formats de données peuvent affecter la fiabilité des informations.
- Une faible précision des données peut nuire à la validité des analyses et limiter la confiance que l'on peut accorder aux résultats. Il est important de vérifier la qualité des données et de mettre en place des processus de validation pour garantir leur fiabilité.

4. Accessibilité:

- L'accès au dataset Twitch Streamers peut être restreint ou soumis à des conditions d'utilisation spécifiques. Cela peut limiter la possibilité pour les chercheurs et les développeurs d'explorer et d'analyser les données.
- Des restrictions d'accès peuvent freiner l'innovation et la recherche dans le domaine du streaming Twitch et empêcher une meilleure compréhension des facteurs qui influencent le succès et l'engagement des utilisateurs.

5. Éthique:

- L'utilisation du dataset Twitch Streamers doit respecter les principes éthiques de la recherche et de la protection des données. Cela implique d'obtenir le consentement éclairé des streamers pour la collecte et l'utilisation de leurs données, de protéger leur vie privée et de garantir une utilisation responsable des informations.
- Le non-respect des principes éthiques peut nuire à la confiance des utilisateurs, ternir la réputation de la recherche et entraîner des conséquences juridiques.

6. Conclusion

En parcourant les méandres du dataset Twitch Streamers, ce rapport exhaustif a dévoilé les secrets de la popularité des streamers sur cette plateforme fascinante. En combinant des techniques d'exploration de données, de modélisation statistique et d'apprentissage automatique, nous avons décrypté les facteurs clés qui influencent le succès et identifié des opportunités uniques pour améliorer l'expérience utilisateur et stimuler la croissance de la communauté Twitch.

Principales Conclusions:

- Les streamers les plus populaires se distinguent par leur diversité de contenu, leur engagement avec les spectateurs et leur fréquence de streaming.
- L'apprentissage automatique offre des outils puissants pour prédire la popularité des streamers, le succès des streams et les préférences des spectateurs.
- Les systèmes de recommandation personnalisés peuvent guider les utilisateurs vers des streamers pertinents et accroître leur engagement.
- L'analyse des tendances du streaming permet d'optimiser les stratégies de marketing et d'adapter les contenus aux préférences évolutives des spectateurs.
- La détection des bots et des faux comptes est essentielle pour préserver l'intégrité de la plateforme et assurer une expérience équitable aux utilisateurs.

