

Список використаних джерел

1. Alpaydin E. Introduction to Machine Learning. Book. The MIT Press. 2014. 600 p.
2. Bishop C. M. Pattern Recognition and Machine Learning. Book. Springer. 2006. 738 p.
3. Goodfellow I., Bengio Y., Courville A. Deep Learning. Book. MIT Press. 2016. 775 p.
4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Book. Springer 2009. 745 p.
5. Mitchell T. M. Machine Learning. Book. McGraw-Hill. 1997. 414 p.
6. Russell S. J., Norvig P. Artificial Intelligence: A Modern Approach. Book. Prentice Hall. 2010. 1152 p.

УДК 004.9

*Лавренюк Б. В., здобувач вищої освіти,
Потапова Н. А., канд. екон. наук, доцент,
доцент кафедри інформаційних технологій*

ВИКОРИСТАННЯ АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТА ВИЯВЛЕННЯ ТРЕНДІВ У СВІТІ КІНЕМАТОГРАФІЇ

Донецький національний університет імені Василя Стуса, м. Вінниця

Сучасний кінематограф все більше інтегрується з технологіями, що дає змогу глибше розуміти вподобання аудиторії та передбачати майбутні тенденції. Одним із найпотужніших інструментів для цього є аналіз даних. Використання великих обсягів інформації дає змогу не лише прогнозувати успіх кінопроектів, але й оперативно реагувати на зміни у запитах глядачів. У цьому контексті розкриття можливостей аналізу даних у кіноіндустрії набуває стратегічної значущості.

З огляду на те, що стримінгові сервіси, соціальні мережі та онлайн-рецензії формують глобальний кіноринок, кінокомпанії зіштовхуються з викликом не лише створення якісного продукту, а й конкурентної боротьби за увагу глядача. Аналіз даних допомагає не лише передбачити успішність майбутніх проєктів, але й знижує фінансові ризики за допомогою обґрунтованого прийняття рішень. У світі, де смаки аудиторії змінюються надзвичайно швидко, здатність виявляти тренди стає визначальним чинником успіху.

Аналіз даних став одним із ключових інструментів у світі кінематографії для виявлення трендів і прогнозування успіху проєктів. Провідні гравці індустрії активно використовують сучасні технології аналізу, як-от машинне навчання, обробка природної мови (NLP), кластеризація даних і алгоритми глибокого навчання, щоб зрозуміти вподобання аудиторії та адаптуватися до змін на ринку.

Netflix, один із лідерів стримінгового ринку, активно застосовує алгоритми машинного навчання для аналізу даних про глядачів. Система аналізує мільярди годин переглядів, включно з даними про час перегляду, пропуски, паузи та повторні перегляди. Наприклад, під час розробки серіалу «Дивні дива» (Stranger Things) компанія враховувала популярність жанру наукової фантастики серед молодшої аудиторії, що була підтверджена аналізом історичних

переглядів та обговорень у соціальних мережах. Використовуються інструменти Apache Spark для обробки великих даних і AWS Redshift для зберігання даних.

Disney+ застосовує аналітичні інструменти Tableau та Snowflake для сегментації аудиторії та визначення потенційно успішних тематик. Наприклад, аналіз даних показав, що аудиторія в Південній Америці має особливий інтерес до локальних героїв та міфології. Це підштовхнуло Disney до створення проєктів, натхнених місцевою культурою, як-от серіал «Союз захисників Латинської Америки». Залучення даних про взаємодії глядачів із контентом також дає змогу адаптувати маркетингові кампанії до кожного регіону.

Amazon Prime Video інтегрує технології машинного навчання через свою платформу AWS SageMaker. Цей інструмент використовується для аналізу відгуків глядачів у реальному часі. Наприклад, перед випуском серіалу «The Boys» було проведено аналіз коментарів на Reddit і Twitter за допомогою обробки природної мови (NLP). Це допомогло визначити, які аспекти супергеройського жанру є найбільш цікавими для цільової аудиторії, як-от темні та іронічні сюжети. На основі цих даних маркетингові матеріали були скориговані для акцентування на брутальності та антигеройських характеристиках персонажів.

Warner Bros. активно використовує платформу IBM Watson для аналізу глядацьких уподобань та формування трейлерів. За допомогою технологій Watson Visual Recognition і Watson Natural Language Understanding компанія аналізує візуальні та текстові відгуки про попередні трейлери. Наприклад, у випадку з фільмом «Джокер» Watson виявив, що аудиторія віддає перевагу глибоким психологічним темам і менш звертає увагу на екшен. Це вплинуло на монтаж трейлерів, які робили акцент на особистій драмі головного героя.

Соціальні медіа, зокрема **Twitter** та **Instagram**, стали важливими джерелами для аналізу трендів. За допомогою інструментів Google Cloud Natural Language API компанії аналізують тональність і зміст постів, щоб виявляти нові теми, які цікавлять глядачів. Наприклад, у випадку фільму «Барбі» студія Warner Bros. використовувала дані з соціальних мереж для розуміння реакції на перші тизери. Аналіз постів, проведений за допомогою NLP, показав, що глядачі з інтересом обговорюють поєднання ностальгії та сучасного гумору. Це дало змогу студії зосередитися на цих аспектах у подальших промоційних кампаніях.

YouTube використовує алгоритми кластеризації даних, щоб аналізувати вподобання користувачів і виявляти нові тренди. За допомогою Google BigQuery компанія відслідковує, які трейлери викликають найбільшу кількість переглядів, лайків і коментарів, щоб спрямувати рекламні бюджети на просування таких фільмів. Успіх трейлера до фільму «Аватар: Шлях води» став можливим завдяки даним, які свідчили про сильний інтерес до візуальних ефектів та екологічної тематики.

Загалом аналіз даних перетворюється на стратегічний інструмент у кіноіндустрії. Популярні сервіси використовують як класичні методи статистики, так і сучасні технології машинного навчання, штучного інтелекту та обробки великих даних. Це дає змогу не лише прогнозувати касові збори, але й формувати контент, який відповідає реальним очікуванням глядачів, виявляючи тренди ще на ранніх етапах їх формування. У майбутньому саме здатність ана-

лізувати та використовувати дані все більше визначатиме успіх кінокомпаній у глобальному масштабі.

Список використаних джерел

1. Kulkarni S. S., Vaishnavi M., Kusuma H. Predicting the Conceptual Appeal of Movies using Data Analytics. *International journal of engineering research & technology (IJERT)*. Vol. 09, iss. 02 (February 2020).
2. Movie Industry Economics: How Data Analytics Can Help Predict Movies' Financial Success / P. C. Murschetz, C. Bruneel, J.-L. Guy, D. Haughton, N. Lemercier, M.-D. McLaughlin, K. Mentzer, et al. *Nordic journal of media management*. 2020. Vol. 1, № 3. P. 339–359.
3. The Role of Data Analytics in Cinema. URL: <https://www.cloudthat.com/resources/blog/the-role-of-data-analytics-in-cinema>

УДК 004.032.26:621.397.331

*Лантєва М. А., здобувачка вищої освіти,
Хмелівський Ю. С., асистент кафедри
інформаційних технологій*

ВИЯВЛЕННЯ УПЕРЕДЖЕНЬ У ШІ ЗА ДОМОПОГОЮ СТАТИСТИЧНОГО НАВЧАННЯ

Донецький національний університет імені Василя Стуса, м. Вінниця

Штучний інтелект (ШІ) дедалі більше стає невід'ємною частиною процесів прийняття рішень у різних галузях – від повсякденних потреб, робочих питань до надважливих для держави сфер. Однак зі зростанням популярності систем штучного інтелекту пропорційно зростає занепокоєння щодо помилок і упереджень, які можуть виникнути під час роботи. Вони можуть виникати на різних етапах конвеєра машинного навчання, включно зі збором даних, попередньою обробкою, вибором функцій, навчанням моделі та розгортанням. Виявлення та пом'якшення цих упереджень має вирішальне значення для забезпечення справедливості, підзвітності та прозорості.

Знаходження упереджень передбачає поєднання статистичного аналізу, знань предметної області та критичного мислення. Використовуючи методи перевірки гіпотез, регресійного аналізу і метрики справедливості, статистичне навчання може допомогти оцінити, чи непропорційно впливають прогнози або класифікації ШІ-моделі на певні групи. У цій статті досліджується роль статистичного навчання у виявленні упередженості в ШІ, його методології та застосування.

Зміщення в моделях ШІ часто перетинаються з різними джерелами, що робить складним завдання їх детекції. Щоб розв'язати проблему, спочатку потрібно виявити її корінь.

По-перше, часто проблеми виникають з навчальними даними. Важливо ретельно переглянути їх на наявність будь-яких властивих упереджень, тому що якщо ці дані відображають наявні суспільні упередження, модель збереже їх і сприйме як щось регулярне. Наприклад, набори даних, що використовуються