АНАЛИЗ НА ПОТРЕБИТЕЛСКАТА НАГЛАСА И ПРОБЛЕМИ ПРИ ОНЛАЙН ПАЗАРУВАНЕ

Гл. ас. д-р Иван Маринов, i.marinov@uni-svishtov.bg
Катедра „Търговски бизнес“
СА „Д. А. Ценов“ - Свищов

Резюме: Изследването анализира 1254 независими отзива за Amazon, публикувани в Trustpilot между октомври 2025 г. и януари 2026 г. Приложени са лексикон-базиран анализ на нагласите (VADER), корелационен и регресионен анализ и тематично моделиране (LDA).

Мненията са поляризирани: 54,63% са с отрицателна нагласа, 41,95% с положителна и 3,43% с неутрална. Поляризацията обаче не е пълна, защото около половината мнения не изразяват крайно отношение. Връзката между композитната оценка на нагласата и рейтинга е умерена и положителна (r = 0,413), тоест текстът и звездите носят сходна, но не еднаква информация. Границите на речниковия подход личат най-ясно при най-ниската оценка: една трета от мненията с една звезда получават положителна композитна оценка, защото VADER отчита учтивите и ироничните формулировки като положителни. Контролът за държавата на автора оставя връзката почти непроменена, но самите рейтинги се различават систематично между държавите.

Тематичният анализ откроява пет области. Най-отрицателни са мненията за доставката до адреса и за контакта с клиентското обслужване, а най-нисък среден рейтинг имат тези за връщането на стоки. Трите области обхващат 57,18% от мненията и се отнасят до етапите след поръчката, а не до избора на продукт.

Ключови думи: потребителска нагласа, потребителско мнение, онлайн пазаруване, Amazon, Trustpilot, VADER, LDA

JEL: M31, L81, D12, C38, C55

DOI: https://doi.org/10.58861/tae.di.2026.3.03

ANALYSIS OF CONSUMER SENTIMENT AND PROBLEMS IN ONLINE SHOPPING

Head Assist. Prof. Ivan Marinov, PhD, i.marinov@uni-svishtov.bg
Department of Commerce
Tsenov Academy of Economics, Svishtov, Bulgaria

Abstract: This study analyses consumer sentiment and the main sources of dissatisfaction in online shopping by examining 1,254 independent customer reviews of Amazon published on Trustpilot between October 2025 and January 2026. The paper applies lexicon-based sentiment analysis (VADER), correlation and regression analysis, and topic modelling (LDA).

Opinions are polarised: 54.63% of reviews carry negative sentiment, 41.95% positive and 3.43% neutral. About half express a categorical attitude, while the rest contain mixed or restrained assessments. The relationship between the composite sentiment score and the rating is moderate and positive (r = 0.413), so the text and the stars convey similar but not identical information. One third of one-star reviews receive a positive composite score, which marks the limits of the lexicon-based approach. Controlling for the reviewer's country leaves this relationship almost unchanged, while the ratings themselves differ systematically across countries.

The topic analysis identifies five areas. Reviews about delivery to the address and about contacting customer service carry the most negative sentiment, while those about returns and refunds have the lowest average rating. Together these three areas cover 57.18% of the reviews and concern stages that follow the order. The findings direct attention to the last mile of delivery and to problem resolution after purchase.

Key words: consumer sentiment, consumer review, online shopping, Amazon, Trustpilot, VADER, LDA

JEL: M31, L81, D12, C38, C55

DOI: https://doi.org/10.58861/tae.di.2026.3.03

АНАЛИЗ НА ПОТРЕБИТЕЛСКАТА НАГЛАСА И ПРОБЛЕМИ ПРИ ОНЛАЙН ПАЗАРУВАНЕ

Гл. ас. д-р Иван Маринов, i.marinov@uni-svishtov.bg
Катедра „Търговски бизнес“
СА „Д. А. Ценов“ - Свищов

Въведение

В условия на дигитализация и засилена конкуренция в онлайн търговията потребителските оценки се утвърждават като ключов източник на информация както за бъдещите купувачи, така и за самите търговци. Онлайн отзивите отразяват реални потребителски преживявания и представляват форма на електронно „от уста на уста“, която оказва измеримо влияние върху продажбите, доверието и намерението за покупка (Chevalier & Mayzlin, 2006). Структурата и съдържанието на клиентските отзиви, особено в платформи като Amazon, оказват пряко влияние върху тяхната полезност и убедителност (Mudambi & Schuff, 2010).

Amazon е сред водещите глобални онлайн търговци с изключително голям продуктов асортимент, високо ниво на логистична интеграция и развита клиентска инфраструктура. Въпреки това мащабът на платформата и разнообразието от участници във веригата на стойността пораждат редица проблеми, които намират отражение в мненията на потребителите в независими платформи като Trustpilot. Сред тях са несъответствията в продуктовите описания и затрудненото следпродажбено обслужване.

Обект на настоящата научна статия са независимите потребителски отзиви за Amazon, публикувани в платформата Trustpilot през пиковия търговски сезон (октомври 2025 г. – януари 2026 г.). **Предмет на изследването** е емоционалната полярност на потребителските отзиви, нейната връзка с техните рейтингови оценки при контрол за междукултурни различия, както и тематичната структура на потребителския опит при пазаруване в онлайн платформата на Amazon.

Тезата гласи следното: *Систематичният емпиричен анализ на независими потребителски отзиви чрез комбиниране на лексикон-базиран анализ на нагласите (VADER), регресионно моделиране с контрол за държавата на автора и вероятностен генеративен модел (LDA) показва, че потребителските оценки за глобалните онлайн платформи са поляризирани, макар и не в крайна степен, и че между текстовата нагласа и числовия рейтинг съществува устойчива, но непълна връзка, като най-отрицателните нагласи се концентрират в етапите след поръчката: получаването на пратката и контактът с обслужването при възникнал проблем.*

Целта на настоящото изследване е чрез емпиричен анализ на независими потребителски отзиви за Amazon, да се измери и интерпретира потребителската нагласа, да се оцени връзката ѝ с рейтинговите оценки при контрол за междукултурни различия и да се идентифицират основните проблемни области в потребителския опит.

За постигането на тази цел са поставени следните конкретни задачи:

1. Да се извлекат и подготвят за анализ независими потребителски отзиви за Amazon от платформата Trustpilot за периода октомври 2025 г. – януари 2026 г.
2. Да се измери емоционалната полярност на отзивите чрез лексикон-базирания метод VADER и да се установи разпределението им по категории на потребителската нагласа.
3. Да се оцени силата и значимостта на връзката между полярността на нагласата в текста и рейтинга чрез корелационен анализ.
4. Да се изгради и интерпретира множествен регресионен модел, отчитащ едновременно връзката на рейтинга с нагласата и с държавата на автора на отзива.
5. Да се идентифицират и характеризират основните тематични области в потребителския опит чрез вероятностно тематично моделиране (LDA).
6. Да се направят изводи и препоръки за оптимизация на ключовите оперативни направления, определящи потребителската удовлетвореност.
7. Да се изведат ограниченията на използваните данни и методи и насоките за бъдещи изследвания.

1. Литературен преглед

Анализът на потребителската нагласа трансформира неструктурираната обратна връзка от купувачите в структурирана информация, която подпомага продуктовите, маркетинговите и управленските решения (Liu, 2015; Aguilar-Moreno, Palos-Sanchez, & Pozo-Barajas, 2024). Лексикон-базирани подходи като VADER (Valence Aware Dictionary and sEntiment Reasoner) са широко използвани за анализ на кратки, неформални онлайн текстове, тъй като отчитат жаргона, съкращенията и усилващите думи, характерни за тях, без да изискват обучителни данни (Hutto & Gilbert, 2014; Qi & Shabrina, 2023).

Тематичното моделиране, и в частност Latent Dirichlet Allocation (LDA), е широко използвано за извличане на повтарящи се проблемни области и модели на удовлетвореност (Ren & Hong, 2017; Sutherland, Sim, Lee, Byun, & Kiatkawsin, 2020; Barrera-Barrera, 2023; Mor, Orsenigo, Soto Gomez, & Vercellis, 2026).

Редица изследвания подчертават систематични проблеми с валидността на онлайн отзивите, породени от самоселективния характер на публикуването: недоволните и изключително доволните потребители са по-склонни да оставят отзиви, което води до поляризация и ограничава репрезентативността (Hu, Pavlou, & Zhang, 2017; Askalidis & Malthouse, 2016). Допълнително предизвикателство произтича от междукултурните различия в начина, по който потребителите формулират и възприемат отзиви, като четимостта, стилът на изразяване и националната принадлежност влияят върху възприеманата стойност на мненията (Fang, Ye, Kucukusta, & Law, 2016; Wang, Wang, Zhang, & Zhang, 2019; Brand, Kopplin, & Rausch, 2022), което е особено приложимо за глобална платформа като Amazon.

2. Методология на изследването.

2.1. Данни и предварителна обработка

За целите на изследването са извлечени 1256 независими клиентски отзиви за Amazon, публикувани в Trustpilot за периода октомври 2025 г. – януари 2026 г., който е пиков, високоинтензивен и стратегически критичен за онлайн търговците. След изчистването на събраната информация остават 1254 отзиви от 61 държави. Данните са извлечени чрез браузърното разширение Instant Data Scraper и включват: потребител (идентификатор, когато е наличен); държава на лицето, което е предоставило отзив; дата на публикуване в платформата Trustpilot; рейтинг (1–5 звезди); текст на отзива. Извадката е изчистена в три стъпки. Първо, езикът на всеки отзив е разпознат автоматично, като изцяло неанглийски отзиви не са открити. Един отзив съдържа английски текст, следван от италиански и френски превод, и преводите са премахнати. Второ, след нормализиране на регистъра и интервалите са открити две двойки отзиви с дословно съвпадащ текст. От всяка двойка е премахнат повторният отзив, а първоначалният е запазен. Трето, проверено е наличието на държава за всеки отзив и такива без държава няма.

Преди прилагането на аналитичните техники е извършена стандартна NLP (Natural Language Processing) обработка, включваща премахване на излишни интервали и базови артефакти, нормализация на регистъра и лематизация чрез spaCy за намаляване на лексикалната вариативност. Лематизираният текст се използва само за тематичното моделиране, където намаляването на лексикалната вариативност подобрява разграничаването на темите. Измерването на нагласата се извършва върху оригиналния текст на отзива. Речникът на VADER отчита пунктуацията, изписването с главни букви, усилващите наречия и отрицанията, а лематизацията и нормализацията на регистъра премахват точно тези сигнали и биха изместили композитната оценка на нагласата. Изчисленията са извършени в среда Python със следните библиотеки: vaderSentiment за оценка на нагласата, spaCy за лематизация и Gensim за тематичното моделиране и анализ на текстови корпуси.

2.2. Аналитичен подход

За количественото измерване на емоционалната потребителска нагласа е приложен методът VADER. Основната използвана метрика е композитната оценка (compound score) на VADER, която е нормализирана претеглена стойност в интервала от −1 (най-крайно отрицателна оценка) до +1 (най-крайно положителна оценка). Категоризацията има стандартни прагове: положителна нагласа (композитна оценка на нагласата ≥ 0,05) отрицателна нагласа (композитна оценка на нагласата ≤ −0,05) и неутрална нагласа (−0,05 < композитна оценка на нагласата < 0,05). Извършен е описателен анализ на разпределението на отзивите по категории на потребителската нагласа и на оценката на нагласите чрез хистограма и Kernel Density Estimation (KDE) за оценка на формата (вкл. поляризация/бимодалност), както и на средните стойности на композитна оценка на нагласата по рейтинг (1–5 звезди).

За оценка на линейната зависимост между рейтинга и отзивите на клиентите на Amazon е приложен коефициентът на Пирсън, със значимост оценена чрез стандартен тест за p-стойност.

За едновременна оценка на връзката на рейтинга с нагласата и с държавата на автора е приложена множествена линейна регресия. Зависимата променлива е рейтингът, а независимите са композитната оценка на нагласата и държавата, кодирана чрез фиктивни променливи. Разпределението на отзивите по държави е силно неравномерно: 26 държави са представени с едно мнение, а 48 с по-малко от пет. Затова като отделни категории са запазени само държавите с поне 10 отзива, а останалите 55 са обединени в категория „Други държави“. Референтна категория е Великобритания като най-голяма група. Качеството на моделите е оценено чрез R^2, Adjusted R^2, остатъчната стандартна грешка и F-статистиката, а приносът на държавата чрез F-тест за разликата в R^2. Използвани са и информационните критерии на Акайке (AIC) и и на Шварц (BIC). Като проверка за устойчивост са оценени модел с отделна фиктивна променлива за всяка от 61-те държави и ординална логистична регресия, която отчита порядковия характер на рейтинга.

За установяване на тематичната структура на отзивите е приложено вероятностно тематично моделиране чрез Latent Dirichlet Allocation (LDA), реализирано с библиотеката Gensim. Моделът е изграден върху лематизирания текст след премахване на стоп думите. От речника отпадат термините, които се срещат в по-малко от 5 мнения или в повече от 20% от тях. Първите са твърде редки, за да оформят тема, а вторите (например клиент, поръчка, обслужване) присъстват в почти всеки отзив и не различават темите помежду им. Параметрите α (разпределение на темите в отделния отзив) и η (разпределение на думите в отделната тема) се оценяват автоматично от данните.

Броят на темите (k) е избран чрез показателя за кохерентност C_v, изчислен за модели с k от 3 до 10. За всяка стойност на k са оценени по десет модела с различни начални стойности, тъй като резултатът от LDA зависи от случайната инициализация. Устойчивостта на решението е проверена чрез съпоставяне на най-характерните термини между отделните стартове.

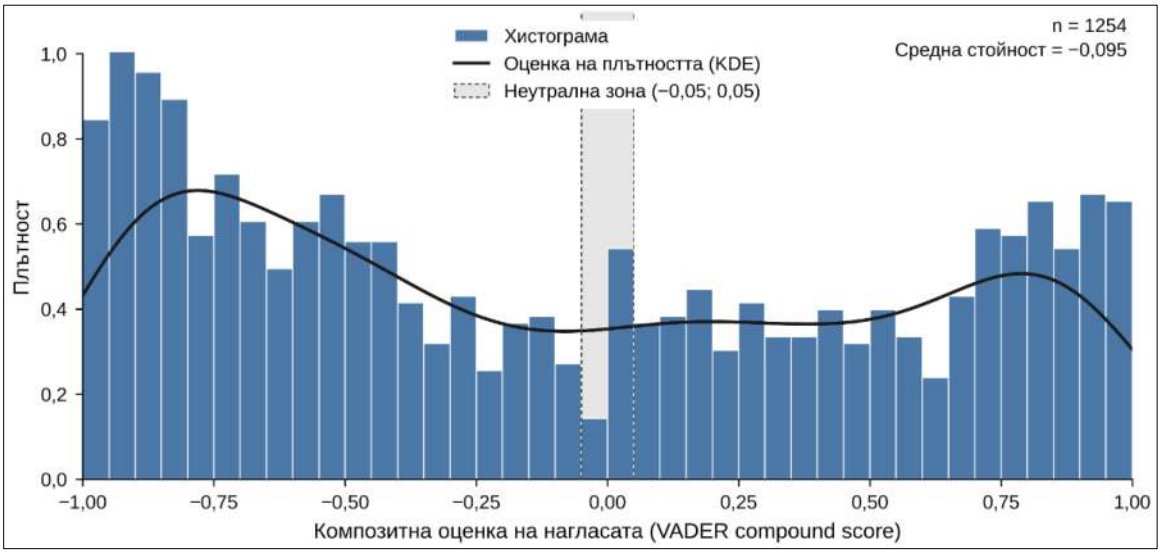
Всеки отзив е отнесен към темата с най-висока вероятност за него, наричана доминираща тема. За всяка тема са изчислени броят и делът на отнесените отзиви, средната композитна оценка на нагласата по формула (4) и средният рейтинг. Наименованията на темите представляват авторска интерпретация на най-характерните термини и не са част от изхода на модела.

Всички приложени методи са корелационни. Регресионните коефициенти показват статистическа връзка между променливите при контрол за останалите предиктори в модела и не се тълкуват като причинно въздействие.

3. Емпирични резултати

След изчистването на данните в анализа остават 1254 отзива. Преобладават тези с отрицателна нагласа (54,63%, 685 отзива), следвани от тези с положителна нагласа (41,95%, 526 отзива). Неутралните отзиви са 3,43% (43 отзива). Превесът на отрицателните отзиви съответства на ефекта на самоселекция при публикуването на отзиви: недоволните клиенти по-често си дават труда да опишат преживяването си (Hu, Pavlou, & Zhang, 2017). Същото личи и от рейтингите, тъй като 76,08% от отзивите са с 1 звезда.

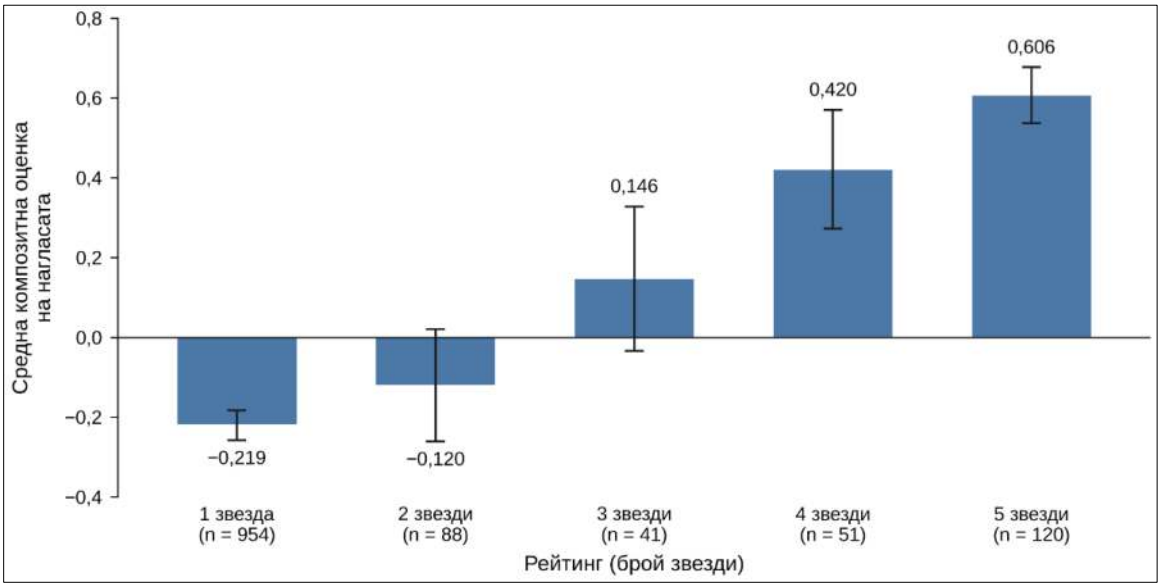
Разпределението на композитната оценка на нагласата е бимодално (Фигура 1). Кривата на плътността (KDE) има основен пик в отрицателната област, около −0,78, и по-нисък пик в положителната, около +0,79. Отзивите с изразена нагласа (под −0,6 или над +0,6) са 52,23% от извадката. Отрицателните сред тях са 30,46%, а положителните са 21,77%. Между двата пика плътността е почти равномерна. Ниският дял на неутралните отзиви затова не означава, че около нулата има празнина. Интервалът от −0,05 до +0,05 съдържа 3,43% от отзивите на клиенти, а съседните интервали със същата ширина съдържат 3,27% и 3,75%. Малкият брой неутрални отзиви идва от тесния праг на VADER оценката за неутрална категория.



Фигура 1. Разпределение на композитната оценка на нагласата в потребителските мнения за Amazon (n = 1254)
Източник: Trustpilot и изчисления на автора, https://www.trustpilot.com/review/www.amazon.com

Поляризацията в нагласите е умерена. Около половината потребители пишат отзиви с изразено отрицателно или положително отношение, но значителна част от тях съдържат смесени или по-сдържани оценки. Тематичният анализ в следващата част показва кои конкретни аспекти от опита с платформата на Amazon пораждат крайните отрицателни нагласи.

Фигура 2 представя средната композитна оценка на нагласата за всяка стойност на рейтинга. Средната оценка расте последователно с увеличаване на броя на звездите: от −0,219 при 1 звезда до +0,606 при 5 звезди. Нарастването обаче не е равномерно. Между 1 и 2 звезди разликата е 0,10, между 2 и 3 звезди и между 3 и 4 звезди е около 0,27, а между 4 и 5 звезди е 0,19. Доверителните интервали показват, че средните оценки при 1 и 2 звезди не се различават статистически значимо (p = 0,18). Двата най-ниски рейтинга следователно съответстват на текстове, които са близки по тон. Отзивите с 3 звезди са със средна оценка +0,146, но интервалът ѝ включва нулата (p = 0,11), така че нагласата в тях се определя най-точно като близка до неутралната. При 4 и 5 звезди нагласата е ясно положителна, а при 5 звезди 90% от отзивите на клиентите са с положителна композитна оценка.



Фигура 2. Средна композитна оценка на нагласата по рейтинг на отзивите на клиенти на Amazon (n = 1254; вертикалните линии показват 95% доверителен интервал).

Източник: Trustpilot и изчисления на автора, https://www.trustpilot.com/review/www.amazon.com

Съответствието между текст и рейтинг е най-слабо при най-ниската оценка. От мненията с 1 звезда 33,8% получават положителна композитна оценка, поради което медианата (−0,394) е значително по-ниска от средната стойност. Част от тези мнения описват очакванията на потребителя или са формулирани учтиво и иронично. Речниковият подход на VADER отчита положителните думи в тях, но не и контекста, в който са употребени. Това е и една от причините връзката между рейтинга и нагласата да е умерена, а не силна. Рейтингът и текстът на отзива носят сходна, но не еднаква информация, затова удовлетвореността на потребителите се оценява по-точно, когато двата показателя се разглеждат заедно.

Корелационният анализ установява умерена положителна линейна зависимост между рейтинга и композитната оценка на нагласата ($r = 0,413$; 95% ДИ: 0,366–0,457; $t = 16,03$; $p < 0,001$; $n = 1254$). Отзивите с повече звезди са систематично по-положителни по тон, но двете измерения не съвпадат. Рейтингът обяснява около 17%

от вариацията в композитната оценка ($r^2 = 0{,}170$). Звездите следователно не предават цялата сила на отношението, изразено в текста. Коефициентът е изчислен по формула (1).

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (1)$$

където:

- r_{XY} – коефициентът на линейна корелация на Пирсън между рейтинга (X) и композитната оценка на нагласата (Y);
- X_i – рейтингът на отзив i (от 1 до 5 звезди);
- Y_i – композитната оценка на нагласата (VADER compound score) на отзив i (в интервала от −1 до +1);
- $\overline{X}$ – средният рейтинг за извадката;
- $\overline{Y}$ – средната композитна оценка на нагласата за извадката;
- n – броят на отзивите в извадката.

Корелацията описва връзката между две променливи, но не показва дали тя се запазва при отчитане на други фактори. Такъв фактор е държавата на автора: културните и езиковите особености могат да влияят както на начина, по който потребителите изразяват отношението си, така и на стандартите, по които поставят оценка. Затова в следващата стъпка е приложен регресионен анализ, в който композитната оценка на нагласата и държавата влизат едновременно като предиктори. Той позволява да се провери дали връзката между нагласата и рейтинга остава значима при контрол за държавата и да се оцени колко добавя държавата към обяснението на рейтинга.

За постигането на тази цел са изградени и сравнени два вложени регресионни модела. Модел 1 представлява проста линейна регресия, която оценява връзката между рейтинга и композитната оценка на нагласата без допълнителни контролни променливи (2). Модел 2 разширява Модел 1 чрез добавяне на категориалната променлива „държава“ под формата на фиктивни променливи (3). Сравнението между двата модела позволява да се изолира собственият принос на нагласата на потребителите и да се оцени количествено допълнителното влияние на културно-специфичните фактори.

Модел 1. Простият регресионен модел има следната спецификация:

$$\text{Rating}_i = \beta_0 + \beta_1 \cdot \text{Compound}_i + \varepsilon_i \quad (2)$$

където:

- Rating_i – рейтингът на отзив i в скалата от 1 до 5 звезди (зависима променлива);
- Compound_i – композитната оценка на нагласата на отзив i в интервала (−1; +1);
- β_0 – константата на модела;

- β_1 – регресионният коефициент, измерващ средната промяна в рейтинга при увеличение на композитната оценка на нагласата с една единица;
- ε_i – случайната грешка.

Модел 2. Модел 2 добавя към Модел 1 държавата, от която е предоставен отзивът като категориален предиктор, кодиран чрез фиктивни променливи. Извадката включва отзиви от 61 държави, но разпределението им е силно неравномерно. Великобритания (647 отзива) и САЩ (393 отзива) заедно представляват 83% от извадката. 26 държави са представени с един-единствен отзив, а 48 – с по-малко от пет. Отделна фиктивна променлива за всяка държава би означавала 60 предиктора, чиито коефициенти за повечето държави се оценяват по едно или две наблюдения. Такива коефициенти нямат стабилност и не подлежат на съдържателно тълкуване, а моделът механично напасва рейтинга на единичните отзиви. Затова като отделни категории са запазени само държавите с поне 10 мнения (Великобритания, САЩ, Канада, Индия, Нидерландия и Австралия). Останалите 55 държави са обединени в категория „Други държави“ (132 мнения). За референтна категория е избрана Великобритания като най-голямата група, спрямо която оценките на останалите коефициенти са най-прецизни. Спецификацията на модела е:

$$\mathrm{Rating_i} = \beta_0 + \beta_1 \cdot \mathrm{Compound_i} + \sum_{k=1}^{6} \gamma_k \cdot \mathrm{D_{k,i}} + \varepsilon_i \qquad (3)$$

където:

- $\mathrm{Rating_i}$ – рейтингът на отзив i (от 1 до 5 звезди);
- $\mathrm{Compound_i}$ – композитната оценка на нагласата на отзив i (в интервала от −1 до +1);
- β_0 – константата на модела, т.е. очакваният рейтинг за отзив от Великобритания с неутрална нагласа ($Compound_i = 0$);
- β_1 – коефициентът на композитната оценка на нагласата при контрол за държавата;
- $\mathrm{D_{k,i}}$ – фиктивна променлива, която приема стойност 1, ако авторът на мнение i е от група k (САЩ, Индия, Канада, Нидерландия, Австралия или „Други държави“), и 0 в противен случай;
- γ_k – разликата в средния рейтинг между група k и Великобритания при еднаква композитна оценка на нагласата;
- ε_i – случайна грешка.

Качеството на моделите е оценено чрез R^2, Adjusted R^2, остатъчната стандартна грешка и F-статистиката. Приносът на държавата е проверен чрез F-тест за разликата в R^2 между Модел 1 и Модел 2. За сравнение между модели с различен брой предиктори са използвани и информационните критерии AIC и BIC, които наказват добавянето на параметри. Като проверка за устойчивост е оценен и модел с отделна фиктивна променлива за всяка от 61-те държави.

Таблица 1.
Резултати от регресионните модели (зависима променлива: рейтинг, 1–5 звезди; n = 1254)

Предиктор	Модел 1	Модел 2	Проверка: всички 61 държави
А. Регресионни коефициенти			
Константа	1,720*** (0,034)	1,552*** (0,046)	–
Композитна оценка на нагласата (β_1)	0,832*** (0,052)	0,807*** (0,051)	0,762*** (0,052)
САЩ (n = 393)	–	0,256*** (0,075)	–
Индия (n = 25)	–	0,774** (0,239)	–
Канада (n = 32)	–	0,398 (0,212)	–
Нидерландия (n = 15)	–	0,467 (0,305)	–
Австралия (n = 10)	–	−0,301 (0,373)	–
Други държави (n = 132)	–	0,533*** (0,112)	–
Б. Показатели за качеството на моделите			
Брой фиктивни променливи за държава	0	6	60
R^2	0,1702	0,1952	0,2521
Adjusted R^2	0,1695	0,1907	0,2138
Остатъчна стандартна грешка	1,1845	1,1693	1,1525
F-статистика	256,81*** (1; 1252)	43,18*** (7; 1246)	6,59*** (61; 1192)
ΔR^2 спрямо Модел 1	–	0,0250	0,0819
F-тест за ΔR^2	–	6,45*** (6; 1246)	2,18*** (60; 1192)
AIC	3985,3	3958,9	3975,0
BIC	3995,6	4000,0	4293,3

*Забележки: В скоби са стандартните грешки (при F-статистиката – степените на свобода). Референтна категория за държава е Великобритания (n = 647). В категорията „Други държави“ са обединени държавите с по-малко от 10 мнения. ** p < 0,01; *** p < 0,001.*
Източник: Trustpilot и изчисления на автора, https://www.trustpilot.com/review/www.amazon.com

Композитната оценка на нагласата е значимо свързана с рейтинга и в двата модела (Таблица 1). В Модел 1 коефициентът е $\beta_1 = 0{,}832$ ($t = 16{,}03$; $p < 0{,}001$), а моделът обяснява 17,0% от вариацията в рейтинга. Два отзива, чиито композитни оценки се различават с 0,5 (например неутрален и умерено положителен отзив), се различават средно с 0,42 звезди в рейтинга.

При включване на държавата коефициентът се понижава слабо, до $\beta_1 = 0{,}807$ ($t = 15{,}68$; $p < 0{,}001$). Композитната оценка на нагласата почти не варира между групите държави: групите обясняват под 1% от вариацията ѝ. Затова контролът за държава не променя съществено връзката между нагласата и рейтинга. Модел 2 обяснява 19,5% от вариацията в рейтинга (Adjusted $R^2 = 0{,}191$). Нарастването спрямо Модел 1 е 2,5 процентни пункта и е статистически значимо ($F(6; 1246) = 6{,}45$; $p < 0{,}001$). Модел 2 има и най-ниската стойност на AIC сред трите модела. По критерия BIC, който наказва допълнителните параметри по-строго, Модел 1 има малко по-ниска стойност (3995,6 срещу 4000,0), което съответства на ограничения принос на държавата към обяснението на рейтинга.

Различията между държавите засягат самия рейтинг. При еднаква композитна оценка на нагласата потребителите от САЩ поставят средно с 0,26 звезди повече от тези от Великобритания ($p < 0{,}001$). Разликата е 0,77 звезди за Индия ($p = 0{,}001$) и 0,53 звезди за групата на останалите държави ($p < 0{,}001$). Разликите за Канада, Нидерландия и Австралия не са статистически значими. Потребителите от Великобритания се открояват с най-строги оценки: при еднакъв тон на текста те поставят по-малко звезди от почти всички останали групи. Регресията установява систематични разлики, но не и причината за тях. Те могат да се дължат на различни стандарти за оценяване, но и на различия в обслужването, доставките и асортимента между отделните пазари на Amazon.

Моделът с отделна фиктивна променлива за всяка от 61-те държави дава по-висок R^2 (0,252), но предимството му е привидно. 60 предиктора при 1254 наблюдения намаляват Adjusted R^2 с почти 4 процентни пункта спрямо R^2. Критерият BIC нараства от 3996 при Модел 1 до 4293, което показва, че допълнителните параметри не се оправдават от подобрението в съответствието с данните. Значимите коефициенти в този модел се отнасят предимно за държави с по един до пет отзива, а кои държави изглеждат значими зависи изцяло от избора на референтна категория. Коефициентът на композитната оценка на нагласата ($\beta_1 = 0{,}762$) остава обаче значим и близък по големина. Основният извод следователно не зависи от начина на кодиране на държавата. Коефициентът на нагласата остава значим и положителен и при ординална логистична регресия, която отчита, че рейтингът е порядкова скала ($p < 0{,}001$).

На следващия етап от анализа е приложено вероятностно тематично моделиране чрез Latent Dirichlet Allocation (LDA), за да се установи за какво пишат потребителите. Броят на темите е избран с помощта на показателя за кохерентност C_v, изчислен за модели с 3 до 10 теми. Стойностите са близки за k от 3 до 8 (между 0,368 и 0,383) и разликите между тях не надвишават вариацията между отделните стартове на модела. Кохерентността следователно не посочва един оптимален брой теми. Избран е моделът с пет теми като най-малкият, при който всяка тема има ясно разграничимо съдържание. От десет модела с различни начални стойности е използван този с най-висока кохерентност ($C_v = 0{,}405$).

За всяка тема е изчислена средната композитна оценка на нагласата по формула (4):

$$\overline{\text{Compound}_t} = \frac{1}{N_t}\sum_{i=1}^{N_t} \text{Compound}_i \quad (4)$$

където:

- $\overline{\text{Compound}_t}$ – средната композитна оценка на нагласата за тема t;
- N_t – броят на отзивите, отнесени към тема t;
- Compound_i – композитната оценка на нагласата на отзив i от тема t.

Резултатите са представени в Таблица 2. Характерните термини са резултат от модела. Наименованията на темите са авторска интерпретация на тези термини.

Таблица 2.

Тематична структура на потребителските отзиви за Amazon (LDA модел с пет теми; n = 1254)

Тема	Най-характерни термини	Брой мнения	Дял (%)	Средна оценка на нагласата	Среден рейтинг (звезди)
Тема 1 Продукти, продавачи и цялостен опит	отзив, продукт, компания, добър, лош, плащане, купуване, хора, опит, пари	298	23,76	+0,070	2,34
Тема 2 Доставка до адреса и работа на куриерите	куриер, пратка, оставям, колет, врата, година, адрес, чакане, днес, указание	274	21,85	−0,234	1,27
Тема 3 Възстановяване на суми и връщане на стоки	възстановяване, карта, връщане, пари, профил, получаване, отказ, чакане, изпращане, седмица	268	21,37	−0,115	1,21
Тема 4 Prime абонамент, плащания и цени	Prime, продукт, плащане, отказ, компания, добър, година, връщане, проблем, покупка	236	18,82	−0,022	1,84
Тема 5 Контакт с клиентското обслужване	чат, възстановяване, питане, свързване, помощ, казус, проблем, телефон, профил, оператор	175	13,96	−0,230	1,41

Забележки: Характерните термини са резултат от модела и са подредени по тегло в темата; наименованията на темите са интерпретация на автора въз основа на тези термини. Всяко мнение е отнесено към темата с най-висока вероятност за него (доминираща тема). Три мнения, в които след предварителната обработка не остават

термини, не са отнесени към тема. Средната композитна оценка на нагласата за всяка тема е изчислена по формула (4). Кохерентност на модела C_v = 0,405.

Източник: Trustpilot и изчисления на автора, https://www.trustpilot.com/review/www.amazon.com

Трите теми с най-голям дял имат сходна тежест, между 21% и 24% от мненията. Темата за доставката до адреса обединява отзиви за куриери, оставени или неоткрити пратки и неспазени инструкции за доставка. Тя е сред най-негативните (средна оценка −0,234; среден рейтинг 1,27 звезди). Със същата отрицателна нагласа е и темата за контакта с клиентското обслужване (−0,230), в която преобладават думи като чат, телефон и служител. Темата за възстановяването на суми и връщането на стоки е с по-умерена нагласа (−0,115), но с най-нисък среден рейтинг (1,21 звезди). Текстовете в нея са по-скоро описателни, а оценката им е категорично ниска.

Темата за продуктите и продавачите е най-голямата (23,76%) и единствената с положителна средна оценка (+0,070). В нея се срещат както оплаквания, така и похвали (добър, лош, страхотен), и тя обхваща и по-голямата част от положителните мнения в извадката. Темата за Prime абонамента, плащанията и цените е с почти неутрална средна оценка (−0,022).

Следователно най-силно отрицателни са отзивите за двата етапа, в които потребителят зависи от действията на Amazon след покупката: получаването на пратката и контакта с обслужването при проблем.

Резултатите съвпадат с литературата в общата картина, но я уточняват в няколко посоки. Превесът на отрицателните отзиви и концентрацията на оценки с една звезда отговарят на самоселективния характер на публикуването (Askalidis & Malthouse, 2016; Hu, Pavlou, & Zhang, 2017), като поляризацията тук е умерена, а ниският дял на неутралните отзиви се дължи на тесния праг на VADER, а не на липса на сдържани клиентски мнения. Умерената корелация между тон и рейтинг потвърждава, че числовата оценка не замества текста на отзива (Mudambi & Schuff, 2010). Разминаването е най-голямо при отзивите с една звезда, където речниковият подход отчита учтивите и ироничните формулировки извън контекста им, което е в съответствие с Qi и Shabrina (2023): VADER се справя добре с неформалния език, но отстъпва по точност на моделите с машинно обучение. Изследванията на междукултурните различия свързват културата главно с начина, по който отзивите се формулират и възприемат (Fang, Ye, Kucukusta, & Law, 2016; Wang, Wang, Zhang, & Zhang, 2019; Brand, Kopplin, & Rausch, 2022). Тук картината е друга. Държавата на автора, който е написал отзива, почти не се свързва с тона на текста, а с начина, по който се използва скалата от една до пет звезди. Значението на връщането на стоки като източник на проблеми в онлайн търговията се потвърждава и от Mor, Orsenigo, Soto Gomez и Vercellis (2026), които чрез тематично моделиране на отзиви извеждат основните причини за връщане. Приносът на настоящото изследване е в това, че най-отрицателните нагласи в отзивите за Amazon се отнасят до етапите след поръчката, а не до избора на продукт.

4. Изводи и препоръки

4.1. Изводи

На база на получените резултати могат да се направят следните изводи:

Първо, отзивите в независимата платформа Trustpilot са поляризирани, но степента на поляризация е умерена. Управленските решения, основани единствено на средния рейтинг, могат да пренебрегнат както смесения характер на клиентските отзиви, така и съдържанието на текстовите отзиви с крайни оценки.

Второ, рейтингът и текстовото съдържание на отзивите се различават в достатъчна степен, за да обосноват необходимостта от тяхното съвместно проследяване. При отзивите с една звезда, които представляват три четвърти от извадката, VADER оценката се разминава с рейтинга в една трета от случаите. Мониторингът, основан само на автоматична оценка на нагласата, крие риск от подценяване на част от негативните потребителски реакции.

Трето, рейтингите, които купувачите дават, не са пряко сравними между отделните държави. При еднакъв тон на текстовите мнения разликата между Великобритания и Индия достига 0,77 звезди. Поради това сравнението на рейтингите между отделните пазари изисква отчитане на съществуващите междудържавни различия.

Четвърто, 57,18% от отзивите се отнасят до три области, които възникват след извършването на поръчката: доставката до адреса, връщането на стоки и контакта с обслужването на клиентите. Двете теми с най-негативна оценка на нагласата са именно доставката и обслужването на клиенти. Резултатите от извадката показват, че основните проблеми не са свързани толкова с избора на продукта, колкото с процесите, които следват след извършването на поръчката.

4.2. Препоръки

Отзивите относно доставката от Amazon съдържат проблеми, които се повтарят, сред които оставяне на пратката пред вратата, неспазване на предварителни указания и доставяне на неправилен адрес. Контролът върху изпълнението на указанията, съчетан с документиране на предаването чрез информация за местоположението и снимков материал, е насочен пряко към ограничаването на такива случаи.

Процесът по връщане на продукти и възстановяване на суми се нуждае от ясно определени срокове и възможност за проследяване на отделните етапи. Тази тема получава най-ниския среден рейтинг в извадката, възлизащ на 1,21 звезди, въпреки че клиентите описват свързаните с нея проблеми сравнително сдържано.

Достъпът до оператор в случай на нерешен проблем следва да бъде по-пряк. Често срещаните термини, свързани с темата, като „чат“, „телефон“, „оператор“ и „помощ“, показват необходимост от достигане до отговорен служител, който да поеме конкретния случай. Очакванията са системата да прехвърля разговора към оператор бързо след два неуспешни опита за автоматично решаване на проблем.

Мониторингът на качеството следва да съчетава анализа на рейтингите с текстов анализ на клиентските отзиви и да се извършва по отделни пазари. Автоматизираната оценка на нагласата е подходяща за идентифициране на общи тенденции, но не може да замести анализа на отделните случаи. Поради това мненията с нисък рейтинг, но с положителна текстова нагласа, следва да бъдат подлагани на допълнителен преглед.

Продуктовият аспект на потребителския опит не се очертава като критичен фактор в наблюдаваната извадка. Този извод се отнася конкретно до изследвания период и до анализираните потребителски отзиви, публикувани в платформата Trustpilot.

5. Ограничения и насоки за бъдещи изследвания

Резултатите трябва да се тълкуват при няколко ограничения.

Извадката обхваща отзиви само в една независима платформа за един пиков сезон. Публикуването в Trustpilot е самоселективно, затова разпределението на рейтингите не представя удовлетвореността на всички клиенти. Периодът октомври 2025 г. – януари 2026 г. е с високо натоварване на доставките, което вероятно повишава дела на отзивите за доставката.

Измерването на нагласата разчита на речников подход. VADER не отчита ирония, сарказъм и условни конструкции, а анализът тук показва, че това се отразява най-силно при отзивите с една звезда.

Рейтингът е порядкова променлива, моделирана с линейна регресия. Проверката с ординална логистична регресия дава същата посока и значимост, но коефициентите на линейния модел следва да се четат като средни разлики, а не като точни величини.

Държавите с малък брой отзиви са обединени в една категория. Изводите за отделни държави извън Великобритания и САЩ се основават на десетки отзиви и са ориентировъчни.

Тематичното моделиране зависи от предварителната обработка и от броя теми. Кохерентността не откроява еднозначно оптимален брой, а при различни начални стойности съставът на темите се променя умерено. Наименованията на темите са авторска интерпретация на характерните термини (Laureate, Buntine, & Linger, 2023).

Анализът е корелационен и не установява причинно-следствени връзки. Бъдещи изследвания могат да съпоставят няколко платформи и по-дълъг период, да приложат контекстни модели за оценка на нагласата и да проверят доколко различията между държавите се дължат на стандарти за оценяване или на реални различия в обслужването.

Заключение

В изследването се прави анализ на 1254 независими отзива за Amazon, публикувани в онлайн платформата Trustpilot между октомври 2025 г. и януари 2026 г. Целта беше да се измери нагласата на потребителите, да се оцени връзката ѝ с рейтинга при контрол за държавата на автора на отзива и да се установи в кои етапи на онлайн покупката се съсредоточава недоволството.

Резултатите защитават тезата на автора. Анализът показа, че отзивите са умерено поляризирани: отрицателната нагласа преобладава, а около половината клиентски мнения за Amazon не изразяват крайно отношение. Връзката между нагласата в текста и рейтинга е умерена: двата показателя носят сходна, но не еднаква информация. Тя почти не се променя при контрол за държавата и при проверките за устойчивост и е най-слаба при отзивите с една звезда, една трета от които получават положителна композитна оценка заради формулировки, които са учтиви или иронични. Най-отрицателни са отзивите за доставката до адреса и за контакта с обслужването, а най-нисък среден рейтинг имат тези за връщането на стоки. Трите области са малко над половината от всички отзиви и се отнасят до етапите на обслужване след поръчката.

Приносът на изследването е преди всичко емпиричен. Изследването показва, че в отзивите за глобална търговска платформа през пиков сезон най-отрицателните

нагласи се отнасят до изпълнението на поръчката и до решаването на проблеми след нея, а не до избора на продукт. Резултатите уточняват и представата за междукултурните различия, които досегашните изследвания свързват главно с начина, по който мненията се формулират и възприемат. В извадката от отзиви държавата на автора почти не е свързана с тона на текста, а с употребата на скалата от една до пет звезди. За измерването следва, че нито рейтингът, нито автоматичната оценка на нагласата описват потребителската удовлетвореност достатъчно точно поотделно. Ниският дял на неутралните отзиви се дължи на тесния праг на VADER, а не на липса на сдържани мнения от клиентите на Amazon.

За онлайн търговците резултатите посочват конкретни мерки. Контролът върху изпълнението на указанията за доставка и ясните срокове и проследимост при връщане на стоки са насочени към двете области с най-нисък среден рейтинг, а бързото прехвърляне към оператор при нерешен проблем засяга контакта с обслужването. Рейтингите от различните пазари не са пряко сравними. Затова мониторингът на качеството трябва да съчетава звездите с текстов анализ и да се води отделно за всеки пазар.

Изводите се отнасят до една платформа и един сезон с високо натоварване на доставките, а приложените методи установяват връзки, а не причини. Сравнение между няколко платформи за по-дълъг период, с контекстни модели за оценка на нагласата, би показало доколко концентрацията на недоволството след поръчката е сезонна и дали различията между държавите отразяват стандарти за оценяване, или реални разлики в обслужването.

Използвани източници

1. Aguilar-Moreno, J. A., Palos-Sanchez, P. R., & Pozo-Barajas, R. (2024). Sentiment analysis to support business decision-making. A bibliometric study. *AIMS Mathematics, 9*(2), 4337-4375. doi:10.3934/math.2024215

2. Askalidis, G., & Malthouse, E. C. (2016). *Understanding and Overcoming Biases in Customer Reviews.* arXiv. Retrieved from https://arxiv.org/abs/1604.00417

3. Barrera-Barrera, R. (2023). Identifying the attributes of consumer experience in Michelin-starred restaurants: a text-mining analysis of online customer reviews. *British Food Journal, 125*(13), 579-598. doi:10.1108/BFJ-05-2023-0408

4. Brand, B. M., Kopplin, C. S., & Rausch, T. M. (2022). Cultural differences in processing online customer reviews: holistic versus analytic thinkers. *Electronic Markets, 32*(3), 1039–1060. doi:10.1007/s12525-022-00543-1

5. Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research, 43*(3), 345-354. doi:10.1509/jmkr.43.3.345

6. Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online reviews: Influence of readability and reviewer characteristics. *Tourism Management, 52*, 498–506. doi:10.1016/j.tourman.2015.07.018

7. Hu, N., Pavlou, P. A., & Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly, 41*(2), 449 - 471. doi:10.25300/MISQ/2017/41.2.06

8. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *8*(1), 216-225. doi:10.1609/icwsm.v8i1.14550

9. Laureate, C., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review, 56*(12), 14223-14255. doi:10.1007/s10462-023-10471-x

10. Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* Cambridge University Press. doi:10.1017/CBO9781139084789

11. Mor, A., Orsenigo, C., Soto Gomez, M., & Vercellis, C. (2026). Shaping the causes of product returns: topic modeling on online customer reviews. *Electronic Commerce Research, 26*(1), 747-781. doi:10.1007/s10660-024-09901-x

12. Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com. *MIS Quarterly, 34*(1), 185-200. doi:10.2307/20721420

13. Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining, 13*(1), 31. doi:10.1007/s13278-023-01030-x

14. Ren, G., & Hong, T. (2017). Investigating Online Destination Images Using a Topic-Based Sentiment Analysis Approach. *Sustainability, 9*(10), 1765. doi:10.3390/su9101765

15. Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation. *Sustainability, 12*(5), 1821. doi:10.3390/su12051821

16. Wang, Y., Wang, Z., Zhang, D., & Zhang, R. (2019). Discovering Cultural Differences in Online Consumer Product Reviews. *Journal of Electronic Commerce Research, 20*, 169. Retrieved from https://www.jecr.org/node/587