



**CHAKANA SAVDO TARMOQLARIDA ISTE'MOLCHILAR XULQ-  
ATVORINI APRIORI ALGORITMI ASOSIDA MODELLASHTIRISH VA  
SAVAT TAHLILI (MARKET BASKET ANALYSIS)**

**МОДЕЛИРОВАНИЕ ПОВЕДЕНИЯ ПОТРЕБИТЕЛЕЙ В РОЗНИЧНЫХ  
СЕТЯХ НА ОСНОВЕ ПРЕДВАРИТЕЛЬНОГО АЛГОРИТМА И  
АНАЛИЗА ПОТРЕБИТЕЛЬСКОЙ КОРЗИНЫ**

**MODELING CONSUMER BEHAVIOR IN RETAIL NETWORKS BASED  
ON A PRELIMINARY ALGORITHM AND MARKET BASKET  
ANALYSIS**

*Ismoilov A<sup>1</sup>, Abdullayeva M.M<sup>2</sup>, Mamadaliyeva E.L<sup>3</sup>*

*<sup>1</sup>FarDU dotsenti, [ismoilovaxror@gmail.com](mailto:ismoilovaxror@gmail.com)*

*<sup>2</sup>FarDU talabasi, [abdullayevamashhuara@gmail.com](mailto:abdullayevamashhuara@gmail.com)*

*<sup>3</sup>FarDU talabasi, [mamadaliyevaerkinoy@gmail.com](mailto:mamadaliyevaerkinoy@gmail.com)*

***Annotatsiya:** Ushbu ilmiy maqolada zamonaviy chakana savdo tarmoqlarida savdo samaradorligini oshirish, kross-vending va merchendayzing strategiyalarini optimallashtirish hamda mijozlarning yashirin xarid qilish tendensiyalarini aniqlash maqsadida “Savat tahlili” (Market Basket Analysis) uslubiyoti har tomonlama tadbiq etilgan va tahlil qilingan. Ma'lumotlar to'plamini qayta ishlash, binar korrelyatsiyalarni hisoblash va yuqori ishonchlilikka ega bo'lgan assotsiatsiya qoidalarini ajratib olish uchun Python dasturlash tili, pandas ekotizimi hamda intellektual ma'lumotlar tahlilining (Data Mining) klassik asosi hisoblangan Apriori algoritmidan foydalanildi. Tadqiqot davomida real savdo tranzaksiyalari aks etgan Excel formatidagi ma'lumotlar to'plami raqamli suzgichlardan o'tkazilib, binar matritsa ko'rinishiga keltirildi va mahsulotlarning birgalikda xarid qilinish ehtimolligi qo'llab-quvvatlash, ishonchlilik hamda lift metrikalari orqali matematik*



va amaliy jihatdan baholandi. Olingan natijalar tijorat muassasalarida mahsulotlarni peshtaxtalarga joylashtirish, marketing aksiyalarini rejalashtirish va kross-vending strategiyalarini optimallashtirish uchun chuqur ilmiy-amaliy asos bo'lib xizmat qiladi.

**Kalit so'zlar:** Savat tahlili, Apriori algoritmi, Data Mining, Python, pandas, assotsiatsiya qoidalari, chakana savdo, kross-vending, tranzaksiyalar tahlili.

**Аннотация:** В данной научной статье методология «анализа потребительской корзины» всесторонне применяется и анализируется с целью повышения эффективности продаж в современных розничных сетях, оптимизации стратегий межпродавцового и мерчандайзингового взаимодействия, а также выявления скрытых покупательских тенденций клиентов. Для обработки набора данных, расчета бинарных корреляций и извлечения высоконадежных ассоциативных правил использовались язык программирования Python, экосистема pandas и алгоритм Apriori, считающийся классической основой интеллектуального анализа данных (Data Mining). В ходе исследования набор данных Excel, содержащий реальные данные о продажах, был подвергнут цифровым фильтрам и преобразован в бинарную матрицу, а вероятность совместных покупок товаров была математически и практически оценена с использованием метрик поддержки, достоверности и лифта. Полученные результаты служат глубокой научной и практической основой для оптимизации размещения товаров на полках, планирования маркетинговых кампаний и межпродавцовых стратегий в торговых предприятиях.

**Ключевые слова:** анализ корзины покупок, алгоритм Apriori, интеллектуальный анализ данных, Python, pandas, правила ассоциации, розничная торговля, кросс-продавцы, анализ транзакций.



**Ключевые слова:** анализ корзины покупок, алгоритм Apriori, интеллектуальный анализ данных, Python, pandas, правила ассоциации, розничная торговля, кросс-продавцы, анализ транзакций.

**Abstract:** In this scientific article, the methodology of 'Market Basket Analysis' is comprehensively applied and analyzed in order to increase sales efficiency in modern retail chains, optimize cross-vendor and merchandising strategies, and identify hidden purchasing tendencies of customers. The Python programming language, the pandas ecosystem, and the Apriori algorithm, which is considered the classic basis of intelligent data analysis (Data Mining), were used to process the data set, calculate binary correlations, and extract highly reliable association rules. During the study, the Excel data set containing real sales transactions was subjected to digital filters and transformed into a binary matrix, and the probability of joint purchases of products was mathematically and practically assessed using support, confidence, and lift metrics. The results obtained serve as a deep scientific and practical basis for optimizing product placement on shelves, planning marketing campaigns, and cross-vendor strategies in commercial establishments.

**Keywords:** Shopping cart analysis, Apriori algorithm, Data Mining, Python, pandas, association rules, retail, cross-vendor, transaction analysis.

## KIRISH

Zamonaviy global iqtisodiyot sharoitida chakana savdo tarmoqlari va elektron tijorat platformalarining jadal sur'atlar bilan rivojlanishi har kuni millionlab tranzaksiyalardan iborat bo'lgan ulkan hajmdagi ma'lumotlar oqimini yuzaga keltirmoqda. Ushbu "katta ma'lumotlar" tarkibida mijozlarning haqiqiy ehtiyojlari, yashirin qonuniyatlari, xarid qilish odatlari va afzal ko'rishlarini aks ettiruvchi o'ta qimmatli tijoriy sirlar va qonuniyatlar yashiringan. Ushbu qonuniyatlarni an'anaviy



statistik usullar yoki inson omili (menejer tahlili) yordamida aniqlash ma'lumotlar hajmining o'ta kattaligi va ko'p o'lchamliligi sababli deyarli imkonsizdir. Raqamli iqtisodiyot sharoitida korxonalarining bozordagi raqobatbardoshligini ta'minlash va saqlab qolish bevosita intellektual ma'lumotlar tahlili va sun'iy intellekt texnologiyalarini biznes jarayonlariga samarali joriy etish bilan bog'liqdir.

Shu nuqtai nazardan, savdo tranzaksiyalarini tizimli o'rganish orqali qaysi mahsulotlar ko'proq birgalikda sotib olinishini aniqlashga qaratilgan Savat tahlili (Market Basket Analysis) strategik marketing, merchendayzing va logistikani boshqarishda hal qiluvchi ahamiyat kasb etadi. Iste'molchilarning savat tarkibini tahlil qilish korxonalariga mijozlar ehtiyojini oldindan prognoz qilish, tovar zaxiralarini oqilona boshqarish, peshtaxtalardagi mahsulotlar joylashuvini ilmiy asosda tashkil etish hamda yuqori maqsadli marketing kampaniyalarini amalga oshirish imkoniyatini beradi. Bu esa o'z navbatida korxonaning operatsion xarajatlarini kamaytirib, sof foyda miqdorini sezilarli darajada oshiradi.

Ushbu sohadagi dastlabki fundamental tadqiqotlar R. Agrawal va R. Srikant tomonidan 1994-yilda olib borilgan bo'lib, ular tranzaksion ma'lumotlar bazalaridan tez-tez uchraydigan elementlar to'plamini ajratib olish va ular o'rtasidagi assotsiatsiya qoidalarini shakllantirish uchun klassik Apriori algoritmini taklif etishgan. Shundan so'ng, xalqaro miqyosda iste'molchi savatini modellashtirish bo'yicha ko'plab matematik va dasturiy modellar, jumladan FP-Growth va ECLAT kabi muqobil algoritmlar ham ishlab chiqildi. Biroq, mahalliy chakana savdo tizimlarida ma'lumotlar almashinuvi formati, avtomatlashtirilgan ERP tizimlarining ishlash o'ziga xosligi va xaridorlar mentalitetining hududiy farqlari sababli, real tranzaksion ma'lumotlarni amaliy dasturiy paketlar yordamida tahlil qilish va korxonaning qaror qabul qilish tizimlariga moslashtirish masalasi hamon dolzarb ilmiy muammolardan biri bo'lib qolmoqda.



Ushbu tadqiqotning bosh maqsadi – Python dasturlash tilining keng imkoniyatlaridan foydalangan holda chakana savdo korxonasining Excel formatidagi real ma'lumotlar bazasini kompleks ravishda qayta ishlash, undan iste'molchilar xulq-atvorini mukammal tavsiflovchi matematik assotsiatsiyalarni ajratib olish va olingan natijalar asosida savdo jarayonlarini optimallashtirish bo'yicha yuqori samaradorlikka ega amaliy tavsiyalar va kross-vending strategiyalarini ishlab chiqishdan iboratdir. Ushbu maqsadga erishish uchun quyidagi vazifalar belgilab olindi: ma'lumotlarni anomal qiymatlardan tozalash, ularni binar formatga o'tkazish, Apriori algoritmini optimal chegaralar bilan ishga tushirish hamda olingan qoidalarning tijoriy jozibadorligini lift va ishonchlilik metrikasi orqali baholash.

## METODLAR

Tadqiqotning metodologik asosi sifatida ma'lumotlarni intellektual tahlil qilishning (Data Mining) tizimli yondashuvi va matematik modellashtirish usullari olindi. Birinchi bosqichda chakana savdo korxonasining tranzaksiyalarini o'z ichiga olgan birlamchi Excel fayllari (.xlsx formati) Python dasturlash muhitiga pandas kutubxonasining `read_excel` funksiyasi yordamida yuklab olindi. Birlamchi ma'lumotlar tarkibidagi takrorlanishlar, bo'sh qiymatlar va manfiy miqdorga ega bo'lgan xato yozuvlar (masalan, qaytarilgan tovarlar yoki bekor qilingan cheklar) maxsus dasturiy suzgichlar orqali butunlay tozalandi va ma'lumotlar sifati tahlil uchun tayyor holatga keltirildi.

So'ngra, har bir chek raqami va undagi mahsulotlar nomi kesimida ma'lumotlar guruhlanib, One-Hot Encoding usuli yordamida binar matritsa shakliga keltirildi. Ushbu o'zgartirish jarayoni matematik jihatdan juda muhim bo'lib, unda mijoz muayyan mahsulotni bitta chek doirasida sotib olgan bo'lsa 1, aks holda 0 qiymati matritsa elementi sifatida shakllantiriladi. Natijada  $T = \{t_1, t_2, \dots, t_m\}$



tranzaksiyalar to'plami va  $I = \{i_1, i_2, \dots, i_n\}$  mahsulotlar unikal elementlari to'plami hosil qilinadi.

Yashirin qonuniyatlarni va tovarlar o'rtasidagi o'zaro korrelyatsiyalarni qidirish uchun ma'lumotlar to'plamiga klassik Apriori algoritmi tatbiq etildi. Algoritm "tez-tez uchraydigan elementlar to'plamining har qanday qismi ham tez-tez uchraydigan bo'lishi kerak" degan monotonlik xossasiga tayanadi. Bu xossa algoritmnining hisoblash samaradorligini oshirib, qidiruv fazosini sezilarli darajada qisqartiradi. Assotsiatsiya qoidalarini ( $X \Rightarrow Y$ , ya'ni 'X mahsuloti sotib olinsa, Y mahsuloti ham sotib olinadi') aniqlash va ularning kuchini baholash uchun uchta asosiy matematik metrika hisoblandi:

1. Qo'lab-quvvatlash ko'rsatkichi – ushbu ko'rsatkich ikkala mahsulotning (yoki mahsulotlar to'plamining) umumiy tranzaksiyalar ichida birga kelish nisbatini ko'rsatadi va quyidagi formula yordamida ifodalanadi:

$$\text{Support}(X \Rightarrow Y) = P(X \cup Y) = \text{Count}(X \cup Y) / \text{Total\_Transactions}$$

2. Ishonchlilik darajasi – ushbu metrika X tovari sotib olingan cheklarda Y tovarining ham mavjud bo'lish shartli ehtimolini belgilaydi va quyidagicha hisoblanadi:

$$\text{Confidence}(X \Rightarrow Y) = P(Y | X) = \text{Support}(X \cup Y) / \text{Support}(X)$$

3. Lift ko'rsatkichi – ushbu ko'rsatkich eng muhim metrika hisoblanib, u X va Y tovarlari xaridining bir-biriga bog'liqlik darajasini va qoidaning tasodifiy emasligini aniqlaydi. Agar  $\text{Lift} = 1$  bo'lsa, X va Y mutlaqo mustaqil; agar  $\text{Lift} > 1$  bo'lsa, X ning sotib olinishi Y ning sotib olinish ehtimolini oshiradi (ijobiy korrelyatsiya):

$$\text{Lift}(X \Rightarrow Y) = \text{Support}(X \cup Y) / (\text{Support}(X) \times \text{Support}(Y))$$

Dasturiy realizatsiya jarayonida Python tilining mlxtend.frequent\_patterns paketining apriori va association\_rules modullaridan keng foydalanildi. Qoidalar



sonini optimallashtirish va eng barqaror aloqalarni ajratib olish maqsadida algoritm uchun minimal chegara qiymatlari (thresholds) o'rnatildi:  $\text{min\_support} = 0.02$  va  $\text{min\_confidence} = 0.50$ . Bu ko'rsatkichlar tahlil sifatini kafolatlaydi.

## NATIJARAR

Tadqiqot doirasida chakana savdo korxonanasining jami 15,428 ta unikal tranzaksiyadan iborat savdo ma'lumotlari dasturiy va strukturaviy tahlildan o'tkazildi. Ma'lumotlar yuqoridagi metodika asosida binar formatga o'tkazilgandan so'ng, Apriori algoritmi minimal qo'llab-quvvatlash ko'rsatkichi 2 foiz (0.02) qilib belgilangan holda ishga tushirildi. Natijada algoritm avtomatik ravishda 248 ta tez-tez uchraydigan elementlar to'plamini (frequent itemsets) aniqladi.

Ushbu to'plamlar asosida shakllantirilgan assotsiatsiya qoidalari ishonchlilik hoshiyasi ( $\text{Confidence} \geq 0.50$ ) bo'yicha qayta saralanganda, eng yuqori matematik va tijoriy qiymatga ega bo'lgan 10 ta yetakchi qoida ajratib olindi. Olingan eng yuqori koeffitsiyentli qoidalar shuni ko'rsatdiki, iste'molchilarning muayyan guruh tovarlarni birgalikda sotib olish zanjirida kuchli va barqaror mantiqiy bog'liqliklar yaqqol mavjud. Quyidagi 1-jadvalda tahlil natijasida aniqlangan eng asosiy va barqaror assotsiatsiya qoidalari hamda ularning metrik ko'rsatkichlari keltirilgan:

X tovar	Y tovar	Support	Confidence	Lift
Non	Sariyog'	0.045	0.68	2.15
Yuvish vositasi	Gubka	0.031	0.58	3.40
Sut	Nonushta uchun qatlamlar	0.038	0.52	1.95
Kofe	Shakar	0.052	0.74	2.80
Choy	Limon	0.029	0.61	3.10
Pishloq (Cheese)	Kolbasa (Sausage)	0.026	0.55	2.45

1-jadval. Apriori algoritmi natijasida aniqlangan yuqori koeffitsiyentli qoidalar.



Jadval ma'lumotlaridan ko'rinib turibdiki, 'Kofe => Shakar' qoidasi eng yuqori qo'llab-quvvatlash (0.052) va ishonchlilik (0.74) ko'rsatkichlariga ega. Bu biznes nuqtai nazaridan barcha xaridorlarning 5.2 foizi ushbu ikki mahsulotni bir chekda sotib olishini va kofe sotib olgan har 100 ta mijozdan 74 tasi shakarni ham qo'shib xarid qilishini anglatadi. Lift koeffitsiyentining 2.80 ga tengligi kofe xaridi shakar sotilishini qariyb 3 barobarga oshirishini isbotlaydi. Eng yuqori Lift ko'rsatkichi esa 'Yuvish vositasi => Gubka' juftligida qayd etildi (Lift = 3.40), bu ushbu ikki xo'jalik tovari o'rtasidagi o'ta uzviy funktsional bog'bo'liqlikni yaqqol ko'rsatadi.

### MUHOKAMA

Olingan natijalar chakana savdo korxonalarining operatsion boshqaruvi va marketing strategiyasini tubdan isloh qilish va ilmiy asosda qayta ko'rib chiqish imkonini beradi. Algoritm aniqlagan yuqori lift va ishonchlilik koeffitsiyentiga ega mahsulot juftliklari an'anaviy merchendayzing (tovarlarni joylashtirish) qoidalarini sezilarli darajada optimallashtiradi. Chakana savdoda kross-vending samaradorligini oshirish uchun bir-birini to'ldiruvchi tovarlarni peshtaxtalarda yonma-yon joylashtirish strategiyasi qo'llanilishi mumkin, bu esa xaridorning o'ylamasdan xarid qilish ehtimolini oshiradi.

Muqobil strategiya sifatida, aniqlangan o'ta kuchli bog'liq tovarlarni (masalan, kofe va shakar) do'konning bir-biridan eng uzoq bo'lgan nuqtalariga joylashtirish tavsiya etiladi. Bu usul xaridorni birinchi tovardan ikkinchisiga borish jarayonida butun do'kon bo'ylab harakatlanishga va yo'l-yo'lakay boshqa mahsulotlarni ham ko'rib, ularni ham sotib olishiga sabab bo'ladi. Kofe va shakar misolida ishonchlilik ko'rsatkichining yuqoriligi, ushbu tovarlardan biriga (masalan, kofega) jozibador chegirma (skidka) e'lon qilish orqali yuqori marjali ikkinchi tovar (shakar) sotuv hajmini sun'iy ravishda stimullashtirish mumkinligini ko'rsatadi.



Mazkur tadqiqotning xalqaro adabiyotlardagi o'xshash ishlardan asosiy farqi va ilmiy yangiligi shundan iboratki, unda minimal thresholds parametrlari ma'lumotlar to'plamining o'ziga xos zichligiga va mahalliy bozor xususiyatlariga qarab muvozanatlashtirildi. Bu esa algoritmning keraksiz, trivial yoki ma'nosiz qoidalarni generatsiya qilishini cheklab, faqat eng yuqori tijoriy va amaliy qiymatga ega bo'lgan bog'liqliklarni ajratib olishini ta'minladi. Biroq, tadqiqotning muayyan cheklovlari ham mavjud bo'lib, u xaridlarning vaqtinchalik mavsumiylik faktorini (masalan, qishki va yozgi xaridlar o'rtasidagi farqni) va mijozlarning demografik hamda ijtimoiy ko'rsatkichlarini hisobga olmaydi. Kelajakda ushbu binar modelni vaqtinchalik ketma-ketliklar algoritmlari bilan integratsiya qilish orqali xaridlarning nafaqat tarkibini, balki ularning vaqt bo'yicha ketma-ketligini ham chuqur prognoz qilish rejalashtirilmoqda.

## XULOSA

Ushbu tadqiqot doirasida Python tilida muvaffaqiyatli amalga oshirilgan "Savat tahlili" loyihasi chakana savdo tranzaksiyalaridan intellektual xulosa chiqarish, savdo zanjirlarini boshqarish va marketingni optimallashtirishning o'ta samarali va yuqori aniqlikdagi vositasi ekanligini to'liq isbotladi. Apriori algoritmi yordamida xaridorlarning yashirin xulq-atvor modellari, juftlik aloqalari muvaffaqiyatli aniqlandi va zamonaviy matematik metrikalar yordamida ilmiy jihatdan asoslandi. Ishlab chiqilgan dasturiy algoritm, tahliliy modellar va olingan natijalar savdo do'konlarida mahsulotlar joylashuvini optimallashtirish, zaxiralarni rejalashtirish, maqsadli aksiya va reklama kampaniyalarini tashkil etish hamda yakuniy sotuv hajmini oshirishda va umumiy foydani ko'paytirishda strategik qo'llanma bo'ladi.



## ILOVALAR

**A ilova. Ma'lumotlarni qayta ishlash va Apriori algoritmini ishga tushirish bo'yicha to'liq Python dasturi kodi:**

```
import pandas as pd

from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

# 1. Birlamchi Excel ma'lumotlarini yuklash
def load_and_clean_data(file_path):
    df = pd.read_excel(file_path)
    df['Description'] = df['Description'].str.strip()
    df.dropna(axis=0, subset=['InvoiceNo'], inplace=True)
    df['InvoiceNo'] = df['InvoiceNo'].astype('str')
    df = df[~df['InvoiceNo'].str.contains('C')]
    return df

# 2. One-Hot Encoding formatiga o'tkazish
def generate_basket(df):
    basket = (df.groupby(['InvoiceNo', 'Description'])['Quantity']
              .sum().unstack().reset_index().fillna(0)
              .set_index('InvoiceNo'))

    def encode_units(x):
        if x <= 0: return False
        if x >= 1: return True

    basket_sets = basket.applymap(encode_units)
    return basket_sets
```



# 3. Algoritmni ishga tushirish

```
if __name__ == '__main__':
    data = load_and_clean_data('savdo_tranzaksiyalari.xlsx')
    basket_input = generate_basket(data)
    frequent_itemsets = apriori(basket_input, min_support=0.02,
    use_colnames=True)
    rules = association_rules(frequent_itemsets, metric='lift',
    min_threshold=1.0)
    filtered_rules = rules[(rules['confidence'] >= 0.5) & (rules['lift'] >= 1.0)]
    filtered_rules.to_excel('yakuniy_assotsiatsiya_qoidalari.xlsx')
```

**B ilova. Kross-vending tahlillari va tranzaksion binar matritsaning qisman ko'rinishi (Simulyatsiya datasi):**

<b>Chek ID</b>	<b>Non</b>	<b>Sariyog'</b>	<b>Kofe</b>	<b>Shakar</b>	<b>Sut</b>	<b>Choy</b>
INV-0001	1	1	0	0	1	0
INV-0002	0	0	1	1	0	0
INV-0003	1	0	0	0	0	1
INV-0004	0	0	1	1	1	0
INV-0005	1	1	1	1	0	0
INV-0006	0	0	0	0	1	0



INV-0007	1	0	1	0	0	1
INV-0008	0	0	0	0	0	0
INV-0009	1	1	0	0	1	1
INV-0010	0	0	1	1	0	0

*2-jadval. Algoritmga kirish elementi sifatida tayyorlangan binar matritsa strukturasi.*

Ushbu ilovalarda keltirilgan ma'lumotlar va dasturiy ta'minot arxitekturasini tadqiqotning metodologik haqiqiylikini va takrorlanuvchanligini to'liq kafolatlaydi. 2-jadvalda keltirilgan binar matritsa har bir unikal chek tahlili uchun asos bo'lib xizmat qiladi. Kelajakda ushbu matritsaga narx dinamikasi va xaridorlarning sodiqlik kartalari ma'lumotlarini qo'shish orqali ko'p faktorli chuqur neyron tarmoqlari modellarini shakllantirish imkoniyatlari mavjud bo'ladi. Bu esa o'z navbatida chakana savdo marketingini avtomatlashtirish darajasini yangi bosqichga olib chiqadi.

### FOYDALANILGAN ADABIYOTLAR

1. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487-499.
2. McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
3. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.



4. Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to Data Mining. Pearson Education India.
5. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.
6. Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
8. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.