

TASNIFLASH VA REGRESSIYA ALGORITMLARI QULAYLIKLARI VA AFZALLIKLARI

УДОБСТВО И ПРЕИМУЩЕСТВА АЛГОРИТМОВ КЛАССИФИКАЦИИ И РЕГРЕССИИ

CONVENIENCE AND ADVANTAGES OF CLASSIFICATION AND REGRESSION ALGORITHMS

Abdug'afforov Dilyorbek Dilshodjon zoda

*Farg'ona Davlat Universiteti Amaliy matematika
yo'nalishi 3 -kurs talabasi Gmail: abdugaforov02@bk.ru*

Onarqulov Maqsadjon

*Farg'ona Davlat Universiteti Amaliy matematika
kafedrasi o'qituvchisi Gmail: maxmaqsad@gmail.com*

Annotatsiya: Tasniflash va regressiya-bu mashinani o'rganish muammolarining eng keng tarqalgan turlaridan biri. Tasniflash elektron pochta spam yoki yo'qligi kabi kategorik natijani bashorat qilishni o'z ichiga oladi, regressiya esa uning xususiyatlariga asoslangan uy narxi kabi raqamli natijani bashorat qilishni o'z ichiga oladi. Ham tasniflash, ham regressiya Real muammolarni hal qilish va aniq bashorat qilish uchun muhim vositadir.

Аннотация: Классификация и регрессия-два наиболее распространенных типа проблем машинного обучения. Классификация включает прогнозирование категориального результата, такого как спам по электронной почте или нет, в то время как регрессия включает прогнозирование численного результата, такого как цена дома, на основе его характеристик. И классификация, и регрессия являются важными инструментами для решения реальных проблем и точного прогнозирования.

Annotation: Classification and regression are some of the most common types of machine learning problems. Classification involves predicting a categorical outcome such as email spam or not, while regression involves predicting a numerical outcome such as home price based on its properties. Both classification and regression are important tools for solving and accurately predicting Real problems.

Kirish(Введение/Introduction) Tasniflash-bu kirish o'zgaruvchilari asosida kategorik yoki diskret chiqish o'zgaruvchisini bashorat qiladigan [boshqariladigan mashinani o'rganish algoritmi](#). Kirish o'zgaruvchilari ko'pincha xususiyatlar yoki bashoratchilar deb ataladi, chiqish o'zgaruvchisi esa sinf yoki yorliq deb ataladi.

Tasniflash algoritmi xaritalash funktsiyasini etiketlangan o'quv ma'lumotlar to'plami asosida kirishdan chiqishga o'zgaruvchilarga o'rganishga qaratilgan. O'quv ma'lumotlar to'plami har biri sinf yorlig'i bilan bog'liq bo'lgan misollardan iborat. Algoritm ushbu yorliqli ma'lumotlardan ularning kirish o'zgaruvchilari asosida yangi, ko'rinmaydigan holatlarning sinf yorlig'ini bashorat qilish uchun model yaratish uchun foydalanadi.

Bir nechta tasniflash algoritmlari mavjud, shu jumladan qaror daraxtlari, [tasodifiy o'rmonlar](#), logistika regressiyasiva qo'llab-[quvvatlash vektor mashinalari \(svm\)](#).

Tasniflash algoritmlari ko'plab dasturlarga ega, jumladan tasvirni aniqlash, tabiiy tilni qayta ishlash, firibgarlikni aniqlash va kredit skoringi.

Tasniflashning afzalliklari

Tasniflash mashinani o'rganish texnikasi sifatida bir nechta afzalliklarga ega:

1. **Interpretatsiya:** tasniflash modellari ko'pincha yuqori darajada talqin qilinadi, bu model qanday bashorat qilishini va qaysi xususiyatlar yoki bashoratchilar eng muhimligini tushunishni osonlashtiradi.
2. **Aniqlik:** tasniflash modellari ko'pincha xaridor mahsulotni sotib oladimi yoki yo'qmi kabi kategorik natijalarni bashorat qilishda yuqori aniqlikka erishishi mumkin.
3. **Foydalanish qulayligi:** tasniflash modellarini amalga oshirish ko'pincha nisbatan oson va minimal ma'lumotlarni qayta ishlashni talab qiladi, bu ularni yangi boshlanuvchilar yoki tezkor prototiplar uchun yaxshi tanlov qiladi.
4. **O'lchovlilik:** tasniflash modellari ko'pincha katta ma'lumotlar to'plamlari yoki yuqori o'lchovli xususiyat bo'shliqlarini boshqarish uchun kengaytirilishi mumkin, bu ularni turli xil ilovalar uchun mos qiladi.
5. **Mustahkamlik:** tasniflash modellari ko'pincha shovqin va etishmayotgan ma'lumotlarga chidamli bo'lib, ularni o'zgaruvchanlik yoki noaniqlik darajasi yuqori bo'lgan ma'lumotlar uchun yaxshi tanlov qiladi.

Regressiya nima?

Regressiya-bu kirish o'zgaruvchilari asosida uzluksiz raqamli chiqish o'zgaruvchisini bashorat qiladigan boshqariladigan mashinani o'rganish algoritmi. Kirish o'zgaruvchilari ko'pincha xususiyatlar yoki bashoratchilar deb ataladi, chiqish o'zgaruvchisi esa maqsadli yoki qaram o'zgaruvchi deb ataladi.

Regressiya algoritmi belgilangan o'quv ma'lumotlar to'plami asosida xaritalash funktsiyasini kirishdan chiqishga o'zgaruvchilarga o'rganishga qaratilgan. O'quv ma'lumotlar to'plami raqamli maqsad qiymati bilan bog'liq misollardan iborat. Algoritm ushbu yorliqli ma'lumotlardan ularning kirish o'zgaruvchilari asosida yangi, ko'rinmaydigan holatlarning maqsad qiymatini taxmin qilish uchun model yaratish uchun foydalanadi.

Bir nechta regressiya algoritmlari mavjud, shu jumladan chiziqli regressiya, polinom regressiyasi, qaror daraxti regressiyasiva tasodifiy o'rmon regressiyasi. Har

bir algoritmnining kuchli va zaif tomonlari bor va algoritmnini tanlash muayyan muammoga va ma'lumotlarning xususiyatlariga bog'liq.

Shuningdek qarang Tuzilgan Va Tuzilmagan Ma'lumotlar O'rtasidagi Farq

Regressiya algoritmlari ko'plab dasturlarga ega, shu jumladan aktsiyalar narxlarini bashorat qilish, uy-joy narxlarini taxmin qilish, sotishni prognoz qilish va mijozlarning xatti-harakatlarini modellashtirish.

Regressiyaning afzalliklari

Regressiya mashinani o'rganish texnikasi sifatida bir nechta afzalliklarga ega:

1. **Bashoratli kuch:** regressiya-bu sotish, aktsiyalar narxi yoki bemorning natijalari kabi raqamli qiymatlarni bashorat qilishning kuchli usuli. Regressiya modellari kelajakdagi qiymatlarni yuqori aniqlik bilan bashorat qilishga yordam beradigan ma'lumotlardagi naqsh va munosabatlarni aniqlay oladi.

2. **Moslashuvchanlik:** regressiya o'zgaruvchilar o'rtasidagi turli xil munosabatlarni modellashtirish uchun ishlatiladigan moslashuvchan texnikadir. Masalan, chiziqli regressiya o'zgaruvchilar orasidagi chiziqli munosabatlarni, polinom regressiya esa murakkabroq chiziqli bo'lmagan munosabatlarni modellashtirishi mumkin.

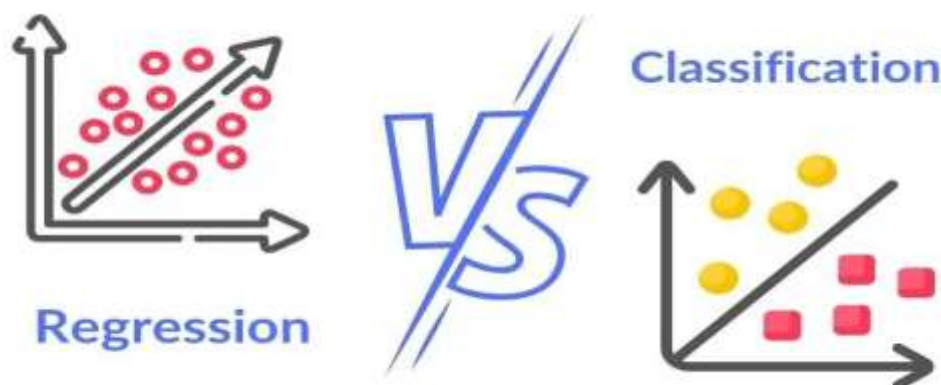
3. **Interpretatsiya:** regressiya modellari ko'pincha juda tushunarli bo'lib, o'zgaruvchilar o'rtasidagi munosabatlarni va ularning natijaga qanday ta'sir qilishini tushunishni osonlashtiradi. Bu, ayniqsa, sog'liqni saqlash yoki moliya kabi sohalarda foydali bo'lishi mumkin, bu erda natijaga olib keladigan asosiy omillarni tushunish juda muhimdir.

4. **Xususiyatning ahamiyati:** regressiya modellari natija o'zgaruvchisini boshqaradigan eng muhim xususiyatlarni yoki bashorat qiluvchilarni aniqlashi mumkin. Ushbu ma'lumot qarorlarni qabul qilishni xabardor qilish va keyingi tekshiruvlarga rahbarlik qilish uchun ishlatilishi mumkin.

5. **Mustahkamlik:** regressiya modellari ko'pincha shovqin va ma'lumotlardan tashqariga chiqadi, bu ularni yuqori o'zgaruvchanlik yoki shovqinli ma'lumotlar uchun yaxshi tanlov qiladi.

Ularning orasidagi farq nima: regressiya va tasniflash?

Regressiya va tasniflash-bu mashinani o'rganishda boshqariladigan ta'limning ikkita muhim turi.



Regressiya uzluksiz qiymatni, tasniflar esa toifani bashorat qiladi.

Regressiya-bu qaram o'zgaruvchi va bir yoki bir nechta mustaqil o'zgaruvchilar o'rtasidagi munosabatni modellashtiradigan bashoratli modellashtirish texnikasi. Regressiya tahlili mustaqil o'zgaruvchilarning ahamiyati asosida bog'liq o'zgaruvchining qiymatini baholashga qaratilgan. Bog'liq o'zgaruvchi uzluksiz, ya'ni u diapazondagi har qanday raqamli qiymatni olishi mumkin.

Regressiya mustaqil o'zgaruvchilarning ma'lum qiymatlari asosida bog'liq o'zgaruvchining kelajakdagi qiymatlarini bashorat qilish uchun ishlatiladi. Regressiya modellariga misol sifatida chiziqli, polinom va logistik regressiya kiradi.

Boshqa tomondan, tasniflash a bashoratli modellashtirish berilgan ma'lumotlar nuqtasini bir nechta oldindan belgilangan sinflardan yoki toifalardan biriga tasniflash yoki tasniflash uchun ishlatiladigan texnika. Tasniflashning maqsadi kirish o'zgaruvchilari va diskret va kategorik bo'lgan chiqish o'zgaruvchilari to'plami o'rtasida xaritalashni o'rganishdir. Tasniflash modellariga misollar kiradi qaror daraxtlari, tasodifiy o'rmonlar, qo'llab-quvvatlash vektorli mashinalarva neyron tarmoqlar.

Xulosa qilib aytganda, regressiya doimiy qiymatlarni bashorat qilish uchun, tasniflash esa diskret yoki kategorik qiymatlarni bashorat qilish uchun ishlatiladi.

Regressiya va tasniflash: ular orasidan qanday tanlash mumkin?

Mashinani o'rganish modelingiz talablariga chuqur sho'ng'iganingizda, siz tasniflash muammosi yoki regressiya muammosi bilan shug'ullanayotganingiz tez-tez ko'rinib turadi. Shunga qaramay, bu qaror toshga qo'yilishi shart emas. Siz tez-tez muammoni o'zgartirishingiz, unga turli burchaklardan qarashingiz va ikkalasini almashtirishingiz mumkin.

Ushbu bo'limda biz sizga misollar va afzalliklarni taqdim etish orqali ikkalasini qanday almashtirishni o'rganishga yordam beramiz deb umid qilamiz. Pirovard maqsad-ikkita texnikani, afzalliklari va kamchiliklarini yaxshiroq tushunish, shunda siz o'zingizning muammoingiz uchun yaxshiroq qaror qabul qilishingiz mumkin.

Regressiya va tasniflash: qanday qilib tasnifni regressiya muammosiga aylantirish mumkin?

Tasniflash va regressiya muammolari o'rtasida to'g'ridan-to'g'ri konvertatsiya qilish har doim ham mumkin emas, chunki ular turli xil ma'lumotlar turlarini o'z ichiga oladi va boshqa maqsadlarga ega. Biroq, muayyan vaziyatlarda ikki turdagi masalalarni aylantirish uchun ba'zi usullardan foydalanish mumkin.

[Shuningdek qarang Bashoratli Analytics Oddiy Qildi & Qanday Python Tutorial Uchun](#)

Tasniflash muammosini regressiya muammosiga aylantirishning usullaridan biri bu har bir sinf yorlig'iga raqamli qiymatlarni belgilash va keyin muammoni regressiya muammosi sifatida ko'rib chiqishdir.

Masalan, agar bizda uchta sinf (A, B va C) bilan tasniflash muammosi bo'lsa, har bir toifaga 1, 2 va 3 qiymatlarni belgilashimiz mumkin. Keyin biz regressiya modelini har bir kirish misoli uchun ushbu raqamli qiymatlarni bashorat qilishga o'rgatishimiz va keyin kutilgan sinfni olish uchun bashorat qilingan qiymatlarni eng yaqin butun songa yaxlitlashimiz mumkin.

Aksincha, biz bog'liq o'zgaruvchining diapazonini diskret toifalar yoki qutilar to'plamiga bo'lish orqali regressiya muammosini tasniflash muammosiga aylantirishimiz mumkin. Keyin biz muammoni [ko'p sinfli tasniflash](#) muammosi sifatida ko'rib chiqishimiz mumkin, bu erda maqsad har bir misolni oldindan belgilangan toifalar yoki qutilardan biriga tasniflashdir. Ushbu uslub binning yoki diskretizatsiya deb nomlanadi.

Tasniflash muammosini regressiya muammosiga aylantirish misoli

Tasniflash muammosini regressiya muammosiga aylantirishning bir misoli ikkilik natija ehtimolini taxmin qilish bo'lishi mumkin, masalan, xaridor mahsulotni sotib oladimi yoki yo'qmi. An'anaviy tasniflash muammosida chiqish ikkilik yorliq bo'ladi (ya'ni sotib olingan yoki sotib olinmagan). Biroq, biz buni 0 dan 1 gacha bo'lgan doimiy o'zgaruvchini sotib olish ehtimolini bashorat qilish orqali regressiya muammosiga aylantirishimiz mumkin.

Buning uchun biz logistik regressiya algoritmidan foydalanishimiz mumkin, bu odatda [ikkilik tasniflash](#) muammolari uchun ishlatiladigan regressiya algoritmining bir turi. Logistik regressiya algoritmi har bir misol uchun ehtimollik balini chiqaradi, bu taxmin qilingan sotib olish ehtimoli sifatida talqin qilinishi mumkin. Keyin biz misollarni ikkita sinfga (ya'ni sotib olingan yoki sotib olinmagan) tasniflash uchun ushbu ehtimollik balli uchun chegara o'rnatishimiz mumkin (masalan, 0,5).

Shu tarzda tasniflash muammosini regressiya muammosiga aylantirish orqali biz ma'lumotlarga qo'shimcha tushunchalar olishimiz va bashoratlarimizning aniqligini oshirishimiz mumkin. Masalan, modelimizning ishlashini o'lchash va yanada takomillashtirish uchun o'rtacha kvadratik xato yoki o'rtacha mutlaq xato kabi regressiyani baholash ko'rsatkichlaridan foydalanishimiz mumkin.

Tasniflash muammosini regressiya muammosiga aylantirishning afzalliklari

Tasniflash muammosini regressiya muammosiga aylantirish bir nechta potentsial foyda keltirishi mumkin, jumladan:

1. **Bashoratli aniqlikning oshishi:** tasniflash muammosini regressiya muammosiga aylantirish orqali biz ma'lumotlarni modellashtirish va bashorat qilish uchun yaxshiroq mos keladigan yanada murakkab regressiya algoritmlaridan foydalanishimiz mumkin. Bu an'anaviy tasniflash algoritmlaridan foydalanishga nisbatan yuqori bashoratli aniqlikka olib kelishi mumkin.

2. **Ma'lumotlarni yaxshiroq tushunish:** regressiya modellari osongina tasavvur qilish va tahlil qilish mumkin bo'lgan doimiy chiqishni ta'minlaydi. Bu mustaqil va qaram o'zgaruvchilar o'rtasidagi munosabatlar haqida yaxshiroq tushuncha berishi va tasniflash modellarida darhol ko'rinmasligi mumkin bo'lgan ma'lumotlarning naqshlari va tendentsiyalarini aniqlashga yordam beradi.

3. **Modellashtirishda ko'proq moslashuvchanlik:** regressiya modellari mustaqil va qaram o'zgaruvchilar o'rtasidagi munosabatlarni modellashtirishda katta moslashuvchanlikni taklif etadi. Masalan, chiziqli bo'lmagan munosabatlarni modellashtirish uchun polinom yoki spline regressiyasidan foydalanishimiz mumkin, bu tasniflash modellari bilan mumkin emas.

4. **Boshqa modellar bilan osonroq integratsiya:** regressiya modellari boshqa regressiya modellari bilan osongina birlashtirilishi mumkin, bu esa murakkabroq modellarni yaratishda foydali bo'lishi mumkin. Odatda mustaqil bo'lgan tasniflash modellarida bu mumkin emas.

Umuman olganda, tasniflash muammosini regressiya muammosiga aylantirish an'anaviy tasniflash modellari yaroqsiz bo'lganda yoki biz ma'lumotlarga yaxshiroq tushuncha olishni xohlaganimizda foydali bo'lishi mumkin.

Regressiya muammosini tasniflash muammosiga aylantirish misoli

Regressiya muammosini tasniflash muammosiga aylantirishning bir misoli doimiy o'zgaruvchiga asoslangan ikkilik natija ehtimolini bashorat qilish bo'lishi mumkin.

[Shuningdek qarang Matn Tasnifi: Pythonda Qanday Qilib \[Eng Yaxshi 2 Usul Mashinani O'rganish Va Chuqur O'rganish\]](#)

Misol uchun, biz bemorning qon shakar darajasiga qarab diabetga chalinganligini taxmin qilmoqchimiz. An'anaviy regressiya muammosida chiqish bemorning qon shakar darajasini ifodalovchi doimiy raqamli qiymat bo'ladi. Biroq, biz bemorning diabetga chalinish ehtimolini taxmin qilish orqali buni ikkilik tasniflash muammosiga aylantirishimiz mumkin.

Buning uchun biz bemorda diabetga chalinish ehtimoli yuqori bo'lgan qon shakar darajasi uchun chegara qiymatini belgilashimiz mumkin. Misol uchun, Pol 140 mg/dL da belgilangan deylik. Keyin qon shakar darajasi 140 mg/dL dan yuqori bo'lgan bemorlar diabetga chalingan, qon shakar darajasi 140 mg/dL dan past bo'lganlar diabetga chalingan deb tasniflanadi.

Ushbu tasniflash vazifasini bajarish uchun logistik regressiya algoritmidan foydalanishimiz mumkin. Birinchidan, algoritm qon shakar darajasidan diabetning ikkilik sinf yorlig'iga yoki diabetga emas, balki xaritalash funksiyasini o'rganadi. Keyinchalik, algoritm har bir misol uchun ehtimollik balini chiqaradi, bu bemorning diabetga chalinishi ehtimoli sifatida talqin qilinishi mumkin.

Regressiya muammosini shu tarzda tasniflash muammosiga aylantirib, biz yanada ta'sirchan va izohlash osonroq bo'lgan bashorat qilishimiz mumkin. Shuningdek, biz modelimizning ishlashini o'lchash va uni yanada yaxshilash uchun [aniqlik, aniqlik, eslash va F1-ball](#) kabi tasnifni baholash ko'rsatkichlaridan foydalanishimiz mumkin.

Regressiya va tasniflash: regressiya muammosini tasniflash muammosiga aylantirishning afzalliklari

Regressiya muammosini tasniflash muammosiga aylantirish bir nechta potentsial foyda keltirishi mumkin, jumladan:

1. **Muammoning soddaligi:** tasniflash muammolarini tushunish odatda regressiya muammolariga qaraganda ancha soddadir. Regressiya muammosini tasniflash muammosiga aylantirish orqali biz uni soddalashtirishimiz va natijalarni talqin qilishni osonlashtirishimiz mumkin.

2. **Yaxshiroq talqin qilish:** tasniflash modellari osongina talqin qilinishi va tushunilishi mumkin bo'lgan kategorik natijalarni beradi. Bu, ayniqsa, natijalar texnik bo'lmagan manfaatdor tomonlarga etkazilishi kerak bo'lgan dasturlarda foydali bo'lishi mumkin.

3. **Chetga chiquvchilarning mustahkamligi:** tasniflash modellari odatda regressiya modellariga qaraganda chetga chiquvchilar va ma'lumotlar xatolariga nisbatan kuchliroqdir. Buning sababi shundaki, tasniflash modellari aniq sonli munosabatlarga emas, balki faqat kirish va chiqish o'zgaruvchilari o'rtasidagi kategorik munosabatlarga qaratilgan.

4. **Balanssiz ma'lumotlar bilan yaxshiroq ishlash: bir sinf boshqalarga qaraganda ancha keng tarqalgan muvozanatsiz ma'lumotlar** tasniflash muammolarida qiyinchilik tug'dirishi mumkin. Biroq, tasniflash modellarida muvozanatsiz ma'lumotlarni boshqarish uchun oversampling va undersampling kabi bir nechta texnikalardan foydalanish mumkin.

5. **Moslashuvchan baholash ko'rsatkichlari:** tasniflash modellari aniqlik, eslash, F1-ball va [AUC-ROC egr](#) chizig'i kabi model ishlashini o'lchash uchun ishlatilishi mumkin bo'lgan baholash ko'rsatkichlarining yanada keng doirasiga ega. Ushbu ko'rsatkichlar tasniflash muammolari uchun aniq ishlab chiqilgan va o'rtacha kvadratik xato kabi an'anaviy [regressiya ko'rsatkichlariga](#) qaraganda modelning ishlashi haqida ko'proq ma'lumot berishi mumkin.

Regressiya muammosini tasniflash muammosiga aylantirish, agar masala kategorik chiqishga yaxshiroq mos keladigan bo'lsa yoki biz muammoni soddalashtirib, uni yanada tushunarli qilishni xohlasak foydali bo'lishi mumkin. Biroq, ushbu texnikadan

foydalanish to'g'risida qaror qabul qilishdan oldin ma'lumotlarning xususiyatlarini va o'ziga xos vaziyatni diqqat bilan ko'rib chiqish kerak.

Xulosa

Tasniflash va regressiya-bu kirish o'zgaruvchilari asosida bashorat qilish uchun ishlatiladigan mashinani o'rganishda ikkita asosiy tushuncha. Tasniflash algoritmlari kategorik yoki diskret chiqish o'zgaruvchilarini bashorat qilish uchun, regressiya algoritmlari esa doimiy raqamli chiqish o'zgaruvchilarini bashorat qilish uchun ishlatiladi.

Ba'zan, tasniflash muammosini regressiya muammosiga yoki aksincha aylantirish foydali bo'lishi mumkin. Shunday qilib, biz ma'lumotlarga qo'shimcha tushunchalar olishimiz va bashoratlarimizning aniqligini oshirishimiz mumkin. Biroq, muammo turini konvertatsiya qilish to'g'risidagi qaror mavjud bo'lgan aniq muammo va ma'lumotlarning xususiyatlariga asoslanishi kerak.

Oxir oqibat, tanlov tasnifi va regressiya biz hal qilmoqchi bo'lgan muammoga va biz ishlayotgan ma'lumotlarning xususiyatiga bog'liq. Shuning uchun farqlarni tushunish va tegishli algoritmlar va baholash ko'rsatkichlarini tanlash juda muhimdir.

FOYDALANILGAN ADABIYOTLAR:

1. "Big Data: A Revolution That Will Transform How We Live, Work, and Think".
2. "Data Science for Business" Avtorlar: Foster Provost, Tom Fawcett
3. "Big Data: Principles and Paradigms"
4. "Big Data Analytics: A Practical Guide for Managers"
5. "Hadoop: The Definitive Guide""Big Data: The Next Frontier for Innovation, Competition, and Productivity"
6. "The Fourth Paradigm: Data-Intensive Scientific Discovery"
7. "Big Data Analytics: Challenges and Opportunities"
8. "The Impact of Big Data on Business Models"