

리더스 앱 기능 제안.





데이터예측모델과
기계학습의응용

— 서론

제공받은 데이터
개발 목적

본론 1
개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

기대 방향

제공받은 데이터.

USER



FOLLOW



BOOK



CAT



USER_CAT



USER_BOOK



BOOK_CAT



SCRAP





데이터예측모델과
기계학습의응용

개발 목적.

개발자 편의 목적 알고리즘



- 이용자들을 두 그룹으로 구분
ACTIVE USER, POTENTIAL ACTIVE USER
- 각 그룹에 개별 추천 알고리즘

이용자 편의 목적 알고리즘



- 특정 책에 대하여 유저가 그 책을 읽을 확률을 예측
- 60% 이상이라면 추천

— 서론

제공받은 데이터
개발 목적

본론 1
개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.





데이터 예측 모델과
기계 학습의 응용

서론

본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

데이터 선택

USER



FOLLOW



BOOK



CAT



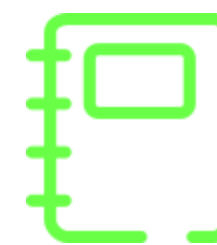
USER_CAT



USER_BOOK



BOOK_CAT



SCRAP





데이터 예측 모델과
기계 학습의 응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

데이터 선택

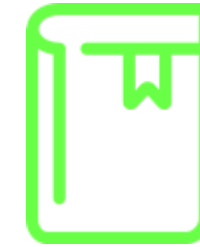
USER



FOLLOW



BOOK



CAT



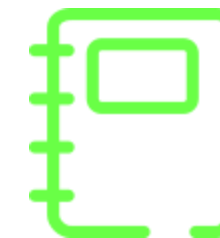
USER_CAT



USER_BOOK



BOOK_CAT



SCRAP





데이터 예측 모델과
기계 학습의 응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

데이터 선택

```
1 str(scrap)
2 str(user_book)
3 str(user_cat)
4 str(book_cat)
5 str(book)
```

Scrap 데이터 가공

```
6
7
8 #스크랩 파일 안 유저 아이디의 빈도 수 계산.
9
10 scrap_table = data.frame(table(scrap$user_id))
11 scrap_table_sort = scrap_table[c(order(-scrap_table$Freq)),]
12
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

user_book 데이터 가공

```
#유저들의 책 평가 중 읽어봤던 책들의 내용만 가져오기.  
read_only = user_book[user_book$read_status == "READ_STATUS_DONE",]  
#그 내용들을 book_id의 빈도 수 즉 얼마나 많이 읽혔는지 가져오기  
read_only_Freq = data.frame(table(read_only$book_id))
```

```
# 그 중 20번 아래로 등장한 데이터들은 없애기  
read_only_Freq_more20 = read_only_Freq[read_only_Freq$Freq >= 20 ,]  
names(read_only_Freq_more20)[1] <- c("book_id")  
# 20번 이상 읽힌 유저_북 정보만 남기기.  
inner <- merge(  
  x = read_only,  
  y = read_only_Freq_more20,  
  by = 'book_id')
```




데이터 예측 모델과
기계 학습의 응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

Book_id 데이터 가공

```
#book_id 평점 평균 데이터 만들기.  
  
inner_groub_mean = aggregate(rate~book_id, inner, mean)  
  
# 평점 평균이 4.8보다 큰 내용들만 추출.  
  
inner_groub_mean_more4.8 = inner_groub_mean[inner_groub_mean$rate > 4.8 ,]  
  
  
my_id <- readline('자신의 유저 id : ')
```

Active User – Potential Active user

개별 추천 알고리즘.





데이터 예측 모델과
기계 학습의 응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

Active user

```
# 현재 유저가 스크랩을 쓴 빈도수 상위 20%안에 든다면(20:80) (240= 1200 × 0.2)

if(which(scrap_table_sort$var1 == my_id)<240){
  #상위 20퍼에 해당(active user) >> 평점 4.8 이상의 책들 중 sample 함수를 이용해 랜덤으로 하나 추천
  recommend_book_id = inner_groub_mean_more4.8[sample(nrow(inner_groub_mean_more4.8),1),]$book_id
  recommend_book = book[book$id == recommend_book_id, ]
  cat("당신에게 추천하는 책은 \'", recommend_book$title, "\" 입니다.")
}

else{
  #80% 해당(potential active user) >> 유저의 관심 항목에 따라 다른 추천
  #유저 아이디에 해당하는 유저 카테고리 정보 로드
  my_cat = user_cat[user_cat$user_id == my_id ,]
  # 모든 유저 카테고리 정보들의 빈도 수를 출력

  user_prefer_cat_freq = data.frame(table(user_cat$title))
```



데이터 예측 모델과
기계 학습의 응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

inner join

```
# inner 함수로 이너조인
```

```
my_cat <- merge(  
  x = my_cat,  
  y = user_prefer_cat_freq,  
  by = 'title')
```

```
my_cat = my_cat[c(order(-my_cat$Freq)),]
```

```
#유저 카테고리 선호
```

```
user_prefer_cat_freq = user_prefer_cat_freq[c(order(-user_prefer_cat_freq$Freq)),]
```

```
names(user_prefer_cat_freq)[1] <- c("title")
```




데이터 예측 모델과
기계 학습의 응용

서론

본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

Potential Active User

```
#관심 카테고리 3개 이상 일 시 (if else 함수 사용)
#가장 많이 선호되는 3가지 카테고리에 대한 도서 정보들을 가져오고 그 도서들 중 sample 함수를 통해 랜덤으로 하나만 추출.
#그 동작을 3가지 카테고리에 대해 동일하게 반복
if(nrow(my_cat) >= 3){
  reco_1_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[1,][1]$title, ]$book_category_id, ]
  reco_1_book_id = reco_1_book_list[sample(nrow(reco_1_book_list),1),]$book_id
  reco_2_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[2,][1]$title, ]$book_category_id, ]
  reco_2_book_id = reco_2_book_list[sample(nrow(reco_2_book_list),1),]$book_id
  reco_3_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[3,][1]$title, ]$book_category_id, ]
  reco_3_book_id = reco_3_book_list[sample(nrow(reco_3_book_list),1),]$book_id

  recommend_1_book = book[book$id == reco_1_book_id, ]
  recommend_2_book = book[book$id == reco_2_book_id, ]
  recommend_3_book = book[book$id == reco_3_book_id, ]
  cat("당신에게 추천하는 책은 \"", recommend_1_book$title, "\", \"", recommend_2_book$title, "\", \"", recommend_3_book$title, "\"입니다.")
}
```




데이터예측모델과
기계학습의응용

서론

— 본론 1

개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

결론
기대 방향

개발자 편의 목적 알고리즘.

Potential Active User

```
#선호하는 카테고리가 두 개라면 그 두개 카테고리에 대한 책들만 추천

}else if(nrow(my_cat) == 2){
  reco_1_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[1,]$title, ]$book_category_id, ]
  reco_1_book_id = reco_1_book_list[sample(nrow(reco_1_book_list),1),]$book_id
  reco_2_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[2,]$title, ]$book_category_id, ]
  reco_2_book_id = reco_2_book_list[sample(nrow(reco_2_book_list),1),]$book_id

  recommend_1_book = book[book$id == reco_1_book_id, ]
  recommend_2_book = book[book$id == reco_2_book_id, ]
  cat("당신에게 추천하는 책은 \"", recommend_1_book$title, "\", \"", recommend_2_book$title, " \"입니다.")
}

#선호하는 카테고리가 한 개라면 한 개만 추천
reco_1_book_list = book_cat[book_cat$book_category_id == cat[cat$name == my_cat[1,]$title, ]$book_category_id, ]
reco_1_book_id = reco_1_book_list[sample(nrow(reco_1_book_list),1),]$book_id

recommend_1_book = book[book$id == reco_1_book_id, ]
cat("당신에게 추천하는 책은 \"", recommend_1_book$title, " \"입니다.")

}
```

책을 읽을 확률을 예측; 이용자 편의 목적 알고리즘.





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

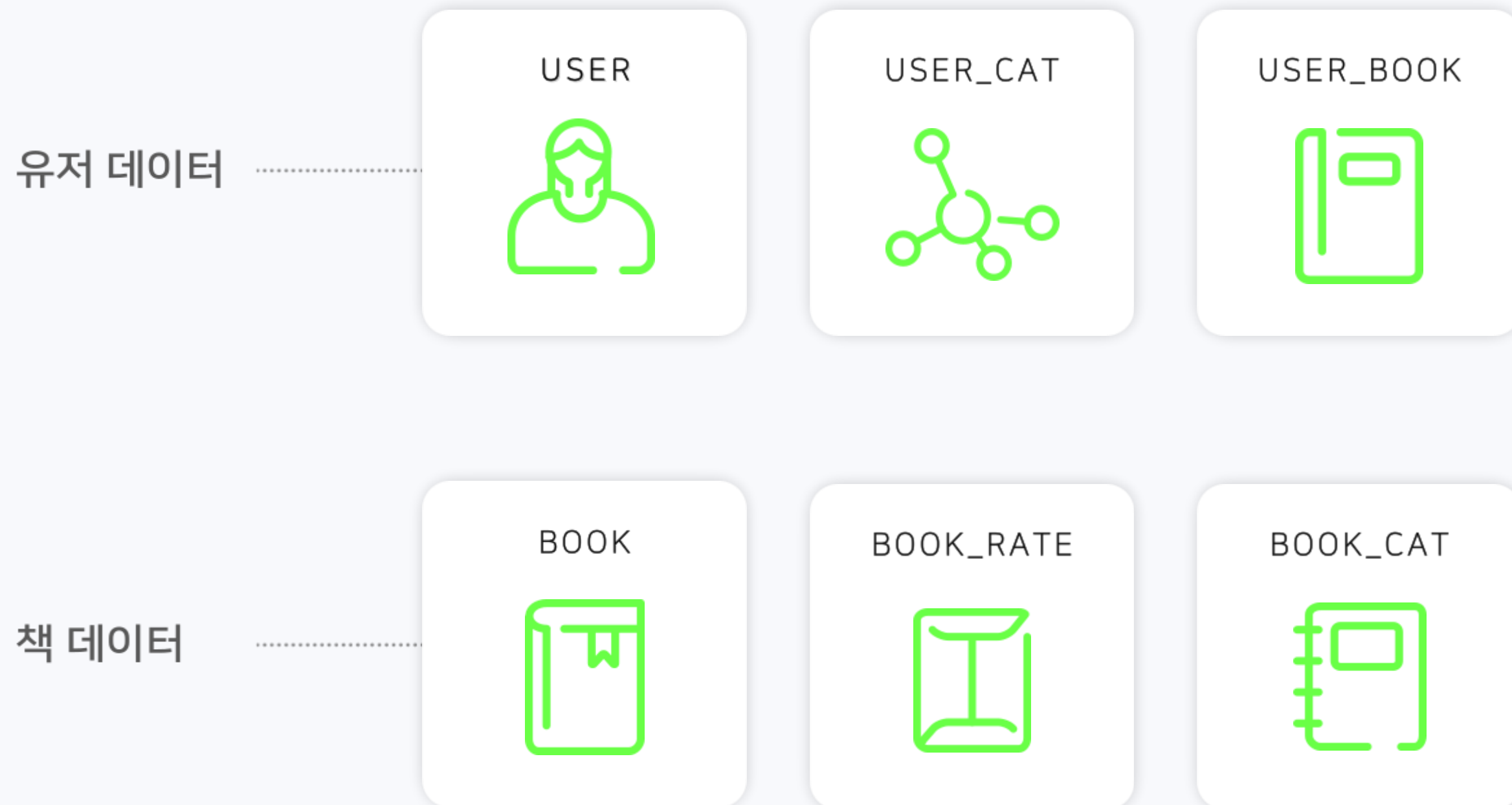
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

궁극적 목표





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

궁극적 목표

USER_CAT



USER_BOOK



BOOK_RATE



BOOK_CAT



해당 책을
읽을 확률을 계산하는
알고리즘



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

데이터 선택

```
4 # 사용할 데이터인 user, user_cat, user_book 구성 확인
5 str(user)
6 str(user_cat)
7 str(user_book)
```

User 데이터 가공

```
14 # user data 중 필요한 것만 추출 후 직업에 대한 NA값 제거
15 user %>% select(user_id, gender, 직업) %>% filter(!is.na(직업)) -> user1
```

```
18 # 여러 직업들을 유사한 직업군으로 대분류
19 user1 <- user1 %>%
20   mutate(직업 = ifelse(직업 %in% c("직장인", "프리랜서", "자영업"), "직장인",
21     ifelse(직업 %in% c("초/중/고등학생", "대학/대학원생"), "학생", "기타"))))
```




데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

User 데이터 가공

```
23 # 직업 및 성별 더미화
24 user1 <- user1 %>%
25   mutate(직업 = ifelse(직업 %in% c("직장인"), 1,
26                         ifelse(직업 %in% c("학생"), 2, 3)))
27 user1 <- user1 %>%
28   mutate(gender = ifelse(gender %in% c("M"), 1, 2))
29 user1
```

```
32 # user_cat data 속 다양한 선호 카테고리들을 유사한 카테고리로 대분류
33 user_cat <- user_cat %>%
34   mutate(title = ifelse(title %in% c("언어", "예술", "디자인", "영어", "음악"), "예술",
35                                   ifelse(title %in% c("에세이", "시", "인문학", "문학"), "리터럴",
36                                           ifelse(title %in% c("습관", "다이어트", "건강", "자기계발",
37                                                             "목표달성", "취미", "운동/레저", "여행",
38                                                             "결혼", "반려동물", "육아", "요리"), "데일리",
39                                           ifelse(title %in% c("부동산", "경제/경영", "재테크", "마케팅",
40                                                             "커뮤니케이션", "인공지능", "컴퓨터/과학",
41                                                             "데이터분석", "교육", "사회/정치"), "전문", "기타"))))))
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

User 데이터 가공

```
45 # 카테고리 중복되는 값 제거
46 user_cat <- distinct(user_cat)
47 user_cat
```

```
49 # 카테고리 더미화
50 user_cat <- user_cat %>%
51   mutate(title = ifelse(title %in% c("예술"), 1,
52                           ifelse(title %in% c("리터럴"), 2,
53                                   ifelse(title %in% c("데일리"), 3,
54                                           ifelse(title %in% c("전문"), 4, 5))))))
55
56 user_cat <- rename(user_cat, cat=title)
57 user_cat
```



데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

Book 데이터 가공

```
67 # 특정 책을 선택한 유저들의 독서 상태를 각각 카운트
68 library(plyr)
69 read_state_count <- ddply(user_book, c("book_id"),
70                             summarize,rsbc =paste(length(which(read_status == "READ_STATUS_BEFORE"))),
71                             rsic =paste(length(which(read_status == "READ_STATUS_ING"))),
72                             rsdc =paste(length(which(read_status == "READ_STATUS_DONE"))),
73                             rssc =paste(length(which(read_status == "READ_STATUS_STOP"))),
74                             rspc =paste(length(which(read_status == "READ_STATUS_PAUSE"))))
75 str(read_state_count)
```

```
77 # numeric으로 변환
78 read_state_count$rsbc<- as.numeric(read_state_count$rsbc)
79 read_state_count$rsic<- as.numeric(read_state_count$rsic)
80 read_state_count$rsdc<- as.numeric(read_state_count$rsdc)
81 read_state_count$rssc<- as.numeric(read_state_count$rssc)
82 read_state_count$rspc<- as.numeric(read_state_count$rspc)
83 str(read_state_count)
```




데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

Book 데이터 가공

```
85 # 책 자체에 대한 완독률
86 read_state_count%>% mutate(done_rate = (rsdc / (rsbc + rsic + rsdc + rssc + rspc))*100) -> read_done_count
87
88 read_done_count %>% select(book_id, done_rate) -> read_done_rate
89
90 head(read_done_rate)
```

```
93 # 책 자체의 평점 결측값 제거
94 user_book %>% select(book_id, rate) %>% filter(!is.na(rate)) -> user_book2
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

Book 데이터 가공

```
96  # 책 자체에 대한 평균 평점
97  detach("package:plyr", unload=TRUE)
98  user_book2 %>%
99    group_by(book_id) %>%
100    summarise(mean_rate = mean(rate)) -> book_rate
101  book_rate
102
103  user1
104  user_book1
105  head(read_done_rate)
106  book_rate
107  user_cat
```




데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

예측할 값을 위한 데이터 가공

```
59 # 어떤 책에 대한 유저의 독서 상태를 분류 후 더미화
60 user_book %>% select(user_id, book_id, read_status) -> user_book1
61 user_book1 <- user_book1 %>%
62   mutate(read_status = ifelse(read_status %in% c("READ_STATUS_DONE", "READ_STATUS_ING"), 1, 0))
63
64 user_book1
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

예측에 사용될 데이터 프레임 제작

```
109 # 유저와 관련된 데이터 모두 병합
110 U0 <- merge(user1, user_book1, by="user_id")
111 U1 <- merge(U0, user_cat, by="user_id")
112 U1 <- as_tibble(U1)
113 U1 %>% relocate(c(gender, 직업), .after = book_id)-> U1
114 U1
```

```
116 # 책과 관련된 데이터 모두 병합
117 B1 <- merge(read_done_rate, book_rate, by="book_id")
118 B1 <- as_tibble(B1)
119 B1
```

```
121 # 유저 데이터, 북 데이터 병합
122 Q1 <- merge(U1, B1)
123 Q1 <- as_tibble(Q1)
124 Q1 %>% relocate(c(done_rate, mean_rate), .after = user_id)-> Q1
125 Q1 %>% relocate(c(book_id), .after = user_id)-> Q1
126 Q1 <- Q1[order(Q1$user_id), ]
127 Q1
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

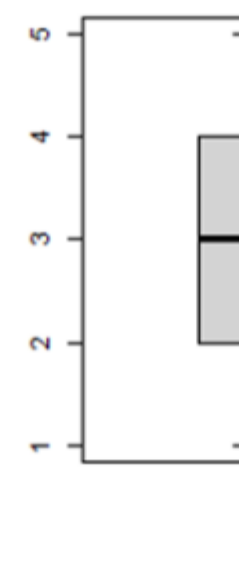
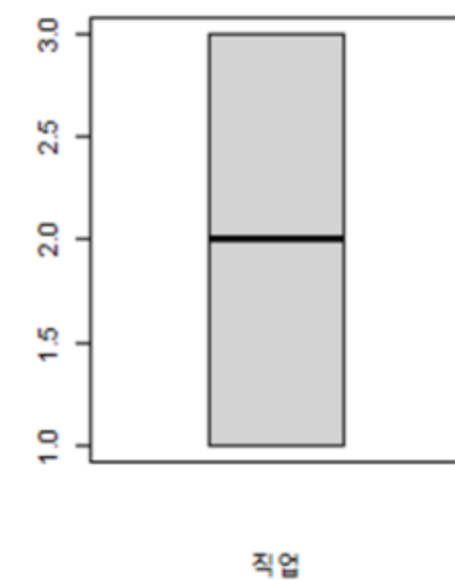
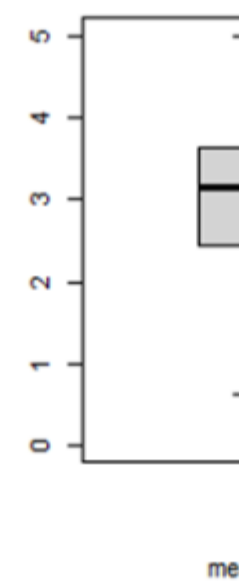
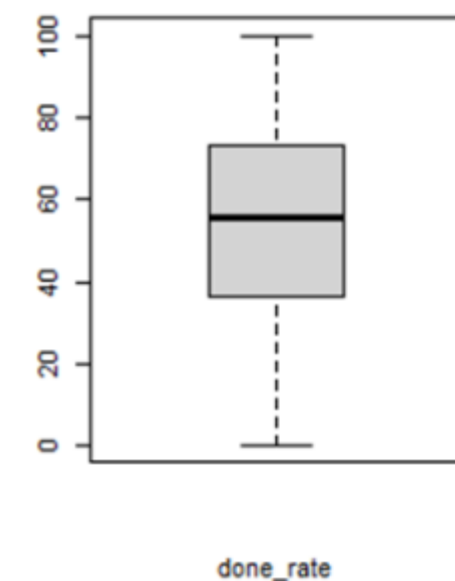
이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

데이터 확인

```
129 # 각 변수의 분포도 확인
130 par(mfrow=c(2,3))
131 boxplot(Q1$done_rate)
132 boxplot(Q1$mean_rate)
133 boxplot(Q1$gender)
134 boxplot(Q1$직업)
135 boxplot(Q1$cat)
```





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

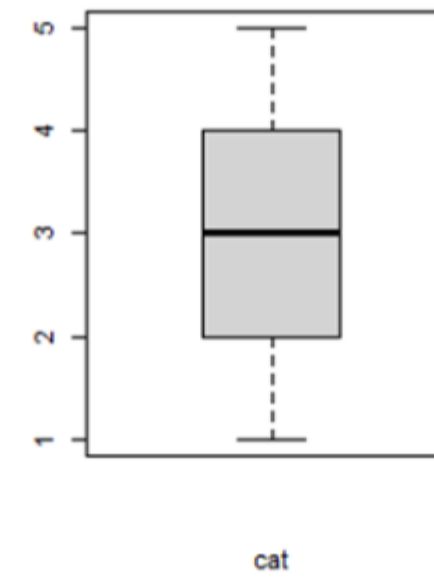
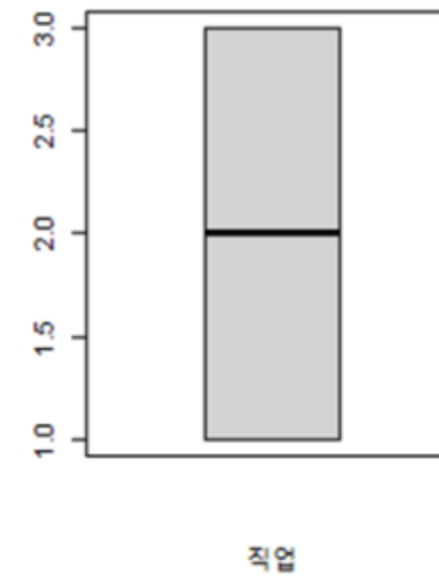
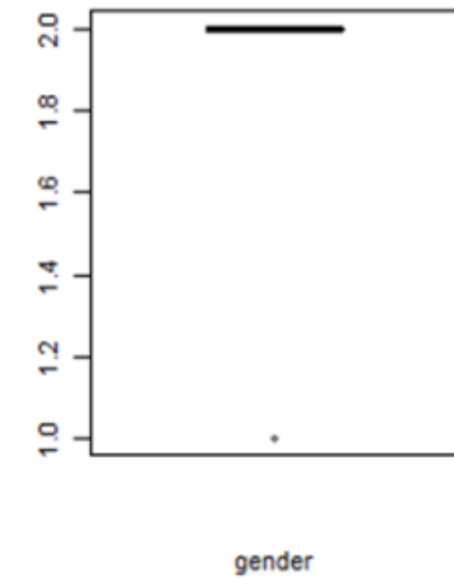
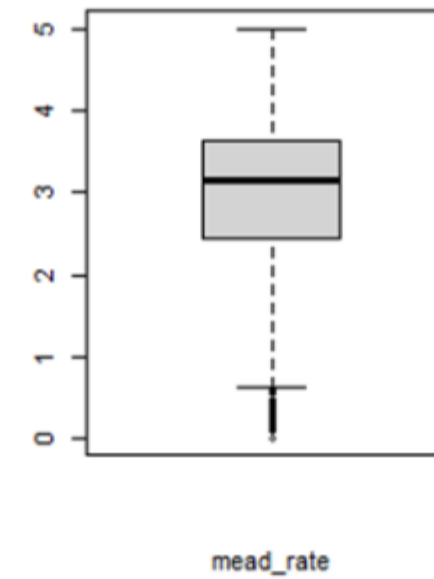
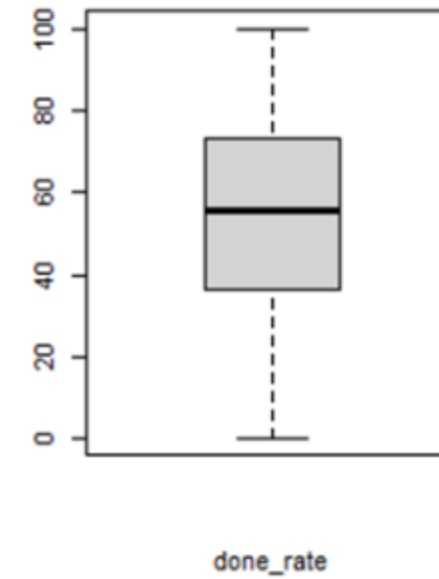
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

데이터 확인





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

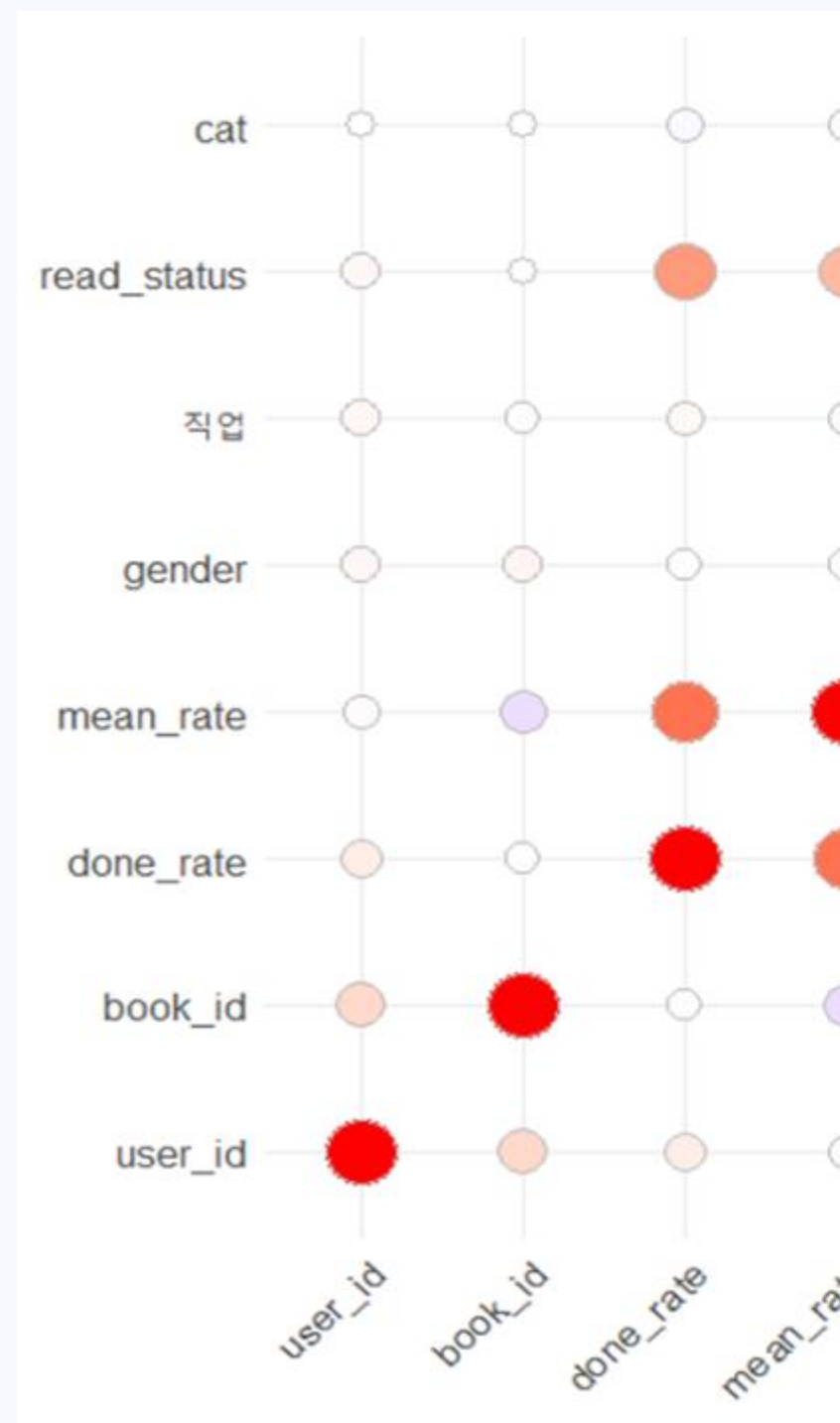
이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

데이터 확인

```
137 # 변수 각각의 상관 관계 확인
138 corr <- round(cor(Q1), 2)
139 corr
140
141 library(ggcorrplot)
142 ggcorrplot(corr, method = "circle")
```





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

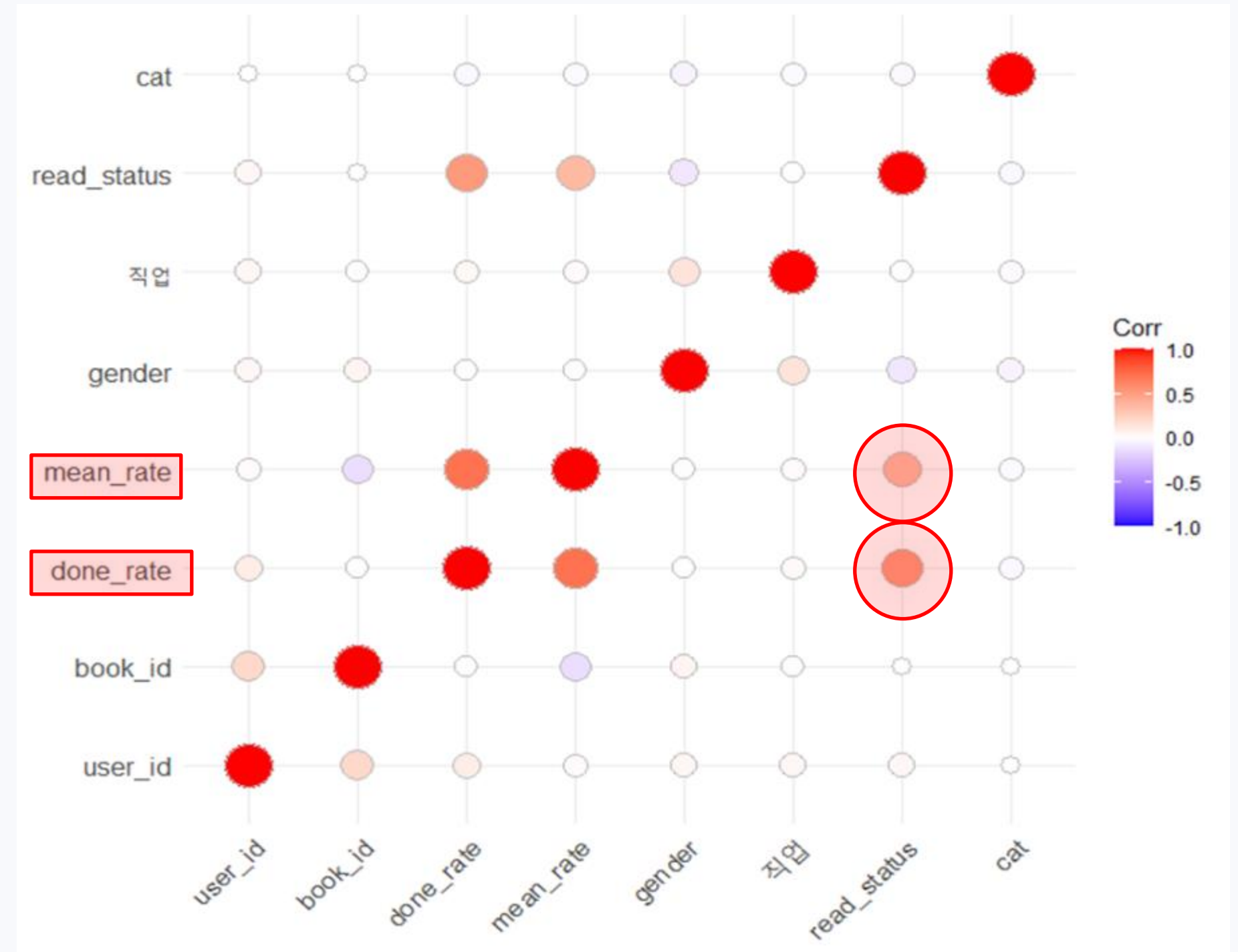
이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

데이터 확인

rcle")





데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

```
145 #모델 신뢰도 확인
146 linear.model<-glm("read_status~.", data=Q1)
147 summary(linear.model)
```



```
> summary(linear.model)

Call:
glm(formula = "read_status~.", data = Q1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max 
-1.07136  -0.36598   0.02397   0.35725   1.07136 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.726e-01  1.909e-03  142.771  <.0001
user_id      -1.248e-12  2.119e-12   -0.589    0.553
book_id       7.138e-09  5.427e-10   13.134  <.0001
done_rate     9.455e-03  1.686e-05  560.841  <.0001
mean_rate     2.491e-03  4.104e-04   6.069  <.0001
gender       -1.098e-01  7.765e-04 -141.341  <.0001
직업         -9.932e-03  3.813e-04 -26.041  <.0001
cat          -7.928e-03  2.600e-04 -30.481  <.0001
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.0000000)

    Null deviance: 418690  on 1710586  degrees of freedom
Residual deviance: 301312  on 1710579  degrees of freedom
AIC: 1884110

Number of Fisher Scoring iterations: 2
```



데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

ta=Q1)



```
> summary(linear.model)

Call:
glm(formula = "read_status~.", data = Q1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.07136  -0.36598   0.02397   0.35725   1.01672

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.726e-01  1.909e-03  142.752  < 2e-16 ***
user_id      -1.248e-12  2.119e-12   -0.589    0.556
book_id       7.138e-09  5.427e-10   13.153  < 2e-16 ***
done_rate     9.455e-03  1.686e-05  560.828  < 2e-16 ***
mean_rate     2.491e-03  4.104e-04   6.069  1.28e-09 ***
gender       -1.098e-01  7.765e-04 -141.383  < 2e-16 ***
직업         -9.932e-03  3.813e-04  -26.048  < 2e-16 ***
cat          -7.928e-03  2.600e-04  -30.492  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.1761459)

Null deviance: 418690  on 1710586  degrees of freedom
Residual deviance: 301312  on 1710579  degrees of freedom
AIC: 1884110

Number of Fisher Scoring iterations: 2
```



데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

```
149 logit.model<-glm("read_status~.", data=Q1,  
150                   family="binomial")  
151 summary(logit.model)
```



```
> summary(logit.model)  
  
Call:  
glm(formula = "read_status~.", family =  
  
Deviance Residuals:  
    Min       1Q   Median       3Q      2.5  
-2.5632  -0.8855   0.3723   0.8638   2.5632  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)    <---> <--->  
(Intercept) -1.094e+00  1.120e-02  -97.148  <2e-16 ***  
user_id      -3.832e-11  1.201e-11   -3.185  0.0015 ***  
book_id       5.896e-08  3.082e-09   19.134  <2e-16 ***  
done_rate     5.240e-02  1.134e-04  462.148  <2e-16 ***  
mean_rate     1.457e-02  2.492e-03   5.847  5.5e-06 ***  
gender       -7.007e-01  4.631e-03 -151.248  <2e-16 ***  
직업         -5.494e-02  2.169e-03 -25.338  <2e-16 ***  
cat          -4.299e-02  1.485e-03 -28.950  <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family using deviance)  
  
Null deviance: 2335422  on 1710586  degrees of freedom  
Residual deviance: 1782288  on 1710579  degrees of freedom  
AIC: 1782304  
  
Number of Fisher Scoring iterations: 4
```




데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

ta=Q1)



```
> summary(logit.model)

Call:
glm(formula = "read_status~.", family = "binomial", data = Q1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5632  -0.8855   0.3723   0.8638   2.4243

Coefficients:
              Estimate Std. Error  z value Pr(>|z|)
(Intercept) -1.094e+00  1.120e-02  -97.679  < 2e-16 ***
user_id      -3.832e-11  1.201e-11   -3.191  0.00142 **
book_id       5.896e-08  3.082e-09   19.132  < 2e-16 ***
done_rate     5.240e-02  1.134e-04  462.041  < 2e-16 ***
mean_rate     1.457e-02  2.492e-03    5.848  4.98e-09 ***
gender       -7.007e-01  4.631e-03 -151.309  < 2e-16 ***
직업         -5.494e-02  2.169e-03  -25.328  < 2e-16 ***
cat          -4.299e-02  1.485e-03  -28.954  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2335422  on 1710586  degrees of freedom
Residual deviance: 1782288  on 1710579  degrees of freedom
AIC: 1782304

Number of Fisher Scoring iterations: 4
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

```
152 exp(logit.model$coefficients)
153 exp(cbind(OR = coef(logit.model), confint(logit.model)))
```



```
> exp(logit.model$coefficients)
(Intercept)      user_id      book_id  done_rate  mean_rate      gender      직업      cat
  0.3348132    1.0000000    1.0000001    1.0537958    1.0146788    0.4962395    0.9465376    0.9579170
> exp(cbind(OR = coef(logit.model), confint(logit.model)))
Waiting for profiling to be done...
              OR      2.5 %      97.5 %
(Intercept) 0.3348132 0.3275413 0.3422446
user_id      1.0000000 1.0000000 1.0000000
book_id      1.0000001 1.0000001 1.0000001
done_rate    1.0537958 1.0535617 1.0540302
mean_rate    1.0146788 1.0097347 1.0196462
gender       0.4962395 0.4917545 0.5007630
직업         0.9465376 0.9425216 0.9505707
cat          0.9579170 0.9551330 0.9607089
```



데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

모델 확인

```
155 #예측값 적용(유저 임의 지정)
156 newdata<-data.frame(book_id=c(1,2,3), user_id=c(1,2,3),
157                      done_rate=c(50,100,30), mean_rate=c(4,5,1),
158                      gender=c(1,1,2), 직업=c(1,1,3), cat=c(1,2,3))
159 newdata
160 predict(linear.model, newdata)
161 predict(logit.model, newdata, type = "response")
```

```
> newdata
  book_id user_id done_rate mean_rate gender 직업 cat
1       1       1       50         4       1     1    1
2       2       2      100         5       1     1    2
3       3       3       30         1       2     3    3
> predict(linear.model, newdata)
      1      2      3
0.6276171 1.0949287 0.2855352
> predict(logit.model, newdata, type = "response")
      1      2      3
0.6868434 0.9669752 0.2309668
```





데이터예측모델과
기계학습의응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

적용

```
163  #예측값 적용(유저 무작위 배정)
164  newdata1<-data.frame(book_id=sample(x=Q1$book_id, size=20, replace=F),
165                        user_id=sample(x=Q1$user_id, size=20, replace=F),
166                        done_rate=sample(x=Q1$done_rate, size=20, replace=T),
167                        mean_rate=sample(x=Q1$mean_rate, size=20, replace=T),
168                        gender=sample(x=Q1$gender, size=20, replace=T),
169                        직업=sample(x=Q1$직업, size=20, replace=T),
170                        cat=sample(x=Q1$cat, size=20, replace=T))
171  newdata1
172  predict(linear.model, newdata1)
173  predict(logit.model, newdata1, type = "response")
174  predict(logit.model, newdata1, type = "response") -> pr_1
```

```
176  newdata2<-data.frame(book_id=sample(x=Q1$book_id, size=20, replace=F),
177                        user_id=sample(x=Q1$user_id, size=20, replace=F),
178                        done_rate=sample(x=Q1$done_rate, size=20, replace=T),
179                        mean_rate=sample(x=Q1$mean_rate, size=20, replace=T),
180                        gender=sample(x=Q1$gender, size=20, replace=T),
181                        직업=sample(x=Q1$직업, size=20, replace=T),
182                        cat=sample(x=Q1$cat, size=20, replace=T))
183  newdata2
184  predict(linear.model, newdata2)
185  predict(logit.model, newdata2, type = "response")
186  predict(logit.model, newdata2, type = "response") -> pr_2
```




데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

```
188 newdata3<-data.frame(book_id=sample(x=Q1$book_id, size=20, replace=F),
189                       user_id=sample(x=Q1$user_id, size=20, replace=F),
190                       done_rate=sample(x=Q1$done_rate, size=20, replace=T),
191                       mean_rate=sample(x=Q1$mean_rate, size=20, replace=T),
192                       gender=sample(x=Q1$gender, size=20, replace=T),
193                       직업=sample(x=Q1$직업, size=20, replace=T),
194                       cat=sample(x=Q1$cat, size=20, replace=T))
195 newdata3
196 predict(linear.model, newdata3)
197 predict(logit.model, newdata3, type = "response")
198 predict(logit.model, newdata3, type = "response") -> pr_3
```



데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

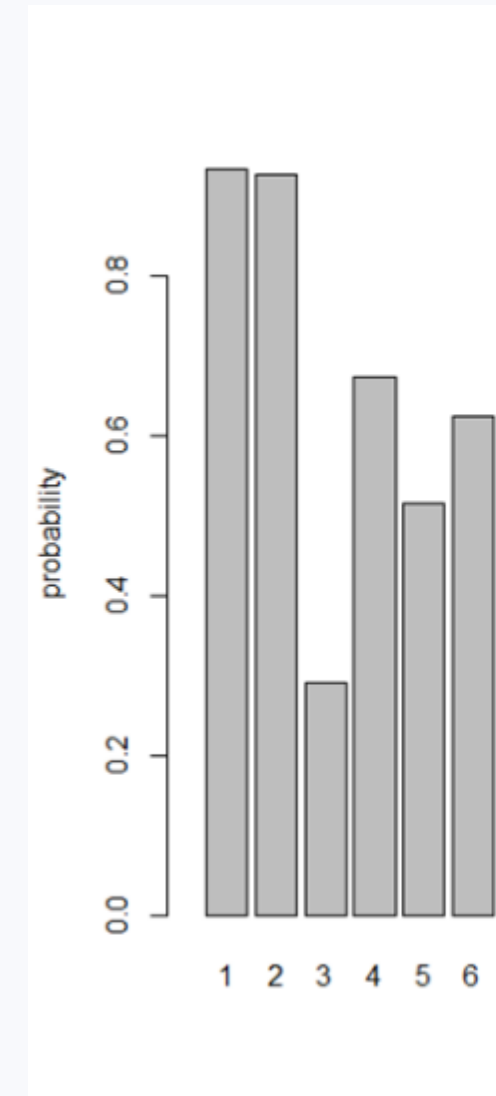
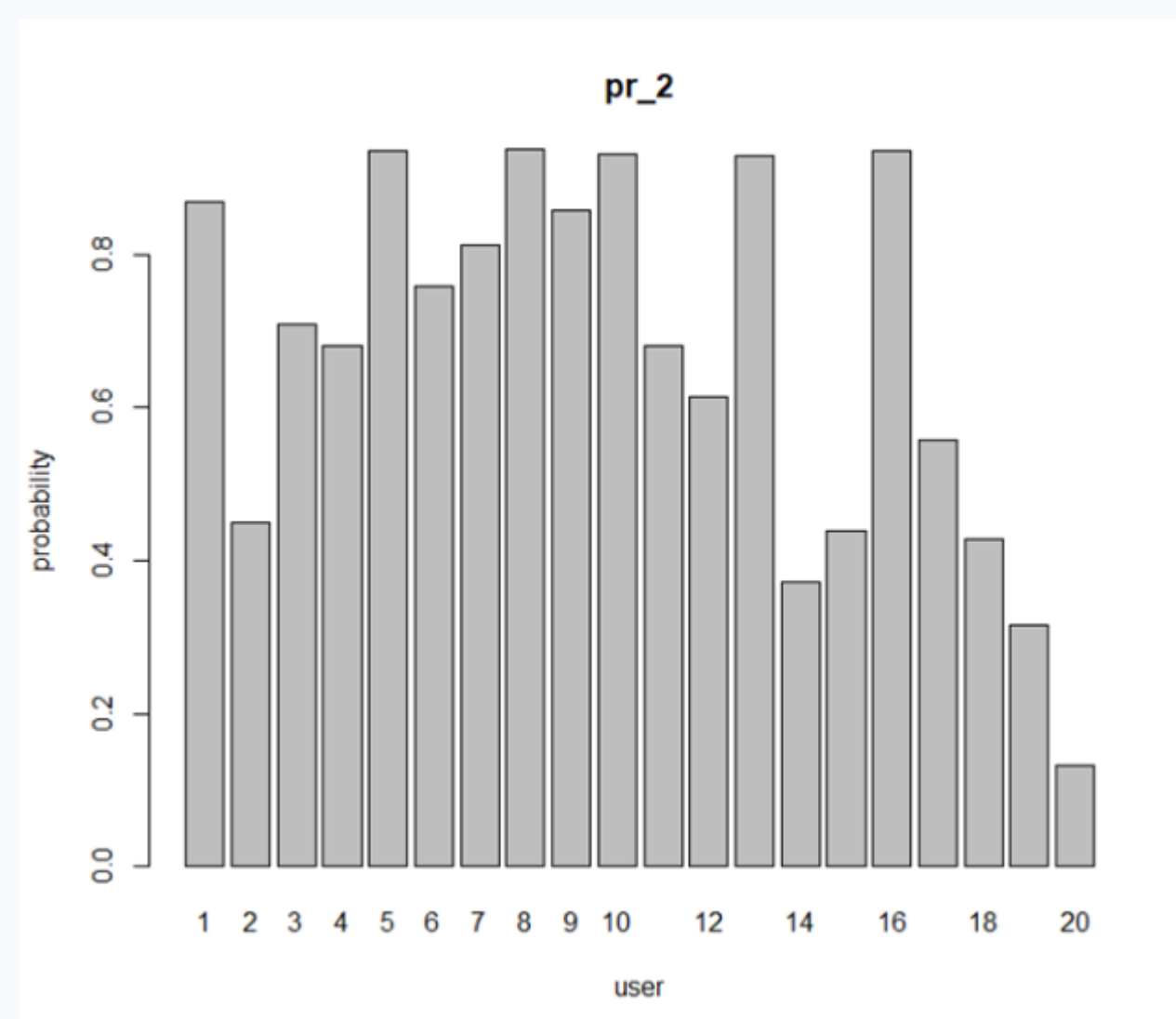
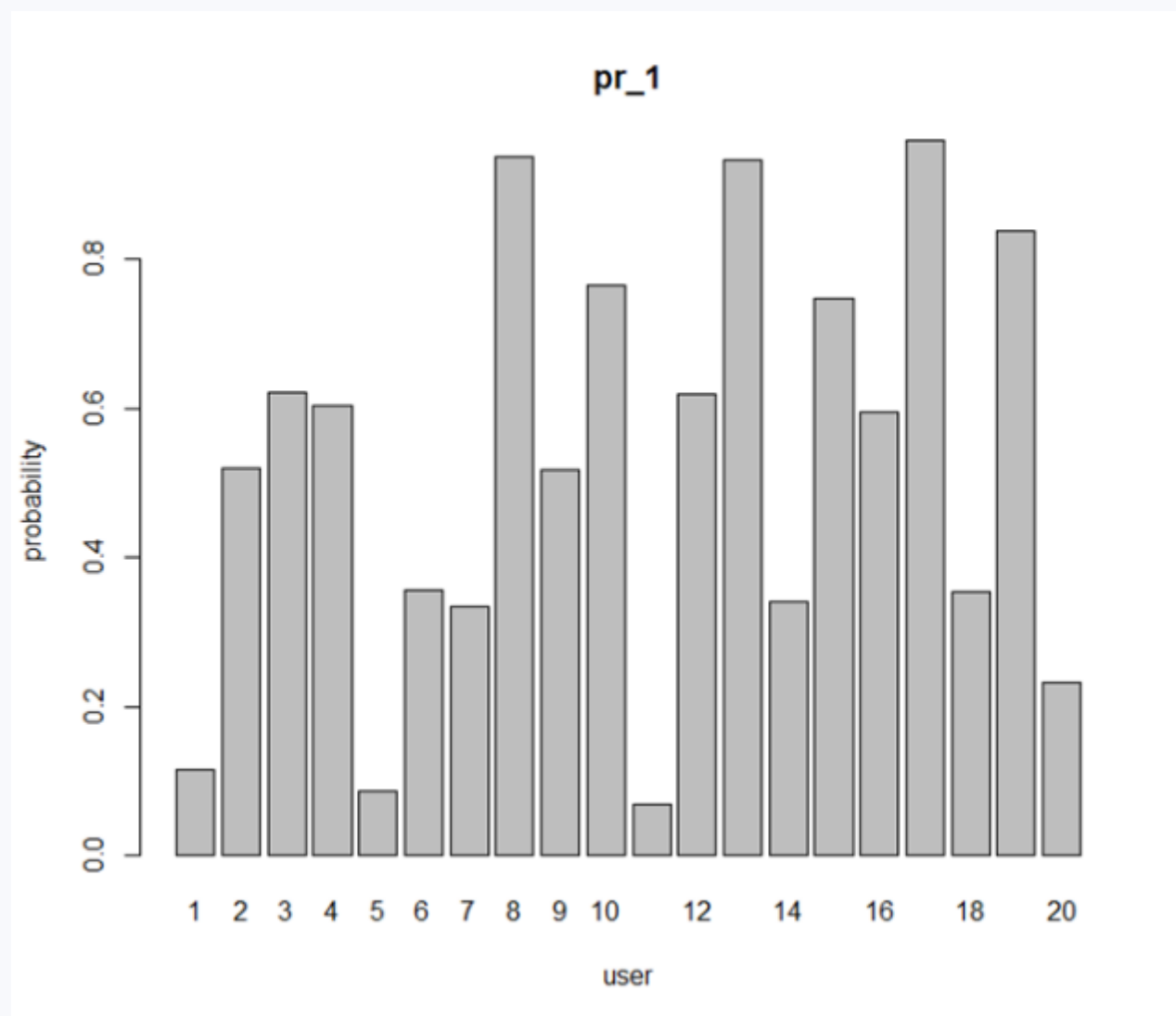
—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

시각화





데이터 예측 모델과
기계 학습의 응용

서론

본론 1
개발자 편의 목적
알고리즘

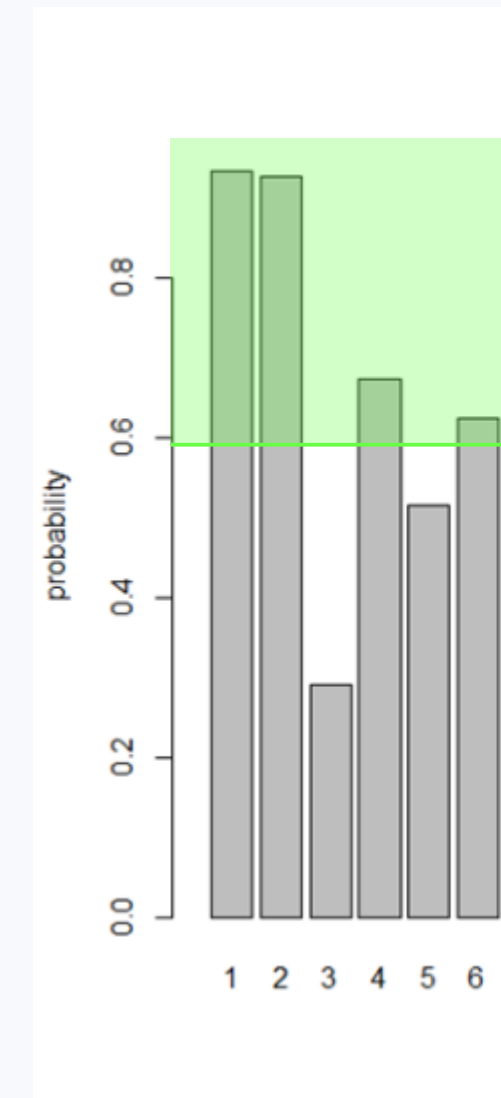
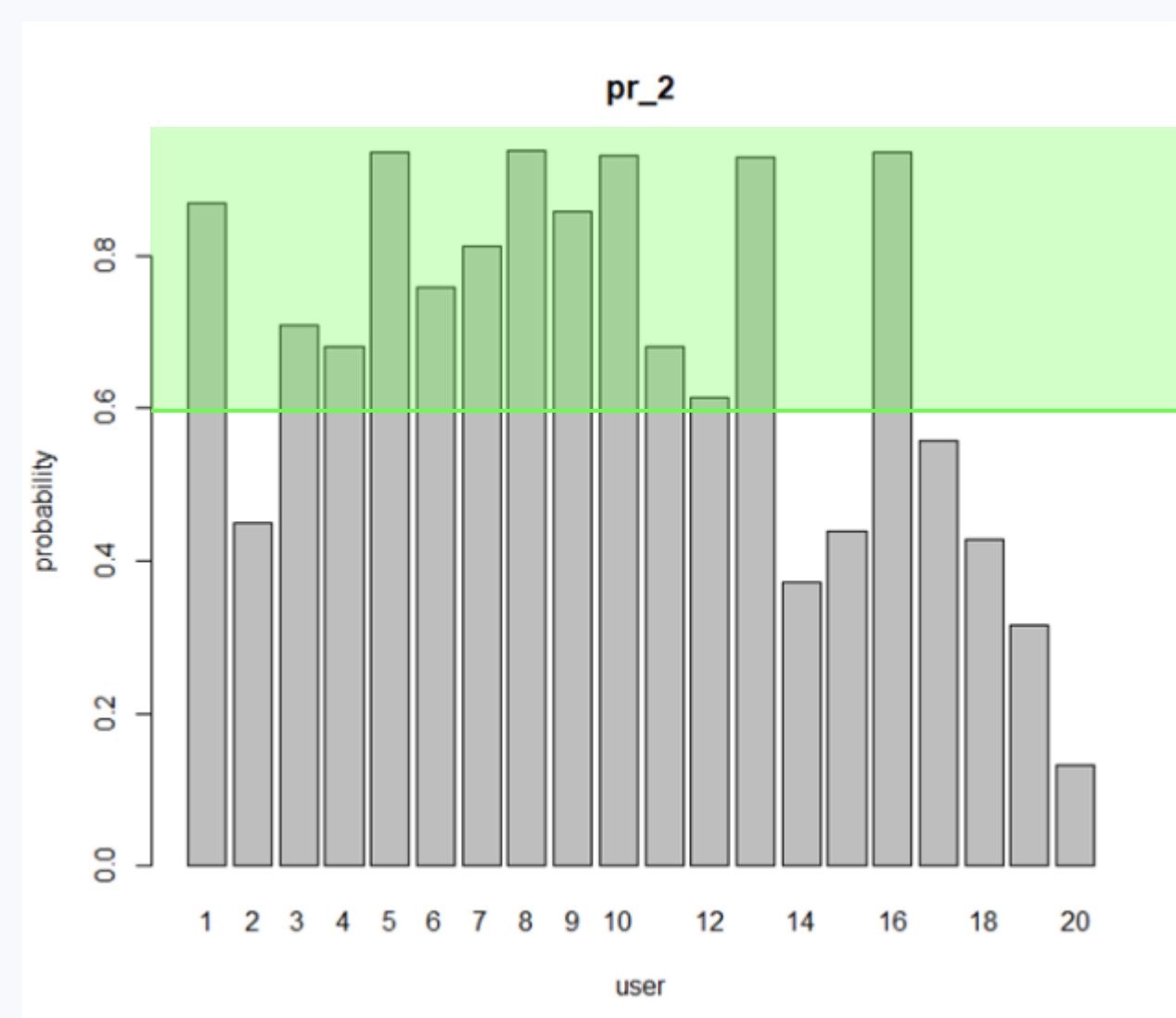
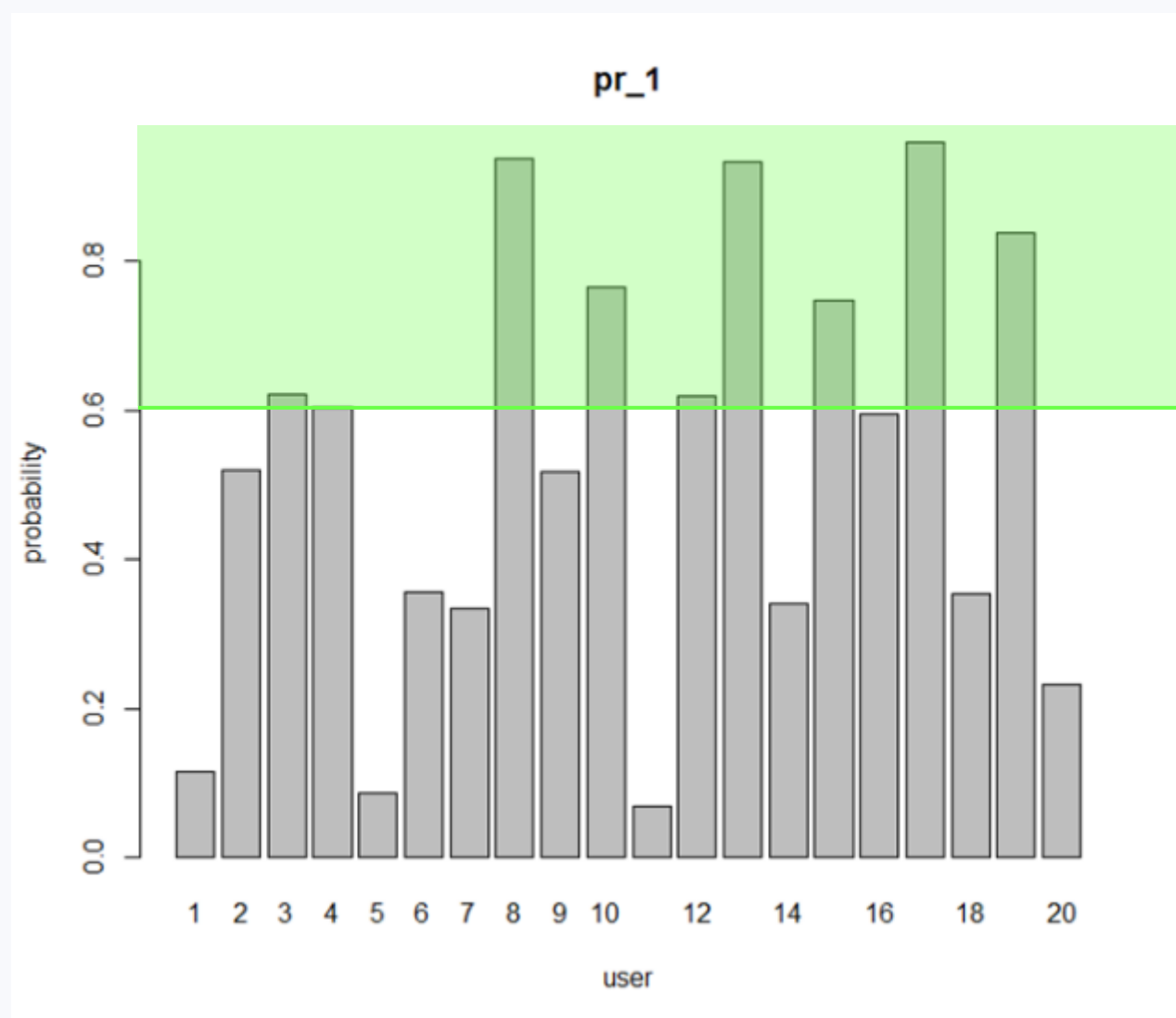
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향

이용자 편의 목적 알고리즘.

시각화





데이터 예측 모델과
기계 학습의 응용

이용자 편의 목적 알고리즘.

시각화

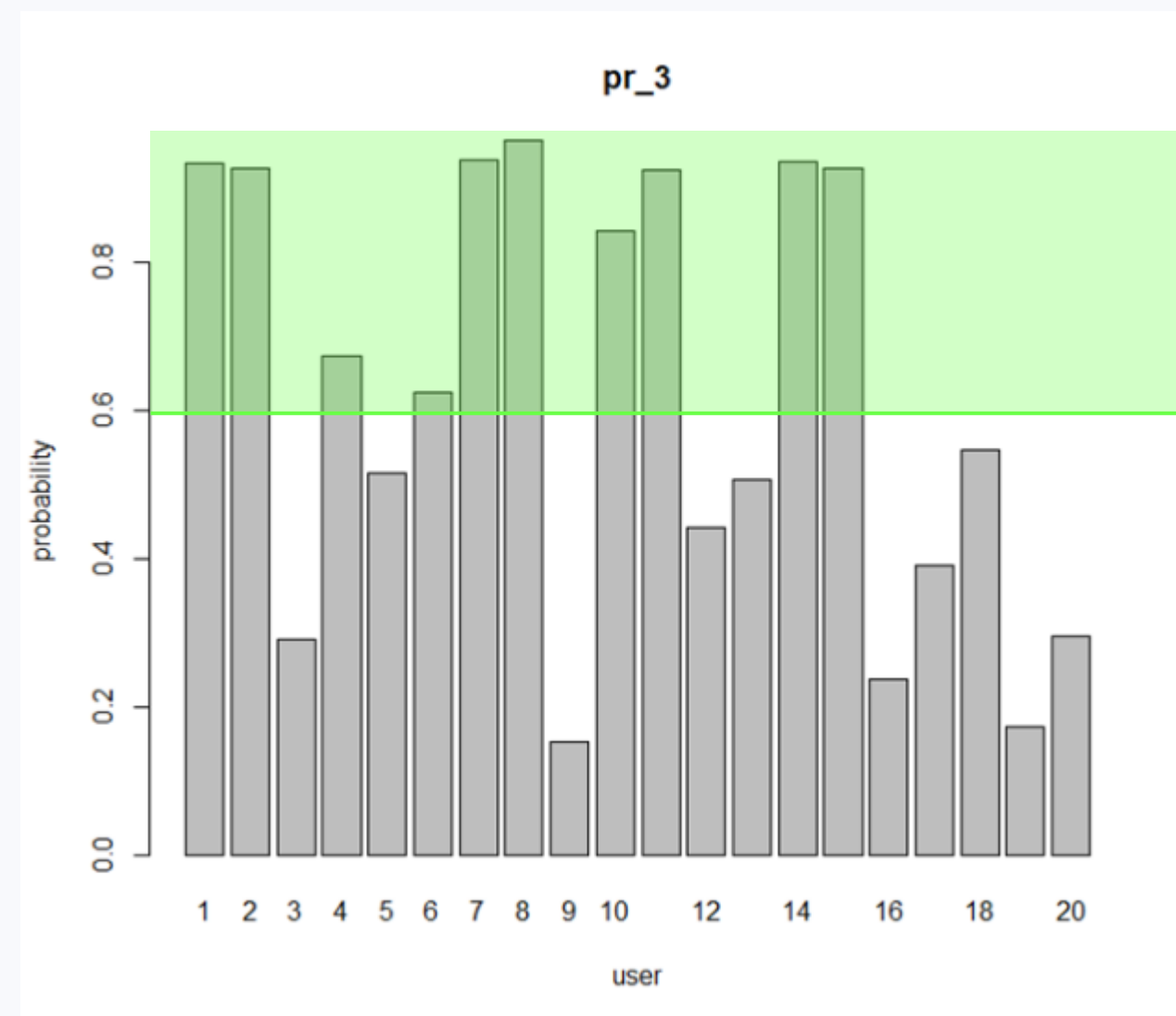
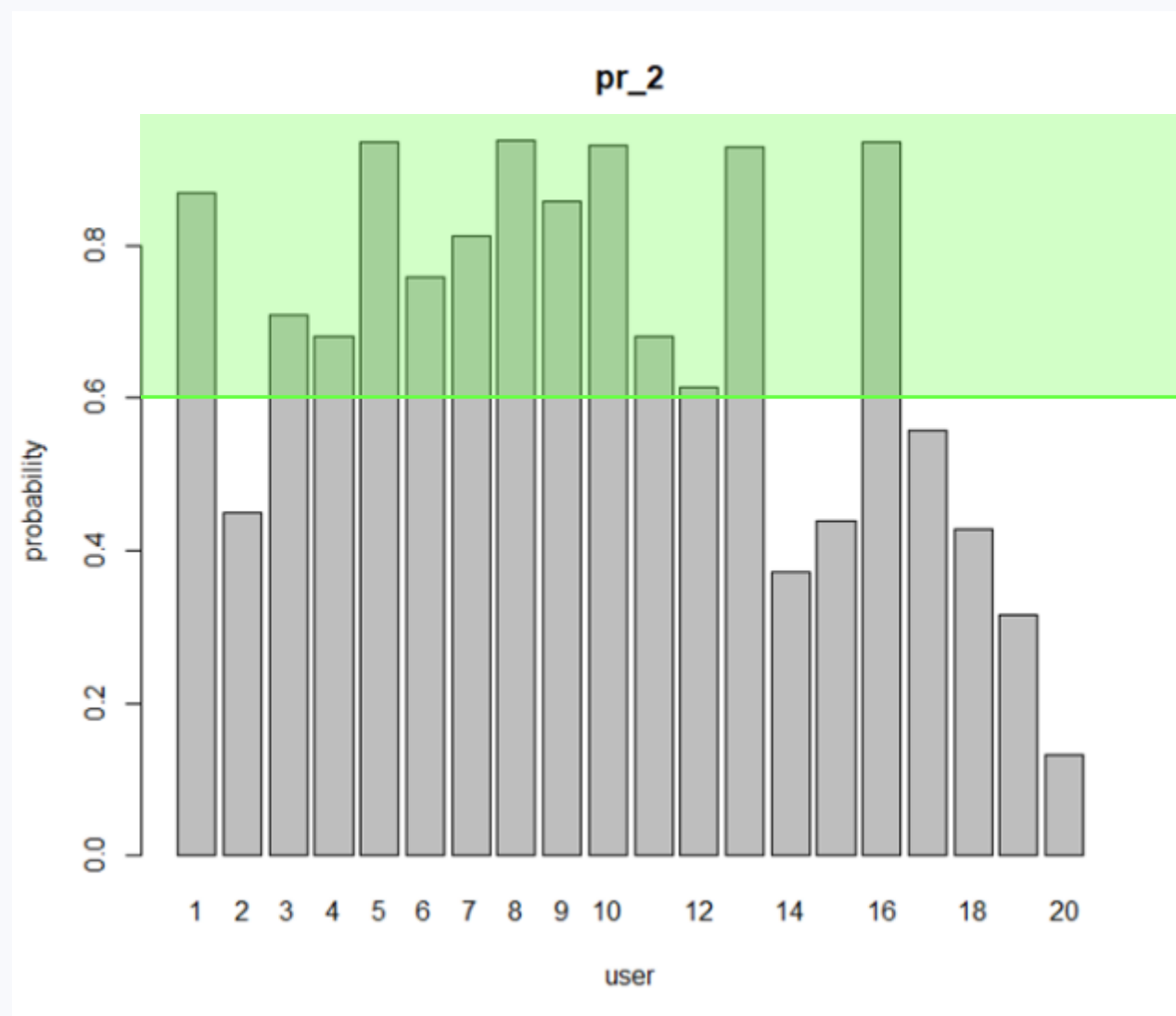
서론

본론 1
개발자 편의 목적
알고리즘

—
본론 2

이용자 편의 목적
알고리즘

결론
기대 방향



기대 효과.





데이터예측모델과
기계학습의응용

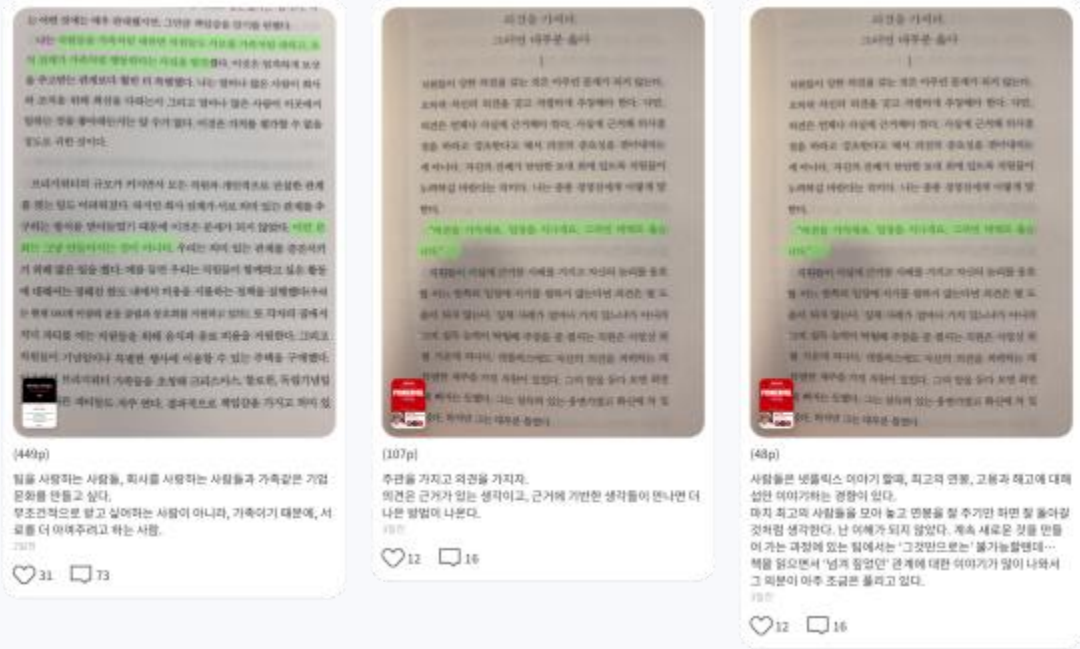
서론

본론 1
개발자 편의 목적
알고리즘

본론 2
이용자 편의 목적
알고리즘

—
결론
기대 방향

기대 효과



개별 추천 알고리즘.

잠재적 액티브 유저를
액티브 유저로 귀속시켜
리더스 어플의 활성화

책을 읽을 확률 예측.

추천 책에 대한 유저의 책을 읽을 확률을 예측해
리더스 어플 유저들에게 해당 책을 추천했을 때
유저들의 만족감 증폭



감사합니다.

을 만든 사람들



윤영훈

IRONMAN



133



78



박영선

youngdaeng



0



27

< 가자, 어디에도 없었던 방법으로



jwjin0330

2019.06.02(일) 11:29



0



4

측정할 수 없는 것

사실 중간까지만 읽고 그만둘 수 없어 하루만에 다 읽
어버렸다. 그 만큼 소설감이 심하게 일었다. 그리고 나책

<



leemiranza

2019.06.01(토) 19:08:16

저도 유년시절의 이야기를 가장 흥
교육 방식을 통해 어떤 가치관을 가
용이 좋더라구요. :)

신고



jwjin0330

2019.06.02(일) 11:21:00