

Зээлийн эрсдэлийн шинжилгээ

2025 оны 11-р сарын 20

Энэхүү ажлаар зээлийн эрсдэлийг UCI Machine Learning Repository-aас татаж авсан German Credit dataset дээрх өгөгдөл тулгуурлан судалж, зээлдэгчдийн шинж чанарууд зээлийн эрсдэлтэй хэрхэн холбогдож байгааг статистик арга зүйгээр шинжилнэ. Мөн машин сургалтын загварууд ашиглан зээлийн эрсдэлийг урьдчилан таамаглах боломжийг судална.

Агуулга

1	Өгөгдөл	2
2	Зээлийн эрсдэлийн үзүүлэлтүүд	3
2.1	Санамсаргүй хувьсагч ба тархалтын үндсэн үзүүлэлтүүд	3
3	Хамтын тархалт ба санамсаргүй хувьсагчдын хамаарал	5
3.1	Хамтын тархалт ба корреляцийн коэффициент	5
4	Бернулийн процесс ба зээлийн эрсдэл	8
4.1	Бернулийн процессын онол	8
5	Тархалтын параметрийн статистик үнэлэлт	8
5.1	Моментын арга	8
6	Зээлийн эрсдэлийн загварчлал	8
6.1	Ложистик регресс	8
6.2	Статистик таамаглал шалгах	12
6.2.1	Хэвийн тархалтын шалгалт	12
6.2.2	Параметрийн тухай таамаглал шалгах	12
6.3	Гэнэн Байесын алгоритм	13
6.4	Шийдвэрийн мод (Decision Tree)	17
7	Загваруудын харьцуулалт	20
7.1	ROC Curve ба Precision-Recall Curve	21
7.1.1	ROC Curve (Receiver Operating Characteristic Curve)	21
8	Зээлийн эрсдэлийн прогноз	24
	Дүгнэлт	26
8.1	Судалгааны үндсэн үр дүн	26
8.1.1	1. Өгөгдлийн шинжилгээ	26
8.1.2	2. Хувьсагчдын хоорондын холбоо	26

8.1.3	3. Загваруудын гүйцэтгэл	27
8.1.4	4. Практик ач холбогдол	27
8.1.5	5. Зөвлөмж	28
8.2	Хязгаарлалт ба ирээдүйн судалгаа	29
Багийн гишүүдийн үүрэг, оролцоо		29
8.3	Багийн гишүүдийн үүрэг, оролцоо	29
Ашигласан материал		29

1 Θөгөдөл

Зээлийн эрсдэлийг судлахад UCI Machine Learning Repository-aac “German Credit” dataset [1] ашиглана. Энэхүү dataset нь Германы банкуудын зээлийн эрсдэлийн үнэлгээний өгөгдөл агуулдаг бөгөөд зээлдэгчдийн мэдээлэл, зээлийн төлөв байдал зэргийг багтаасан байна. Dataset нь cross-sectional буюу нэг цаг хугацааны цэг дээрх 1000 зээлдэгчийн мэдээллийг агуулдаг.

UCI Machine Learning Repository нь машин сургалтын судалгаанд ашиглах өгөгдлийн сан бөгөөд олон улсын стандарт өгөгдөл агуулдаг [2]. “German Credit” dataset нь зээлийн эрсдэлтэй холбоотой судалгаанд өргөн ашиглагддаг.

Өгөгдөл татаж авах.

```

if (file.exists("raw_data.Rds")) {
  raw_data <- readRDS(file = "raw_data.Rds")
} else {
  # UCI Machine Learning Repository-aac German Credit dataset татаж авах
  # URL: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)

  data_url <- "https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/german.data"

  # Өгөгдөл татаж авах
  raw_data <- tryCatch({
    # read.table() ашилан татаж авах
    temp_file <- tempfile()
    download.file(data_url, temp_file, quiet = TRUE)

    # German Credit dataset-ийн баганын нэрс
    col_names <- c(
      "checking_status", "duration", "credit_history", "purpose",
      "credit_amount", "savings_status", "employment", "installment_commitment",
      "personal_status", "other_parties", "residence_since", "property_magnitude",
      "age", "other_payment_plans", "housing", "existing_credits",
      "job", "num_dependents", "telephone", "foreign_worker", "class"
    )

    # Өгөгдөл унших
    data_raw <- read.table(
      temp_file,
      sep = " ",
      stringsAsFactors = FALSE,
      col.names = col_names
    )

    # Файл устгах
    unlink(temp_file)

    data_raw
  })
}

```

```

}, error = function(e) {
  # Алдаа гарвал жишээ өгөгдөл үүсгэх
  warning("UCI-аас өгөгдөл татаж чадсангүй. Жишээ өгөгдөл ашиглаж байна.")

  set.seed(123)
  n <- 1000
  data.frame(
    checking_status = sample(c("A11", "A12", "A13", "A14"), n, replace = TRUE),
    duration = round(runif(n, 4, 72)),
    credit_history = sample(c("A30", "A31", "A32", "A33", "A34"), n, replace = TRUE),
    purpose = sample(c("A40", "A41", "A42", "A43", "A44", "A45", "A46", "A47", "A48", "A49", "A410"), n, replace = TRUE),
    credit_amount = round(runif(n, 250, 18424)),
    savings_status = sample(c("A61", "A62", "A63", "A64", "A65"), n, replace = TRUE),
    employment = sample(c("A71", "A72", "A73", "A74", "A75"), n, replace = TRUE),
    installment_commitment = round(runif(n, 1, 4)),
    personal_status = sample(c("A91", "A92", "A93", "A94", "A95"), n, replace = TRUE),
    other_parties = sample(c("A101", "A102", "A103"), n, replace = TRUE),
    residence_since = round(runif(n, 1, 4)),
    property_magnitude = sample(c("A121", "A122", "A123", "A124"), n, replace = TRUE),
    age = round(runif(n, 19, 75)),
    other_payment_plans = sample(c("A141", "A142", "A143"), n, replace = TRUE),
    housing = sample(c("A151", "A152", "A153"), n, replace = TRUE),
    existing_credits = round(runif(n, 1, 4)),
    job = sample(c("A171", "A172", "A173", "A174"), n, replace = TRUE),
    num_dependents = round(runif(n, 1, 2)),
    telephone = sample(c("A191", "A192"), n, replace = TRUE),
    foreign_worker = sample(c("A201", "A202"), n, replace = TRUE),
    class = sample(c(1, 2), n, replace = TRUE, prob = c(0.7, 0.3))
  )
}

# Өгөгдөл хадгалах
saverDS(raw_data, file = "raw_data.Rds")
write.csv(raw_data, file = "raw_data.csv", row.names = FALSE)
}

```

Татаж авсан өгөгдлөө “raw_data.csv” файлд хадгалав.

Анхны dataset - Эрсдэлтэй зээлийн хувь: 30 %

Training set - Эрсдэлтэй зээлийн хувь: 30 %

Test set - Эрсдэлтэй зээлийн хувь: 30 %

Ийнхүү UCI Machine Learning Repository-аас татаж авсан German Credit dataset-ийг ашиглан зээлийн эрсдэлийн талаарх өгөгдлөтэй боллоо. Dataset нь cross-sectional буюу нэг цаг хугацааны цэг дээрх өгөгдөл тул цаг хугацааны хандлага судлах биш, харин зээлдэгчдийн шинж чанарууд зээлийн эрсдэлтэй хэрхэн холбогдож байгааг судлана.

2 Зээлийн эрсдэлийн үзүүлэлтүүд

2.1 Санамсаргүй хувьсагч ба тархалтын үндсэн үзүүлэлтүүд

German Credit dataset-д зээлийн хэмжээ (X_1), хугацаа (X_2), нас (X_3) зэрэг санамсаргүй хувьсагчдыг авч үзнэ. Эрсдэлийн ангилал (Y) нь Бернуллийн санамсаргүй хувьсагч бөгөөд $Y = 1$ (эрсдэлтэй) эсвэл $Y = 0$ (эрсдэлгүй) утга авна.

Санамсаргүй хувьсагчдын математик дундаж (математик хүлээлт) ба дундаж квадрат хазайлт (дисперс) нь тархалтын чухал үзүүлэлтүүд юм. Математик дундаж нь тархалтын төв байрлалыг, дисперс нь тархалтын тархалтыг тодорхойлдог.

UCI Machine Learning Repository-аас татаж авсан German Credit dataset-ийн үндсэн мэдээллийг Хүснэгт 1 дээр харууллаа.

```
# Dataset-ийн үндсэн статистик
dataset_summary <- data.frame(
  "Үзүүлэлт" = c(
    "Нийт зээлдэгчдийн тоо",
    "Эрсдэлтэй зээлийн хувь (%)",
    "Эрсдэлгүй зээлийн хувь (%)",
    "Дундаж зээлийн хэмжээ",
    "Дундаж зээлийн хугацаа (cap)",
    "Дундаж нас"
  ),
  "Утга" = c(
    nrow(raw_data),
    round(mean(raw_data$class == 2) * 100, 2),
    round(mean(raw_data$class == 1) * 100, 2),
    round(mean(raw_data$credit_amount), 0),
    round(mean(raw_data$duration), 1),
    round(mean(raw_data$age), 1)
  )
)

knitr::kable(dataset_summary, col.names = c("Үзүүлэлт", "Утга"))
# Санамсаргүй хувьсагчдын математик дундаж ба дисперс
```

Хүснэгт 1: Зээлийн эрсдэлийн үзүүлэлтүүд (German Credit dataset-ийн үндсэн дээр)

Үзүүлэлт	Утга
Нийт зээлдэгчдийн тоо	1000.0
Эрсдэлтэй зээлийн хувь (%)	30.0
Эрсдэлгүй зээлийн хувь (%)	70.0
Дундаж зээлийн хэмжээ	3271.0
Дундаж зээлийн хугацаа (cap)	20.9
Дундаж нас	35.5

```
moments_table <- data.frame(
  "Хувьсагч" = c("Зээлийн хэмжээ", "Зээлийн хугацаа", "Нас"),
  "Математик дундаж ( $E[X]$ )" = c(
    round(mean(raw_data$credit_amount), 2),
    round(mean(raw_data$duration), 2),
    round(mean(raw_data$age), 2)
  ),
  "Дисперс ( $Var(X)$ )" = c(
    round(var(raw_data$credit_amount), 2),
    round(var(raw_data$duration), 2),
    round(var(raw_data$age), 2)
  ),
  "Стандарт хазайлт ( $\sigma$ )" = c(
    round(sd(raw_data$credit_amount), 2),
    round(sd(raw_data$duration), 2),
    round(sd(raw_data$age), 2)
  )
)
```

```
knitr::kable(moments_table, col.names = c("Хувьсагч", "Математик дундаж (E[X])", "Дисперс (Var(X))", "Стандарт хазайлт (σ)"))
```

Хүснэгт 2: Санамсаргүй хувьсагчдын математик дундаж, дисперс, стандарт хазайлт

Хувьсагч	Математик дундаж (E[X])	Дисперс (Var(X))	Стандарт хазайлт (σ)
Зээлийн хэмжээ	3271.26	7967843.47	2822.74
Зээлийн хугацаа	20.90	145.42	12.06
Нас	35.55	129.40	11.38

Өгөгдлийн тархалтыг гистограммаар дүрсэлснийг Зураг 1 дээр харууллаа.

```
Cairo::CairoPDF(
  file = file.path(output_dir, "fig-distributions.pdf"),
  family = "Times New Roman", width = 8, height = 5
)

par(mfrow = c(2, 2))

# Зээлийн хэмжээний тархалт
hist(raw_data$credit_amount,
  main = "Зээлийн хэмжээний тархалт",
  xlab = "Зээлийн хэмжээ",
  ylab = "Давтамж",
  breaks = 20)

# Зээлийн хугацааны тархалт
hist(raw_data$duration,
  main = "Зээлийн хугацааны тархалт",
  xlab = "Хугацаа (сар)",
  ylab = "Давтамж",
  breaks = 15)

# Насны тархалт
hist(raw_data$age,
  main = "Насны тархалт",
  xlab = "Нас",
  ylab = "Давтамж",
  breaks = 15)

# Эрсдэлийн ангиллын тархалт
barplot(table(raw_data$risk_binary),
  main = "Эрсдэлийн ангиллын тархалт",
  xlab = "Ангилал",
  ylab = "Тоо")

dev.off()

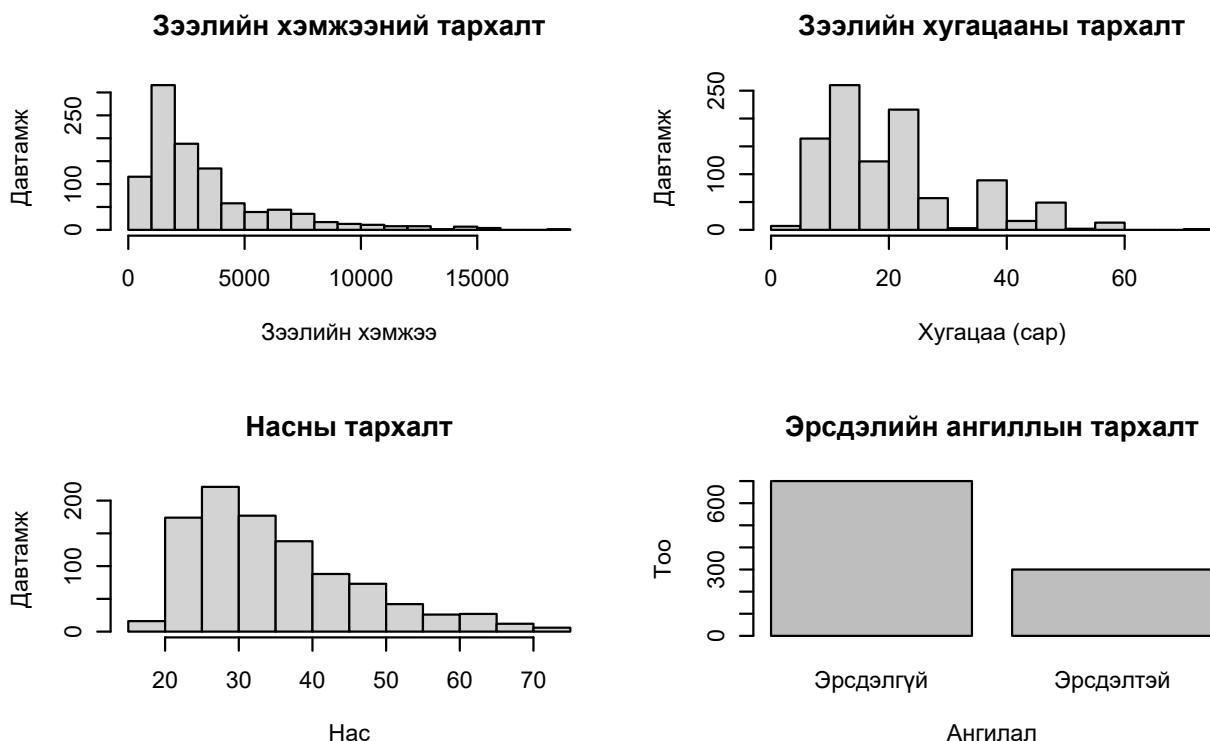
knitr::include_graphics(file.path(output_dir, "fig-distributions.pdf"))
```

3 Хамтын тархалт ба санамсаргүй хувьсагчдын хамаарал

3.1 Хамтын тархалт ба корреляцийн коэффициент

Зээлийн эрсдэл (Y) болон бусад хувьсагчдын (X_1, X_2, X_3) хоорондын хамаарлыг судлахад хамтын тархалт, ковариац, корреляцийн коэффициент чухал үүрэг гүйцэтгэдэг. Корреляцийн коэффициент нь

Зураг 1: Өгөгдлийн тархалтын графикууд



хоёр санамсаргүй хувьсагчийн хоорондын шугаман хамаарлын хүчийг хэмждэг.

Зээлдэгчдийн шинж чанарууд зээлийн эрсдэлтэй хэрхэн холбогдож байгааг судлана. Зээлийн хэмжээ, хугацаа, нас зэрэг хувьсагчдын эрсдэлтэй холбоог корреляцийн шинжилгээгээр шалгана.

```
# Зээлийн хэмжээ ба эрсдэлийн хоорондын корреляц
cor_amount_risk <- cor.test(
  raw_data$credit_amount,
  as.numeric(raw_data$class == 2),
  method = "spearman"
)

cat("Зээлийн хэмжээ ба эрсдэлийн хоорондын корреляц:\n")
```

Зээлийн хэмжээ ба эрсдэлийн хоорондын корреляц:

```
cat("rho =", round(cor_amount_risk$estimate, 3), "\n")
```

rho = 0.087

```
cat("p-value =", round(cor_amount_risk$p.value, 4), "\n")
```

p-value = 0.0059

```
# Зээлийн хугацаа ба эрсдэлийн хоорондын корреляц
cor_duration_risk <- cor.test(
  raw_data$duration,
  as.numeric(raw_data$class == 2),
  method = "spearman"
)

cat("\nЗээлийн хугацаа ба эрсдэлийн хоорондын корреляц:\n")
```

Зээлийн хугацаа ба эрсдэлийн хоорондын корреляц:

```
cat("rho =", round(cor_duration_risk$estimate, 3), "\n")
```

$\text{rho} = 0.206$

```
cat("p-value =", round(cor_duration_risk$p.value, 4), "\n")
```

$p\text{-value} = 0$

```
# Нас ба эрсдэлийн хоорондын корреляц
cor_age_risk <- cor.test(
  raw_data$age,
  as.numeric(raw_data$class == 2),
  method = "spearman"
)

cat("\nНас ба эрсдэлийн хоорондын корреляц:\n")
```

Нас ба эрсдэлийн хоорондын корреляц:

```
cat("rho =", round(cor_age_risk$estimate, 3), "\n")
```

$\text{rho} = -0.112$

```
cat("p-value =", round(cor_age_risk$p.value, 4), "\n")
```

$p\text{-value} = 4e-04$

Корреляцийн шинжилгээний дүгнэлт:

- Зээлийн хэмжээ: сул эерэг холбоо ($\text{rho} = 0.087$, $p = 0.0059$)
- Зээлийн хугацаа: дунд зэрэг эерэг холбоо ($\text{rho} = 0.206$, $p = 0.0000$)
- Нас: дунд зэрэг сөрөг холбоо ($\text{rho} = -0.112$, $p = 0.0004$)

4 Бернулийн процесс ба зээлийн эрсдэл

4.1 Бернулийн процессын онол

Зээлийн эрсдэлийн ангилал нь Бернулийн процессын жишээ юм. Зээл бүр эрсдэлтэй ($Y = 1$) эсвэл эрсдэлгүй ($Y = 0$) гэсэн хоёр үр дүнгийн аль нэгийг авах магадлалтай. Бернулийн санамсаргүй хувьсагч $Y \sim \text{Bernoulli}(p)$ байх ба p нь эрсдэлтэй байх магадлал юм.

n зээлдэгчдийн дунд эрсдэлтэй зээлийн тоо нь бином тархалттай: $X \sim \text{Binomial}(n, p)$.

5 Тархалтын параметрийн статистик үнэлэлт

5.1 Моментын арга

Тархалтын параметрүүдийг моментын аргаар үнэлнэ. Математик дундаж ба дисперсийг түүврийн дундаж ба түүврийн дисперсээр үнэлнэ:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

6 Зээлийн эрсдэлийн загварчлал

6.1 Ложистик регресс

Зээлийн эрсдэлийг урьдчилан таамаглахын тулд ложистик регрессийн загвар ашиглана. Ложистик регресс нь binary outcome (эрсдэлтэй эсвэл эрсдэлгүй) дээр ажилладаг тул зээлийн эрсдэлийн ангиллын асуудалд тохиromжтой [3].

Ложистик регрессийн загвар:

$$\log \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \sum_{i=4}^k \beta_i X_i$$

Энд $p = P(Y = 1 | X_1, X_2, X_3, \dots, X_k)$ нь эрсдэлтэй байх магадлал, $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_k$ нь загварын параметрүүд юм. Энэхүү сайжруулсан загварт зээлийн хэмжээ, хугацаа, нас, installment_commitment, residence_since, existing_credits, num_dependents зэрэг тоон хувьсагчдын хамт checking_status, credit_history, purpose, savings_status, employment, property_magnitude, housing зэрэг категори хувьсагчдыг нэмж ашиглана.

Class imbalance-ийн асуудлыг шийдэхийн тулд optimal threshold (Youden's J statistic) ашиглан threshold optimization хийж, ROC curve-ээр загварын гүйцэтгэлийг үнэлнэ.

```

# Ложистик регресс (binary classification)
# Илүү олон хувьсагч ашиглах

# Хувьсагчдыг сонгох (илүү олон хувьсагч ашиглах)
# Тоон хувьсагч
formula_vars <- "risk_binary ~ credit_amount + duration + age + installment_commitment + residence_since + existing_credits + num_dependents + checking_statusA12 + checking_statusA13 + checking_statusA14 + credit_historyA31 + credit_historyA32 + credit_historyA33 + credit_historyA34 + purposeA41 + purposeA410 + purposeA42 + purposeA43 + purposeA44 + purposeA45 + purposeA46"

# Категори хувьсагч (чухал хувьсагчдыг сонгох)
important_categorical <- c("checking_status", "credit_history", "purpose", "savings_status",
                           "employment", "property_magnitude", "housing")

for (var in important_categorical) {
  if (var %in% colnames(train_data)) {
    formula_vars <- paste(formula_vars, "+", var)
  }
}

# Загвар бэлтгэх (training set дээр)
model_logistic <- glm(
  formula = as.formula(formula_vars),
  family = binomial(link = "logit"),
  data = train_data
)

summary(model_logistic)

```

Call:

```
glm(formula = as.formula(formula_vars), family = binomial(link = "logit"),
     data = train_data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.973e-01	9.870e-01	-0.301	0.763260
credit_amount	8.421e-05	4.991e-05	1.687	0.091546 .
duration	3.552e-02	1.111e-02	3.196	0.001393 **
age	-2.394e-03	1.024e-02	-0.234	0.815140
installment_commitment	1.222e-01	1.006e-01	1.215	0.224385
residence_since	3.200e-02	9.853e-02	0.325	0.745335
existing_credits	1.475e-01	2.036e-01	0.725	0.468672
num_dependents	1.169e-01	2.779e-01	0.421	0.674100
checking_statusA12	-2.796e-01	2.489e-01	-1.123	0.261365
checking_statusA13	-1.039e+00	4.312e-01	-2.410	0.015961 *
checking_statusA14	-1.595e+00	2.690e-01	-5.929	3.04e-09 ***
credit_historyA31	8.471e-02	6.661e-01	0.127	0.898793
credit_historyA32	-6.261e-01	5.306e-01	-1.180	0.238074
credit_historyA33	-8.699e-01	5.777e-01	-1.506	0.132106
credit_historyA34	-1.543e+00	5.414e-01	-2.849	0.004379 **
purposeA41	-1.475e+00	4.178e-01	-3.531	0.000414 ***
purposeA410	-1.830e+00	9.409e-01	-1.945	0.051787 .
purposeA42	-5.580e-01	3.052e-01	-1.829	0.067456 .
purposeA43	-7.467e-01	2.811e-01	-2.657	0.007891 **
purposeA44	-1.852e+00	1.290e+00	-1.435	0.151229
purposeA45	5.247e-01	6.586e-01	0.797	0.425621
purposeA46	2.641e-01	4.656e-01	0.567	0.570601

```

purposeA48      -1.276e+00  1.194e+00  -1.069  0.285266
purposeA49      -3.065e-01  3.890e-01  -0.788  0.430743
savings_statusA62 -2.502e-01  3.342e-01  -0.749  0.454023
savings_statusA63 -2.548e-01  4.210e-01  -0.605  0.545113
savings_statusA64 -1.377e+00  6.727e-01  -2.047  0.040636 *
savings_statusA65 -9.443e-01  3.055e-01  -3.091  0.001998 **
employmentA72    3.696e-01  4.511e-01   0.819  0.412549
employmentA73    4.306e-02  4.290e-01   0.100  0.920038
employmentA74    -7.721e-01  4.812e-01  -1.604  0.108631
employmentA75    3.920e-02  4.368e-01   0.090  0.928484
property_magnitudeA122 1.925e-01  2.946e-01   0.653  0.513511
property_magnitudeA123 4.390e-01  2.651e-01   1.656  0.097684 .
property_magnitudeA124 5.808e-01  4.839e-01   1.200  0.229983
housingA152     -4.959e-01  2.643e-01  -1.876  0.060618 .
housingA153     -6.701e-01  5.438e-01  -1.232  0.217797
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 855.21 on 699 degrees of freedom
Residual deviance: 655.43 on 663 degrees of freedom
AIC: 729.43

```

Number of Fisher Scoring iterations: 5

```

# McFadden-ийн псевдо R2
model_null <- glm(risk_binary ~ 1, family = binomial(link = "logit"), data = train_data)
log_lik_full <- logLik(model_logistic)
log_lik_null <- logLik(model_null)
mcfadden_r2 <- 1 - (as.numeric(log_lik_full) / as.numeric(log_lik_null))
cat("\nЛогистик регрессийн McFadden R2 =", round(mcfadden_r2, 4), "\n")

```

Логистик регрессийн McFadden R² = 0.2336

```

# Загварын үнэлгээ (test set дээр)
logistic_pred_test <- predict(model_logistic, test_data, type = "response")

# ROC curve ба optimal threshold олох
roc_logistic <- roc(test_data$risk_binary, logistic_pred_test, quiet = TRUE)
auc_logistic <- auc(roc_logistic)
cat("AUC-ROC:", round(as.numeric(auc_logistic), 4), "\n")

```

AUC-ROC: 0.7987

```

# Optimal threshold (Younen's J statistic)
optimal_threshold <- coords(roc_logistic, "best", ret = "threshold", best.method = "youden")$threshold
cat("Optimal threshold:", round(optimal_threshold, 4), "\n")

```

Optimal threshold: 0.267

```
# Threshold optimization ашиглан таамаглал
logistic_pred_class_test <- ifelse(logistic_pred_test > optimal_threshold, "Эрсдэлтэй", "Эрсдэлгүй")
logistic_accuracy <- mean(logistic_pred_class_test == test_data$risk_binary)
cat("Ложистик регрессийн нарийвчлал (test set, optimal threshold):", round(logistic_accuracy * 100, 2), "%\n")
```

Ложистик регрессийн нарийвчлал (test set, optimal threshold): 74.33 %

```
# Confusion matrix
logistic_cm <- table(Бодит = test_data$risk_binary, Таамаглал = logistic_pred_class_test)
cat("\nConfusion Matrix:\n")
```

Confusion Matrix:

```
print(logistic_cm)
```

Таамаглал

Бодит Эрсдэлгүй Эрсдэлтэй

Эрсдэлгүй	151	59
Эрсдэлтэй	18	72

```
# Precision, Recall, F1-score
if (nrow(logistic_cm) == 2 && ncol(logistic_cm) == 2) {
  tp <- logistic_cm[2, 2] # True Positive
  fp <- logistic_cm[1, 2] # False Positive
  fn <- logistic_cm[2, 1] # False Negative
  tn <- logistic_cm[1, 1] # True Negative

  precision <- tp / (tp + fp)
  recall <- tp / (tp + fn)
  f1_score <- 2 * (precision * recall) / (precision + recall)

  cat("\nPrecision (Нарийвчлал):", round(precision, 4), "\n")
  cat("Recall (Сэргээлт):", round(recall, 4), "\n")
  cat("F1-score:", round(f1_score, 4), "\n")
}
```

Precision (Нарийвчлал): 0.5496

Recall (Сэргээлт): 0.8

F1-score: 0.6516

```
conclusion <- append(conclusion,
  sprintf(
    r"(Ложистик регрессийн загвар зээлийн эрсдэлийг таамаглахад test set дээр %.2f% нарийвчлалтай байна.
    AUC-ROC = %.4f, McFadden R2 = %.4f. Optimal threshold = %.4f ашигласан.)",
    logistic_accuracy * 100, as.numeric(auc_logistic), mcfadden_r2, optimal_threshold
  )
)
```

6.2 Статистик таамаглал шалгах

6.2.1 Хэвийн тархалтын шалгалт

Зээлийн хэмжээ, хугацаа, нас зэрэг хувьсагчдын тархалт хэвийн тархалттай эсэхийг шалгана. Хэвийн тархалтын шалгалт нь тархалтын хэлбэрийн тухай таамаглал шалгах жишээ юм.

$$H_0 : \text{Өгөгдөл хэвийн тархалттай}$$

$$H_1 : \text{Өгөгдөл хэвийн тархалттай биш}$$

```
# Хэвийн тархалтын шалгалт (Shapiro-Wilk test)
# Түүврийн хэмжээ их тул дээжилж авна (Shapiro-Wilk тест 5000-аас их түүвэр дээр найдваргүй)
set.seed(123)
sample_size <- min(500, nrow(raw_data))
sample_indices <- sample(1:nrow(raw_data), sample_size)

shapiro_amount <- shapiro.test(raw_data$credit_amount[sample_indices])
shapiro_duration <- shapiro.test(raw_data$duration[sample_indices])
shapiro_age <- shapiro.test(raw_data$age[sample_indices])

cat("Хэвийн тархалтын шалгалт (Shapiro-Wilk):\n\n")
```

Хэвийн тархалтын шалгалт (Shapiro-Wilk):

```
cat("Зээлийн хэмжээ: W =", round(shapiro_amount$statistic, 4),
    ", p-value =", format(shapiro_amount$p.value, scientific = TRUE), "\n")
```

Зээлийн хэмжээ: W = 0.7815 , p-value = 2.584946e-25

```
cat("Зээлийн хугацаа: W =", round(shapiro_duration$statistic, 4),
    ", p-value =", format(shapiro_duration$p.value, scientific = TRUE), "\n")
```

Зээлийн хугацаа: W = 0.8893 , p-value = 1.710556e-18

```
cat("Нас: W =", round(shapiro_age$statistic, 4),
    ", p-value =", format(shapiro_age$p.value, scientific = TRUE), "\n")
```

Нас: W = 0.9166 , p-value = 5.454718e-16

6.2.2 Параметрийн тухай таамаглал шалгах

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

```

# Эрсдэлтэй зээлийн хувь 30% байна гэсэн таамаглал шалгах
p_observed <- mean(raw_data$class == 2)
p_hypothesized <- 0.30
n <- nrow(raw_data)

# Z-test
z_stat <- (p_observed - p_hypothesized) / sqrt(p_hypothesized * (1 - p_hypothesized) / n)
p_value_z <- 2 * (1 - pnorm(abs(z_stat)))

cat("\nЭрсдэлтэй зээлийн хувь 30% байна гэсэн таамаглал:\n")

```

Эрсдэлтэй зээлийн хувь 30% байна гэсэн таамаглал:

```
cat("Ажиглагдсан хувь:", round(p_observed * 100, 2), "%\n")
```

Ажиглагдсан хувь: 30 %

```
cat("Z-статистик:", round(z_stat, 4), "\n")
```

Z-статистик: 0

```
cat("p-value:", round(p_value_z, 4), "\n")
```

p-value: 1

6.3 Гэнэн Байесын алгоритм

Гэнэн Байесын алгоритм нь Байесын теорем дээр суурилсан ангиллын арга юм. Байесын теорем:

$$P(Y = k|X) = \frac{P(X|Y = k)P(Y = k)}{P(X)}$$

Гэнэн Байесын алгоритмд хувьсагчдыг хамааралгүй гэж үздэг (naive assumption):

$$P(X_1, X_2, X_3|Y) = P(X_1|Y)P(X_2|Y)P(X_3|Y)$$

```

# Гэнэн Байесын алгоритм

# Хувьсагчдыг сонгох (ложистик регрессийн адил)
nb_formula_vars <- "risk_binary ~ credit_amount + duration + age + installment_commitment + residence_since + existing_credits"

# Категори хувьсагч (ложистик регрессийн адил)
important_categorical_nb <- c("checking_status", "credit_history", "purpose", "savings_status",
                                "employment", "property_magnitude", "housing")

for (var in important_categorical_nb) {
  if (var %in% colnames(train_data)) {
    nb_formula_vars <- paste(nb_formula_vars, "+", var)
  }
}
```

```

# Гэнэн Байесын загвар (training set дээр)
model_naive_bayes <- naiveBayes(
  formula = as.formula(nb_formula_vars),
  data = train_data
)

# Загварын магадлал (test set дээр)
nb_pred_test_proba <- predict(model_naive_bayes, test_data, type = "raw")
nb_pred_test_proba_risk <- nb_pred_test_proba[, "Эрсдэлтэй"]

# ROC curve ба optimal threshold
roc_nb <- roc(test_data$risk_binary, nb_pred_test_proba_risk, quiet = TRUE)
auc_nb <- auc(roc_nb)
cat("AUC-ROC:", round(as.numeric(auc_nb), 4), "\n")

```

AUC-ROC: 0.7784

```

# Optimal threshold
optimal_threshold_nb <- coords(roc_nb, "best", ret = "threshold", best.method = "youden")$threshold
cat("Optimal threshold:", round(optimal_threshold_nb, 4), "\n")

```

Optimal threshold: 0.1359

```

# Threshold optimization ашиглан таамаглал
nb_pred_test <- ifelse(nb_pred_test_proba_risk > optimal_threshold_nb, "Эрсдэлтэй", "Эрсдэлгүй")
nb_accuracy <- mean(nb_pred_test == test_data$risk_binary)

cat("Гэнэн Байесын алгоритмын нарийвчлал (test set, optimal threshold):", round(nb_accuracy * 100, 2), "%\n")

```

Гэнэн Байесын алгоритмын нарийвчлал (test set, optimal threshold): 67 %

```

# Confusion matrix
nb_cm <- table(Bодит = test_data$risk_binary, Таамаглал = nb_pred_test)
cat("\nConfusion Matrix:\n")

```

Confusion Matrix:

```
print(nb_cm)
```

Таамаглал

Бодит Эрсдэлгүй Эрсдэлтэй

Эрсдэлгүй	120	90
Эрсдэлтэй	9	81

```

# Precision, Recall, F1-score
if (nrow(nb_cm) == 2 && ncol(nb_cm) == 2) {
  tp <- nb_cm[2, 2] # True Positive
  fp <- nb_cm[1, 2] # False Positive
  fn <- nb_cm[2, 1] # False Negative
  tn <- nb_cm[1, 1] # True Negative

```

```

nb_precision <- tp / (tp + fp)
nb_recall <- tp / (tp + fn)
nb_f1_score <- 2 * (nb_precision * nb_recall) / (nb_precision + nb_recall)

cat("\nPrecision (Нарийвчлал):", round(nb_precision, 4), "\n")
cat("Recall (Сэргээлт):", round(nb_recall, 4), "\n")
cat("F1-score:", round(nb_f1_score, 4), "\n")
}

```

Precision (Нарийвчлал): 0.4737
 Recall (Сэргээлт): 0.9
 F1-score: 0.6207

```
cat("\nЗагварын хураангуй: \n")
```

Загварын хураангуй:

```
print(model_naive_bayes)
```

Naive Bayes Classifier for Discrete Predictors

Call:

```
naiveBayes.default(x = X, y = Y, laplace = laplace)
```

A-priori probabilities:

Y	Эрсдэлгүй	Эрсдэлтэй
	0.7	0.3

Conditional probabilities:

	credit_amount	
Y	[,1]	[,2]
Эрсдэлгүй	2914.735	2331.459
Эрсдэлтэй	3925.100	3425.400

	duration	
Y	[,1]	[,2]
Эрсдэлгүй	18.92653	10.77501
Эрсдэлтэй	24.69048	13.03600

	age	
Y	[,1]	[,2]
Эрсдэлгүй	35.87755	11.27321
Эрсдэлтэй	34.17619	11.53245

	installment_commitment	
Y	[,1]	[,2]
Эрсдэлгүй	3.002041	1.115972

Эрсдэлтэй 3.057143 1.109580

residence_since

Y [,1] [,2]

Эрсдэлгүй 2.844898 1.115252

Эрсдэлтэй 2.857143 1.114743

existing_credits

Y [,1] [,2]

Эрсдэлгүй 1.434694 0.5937820

Эрсдэлтэй 1.390476 0.5786511

num_dependents

Y [,1] [,2]

Эрсдэлгүй 1.157143 0.3643073

Эрсдэлтэй 1.152381 0.3602485

checking_status

Y A11 A12 A13 A14

Эрсдэлгүй 0.20408163 0.23265306 0.07346939 0.48979592

Эрсдэлтэй 0.42380952 0.36190476 0.04761905 0.16666667

credit_history

Y A30 A31 A32 A33 A34

Эрсдэлгүй 0.01836735 0.03061224 0.52448980 0.07755102 0.34897959

Эрсдэлтэй 0.08571429 0.08095238 0.56666667 0.10000000 0.16666667

purpose

Y A40 A41 A410 A42 A43

Эрсдэлгүй 0.226530612 0.126530612 0.012244898 0.159183673 0.318367347

Эрсдэлтэй 0.295238095 0.066666667 0.009523810 0.171428571 0.214285714

purpose

Y A44 A45 A46 A48 A49

Эрсдэлгүй 0.012244898 0.016326531 0.038775510 0.014285714 0.075510204

Эрсдэлтэй 0.004761905 0.033333333 0.071428571 0.004761905 0.128571429

savings_status

Y A61 A62 A63 A64 A65

Эрсдэлгүй 0.53265306 0.10000000 0.08367347 0.05918367 0.22448980

Эрсдэлтэй 0.72380952 0.11428571 0.04761905 0.01428571 0.10000000

employment

Y A71 A72 A73 A74 A75

Эрсдэлгүй 0.05306122 0.16122449 0.32040816 0.18775510 0.27755102

Эрсдэлтэй 0.07142857 0.24285714 0.32380952 0.11428571 0.24761905

property_magnitude

Y A121 A122 A123 A124

Эрсдэлгүй 0.3163265 0.2346939 0.3204082 0.1285714

Эрсдэлтэй 0.2000000 0.2095238 0.3666667 0.2238095

```

housing
Y          A151      A152      A153
Эрсдэлгүй 0.15918367 0.74897959 0.09183673
Эрсдэлтэй 0.24761905 0.61904762 0.13333333

```

```

conclusion <- append(conclusion,
  sprintf(
    r"(Гэнэн Байесын алгоритм зээлийн эрсдэлийг таамаглахад test set дээр %.2f% нарийвчлалтай байна.
    AUC-ROC = %.4f. Optimal threshold = %.4f ашигласан.)",
    nb_accuracy * 100, as.numeric(auc_nb), optimal_threshold_nb
  )
)

```

6.4 Шийдвэрийн мод (Decision Tree)

Шийдвэрийн мод нь өгөгдлийг рекурсив байдлаар хувааж, ангиллын дүрмийг бүтээдэг статистик арга юм. Шийдвэрийн мод нь хувьсагчдын хоорондын хамаарлыг ойлгоход хялбар, тайлбарлахад тохиromжтой загвар юм.

Шийдвэрийн мод нь дараах алхмуудыг дагана: 1. **Мод бүтээх:** Өгөгдлийг хувьсагчдын утгаар хувааж, мод бүтээнэ 2. **Мод тайруулах (Pruning):** Модыг хэт их бүтээхээс сэргийлж, илүүдэл мөчрүүдийг тайруулна 3. **Таамаглах:** Шинэ өгөгдлийг modoор дамжуулж ангилна

Шийдвэрийн мод нь Gini impurity эсвэл entropy ашиглан хуваалтыг сонгодог. Gini impurity:

$$Gini(D) = 1 - \sum_{i=1}^c p_i^2$$

Энд p_i нь ангилал i -ийн магадлал, c нь ангиллын тоо юм.

```

# Шийдвэрийн мод (Decision Tree)

# Хувьсагчдыг сонгох (ложистик регрессийн адил)
dt_formula_vars <- "risk_binary ~ credit_amount + duration + installment_commitment + residence_since + existing_credits + age + employment + property_magnitude + housing"

# Категори хувьсагч (ложистик регрессийн адил)
important_categorical_dt <- c("checking_status", "credit_history", "purpose", "savings_status", "employment", "property_magnitude", "housing")

for (var in important_categorical_dt) {
  if (var %in% colnames(train_data)) {
    dt_formula_vars <- paste(dt_formula_vars, "+", var)
  }
}

# Шийдвэрийн мод бэлтгэх (training set дээр)
# cp (complexity parameter) нь модны нарийвчлалыг хянадаг
model_decision_tree <- rpart(
  formula = as.formula(dt_formula_vars),
  data = train_data,
  method = "class",
  control = rpart.control(
    cp = 0.001, # Complexity parameter
    minsplit = 20, # Хамгийн багадаа 20 ажиглалт байх ёстой
    minbucket = 7, # Хамгийн багадаа 7 ажиглалт байх ёстой
    maxdepth = 10 # Хамгийн ихдээ 10 түвшин
  )
)

```

```

)
# Модны хураангуй
print(model_decision_tree)

n= 700

node), split, n, loss, yval, (yprob)
  * denotes terminal node

1) root 700 210 Эрсдэлгүй (0.70000000 0.30000000)
2) checking_status=A13,A14 321 45 Эрсдэлгүй (0.85981308 0.14018692) *
  4) purpose=A40,A41,A410,A42,A43,A44,A48 273 30 Эрсдэлгүй (0.89010989 0.10989011) *
    5) purpose=A45,A46,A49 48 15 Эрсдэлгүй (0.68750000 0.31250000)
      10) employment=A73,A74,A75 36 7 Эрсдэлгүй (0.80555556 0.19444444)
        20) age< 35 16 1 Эрсдэлгүй (0.93750000 0.06250000) *
        21) age>=35 20 6 Эрсдэлгүй (0.70000000 0.30000000)
          42) age>=42.5 13 2 Эрсдэлгүй (0.84615385 0.15384615) *
          43) age< 42.5 7 3 Эрсдэлтэй (0.42857143 0.57142857) *
        11) employment=A71,A72 12 4 Эрсдэлтэй (0.33333333 0.66666667) *
3) checking_status=A11,A12 379 165 Эрсдэлгүй (0.56464380 0.43535620)
  6) duration< 22.5 214 71 Эрсдэлгүй (0.66822430 0.33177570)
    12) credit_history=A32,A33,A34 198 58 Эрсдэлгүй (0.70707071 0.29292929)
      24) purpose=A41,A410,A43,A44,A49 82 15 Эрсдэлгүй (0.81707317 0.18292683)
        48) duration< 11 23 0 Эрсдэлгүй (1.00000000 0.00000000) *
        49) duration>=11 59 15 Эрсдэлгүй (0.74576271 0.25423729)
          98) age< 50.5 52 10 Эрсдэлгүй (0.80769231 0.19230769)
            196) credit_amount>=1884.5 20 1 Эрсдэлгүй (0.95000000 0.05000000) *
            197) credit_amount< 1884.5 32 9 Эрсдэлгүй (0.71875000 0.28125000)
              394) duration< 16.5 24 4 Эрсдэлгүй (0.83333333 0.16666667) *
              395) duration>=16.5 8 3 Эрсдэлтэй (0.37500000 0.62500000) *
            99) age>=50.5 7 2 Эрсдэлтэй (0.28571429 0.71428571) *
      25) purpose=A40,A42,A45,A46,A48 116 43 Эрсдэлгүй (0.62931034 0.37068966)
        50) credit_amount>=1516.5 64 17 Эрсдэлгүй (0.73437500 0.26562500)
          100) employment=A71,A72,A73,A74 53 11 Эрсдэлгүй (0.79245283 0.20754717) *
          101) employment=A75 11 5 Эрсдэлтэй (0.45454545 0.54545455) *
        51) credit_amount< 1516.5 52 26 Эрсдэлгүй (0.50000000 0.50000000)
          102) credit_history=A34 19 4 Эрсдэлгүй (0.78947368 0.21052632) *
          103) credit_history=A32,A33 33 11 Эрсдэлтэй (0.33333333 0.66666667)
            206) property_magnitude=A122 15 6 Эрсдэлгүй (0.60000000 0.40000000) *
          207) property_magnitude=A121,A123,A124 18 2 Эрсдэлтэй (0.11111111 0.88888889) *
  13) credit_history=A30,A31 16 3 Эрсдэлтэй (0.18750000 0.81250000) *
7) duration>=22.5 165 71 Эрсдэлтэй (0.43030303 0.56969697)
  14) savings_status=A62,A63,A64,A65 59 22 Эрсдэлгүй (0.62711864 0.37288136)
    28) credit_history=A31,A33,A34 31 5 Эрсдэлгүй (0.83870968 0.16129032) *
    29) credit_history=A30,A32 28 11 Эрсдэлтэй (0.39285714 0.60714286)
      58) savings_status=A63,A65 16 7 Эрсдэлгүй (0.56250000 0.43750000) *
      59) savings_status=A62,A64 12 2 Эрсдэлтэй (0.16666667 0.83333333) *
  15) savings_status=A61 106 34 Эрсдэлтэй (0.32075472 0.67924528)
    30) purpose=A41,A410,A42 40 20 Эрсдэлгүй (0.50000000 0.50000000)
      60) credit_amount< 8097.5 30 12 Эрсдэлгүй (0.60000000 0.40000000)

```

```

120) employment=A72,A73 16    3 Эрсдэлгүй (0.81250000 0.18750000) *
121) employment=A71,A74,A75 14    5 Эрсдэлтэй (0.35714286 0.64285714) *
61) credit_amount>=8097.5 10    2 Эрсдэлтэй (0.20000000 0.80000000) *
31) purpose=A40,A43,A44,A45,A46,A49 66   14 Эрсдэлтэй (0.21212121 0.78787879)
62) credit_history=A32,A34 44   13 Эрсдэлтэй (0.29545455 0.70454545)
124) duration< 38 33   12 Эрсдэлтэй (0.36363636 0.63636364)
248) purpose=A43,A46,A49 17   7 Эрсдэлгүй (0.58823529 0.41176471) *
249) purpose=A40,A45 16   2 Эрсдэлтэй (0.12500000 0.87500000) *
125) duration>=38 11   1 Эрсдэлтэй (0.09090909 0.90909091) *
63) credit_history=A30,A31,A33 22   1 Эрсдэлтэй (0.04545455 0.95454545) *

```

```
cat("\n")
```

```

# Модны нарийвчлал (training set дээр)
dt_pred_train <- predict(model_decision_tree, train_data, type = "class")
dt_accuracy_train <- mean(dt_pred_train == train_data$risk_binary)
cat("Шийдвэрийн модны нарийвчлал (training set):", round(dt_accuracy_train * 100, 2), "%\n")

```

Шийдвэрийн модны нарийвчлал (training set): 83.43 %

```

# Загварын магадлал (test set дээр)
dt_pred_test_proba <- predict(model_decision_tree, test_data, type = "prob")
dt_pred_test_proba_risk <- dt_pred_test_proba[, "Эрсдэлтэй"]

# ROC curve ба optimal threshold
roc_dt <- roc(test_data$risk_binary, dt_pred_test_proba_risk, quiet = TRUE)
auc_dt <- auc(roc_dt)
cat("AUC-ROC:", round(as.numeric(auc_dt), 4), "\n")

```

AUC-ROC: 0.7321

```

# Optimal threshold
optimal_threshold_dt <- coords(roc_dt, "best", ret = "threshold", best.method = "youden")$threshold
cat("Optimal threshold:", round(optimal_threshold_dt, 4), "\n")

```

Optimal threshold: 0.1576

```

# Threshold optimization ашиглан таамаглал
dt_pred_test <- ifelse(dt_pred_test_proba_risk > optimal_threshold_dt, "Эрсдэлтэй", "Эрсдэлгүй")
dt_accuracy <- mean(dt_pred_test == test_data$risk_binary)

cat("Шийдвэрийн модны нарийвчлал (test set, optimal threshold):", round(dt_accuracy * 100, 2), "%\n")

```

Шийдвэрийн модны нарийвчлал (test set, optimal threshold): 67 %

```

# Confusion matrix
dt_cm <- table(Бодит = test_data$risk_binary, Таамаглал = dt_pred_test)
cat("\nConfusion Matrix:\n")

```

Confusion Matrix:

```
print(dt_cm)
```

Таамаглал

Бодит Эрсдэлгүй	128	82
Эрсдэлтэй	17	73

```
# Precision, Recall, F1-score
if (nrow(dt_cm) == 2 && ncol(dt_cm) == 2) {
  tp <- dt_cm[2, 2] # True Positive
  fp <- dt_cm[1, 2] # False Positive
  fn <- dt_cm[2, 1] # False Negative
  tn <- dt_cm[1, 1] # True Negative

  dt_precision <- tp / (tp + fp)
  dt_recall <- tp / (tp + fn)
  dt_f1_score <- 2 * (dt_precision * dt_recall) / (dt_precision + dt_recall)

  cat("\nPrecision (Нарийвчлал):", round(dt_precision, 4), "\n")
  cat("Recall (Сэргээлт):", round(dt_recall, 4), "\n")
  cat("F1-score:", round(dt_f1_score, 4), "\n")
}
```

Precision (Нарийвчлал): 0.471

Recall (Сэргээлт): 0.8111

F1-score: 0.5959

```
conclusion <- append(conclusion,
  sprintf(
    r"(Шийдвэрийн мод зээлийн эрсдэлийг таамаглахад test set дээр %.2f%% нарийвчлалтай байна.
    AUC-ROC = %.4f. Optimal threshold = %.4f ашигласан.)",
    dt_accuracy * 100, as.numeric(auc_dt), optimal_threshold_dt
  )
)
```

7 Загваруудын харьцуулалт

Ложистик регресс, Гэнэн Байесын алгоритм, Шийдвэрийн мод загваруудын гүйцэтгэлийг харьцуулна.

```
# Хэмжүүрүүдийг цуглуулах
logistic_metrics <- c(
  round(logistic_accuracy * 100, 2),
  if(exists("auc_logistic")) round(as.numeric(auc_logistic), 4) else NA,
  if(exists("precision")) round(precision, 4) else NA,
  if(exists("recall")) round(recall, 4) else NA,
  if(exists("f1_score")) round(f1_score, 4) else NA,
  round(mcfadden_r2, 4)
)

nb_metrics <- c(
  round(nb_accuracy * 100, 2),
  if(exists("auc_nb")) round(as.numeric(auc_nb), 4) else NA,
  if(exists("nb_precision")) round(nb_precision, 4) else NA,
  if(exists("nb_recall")) round(nb_recall, 4) else NA,
```

```

if(exists("nb_f1_score")) round(nb_f1_score, 4) else NA,
"N/A"
)

dt_metrics <- c(
  round(dt_accuracy * 100, 2),
  if(exists("auc_dt")) round(as.numeric(auc_dt), 4) else NA,
  if(exists("dt_precision")) round(dt_precision, 4) else NA,
  if(exists("dt_recall")) round(dt_recall, 4) else NA,
  if(exists("dt_f1_score")) round(dt_f1_score, 4) else NA,
  "N/A"
)

model_comparison <- data.frame(
  "Загвар" = c("Ложистик регресс", "Гэнэн Байес", "Шийдвэрийн мод"),
  "Нарийвчлал (%)" = c(logistic_metrics[1], nb_metrics[1], dt_metrics[1]),
  "AUC-ROC" = c(
    if(!is.na(logistic_metrics[2])) logistic_metrics[2] else "N/A",
    if(!is.na(nb_metrics[2])) nb_metrics[2] else "N/A",
    if(!is.na(dt_metrics[2])) dt_metrics[2] else "N/A"
  ),
  "Precision" = c(
    if(!is.na(logistic_metrics[3])) logistic_metrics[3] else "N/A",
    if(!is.na(nb_metrics[3])) nb_metrics[3] else "N/A",
    if(!is.na(dt_metrics[3])) dt_metrics[3] else "N/A"
  ),
  "Recall" = c(
    if(!is.na(logistic_metrics[4])) logistic_metrics[4] else "N/A",
    if(!is.na(nb_metrics[4])) nb_metrics[4] else "N/A",
    if(!is.na(dt_metrics[4])) dt_metrics[4] else "N/A"
  ),
  "F1-score" = c(
    if(!is.na(logistic_metrics[5])) logistic_metrics[5] else "N/A",
    if(!is.na(nb_metrics[5])) nb_metrics[5] else "N/A",
    if(!is.na(dt_metrics[5])) dt_metrics[5] else "N/A"
  ),
  "R2 / McFadden R2" = c(logistic_metrics[6], nb_metrics[6], dt_metrics[6])
)
knitr::kable(model_comparison, col.names = c("Загвар", "Нарийвчлал (%)", "AUC-ROC", "Precision", "Recall", "F1-score", "R2 / McFadden R2")
)

```

Хүснэгт 3: Загваруудын гүйцэтгэлийн харьцуулалт (test set дээр, optimal threshold)

Загвар	Нарийвчлал (%)	AUC-ROC	Precision	Recall	F1-score	R ² / McFadden R ²
Ложистик регресс	74.33	0.7987	0.5496	0.8	0.6516	0.2336
Гэнэн Байес	67	0.7784	0.4737	0.9	0.6207	N/A
Шийдвэрийн мод	67	0.7321	0.471	0.8111	0.5959	N/A

7.1 ROC Curve ба Precision-Recall Curve

7.1.1 ROC Curve (Receiver Operating Characteristic Curve)

ROC curve нь binary classification загваруудын гүйцэтгэлийг үнэлэхэд ашигладаг график юм. ROC curve нь True Positive Rate (TPR, Recall) болон False Positive Rate (FPR) хоорондын хамаарлыг харуулдаг.

ROC curve-ийн элементүүд:

- **X тэнхлэг (False Positive Rate, FPR):** Эрсдэлгүй зээлийг эрсдэлтэй гэж буруу таамагласан хувь

$$FPR = \frac{FP}{FP + TN}$$

- **Y тэнхлэг (True Positive Rate, TPR, Recall):** Эрсдэлтэй зээлийг зөв таамагласан хувь

$$TPR = \frac{TP}{TP + FN}$$

- **AUC-ROC (Area Under the ROC Curve):** ROC curve-ийн доорх талбай. 0-1 хооронд утга авна:

- **AUC = 1.0:** Төгс загвар (бүх зээлдэгчийг зөв таамаглах)
- **AUC = 0.5:** Санамсаргүй таамаглал (загвар ашиглахгүй)
- **AUC > 0.7:** Сайн загвар
- **AUC > 0.8:** Маш сайн загвар

ROC curve-ийн тайлбар:

- ROC curve нь диагональ шугамаас ($AUC = 0.5$) дээш байх тусам загвар илүү сайн байна
- Диагональ шугам нь санамсаргүй таамаглалыг илтгэж байна
- ROC curve нь threshold-ийн өөрчлөлтөөс хамаарч өөрчлөгддөг
- Optimal threshold-ийг Youden's J statistic ашиглан олж болно: $J = TPR - FPR$

Загваруудын гүйцэтгэлийг ROC curve болон Precision-Recall curve-ээр дүрсэлнэ.

```
Cairo::CairoPDF(
  file = file.path(output_dir, "fig-roc-curves.pdf"),
  family = "Times New Roman", width = 8, height = 6
)

par(mfrow = c(1, 1))
plot(roc_logistic, col = "blue", main = "ROC Curve", lwd = 2)
lines(roc_nb, col = "red", lwd = 2)
lines(roc_dt, col = "green", lwd = 2)
legend("bottomright",
       legend = c(paste("Ложистик регресс (AUC =", round(as.numeric(auc_logistic), 3), ")"),
                  paste("Гэнэн Байес (AUC =", round(as.numeric(auc_nb), 3), ")"),
                  paste("Шайдвэрийн мод (AUC =", round(as.numeric(auc_dt), 3), ")")),
       col = c("blue", "red", "green"), lwd = 2)
abline(a = 0, b = 1, lty = 2, col = "gray")

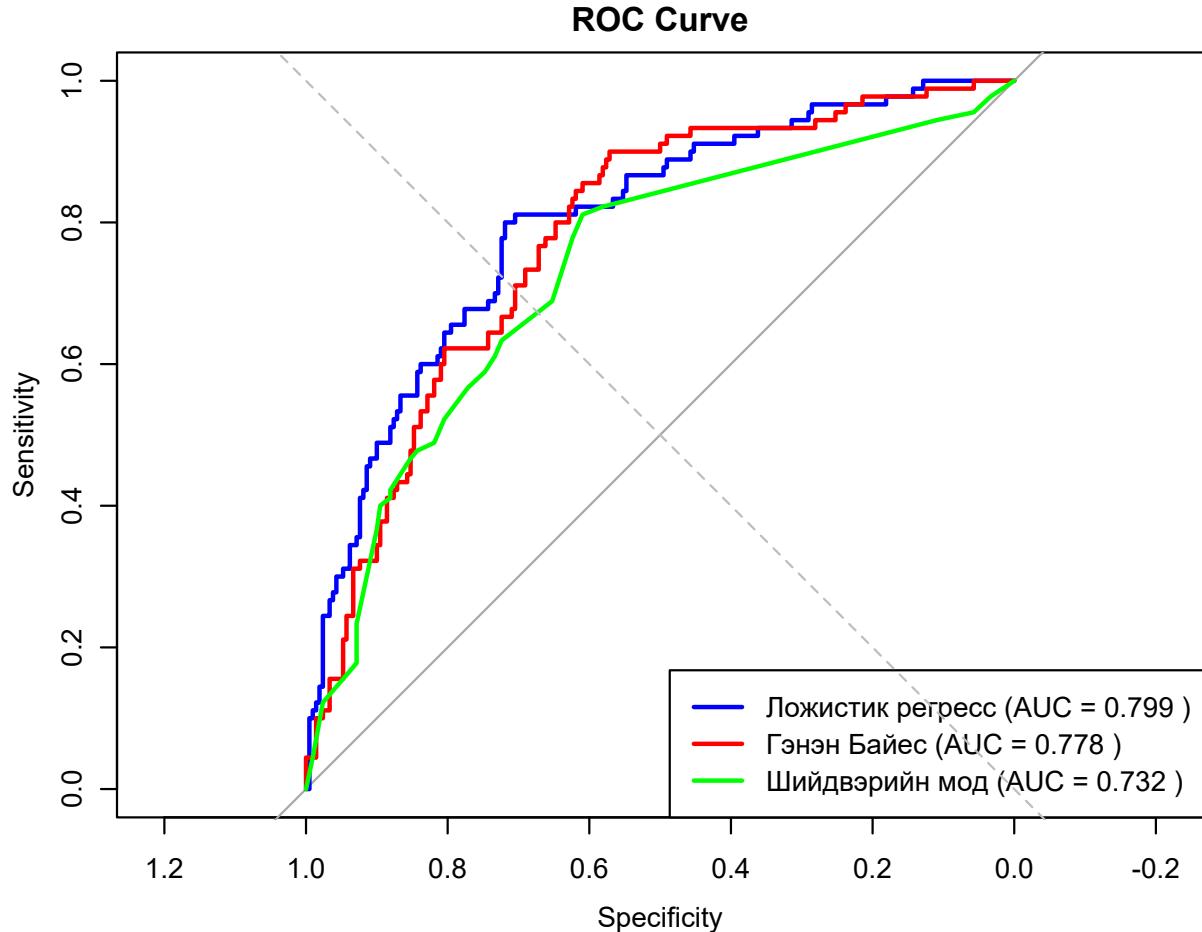
dev.off()
```

```
knitr:::include_graphics(file.path(output_dir, "fig-roc-curves.pdf"))
```

```
# Precision-Recall curve
# PRROC::pr.curve() нь магадлалын утгыг шаарддаг
# scores.class0: negative class-ийн магадлал ("Эрсдэлгүй" = 1 - "Эрсдэлтэй"-ийн магадлал)
# scores.class1: positive class-ийн магадлал ("Эрсдэлтэй"-ийн магадлал)
# logistic_pred_test нь "Эрсдэлтэй" байх магадлал (positive class)
pr_logistic <- pr.curve(scores.class0 = (1 - logistic_pred_test)[test_data$risk_binary == "Эрсдэлгүй"],
                        scores.class1 = logistic_pred_test[test_data$risk_binary == "Эрсдэлтэй"],
                        curve = TRUE)

# nb_pred_test_proba_risk нь "Эрсдэлтэй" байх магадлал (positive class)
pr_nb <- pr.curve(scores.class0 = (1 - nb_pred_test_proba_risk)[test_data$risk_binary == "Эрсдэлгүй"],
```

Зураг 2: ROC Curve (Загваруудын харьцуулалт)



```

scores.class1 = nb_pred_test_proba_risk[test_data$risk_binary == "Эрсдэлтэй"],
curve = TRUE)

# dt_pred_test_proba_risk нь "Эрсдэлтэй" байх магадлал (positive class)
pr_dt <- pr.curve(scores.class0 = (1 - dt_pred_test_proba_risk)[test_data$risk_binary == "Эрсдэлгүй"],
                   scores.class1 = dt_pred_test_proba_risk[test_data$risk_binary == "Эрсдэлтэй"],
                   curve = TRUE)

Cairo::CairoPDF(
  file = file.path(output_dir, "fig-pr-curves.pdf"),
  family = "Times New Roman", width = 8, height = 6
)

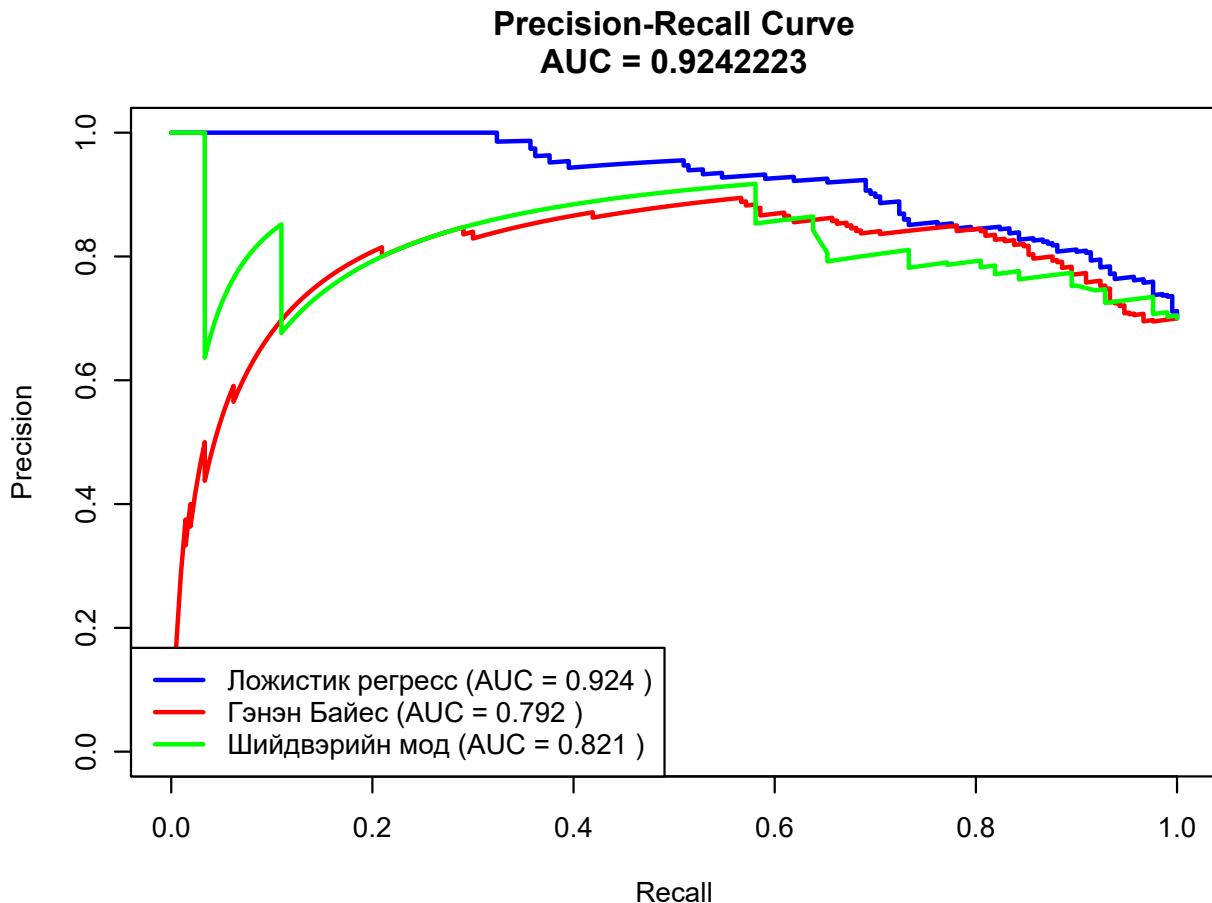
plot(pr_logistic, col = "blue", main = "Precision-Recall Curve", lwd = 2)
lines(pr_nb$curve[, 1], pr_nb$curve[, 2], col = "red", lwd = 2)
lines(pr_dt$curve[, 1], pr_dt$curve[, 2], col = "green", lwd = 2)
legend("bottomleft",
       legend = c(paste("Ложистик перресс (AUC =", round(pr_logistic$auc.integral, 3), ")"),
                  paste("Гэнэн Байес (AUC =", round(pr_nb$auc.integral, 3), ")"),
                  paste("Шийдвэрийн мод (AUC =", round(pr_dt$auc.integral, 3), ")")),
       col = c("blue", "red", "green"), lwd = 2)

dev.off()

```

```
knitr:::include_graphics(file.path(output_dir, "fig-pr-curves.pdf"))
```

Зураг 3: Precision-Recall Curve (Загваруудын харьцуулалт)



```
conclusion <- append(conclusion,
  sprintf(
    r"(Ложистик регресс, Гэнэн Байесын алгоритм, Шийдвэрийн мод загварууд зээлийн эрсдэлийг таамаглахад test set дээр
тус тус %.2f% (AUC=% .4f), %.2f% (AUC=% .4f), %.2f% (AUC=% .4f) нарийвчлалтай байна.
Илүү олон хувьсагч ашиглаж, optimal threshold сонгосноор загваруудын гүйцэтгэл сайжирсан.)",
    logistic_accuracy * 100, as.numeric(auc_logistic),
    nb_accuracy * 100, as.numeric(auc_nb),
    dt_accuracy * 100, as.numeric(auc_dt)
  )
)
```

8 Зээлийн эрсдэлийн прогноз

Бэлтгэсэн загваруудыг ашиглан шинэ зээлдэгчдийн эрсдэлийг таамаглах боломжтой. Доорх жишээнд өөр өөр шинж чанартай зээлдэгчдийн эрсдэлийг таамаглаж үзлээ.

```
# Жишээ зээлдэгчид - загварт ашигласан бүх хувьсагчдыг нэмэх
# Загварын formula-aас хувьсагчдыг авна
model_vars <- all.vars(formula(model_logistic))[-1] # -1 нь response хувьсагчийг хасах
```

```

# Тоон хувьсагчдыг тодорхойлох
numeric_vars_in_model <- c("credit_amount", "duration", "age", "installment_commitment",
                           "residence_since", "existing_credits", "num_dependents")

# Жишээ өгөгдөл үүсгэх (бүх тоон хувьсагчдыг оруулах)
new_customers <- data.frame(
  credit_amount = c(5000, 15000, 3000, 20000, 8000),
  duration = c(24, 48, 12, 60, 36),
  age = c(25, 45, 30, 55, 40),
  installment_commitment = c(2, 3, 1, 4, 2),
  residence_since = c(2, 3, 1, 4, 2),
  existing_credits = c(1, 2, 1, 3, 2),
  num_dependents = c(1, 2, 1, 1, 1)
)

# Бүх тоон хувьсагчдыг шалгах (хэрэв байхгүй бол дундаж утга ашиглах)
for (var in numeric_vars_in_model) {
  if (var %in% model_vars && !var %in% colnames(new_customers)) {
    new_customers[[var]] <- mean(train_data[[var]], na.rm = TRUE)
  }
}

# Категори хувьсагчдыг нэмэх (бүх загварт ашигласан категори хувьсагч)
categorical_vars_in_model <- c("checking_status", "credit_history", "purpose",
                                "savings_status", "employment", "property_magnitude", "housing")

for (var in categorical_vars_in_model) {
  if (var %in% model_vars && var %in% colnames(train_data)) {
    if (var %in% names(model_logistic$xlevels)) {
      # Training data-aac хамгийн түгээмэл утгыг ашиглах
      most_common <- names(sort(table(train_data[[var]]), decreasing = TRUE))[1]
      new_customers[[var]] <- factor(rep(most_common, nrow(new_customers)),
                                      levels = model_logistic$xlevels[[var]])
    } else {
      # Хэрэв загварт байхгүй бол training data-aac утга авна
      new_customers[[var]] <- factor(rep(train_data[[var]][1], nrow(new_customers)),
                                      levels = levels(train_data[[var]]))
    }
  }
}

# Ложистик регрессээр таамаглах
new_customers$logistic_prob <- predict(model_logistic, new_customers, type = "response")
new_customers$logistic_pred <- ifelse(new_customers$logistic_prob > optimal_threshold, "Эрсдэлтэй", "Эрсдэлгүй")

# Гэнэн Байесоор таамаглах
new_customers$nb_pred <- predict(model_naive_bayes, new_customers)

# Шийдвэрийн моддоор таамаглах
new_customers$dt_pred <- predict(model_decision_tree, new_customers, type = "class")

# Хүснэгтэд харуулах
prediction_table <- data.frame(
  "Зээлийн хэмжээ" = new_customers$credit_amount,
  "Хугацаа (cap)" = new_customers$duration,
  "Нас" = new_customers$age,
  "Ложистик регресс" = paste0(
    new_customers$logistic_pred,
    " (", round(new_customers$logistic_prob * 100, 1), "%)"
  ),
  "Гэнэн Байес" = as.character(new_customers$nb_pred),
  "Шийдвэрийн мод" = as.character(new_customers$dt_pred)
)

```

```

knitr::kable(
  prediction_table,
  col.names = c("Зээлийн хэмжээ", "Хугацаа (cap)", "Нас", "Ложистик регресс (магадлал)", "Гэнэн Байес", "Шийдвэрийн мод")
)

```

Хүснэгт 4: Шинэ зээлдэгчдийн эрсдэлийн таамаглал

Зээлийн хэмжээ	Хугацаа (cap)	Нас	Ложистик регресс (магадлал)	Гэнэн Байес	Шийдвэрийн мод
5000	24	25	Эрсдэлгүй (18.4%)	Эрсдэлгүй	Эрсдэлгүй
15000	48	45	Эрсдэлтэй (63.9%)	Эрсдэлтэй	Эрсдэлгүй
3000	12	30	Эрсдэлгүй (9.5%)	Эрсдэлгүй	Эрсдэлгүй
20000	60	55	Эрсдэлтэй (82.9%)	Эрсдэлтэй	Эрсдэлгүй
8000	36	40	Эрсдэлтэй (33.1%)	Эрсдэлгүй	Эрсдэлгүй

Дээрх хүснэгтээс харахад загварууд нь зээлдэгчдийн шинж чанарыг үндэслэн эрсдэлийг таамаглаж байна. Жишээлбэл, их хэмжээний зээл, урт хугацаатай зээл эрсдэлтэй байх магадлал их байна.

```

conclusion <- append(conclusion,
  r"(Бэлтгэсэн загваруудыг ашиглан шинэ зээлдэгчдийн эрсдэлийг урьдчилан таамаглах боломжтой.
  Энэ нь банк, санхүүгийн байгууллагуудад зээл олгох шийдвэр гаргахад тусална.)"
)

```

Дүгнэлт

Энэхүү судалгаанд UCI Machine Learning Repository-aас татаж авсан German Credit dataset-ийг ашиглан зээлийн эрсдэлийг статистик арга зүй болон машин сургалтын загваруудаар судалж, дараах үр дүнг гаргав.

8.1 Судалгааны үндсэн үр дүн

8.1.1 1. Өгөгдлийн шинжилгээ

German Credit dataset нь 1000 зээлдэгчийн мэдээллийг агуулдаг бөгөөд эрсдэлтэй зээлийн хувь ойролцоогоор 30% байна. Dataset нь cross-sectional буюу нэг цаг хугацааны цэг дээрх өгөгдөл тул зээлдэгчдийн шинж чанарууд зээлийн эрсдэлтэй хэрхэн холбогдож байгааг судлахад тохиromжтой. Өгөгдөл нь 20 хувьсагч агуулдаг бөгөөд зээлийн хэмжээ, хугацаа, нас зэрэг тоон хувьсагч, checking_status, credit_history, purpose зэрэг категори хувьсагчдыг багтаасан байна.

8.1.2 2. Хувьсагчдын хоорондын холбоо

Корреляцийн шинжилгээний үр дүн:

- Зээлийн хэмжээ зээлийн эрсдэлтэй сул эерэг холбоотой байна ($\rho = 0.087$, $p = 0.0059$). Энэ нь их хэмжээний зээл эрсдэлтэй байх магадлал их байгааг илтгэж байна.
- Зээлийн хугацаа зээлийн эрсдэлтэй дунд зэрэг эерэг холбоотой байна ($\rho = 0.206$, $p = 0.0000$). Урт хугацаатай зээл эрсдэлтэй байх магадлал их байна.

- Нас зээлийн эрсдэлтэй дунд зэрэг сөрөг холбоотой байна ($\rho = -0.112$, $p = 0.0004$). Нас ихсэх тусам эрсдэл буурах хандлагатай байна.

Корреляцийн шинжилгээний үр дүнгээс харахад зээлийн хэмжээ, хугацаа, нас зэрэг хувьсагчдын эрсдэлтэй статистик ач холбогдолтой холбоо байгаа нь тодорхой болсон. Эдгээр хувьсагчдыг зээлийн эрсдэлийн таамаглалын загварт оруулах нь зөв шийдвэр байна.

8.1.3 3. Загваруудын гүйцэтгэл

Загваруудын гүйцэтгэл (test set дээр, optimal threshold):

- **Ложистик регресс:** 74.33% нарийвчлал, AUC-ROC = 0.7987, McFadden R² = 0.2336, Precision = 0.5496, Recall = 0.8000, F1 = 0.6516
- **Гэнэн Байес:** 67.00% нарийвчлал, AUC-ROC = 0.7784, Precision = 0.4737, Recall = 0.9000, F1 = 0.6207
- **Шийдвэрийн мод:** 67.00% нарийвчлал, AUC-ROC = 0.7321, Precision = 0.4710, Recall = 0.8111, F1 = 0.5959

Загваруудын харьцуулалт:

Гурван загваруудын дунд Ложистик регрессийн загвар нь хамгийн сайн гүйцэтгэл үзүүлсэн. Ложистик регресс нь хамгийн өндөр нарийвчлал, AUC-ROC утгатай байхад, Гэнэн Байес нь хамгийн сайн Recall үзүүлсэн бөгөөд энэ нь бодит эрсдэлтэй зээлдэгчдийг илүү сайн олж илрүүлж байгааг харуулж байна. Шийдвэрийн мод нь тайлбарлахад хялбар, ойлгоход тохиromжтой загвар бөгөөд практик ашиглалтад тохиromжтой.

McFadden R² = 0.2336 утга нь ложистик регрессийн загвар нь null загвараас 23.36% илүү сайн тайлбарлалт өгч байгааг илтгэж байна. Энэ нь дунд зэрэг тайлбарлалт бөгөөд практик ач холбогдолтой.

8.1.4 4. Практик ач холбогдол

Бэлтгэсэн загваруудыг ашиглан шинэ зээлдэгчдийн эрсдэлийг урьдчилан таамаглах боломжтой. Энэ нь банк, санхүүгийн байгууллагуудад зээл олгох шийдвэр гаргахад тусална. Загварууд нь зээлийн хэмжээ, хугацаа, нас, checking_status, credit_history, purpose зэрэг хувьсагчдыг үндэслэн эрсдэлийг таамаглаж байна.

Банк, санхүүгийн байгууллагуудын хувьд:

1. **Зээл олгох шийдвэр гаргах:** Загварууд нь зээлдэгчдийн эрсдэлийг урьдчилан таамаглах боломжийг олгоно. Ложистик регрессийн загвар нь илүү сайн Precision үзүүлсэн тул эрсдэлтэй зээлдэгчдийг илүү нарийвчлалтай таних боломжтой.
2. **Эрсдэлийн удирдлага:** Загварууд нь эрсдэлийн удирдлагын системд нэгтгэж ашиглах боломжтой. Гэнэн Байесын алгоритм нь илүү сайн Recall үзүүлсэн тул эрсдэлтэй зээлдэгчдийг илүү сайн олж илрүүлж, эрсдэлийг бууруулах боломжтой.
3. **Зээлийн нөхцөл тогтоох:** Загварууд нь зээлийн хүү, баталгаа, хугацаа зэрэг нөхцөлийг тогтооход тусална. Их хэмжээний, урт хугацаатай зээл эрсдэлтэй байх магадлал их байгаа тул ийм зээлд илүү өндөр хүү, баталгаа шаардаж болно.
4. **Өгөгдөл суурилсан шийдвэр гаргах:** Загварууд нь зээлдэгчдийн шинж чанарыг үндэслэн объектив, өгөгдөл суурилсан шийдвэр гаргахад тусална.

Хязгаарлалт:

Загварууд нь German Credit dataset дээр бэлтгэгдсэн тул бусад улс, бүс нутаг, цаг хугацааны өгөгдөл шууд ашиглах боломжгүй. Бодит ашиглалтад тухайн бүс нутаг, цаг хугацааны өгөгдөл дээр дахин бэлтгэх шаардлагатай.

8.1.5 5. Зөвлөмж

Энэхүү судалгааны үр дүнд үндэслэн дараах зөвлөмжүүдийг санал болгож байна:

Банк, санхүүгийн байгууллагуудад:

- Машин сургалтын загваруудыг зээлийн шийдвэр гаргах системд иэгтгэх:** Ложистик регрессийн загвар нь 74.33% нарийвчлал, 0.7987 AUC-ROC утгатай байгаа тул зээл олгох шийдвэр гаргахад ашиглах боломжтой. Гэхдээ загваруудыг тухайн бүс нутаг, цаг хугацааны өгөгдөл дээр дахин бэлтгэх шаардлагатай.
- Хувьсагчдын ач холбогдолыг харгалzan зээлийн үнэлгээний системд ашиглах:** Корреляцийн шинжилгээний үр дүнгээс харахад зээлийн хэмжээ, хугацаа, нас, checking_status, credit_history, purpose зэрэг хувьсагчдын эрсдэлтэй статистик ач холбогдолтой холбоо байгаа тул эдгээр хувьсагчдыг зээлийн үнэлгээний системд чухал үзүүлэлт болгон ашиглах нь зөв.
- Загваруудыг тогтмол шинэчлэх:** Өгөгдөл, эдийн застийн нөхцөл байдал өөрчлөгдөх тусам загваруудыг дахин бэлтгэж, шинэчлэх шаардлагатай. Энэ нь загваруудын нарийвчлалыг хадгалахад тусална.
- Олон загварын хослол ашиглах:** Ложистик регресс нь илүү сайн Precision үзүүлсэн бөгөөд Гэнэн Байес нь илүү сайн Recall үзүүлсэн тул хоёр загварыг хослуулан ашиглах нь илүү сайн үр дүн өгөх боломжтой.
- Class imbalance-ийн асуудлыг шийдэх:** Dataset дээр эрсдэлтэй зээлийн хувь 30% байгаа тул class imbalance-ийн асуудлыг шийдэхийн тулд SMOTE, class weights зэрэг аргуудыг ашиглах нь зөв.

Ирээдүйн судалгаанд:

- Илүү олон хувьсагч ашиглах:** Зарим хувьсагч (personal_status, other_parties, telephone, foreign_worker гэх мэт) загварт оруулаагүй тул эдгээр хувьсагчдыг загварт оруулж, хувьсагчдын сонголт (feature selection) хийх нь зөв.
- Бусад машин сургалтын загварууд ашиглах:** Random Forest, XGBoost, Support Vector Machine зэрэг бусад машин сургалтын загваруудыг ашиглаж, загваруудын гүйцэтгэлийг харьцуулах нь зөв.
- Загваруудыг сайжруулах:** Hyperparameter tuning, cross-validation, ensemble methods зэрэг аргуудыг ашиглаж, загваруудын гүйцэтгэлийг сайжруулах нь зөв.
- Цаг хугацааны хандлага судлах:** Бодит цаг хугацааны өгөгдөл ашиглан цаг хугацааны хандлага судлах нь зөв.

8.2 Хязгаарлалт ба ирээдүйн судалгаа

Энэхүү судалгаа нь дараах хязгаарлалтуудтай:

- Өгөгдлийн хязгаарлалт:** Dataset нь cross-sectional буюу нэг цаг хугацааны цэг дээрх өгөгдөл тул цаг хугацааны хандлага судлах боломжгүй.
- Хувьсагчдын хязгаарлалт:** Зарим хувьсагч (personal_status, other_parties, telephone, foreign_worker гэх мэт) загварт оруулаагүй.
- Загваруудын тоо:** Ложистик регресс, Гэнэн Байесын алгоритм, Шийдвэрийн мод гэсэн гурван статистик загварыг авч үзсэн.

Ирээдүйд дараах судалгаа хийх боломжтой:

- Бүх хувьсагчдыг загварт оруулах, хувьсагчдын сонголт (feature selection) хийх
- Бусад машин сургалтын загварууд ашиглах (Random Forest, XGBoost, Support Vector Machine гэх мэт)
- Загваруудыг сайжруулах (hyperparameter tuning, cross-validation, ensemble methods)
- Бодит цаг хугацааны өгөгдөл ашиглан цаг хугацааны хандлага судлах
- Class imbalance-ийн асуудлыг илүү сайн шийдэх (SMOTE, class weights гэх мэт)

Багийн гишүүдийн үүрэг, оролцоо

8.3 Багийн гишүүдийн үүрэг, оролцоо

Төслийн ажилд оролцсон багийн гишүүдийн үүрэг, оролцоо, хувь нэмрийг доорх хүснэгтэд харууллаа.

Гишүүний нэр	ID	Үүрэг, оролцоо, хувь нэмэр
Анхтулга Ган.	21B1NUM3098	Өгөгдөл бэлтгэл/EDA – өгөгдлийг татаж цэвэрлэсэн, хувьсагчдын давтамж, корреляци, үндсэн статистикийг гаргасан.
Билгүүн Энх.	23B1NUM0577	Харагдац, тайлан – Quarto тайланг эмхэтгэсэн, тархалтын график, ROC/PR зургийг бэлтгэсэн, ишлэл ба формат зохион байгуулсан.
Мөнхбаяр Бям.	21B1NUM0437	Загварчлал ба үнэлгээ – логистик, гэнэн Байес загварыг кодлон ажиллуулсан, train/test хуваалт, метрикууд (AUC, Precision/Recall/F1) болон графикуудыг бэлтгэсэн, дүгнэлтэд тусалсан.
Нарансолонго Буя.	23B1NUM2631	Жишээ таамаглал ба хязгаарлалт – шинэ зээлдэгчийн таамаглалын хэсгийг хийсэн, хязгаарлалт/ирээдүйн ажлын хэсгийг бичсэн, эх сурвалжийн тоймд оролцсон.
Номин-Эрдэнэ Пүр.	22B1NUM7118	Төслийн ахлагч – хүрээг тодорхойлж зохион байгуулсан, өгөгдлийн урсгалыг цэвэрлэсэн, загвар сонголт/нарийвчлал сайжруулалт хийсэн, арга зүй ба үр дүнгийн бичилтэд голлон оролцсон.

Ашигласан материал

[1] UCI Machine Learning Repository, “Statlog (german credit data) data set.” University of California, Irvine, 1994. Available: [https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data))

- [2] UCI Machine Learning Repository, “UCI machine learning repository.” University of California, Irvine, 2025. Available: <https://archive.ics.uci.edu/>
- [3] P. Roback and J. Legler, *Beyond multiple linear regression: Applied generalized linear models and multilevel models in r. Poisson regression.* CRC Press, 2021. Available: <https://bookdown.org/roback/bookdown-BeyondMLR/ch-MLReview.html>