

Оюутны эцсийн дүнг урьдчилан таамаглах

Багийн гишүүдийн нэрс

2025 оны 12-р сарын 3

Table of contents

Хураангуй	2
1 1. Оршил	2
1.1 1.1 Төслийн ажлын зорилго	2
2 2. Өгөгдлийн эх сурвалж ба тайлбар	2
2.1 2.1 Эх сурвалж	2
2.2 2.2 Өгөгдлийн тайлбар	2
2.3 2.3 Өгөгдлийг уншиж танилцах	3
3 3. Хэрэглэсэн арга загварын танилцуулга	3
3.1 3.1 Шугаман регресс	3
3.2 3.2 Математик загвар	3
3.3 3.3 Загварын давуу тал	3
4 4. Өгөгдөлтэй танилцах шинжилгээ	3
4.1 4.1 Зорилтот хувьсагчийн шинжилгээ (G3)	3
4.2 4.2 Корреляцийн шинжилгээ	4
4.3 4.3 Категори хувьсагчдын шинжилгээ	4
4.4 4.4 Тоон хувьсагчдын корреляци	4
5 5. Загварыг хэрэгжүүлсэн алхмууд	4
5.1 5.1 Шаардлагатай сангуудыг импортлох	4
5.2 5.2 Өгөгдөл боловсруулалт	4
5.3 5.3 Сургалт ба тестийн олонлог	5
5.4 5.4 Загварыг сургах	5
5.5 5.5 Шинж чанаруудын коэффициентүүд	5
6 6. Үр дүн ба загварын үнэлгээ	5
6.1 6.1 Таамаглал	5
6.2 6.2 Загварын гүйцэтгэл	5
6.3 6.3 Үлдэгдлийн шинжилгээ	6
6.4 6.4 Хөндлөн баталгаажуулалт	6
7 7. Дүгнэлт	6
7.1 7.1 Үндсэн дүгнэлтүүд	6
7.2 7.2 Практик хэрэглээ	6

7.3	7.3 Хязгаарлалтууд	6
7.4	7.4 Цаашдын судалгаа	6
8	8. Багийн гишүүдийн үүрэг оролцоо	7
	Ашигласан материал	7

Энэхүү төслийн ажлын зорилго нь Португалийн дунд сургуулийн математикийн хичээлийн оюутнуудын эцсийн дүн (G3)-ийг урьдчилан таамаглахад шугаман регрессийн загварыг ашиглан статистикийн арга зүйг практикт хэрэглэх юм. Өгөгдлийн олонлогт 395 оюутны 33 шинж чанарын мэдээлэл багтсан бөгөөд загварын үр дүнг үнэлсэн. Үр дүнгээр оюутнуудын хоёр дахь үеийн дүн (G2) нь эцсийн дүнг урьдчилан таамаглахад хамгийн чухал хүчин зүйл бөгөөд загварын R^2 үзүүлэлт 0.75 байв.

1 1.

[Яагаад оюутны дүнг урьдчилан таамаглах чухал вэ? Машин сургалт боловсролд хэрхэн тусалдаг талаар]

1.1 1.1

Төслийн ажлын үндсэн зорилгууд:

1. [Зорилго 1]
2. [Зорилго 2]
3. [Зорилго 3]
4. [Зорилго 4]

2 2.

2.1 2.1

[Kaggle холбоос, Португалийн оюутнууд, математикийн хичээл]

2.2 2.2

[395 оюутан, 33 шинж чанар]

Демографик мэдээлэл:

[school, sex, age, address, famsize, Pstatus]

Боловсролын мэдээлэл:

[Medu, Fedu, studytime, failures, schoolsup, absences]

Сурлагын дүн:

[G1, G2, G3 тайлбар]

2.3 2.3

```
# Сангуудыг импортлох
# pandas, numpy, matplotlib, seaborn
# Өгөгдлийг уншиж, хэмжээг харуулах
```

[Өгөгдөл амжилттай ачаалагдсан тухай]

```
# df.head() - эхний 5 мөр
```

```
# df.describe() - статистик үзүүлэлтүүд
```

```
# Дутуу утга шалгах
```

[Дутуу утга байхгүй]

3 3.

3.1 3.1

[Шугаман регресс юу вэ, яагаад хэрэглэх вэ]

3.2 3.2

Шугаман регрессийн ерөнхий загвар:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Энд:

- [y тайлбар]
- [β_0 тайлбар]
- [β_i тайлбар]
- [x_i тайлбар]
- [ϵ тайлбар]

3.3 3.3

[Коэффициент ойлгогдох, хэрэгжүүлэхэд хялбар, хурдан, статистик үндэслэлтэй гэх мэт]

4 4.

4.1 4.1 (G3)

```
# G3 статистик (дундаж, медиан, std, min, max)
# Histogram + boxplot зурах
```

[G3 тархалтын тайлбар]

4.2 4.2

```
# G1, G2, G3 корреляцийн матриц
# Heatmap зурах
```

[G2-G3 корреляци өндөр тухай]

```
# G1 vs G3, G2 vs G3 scatter plots
```

[Шугаман хамаарал харагдаж байна]

4.3 4.3

```
# sex, school, internet, higher гэх мэтийн boxplot-ууд
```

[Дээд боловсрол хүсч буй оюутнууд илүү өндөр дүнтэй]

4.4 4.4

```
# Тоон хувьсагчдын G3-тай корреляци
# Bar chart
```

[failures хамгийн сөрөг корреляцитай]

5 5.

5.1 5.1

```
# sklearn импортлох
# train_test_split, LinearRegression, LabelEncoder
# mean_squared_error, r2_score гэх мэт
```

[Эдгээр сангууд юунд ашиглагдах]

5.2 5.2

```
# Категори хувьсагчдыг кодлох (Label Encoding)
```

[Текст утгуудыг тоонд хөрвүүлсэн]

```
# X (features) болон y (target) ялгах
```

5.3 5.3

```
# train_test_split 80/20
```

[316 сургалт, 79 тест]

5.4 5.4

```
# LinearRegression()  
# model.fit()
```

5.5 5.5

```
# Коэффициентүүдийг dataframe-д оруулах  
# Bar chart зурах  
# Топ 10 харуулах
```

[G2 хамгийн өндөр коэффициенттэй (0.95)]

6 6.

6.1 6.1

```
# y_pred = model.predict()  
# Comparison table (бодит vs таамаглал)
```

6.2 6.2

```
# MSE, RMSE, MAE, R2 тооцоолох  
# Үр дүнг хэвлэх
```

[R² = 0.75, MAE = 1.5 тайлбар]

```
# Бодит vs таамаглал scatter plot
```

[Улаан шугамд ойрхон = сайн таамаглал]

6.3 6.3

```
# Residual scatter plot  
# Residual histogram
```

[Үлдэгдэл нормал тархалттай = таамаглал хангагдсан]

6.4 6.4

```
# 5-fold cross-validation  
#  $R^2$  оноонуудыг харуулах
```

[Дундаж $R^2 = 0.79$, тогтвортой]

7 7.

7.1 7.1

Төслийн ажлын үндсэн дүгнэлтүүд:

1. [G2-ийн тухай дүгнэлт]
2. [Загварын нарийвчлалын тухай дүгнэлт]
3. [Failures-ийн сөрөг нөлөөний тухай]
4. [Бусад хүчин зүйлсийн тухай]
5. [Cross-validation-ий тухай]

7.2 7.2

[Хэрхэн хэрэглэж болох - эрт таних, багш дэмжлэг гэх мэт]

7.3 7.3

[Зөвхөн Португалийн өгөгдөл, корреляци≠шалтгаан, G2 шаардлагатай]

7.4 7.4

[Бусад загвар туршиж үзэх, feature engineering, илүү их өгөгдөл]

8 8.

Гишүүний нэр	Үүрэг	Хувь нэмэр
[Нэр 1]	[Үүрэг 1]	25%
[Нэр 2]	[Үүрэг 2]	30%
[Нэр 3]	[Үүрэг 3]	25%
[Нэр 4]	[Үүрэг 4]	20%

Тэмдэглэл: [Бүх гишүүд идэвхтэй оролцсон]