

ЗЭЭЛИЙН БАТЛАМЖИЙН ТААМАГЛАЛ

Машин сургалтын төсөл

Багийн гишүүд:

- [Нэр 1] - Өгөгдөл боловсруулалт, цэвэрлэгээ
- [Нэр 2] - Машин сургалтын загвар ажиллуулалт
- [Нэр 3] - Визуализаци, график үүсгэлт
- [Нэр 4] - Дүн шинжилгээ, дүгнэлт
- [Нэр 5] - Тайлан бичилт, танилцуулга

Огноо: 2025-12-05

Эх сурвалж: Kaggle - Loan Approval Prediction Dataset

ХУРААНГУЙ

Энэхүү төсөл нь зээлийн батламжийг машин сургалтын аргаар таамаглах зорилготой. Kaggle-аас авсан 614 мөрөөс бүрдсэн өгөгдлийг ашигласан.

ГҮЙЦЭТГЭСЭН АЖЛУУД:

- Өгөгдлийн цэвэрлэлт, боловсруулалт
- 11 хувьсагчийн шинжилгээ
- 4 төрлийн машин сургалтын загвар ажиллуулсан
- Загваруудын үр дүнг харьцуулсан

ҮР ДҮН:

Хамгийн сайн загвар: Логистик регресс

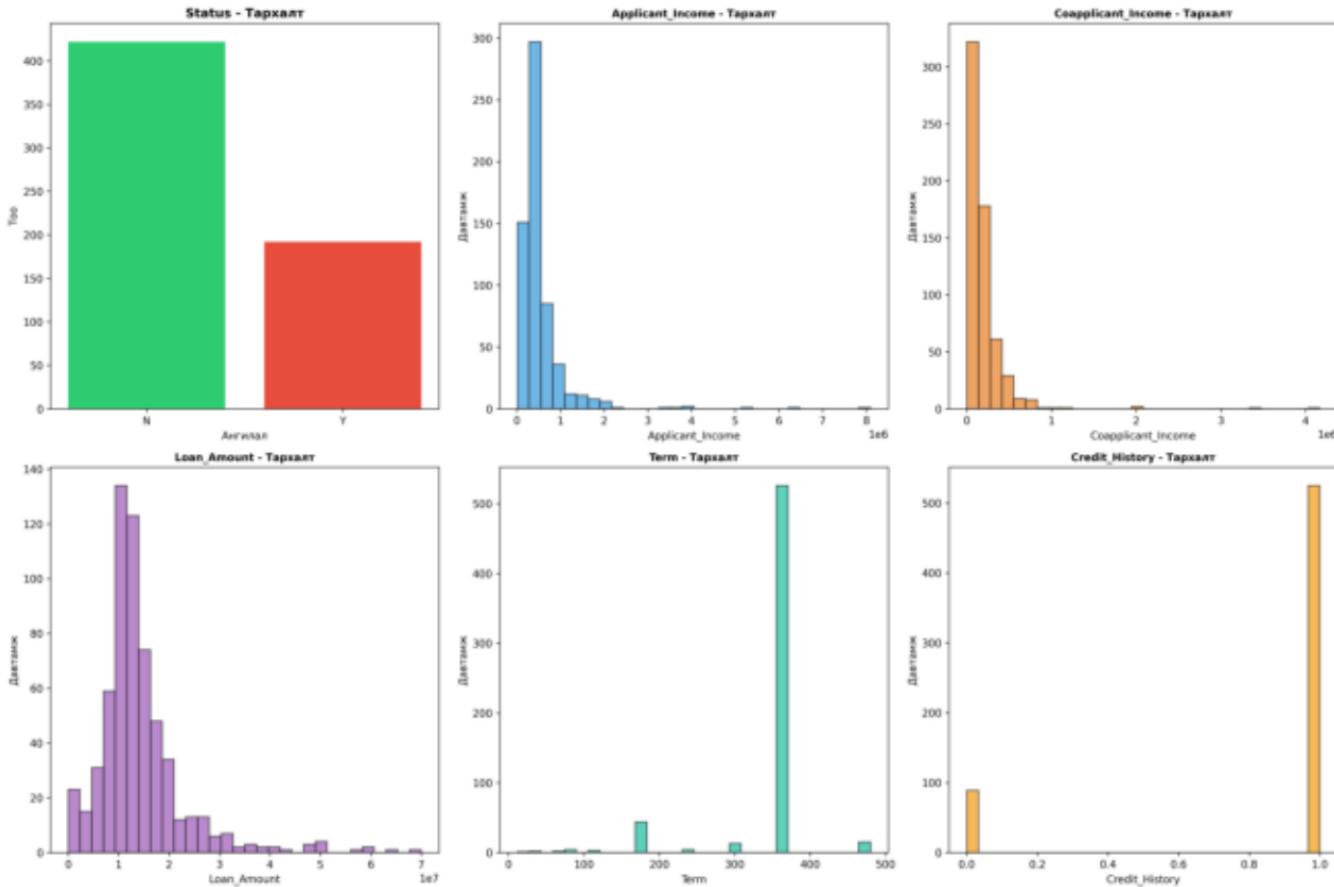
Нарийвчлал: 86.18%

F1-Score: 85.04%

ДҮГНЭЛТ:

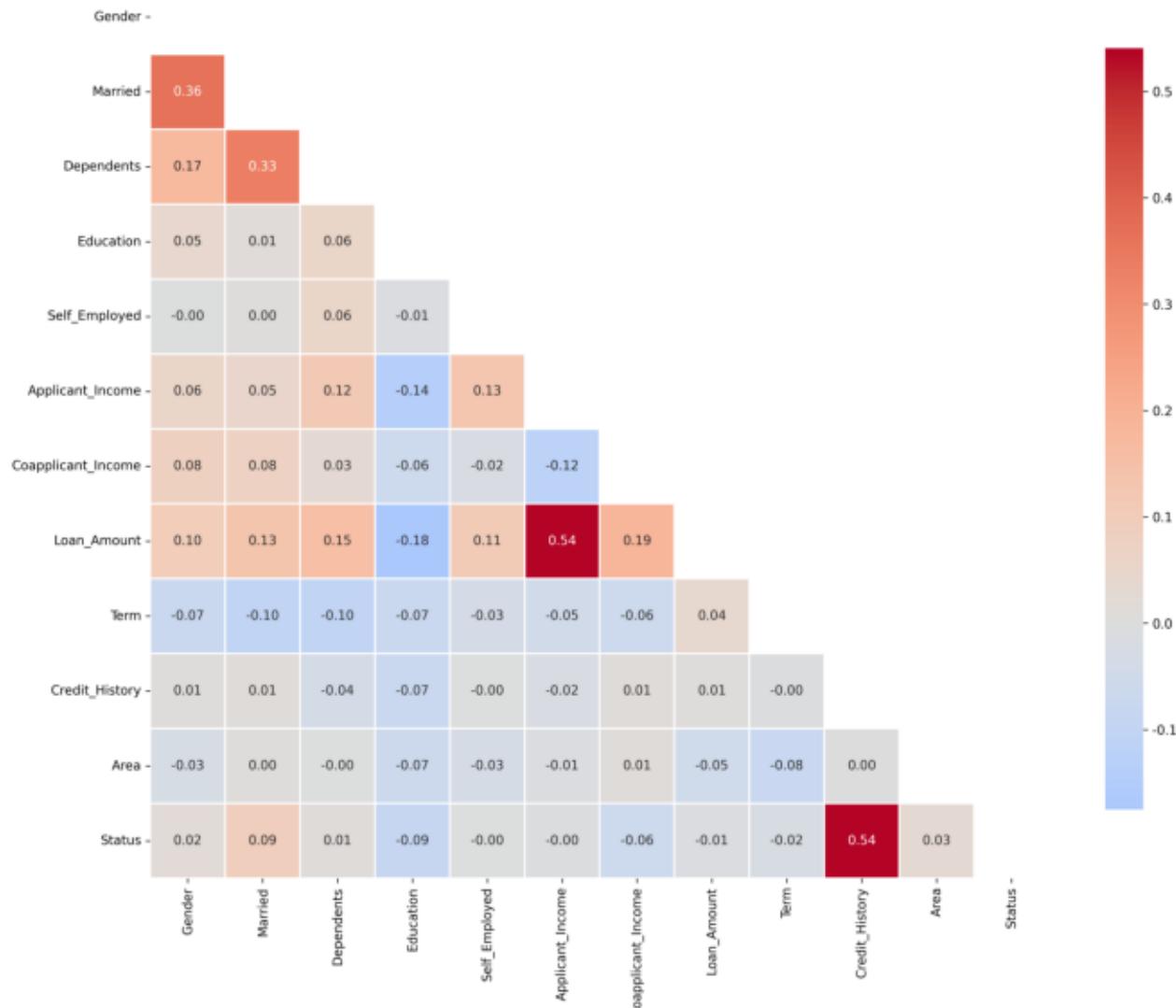
Зээлийн батламжийг 86.2% нарийвчлалаар таамаглах боломжтой болсон. Энэ нь банк, санхүүгийн байгууллагуудад зээл олгох шийдвэр гаргахад тусална.

Өгөгдлийн тархалт



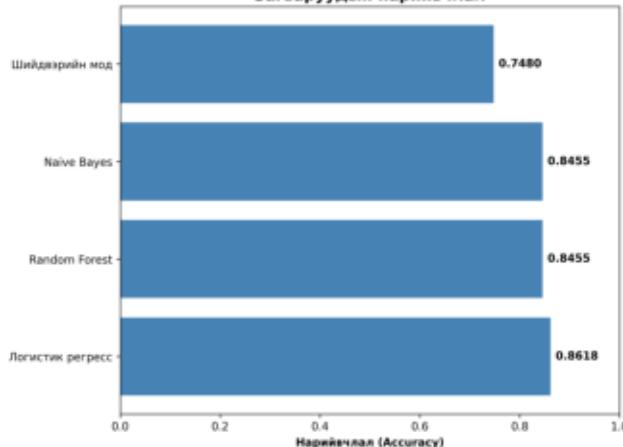
Хувьсагчдын хоорондын корреляци

Хувьсагчдын хоорондын корреляци

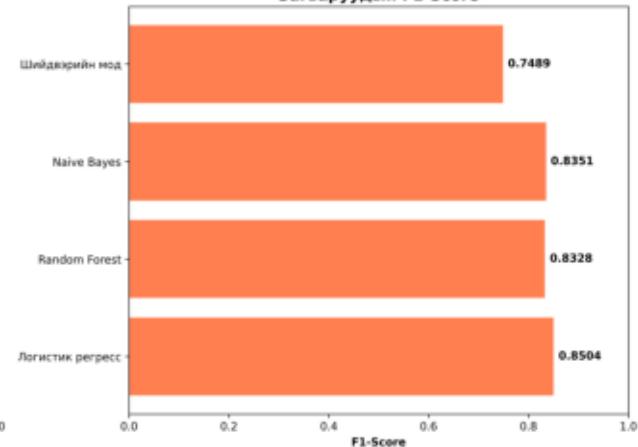


Загваруудын харьцуулалт

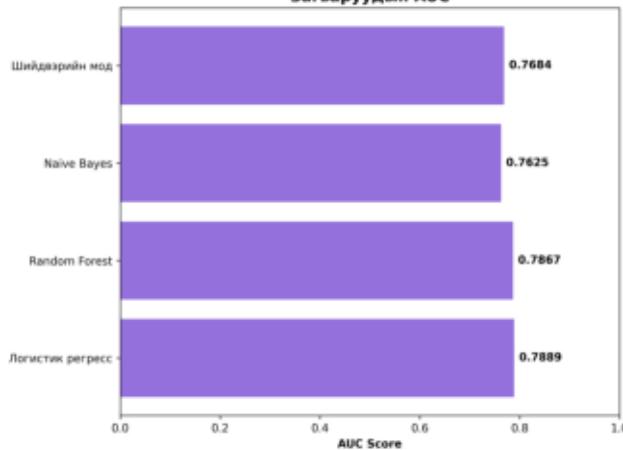
Загваруудын нарийвчлал



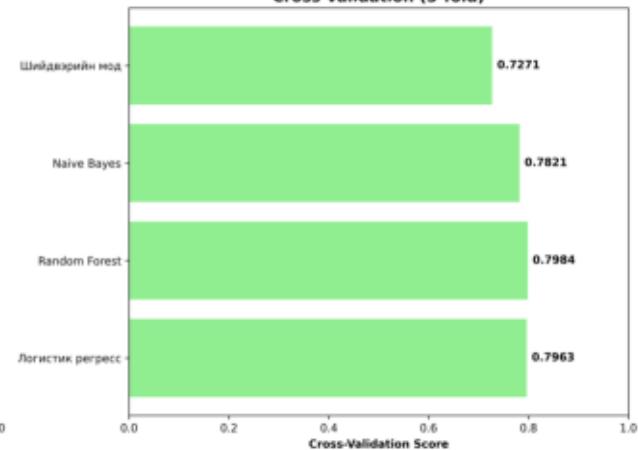
Загваруудын F1-Score



Загваруудын AUC

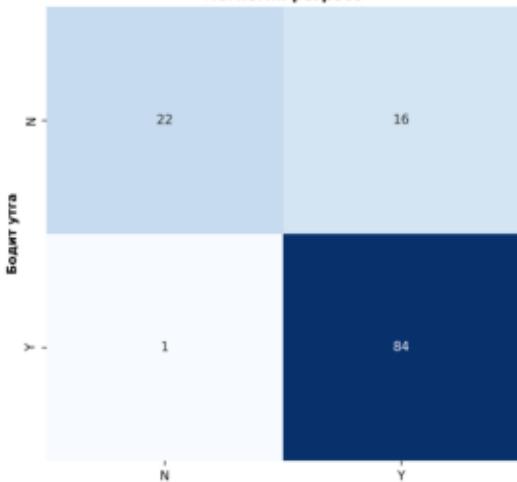


Cross-Validation (5-fold)

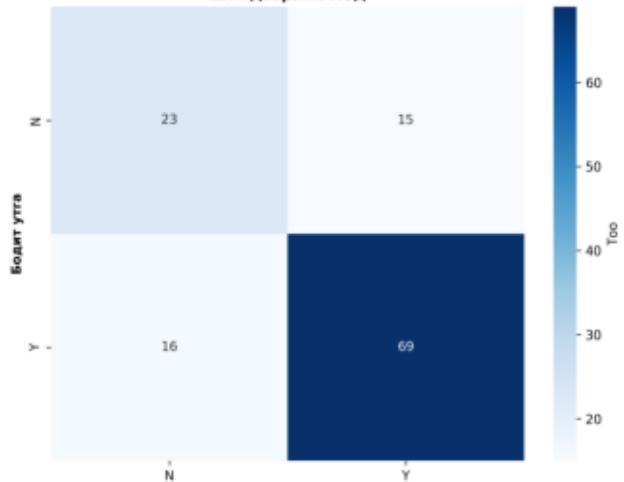


Confusion Matrices

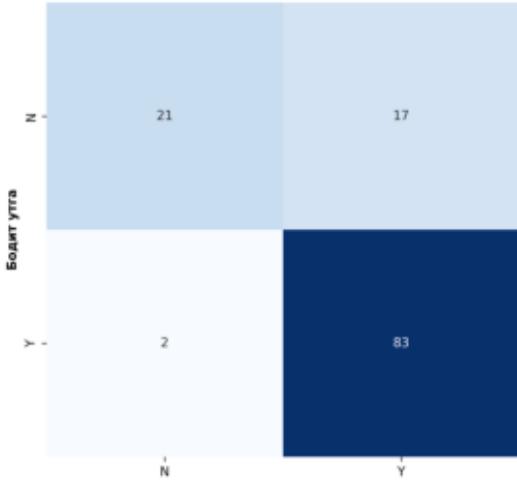
Логистик регресс



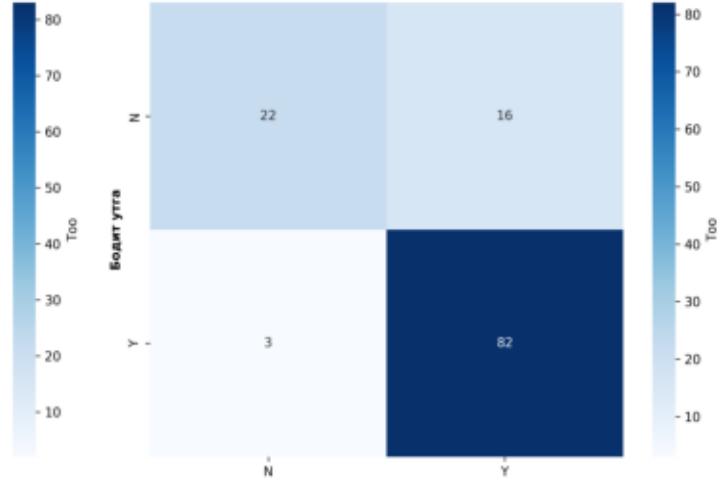
Шийдвэрийн мод



Бодлутга

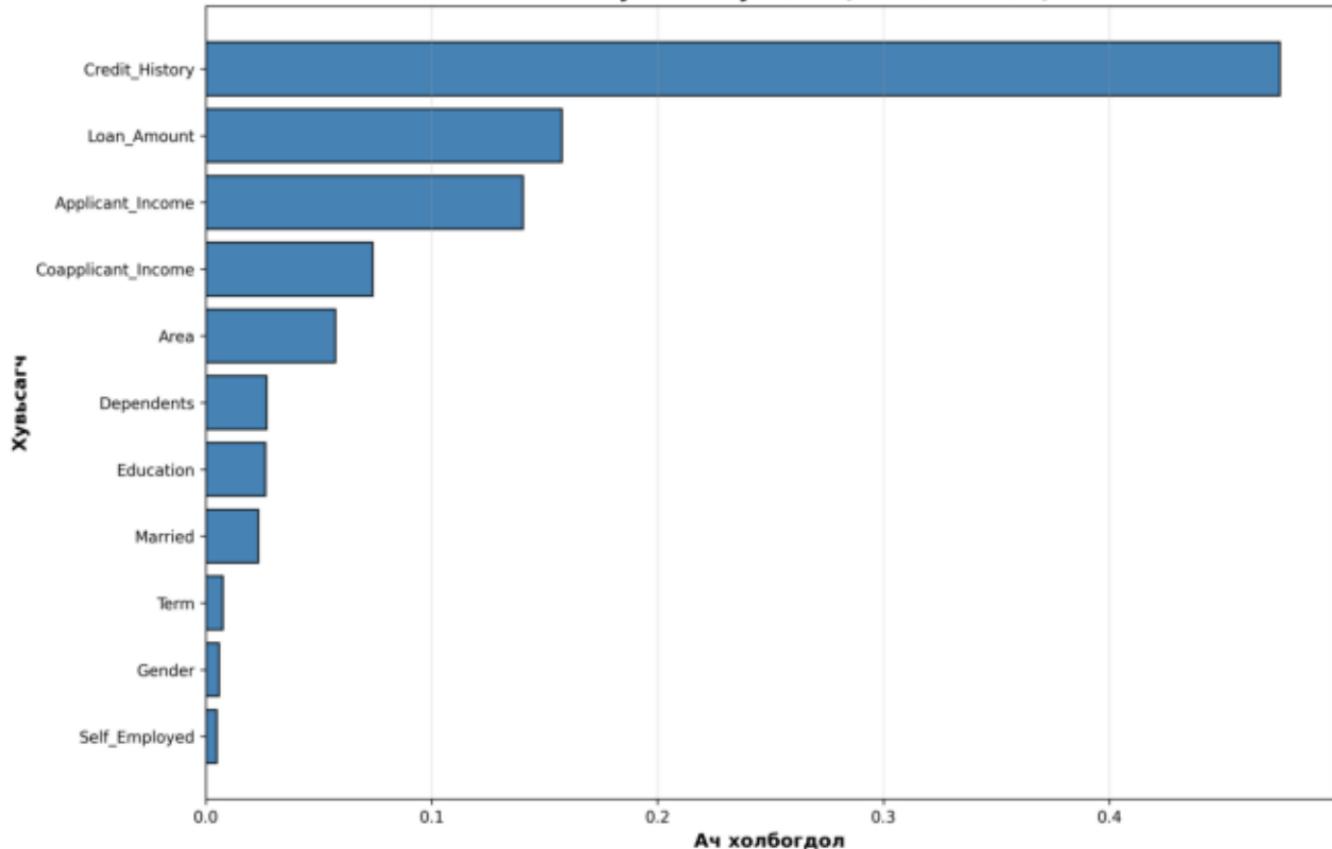


Бодлутга



Хувьсагчдын ач холбогдол

Хамгийн чухал 11 хувьсагч (Random Forest)



ЗАГВАРУУДЫН ҮР ДҮНГИЙН ХҮСНЭГТ

Загвар	Accuracy	F1-Score	AUC	CV Score
Логистик регресс	0.8618	0.8504	0.7889	0.7963 (± 0.0145)
Random Forest	0.8455	0.8328	0.7867	0.7984 (± 0.0119)
Naive Bayes	0.8455	0.8351	0.7625	0.7821 (± 0.0154)
Шийдвэрийн мод	0.7480	0.7489	0.7684	0.7271 (± 0.0407)

ДҮГНЭЛТ БА ЗӨВЛӨМЖ

1. ӨГӨГДЛИЙН МЭДЭЭЛЭЛ:

- Сургалтын өгөгдөл: 614 мөр
- Шинжийн тоо: 11
- Зорилтот хувьсагч: Status

2. ХАМГИЙН САЙН ЗАГВАР:

- Загвар: Логистик регресс
- Нарийвчлал: 86.18%
- F1-Score: 85.04%

3. ХАМГИЙН ЧУХАЛ ХУВЬСАГЧИД:

- Credit_History: 0.4756
- Loan_Amount: 0.1576
- Applicant_Income: 0.1405
- Coapplicant_Income: 0.0739
- Area: 0.0574

4. ДҮГНЭЛТ:

Зээлийн батламжийг 86.2% нарийвчлалаар таамаглах боломжтой болсон. Логистик регресс загвар хамгийн сайн үр дүн үзүүлсэн.

5. ПРАКТИКТ ХЭРЭГЛЭХ:

- Банкны зээл батлах/татгалзах шийдвэрт туслах
- Эрсдэлийн үнэлгээг автоматжуулах
- Зээлийн процессыг хурдаасгах

6. ЦААШДЫН САЙЖРУУЛАЛТ:

- Илүү олон өгөгдөл цуглуулах
- Feature engineering хийх
- Hyperparameter tuning хийх
- Ensemble методууд туршиж үзэх

АШИГЛАСАН МАТЕРИАЛЫН ЖАГСААЛТ

ЭШЛЭЛ (АРА ФОРМАТ):

- [1] Kaggle. (2024). Loan Approval Prediction Dataset.
Retrieved from <https://www.kaggle.com/datasets/>
- [2] Scikit-learn Developers. (2024). Scikit-learn: Machine Learning in Python.
Retrieved from <https://scikit-learn.org/>
- [3] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python.
Journal of Machine Learning Research, 12, 2825-2830.
- [4] McKinney, W. (2010). Data Structures for Statistical Computing in Python.
Proceedings of the 9th Python in Science Conference, 56-61.
- [5] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment.
Computing in Science & Engineering, 9(3), 90-95.
- [6] Waskom, M. (2021). seaborn: statistical data visualization.
Journal of Open Source Software, 6(60), 3021.

АШИГЛАСАН ПРОГРАМ ХАНГАМЖ:

- Python 3.8+
- pandas 1.3.0+
- numpy 1.21.0+
- matplotlib 3.4.0+
- seaborn 0.11.0+
- scikit-learn 1.0.0+

ӨГӨГДЛИЙН ЭШ СУРВАЛЖ:

Төслийн ажилд ашигласан өгөгдөл нь Kaggle платформ дээрх "Loan Approval Prediction Dataset" юм. Энэ өгөгдөл нь зээл авагчдын хувийн мэдээлэл, орлого, өр төлбөр болон бусад санхүүгийн мэдээллүүдийг агуулдаг.