іністерство освіти і науки України

Західноукраїнський національний університет

Факультет комп’ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Звіт до лабораторної роботи №7

З дисципліни “Інтелектуальний аналіз даних”

Студента групи КН-31

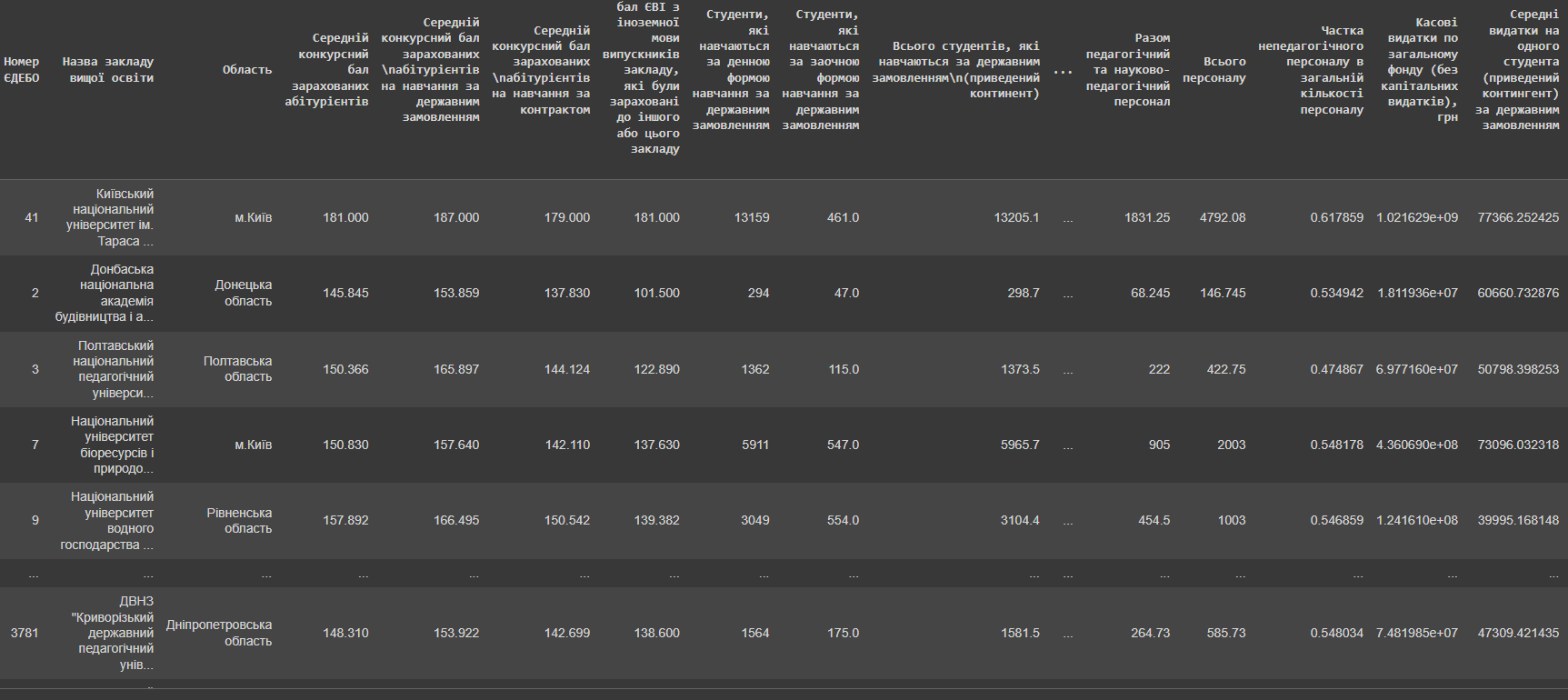
Ковальковський В. В.

Тернопіль 2023

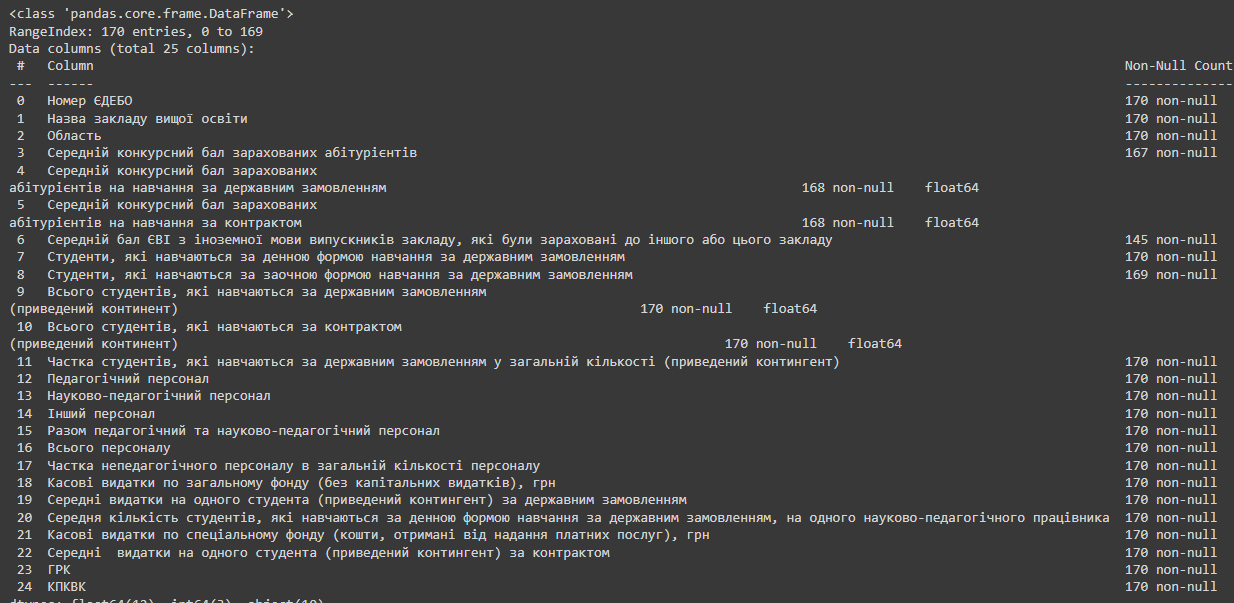
ЛАБОРАТОРНА РОБОТА 7:

Регуляризована регресія (Витрати на ВНЗ за 2019)

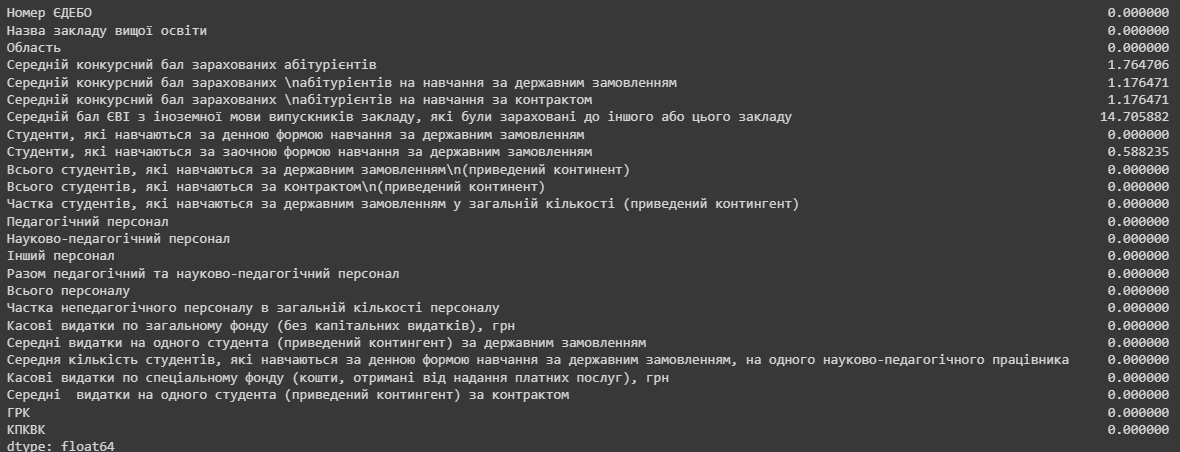
**Читання датасету:**



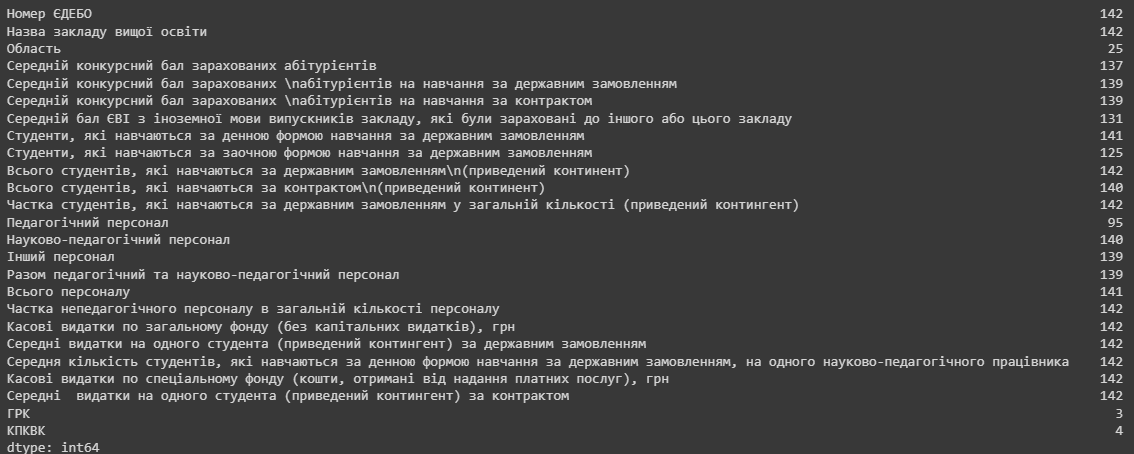
**Перевірка даних:**



**Очищення даних:**

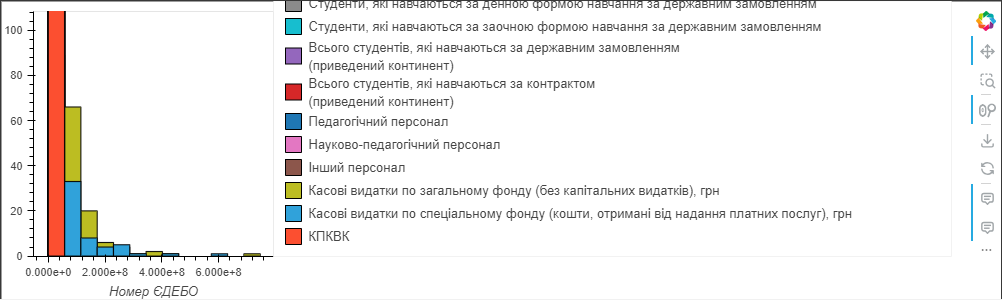


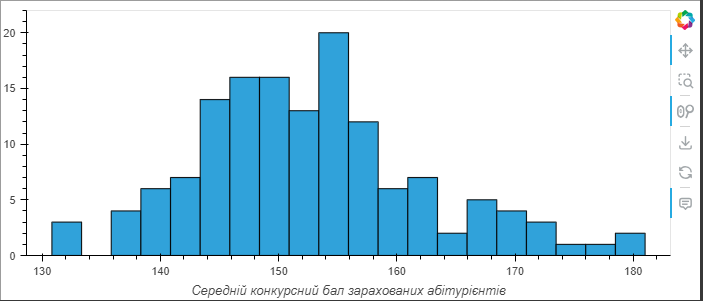
**Аналіз унікальності даних:**



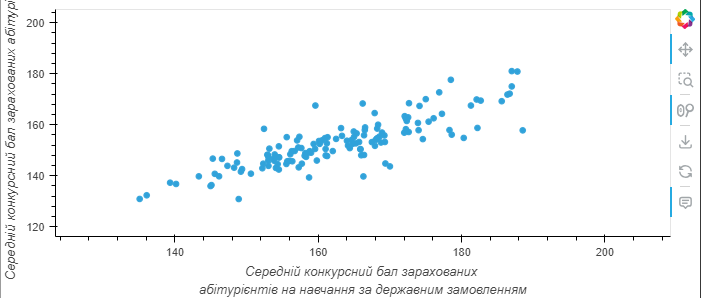
**Розвідувальний аналіз данних**

****

****

****

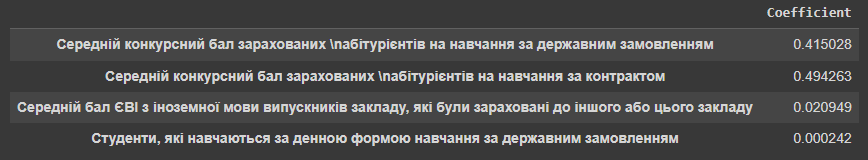
Даний графік зображає частотну залежність студентів від середнього балу.

****

З графіка видно, що середній конкурсний бал зарахованих абітурієнтів на навчання за державним замовленням у 2023 році був вищим, ніж у 2022 році. У 2023 році середній конкурсний бал становив 160 балів, а у 2022 році – 140 балів.

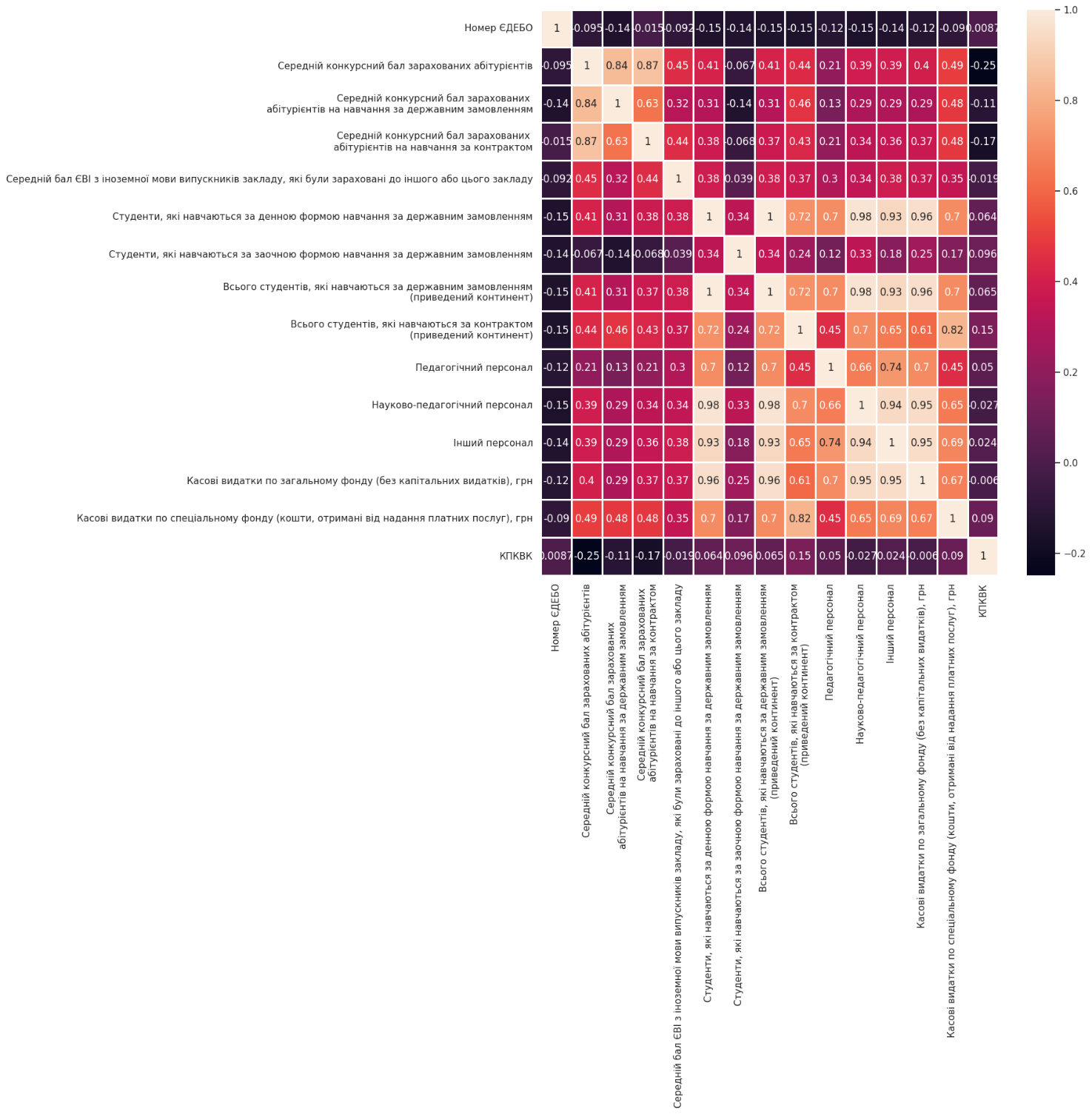
Це означає, що для вступу на навчання за державним замовленням у 2023 році абітурієнтам потрібно було мати вище середнє значення конкурсного балу, ніж у 2022 році.

**Коефіцієнти регресії:**

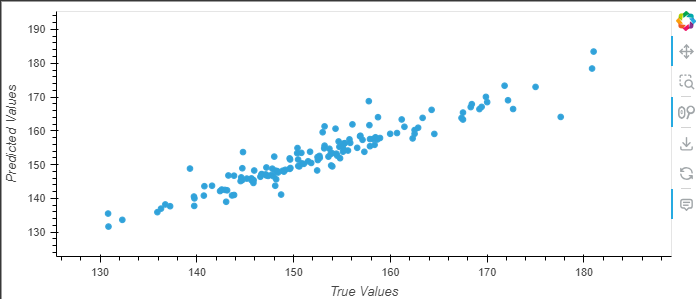
****

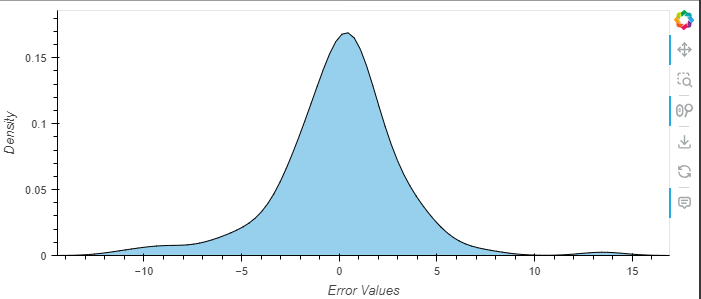
За цими коефіцієнтами, можна зробити такі висновки:

* Середній бал випускників закладу зростає на 0,415 бала за кожну одиницю підвищення середнього конкурсного балу зарахованих абітурієнтів на навчання за державним замовленням. Це означає, що заклади, які приймають абітурієнтів з високими конкурсними балами, випускають більш успішних студентів.
* Середній бал випускників закладу зростає на 0,494 бала за кожну одиницю підвищення середнього конкурсного балу зарахованих абітурієнтів на навчання за контрактом. Це означає, що заклади, які приймають абітурієнтів з високими конкурсними балами, незалежно від того, навчаються вони за державним замовленням чи за контрактом, випускають більш успішних студентів.
* Середній бал випускників закладу зростає на 0,021 бала за кожну одиницю підвищення середнього балу ЄВІ з іноземної мови випускників закладу, які були зараховані до іншого або цього закладу. Це означає, що випускники закладів, які мають високий середній бал ЄВІ з іноземної мови, мають більші шанси на успішне навчання та подальшу кар'єру.
* Середній бал випускників закладу зростає на 0,0002 бала за кожного студента, який навчається за денною формою навчання за державним замовленням. Цей коефіцієнт є дуже малим, тому його можна вважати незначним.

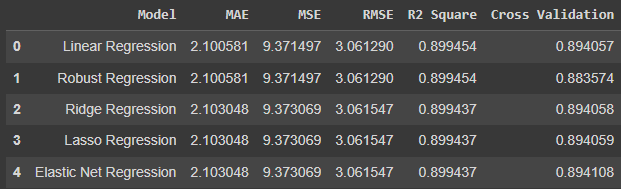
**Кореляція: **

**Прогнози моделі**

****

****

**Результат моделювання різними методами**

****

На основі даних, наведених у таблиці, можна зробити висновок, що всі п'ять моделей регресії (лінійна, robust, ridge, lasso, elastic net) забезпечують досить високу точність прогнозування.

Характеристика результатів

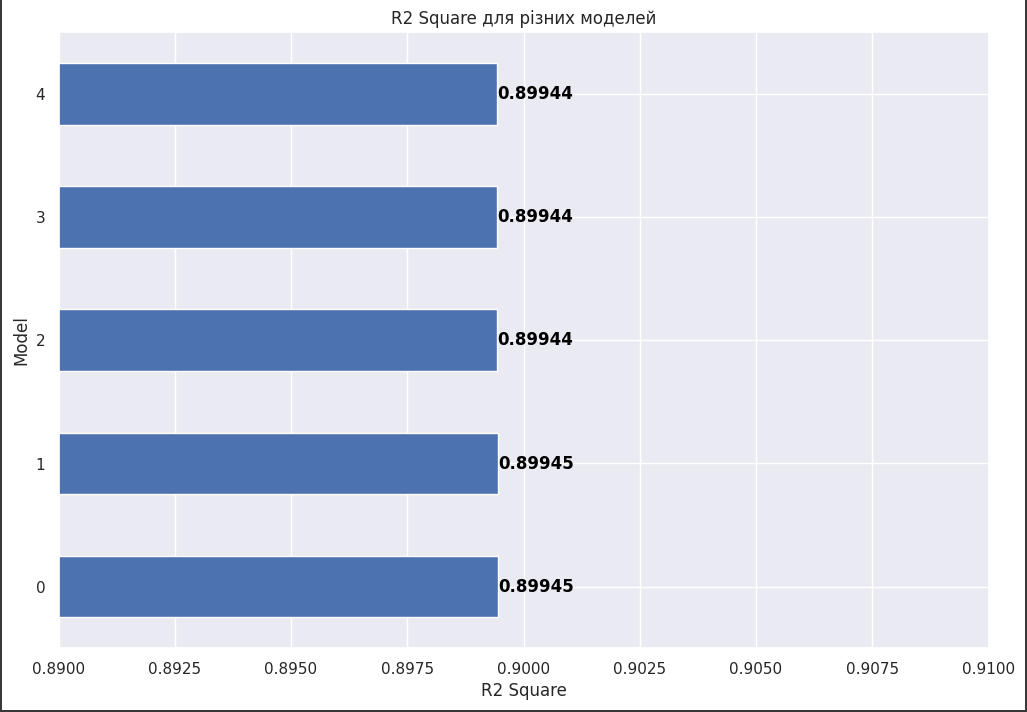
* MAE (середнє абсолютне відхилення) для всіх моделей становить близько 2,1 бала. Це означає, що середнє значення помилки прогнозу не перевищує 2,1 бала.
* MSE (середнє квадратичне відхилення) для всіх моделей становить близько 9,4 бала. Це означає, що середнє значення квадрата помилки прогнозу не перевищує 9,4 бала.
* RMSE (корень середнього квадратичного відхилення) для всіх моделей становить близько 3,1 бала. Це означає, що середнє значення помилки прогнозу не перевищує 3,1 бала.
* R2 Square (коефіцієнт детермінації) для всіх моделей становить близько 0,89. Це означає, що незалежні змінні пояснюють близько 89% дисперсії залежної змінної.
* Cross Validation (перевірка на кросовій перевірці) для всіх моделей становить близько 0,89. Це означає, що модель добре узгоджується з даними, на яких вона не навчалася.

Порівняння моделей

Якщо порівнювати результати п'яти моделей, то можна зробити такі висновки:

* Лінійна регресія забезпечує найвищу точність прогнозування, але вона є найменш стійкою до шуму в даних.
* Robust Regression забезпечує трохи меншу точність прогнозування, ніж лінійна регресія, але вона є більш стійкою до шуму в даних.
* Ridge Regression забезпечує трохи меншу точність прогнозування, ніж лінійна регресія, але вона робить це шляхом зменшення впливу великих коефіцієнтів регресії.
* Lasso Regression забезпечує трохи меншу точність прогнозування, ніж лінійна регресія, але вона робить це шляхом зменшення кількості незначних коефіцієнтів регресії.
* Elastic Net Regression забезпечує точність прогнозування, яка є приблизно такою ж, як у лінійної регресії, але вона є більш стійкою до шуму в даних.

**Порівняння R^2**

****

Висновок: у ході виконання лабораторної роботи провів ретельний розвідувальний аналіз накопичених даних, використовуючи різноманітні методи візуалізації та статистичного аналізу, виконав обчислення регресії різними методами.