## Лабораторная работа № 1

**Тема:** **«Прогнозирование данных на основе множественной линейной регрессии»**

**Цель работы:** разработка приложения для прогнозирования цен на недвижимость.

**Задачи:**

1. Загрузить данные с сайта: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>
2. Построить тренды по паре целевых переменных (взять случайные 200 значений из массива).
3. Проверить данные на нормальность распределения.
4. Задание в зависимости от варианта.
5. Отобразить среднеквадратическую ошибку (RMSE).
6. Визуализировать тренды прогнозируемых значений и реальных значений.

**Варианты заданий и ссылки на библиотеки**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | Модель | Нормализация | Cross Validation/  Масштабирование | Ссылки на библиотеки и ресурсы | Ф.И.О студентов (группы) |
| 1 | Linear Regression | да | Да/Да | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.LinearRegression.html |  |
| 2 | Linear Regression | нет | Нет/Да | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.LinearRegression.html |  |
| 3 | Linear Regression | да | Да/ Нет | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.LinearRegression.html |  |
| 4 | Linear Regression | нет | Нет/Да | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.LinearRegression.html |  |
| 5 | Lasso regression | Да | Нет/ Нет | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.Lasso.html#sklearn.linear\_model.Lasso |  |
| 6 | Lasso regression | Нет | Да/ Нет | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.Lasso.html#sklearn.linear\_model.Lasso |  |
| 7 | Ridge Regression | Да | Да/ Нет | <https://scikit-learn.org/0.21/modules/generated/sklearn.linear_model.Ridge.html#sklearn.linear_model.Ridge> |  |
| 8 | Ridge Regression | Нет | Нет/ Нет | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.Ridge.html#sklearn.linear\_model.Ridge |  |
| 8 | Ridge Regression | Нет | Да/Да | https://scikit-learn.org/0.21/modules/generated/sklearn.linear\_model.Ridge.html#sklearn.linear\_model.Ridge |  |

**Вопросы к лабораторной работе:**

* + - 1. Дать определение термину «нормального распределения»;
      2. Дать определение термину «критерий Фишера»;
      3. Дать определение термину «мультиколлинеарность»;
      4. Дать определение термину «гомоскедастичность» и «гетероскедастичность»
      5. Формула функции линейной регрессии в общем виде;
      6. Формула функции нелинейной регрессии в общем виде;
      7. Формула математического ожидания дискретной случайной величины;
      8. Формула расчёта дисперсии случайной величины;
      9. Формула функции плотности нормального распределения.
      10. Проверка значимости и валидности линейных моделей