

Прогнозирование продаж

Описание проекта

Данный проект направлен на построение модели для прогнозирования продаж на основе данных о товарах и магазинах. Модель использует линейную регрессию для предсказания количества продаж, основываясь на характеристиках товаров и времени.

Используемые технологии

- **Julia:** Язык программирования для научных и численных вычислений.
- **CSV.jl:** Библиотека для работы с CSV файлами.
- **DataFrames.jl:** Библиотека для работы с табличными данными.
- **GLM.jl:** Библиотека для построения моделей линейной регрессии.
- **Statistics.jl:** Библиотека для выполнения статистических вычислений.
- **Plots.jl:** Библиотека для построения графиков.

Шаги реализации

1. **Загрузка данных:**
 - Данные загружаются из CSV файла `sales.csv`, который содержит информацию о продажах.
1. **Предобработка данных:**
 - Заполнение пропусков в числовых данных средними значениями.
 - Выбор необходимых колонок для дальнейшего анализа (`item_id`, `store_id`, `date`, `sales`).
1. **Разделение данных:**
 - Данные делятся на обучающую (80%) и тестовую (20%) выборки для оценки производительности модели.
1. **Моделирование:**
 - Строится модель линейной регрессии с использованием библиотеки `GLM.jl`.
1. **Предсказание:**
 - Модель используется для предсказания продаж на тестовой выборке.
1. **Оценка модели:**
 - Рассчитывается среднеквадратичная ошибка (MSE) для оценки точности предсказаний.
 - Выводятся коэффициенты модели и статистическая информация о ней.
1. **Визуализация результатов:**
 - Строятся графики, показывающие сравнение фактических и предсказанных значений, а также распределение ошибок.

Результаты

- Среднеквадратичная ошибка (MSE) показывает, насколько предсказания модели отклоняются от фактических значений.
- Коэффициенты модели позволяют понять, какие факторы влияют на продажи.

- Графики визуализируют результаты и помогают интерпретировать данные.

```

Среднеквадратичная ошибка: 1.2750000000000055
Коэффициенты модели:
[16.500000000000043, -4.4000000000000155, -2.100000000000001, 1.7499999999999918, 0.3999999999999961, 3.2999999999999945, 2.2499999999999907]
Статистическая информация:

```

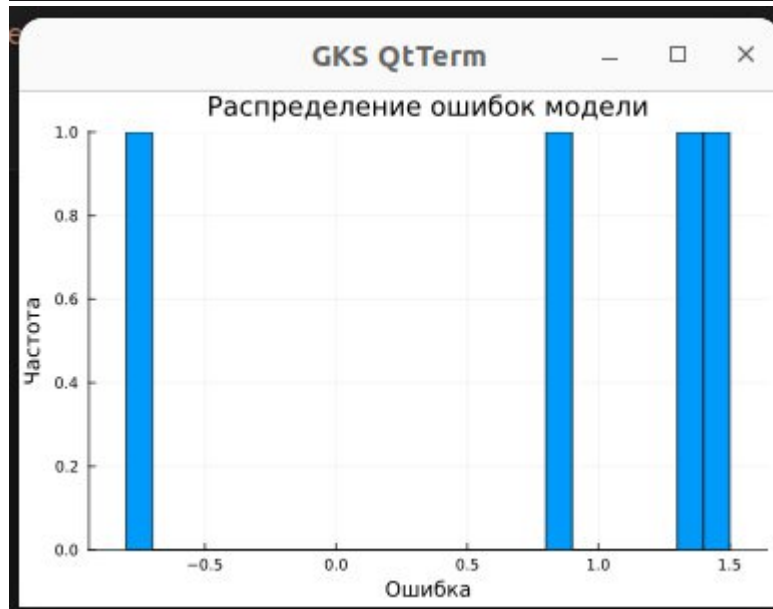
	Coef.	Std. Error	t	Pr(> t)	Lower 95%	Upper 95%
(Intercept)	16.5	1.30224	12.67	<1e-06	13.5541	19.4459
item_id	-4.4	0.58238	-7.56	<1e-04	-5.71744	-3.08256
store_id	-2.1	0.58238	-3.61	0.0057	-3.41744	-0.782565
date: 2023-01-02	1.75	0.785281	2.23	0.0528	-0.0264296	3.52643
date: 2023-01-03	0.4	1.31403	0.30	0.7677	-2.57254	3.37254
date: 2023-01-04	3.3	0.860233	3.84	0.0040	1.35402	5.24598
date: 2023-01-05	2.25	0.785281	2.87	0.0186	0.47357	4.02643

```

R-квадрат: 0.8807017543859644

```

-



-

Запуск проекта

1. Убедитесь, что у вас установлен Julia.
2. Установите необходимые библиотеки, если они не установлены:
3. `julia`
4. Копировать код
5. `using Pkg`
6. `Pkg.add("CSV")`
7. `Pkg.add("DataFrames")`
8. `Pkg.add("GLM")`
9. `Pkg.add("Statistics")`
10. `Pkg.add("Plots")`
- 11.
12. Сохраните код в файл `main.jl`.
13. Запустите код в Julia REPL или VS Code.

Заключение

Этот проект демонстрирует, как можно использовать линейную регрессию для прогнозирования продаж, анализируя данные и визуализируя результаты. Вы можете расширять его, добавляя новые признаки или используя другие методы машинного обучения для улучшения точности.