

Министерство образования Республики Беларусь
Учреждение образования
«Брестский Государственный технический университет»
Кафедра ИИТ

Лабораторная работа №2
По дисциплине «Основы машинного обучения»
Тема: **«Линейные модели
для задач регрессии и классификации»**

Выполнил:
Студент 3 курса
Группы АС-65
Хвисюк К. Г.
Проверил:
Крощенко А. А.

Цель: изучить применение линейной и логистической регрессии для решения практических задач. Научиться обучать модели, оценивать их качество с помощью соответствующих метрик и интерпретировать результаты.

Вариант 9

- **Регрессия (Прогнозирование цены автомобиля)**

1. Car Price Prediction

2. Предсказать цену автомобиля (price)

3. Задания:

- загрузите данные. Выберите 5-6 (например, horsepower, citympg, enginesize);
- обучите модель линейной регрессии;
- рассчитайте R2 и MAE;
- визуализируйте зависимость price от horsepower с линией регрессии.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error

df = pd.read_csv("CarPrice_Assignment.csv")

features = ['horsepower', 'citympg', 'enginesize', 'curbweight', 'carwidth',
            'carlength']

X = df[features]
y = df['price']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

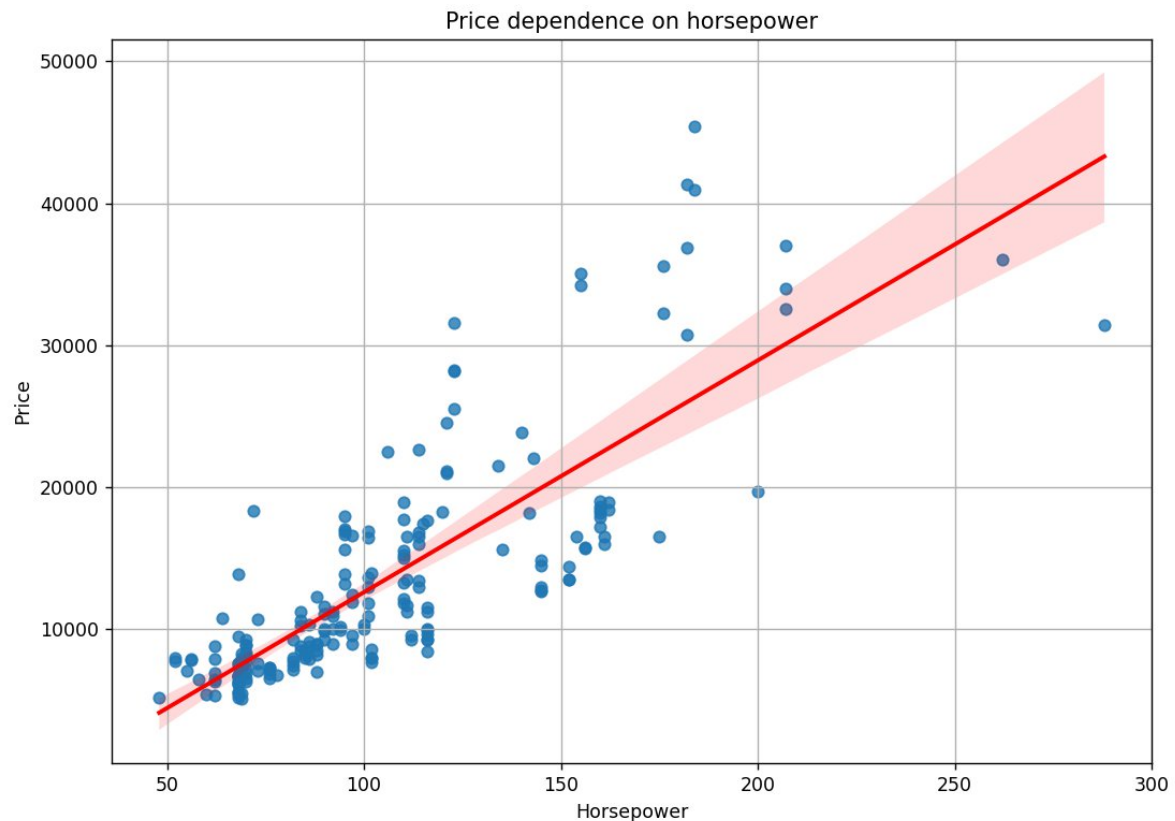
y_pred = model.predict(X_test)

r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)

print(f"R^2: {r2:.3f}")
print(f"MAE: {mae:.2f}")

plt.figure(figsize=(10, 7))
sns.regplot(x=df['horsepower'], y=df['price'], line_kws={"color": "red"})
plt.title("Price dependence on horsepower")
plt.xlabel("Horsepower")
plt.ylabel("Price")
plt.grid(True)
plt.show()
```

R^2 : 0.814
MAE: 2772.52



- **Классификация (Детекция мошеннических транзакций)**

1. Credit Card Fraud Detection

2. Определить, является ли транзакция мошеннической (Class = 1)

3. Задания:

- загрузите данные. Обратите внимание на сильный дисбаланс классов;
- обучите модель логистической регрессии;
- рассчитайте Accuracy, Precision и Recall. Объясните, почему Accuracy в данном случае не является показательной метрикой;
- постройте матрицу ошибок.

```
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score,
confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("creditcard.csv")
```

```
X = df.drop(columns=['Class'])
y = df['Class']
```

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42, stratify=y)

model = LogisticRegression(max_iter=12345)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print(f"Accuracy: {accuracy:.3f}")
print(f"Precision: {precision:.3f}")
print(f"Recall: {recall:.3f}")

cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=["Not scam",
"Scam"], yticklabels=["Not scam", "Scam"])
plt.xlabel("Predicted")
plt.ylabel("True")
plt.show()

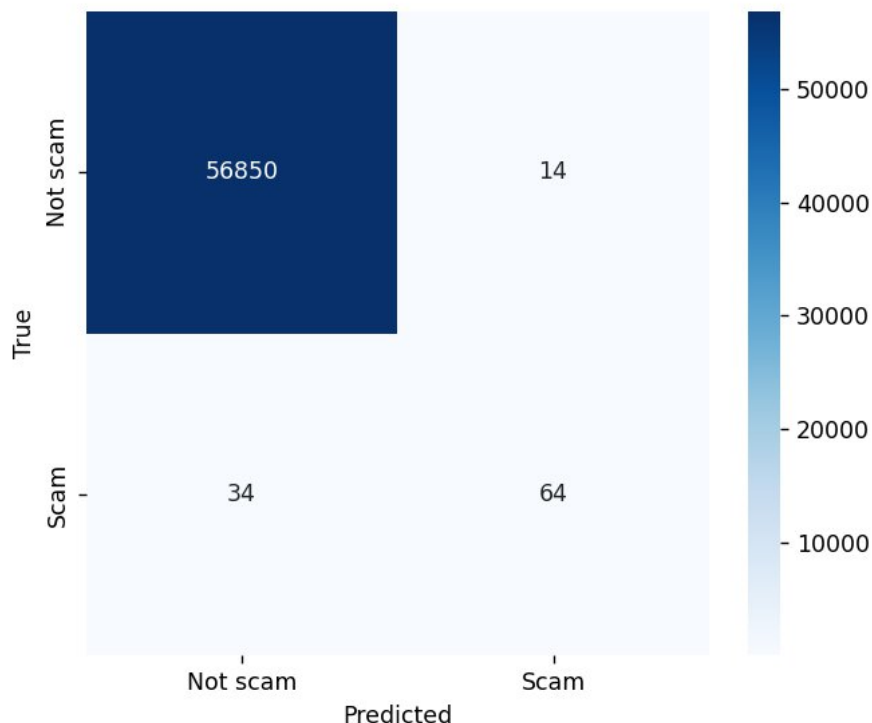
```

```

Accuracy: 0.999
Precision: 0.821
Recall: 0.653

```

Ассурасу не является показательной метрикой при дисбалансе классов, так как большинство транзакций - не мошеннические. Модель может просто предсказывать '0' и получать высокий Accuracy.



Вывод: изучили применение линейной и логистической регрессии для решения практических задач. Научились обучать модели, оценивать их качество с помощью соответствующих метрик и интерпретировать

результаты.