

Тимошенко Є. С.,
здобувач вищої освіти спеціальності 122 Комп'ютерні науки
Науковий керівник: **Пархоменко О. Ю.,**
канд.фіз.-мат.наук., доцент кафедри економічної кібернетики,
комп'ютерних наук та інформаційних технологій
Миколаївський національний аграрний університет, м. Миколаїв

СИСТЕМА ПРОГНОЗУВАННЯ ЦІН НА СІЛЬСЬКОГОСПОДАРСЬКУ ПРОДУКЦІЮ З ВИКОРИСТАННЯМ PYTHON-БІБЛІОТЕКИ SCIKIT-LEARN

В дослідженні розглянуто створення системи прогнозування цін на сільськогосподарську продукцію і робота з нею в Python бібліотеці scikit-learn.

Бібліотека scikit-learn – це безплатна бібліотека машинного навчання, написана на Python. Вона надає широкий вибір алгоритмів навчання з учителем і без нього. Одна з основних переваг бібліотеки полягає в тому, що вона працює на основі декількох поширених математичних бібліотек і легко інтегрує їх одна з одною. Ще однією перевагою є широка спільнота й докладна документація. Scikit-learn спеціалізується на алгоритмах машинного навчання для вирішення задач навчання з учителем: класифікації й регресії, а також для завдань навчання без учителя: кластеризації, зменшення розмірності й детектування аномалій.

Прогнозування цін на сільськогосподарську продукцію є ключовим фактором для прийняття стратегічних рішень у галузі сільського господарства. Ефективне передбачення майбутніх цін допомагає фермерам планувати посіви, оптимізувати витрати та максимізувати прибутки. Однак, ціни на сільськогосподарську продукцію залежать від багатьох факторів, таких як попит, пропозиція, погодні умови, політичні рішення та глобальні тенденції, що робить завдання прогнозування складним.

Саме за допомогою таких алгоритмів регресії, кластеризації, розмірності та детектування аномалій ми можемо легко прогнозувати ціни на сільськогосподарську продукцію, використовуючи бібліотека scikit-learn за таким алгоритмом:

- збір та підготовка даних;
- вибір та налаштування моделі машинного навчання;
- тренування та оцінка моделі;
- розгортання та використання моделі.

Розглянемо можливі коди реалізації деяких етапів роботи з бібліотекою.

1. Підготовка даних

```
pythonCopy codeimport pandas as pd
# Завантаження історичних даних
data = pd.read_csv('crop_prices.csv')
# Обробка відсутніх значень
data = data.dropna()
# Розділення даних на вхідні ознаки (X) та цільову змінну (y)
X = data.drop('price', axis=1)
```

```

y = data['price']
# Розділення даних на навчальний та тестовий набори
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

```

2. Тренування моделі випадкового лісу

```

pythonCopy codefrom sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
# Створення та тренування моделі
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
# Оцінка точності моделі на тестовому наборі
y_pred = rf_model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f'Mean Squared Error: {mse:.2f}')
print(f'R-squared: {r2:.2f}')

```

3. Прогнозування цін

```

pythonCopy code# Нові дані для прогнозування
new_data = [[...]] # Вхідні ознаки
# Прогнозування ціни
predicted_price = rf_model.predict(new_data)
print(f'Прогнозована ціна: {predicted_price[0]:.2f}')

```

Створення системи прогнозування цін на сільськогосподарську продукцію з використанням бібліотеки `scikit-learn` у Python є ефективним рішенням для аналізу та передбачення майбутніх цін. Бібліотека `scikit-learn` пропонує різноманітні алгоритми машинного навчання, такі як лінійна регресія, дерева рішень, випадковий ліс, градієнтний бустинг та нейронні мережі, які можна застосовувати для вирішення завдань регресії та прогнозування цін.

Переваги бібліотеки `scikit-learn`:

1. Бібліотека має зрозумілий та стандартизований інтерфейс для різних алгоритмів машинного навчання, що полегшує процес навчання та використання.

2. `Scikit-learn` пропонує велику кількість алгоритмів для класифікації, регресії, кластеризації, зменшення розмірності даних та інших задач машинного навчання.

3. Бібліотека оптимізована для обробки великих наборів даних і забезпечує швидке виконання алгоритмів.

4. `Scikit-learn` тісно інтегрована з бібліотеками `NumPy` та `SciPy`, що полегшує обробку та маніпуляції з даними.

5. Бібліотека є відкритою та безкоштовною, що сприяє її активному розвитку та підтримці спільнотою.

6. `Scikit-learn` має детальну документацію, приклади коду та активну спільноту користувачів.

Недоліки бібліотеки `scikit-learn`:

1. Scikit-learn зосереджена на класичних алгоритмах машинного навчання і має обмежену підтримку для глибоких нейронних мереж.

2. Бібліотека не підтримує розподілене навчання на кластерах або в хмарних середовищах.

3. Scikit-learn в основному орієнтована на навчання з використанням статичних наборів даних, а не потоків даних.

4. Бібліотека не має вбудованих інструментів для візуалізації даних та результатів моделей. Для візуалізації необхідно використовувати сторонні бібліотеки, такі як Matplotlib або Seaborn.

5. Scikit-learn не має вбудованих інструментів для автоматичного вибору та налаштування найкращої моделі на основі даних.

6. Процес налаштування гіперпараметрів моделей може бути трудомістким і вимагати значних ресурсів та експериментів.

Ключовими етапами у створенні системи прогнозування є збір та підготовка даних, вибір та налаштування відповідної моделі машинного навчання, тренування та оцінка моделі з використанням метрик точності, таких як середньоквадратична помилка та R-квадрат. Після тренування, модель можна інтегрувати в систему прогнозування цін та використовувати для передбачення майбутніх цін на основі нових вхідних даних.

Важливим аспектом є моніторинг та періодичне оновлення моделі, оскільки ринкові тенденції та фактори впливу на ціни можуть змінюватися з часом. Регулярне переналаштування моделі з урахуванням нових даних дозволить забезпечити більш точні та актуальні прогнози.

Загалом, застосування Python та бібліотеки scikit-learn для створення системи прогнозування цін на сільськогосподарську продукцію є перспективним підходом, який може надати фермерам та агропромисловим підприємствам важливу інформацію для прийняття стратегічних рішень, оптимізації витрат та підвищення рентабельності виробництва.

Список використаних джерел

1. Інтелектуальна інформаційна система прогнозування попиту на сільськогосподарську продукцію. Юр Р.Ю. – Кваліфікаційна робота бакалавра. Кафедра інформаційних технологій – Дубляни, ЛНУП, 2023. 50 с.

URI: <https://repository.lnup.edu.ua/jspui/handle/123456789/650>

2. <https://dou.ua/lenta/articles/first-steps-in-nlp-scikit-learn/>

3. <https://uk.wikipedia.org/wiki/Scikit-learn>

4. http://cloud-5.bitp.kiev.ua/?page_id=640