



Національний технічний університет України “КПІ”
“Інститут прикладного системного аналізу”

Дипломна робота

тема:

*Прогноз попиту на харчову продукцію методом
екстремального градієнтного бустінгу*

Виконала: студентка групи КА-41
Снігірьова Валерія Юріївна

Київ 2018

Науковий керівник: д.ф.-м.н., професор,
Богданський Юрій Вікторович

Об'єкт, предмет і мета дослідження

- **Об'єкт дослідження** – нелінійні динамічні процеси, що характеризують попит на .
- **Мета дослідження** – розробка системи підтримки прийняття рішень та її застосування для побудови адекватних моделей вибраних процесів, оцінювання прогнозів і порівняльний аналіз альтернативних методів прогнозування.
- **Предмет дослідження** – методи та інформаційні технології математичного моделювання і прогнозування нелінійних динамічних процесів, що характеризують попит на споживчі товари.

Актуальність дослідження

- Задача прогнозування попиту є критично важливою для кожного підприємства, що займається роздрібною торгівлею, оскільки є невід'ємною частиною планування закупівлі та використання ресурсів.
- Якщо буде закуплено замало товарів, це може призвести при великому попиті до нестачі товарів на полицях та недоотриманого прибутку.
- Забагато закуплених товарів потребує відповідних витрат на зберігання та при недостатньому попиті зайві товари можуть зіпсуватися, що спричинить збитки для компанії.

Постановка задачі дослідження

1. **Вибрати** та дослідити історичні **дані** стосовно попиту на споживчі товари, отримані з **діючих підприємств**.
2. **Розробити методiku** побудови **моделей** і прогнозування нелінійних динамічних процесів, а також оцінювання їх короткострокових та довгострокових **прогнозів**.
3. **Вибрати методи** оцінювання параметрів моделей НДП.
4. **Розробити** програмне забезпечення для виконання обчислювальних експериментів – **системи** для моделювання і прогнозування попиту.
5. **Виконати** і проаналізувати результати обчислювальних **експериментів** з використанням фактичних даних; зробити порівняльний **аналіз** результатів, отриманих за допомогою альтернативних методів.

Перелік використаних методів та моделей прогнозування

Методи регресійного аналізу:

- GLMNET

Методи дерев рішень:

- XGBOOST

Нейронні мережі:

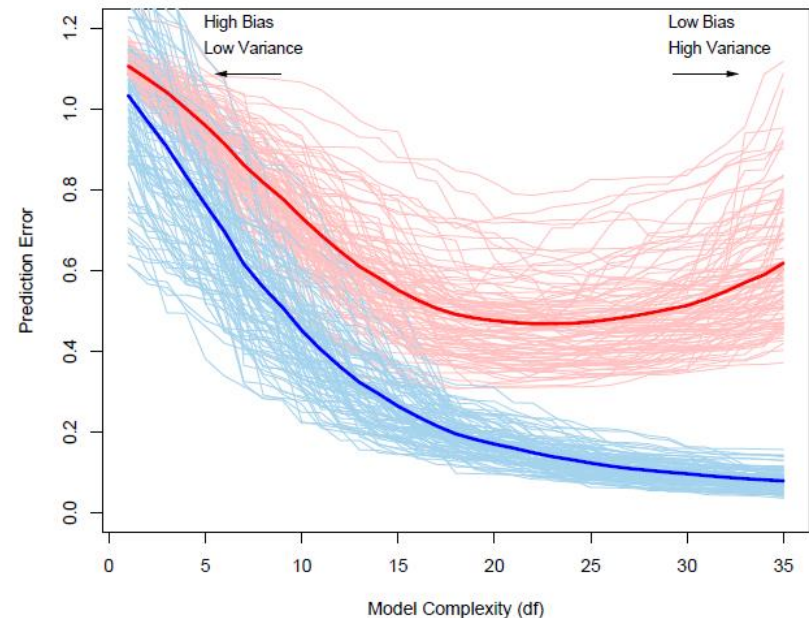
- Neuralnet

Навчання з учителем

Емпіричний ризик

$$\begin{aligned} R_{emp} &= \mathbf{E}[L(y_i, g(x_i))] \\ &= \frac{1}{N} \sum_i L(y_i, g(x_i)) \\ &= \frac{1}{N} \sum_i (y_i - g(x_i))^2 \end{aligned}$$

Структурний ризик



Математичні моделі для прогнозування вибраних процесів

1. Методи лінійної регресії

Пошук апроксимуючої функції серед класу:

$$\hat{f}(X) = X\theta = \hat{y}$$

Функція похибки:

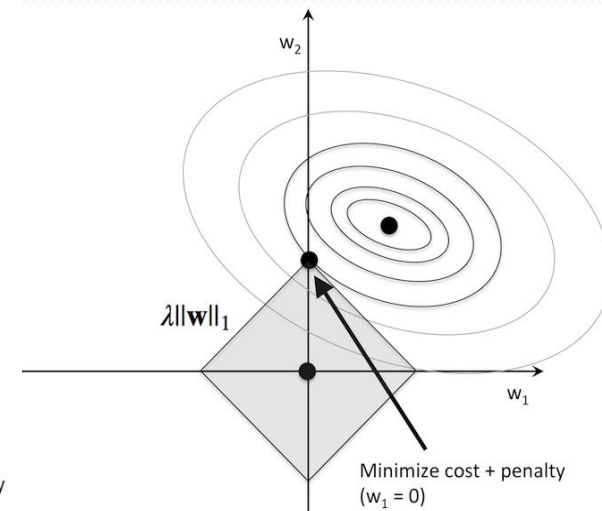
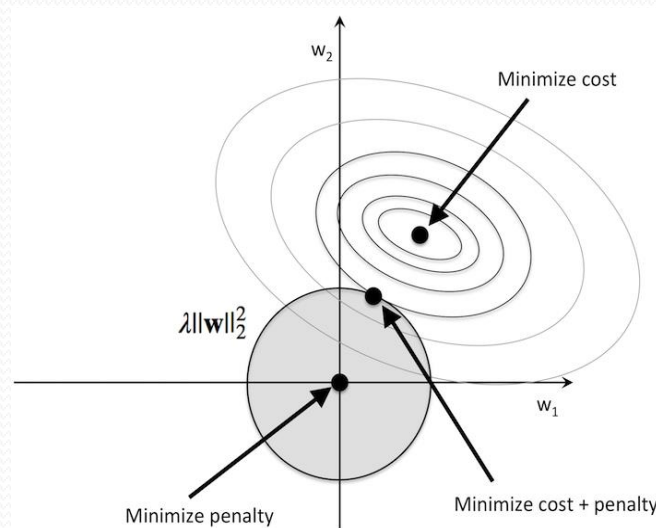
$$J(\theta) = (y - X\theta)^T (y - X\theta)$$

Оцінка параметрів:

$$\theta = (X^T X)^{-1} X^T y$$

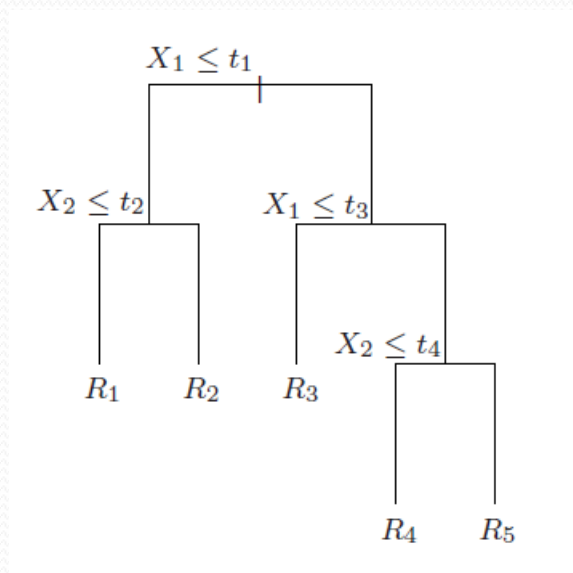
Регуляризація

$$J(\theta) = (y - X\theta)^T (y - X\theta) + \delta^2 \|\theta\|_p$$



Математичні моделі для прогнозування вибраних процесів

2. Дерева рішень



$$T(x) = c_j I_{(x \in R_j)}$$

Визначення розбиття

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

Критерій вартості складності

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T|$$

Математичні моделі для прогнозування вибраних процесів

3. Екстремальний градієнтний бустінг

- Поеднання результатів декількох "слабких" класифікаторів для створення потужного "ансамблю".

$$\hat{f}(x) = \sum_{i=0}^M \hat{f}_i(x)$$

- Пошук $\hat{f}$ алгоритмом градієнтного бустінгу

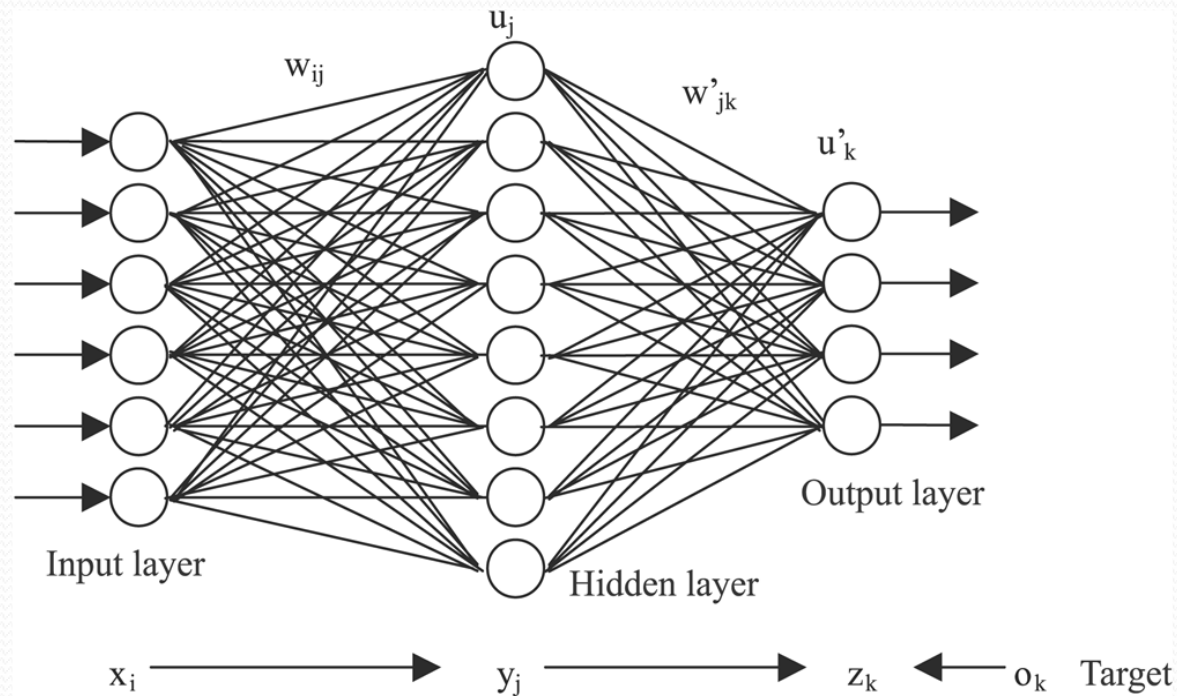
$$\widehat{f_m} = \widehat{f_{m-1}} + h_m, \quad h_m = -\rho_m g_m \quad g_{mi} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}, \quad \rho_m = \arg \min_{\rho} L(f_{m-1} - \rho g_m),$$

- Емпірично виведена регуляризаційна складова функції вартості та підтримка будь-якої двічі диференційовної функції похибки

$$J(m)^* = -\frac{1}{2} \sum_{j=1}^T \frac{Q_j^2}{D_j + \lambda} + \gamma |T|.$$

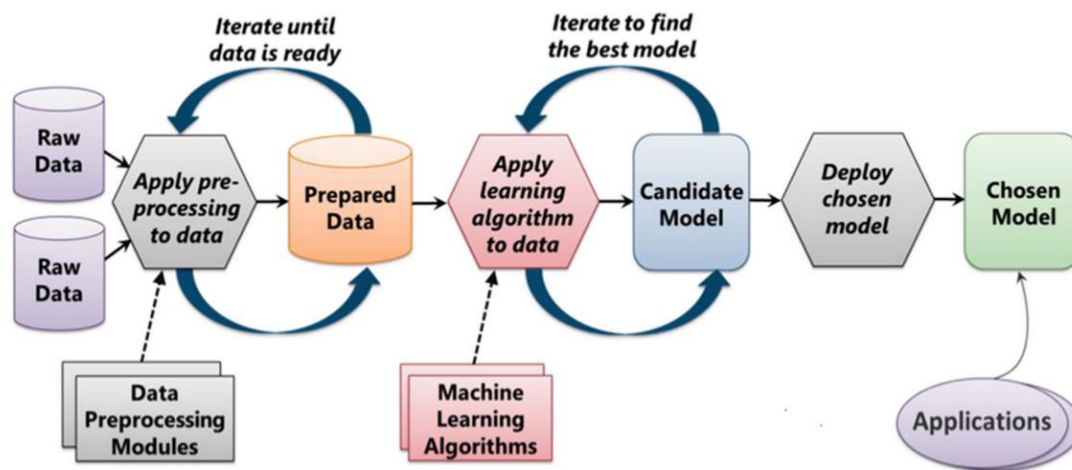
Математичні моделі для прогнозування вибраних процесів

3. Нейронні мережі



Архітектура системи для прогнозування попиту на споживчі товари

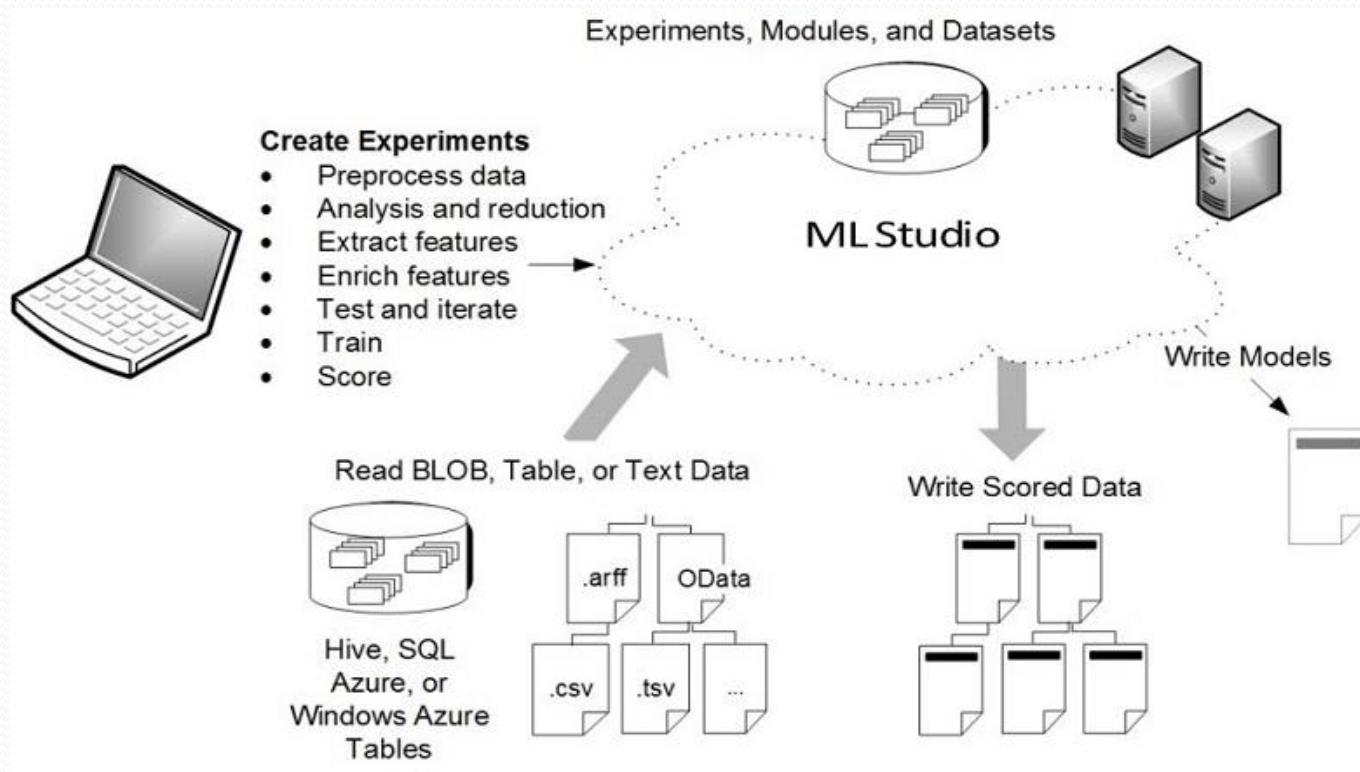
The Machine Learning Process



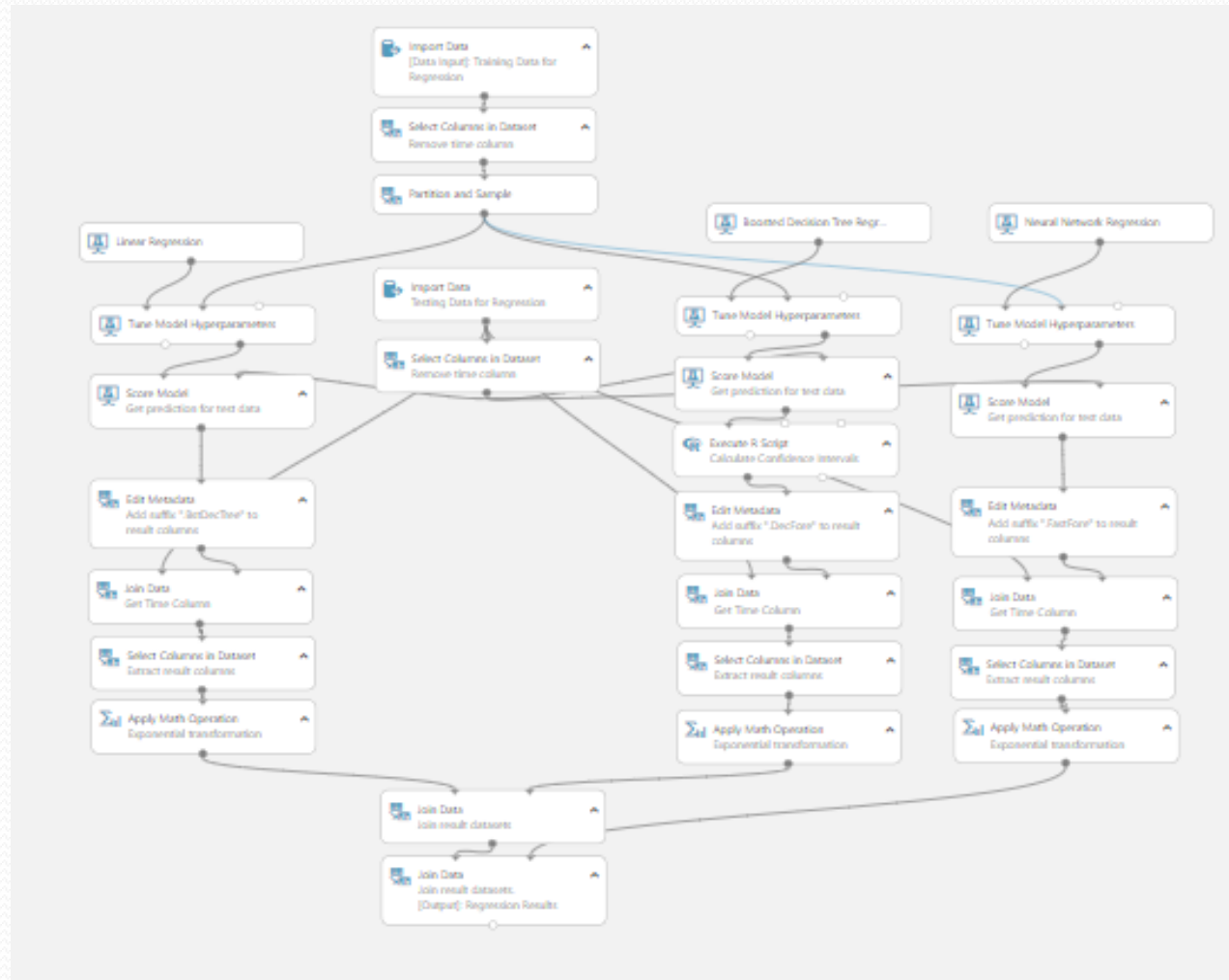
From "Introduction to Microsoft Azure" by David Chappell

Призначення: аналіз даних; моделювання; прогнозування; прийняття рішень

Архитектура Azure ML Studio

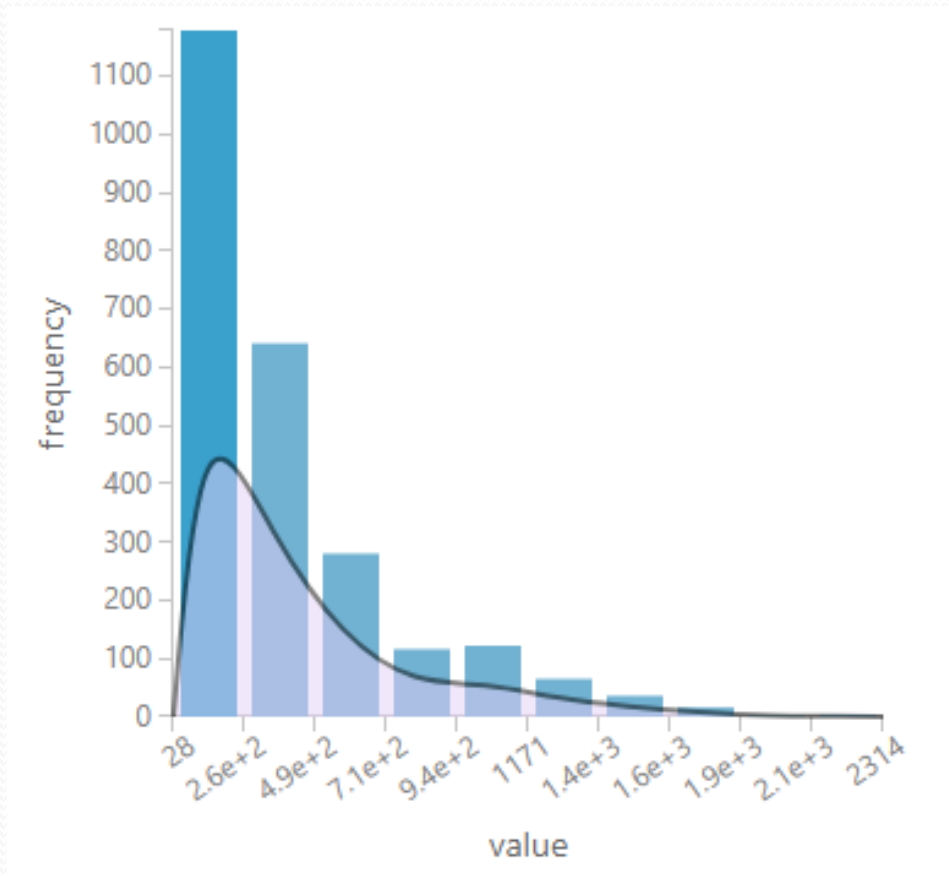


Остаточный эксперимент



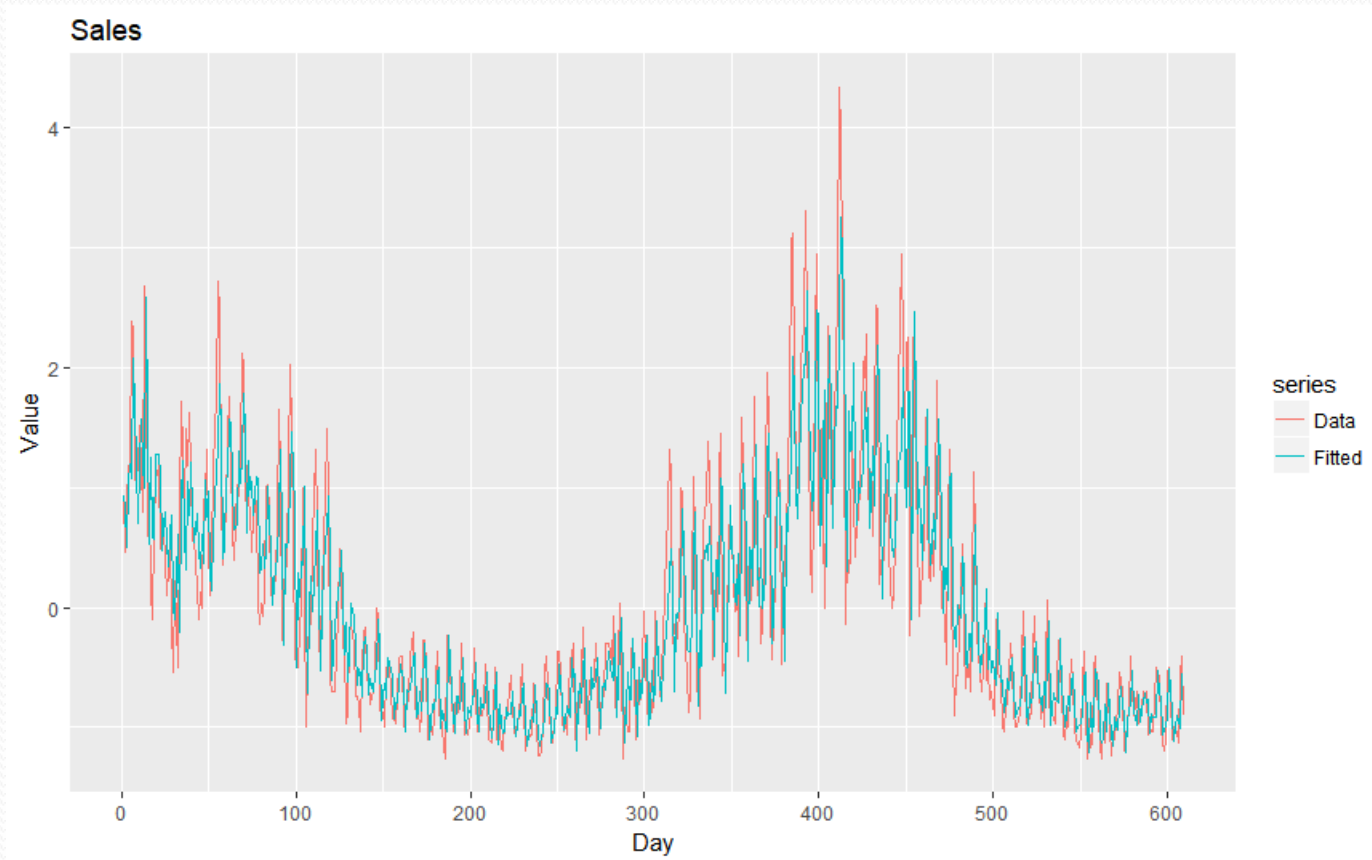
Моделювання попиту на харчову продукцію

Для побудови моделей візьмемо щотижневі дані з продажу харсової продукції одного з магазинів деякого українського підприємства. Розподіл продажів зображено на рисунку.



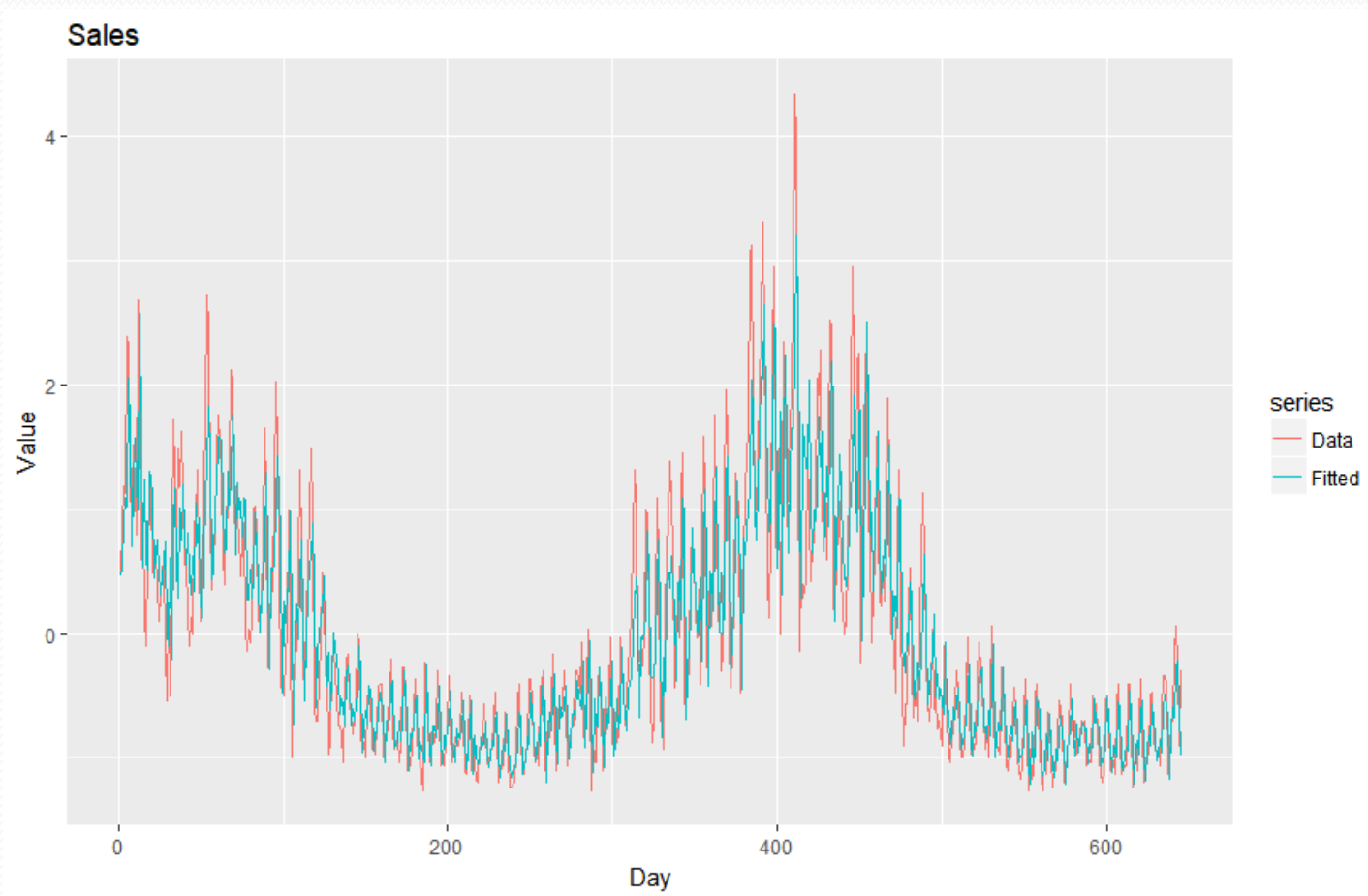
Моделювання попиту на харчову продукцію : порівняння реальних і прогнозованих продаж

Результати роботи розробленої системи (NeuralNet)



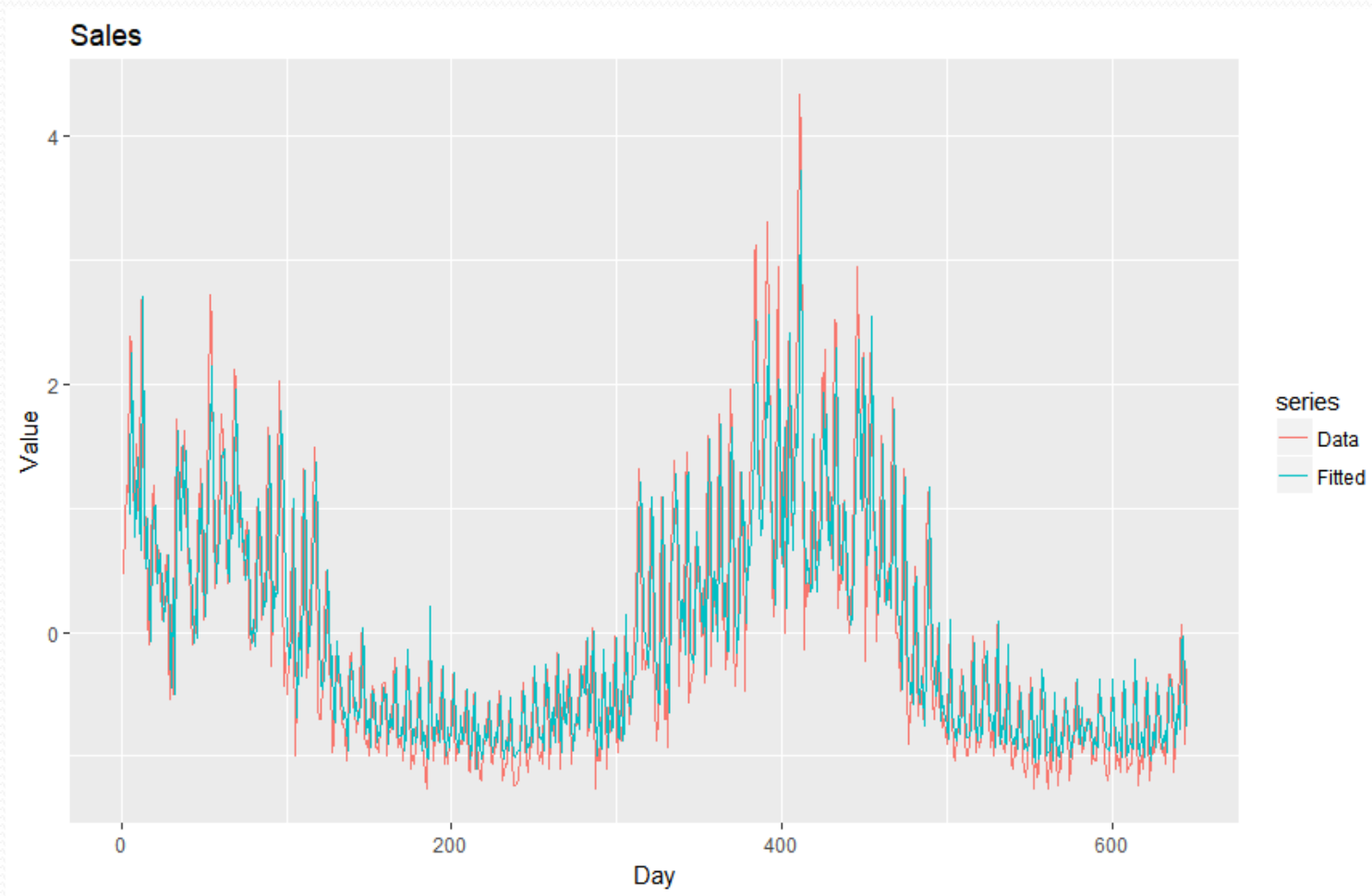
Моделювання попиту на харчову продукцію: Glmnet

Результати роботи розробленої системи (Glmnet)



Моделювання попиту на харчову продукцію : XGBoost

Результати роботи розробленої системи (XGBoost)



Моделювання попиту на харчову продукцію: порівняння моделей

Тип моделі	Характеристики моделі		Характеристики прогнозу		
	R ²	AIC	RMSE	MAE	MAPE
glmnet	0.7955346	900.03	0.5392346	0.404728	9.289125
XGBoost	0.8675924	822.15	0.3001305	0.230912	2.412832
neuralnet	0.8481675	836.27	0.3718906	0.288897	1.988698
Тип моделі	Характеристики моделі		Характеристики прогнозу		
	R ²	AIC	RMSE	MAE	MAPE
glmnet	0.8352460	958.24	0.6349195	0.474474	7.587532
XGBoost	0.915846	917.04	0.5518013	0.437155	1.128447
neuralnet	0.9348357	916.02	0.4224225	0.315366	0.961578

Найкращі результати прогнозування показали XGBoost та neuralnet моделі

ВИСНОВКИ

- Виконано огляд існуючих методів аналізу, моделювання і прогнозування нестационарних процесів. Розглянуто основні фактори, що впливають на попит на харчову продукцію. Виявлено особливості моделювання динамічних процесів.
- Створена система підтримки прийняття рішень з використанням платформи Azure ML Studio для проведення обчислювальних експериментів.
- Досліджено теоретичні засади основних методів вирішення задач регресії.
- На основі моделювання і прогнозування визначені кращі моделі для вибраних процесів: XGBoost та neuralnet.

Перспективи щодо подальших досліджень

- Підвищення адекватності побудованих моделей та покращення якості прогнозів
- Аналіз факторів, які впливають на зміну динаміки вибраних процесів
- Застосування інших методів та моделей (нейронні мережі типу LSTM, Random Forest, LGBM, ансамблювання моделей)

Дякую за увагу!