

Індивідуальне завдання з курсу «Статистичний аналіз та прогнозування»

Мета роботи: навчитися застосовувати прикладні методи прогнозування для вирішення задач аналізу даних.

Порядок виконання роботи:

1. Постановка власної задачі моделювання.
 2. Виконати пошук даних та формування файлу навчальних даних. Дані повинні задовольняти наступним вимогам:
 - Обсяг даних – щонайменше 1000 записів;
 - Наявність причинно-наслідкового зв'язку між компонентами даних.
 - 2.1. Побудувати графіки розподілів значень для числових даних.
 - 2.2. Визначити характеристики розподілів для числових даних (мода, медіана, математичне сподівання, дисперсія, асиметрія, ексцес, тощо).
 - 2.3. Здійснити аналіз пропущених значень.
 3. Розбити навчальні дані на навчальну та перевіірочну вибірку (крос-валідація). Рекомендується використовувати такі системи ІАД для виконання інтелектуального завдання, як SAS Enterprise Miner, SPSS, R, Python).
 4. За навчальною вибіркою побудувати структури моделей різними методами інтелектуального аналізу даних (логістична регресія, лінійна регресія, нейронні мережі, мережі Байєса, дерева рішень тощо) і обрати кращу з них.
- Побудова математичної моделі у вигляді регресії здійснюється за наступним алгоритмом:
- 4.а. Перевірити наявність сезонного ефекту (обчислюються сезонні різниці ряду відповідно до періодичності сезонного ефекту з метою «згладжування» дисперсії процесу).
 - 4.б. Розрахувати АКФ та ЧАКФ для ряду.

- 4.в. Визначити можливі структури моделей-кандидатів (АР або АРКС).
- 4.г. Обчислюють оцінки параметрів моделей кандидатів.
- 4.д. Вибрати кращу модель за допомогою статистичних критеріїв.

Для всіх побудованих моделей:

- 4.1. За перевіркою вибіркою оцінити якість моделей.
- 4.2. Покращити якість побудованих моделей.
- 4.3. Обрати кращу модель.

5. Порівняти побудовані моделі та зробити висновки щодо доцільності використання прикладних методів прогнозування та типів задач та прогнозних змінних, в яких це доцільно виконувати.

6. Оформити звіт.

7. Захист роботи (при наявності кодів програми, вибірки даних та протоколу).

Зауваження

Вибір даних та постановка задачі – є частиною завдання і здійснюється студентом самостійно.

В якості прикладу студентам надаються вибірки вхідних даних, які вони можуть використати для виконання індивідуального завдання, проте постановка задачі не може повторюватись.