

ПРОЕКТУВАННЯ ТА РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ ПОГОДИ

А. О. Білич, Т. В. Нескородєва

Анотація. У даному дослідженні відображенні основні міркування у побудові систем з прогнозування погоди, показані основні методи прогнозування які застосовуються у побудові систем такого роду. Загалом у цьому дослідженні представлені основні статистичні методи, а саме, методи регресійного аналізу, які ґрунтуються на тому, щоб створювати системи прогнозування погоди. Представлені такі методи як : лінійна регресія, множино-лінійна регресія та метод часових рядів. Основна мета дослідження створення додатку з прогнозування погоди з використанням цих регресійних методів прогнозу.

Ключові слова: Регресійна модель, прогнозування, погода, аналіз даних, додаток.

Постановка проблеми. Системи з прогнозування погоди є важливим елементом для сучасного людського життя, вони є досить важливим чинником у сучасному житті і використовуються у багатьох галузях, наприклад [2]:

- Прогнозування загального користування – системи прогнозування погоди, які створені для простих користувачів, аби вони змогли переглянути прогноз погоди, наприклад, на наступні 7 діб.
- Авіаційні системи прогнозування – системи створені для чіткого прогнозу атмосферних явищ у природі таких, як детальну характеристику вітру, вірогідність шторму, опади, осадки, атмосферний тиск тощо.

Прогнозуванням погоди є обґрунтоване припущення щодо майбутнього стану погоди, ось це є основною концепцією усіх існуючих систем та додатків, які виконують функції діагностики погоди. Зазвичай будь які системи з прогнозування створюють свій прогноз на основі аналізу відомих даних таких як, швидкість вітру, атмосферний тиск, вологість, хмарність тощо. Дані є великим компонентом з прогнозування. Знання метеорології лежать в основі наукового прогнозування погоди, яка обертається навколо передбачення стану атмосфери для певного місця [1].

Аналіз останніх досліджень та публікацій. Загалом у світі є досить багато досліджень на тематику прогнозування погоди. Усі вони різні та представлені у різних ключах та досліджуються за різними методологіями. В основі цих досліджень часто використовують регресійний аналіз даних за допомогою якого можливий прогноз погоди. Саме завдяки основним регресійним моделям можливий прогноз погоди на основі аналізу даних, тобто на основі даних які ми уже отримували раніше. Зазвичай у сучасних дослідженнях прогнозу погоди використовуються такі основні регресійні моделі для аналізу даних [5]:

- Лінійна регресійна модель
- Множино-лінійна регресійна модель
- Прогнозування на основі часових рядів.

Виклад основного матеріалу дослідження. В основі дослідження пов'язаного з можливістю створення системи з прогнозування погоди ми використовували три основних моделі для порівняння та вибору моделей з найкращими можливостями для реалізації десктоп додатку з прогнозування погоди.

Основною ціллю дослідження були певні мети та цілі. Основною задачею були дослідження регресійних моделей та вибір найкращою моделі для створення додатку з прогнозування погоди [5].

Дослідження складалось з трьох основних регресійних моделей:

- Лінійна регресійна модель
- Множино-лінійна регресійна модель
- Прогнозування на основі часових рядів

Проста лінійна регресійна модель є самим простим методом прогнозування на основі часових рядів. У найпростішому випадку регресійна модель допускає лінійну залежність між змінною прогнозу y і однією змінною предиктора x . Загалом типова проста лінійна регресія записується за формулою (1) та записується у такому вигляді :

$$Y_t = \beta_0 + \beta_1 * x_t + \varepsilon_t \quad (1)$$

Коефіцієнти β_0 та β_1 позначають перехоплення та нахил лінії відповідно. Перехоплення β_0 представляє передбачуване значення y коли $x = 0$. Схил β_1 являє собою середню прогнозовану зміну y одержуючи від однієї одиниці збільшення x .

Загалом у концепції статистичного прогнозування проста лінійна регресія є емпіричним підходом, і вона може вирішувати завдання, беручи до уваги набір історичних даних про погоду чи параметри температури. Він може складатися лише з однієї залежної змінної та незалежної змінної. У простій моделі лінійної регресії можуть існувати лише дві змінні [5]. Загалом коли говорять про просту лінійну регресію, мають на увазі таку формулу (2.), і має такий вигляд:

$$Y = a + \beta_x \quad (2)$$

Де:

- Y – залежна змінна,
- X – незалежна змінна,
- a, β – коефіцієнти регресії

Загалом для вирішення проблеми методом SLR (лінійна регресійна модель) потрібно враховувати такий простий алгоритм, який рекомендують для використання [3]. Алгоритм виконання дій по відношенню до SLR зображений на рисунку 2:

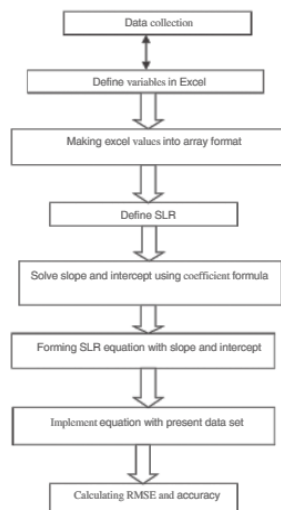


Рис. 1. Алгоритм вирішення задачі методом простої лінійної регресії

Ще однією важливою регресійною моделлю є множино-лінійна регресія. Прогнозування погоди за допомогою множинної регресійної моделі є одним із актуальних напрямків прогнозування погоди на основі статистичного аналізу даних, які передбачають складний процес у метеорологічних дослідженнях, які стають можливими на основі статистичних даних, які були отримані раніше [4].

Загалом модель прогнозування погоди, яка буде використовувати цю методологію прогнозування представляється у вигляді блок-схеми та включає у себе основні пункти [4], а саме:

- Input Data – Вхідні дані, які необхідні алгоритму для обробки
- Filter – Етап на якому вхідні дані повинні поділитись на тестові та тренувані вибірки
- Scaling Conversion – перетворення даних у зручний підбір для подальшого аналізу даних та створення прогнозованих моделей для різних явищ у природі
- Semantic of conversion errors – елемент який використовується для вирахування можливої помилки в програмі, відповідає за вирахування RSME помилки.
- Testing Data – елемент в якому зберігаються тестові вихідні дані
- Training Data – елемент в якому зберігаються тренувальні набори даних для навчання системи, щоб вона змогла у подальшому давати точніші прогнози.

Загалом блок-схема (алгоритм) виконання задачі з прогнозування показана на рисунку 3, та має такий вигляд:

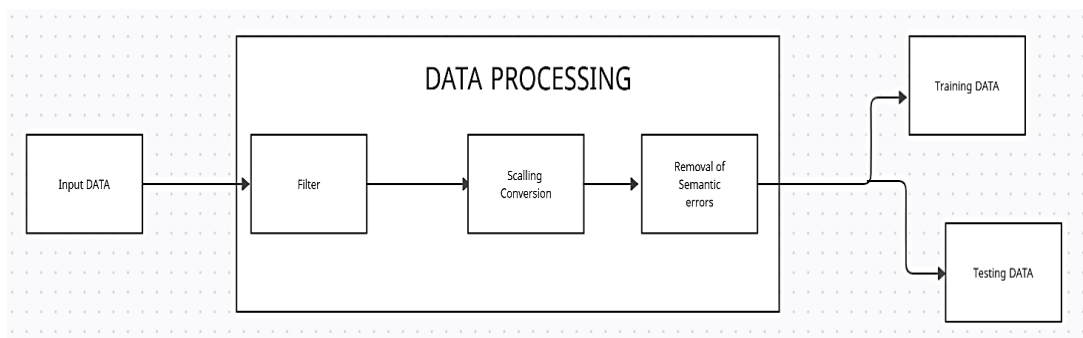


Рис. 2. Блок-схема процесу вирішення задачі з прогнозування на основі множино-лінійної регресії

Основна частина програми – написана на мові С#. Містить такі елементи як оболонку програми яка містить в собі основний інтерфейс програми та відповідає за збір даних. Також в цій частині міститься основний елемент збору даних для подальшого аналізу за допомогою мови R. Проте основна частина програми в змозі сама відтворити примітивні дані, щодо аналізу.

Допоміжна частина програми – містить елементи статистичного аналізу, створює дата-фрейми на основі зібраних раніше даних для подальшого аналізу даних. На основі методів, які є в статистичному аналізі, програма буде здатна відтворити графіки любых типів які необхідні для виявлення даних.

Загалом була вибрана модель лінійної регресії. За результатами дослідження методів була побудована блок-схема та алгоритм дій таким чином, щоб можна було спроектувати додаток з прогнозування погоди та отримати чітку модель прогнозу на основі регресійних аналізів які було проаналізовано.

Блок-схема та алгоритм представлений на рисунку 3 та має такий вигляд:

Weather Forecast report block scheme

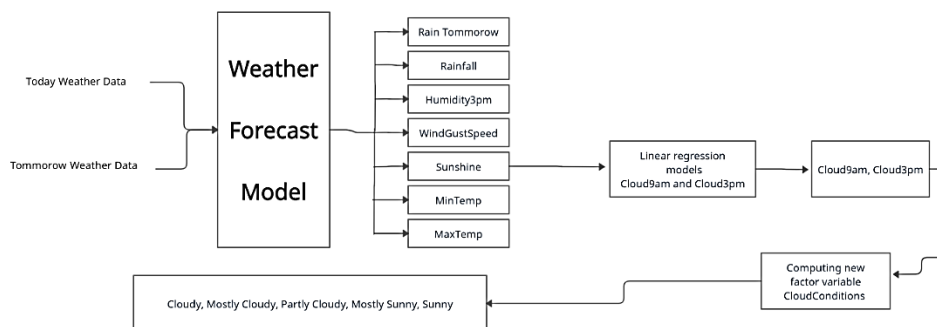


Рис. 3. Блок схема додатку з прогнозування додатку

Блок схема містить у собі проектування моделей прогнозу погоди на різні етапи доби: ранок, день, вечір. Також модель прогнозу має у собі основні елементи прогнозу погоди різних явищ, таких як: опади, температура, вологість тощо.

В основі додатку міститься зручний інтерфейс програми написаний на C#, основна ціль основної частини програми формування вхідних даних з використанням програмного API Open Weather та формування дата-сету з вхідних даних за допомогою якого створюється аналіз даних за допомогою мови R та прогнозування за допомогою регресійних моделей, які обрані для побудови додатку.

Інтерфейс програми представлений таким чином, щоб простий користувач міг би користуватись програмою. Інтерфейс програми представлений на рисунку 4:

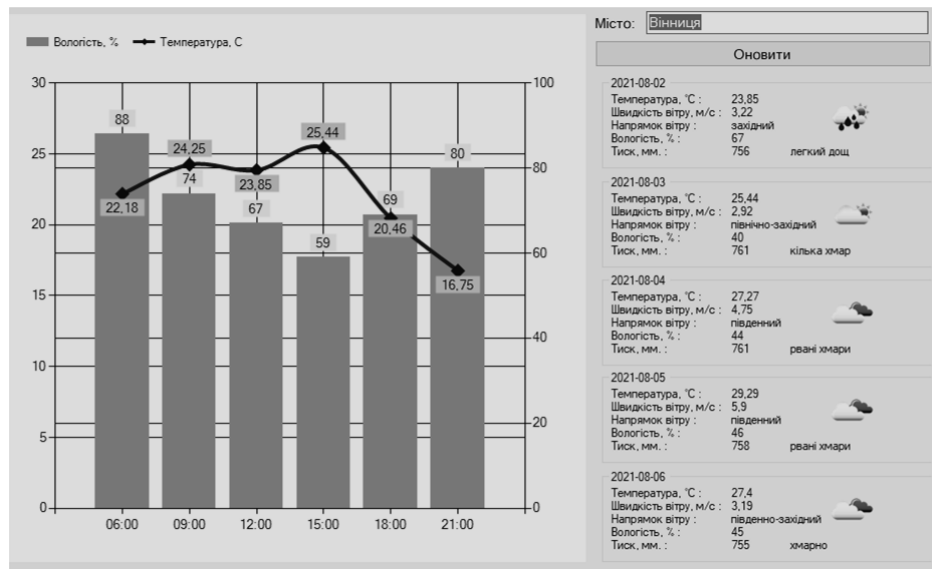


Рис. 4. Інтерфейс додатку з прогнозування погоди

Головна у додатку це формування дата-сету і отримання вхідних даних для вибраного міста і створення аналізу і прогнозування на основі уже відомих вхідних даних, після чого, отримати вихідні дані у вигляді прогнозу основних явищ, як показано на рисунку 4.

Зокрема додаток з прогнозування після отримання вхідних даних та генеруванні дата-сету має такий вигляд і представляє такий список перед тим як буде проводитись статистичний аналіз та прогнозування, Рисунок вхідних даних показаний на рисунку 5:

Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine (h)
..	Vinnitza	100
..	Vinnitza	800
..	Vinnitza	100
..	Vinnitza	100
..	Vinnitza	990
..	Vinnitza	90

Рис. 5. Генерований дата-сет, вхідні дані для додатку

Додаток проводить статистичний аналіз, в тому числі буде певні візуалізації, порівняння даних, категоріальних і множинних змін. Одним із цих моментів є побудова матриці кореляції для цього дата-сету. Матриця кореляції представлена на рисунку 6 та має такий вигляд:

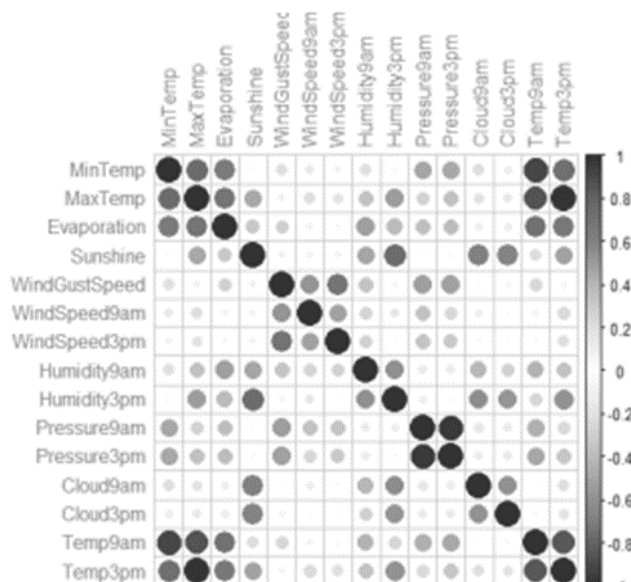


Рис. 6. Матриця кореляції вхідних даних для додатку з прогнозування погоди

Також на основі того, що ми знайшли матрицю кореляції для створеного дата-сету, ми також можемо зробити такі висновки та припущення.

Поглянувши на наведений вище графік кореляції, можна констатувати, що:

- Temp9am сильно позитивно корелює з MinTemp
- Temp9am сильно позитивно корелює з MaxTemp
- Temp9am сильно позитивно корелює з Temp3pm
- Temp3pm сильно позитивно корелює з MaxTemp
- Pressure9am позитивно корелює з Pressure3pm
- Humidity3pm сильно негативно корелює з Sunshine
- Cloud9am сильно негативно корелює з Sunshine
- Cloud3pm сильно негативно корелює з Sunshine

На основі аналізу були побудовані моделі для різних погодних чинників погодних умов. Зокрема були побудовані моделі для прогнозування опадів, температури, вологості, швидкості вітру тощо.

Приклади роботи моделей представлені на рисунку 7 і має такий вигляд:

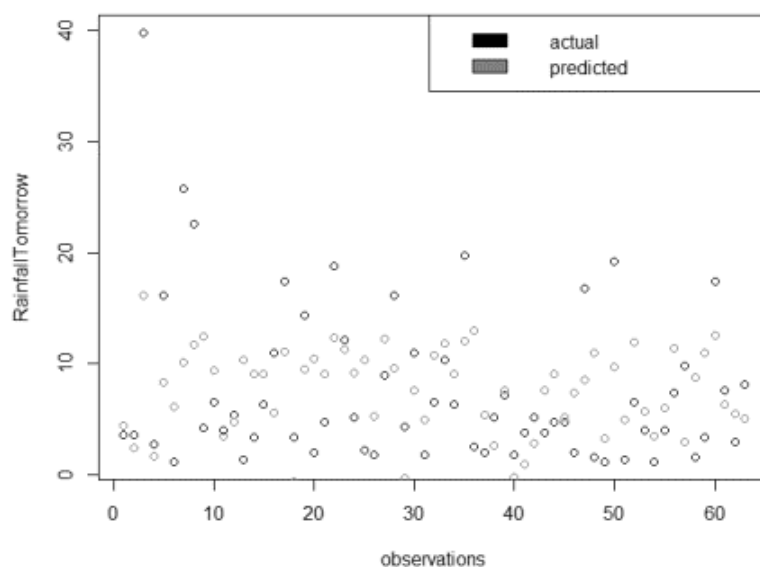


Рис. 7. Приклад роботи моделі прогнозування опадів. Візуалізація даних щодо прогнозування опадів, чорні – реальні отримані дані, сірі – прогнозовані дані

Висновок. Розроблений програмний додаток спроектований за допомогою двох мов програмування, одна з яких мова програмування високого рівня C# та мови R для статистичного аналізу та прогнозування.

Програмна реалізація представляє собою модель програми складеною з двох частин в якій присутні програмування програмного інтерфейсу на мові C# та друга частина в якій відбувається статистичний аналіз даних R. Зокрема програма використовує елементи кросс-платформного програмування та має такі бібліотеки в своїй основі, як Entity Framework, Apache Shark та ін. Зокрема програма використовує відкритий API Open Weather для отримання початкових даних та формування свого дата-сету на якому будуть створюватись прогнози, щодо наступних даних. Основою для прогнозування і аналізу є такі основні моделі, як регресійна модель, методи динамічних кіл та основи статистичного аналізу, саме вони є основою для створення прогнозу.

Основною задачею, яку я поставив при розробці додатку, це створення простого і зручного додатку, який би не був проблемою для нового користувача. Але і на далі, після захисту я планую його подальшу розробку да вдосконалення. Основними кроками поліпшення додатку є :

- Покращення інтерфейсу
- Додання більше функціональних кнопок
- Додання форми логіну та паролю
- Створення кабінету та адмін контролю програми
- Дата-сет який збирає дані перетворити на повноцінну базу даних, для більш зручного контролю даних
- Мобільна версія додатку або аналог у вигляді сайту.

Аннотация. В данном исследовании отображены основные соображения в построении систем по прогнозированию погоды, показаны основные методы прогнозирования, применяемые в построении систем такого рода. Всего в этом исследовании представлены основные статистические методы, в частности, методы регрессионного анализа, основанные на том, чтобы создавать системы прогнозирования погоды. Представлены такие методы как: линейная регрессия, множественная линейная регрессия и метод временных рядов. Основная цель исследования создания приложения по прогнозированию погоды с использованием этих регрессионных методов прогноза.

Ключевые слова: регрессионные модели, прогнозирование, погода, анализ данных, приложение.

Abstract. This study displays the main considerations in building weather forecasting systems, showing the main forecasting methods used in building systems of this kind. In total, this study presents the main statistical methods, in particular regression analysis methods, based on which weather forecasting systems are built. The methods presented are linear regression, multiple linear regression, and time series method. The main purpose of the study is to create a weather forecasting application using these regression-forecasting methods.

Key words: regression models, forecasting, weather, data analysis, application.

СПИСОК ЛІТЕРАТУРИ

1. Хамід Ісмаїл., Статистичне моделювання, лінійна регресія та ANOVA, практична обчислювальна перспектива в м'якій обкладинці. - lulu.com; First Edition, 2018. 466 с.
2. Саймон Н. Вуд., Узагальнені адаптивні моделі: Вступ до R, друге видання. - Chapman and Hall/CRC; 2nd edition, 2017. 496 с.
3. Джейсон У. Осборн. Найкращі практики з логістичної регресії. SAGE Publications, Inc, 2014. 500 с.
4. Пірбі Р., Хелер Г., Розподіли для моделювання розташування, масштабу та форми: використання GAMLSS у R (Chapman & Hall/CRC The R Series) 1-е видання. Chapman and Hall/CRC; 1st edition, 2019. 588 с.
5. Ануша Люккумбура., Вступ до регресійного аналізу. - Independently published, 2020 – 121 с.
6. Алан Агрешті., Основи лінійної та узагальненої лінійних моделей. Wiley; 1st edition, 2015. 469 с.
7. Джеймс Г., Вітнер Д., Хасті Т. Вступ до статистичного навчання. Springer; 1st ed. 2013, Corr. 7th printing 2017 edition, 2017. 440 с.