

ПРОГНОЗУВАННЯ ОБСЯГІВ СТОВРЕНИХ, ЗІБРАНИХ, СКОПІЙОВАНИХ І СПОЖИТИХ ДАНИХ У СВІТІ

UDC 004.62

S. R. Myskiv

FORECASTING THE AMOUNT OF DATA CREATED, COLLECTED, COPIED AND CONSUMED IN THE WORLD

Прогнозування обсягів даних – це критично важливе завдання в епоху цифрової трансформації. Воно дозволяє підготуватися до викликів, які супроводжують швидке зростання інформаційних потоків, і максимально ефективно використовувати потенціал великих даних для розвитку суспільства та економіки [1].

В середньому, у світі, за секунду створюється 29 терабайтів даних, за хвилину – 1736 терабайтів, за годину – 104000 терабайтів, за день – 2,5 мільйонів терабайтів [2].

Вхідні дані для прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно наведено у таблиці 1 [3].

Таблиця 1. Вхідні дані для прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно

Рік	Обсяг даних	Зміна у % порівняно з попереднім роком
2024	147 zettabytes	22.5%
2023	120 zettabytes	23.71%
2022	97 zettabytes	22.78%
2021	79 zettabytes	23.05%
2020	64.2 zettabytes	56.59%
2019	41 zettabytes	24.24%
2018	33 zettabytes	26.92%
2017	26 zettabytes	44.44%
2016	18 zettabytes	16.13%
2015	15.5 zettabytes	24%

Аналіз вибірки: найбільші зміни у відсотках припали на роки COVID-19 – 2020 рік.

Для задачі прогнозування було використано метод поліноміальну регресії (2-го ступеня). Аргументація вибору даного методу наступна: дані про обсяги даних у світі показують нелінійний, прискорений ріст.

Python-скрипт реалізує задачу прогнозування. Основні етапи реалізації:

Імпорт бібліотек – бібліотеки pandas, numpy, matplotlib, sklearn для аналізу та візуалізації даних.

Дані – обсяги даних (у зетабайтах) за 2015–2024 роки.

Модель – застосовано поліноміальну регресію 2-го степеня..

Прогноз – обчислюється прогноз на 2025–2030 роки.

Візуалізація – будується графік з фактичними даними, трендом і прогнозами.

Результати – виводиться графік та таблиця з прогнозованими значеннями.

Результат прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно графічно наведено на рисунку 1.

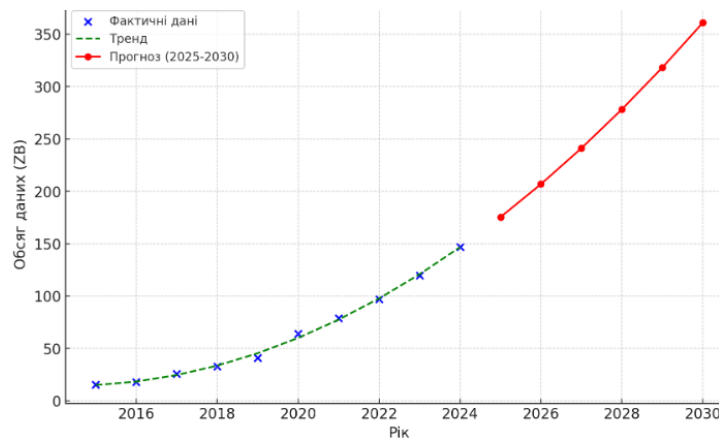


Рисунок 1. Результат прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно (графічне представлення)

Результат прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно представлених в табличному форматі наведено у таблиці 2.

Таблиця 2. Результат прогнозування обсягів даних, створених, зібраних, скопійованих і спожитих в світі сумарно представлених в табличному форматі

Рік	Прогнозований обсяг даних (ЗВ)
2025	175.46
2026	206.96
2027	241.27
2028	278.40
2029	318.34
2030	361.09

Висновки: аналіз показує, що обсяги інформації зростають прискореними темпами, особливо помітно під час пандемії COVID-19 у 2020 році. За прогнозом, дані створені, зібрані, скопійовані і спожиті в світі сумарно у 2030 році буде у 23,296 раз більше порівняно з 2015.

Література

1. Kumar N. Big Data Statistics 2024: Growth and Market Data. Demandsage. URL: <https://www.demandsage.com/big-data-statistics/> (date of access: 19.01.2024)
2. Marr B. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Forbs. 2021. URL: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/> (date of access: 30.01.2024).
3. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024. Statista. URL: <https://www.statista.com/statistics/871513/worldwide-data-created/> (date of access: 05.02.2024).