



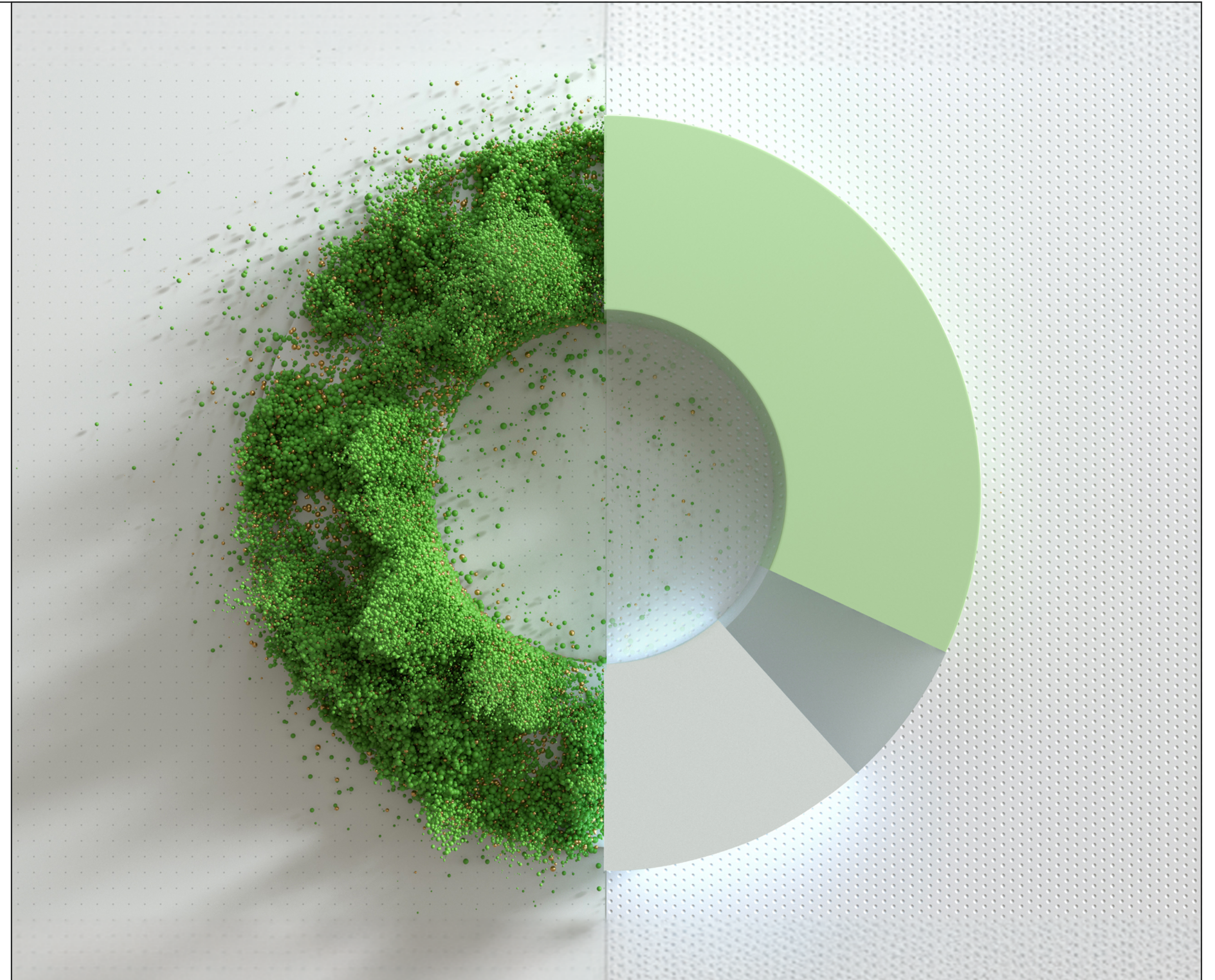
# Accurate Data Matching Solution to Accelerate Sales Analytics for Darnytsia\_

Data Matching Solution for Darnytsia Aligns Data from Different  
Sources with Above 96% Accuracy

**Industry: Pharmaceuticals**

**Location: Ukraine**

**Employees: 1,000+**



## Client Background

**Website:**

<https://darnytsia.ua/en>

Pharmaceutical Company “Darnytsia” is Ukraine’s largest producer of medicinal products, having long-standing pharmaceutical traditions. With 90+ years of history, Darnytsia has had the leading position in the domestic pharmaceutical market by volume for more than 20 years, which proves to be true by year-to-year indicators of the industrial and market activity of the enterprise, as well as authoritative branch ratings. Darnytsia shows significant progress in the process of digital transformation, succeeding in the implementation of market-leading technologies.

## Executive Summary



### Goals

Design a solution to match data tables from different sources and adjust data to a single format for further sales and market analytics, accelerating data processing and addressing human error risks.



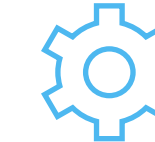
### Solution

Infopulse data scientists implemented a hybrid data matching algorithm with above 96% matching accuracy for drug names and above 85% accuracy for pharmacy names and addresses.



### Benefits

Automated data matching with zero human error risk, faster data analytics and decision-making, streamlined data processing with reduced labor and cost.



### Services delivered

Innovation Services, Intelligent Business, Intelligent Automation, Smart Insights

# Business Challenge

Darnytsia and Infopulse have a long history of collaboration, having implemented numerous digital transformation projects together. After successfully implementing [a GenAI data analytics bot](#) and a sales prediction solution in cooperation with Infopulse, the company requested our assistance in addressing another challenge related to data used for sales and market analytics:

- Matching the drug information between the company's internal database and external tables received from pharmacies collaborating with Darnytsia
- Matching the pharmacies' names and addresses between a table from the internal database and external sources

Initially, data matching was conducted manually; yet such an approach is time-consuming and is associated with risks of human errors that may negatively affect the results. The pharma giant was not satisfied with a ready-made solution they had previously tested and requested Infopulse to develop a custom matching algorithm.

# Solution & Business Value

Together with Darnytsia, Infopulse developed a **hybrid matching algorithm** that matches records between data tables (internal and external) with a **total error of less than 5%**.

The business value from such an output for Darnytsia includes:

- **Complete automation** of the data matching process with 20x less employees' time required
- **Reduced time and cost** for the related operations: matching tens of thousands of records takes less than 10 minutes

- **85-96% accuracy of matching**
- **Zero human error risk** due to automation and no manual job involved
- **Faster data analytics**, decision-making, and insights generation

**The project implementation took two weeks.**

## Technical Details

The client provided us with two Excel data tables containing drug information (300-400 records each), and three tables containing pharmacy data (up to 10,000 records each). Data was inconsistent and set in different formats, making the initial task more complicated. The goal was to design a solution that would find matches between records, taking into account disparate formats and typos.

Using these initial datasets, we searched for suitable matching metrics and tested various approaches and available algorithms. As a result, Infopulse created a **hybrid matching algorithm** that looks for similarities between two data strings and calculates the total similarity score.

### The following metrics were used:

- Ratcliff similarity
- Levenshtein distance

- TF-IDF (term frequency-inverse document frequency)

As initially agreed with Darnytsia, the output is presented in the form of Python code that our client can run when needed.

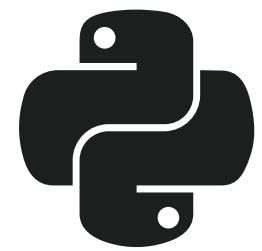
### Matching algorithm tests showed the following results:

- **96% accuracy** for matching two drug tables of 300-400 records each
- **Above 85% accuracy** for matching three pharmacy tables (names and addresses matched) of **10,000 records** each

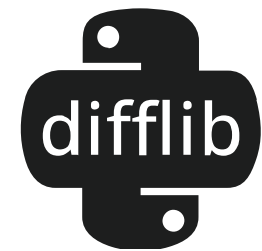
The time required for running the matching algorithm is **under one minute** for drug tables, and **under 10 minutes** for pharmacy tables. The solution is not limited by a specific data scale and can work for any data volume.



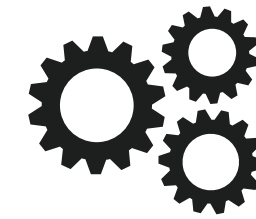
# Technologies & Tools



Python



DiffLib



SequenceMatcher



RapidFuzz



Sklearn



## About Infopulse

Infopulse, part of the leading Nordic digital services company Tietoevry OneCreate, is an international vendor of services in the areas of Intelligent Business, Smart Insights, Intelligent Automation, and Advanced Data Analytics. Our services cover digital transformation and data-driven solutions for business management to reach operational excellence and drive growth with technology.

Infopulse is trusted by many established brands, such as Darnytsia, Delta Medical, Allianz Bank, Credit Agricole, ING Bank, OTP Bank, Santander, BICS, Bosch, LMT, Microsoft, Metinvest, Offshore Norge, Delta Wilmar, OLX, SAP, UkrSibbank BNP Paribas Group, VEON, Vodafone, Zeppelin, and others.

For more information, please visit [www.infopulse.com](http://www.infopulse.com)

## Contact us

**PL** +48 (221) 032-442

**DE** +49 (69) 505-060-4719

**US** +1 (888) 339-75-56

**UA** +38 (044) 585-25-00

**BG** +359 (876) 92-30-90

**NL** +31 (70) 89-10-8534

**BR** +55 (21) 99298-3389

 [info@infopulse.com](mailto:info@infopulse.com)

