Database-Agnostic Machine Learning

Vinay Valson

2026-01-20

Table of contents

Do you have a database? Do you want to make predictions without the hassles of having data scientists? I got you covered!

The Challenge with Relational Data

Most companies store their data in relational databases. These are systems that organize information into tables, like spreadsheets, which can be linked together using keys - like connecting a "Customers" table with an "Orders" table. While this structure is great for storing large amounts of data efficiently, it can be challenging when you want to use machine learning to make predictions or gain insights.

Traditionally, machine learning models expect all the data in a single, flat table. This means analysts have to manually combine tables, calculate summary statistics, and create new "features" from the raw data. This is a process that is very time consuming, error-prone, and specific to each database. If a database changes even slightly, the model may break, and all the work has to be redone - fyi job of the data scientist.

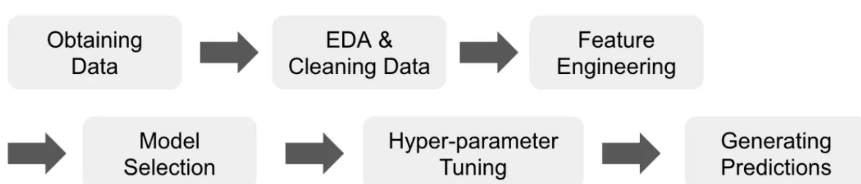

Figure 1: Data Science Pipeline

The Solution: Database-Agnostic Machine Learning

The solution is database-agnostic machine learning, which allows models to work directly with relational data without manual preparation. "Agnostic" here means flexible and the model can work across different database structures without needing to be customized for each one.

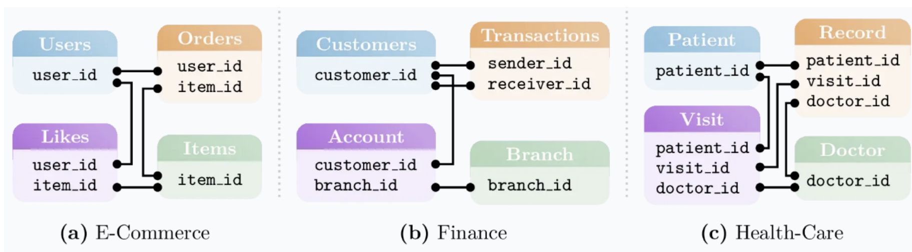

Figure 2: Relational Database Schema

Graph Transformers: Beyond Databases

Graph transformers are a powerful architecture that extends the transformer model (famous from GPT and BERT) to work with graph-structured data. While traditional transformers work on sequences (like sentences), graph transformers can process nodes and edges in a graph.

Where else are graph transformers used?

- **Social Networks:** Predicting user behavior and connections on platforms like Facebook or LinkedIn, where users are nodes and friendships are edges
- **Drug Discovery:** Modeling molecular structures where atoms are nodes and chemical bonds are edges, helping researchers predict drug properties
- **Traffic Prediction:** Understanding road networks where intersections are nodes and roads are edges, enabling better route planning
- **Recommendation Systems:** Connecting users, products, and interactions in e-commerce platforms to suggest relevant items
- **Knowledge Graphs:** Organizing and reasoning over facts and relationships, like Google's Knowledge Graph

Figure 3: Graph Transformer Visulization

KumoRFM: A Real-World Example

A real-world example of this approach is **KumoRFM**, a tool that works like a "foundation model" for relational data. Think of it as a large language model but for structured business data instead of text. KumoRFM represents tables and their relationships as a graph, where each row is a point (node) and connections between tables are lines (edges). Using a graph transformer, the model can automatically learn patterns and relationships across tables, so it understands the structure and meaning of the data.

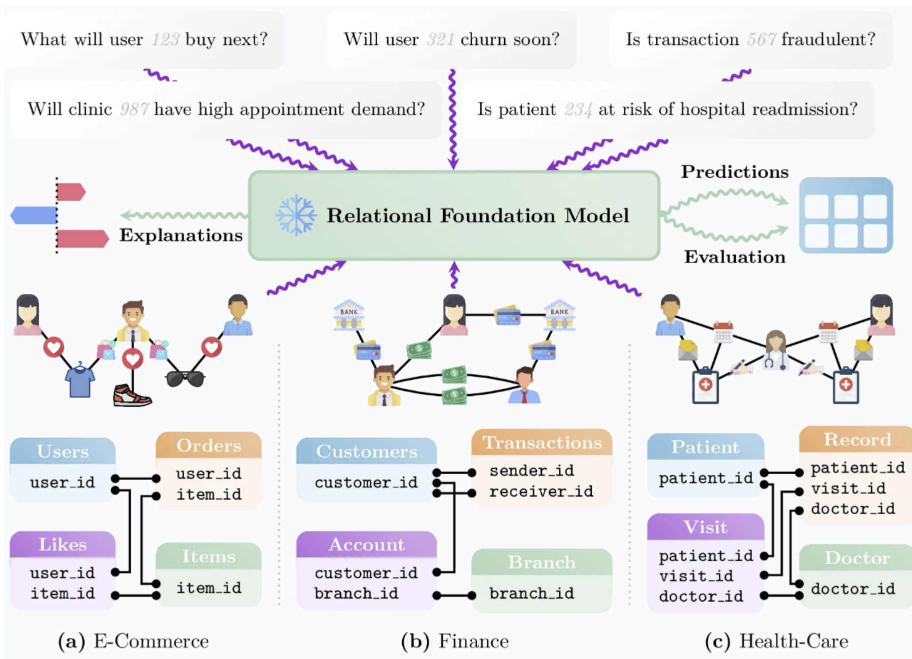

Figure 4: Key capabilities of RFM

Key Advantages

This approach has several advantages:

- **No Complex Queries:** Analysts no longer need to write complex SQL queries or manually engineer features
- **Instant Predictions:** Models can generate predictions instantly for new questions
- **Universal Application:** The same approach can be applied across different databases and organizations
- **Reduced Errors:** Eliminates manual data preparation mistakes
- **Time Savings:** Allows analysts to focus on interpreting results rather than preparing data

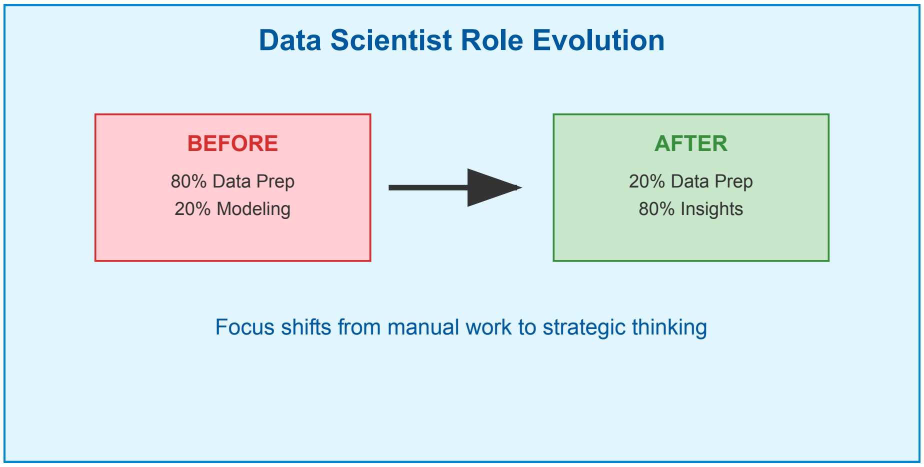

Figure 5: Role Transformation

The Impact

In short, database-agnostic machine learning allows businesses to unlock the full value of their relational data in a way that is fast, reliable, and scalable. Tools like KumoRFM demonstrate that it's possible to combine advanced machine learning with the complex structure of relational databases, making insights more accessible and actionable for decision makers.

As a result, this shift transforms the role of data scientists from manual feature engineering to higher-level problem formulation and decision-making.

References

- KumoRFM Video Introduction
- KumoRFM Getting Started
- Case Study: Slashing Utility Costs with KumoRFM