

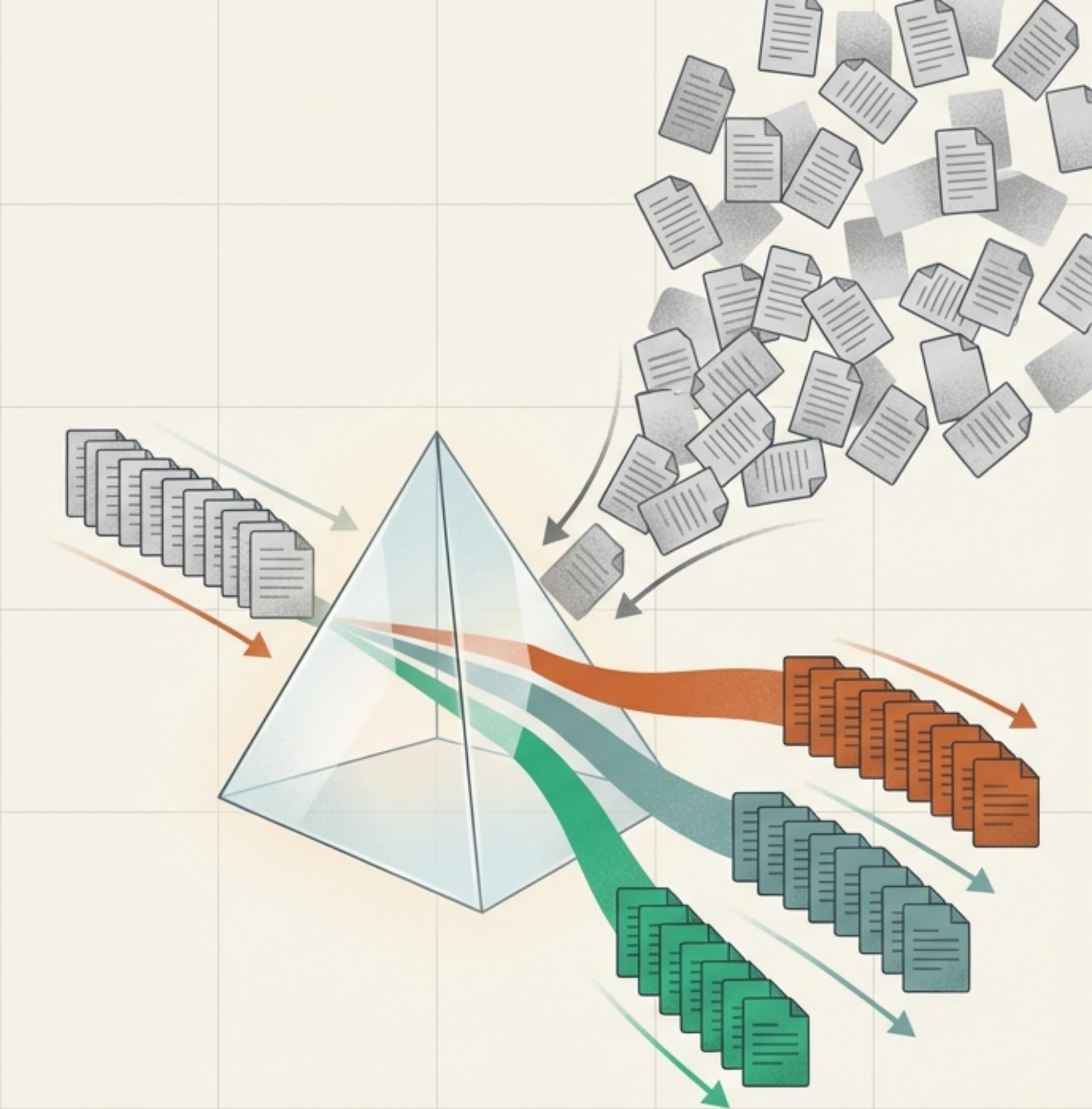
Unlocking Hidden Patterns

A Simple Guide to Topic Modeling & LDA

Explained in simple English with real examples.

What's Inside

1. What is Topic Modeling? (The intuition)
2. Real-world examples (Biryani vs. Delivery)
3. Understanding LDA (The engine)
4. The Mathematics (How it calculates)
5. Why it matters for business.



The Problem: Too Many Reviews, Too Little Time



Imagine you own a popular restaurant. You have received 10,000 customer reviews.

You cannot possibly read them all yourself.

How do you know what your customers are saying? Are they happy with the food? Are they angry about the service?

If you try to read them one by one, it will take forever.

You need a way to automate this.

The Solution: Topic Modeling

In simple words:
Topic Modeling is a technique in NLP (Natural Language Processing) that helps a computer automatically find “hidden topics” in a large collection of text.



Instead of you reading every single review, the model scans the text and tells you:

“Most of these documents are talking about Food, Delivery, and Price.”

The key is that the computer does this *without* you telling it what the topics are beforehand.

Real Life Example: Grouping Reviews

The Raw Reviews

The **biryani** was **spicy** and **tasty**

Chicken kebab was juicy and delicious

Delivery was **late** and **cold** food arrived

Packaging was bad and **delivery** was **slow**

The Hidden Topics

Topic 1:
Food Quality

biryani, spicy,
chicken, kebab,
tasty

Topic 2:
Service Issues

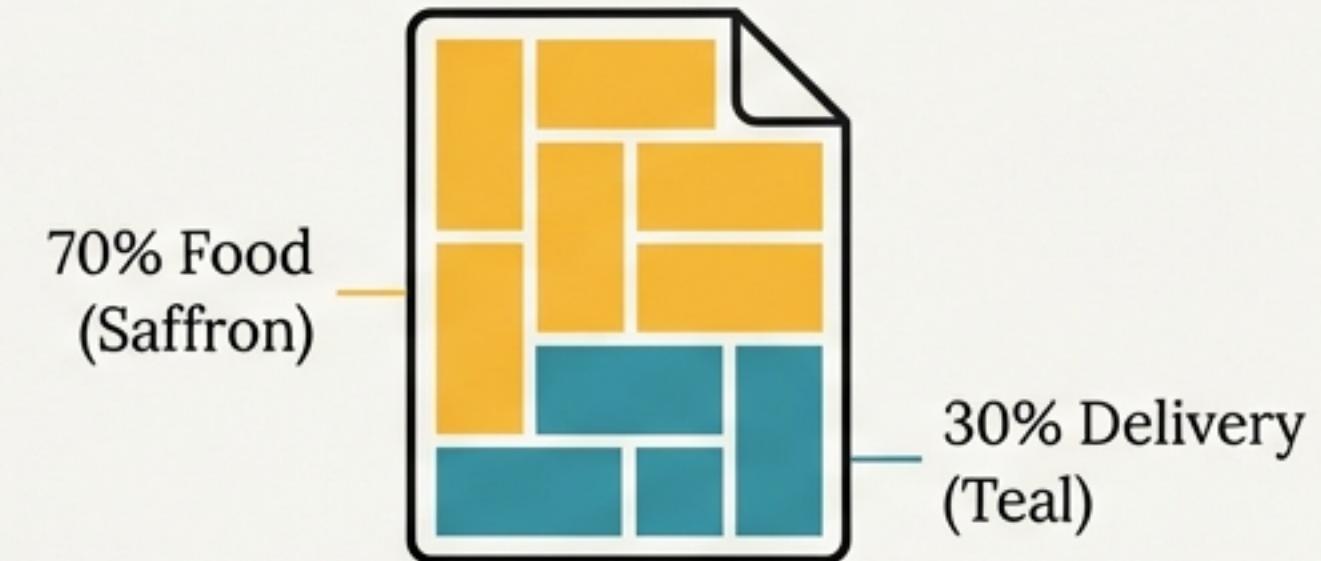
delivery, late, slow,
packaging, cold

Note: The model identifies these groups without needing labels.

Introducing LDA (Latent Dirichlet Allocation)

LDA is the most popular algorithm used to perform Topic Modeling.

1. Every document is a mix of topics.



How does it think?

LDA makes two simple assumptions about every piece of text:

2. Every topic is a mix of words.



The Intuition: Think of a Restaurant Menu

The Menu (Topics)



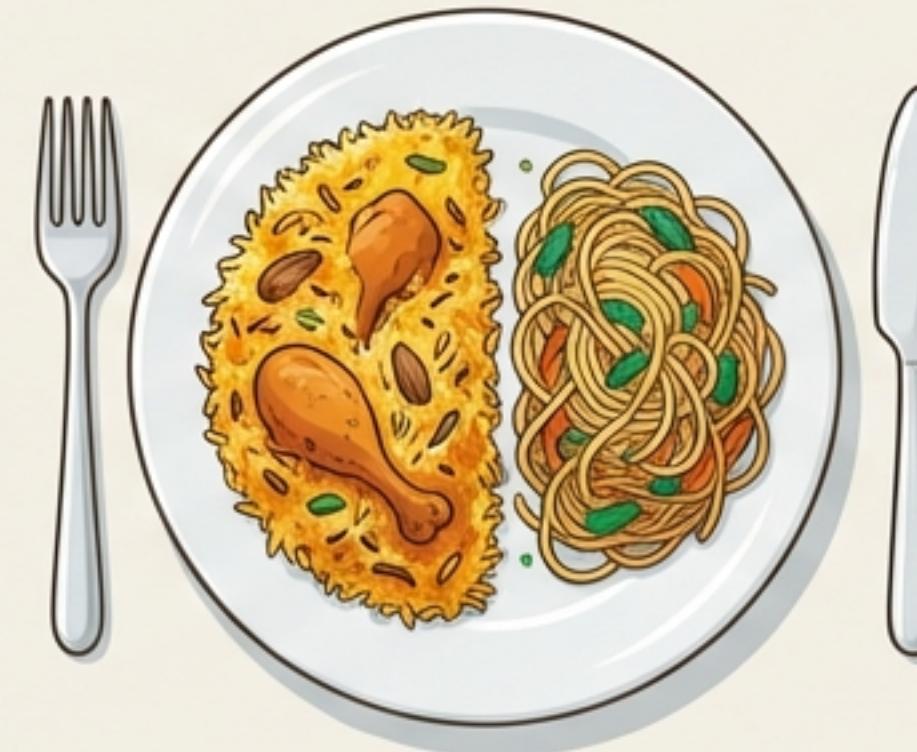
Just as a **Dish** is made of ingredients:

- **Biryani** = rice + chicken + spices
- **Noodles** = noodles + soy sauce + veggies

A **Topic** is made of words:

- **Food Topic** = biryani, chicken, spicy (words in saffron color)
- **Delivery Topic** = late, slow, cold (words in teal color)

The Customer's Plate (Document)



A customer's plate is a mix of dishes. They might take some Biryani and some Noodles.

Similarly, a **Document** is a mix of different topics.

How LDA Works: The Step-by-Step Logic

This is the generative process the algorithm assumes.



The algorithm repeats this loop for every single word in every document to learn the patterns.

The Math: Setting Up the Matrices

Let's look at the math with a very small example.

Table 1: Topic-Word Probabilities (Beta)

Probability of a word appearing in a specific topic.

Word	Topic 1 (Food)	Topic 2 (Delivery)
biryani	0.30	0.01
chicken	0.25	0.01
spicy	0.20	0.01
delivery	0.01	0.30
late	0.01	0.25
slow	0.01	0.20

Table 2: Document-Topic Probabilities (Theta)

For a specific review: "biryani was spicy but delivery was late"

Topic	Probability
Food	0.6
Delivery	0.4

The Math: Calculating the Probability

Let's calculate the probability of seeing the word "biryani" in this document.

$$P(\text{biryani}) = [P(\text{Food}) \times P(\text{biryani} \mid \text{Food})] + [P(\text{Delivery}) \times P(\text{biryani} \mid \text{Delivery})]$$

$$P(\text{biryani}) = [0.6 \times 0.30] + [0.4 \times 0.01]$$

$$P(\text{biryani}) = [0.18] + [0.004]$$

Total Probability = 0.184

This calculation determines the likelihood of a word appearing based on the mixture of topics.

Why is LDA Useful in the Real World?



Customer Reviews

Find the main complaints without reading thousands of comments.

Example: ‘The main issue this month is cold food.’



News Articles

Automatically group articles by hidden themes to recommend content.

Example: Grouping stories into Sports, Politics, or Finance.



Support Tickets

Auto-tag incoming IT issues so they are routed to the right team immediately.

Example: Tagging tickets as ‘Login Failure’ or ‘Hardware Crash’.

Summary: What to Remember



- ✓ **The Goal:** Topic Modeling finds hidden themes in large collections of text without human help.
- ✓ **The Logic:** LDA is the main algorithm. It assumes every Document is a mixture of topics, and every Topic is a mixture of words.
- ✓ **The Math:** It uses probability (matrices) to calculate which words belong to which topics.

The Power of Patterns



The beauty of Topic Modeling is that it doesn't need to be taught what 'Food' or 'Service' means. It simply looks at the math, finds the patterns, and reveals the story hidden inside your data.