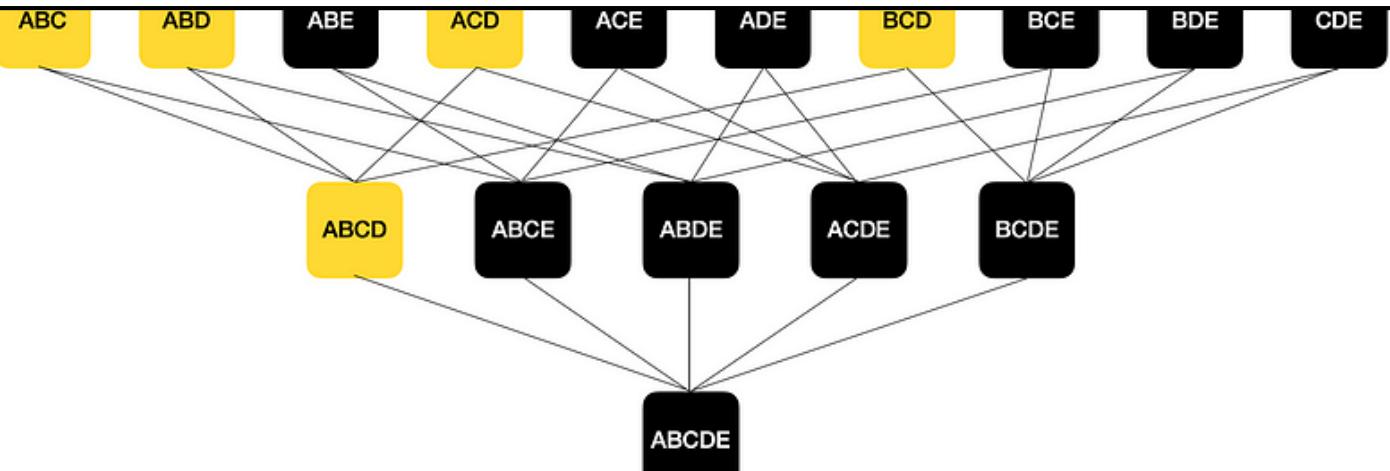


Gyakori elemhalmazok előállítása // FP-Growth algoritmus



BEVEZETÉS

Mi az FP-Growth algoritmus?

- Egy adatbányászati algoritmus **asszociációs szabályok keresésére**.
- Azonosítja a **gyakran együtt előforduló elemeket** egy nagy adathalmazban.

Hol használják?

- Leggyakrabban: bevásárlókosár-elemzés (szupermarketek, e-kereskedelem).
- Csalásfelismerés (banki tranzakciók).
- Orvosi diagnosztika (tünetek és betegségek kapcsolatai).
- Weboldalak elemzése (milyen tartalmak érdeklik a felhasználókat).

ELŐNYÖK ÉS HÁTRÁNYOK



Nincs szükség kandidátushalmaz előállítására, ezért skálázható nagy adatokra.



Hatékonyabb futási idő.



Ha az elemhalmazok ritkák vagy az adathalmaz nagyon változatos, akkor az FP-Tree nagy és komplex lesz, ami növeli a tárhelyigényt és a feldolgozási időt.

ALAPFOGALMAK

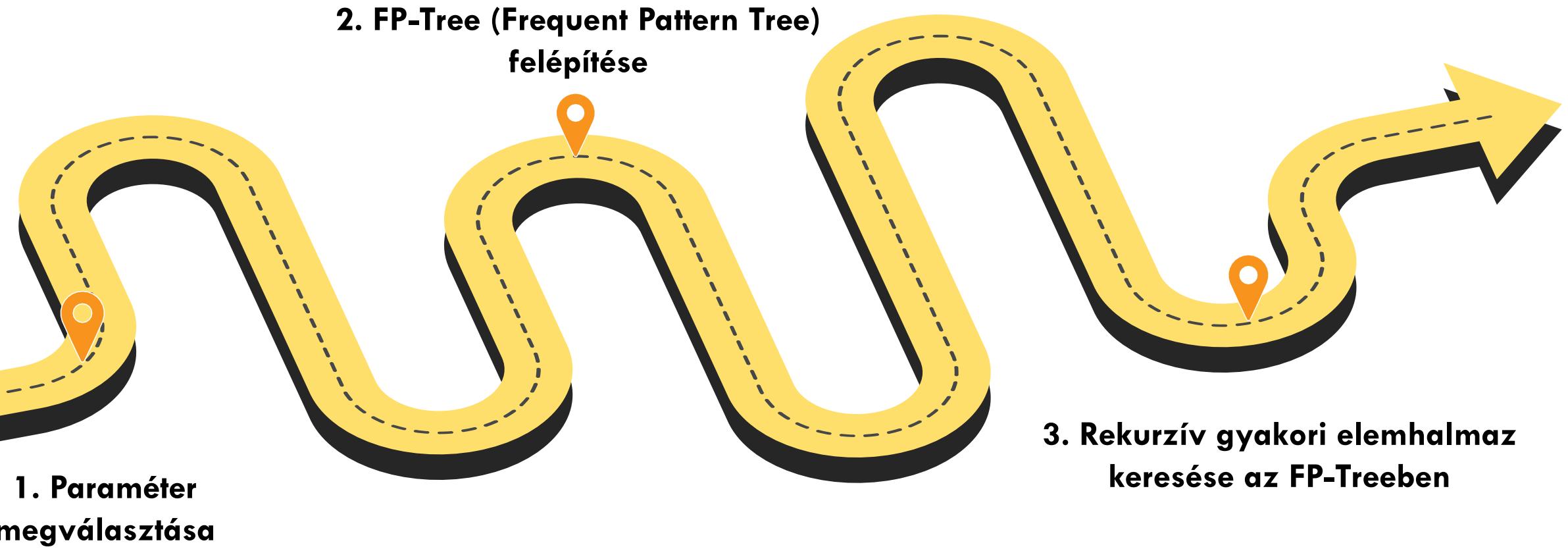
Gyakori elemhalmazok: Azok az elemek, amelyek gyakran szerepelnek együtt a tranzakciókban.

Support: Egy elemhalmaz előfordulásának aránya az összes tranzakcióhoz képest.

$$support(A) = \frac{A - t \text{ tartalmazó tranzakciók száma}}{\text{összes tranzakció száma}}$$

FP-Tree: Kompakt, fa alapú adatszerkezet, amely az input tranzakciós adatokat hatékonyan tárolja (prefixek alapján épül fel).

ALGORITMUS LÉPÉSEI



FP-TREE ADATSZERKEZET

FP-Tree létrehozása

Az elemek gyakoriságának meghatározása

- Az adathalmazból először minden elem gyakoriságát kiszámítjuk.
- Azokat az elemeket, amelyek nem érik el a minimum *support* értéket, elhagyjuk.

Az elemek sorrendbe állítása

- A tranzakciókat úgy rendezzük, hogy a leggyakoribb elemek kerüljenek előre.

FP-Tree létrehozása

- A tranzakciókat egy fa adatszerkezetbe illesztjük, ahol az azonos prefixeket megosztjuk.

PÉLDA

| Tranzakció sorszáma | Vásárolt termékek |
|---------------------|-------------------|
| T1 | A, B, C |
| T2 | A, B |
| T3 | A, B, C |
| T4 | C |
| T5 | A, C |

Adottak a fenti tranzakciók. Keressük ki a ritka elemhalmazokat. Hozzuk létre az asszociációs szabályokat. Értékeljük a szabályokat *confidence* és *lift* alapján.

PARAMÉTEREK MEGVÁLASZTÁSA

| Tranzakció sorszáma | Vásárolt termékek |
|---------------------|-------------------|
| T1 | A, B, C |
| T2 | A, B |
| T3 | A, B, C |
| T4 | C |
| T5 | A, C |

Választott paraméterek:

- **Minimum support érték:** 50%, azaz a terméknek a tranzakciók legalább 3/5-ben szerepelnie kell.
- **Minimum confidence érték:** 70%.

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

| Tranzakció sorszáma | Vásárolt termékek |
|---------------------|-------------------|
| T1 | A, B, C |
| T2 | A, B |
| T3 | A, B, C |
| T4 | C |
| T5 | A, C |

Számoljunk support értéket. Ami 0.5 alatt lenne, azt eldobnánk.

| Termék | Support (%) |
|--------|----------------------------|
| A | $4/5 = 0.8 \text{ (80\%)}$ |
| B | $3/5 = 0.6 \text{ (60\%)}$ |
| C | $4/5 = 0.8 \text{ (80\%)}$ |

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján. (gyökér)

A leggyakoribb elemek: **A (4), C (4), B (3)**

Újrarendezett tranzakciós lista:

| | | |
|---------|---|---------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |

Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

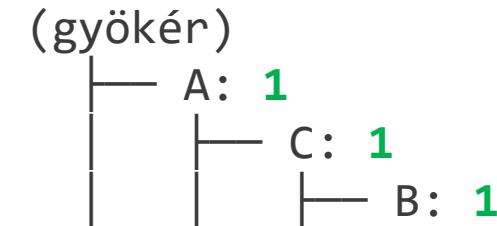
FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján.

A leggyakoribb elemek: **A (4), C (4), B (3)**

Újrarendezett tranzakciós lista:

| | | |
|---------|---|---------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

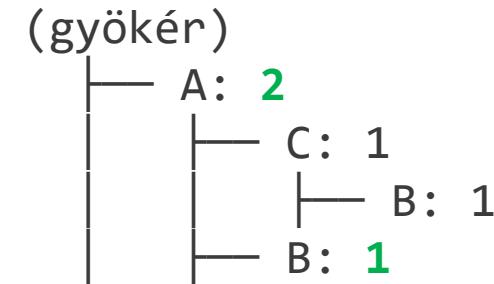
FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján.

A leggyakoribb elemek: **A (4)**, **C (4)**, **B (3)**

Újrarendezett tranzakciós lista:

| | | |
|---------|---|-------------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

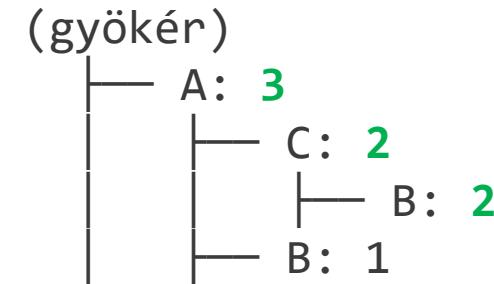
FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján.

A leggyakoribb elemek: **A (4), C (4), B (3)**

Újrarendezett tranzakciós lista:

| | | |
|---------|---|----------------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

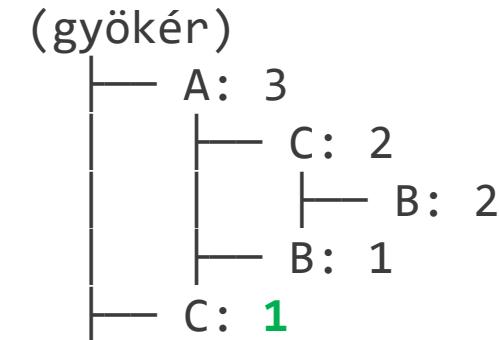
GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján.

A leggyakoribb elemek: **A (4)**, **C (4)**, **B (3)**
Újrarendezett tranzakciós lista:

| | | |
|---------|---|----------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

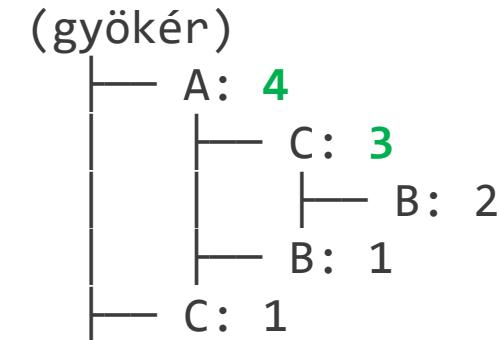
GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

FP-Tree létrehozása:

Először **csökkenő sorrendbe rendezzük** az elemeket minden tranzakcióban az előfordulási gyakoriságuk alapján.

A leggyakoribb elemek: **A (4)**, **C (4)**, **B (3)**
Újrarendezett tranzakciós lista:

| | | |
|---------|---|-------------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



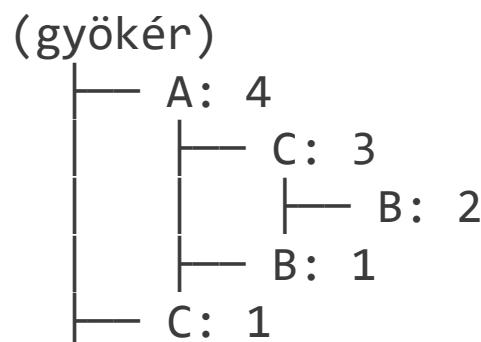
Fa építés látványosan:

<https://planktonfun.github.io/FPTreeSimulator/>

GYAKORI ELEMHALMAZOK ELŐÁLLÍTÁSA

FP-Tree létrehozása:

| | | |
|---------|---|---------|
| A, B, C | → | A, C, B |
| A, B | → | A, B |
| A, B, C | → | A, C, B |
| C | → | C |
| A, C | → | A, C |



Gyakori mintázatok kinyerése:

| Gyakori elemhalmaz | Support érték |
|--------------------|-------------------------------------|
| {A} | $4/5 = 80\% \checkmark$ |
| {A, B} | $3/5 = 60\% \checkmark$ |
| {A, B, C} | $2/5 = 40\% \times \text{(kiesik)}$ |
| {A, C} | $3/5 = 60\% \checkmark$ |
| {C} | $4/5 = 80\% \checkmark$ |
| {B, C} | $2/5 = 40\% \times \text{(kiesik)}$ |
| {B} | $3/5 = 60\% \checkmark$ |

MIT KAPTUNK A MÚLTKOR?

Apriori

| Gyakori elemhalmaz | Support (%) |
|--------------------|----------------------------|
| A | $4/5 = 0.8 \text{ (80\%)}$ |
| B | $3/5 = 0.6 \text{ (60\%)}$ |
| C | $4/5 = 0.8 \text{ (80\%)}$ |

| Gyakori elemhalmaz | Support (%) |
|--------------------|----------------------------|
| A, B | $3/5 = 0.6 \text{ (60\%)}$ |
| A, C | $3/5 = 0.6 \text{ (60\%)}$ |
| B, C | $2/5 = 0.4 \text{ (40\%)}$ |

| Gyakori elemhalmaz | Support (%) |
|--------------------|----------------------------|
| A, B, C | $2/5 = 0.4 \text{ (40\%)}$ |

FP-Growth

| Gyakori elemhalmaz | Support érték |
|--------------------|--------------------------------------|
| {A} | $4/5 = 80\% \checkmark$ |
| {A, B} | $3/5 = 60\% \checkmark$ |
| {A, B, C} | $2/5 = 40\% \times \text{ (kiesik)}$ |
| {A, C} | $3/5 = 60\% \checkmark$ |
| {C} | $4/5 = 80\% \checkmark$ |
| {B, C} | $2/5 = 40\% \times \text{ (kiesik)}$ |
| {B} | $3/5 = 60\% \checkmark$ |