

SELECTING IMPACTFUL PRODUCT FEATURES

using statistics and machine learning

ȘTEFAN NICULAE

C U P R I N S

Enunțul problemei

Vedere de ansamblu

Context

Analiza datelor

Metode statistice

Pipeline optimizare învățare

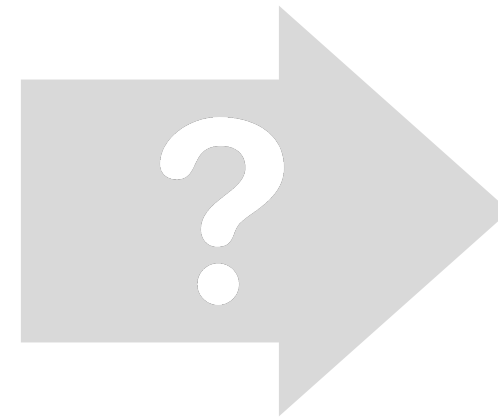
Meta-clasificator

Comparare modele

Clasament caracteristici

ENUNȚUL PROBLEMEI

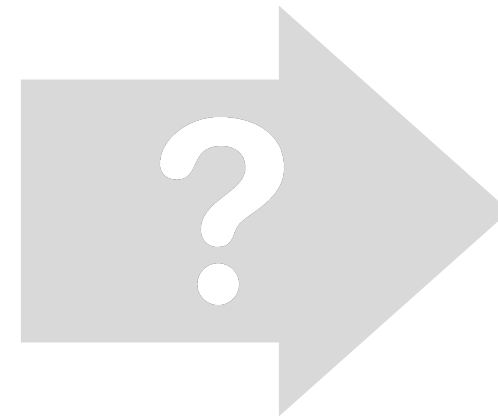
o aplicație



mai bună

ENUNȚUL PROBLEMEI

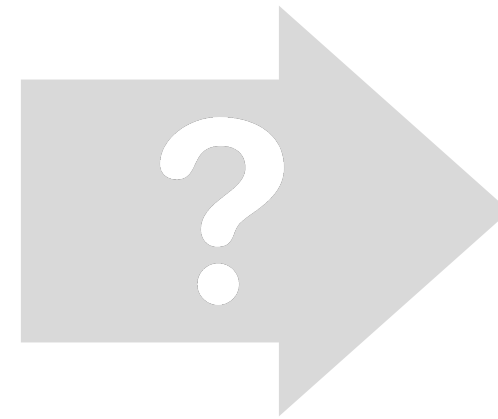
o aplicație



mai bună
mai mult succes

ENUNȚUL PROBLEMEI

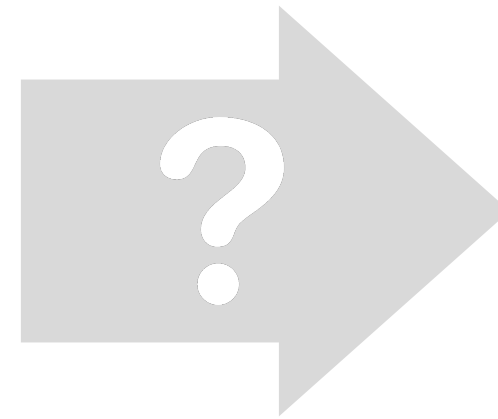
o aplicație



mai bună
mai mult succes
mai mulți clienți

ENUNȚUL PROBLEMEI

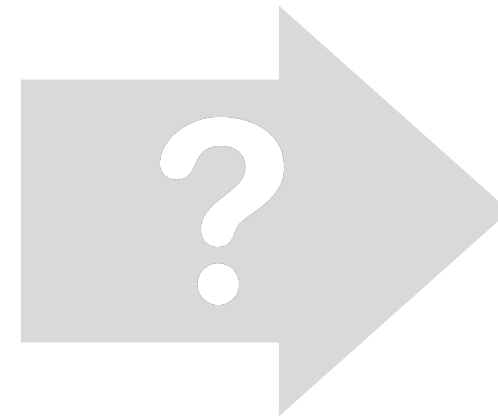
o aplicație



mai bună
mai mult succes
mai mulți clienți
retenție sporită

ENUNȚUL PROBLEMEI

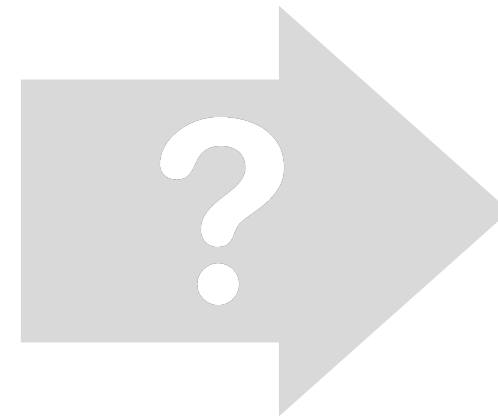
o aplicație



mai bună
mai mult succes
mai mulți clienți
retenție sporită
funcționalități care impactează retenția

ENUNȚUL PROBLEMEI

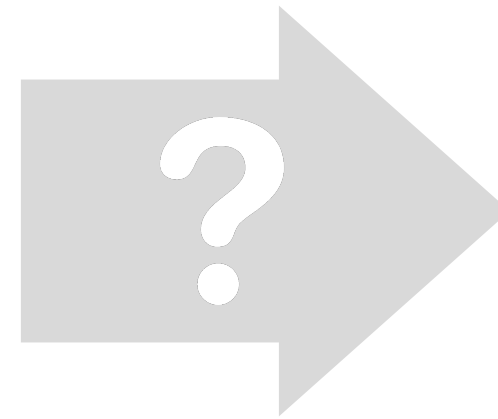
o aplicație
modul în care s-a folosit o aplicație



mai bună
mai mult succes
mai mulți clienți
retenție sporită
funcționalități care impactează retenția

ENUNȚUL PROBLEMEI

o aplicație
modul în care s-a folosit o aplicație
loguri ale utilizatorilor într-o aplicație



mai bună
mai mult succes
mai mulți clienți
retenție sporită
funcționalități care impactează retenția

Problemă de machine learning!

MACHINE LEARNING

antrenăm pe o mulțime de exemple etichetate,
apoi obținem eticheta pentru un exemplu nou

CAZUL NOSTRU

pornind de la o mulțime mare de date,
care sunt cele mai discriminatorii funcționalități?

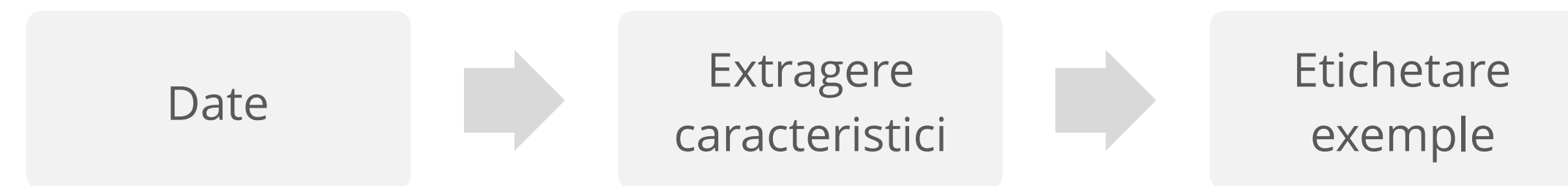
A B O R D A R E

antrenăm un model
care va învăța relațiile dintre date;
întrebăm ce l-a ajutat cel mai mult să decidă

cu cât mai precis modelul,
cu atât mai mare încredere în "opinia" sa

VEDERE DE ANSAMBLU

PREGĂTIRE

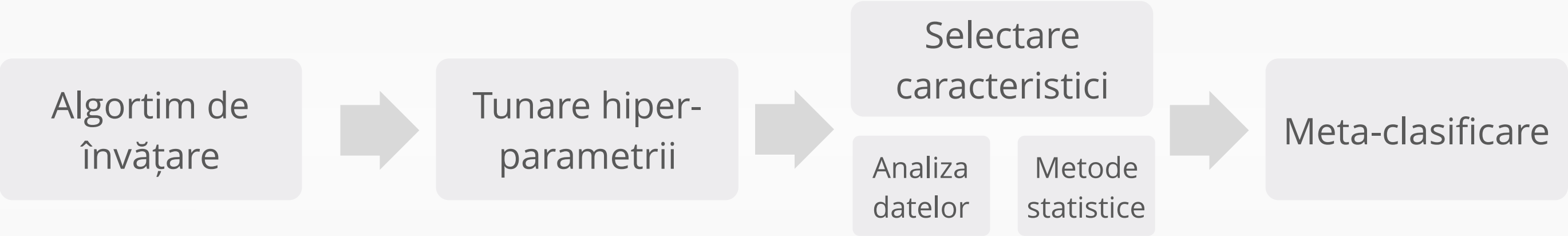


VEDERE DE ANSAMBLU

PREGĂTIRE



OPTIMIZARE

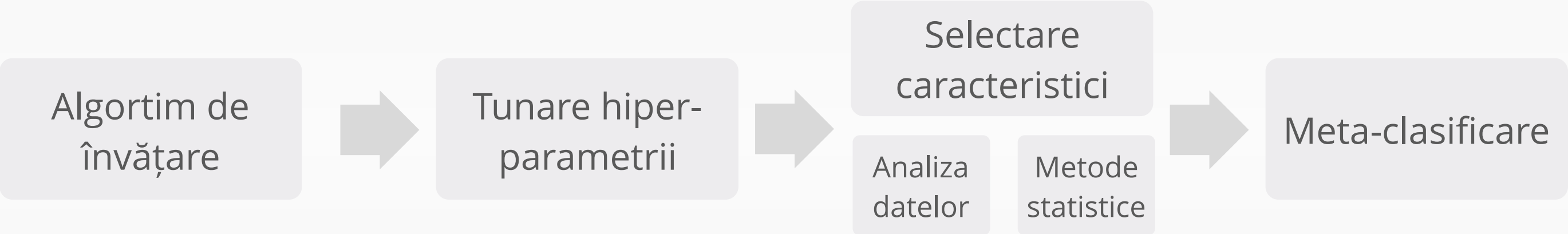


VEDERE DE ANSAMBLU

PREȚĂTIRE



OPTIMIZARE



REZULTATE

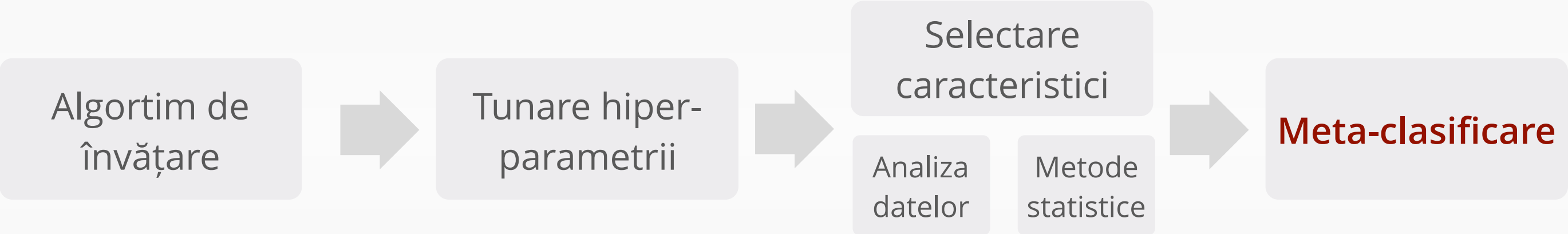


VEDERE DE ANSAMBLU

PREȚĂTIRE



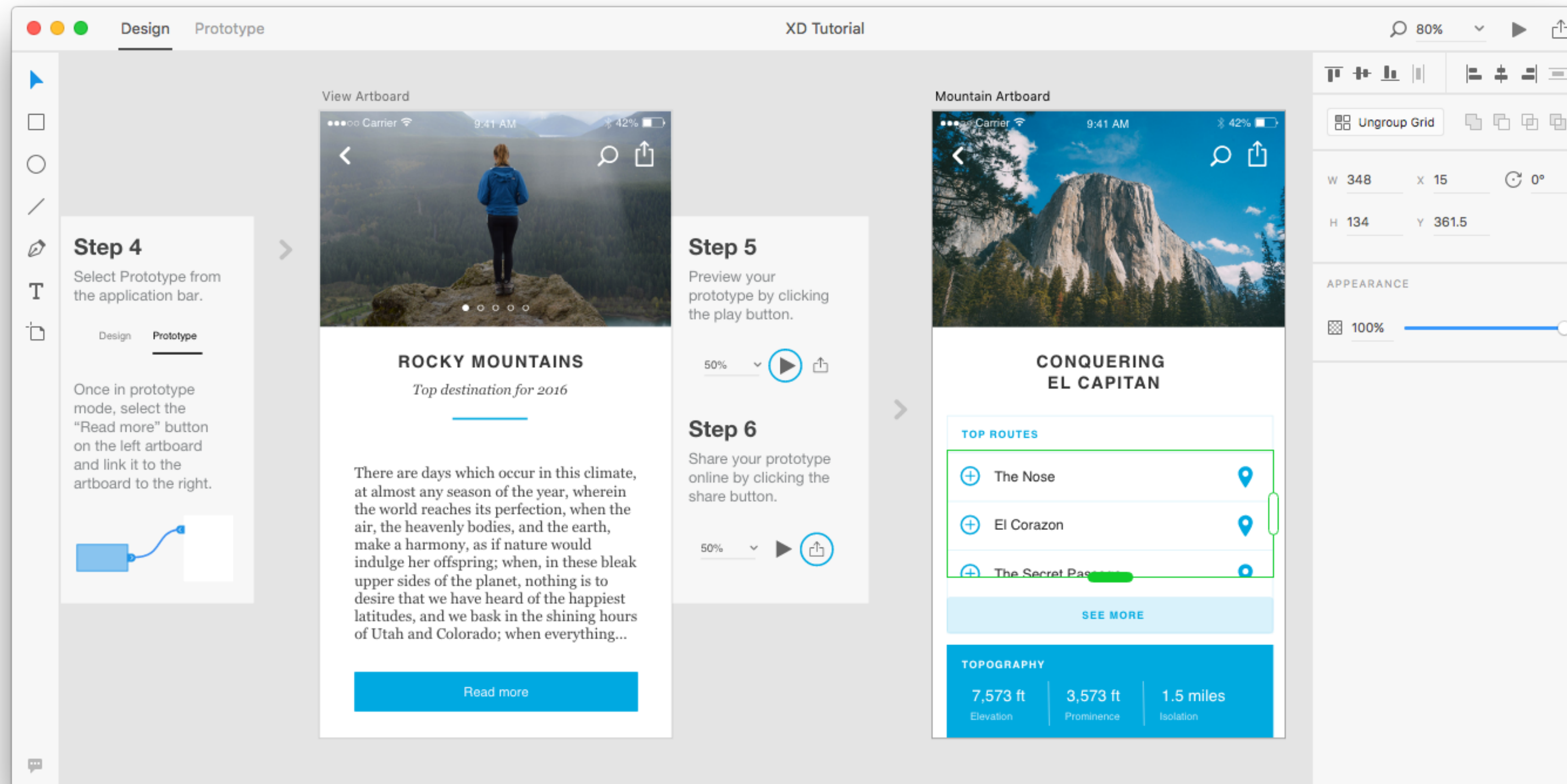
OPTIMIZARE



REZULTATE



APLICAȚIA



Xd Adobe Experience Design

DATE I M E D I A T E

pentru mai 2016

43 k

utilizatori

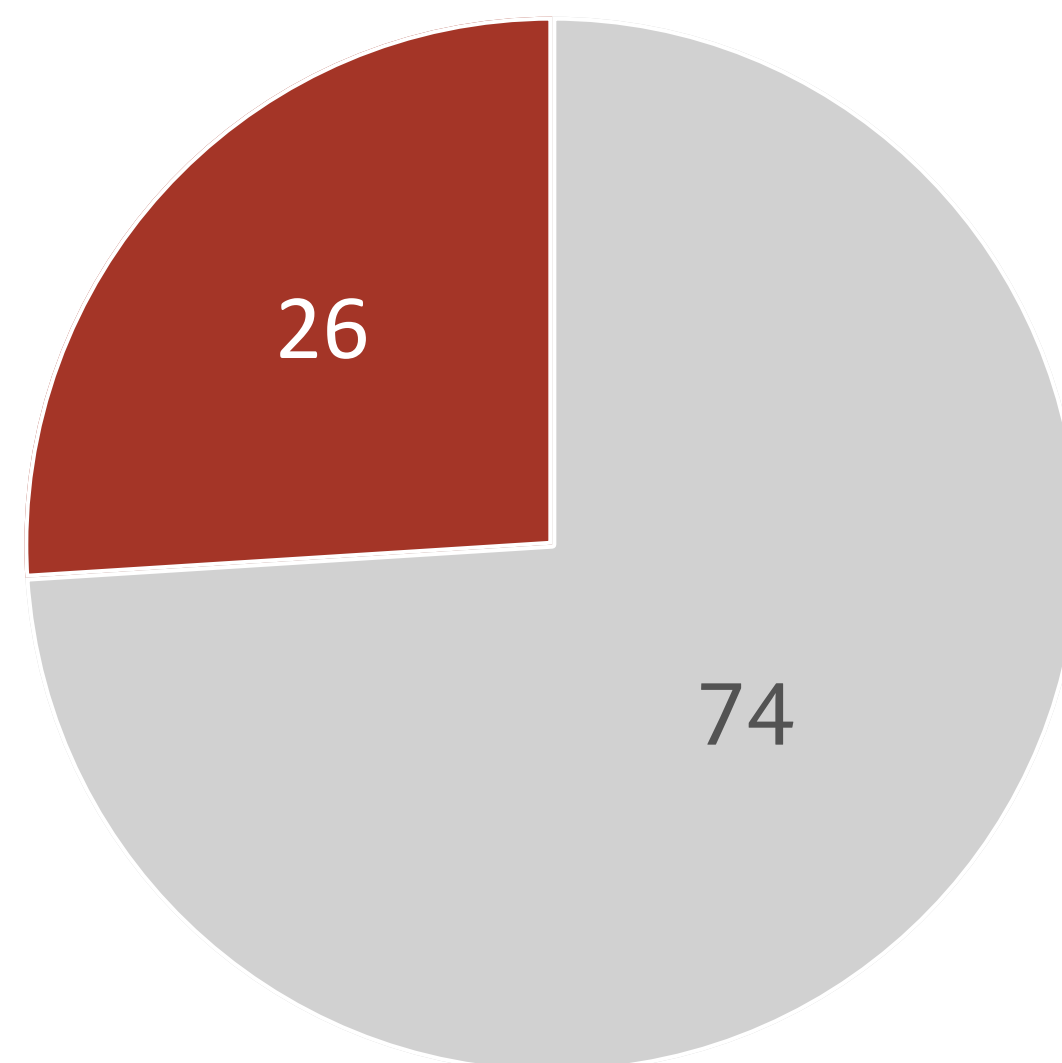
115 k

sesiuni

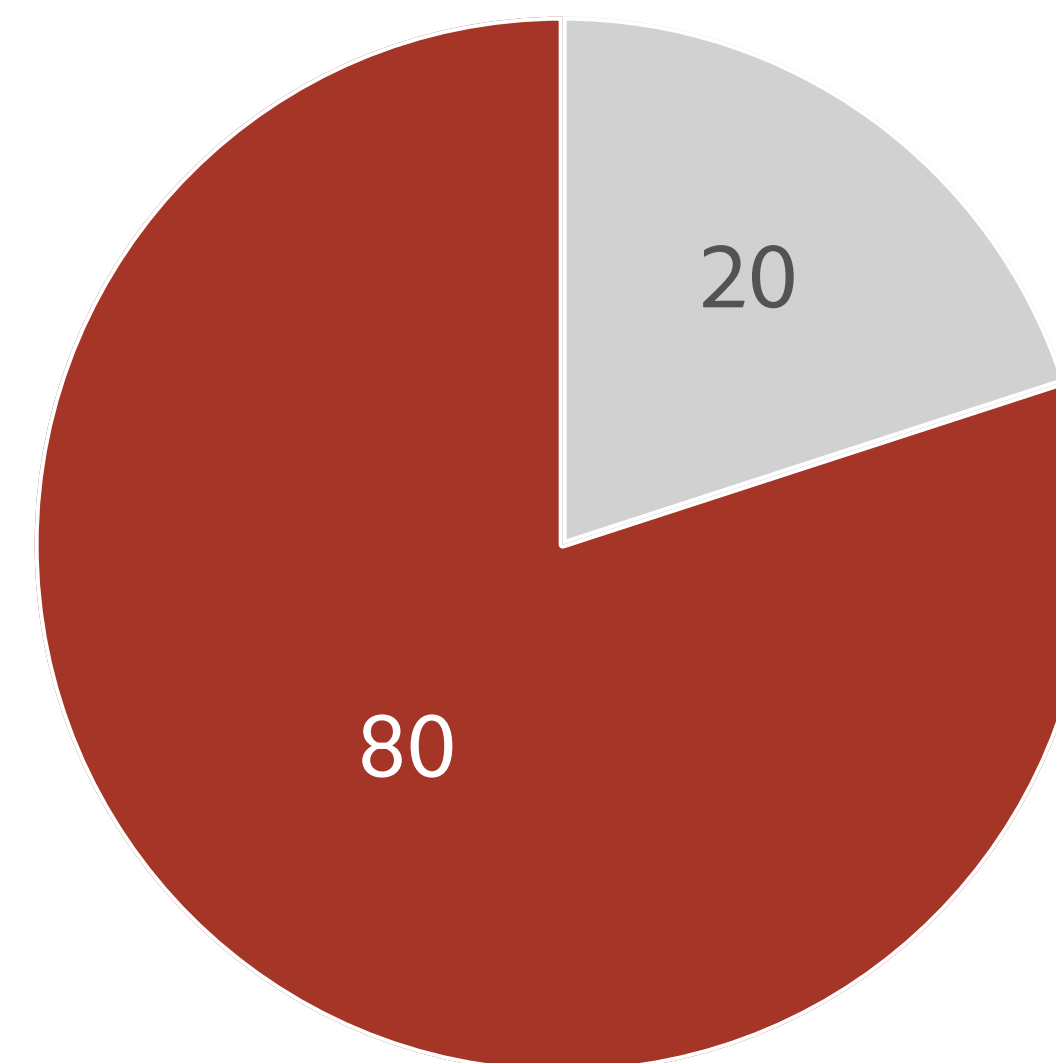
4.8 m

log-uri

ÎMPĂRȚIRE UTILIZATORI



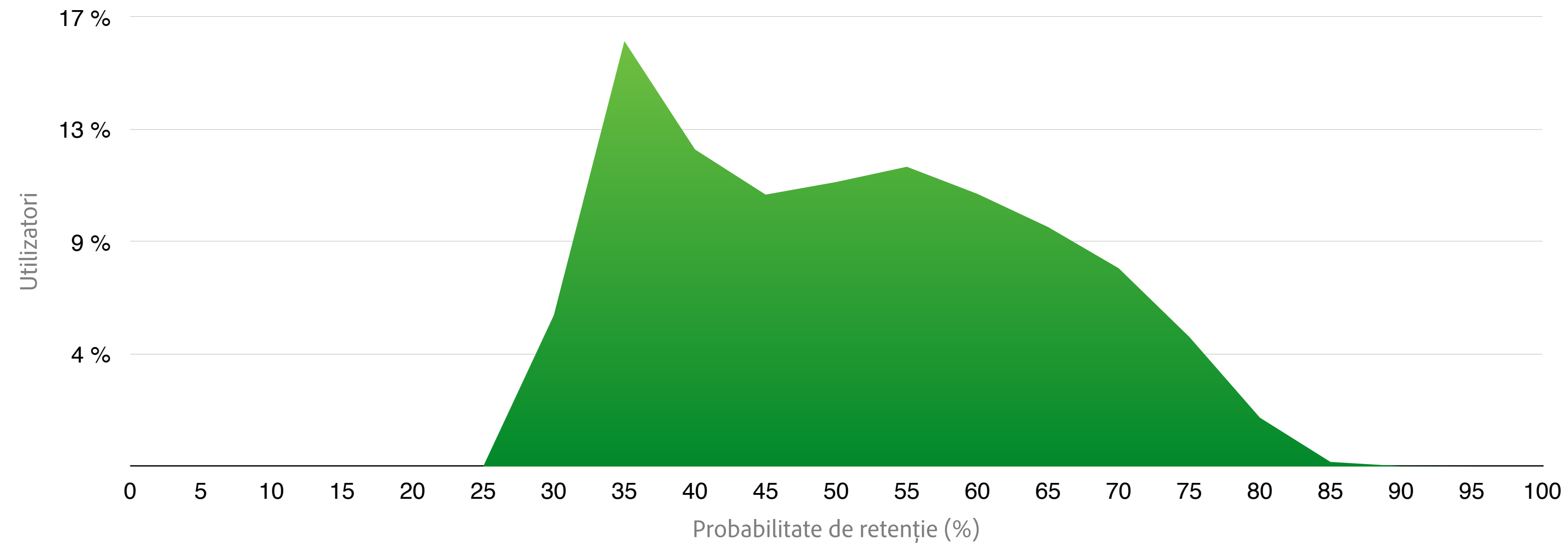
utilizatori



timp petrecut

● churneri
● retaineri

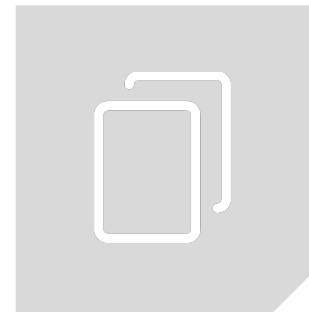
DISTRIBUȚIA RETENȚIEI



C A R A C T E R I S T I C I E X T R A S E

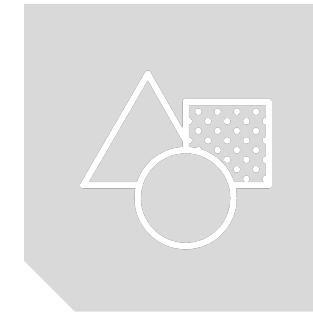
DOCUMENT

deschidere, creare, salvare
importare, exportare, partajare



DESEN

dreptunghiuri, elipse, drepte, traiectorii,
text, planșe, repeat-grid-uri, fire prototipare



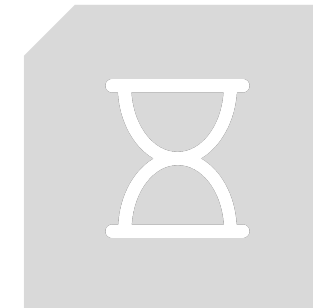
ISTORICE

timp total, număr de lansări,
perioadă de timp



TIMP

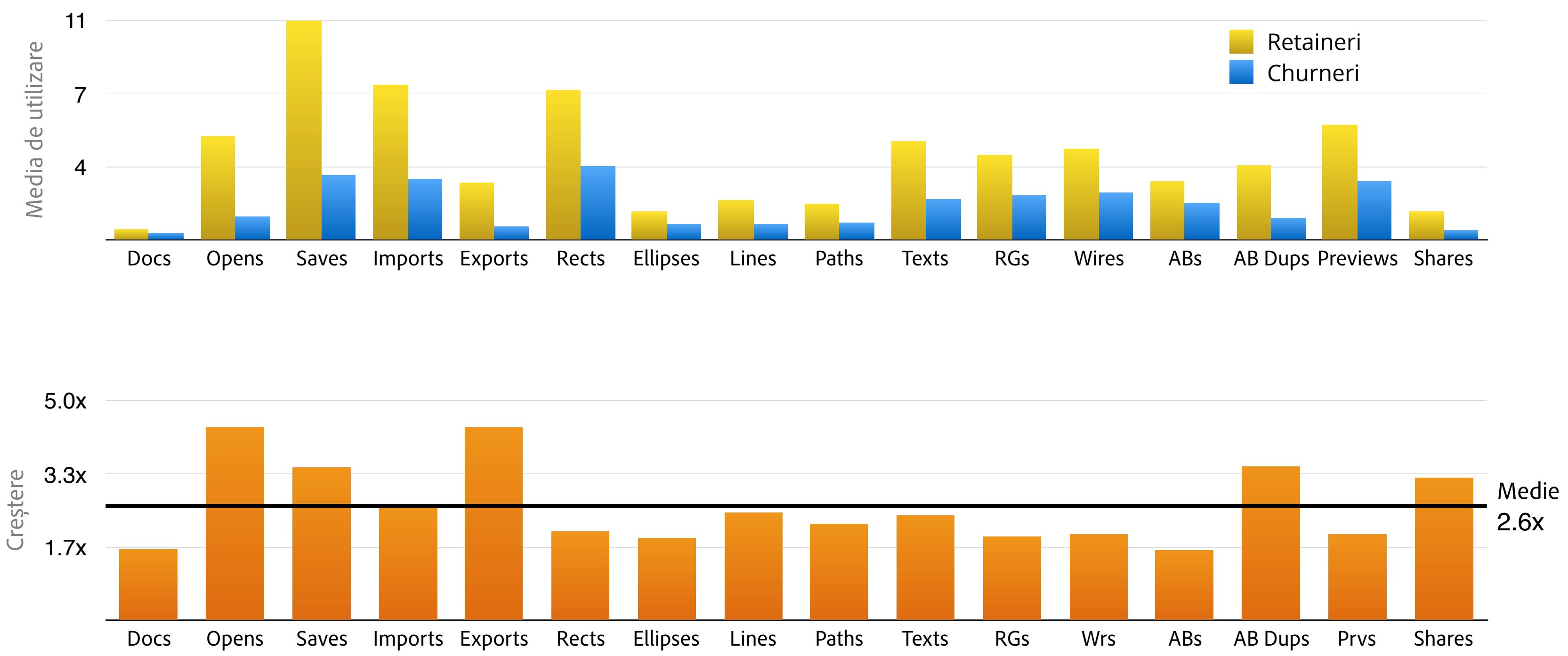
în design, prototipare, navigare,
prima sesiune; frecvența acțiunilor



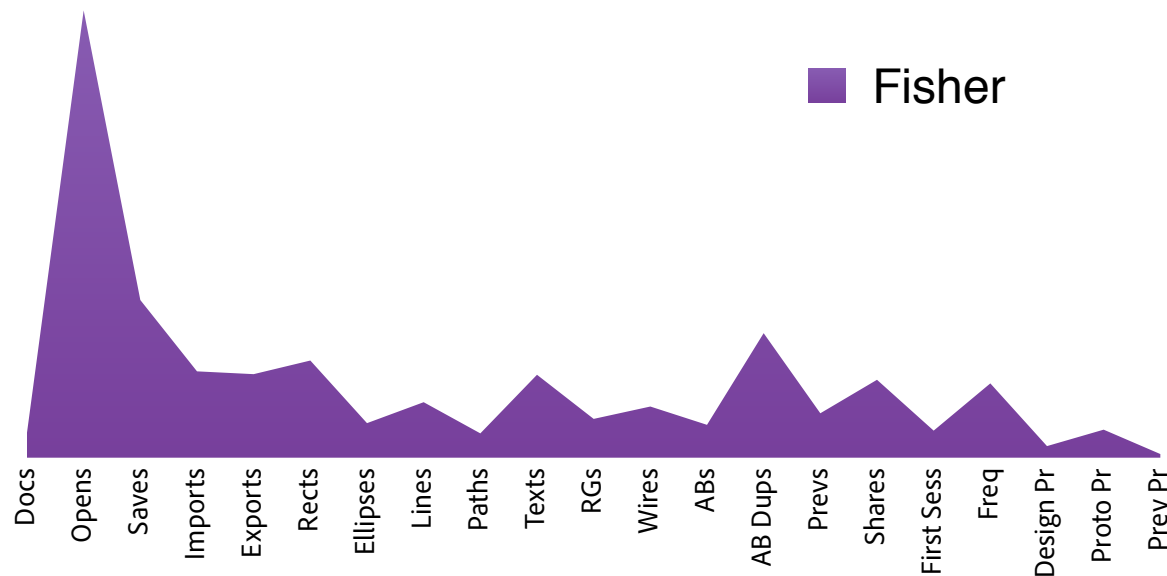
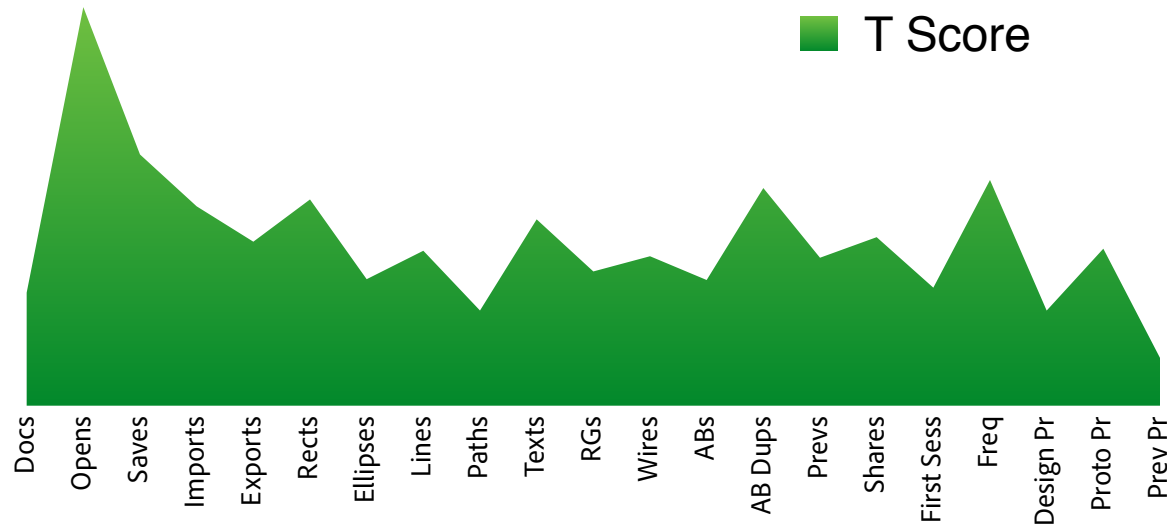
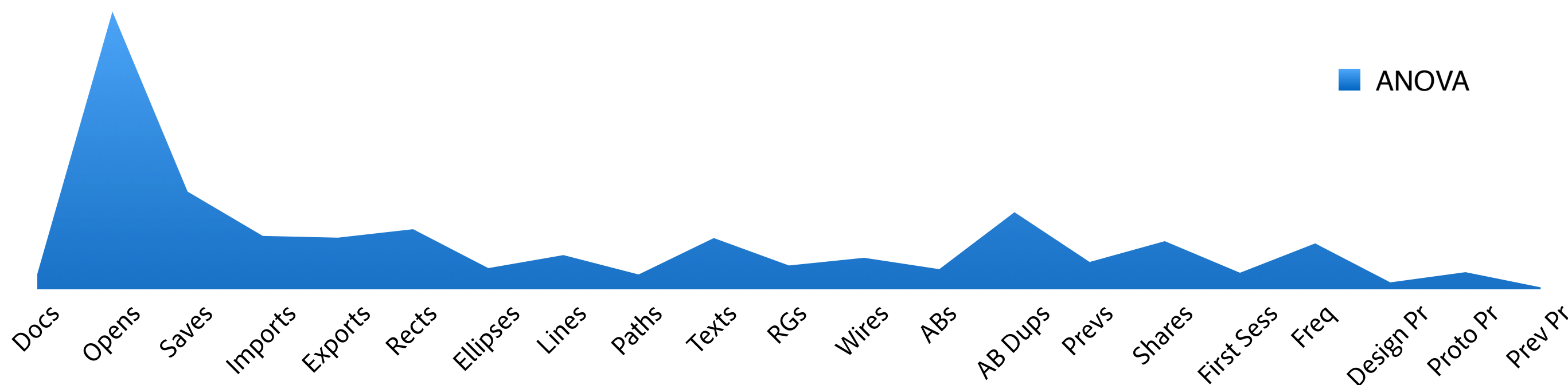
Doar numărând ignorăm secvențialitatea acțiunilor.

Construim **secvențe de acțiuni**

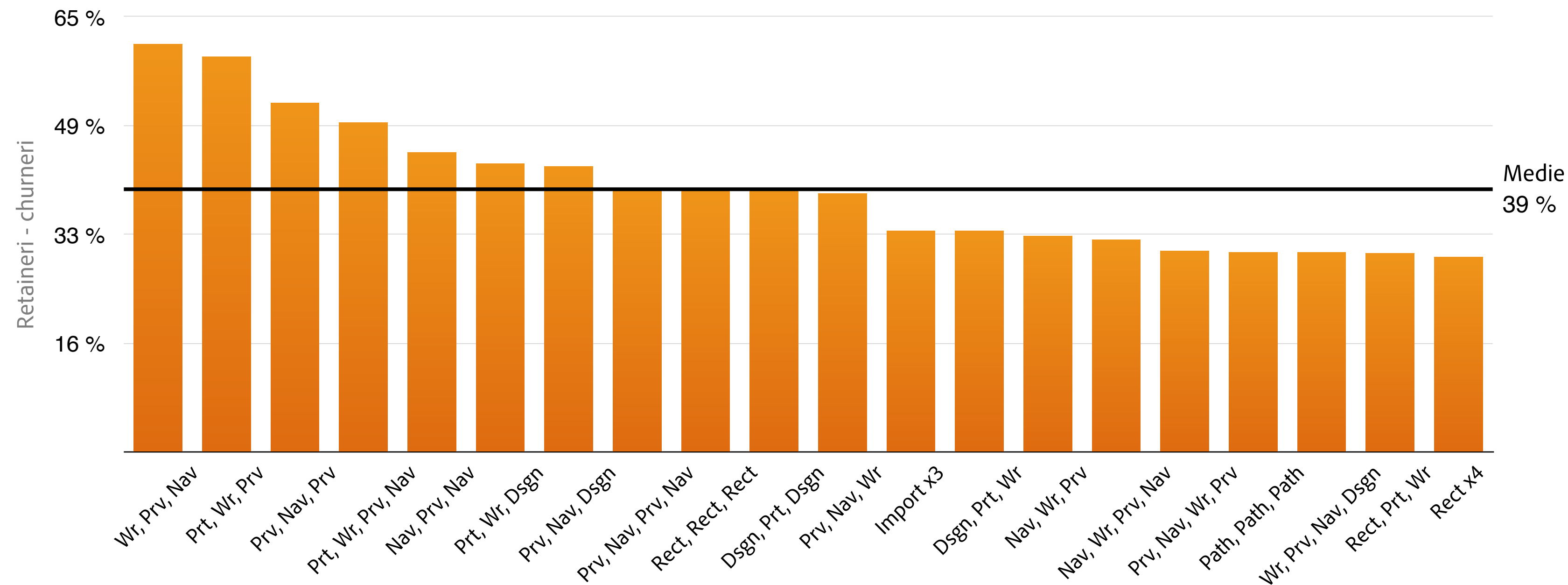
NUMĂR MEDIU DE UTILIZĂRI



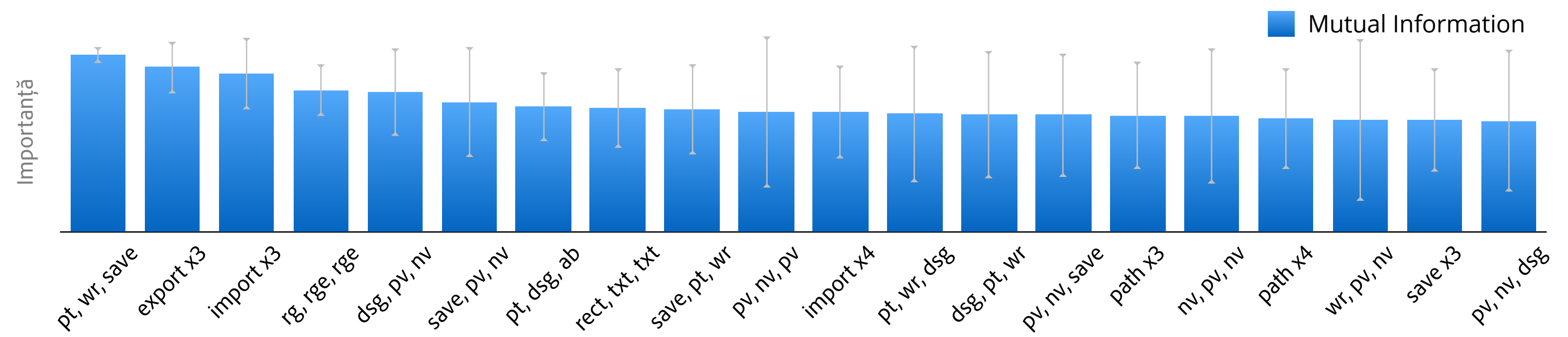
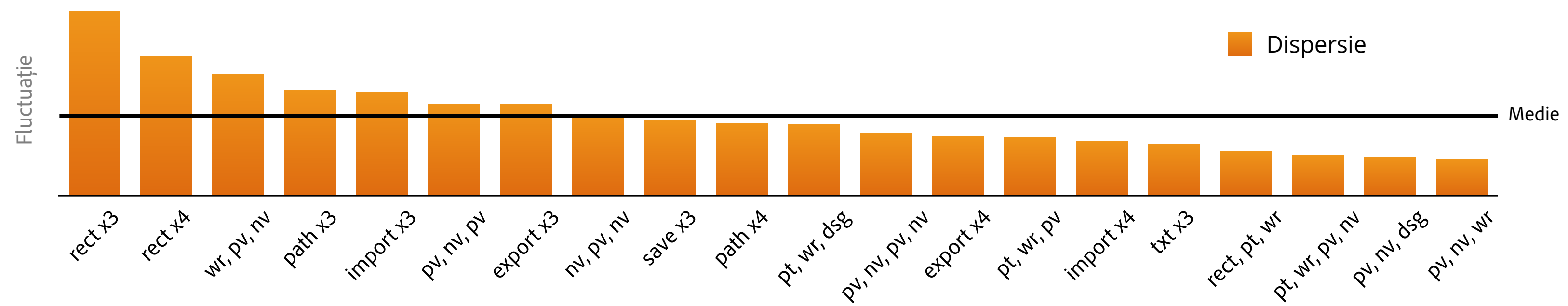
SCORURI STATISTIC



DIFERENȚA DINTRE UTILIZATORI



M E T O D E S T A T I S T I C E



OPTIMIZAREA ÎNVĂȚĂRII

PIPELINE

- găsește cel mai bun model pentru o problemă dată
- reduce necesitatea input-ului uman
- compensează pentru lipsa de experiență

OPTIMIZAREA ÎNVĂȚĂRII

P A Ș I

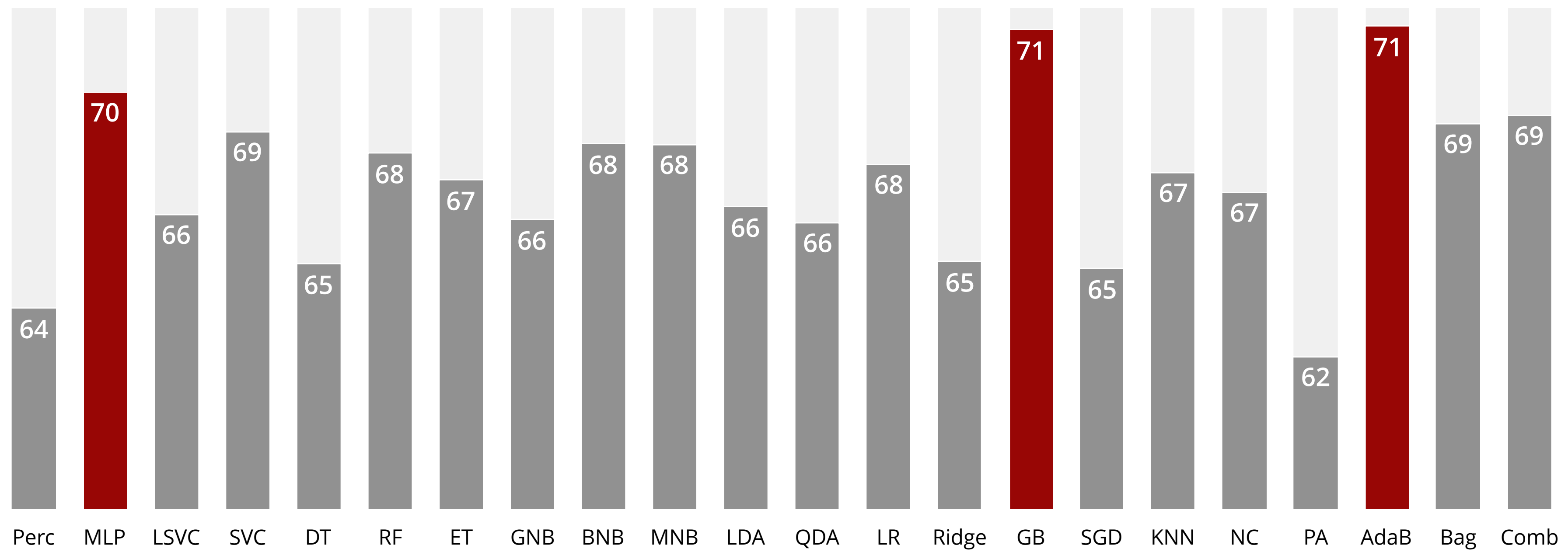
- 1 **Restrânge** precizia cautării în funcție de timpul rulării pe un eșantion
- 2 Ajustează **hiper-parametrii**, prin căutare aleatoare
- 3 Alege cea mai bună **submulțime de caracteristici** prin:
 - recursive feature elimination
 - submulțimea rezultată din **analiza datelor**
 - submulțimea rezultată din **metode statistice**
- 4 Combină cele mai bune modele cu un **meta-clasificator**

M E T A - C L A S I F I C A T O R

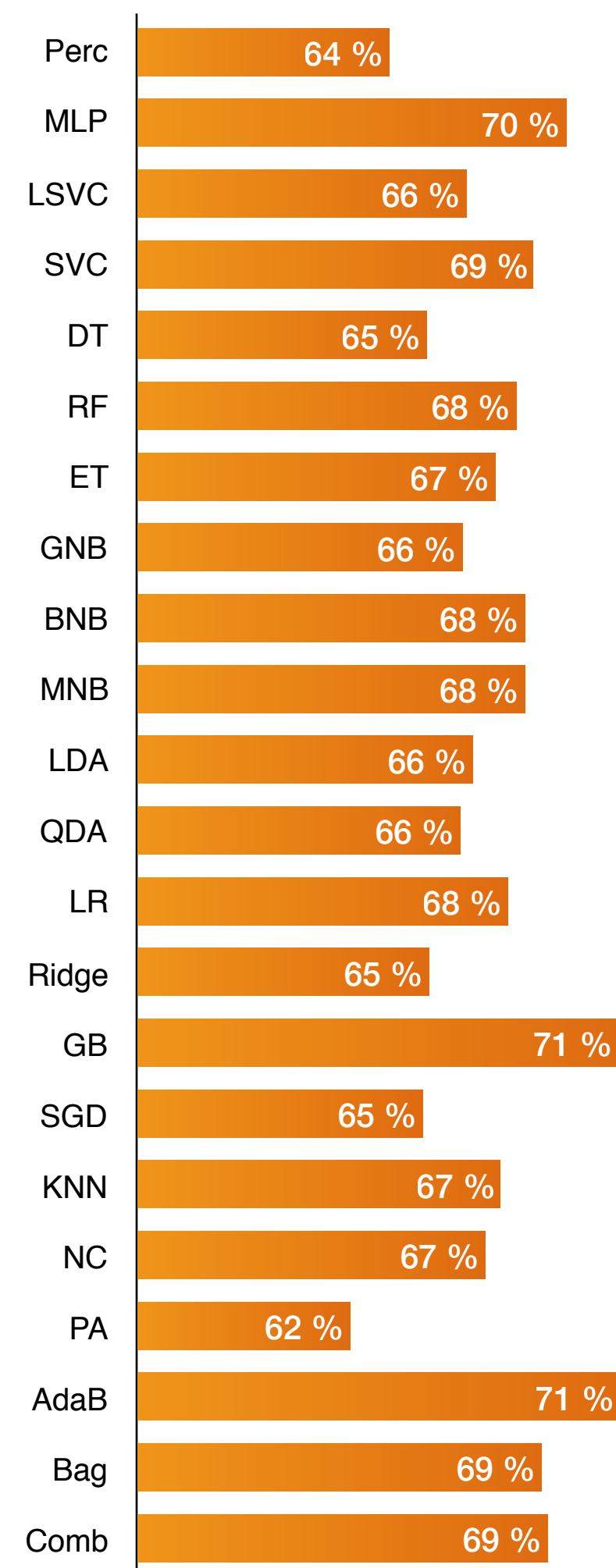
clasifică pe baza deciziilor altor modele,
nu pe caracteristicile exemplelor în sine

- învață **cum** să învețe
- am antrenat deja multe modele
- învață cum să le combine
output-ul în cel mai bun mod

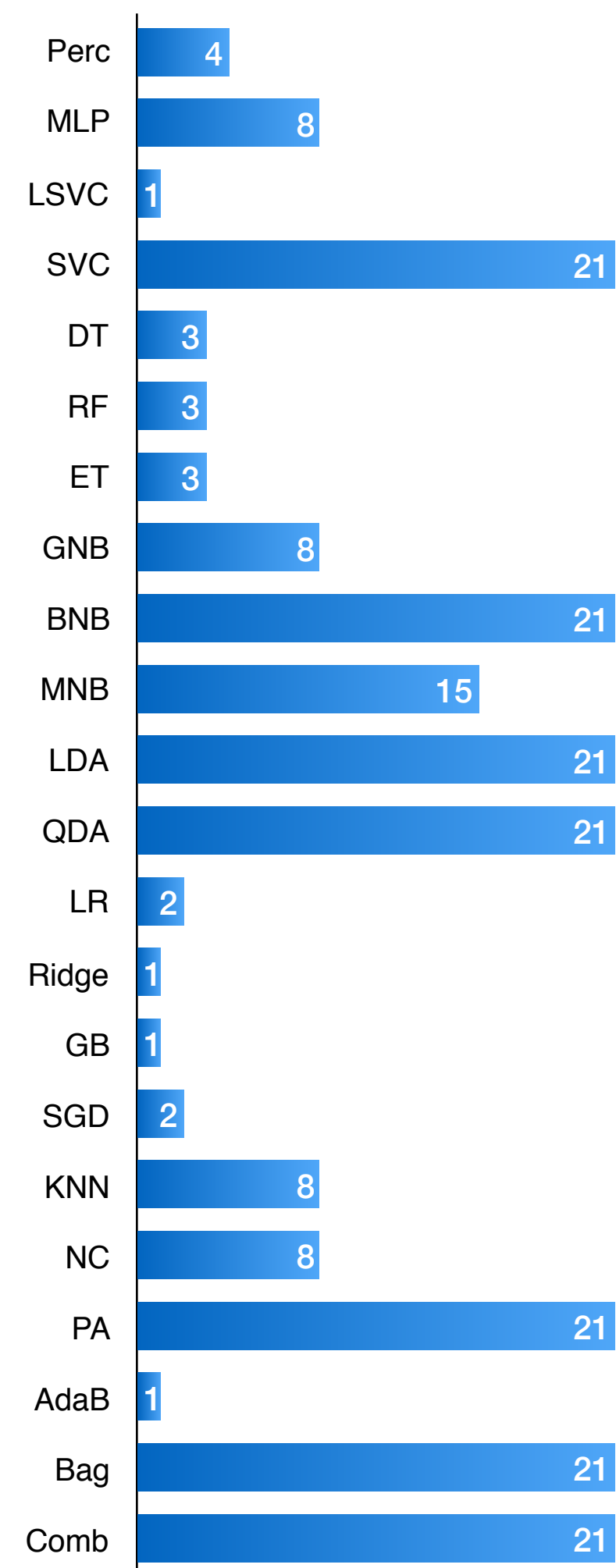
PERFORMANȚA MODELELOR



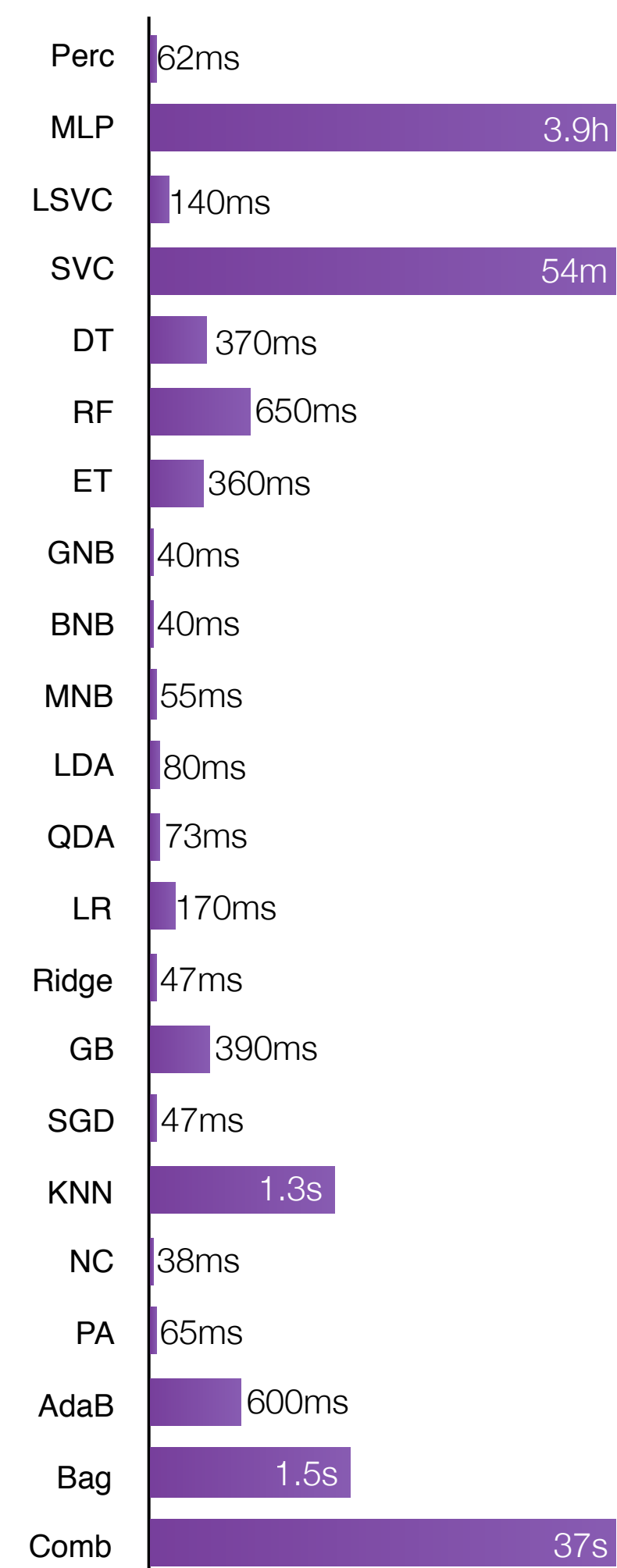
COMPARAȚIE MODELE



performanță

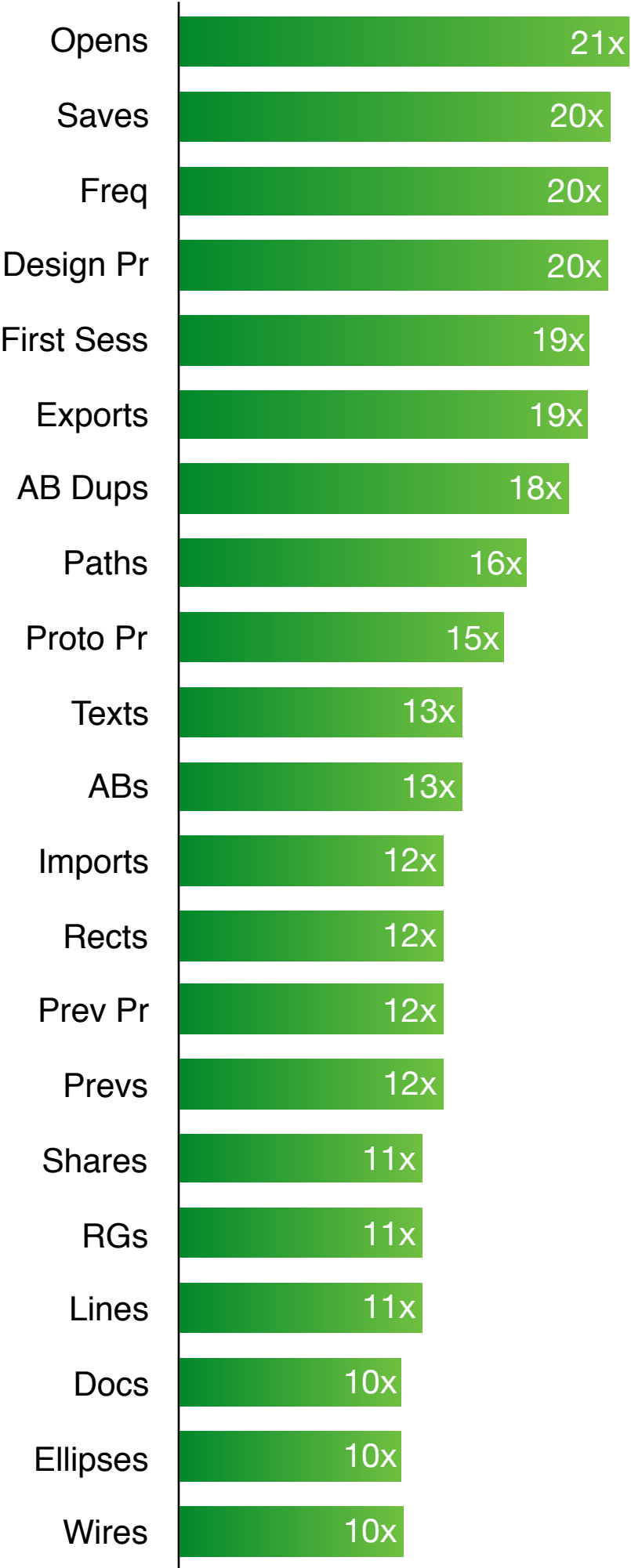


nr caracteristici

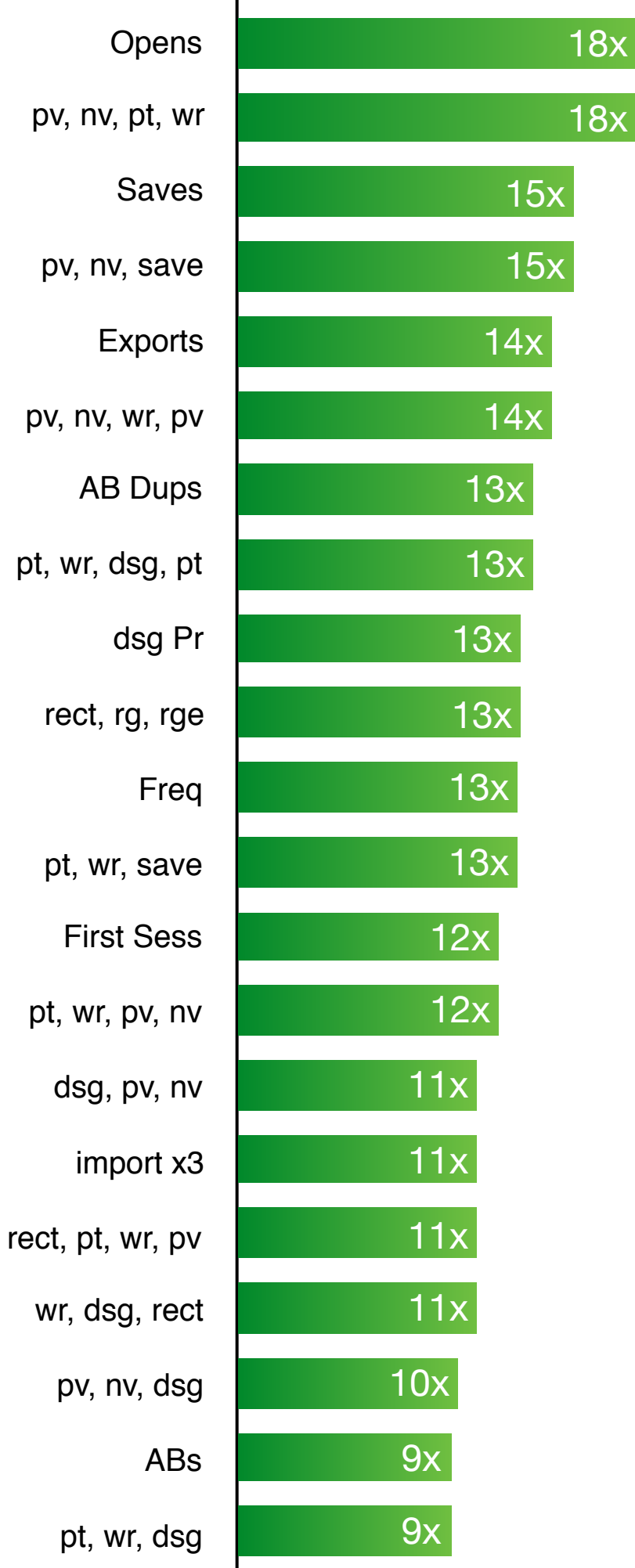


time învățare

CLASAMENT CARACTERISTICI



modele ce îl folosesc



& secvențe de acțiuni

CONCLUZII

FUNCȚIONALITĂȚI PRODUS

- prototipare, navigare
- planșe, repeat-grid-uri

CONTRIBUȚIA ÎN DOMENIU

- pipeline optimizare învățare
- meta-clasificator

U R M A T O R I I P A Ş I

RAFINAREA TEHNICII

- deep learning
- moduri noi de vizualizare

O NOUĂ DIRECȚIE

- sequence learning
- sistem predictiv pt acțiuni

M U L Ţ U M E S C

pentru atenție

B I B L I O G R A F I E

C. Bishop - *Pattern Recognition and Machine Learning*

T. Hastie et al - *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*

lista completă în lucrare

“An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem”

— John Tukey, mathematician