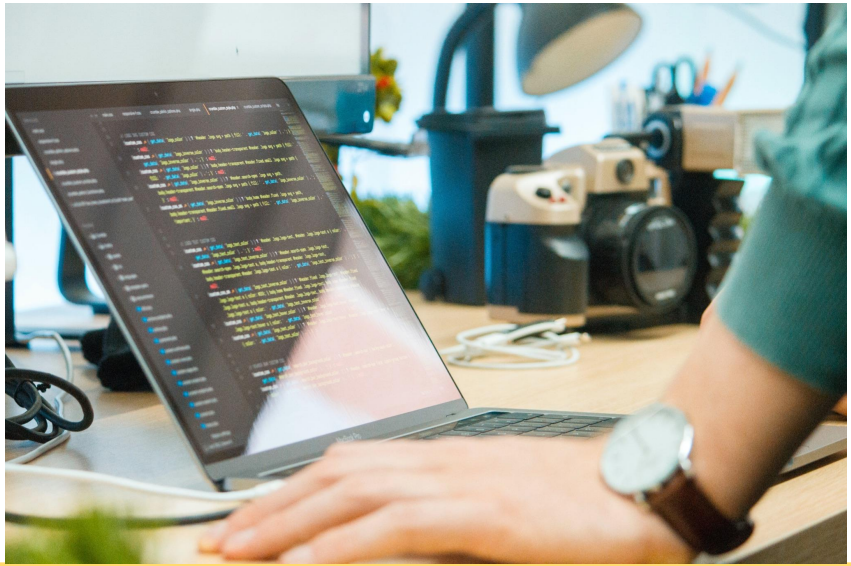
Indonesia AI
AI for Everyone, AI for Indonesia



MENTOR TALKS

Ngobrol topik santai tentang AI bareng mentor dan member di telegram

Tentang Komunitas Indonesia AI

Indonesia AI
AI for Everyone, AI for Indonesia

Bertumbuh Bersama Komunitas

Mari bergabung bersama ribuan lebih para pembelajar dan praktisi AI lainnya di komunitas Indonesia AI!

We are a community

Komunitas Indonesia AI berkomitmen untuk mengadakan program-program menarik dan tanpa biaya yang mengangkat ragam topik di dunia AI. Kamu juga bisa bertemu dengan para mentor kami di komunitas ini. Dari komunitas untuk komunitas.

We are a community, together we can!

Benefits

Networking

Berkenalan dengan para member lainnya melalui ragam program hingga event rutin yang diadakan.

Troubleshooting

Merasa sulit ketika coba buat AI sendiri? Kamu bisa coba konsultasikan masalah yang kamu hadapi ke para mentor kami atau member lainnya.

Collaboration

Kamu akan memiliki kesempatan untuk berkolaborasi bersama untuk mengadakan event yang mengundang para ahli AI di Indonesia.

Inspiration

Kami percaya kalau kamu bergerak bersama kamu bisa mendapatkan banyak inspirasi untuk hal yang tengah kamu lakukan dengan teknologi AI, termasuk informasi karir.

Tim Pengurus

Indonesia AI
AI for Everyone, AI for Indonesia

- **Registered members**

  Komunitas Indonesia AI terdiri dari pengurus inti dan para anggota yang teregistrasi sebanyak lebih dari 1000 members termasuk kamu!

- **Open Recruitment**

  Kedepan kami akan buka program Open Recruitment Pengurus Komunitas Indonesia AI 2021/2022, stay tuned terus yah!

Tim Pengurus

Muhammad Angga Muttaqien

Lead dan Mentor

Minat: Computer Vision, Natural Language Processing, Reinforcement Learning, Robotics

Domisili: Tokyo, Jepang

Muhammad Vikri

Vice Lead dan Mentor

Minat: Python Programming, Data Analytics

Domisili: Depok, Jawa Barat

Faris Dzaudan Qadri

Data Scientist dan Mentor

Minat: Data Science, Machine Learning

Domisili: Bremen, Jerman

Muhammad Iqbal

Software Engineer dan Mentor

Minat: Python Programming, Javascript Programming

Domisili: Kairo, Mesir

Muhammad Hendrawan Hidayat

Mentor

Minat: Python Programming, Computer Vision

Domisili: Surabaya, Jawa Timur

Hari Purnomo Sidik

Mentor

Minat: Deep Learning, Computer Vision, Natural Language Processing

Domisili: Makassar, Sulawesi Selatan

Networking Session

Indonesia AI
AI for Everyone, AI for Indonesia

Sesi networking (10 menit)

Yuk silahkan para member bisa memperkenalkan dirinya dulu :)

Mentor Talks
Catboost Classifier Algorithm

Catboost

Indonesia AI
AI for Everyone, AI for Indonesia

- What is Catboost?
  - A boosting method that focuses on processing **categorical** features and boosting trees some **“ordering principle”**.
  - The main take-away is to apply **ordering principle** in :
    1. Target encoding categorical features
    2. Boosting trees

CatBoost

Target Encoding

Indonesia AI
AI for Everyone, AI for Indonesia

- An efficient way to deal with categorical variables is to **substitute** them with **numerical values (usually some target statistics).**
- **Mean Target Encoding:** Replace categoricals with **mean target** value for them.

- Example

Color	Target
blue	0
red	1
blue	1
blue	1
green	0
red	0

Color	Mean_Target
blue	(0+1+1)/3 = 0.67
red	(1+0)/2 = 0.5
green	0/1 = 0

Color	enc_color	Target
blue	0.67	0
red	0.5	1
blue	0.67	1
blue	0.67	1
green	0	0
red	0.5	0

Target Encoding with Smoothing

Indonesia AI
AI for Everyone, AI for Indonesia

- We usually apply some smoothing in the calculation with a **prior term.**
- avg_target = (*count inclass* + *prior*) / (*total count* + 1)

- Example: Assume prior = 0.05

Color	Target
blue	0
red	1
blue	1
blue	1
green	0
red	0

Color	Mean_Target
blue	(2+0.05)/4 = 0.51
red	(1+0.05)/3 = 0.35
green	(0+0.05/2 = 0.025

Color	enc_color	Target
blue	0.51	0
red	0.35	1
blue	0.51	1
blue	0.51	1
green	0.025	0
red	0.35	0

Ordered Target Encoding

Indonesia AI
AI for Everyone, AI for Indonesia

- Why use **“ordered”** encoding? It helps prevent overfitting due to **“target leakage”.**
- Target statistics rely on the observed history.

- Example: Assume prior = 0.05

Color	Target
blue	0
red	1
blue	1
blue	1
green	0
red	0

Color	Mean_Target
blue	(0+0.05)/(0+1) = 0.05
red	(0+0.05)/(0+1) = 0.05

Still no red 1 before

Still no red = 1 before

Ordered Target Encoding

Indonesia AI
AI for Everyone, AI for Indonesia

- Why use **“ordered”** encoding? It helps prevent overfitting due to **“target leakage”.**
- Target statistics rely on the observed history.

- Example: Assume prior = 0.05

Color	Target
blue	0
red	1
blue	1
blue	1
green	0
red	0

Color	Mean_Target
blue	(0+0.05)/(0+1) = 0.05
red	(0+0.05)/(0+1) = 0.05
blue	(0+0.05)/(1+1) = 0.025

Now one blue before

Still no blue = 1 before

Ordered Target Encoding

Indonesia AI
AI for Everyone, AI for Indonesia

- Why use **“ordered”** encoding? It helps prevent overfitting due to **“target leakage”.**
- Target statistics rely on the observed history.

- Example: Assume prior = 0.05

Color	Target
blue	0
red	1
blue	1
blue	1
green	0
red	0

Now one red = 1 before

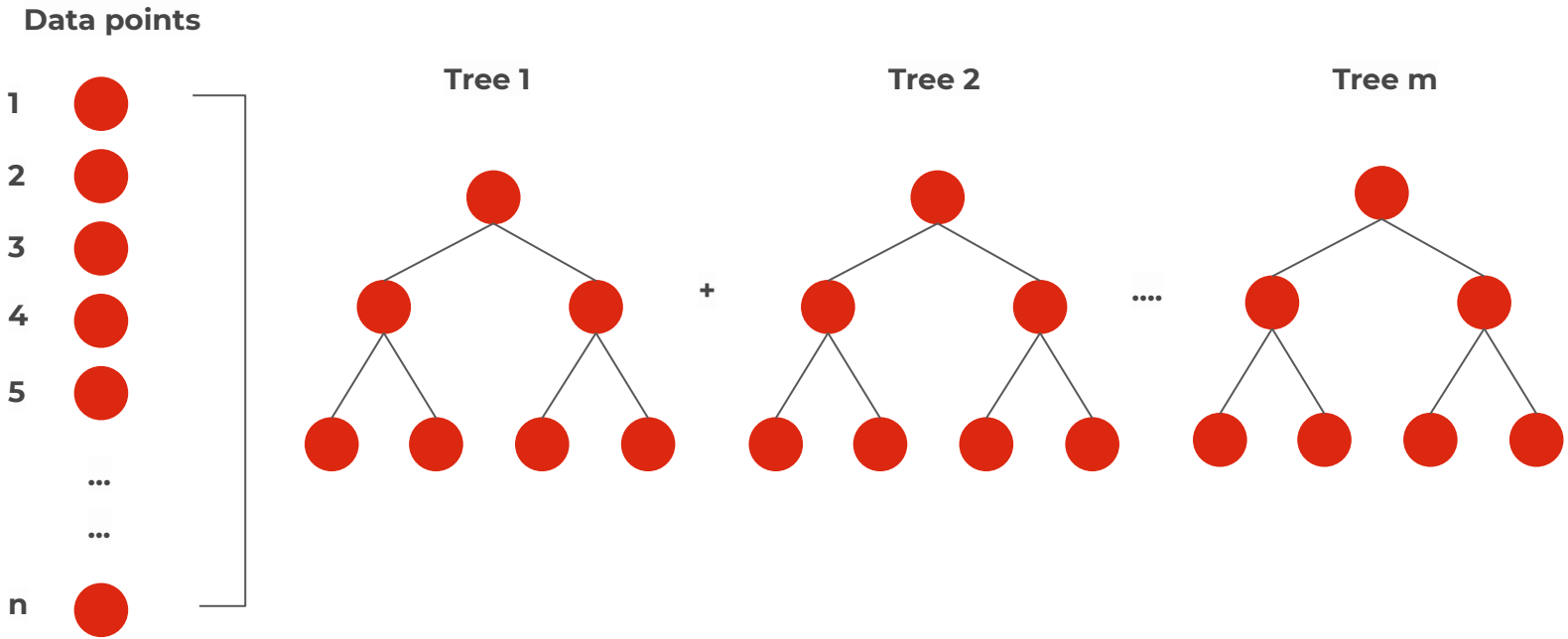
Color	Mean_Target
blue	(0+0.05)/(0+1) = 0.05
red	(0+0.05)/(0+1) = 0.05
blue	(0+0.05)/(1+1) = 0.025
blue	(1+0.05)/(2+1) = 0.35
green	(0+0.05)/(0+1) = 0.05
red	(1+0.05)/(1+1) = 0.05

Now one red before

Classical Boosting

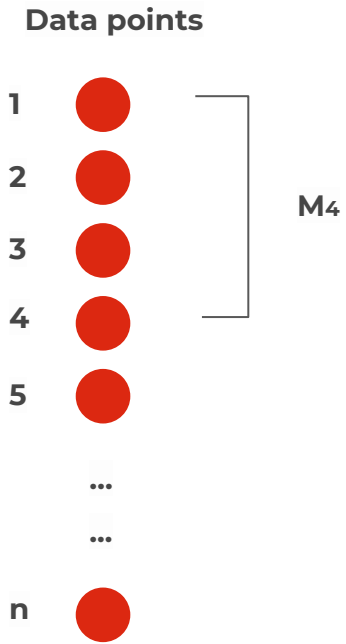
Indonesia AI
AI for Everyone, AI for Indonesia

In classical boosting, we fit multiple trees using the **whole dataset (x_n).** This can be lead to **overfitting.**



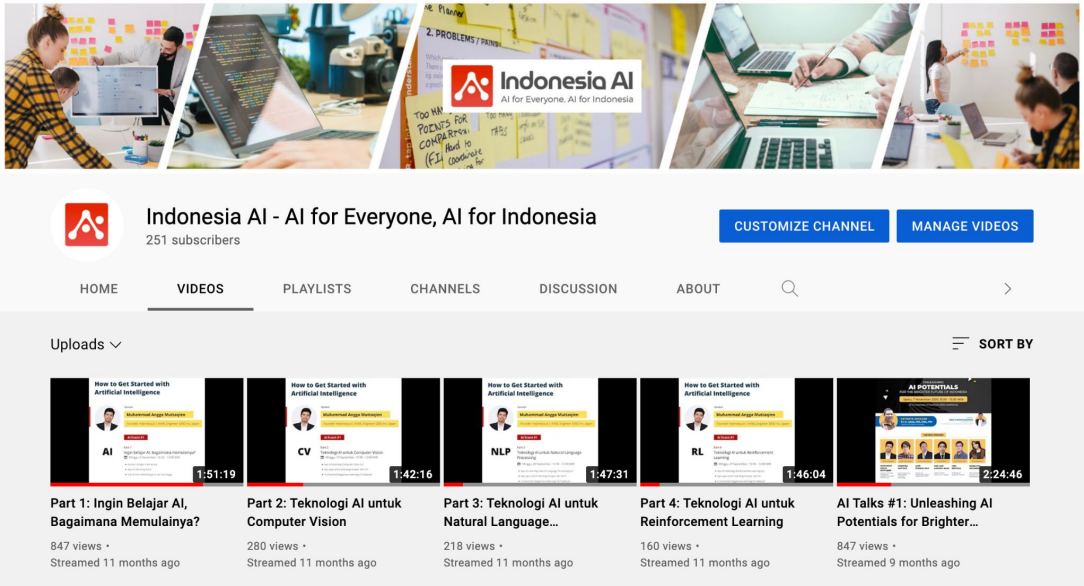
Ordered Boosting

- Assume model M_i was trained on the first i data points.
- We compute residuals at each data point i using model M_{i-1} (use a tree that didn't see that data point before)



Our Channel

Indonesia AI
AI for Everyone, AI for Indonesia



Find us on [Youtube]

Indonesia AI
AI for Everyone, AI for Indonesia

Terima Kasih!

Indonesia AI | AI for Everyone, AI for Indonesia

contact@aiforindonesia.org