Real Life vs Competitions

Real World ML Pipeline

It's a complicated process included:

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

- Way to evaluate model in real life

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

- Way to evaluate model in real life

- Way to deploy model

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

- Way to evaluate model in real life

- Way to deploy model

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

- Way to evaluate model in real life

- Way to deploy model

Real World ML Pipeline

It's a complicated process included:

- Understanding of business problem

- Problem formalization

- Data collecting

- Data preprocessing

- Modelling

- Way to evaluate model in real life

- Way to deploy model

Things we need to care about

Aspect	Real Life	Competition
Problem formalization	Y	N
Choice of target metric	Y	N
Deployment issues	Y	N
Inference speed	Y	N
Data collecting	Y	N/Y
Model complexity	Y	N/Y
Target metric value	Y	Y

Things we need to care about

Aspect	Real Life	Competition
Problem formalization	Y	N
Choice of target metric	Y	N
Deployment issues	Y	N
Inference speed	Y	N
Data collecting	Y	N/Y
Model complexity	Y	N/Y
Target metric value	Y	Y

The only thing we really care about

Conclusion

- Real-world problems are quite complicated

- Competition are a great way to learn

- But they don't address the questions of formalization, deployment and testing

Philosophy of competitions

It's not only about algorithms

- It is all about data and making things work, not about algorithms itself
    - Everyone can and will tune classic approaches
    - We need some insights to win
- Sometimes there is no ML

Do not limit yourself

It is ok to use
- Heuristics
- Manual data analysis

Do not be afraid of
- Complex solutions
- Advanced feature engineering
- Doing huge calculation

Be creative

It's ok to modify or hack existing algorithms or even to design completely new algorithm

Do not be afraid of reading source codes and changing them

Enjoy!

- Do your best

- Enjoy the competition

- Get experience and fun

    They are much more valuable than prize money