# How to Avoid Machine Learning Pitfalls

Talha Ünal
Sabahattin Timur Uzun

# Contents

- Before Building Models
- How to Build Models Reliably?
- How to Evaluate Models Robustly?
- How to Compare Models Fairly?
- How to Report the Results?

# Before Building Models

## 1. Take time to understand the data

- Reliable Source of Data
- "*Garbage in Garbage Out*"
- Exploratory Data Analysis: Missing and inconsistent data

## 2. Don't look at all the data

- Good to spot patterns and make insights.
- Don't make **untestable** assumptions.
- Data leakage from testing set to training set.

# Before Building Models (cont'd)

## 3. Make sure data is enough

- Inverse Correlation between **noise ratio** and the **needed size** of the data
- Cross Validation and Data Augmentation for lack of data
- Overfitting is common problem in small data sets, identifying this issue early on is significiant.

## 4. Domain Expert consulting

- They can help to choose the appropriate feature set and ML model
- For example, an explainable model for medical or financial decisions may be desired.
- They can help to reach right people.

# Before Building Models (cont'd)

## 5. Survey Literature

- Academic progress is an iterative process.

## 6. Considering Model Deployment

- For many researches, eventual goal is to produce an ML model that can be deployed in a real world situation.
- Limited Resource Environment -> Choose Less Complex Models
- Limited Time Environment -> Choose Faster Models

# How to Reliably Build Models ?

## 1. Test data leakage into the training

- Such models often fail to generalize to real world data.
- To prevent this, partition off a subset at the beginning.
- Be careful partitioning time series data, random splits may cause leakage and overfitting

## 2. Try different models

- **No Free Lunch Theorem**
- Try to use appropriate models according to the problem.
- Use validation set, and optimize hyperparameters.

# How to Reliably Build Models ? (cont'd)

## 3. Don't use inappropriate models

- Choose proper type of features. If features types are not appropriate, transform the  features to prevent loss of information
- Common mistake: Choosing classification model where regression model would make more sense.
- Avoid using unnecessarily complex models and consider the deployment phase.

## 4. Deep Learning isn't always the best

- Weak in limited data.
- Consider the actual complexity.

# How to Reliably Build Models ? (cont'd)

## 5. Optimize hyperparameters

- There is no one-size-fits-all.
- Use traditional HPO techniques, such as Bayesian Optimizaiton and Gradient-based optimization
- Use AutoML techniques to optimize model choice and hyperparameters.

## 6. Feature selection

- Feature selection and HPO are part of model training.
- Don't apply to whole data set (Data leakage)
- **Nested Cross-Validation** as a solution.

# How to Robustly Evaluate  Models ?

## 1. Appropriate Test Set

- Shouldn't overlap the training set, should be representative of the wider population

## 2. When to apply data augmentation?

- **Apply** data augmentation **after** splitting the data.
- Apply on the training set, not test set.
- If we apply before splitting data, it causes the most dangerous data leakage scenario (Test samples are mostly variants of the training samples).

# How to Robustly Evaluate Models(cont'd)

## 3. Use a validation set

- Do not use the test set within training
- Separate validation set should be used to measure performance.
- Avoid using unnecessarily complex models and consider the deployment phase.

## 4 .Evaluate a model multiple times

If you train them multiple times, or if you make small changes to the training data, then their performance varies significantly So carry out multiple evaluations.

- Use Cross-validation (CV)
- Stratification is important
- report the mean and standard deviation of the multiple evaluations

# How to Robustly Evaluate Models(cont'd)

## 5.Save some data to evaluate your final model instance

The only way of getting a reliable estimate of a model instance's generality may be to use another test set. So, if you have enough data, it's better to keep some aside and only use it once to provide an unbiased estimate of the final selected model instance.

## 6 .Don't use accuracy with imbalanced data sets

Be careful which metrics you use to evaluate your ML models. For classification models, the most commonly used metric is accuracy ut many data sets are not balanced, and in this case accuracy can be a very misleading metric

# How to compare models fairly?

## 1. Don’t assume a bigger number means a better model

- Models migth be trained or evaluated on different partitions of the same data set
- Entirely different data sets might be used
- Failure to carry out the same amount of hyperparameter optimisation

## 2. Use statistical tests when comparing models

There are two categories of tests:

- Compare individual model instances
- Compare two models more genally

# How to compare models fairly?(cont'd)

## 3. Correct for multiple comparisons

Things get a bit more complicated when you want to use statistical tests to compare more than two models

- Multiplicity Effect

## 4. Don’t always believe results from community benchmarks

The idea is that, because everyone is using the same data to train and test their models, then comparisons will be more transparent.

- Developing to the test set

# How to compare models fairly?(cont'd)

## 5. Consider combinations of models

Ensemble model types:

- Form of same base model type: Approaches
- Combine different kinds of ML models

# How to report your results

## 1. Be transparent

- Share your models in an accessible way.
- Share the script when you publish the results.
- Use experiment tracking frameworks, such as MLflow

## 2. Report performance in multiple ways

There are two categories of tests:

- Use multiple data sets.
- Report multiple metrics for each data set

# How to report your results (cont'd)

## 3.Don’t generalise beyond the data

A common mistake is to make general statements that are not supported by the data used to train and evaluate models. Issues:

- Bias, or sampling error
- Overlap
- Issue of quality

## 4. Be careful when reporting statistical significance

Statistical tests are not perfect. They can be;

- Conservative
- Liberal

# How to compare models fairly?(cont'd)

## 5. Look at your models

Look inside your models and do try to understand how they reach a decision. The aim of research is not to get a slightly higher accuracy than everyone else. Rather, it’s to generate knowledge and understanding and share this with the research community.

- Visualise models,
- Use explainable AI (XAI) techniques to extract knowledge.

# Final thoughts

How to avoid machine learning pitfalls: a guide for academic researches

# Thank you for listening!

Talha Ünal
Sabahattin Timur Uzun