**Prof. Dr. Peer Küppers**

Summer School 2021

# Bias-Variance-Tradeoff

# Use Case

- **Simulated Webshop**

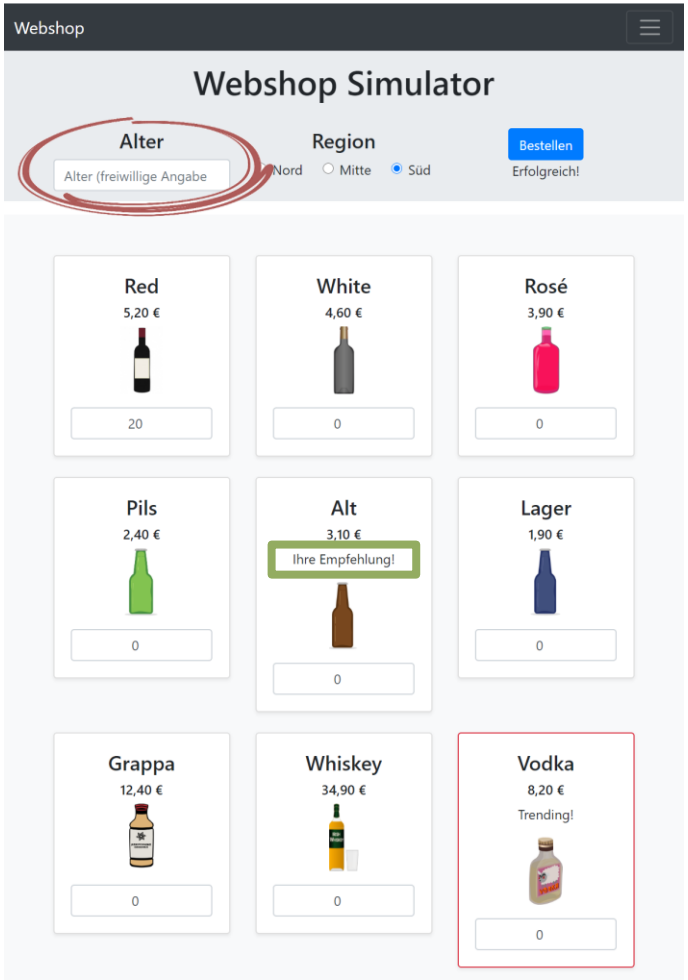
  https://pk-bigdata.appspot.com

- **Business Context**
  - "Personalized recommendations increase turnover."

  → **Machine learning-based personalized recommendation**

- **Problem**
  - Personalization requires knowledge about the age group of a customer

  → **Not all users provide their age.**



# Use Case

- **Task**
  - Age (group) missing? → **predict it!**

- **Approach**

*based on …*

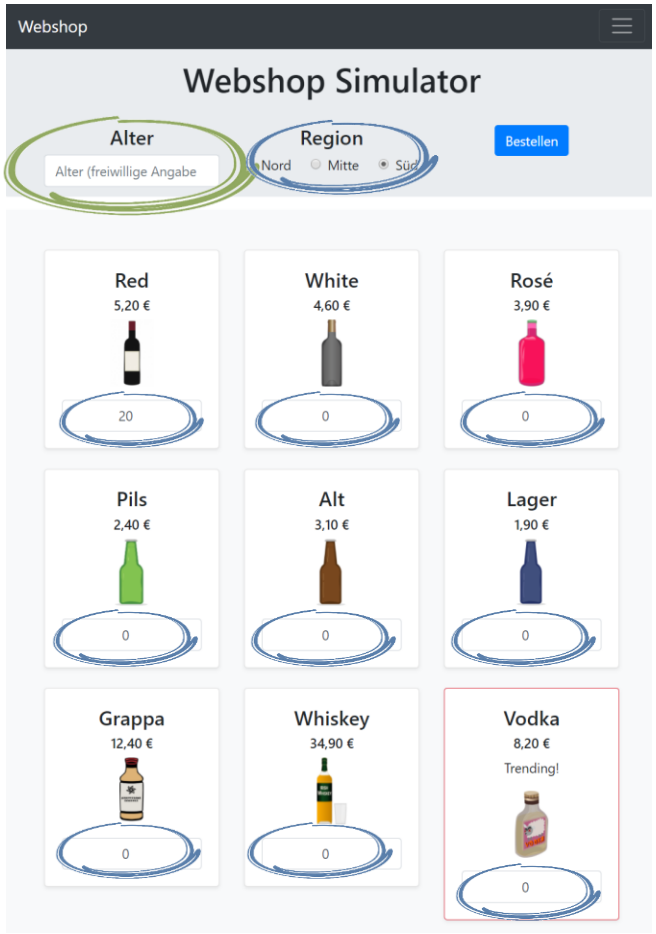
  - Region
  - Order date
  - Contents in the shopping cart

**Features**

*predict …*

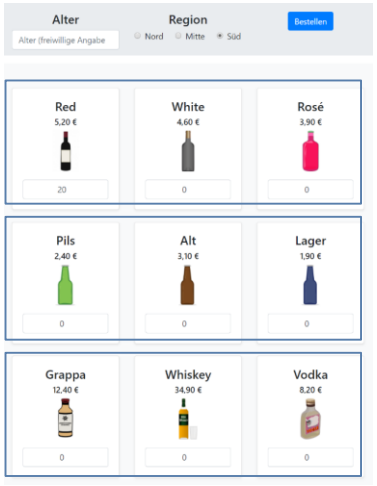
  - Age group

**Target**




+ Order Date

# Use Case: Data Understanding




Product group 1: wine

Product group 2: beer

Product group 3: spirits

| weekday | region | sum_sales_pg1 | sum_sales_pg2 | sum_sales_pg3 | age_group |
|---|---|---|---|---|---|
| Sunday | Middle | 99.0 | 0.0 | 0.0 | 18-29 |
| Wednesday | North | 8.1 | 91.9 | 0.0 | 18-29 |
| Friday | South | 68.8 | 19.3 | 0.0 | 18-29 |
| Thursday | South | 65.5 | 13.9 | 15.7 | 30-49 |
| Friday | Middle | 51.6 | 3.8 | 34.6 | 50- |

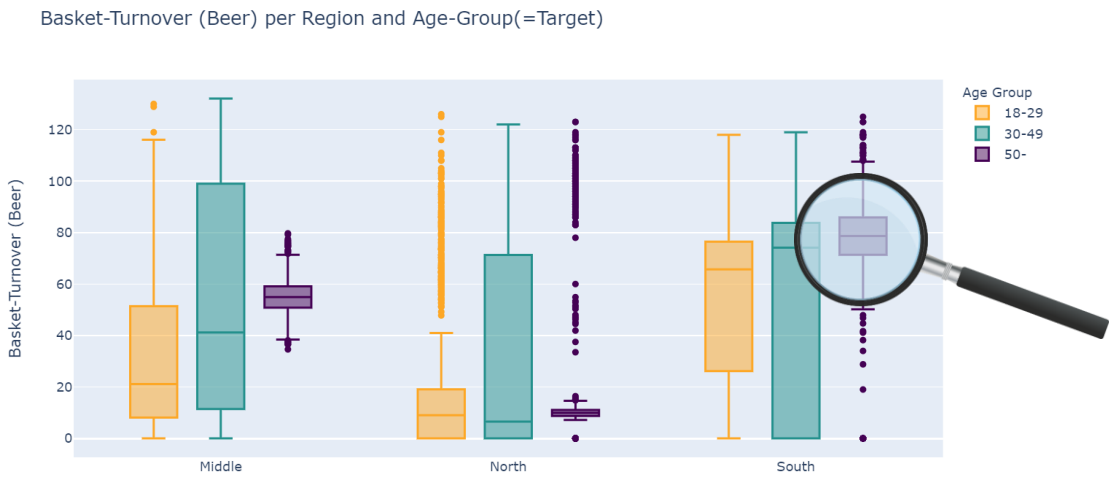
One shopping cart

Features

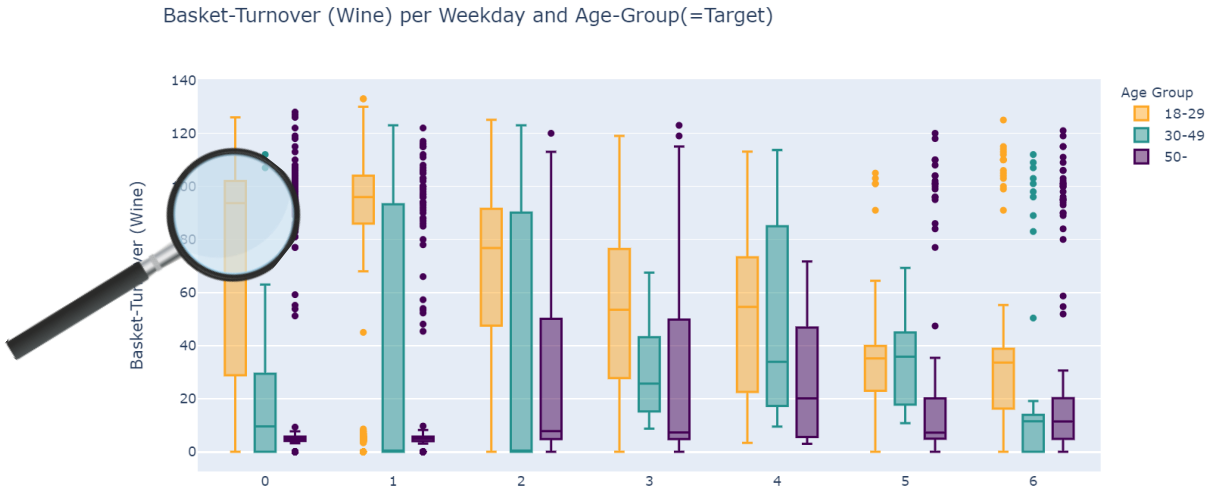
Target

# Use Case: Data Understanding

- Patterns in the data, e.g.




Older customers tend to buy more beer in the south compared to the other regions.

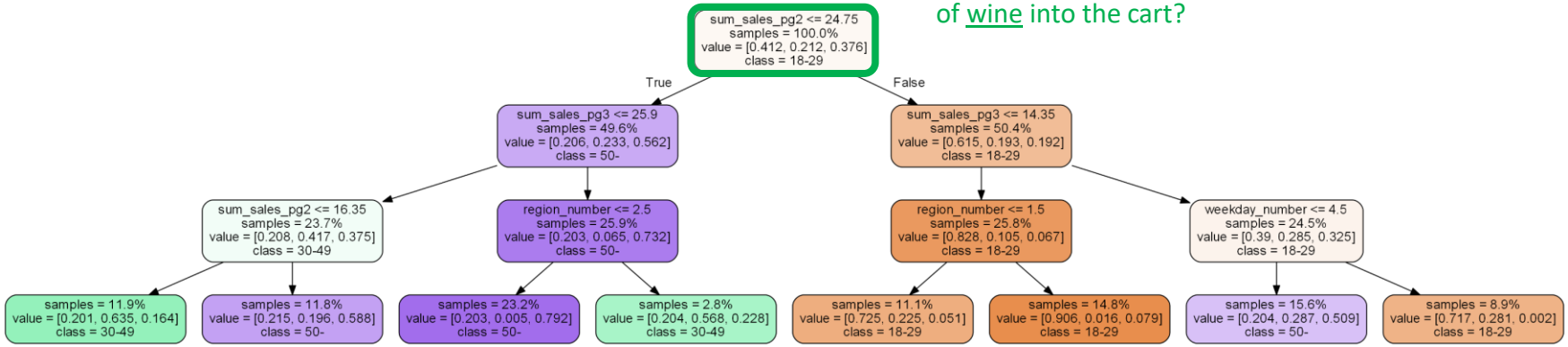



Younger customers seem to fill up their wine stocks at the beginning of the week.

# Use Case: Current Solution

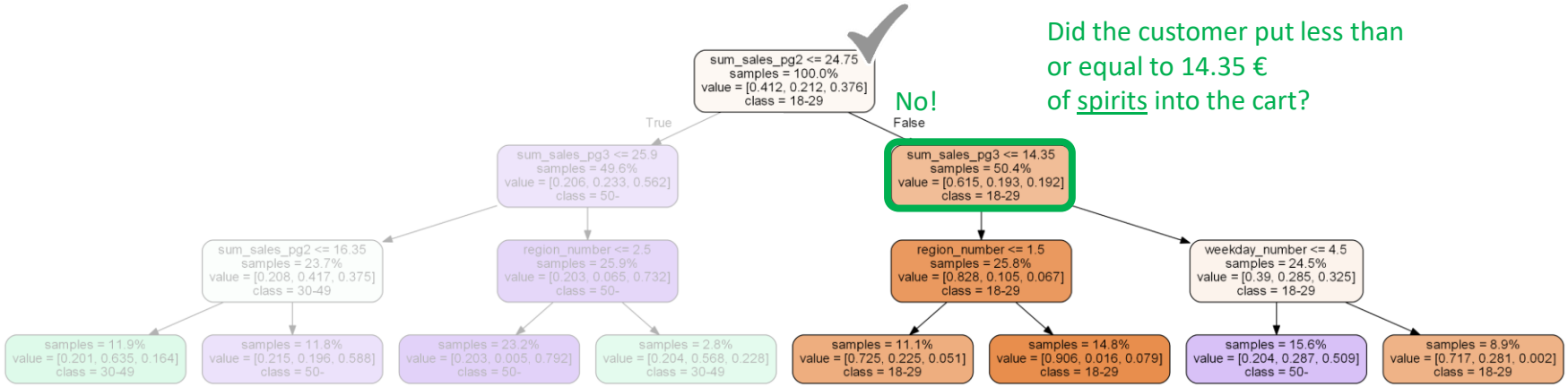
- Patterns identified by machine learning
  - Current solution: Decision Tree
    - Example: Customer from the “South” (region 3) buys 50€ wine and 8€ spirits.
    - What is the predicted age group?

Did the customer put less than
or equal to 24.75 €
of wine into the cart?



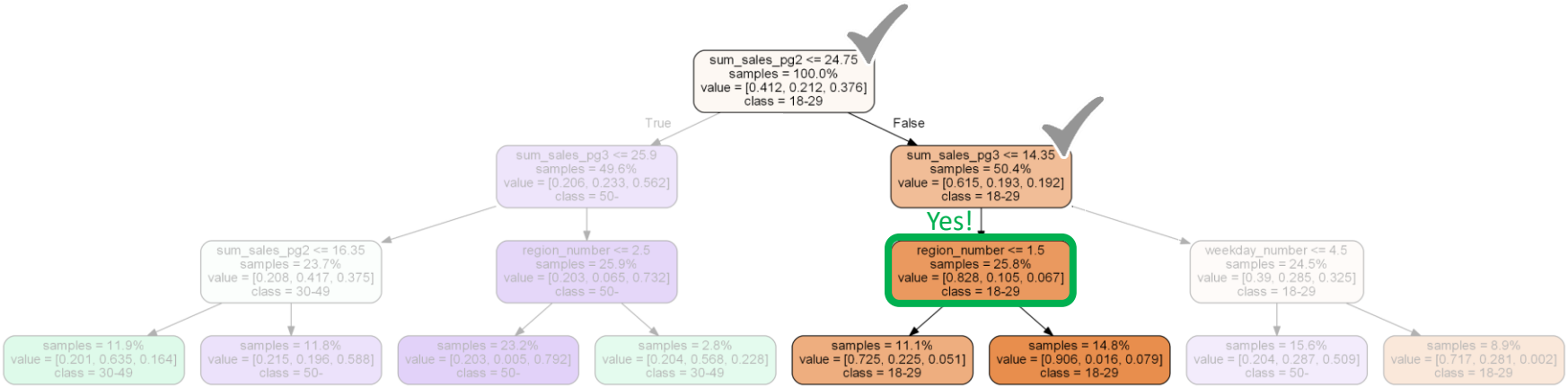
# Use Case: Current Solution

- Patterns identified by machine learning
  - Current solution: Decision Tree
    - Example: Customer from the “South” (region 3) buys 50€ wine and 8€ spirits.
    - What is the predicted age group?



# Use Case: Current Solution

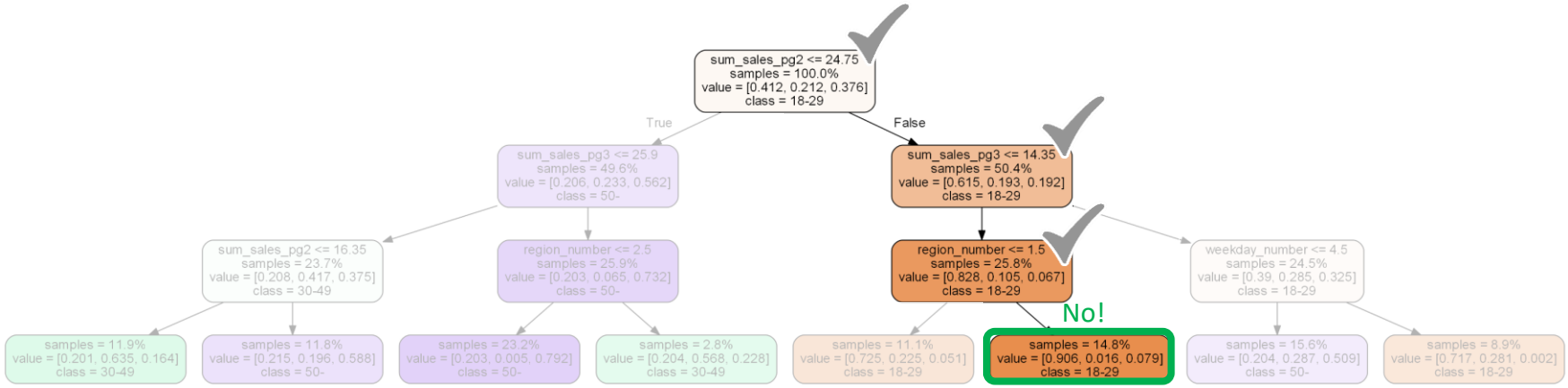
- Patterns identified by machine learning
  - Current solution: Decision Tree
    - Example: Customer from the “South” (region 3) buys 50€ wine and 8€ spirits.
    - What is the predicted age group?




Does the customer stem from the “North” (region 1)?

# Use Case: Current Solution

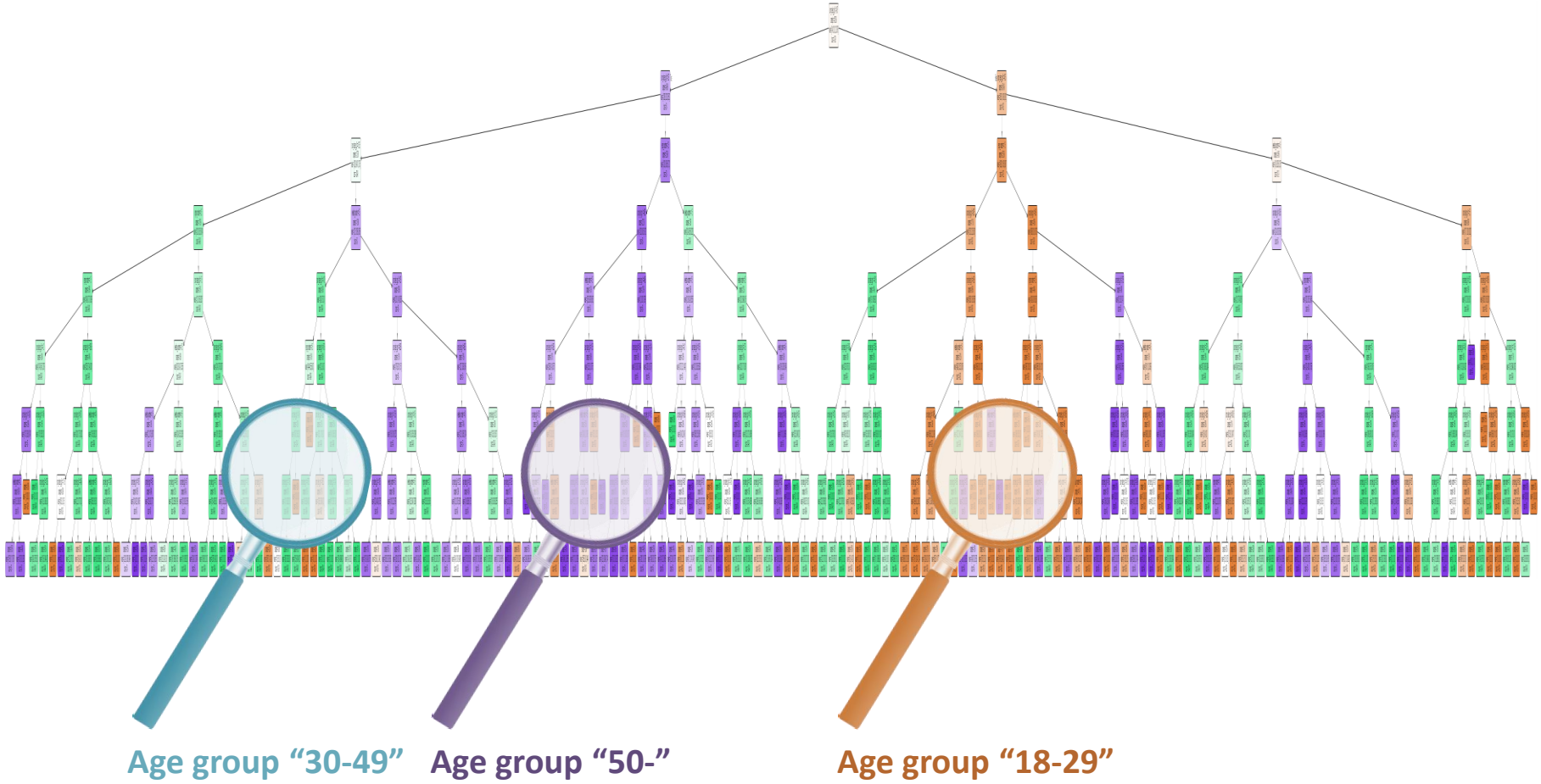
- Patterns identified by machine learning
  - Current solution: Decision Tree
    - Example: Customer from the “South” (region 3) buys 50€ wine and 8€ spirits.
    - **What is the predicted age group?** **18-29**




Predict age group “18-29”

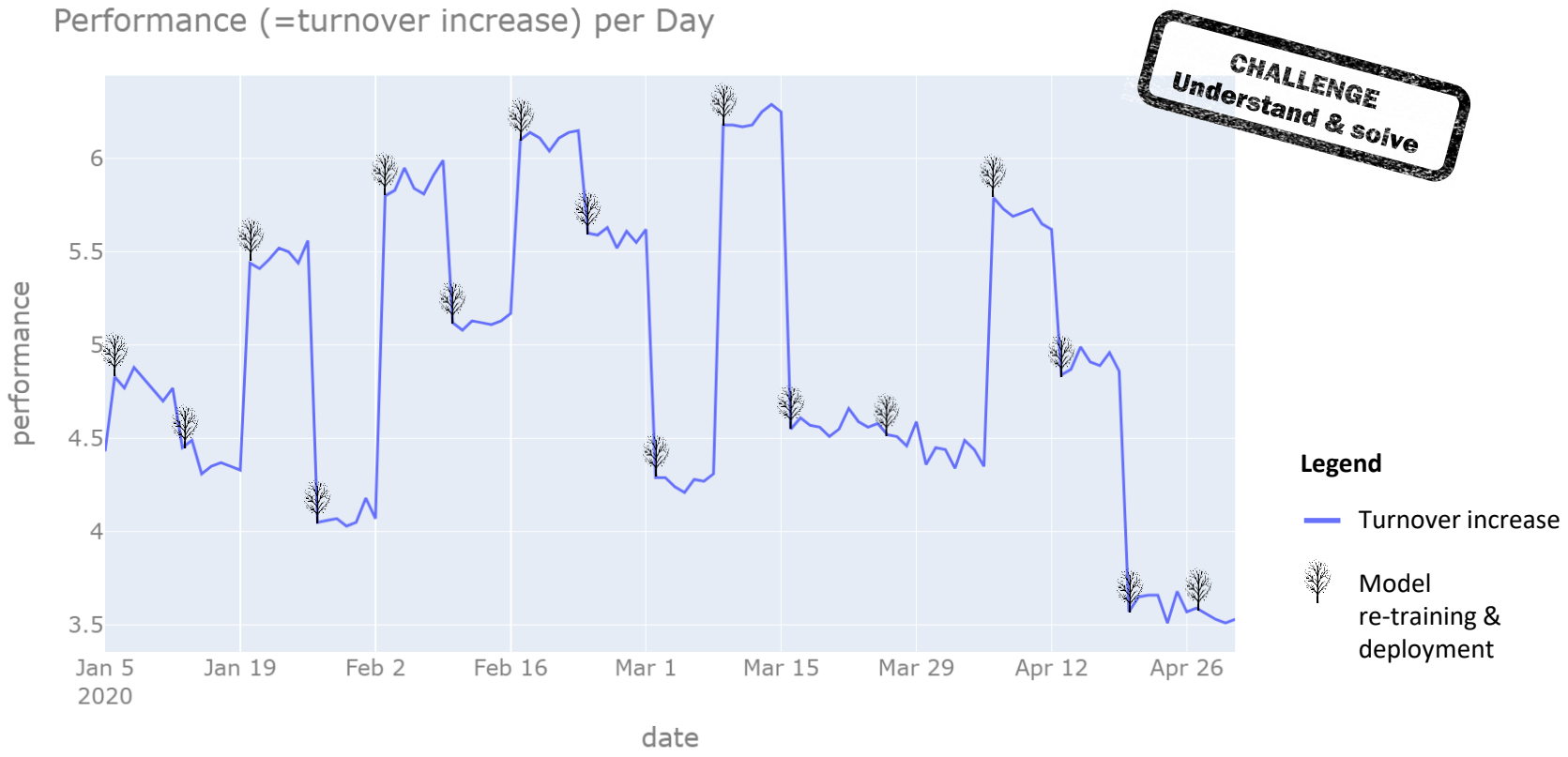
# Use Case: Initial Solution

- Initial solution: Decision Tree
  - It is capable of identifying the patterns in the data.



# Use Case: The Challenge

- Our model achieves an accuracy of ~80%
  - This should lead to ~5.00 % increase in turnover.
- Result:



# Goals

- Present the central challenge in machine learning:
  the **“Bias-Variance Tradeoff”**
- See its effects in an exemplary use case.
- Learn about the idea to design more **effective** machine learning solutions with ensemble methods.

# Agenda

- [x] **Use Case & Goals**
- [ ] **Bias-Variance Tradeoff**
- [ ] **Reducing Variance with Ensemble Methods**
- [ ] **Summary & Outlook**

# Over- and Underfitting

- **Fundamental challenge in machine learning**

**How can a model**

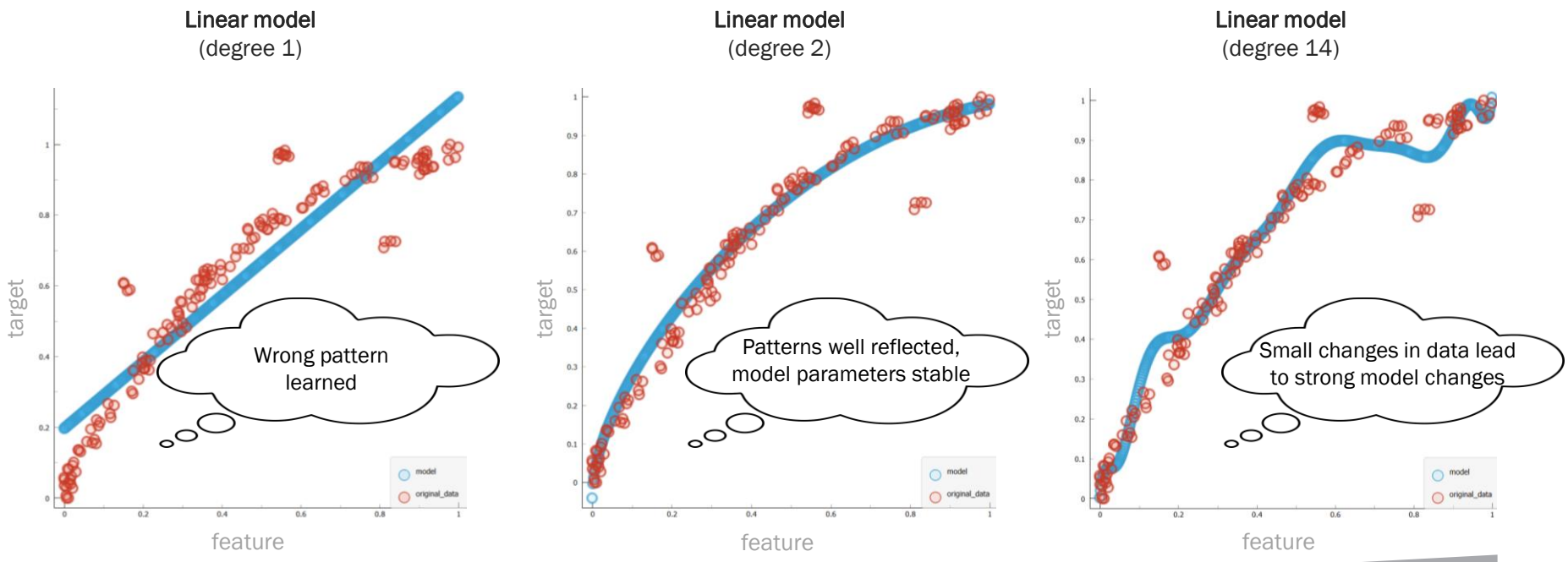
(a) catch the **right patterns** in historical data

(b) work well on new data (i.e. **generalize**)?


Cf. Fortmann-Roe (2012)

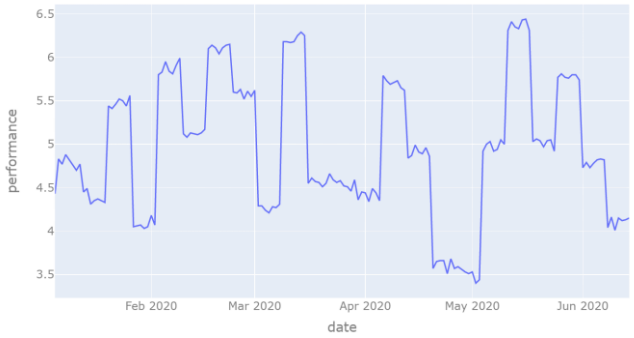
# Bias-Variance Tradeoff

- **Fundamental challenge in machine learning**

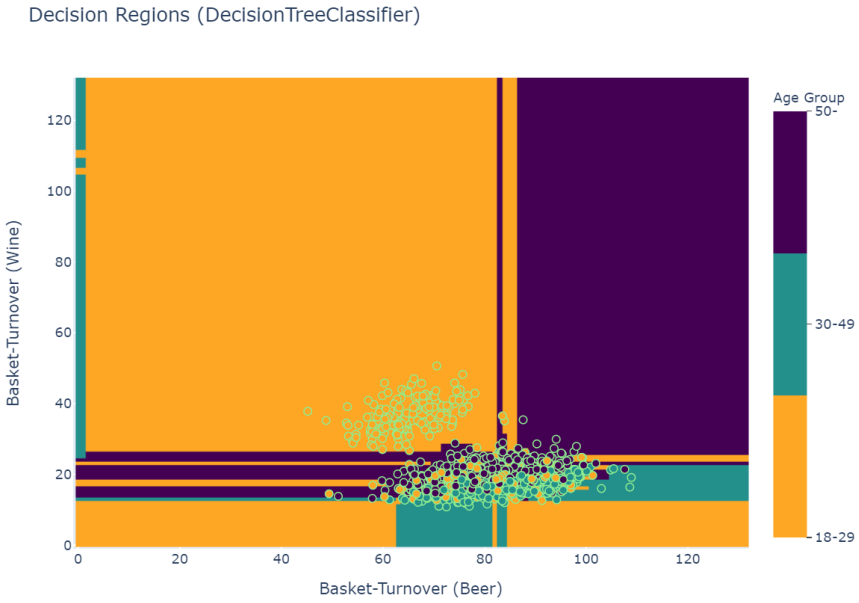


# Use Case: Variance in Current Solution

- Probably, the model has a high variance
  - How does variance look like in classification?
  - Let's take a look at the decision boundaries.



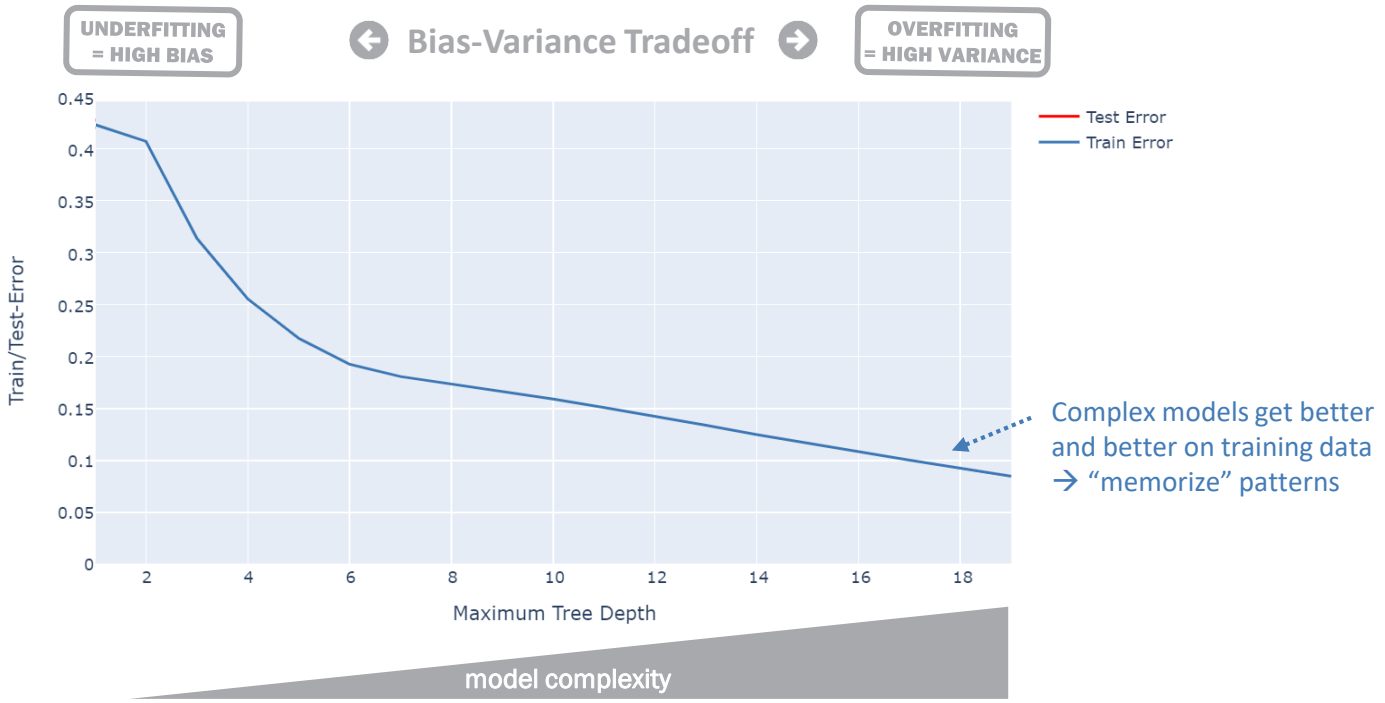

**Fix values of 3 features, then a scatterplot is possible, e.g.:**




**Train model multiple times and watch decision boundary**

# Bias-Variance Tradeoff

- **Approaching the Bias-Variance Tradeoff numerically**
  - Vary model parameters & measure train-test-error

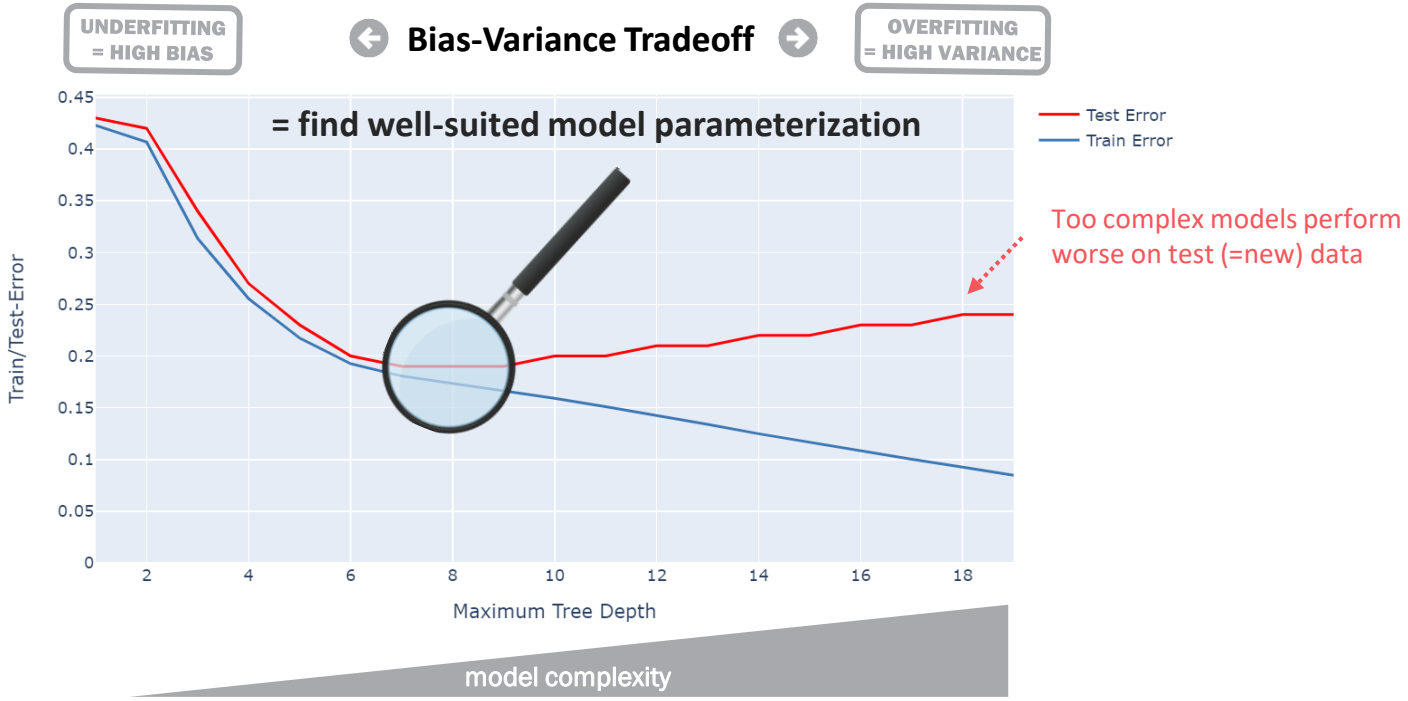




Source: adapted from Fortmann-Roe (2012)

# Bias-Variance Tradeoff

- **Approaching the Bias-Variance Tradeoff numerically**
  - Vary model parameters & measure train-test-error





Source: adapted from Fortmann-Roe (2012)

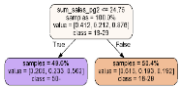
# Over- and Underfitting

- There is a **fundamental challenge** in machine learning

**How should a model be parameterized to work well in its use case?**

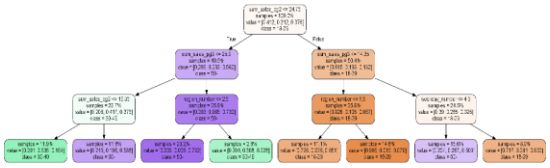

The model is (potentially) …

… too simple



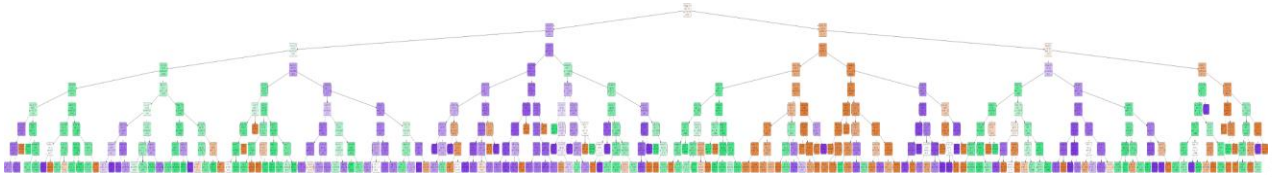
*maximum depth 1*

… reasonable



*maximum depth 3*

… too complex



*maximum depth 8*

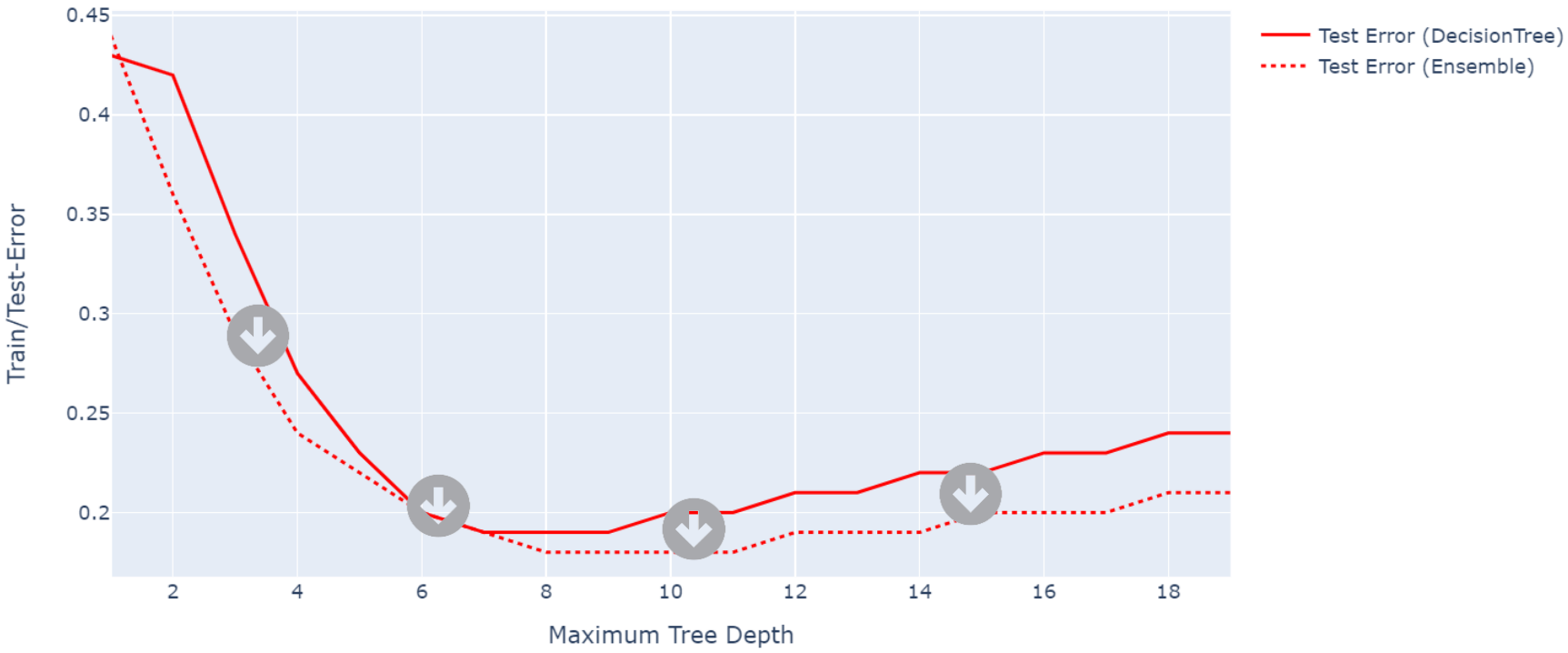
Model complexity

# Agenda

- ☐ Use Case & Goals
- ☐ Bias-Variance Tradeoff
- ■ Reducing Variance with Ensemble Methods
- ☐ Summary & Outlook

# Reduce Variance with Ensemble Methods

- **Idea: combine predictions of <u>multiple</u> models**
  - Goal: improve model performance

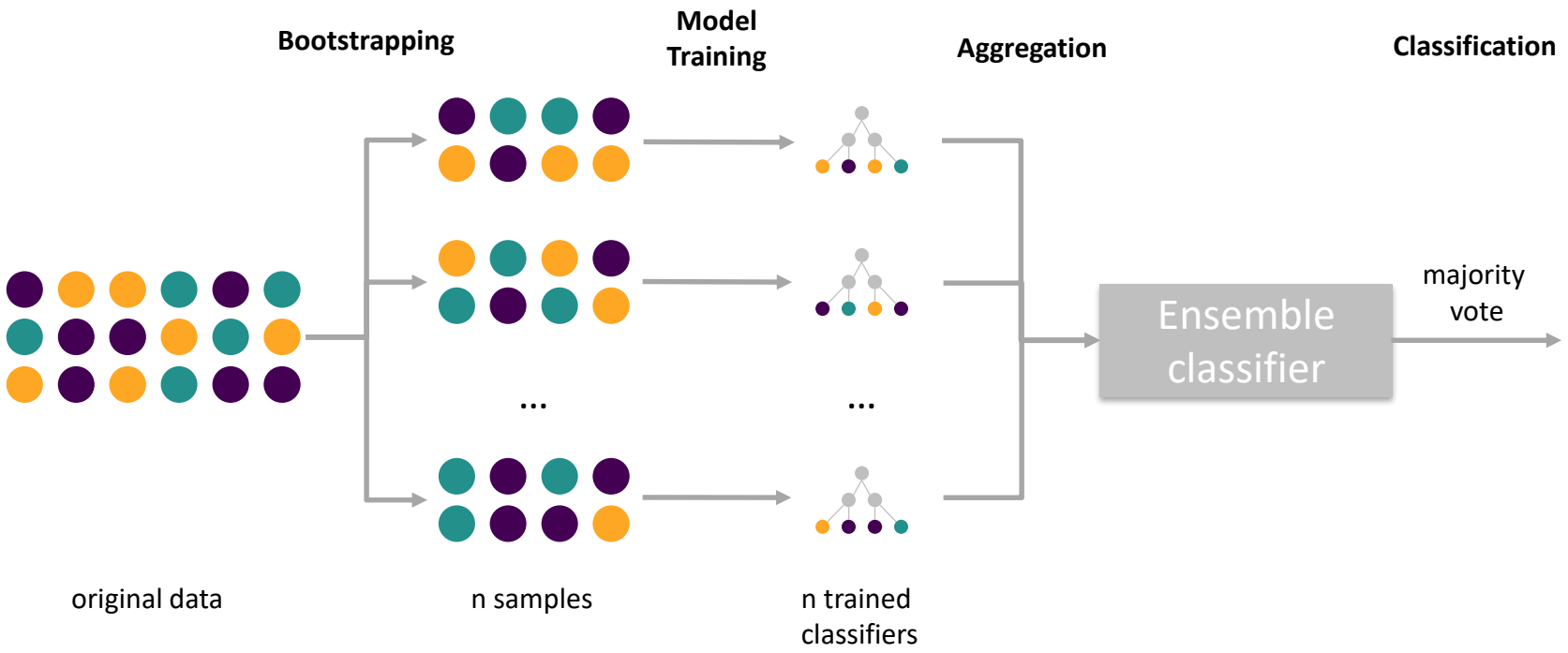




Cf. Zhou (2012) "Ensemble Methods. Foundations and Algorithms"

# Reduce Variance with Ensemble Methods

- **Example: Bagging**

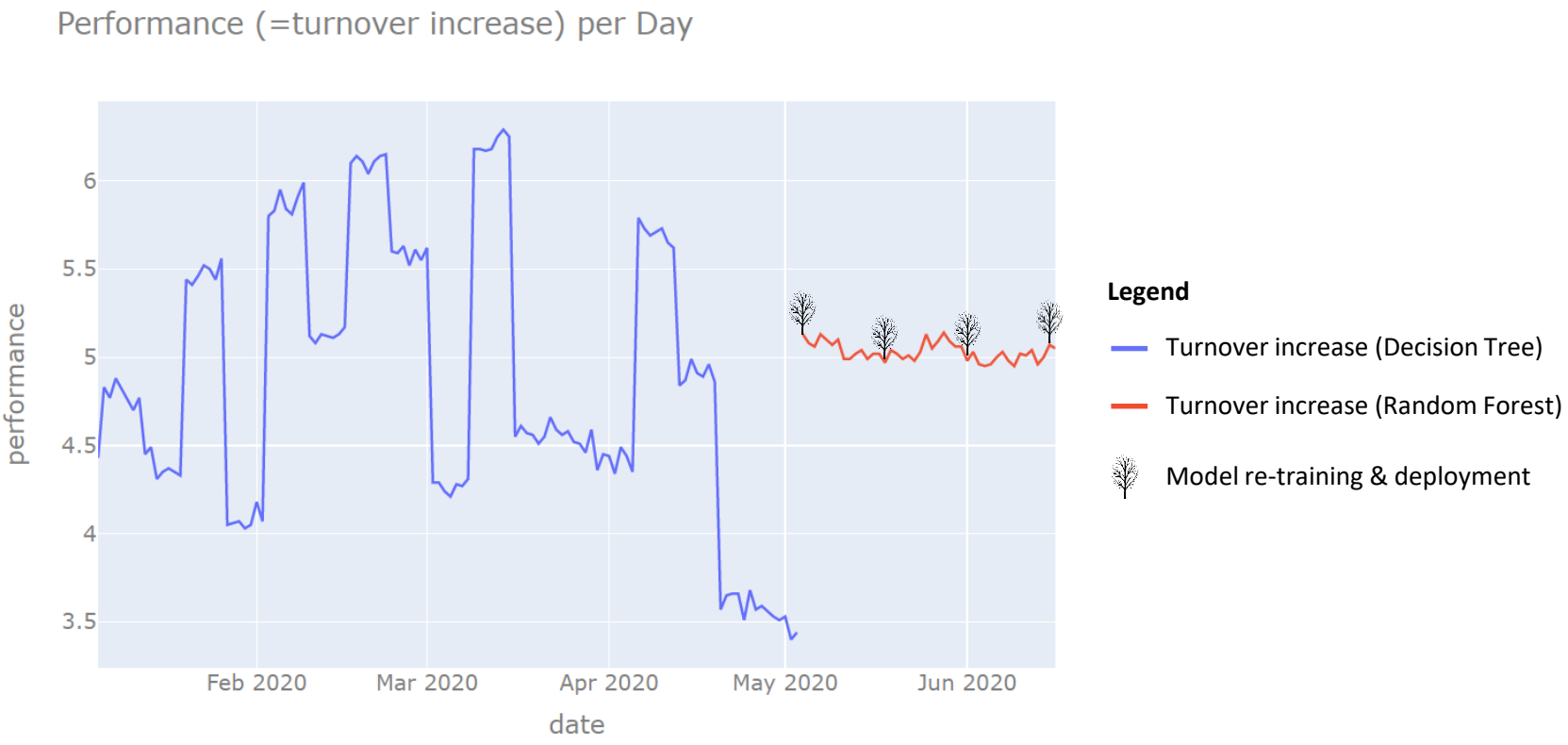



**How do such models perform regarding the bias-variance tradeoff?**


Cf. Zhou (2012) "Ensemble Methods. Foundations and Algorithms"

# Use Case: The Improvement

- Our model now achieves an accuracy of ~82%
  - This leads to ~5.15 % increase in turnover.
  - Our tests show improved performance and less variance → **more effective model.**

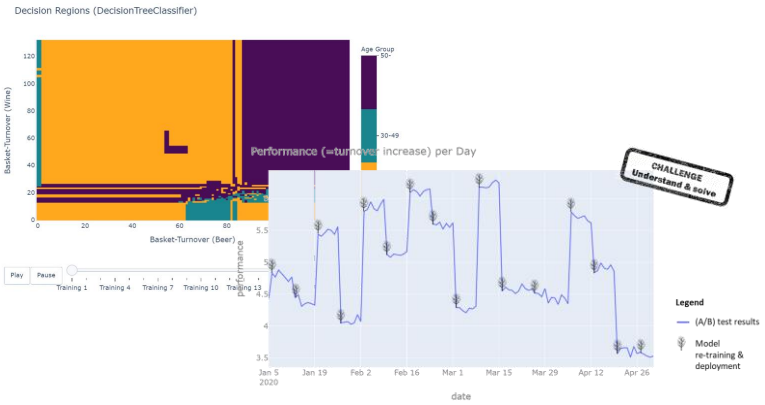


# Agenda

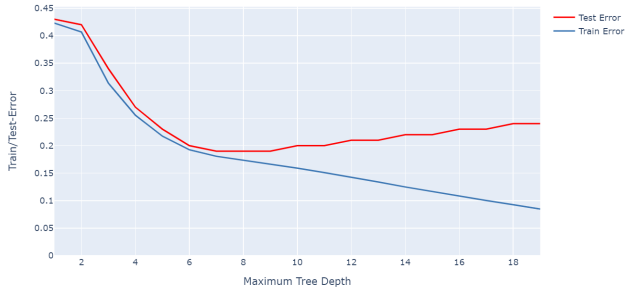
- [ ] Use Case & Goals
- [ ] Bias-Variance Tradeoff
- [ ] Reducing Variance with Ensemble Methods
- [x] Summary & Outlook

# Summary & Outlook

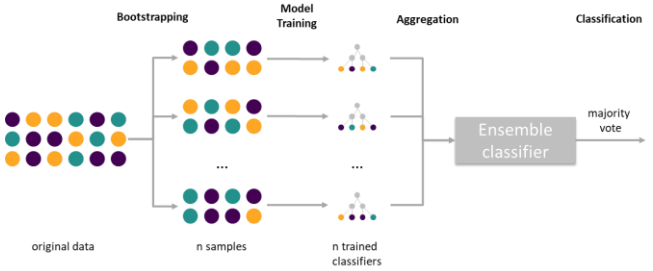
- Effects of overfitting in a predictive application




- Methods for evaluating bias vs. variance

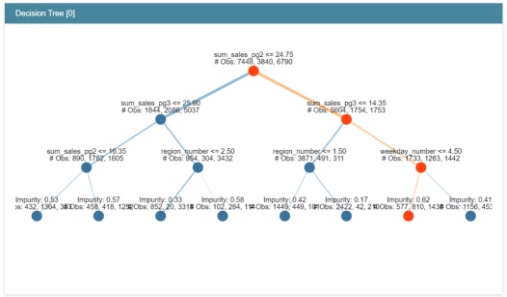



- Ensemble methods to improve model performance

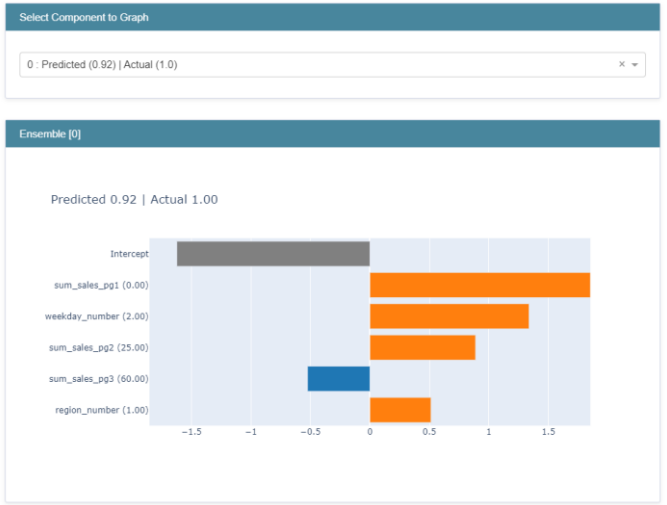


# Summary & Outlook

- Approach the next challenge in effective machine learning:
  - **Explainability** *"… can be a mandatory legal requirement […and] is highly desirable […] to build trust in the machine learning model."*
    Wick et al. (2020)




- Example
  - "Explainable Gradient Boosting" with "Microsoft interpret"
    Cf. Microsoft (2020)



# References


Fortmann-Roe, Scott (2012). “Understanding the Bias-Variance Tradeoff.” Accessed June 04, 2020. http://scott.fortmann-roe.com/docs/BiasVariance.html.

Microsoft (2020). “Fit Interpretable Models. Explain Blackbox Machine Learning.” Accessed June 08, 2020. https://github.com/interpretml/interpret.

Wick, Felix, Ulrich Kerzel, and Michael Feindt (2020). *Cyclic Boosting - an Explainable Supervised Machine Learning Algorithm*.

Zhou, Zhi-Hua (2012). *Ensemble Methods: Foundations and Algorithms.* Chapman & Hall / CRC Machine Learning & Pattern Recognition. Hoboken: CRC Press. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=465619.

# Thank you!