



## Predicting Tax Treaty Formation Using Machine Learning: Implications for Parliamentary Practice

- Dmitry Erokhin, PhD
- 20 June 2024

Erokhin, D., & Zagler, M. (2023). Explaining and Predicting Double Tax Treaty Formation with Machine Learning Algorithms. *WU International Taxation Research Paper Series*, (2023-03).

# Content

Introduction

Data and methodology

Results

Conclusion

# Tax treaty formation

- Main goals of tax treaties:
  - Boost **trade and investment** by eliminating tax barriers;
  - Fight tax evasion and double non-taxation with **anti-avoidance** measures;
  - **Exchange of information**, a primary focus of new tax treaties.
- **Complex decision-making process** precedes tax treaty formation:
  - Historically driven by “**chess-games between superpowers**”, **key persons' decisions, and corporate lobbyism**;
  - **Policy diffusion** and other factors may influence tax treaty formation.
- Double tax treaty probability can be predicted using machine learning approach.
- Highly relevant for policymakers:
  - Identifies advantageous treaties to pursue given **limited capacity of negotiators**;
  - Helps **assess risk of FDI diversion** to neighboring jurisdictions;
  - Aids in understanding level playing field for multinational firms in foreign markets.

# Literature – tax treaty formation

## **Economic factors:**

- Personal tax rates, non-resident withholding tax rates on dividends and interest
- FDI stock, symmetric allocation of FDI
- Transfer price, auditing costs, production costs
- Revenue sharing differences

## **Governance and policy factors:**

- Tax haven bargaining power
- Good governance
- Experience in entering into tax treaties

## **International relations factors:**

- Spatial spillovers and dependence in tax treaty formation

- Export-product competitors
- Power asymmetries between signatories
- Developed countries compensating developing countries' tax base losses through foreign aid

## **Other factors:**

- Common language
- Information sharing, tax audit, and revenue sharing agreements

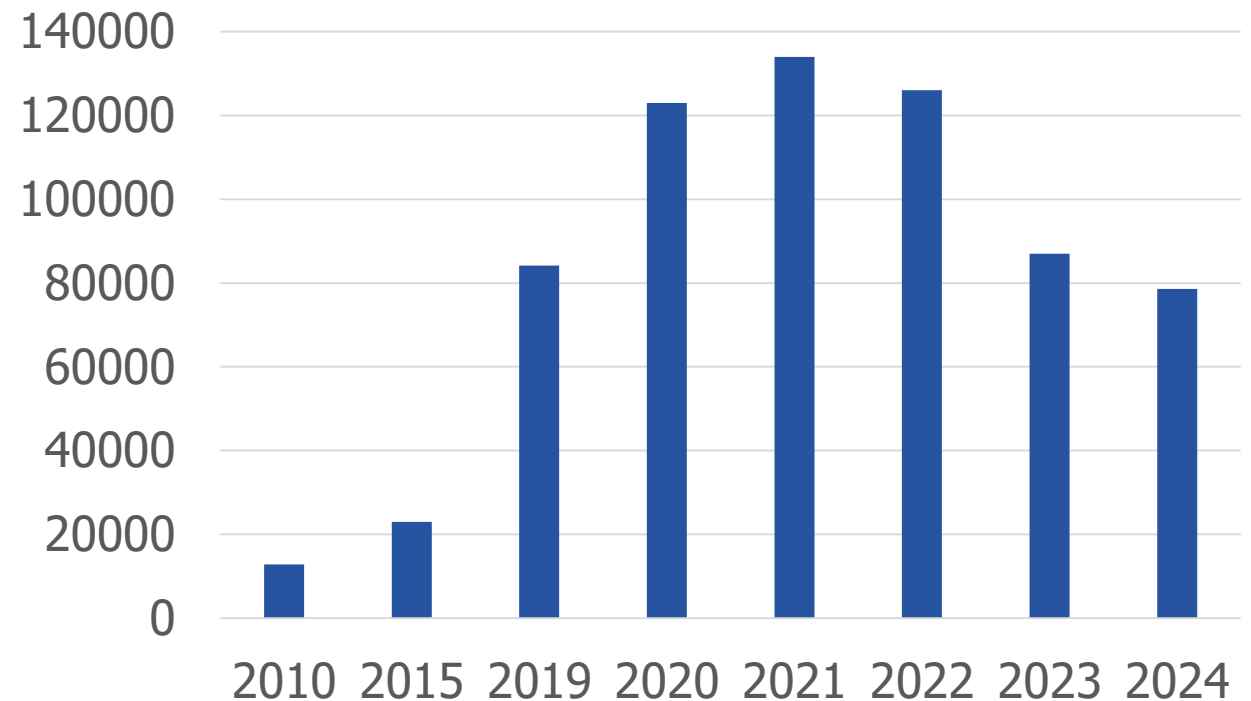
## **Withholding tax rates:**

- Positive relationship with withholding tax rates negotiated in past tax treaties
- Spatial dependence of dividends withholding tax rates
- Increase in withholding tax rates with asymmetric FDI activities

# Machine learning in economics

- Increased application in recent years.
- Various areas of economics such as energy economics, growth economics, crypto economics, urban economics, and taxation.
- Modeling **complex and flexible relationships**.
- Uncovering **generalizable patterns** and finding functions with high **out-of-sample predictive power**, which is important for policy makers.
- Limited prior theoretical assumptions or major assumptions on variable distribution.

"Machine learning" economics searches on Google scholar

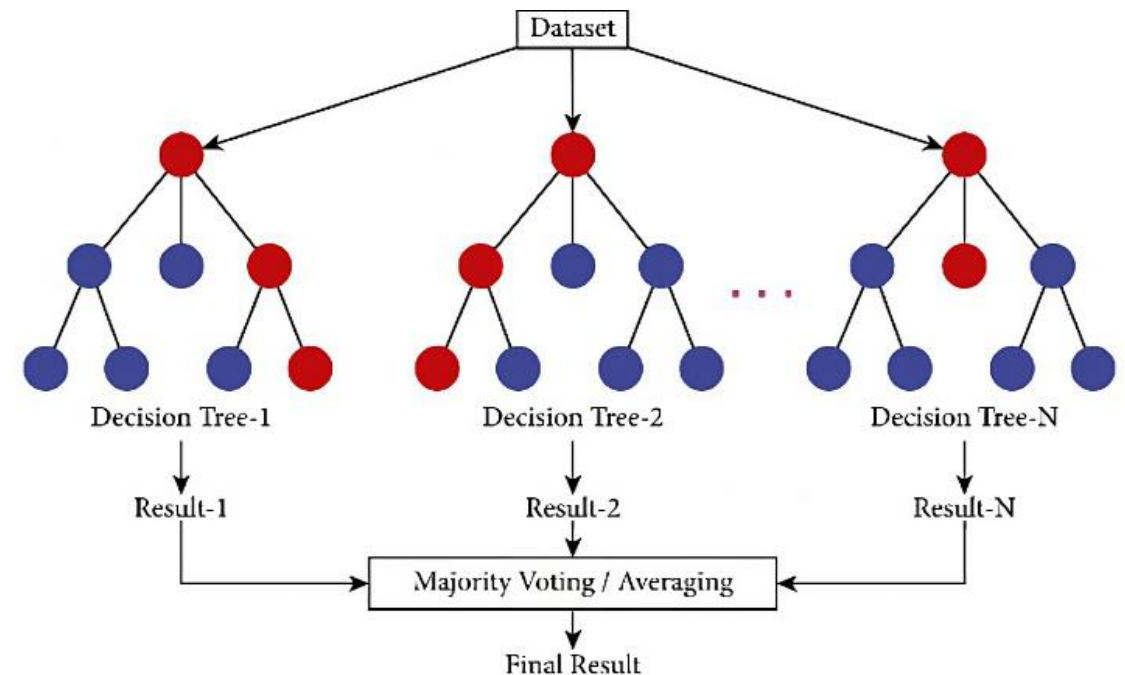


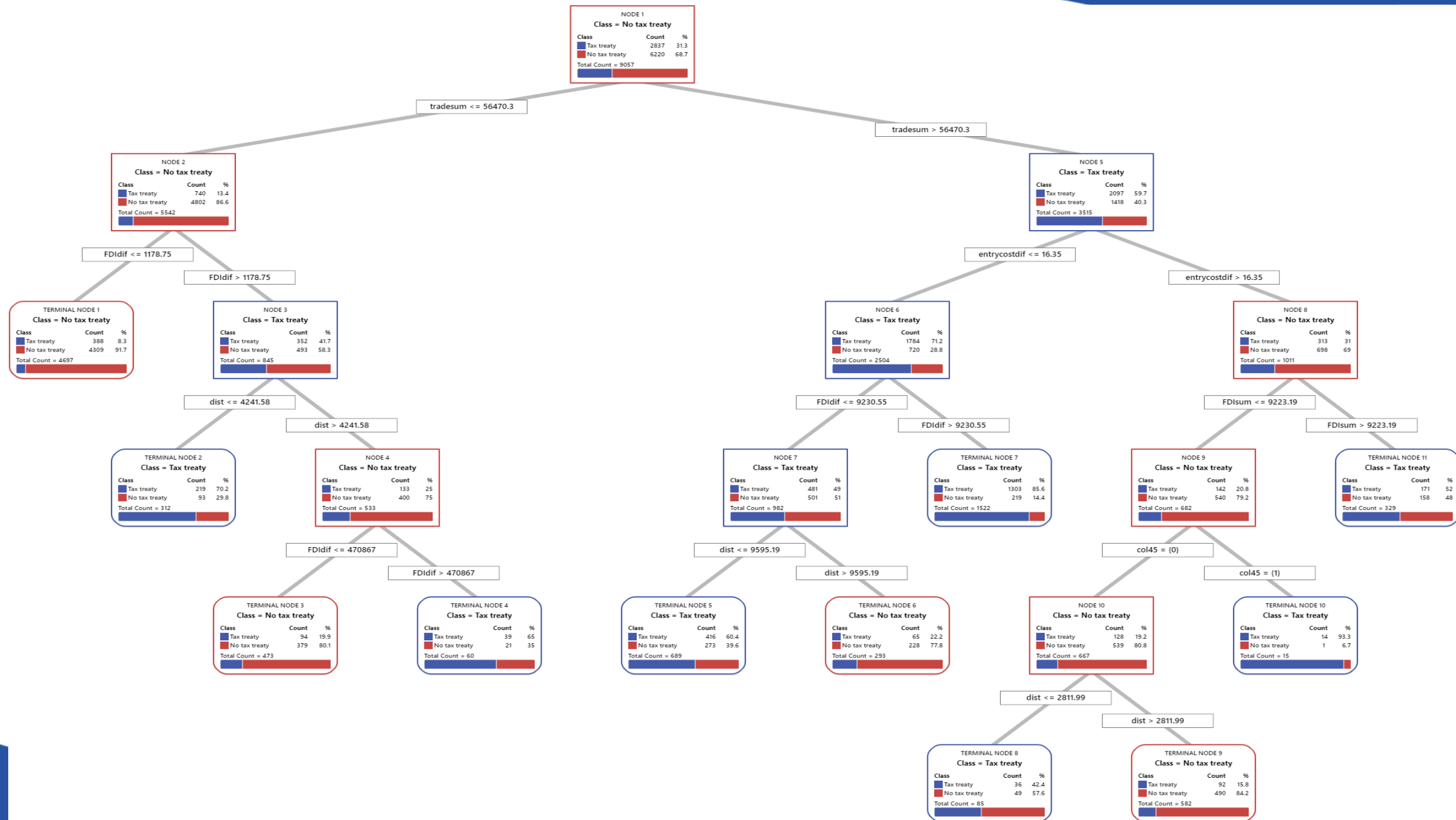
# Methodology and data

- **Binary classification problem:** country pairs as “having tax treaties” or “not having tax treaties”.
- `c_ml_stata_cv` command, Stata/Python integration, scikit-learn API.
- Supervised machine learning algorithms: classification tree, random forest, boosting, regularized multinomial, nearest neighbor, neural network, naive Bayes, support vector machine, and standard multinomial algorithms.
- Classification error rate on the test data, sensitivity, precision, specificity, F1-score, and area-under-the-curve.
- Two periods: 2018 for training the machine and 2019 for testing it.
- 2800 country pairs with tax treaties and 6200 without tax treaties.
- 29 gravity features constructed as bilateral variables.

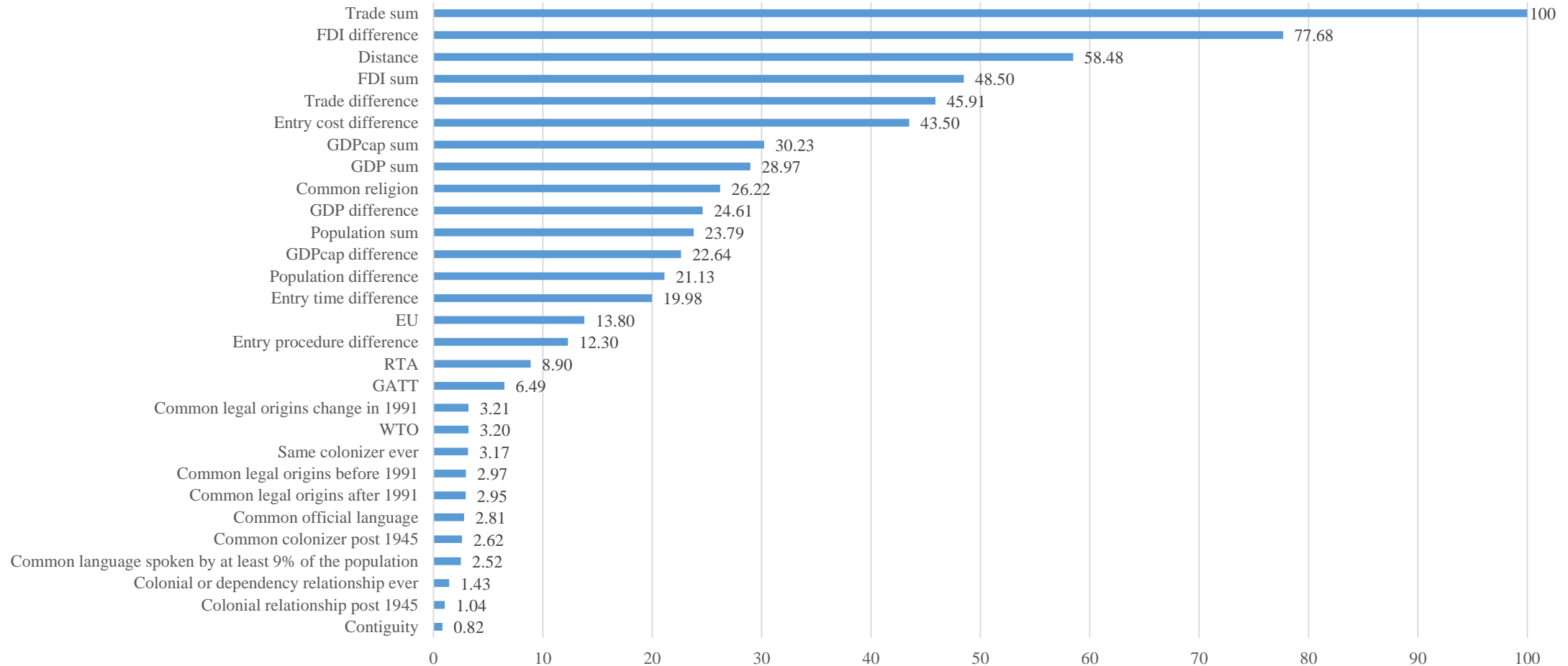
# Algorithm selection

- The selection of the most accurate algorithm is based on the testing classification error rate, and the **random forest** algorithm has the lowest rate (0.057).
- The random forest algorithm outperforms other algorithms according to all the evaluation metrics.
- Both the random forest algorithm and the classification tree algorithm exhibit a zero training classification error rate and the lowest two testing classification error rates.
- A specific classification tree with 11 nodes drawn using the CART® Classification of the Minitab statistics package.

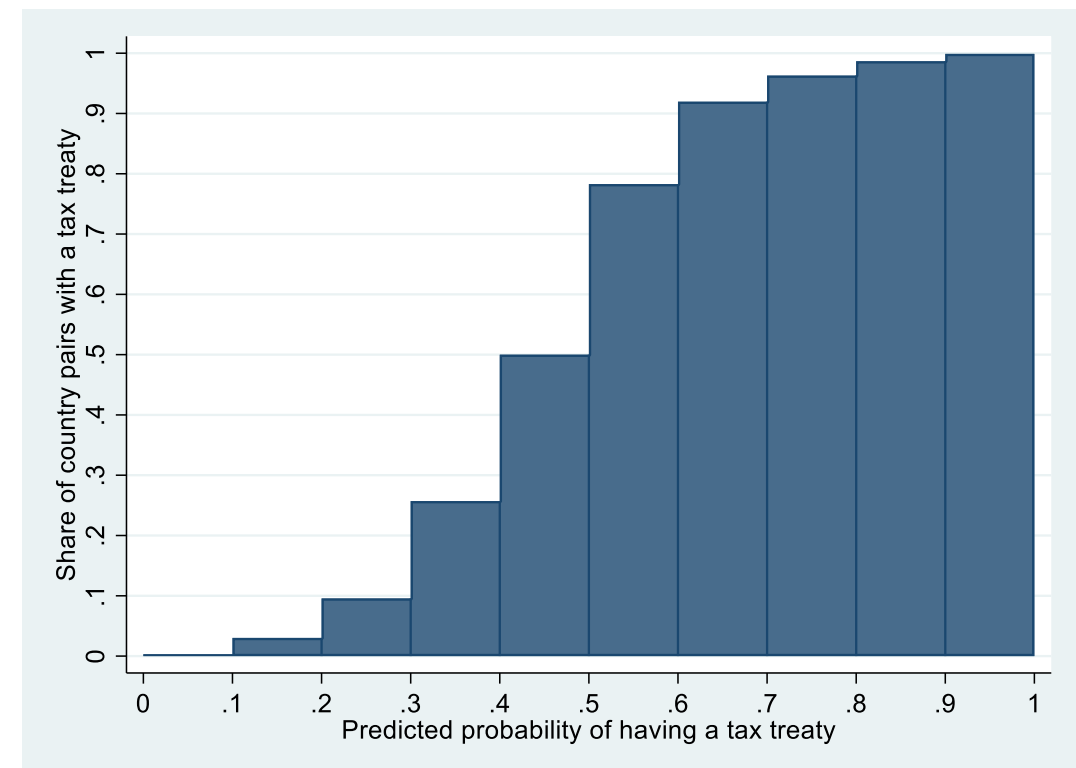
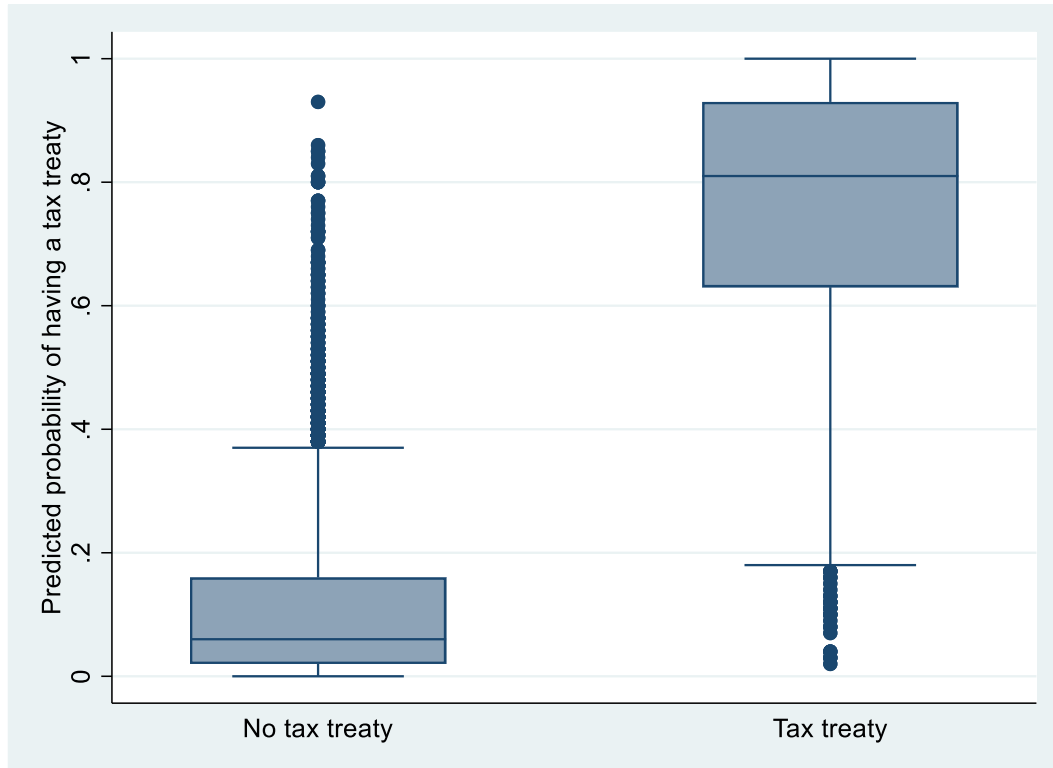




# Relative variable importance (random forest)



# Prediction accuracy



# Tax treaty formation prediction

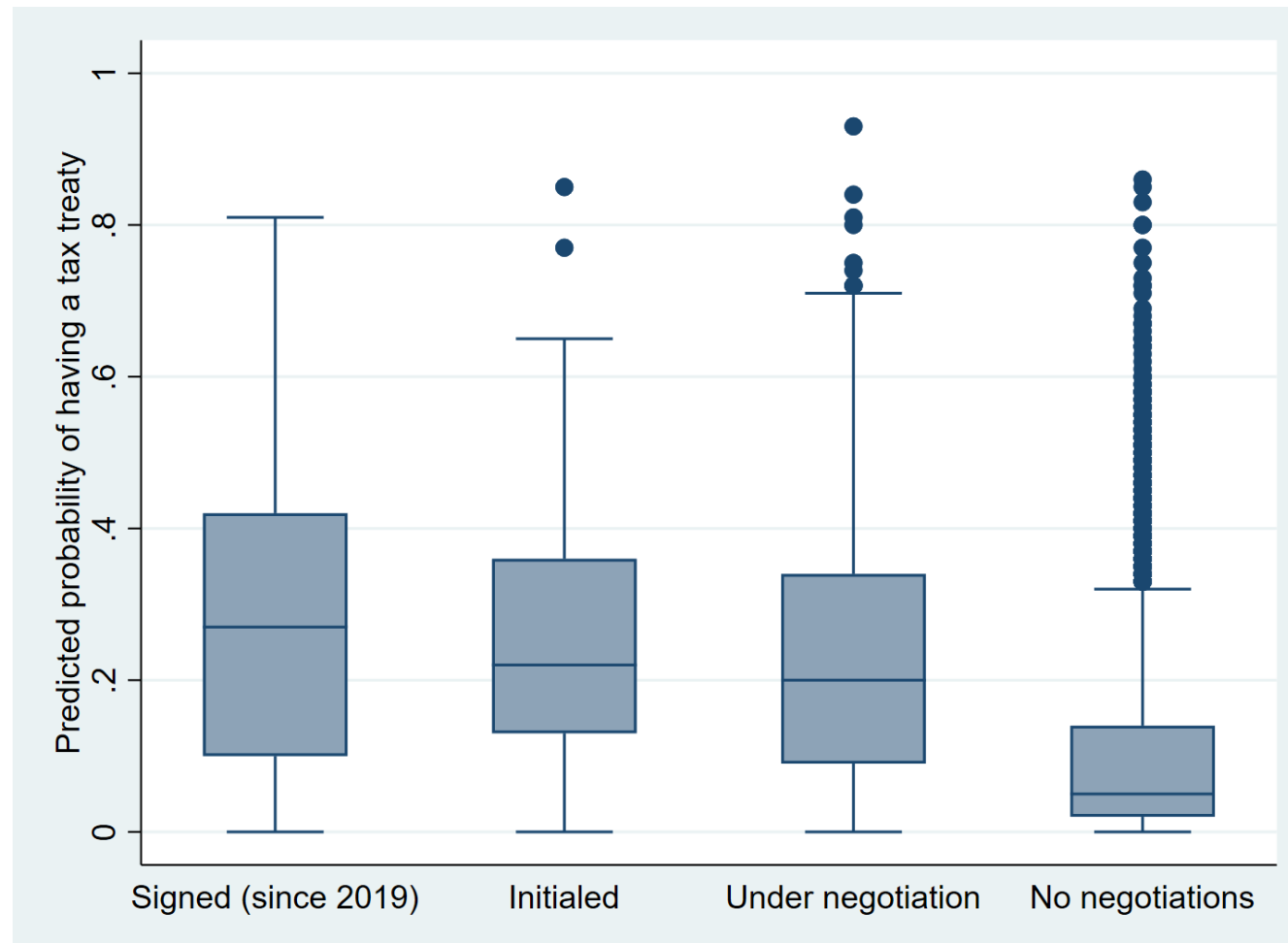
- **59 country pairs** likely to have tax treaties based on 2019 gravity features.
- Germany (9), Saudi Arabia (8), Brazil (7), Myanmar (7), and Hong Kong (6).
- 24 country pairs in negotiation process, 4 signed a tax treaty, 3 initialed a tax treaty, 6 signed or negotiating exchange of information agreement or transport tax treaty, and 3 terminated tax treaties.

Country/region A	Country/region B	Predicted probability of a tax treaty in 2019	Current status of a tax treaty in 2023 (IBFD, 2023; Orbitax, 2023)
Croatia	United States	0.62	Signed (December 2022)
Denmark	France	0.81	Signed (February 2022)
Brazil	United Kingdom	0.76	Signed (November 2022)
Brazil	Poland	0.67	Signed (September 2022)
Austria	Bangladesh	0.55	

Austria and Bangladesh Seek to Conclude Tax Treaty

May 1, 2024

# Predicted probability vs actual status after 2019



# Implications for parliaments

- Clear **guideline** on how machine learning algorithms can inform policymakers on treaty policy to follow.
- Policymakers in countries with missed opportunities for negotiation should take advantage of **potential treaties** to improve their treaty policy.
- **Neighboring countries** should check their existing treaty network and tax policy with respect to potential partners of their neighbors, given predicted treaties between other countries.
- Other potential applications:
  - Predictive analysis of policy outcomes
  - Economic forecasting
  - Analyzing public opinion
  - Resource allocation
  - Fraud detection
  - ...

Thank you for your time.

Questions.