Estimating the Condition of Streams & Rivers: An Approach Using Supervised Machine Learning Methodologies

Daniel S. Adams

INTRODUCTION

In this study, a habitat system condition index is developed and modeled to be **representative of the relative departure of a current wildlife habitat condition from a desired condition** to identify where there are conservation opportunities available across the landscape.

Whereas similar efforts deploy expert systems or multi-criterion decision making modeling approaches, **this study explores the usage of supervised machine learning to classify rivers and streams** into habitat condition categories.

METHODS

Training Data:
Training data is sourced from digitized aquatic biologist field work, using social science derived shared language to describe the four habitat classes.

Indicator Data:
The condition index was modeled using data that indicates stream/river **functional network length, riparian buffers, network complexity, sinuosity, dam density,** and **road crossing density**.

Model Testing, Comparisons, and Selection:
Naïve Bayes Classifier (Baseline for comparison)
Support Vector Machine Classifier (SVM)
Decision Tree Classifier
Random Forest Classifier

Accuracy reports can be seen in **table 1a and 1b**.

DISCUSSION

Interpreting Results:
All three proposed supervised machine learning methods outperform the Naïve Bayes Classifier.

Both Random Forest and Decision Tree Classifiers classify streams and rivers into habitat conditions to a high degree of accuracy.

The Random Forest Classifier marginally outperforms the Decision Tree Classifier.

The Decision Tree Classifier is selected for final streams & rivers classification due to the ease of interpreting results and comparable accuracy metrics.

CONCLUSIONS & USE CASES

Conclusions:
Results suggest that supervised learning approaches to classifying streams and rivers show great promise.

What makes this study novel is the ability to classify streams and rivers by the usage of aquatic biologist field experiences. Related works focus on classifying streams and rivers using relevant indicator data, whereas this study has shown streams and rivers can be classified by extrapolating subject domain field experience to areas that they have not worked within.

Use Cases:
Currently in use by the states of Louisiana and Mississippi for aquatic conservation planning. Also in use by the Southeast Aquatic Resources Partnership for barrier removal project identification.

RESULTS

Table 1a. Model Comparisons

	Precision	Recall	F-1	Accuracy
Naïve Bayes	0.260	0.380	0.298	0.382
Decision Tree	0.888	0.888	0.890	0.888
Random Forest	0.903	0.905	0.905	0.907
SVM	0.558	0.553	0.548	0.553

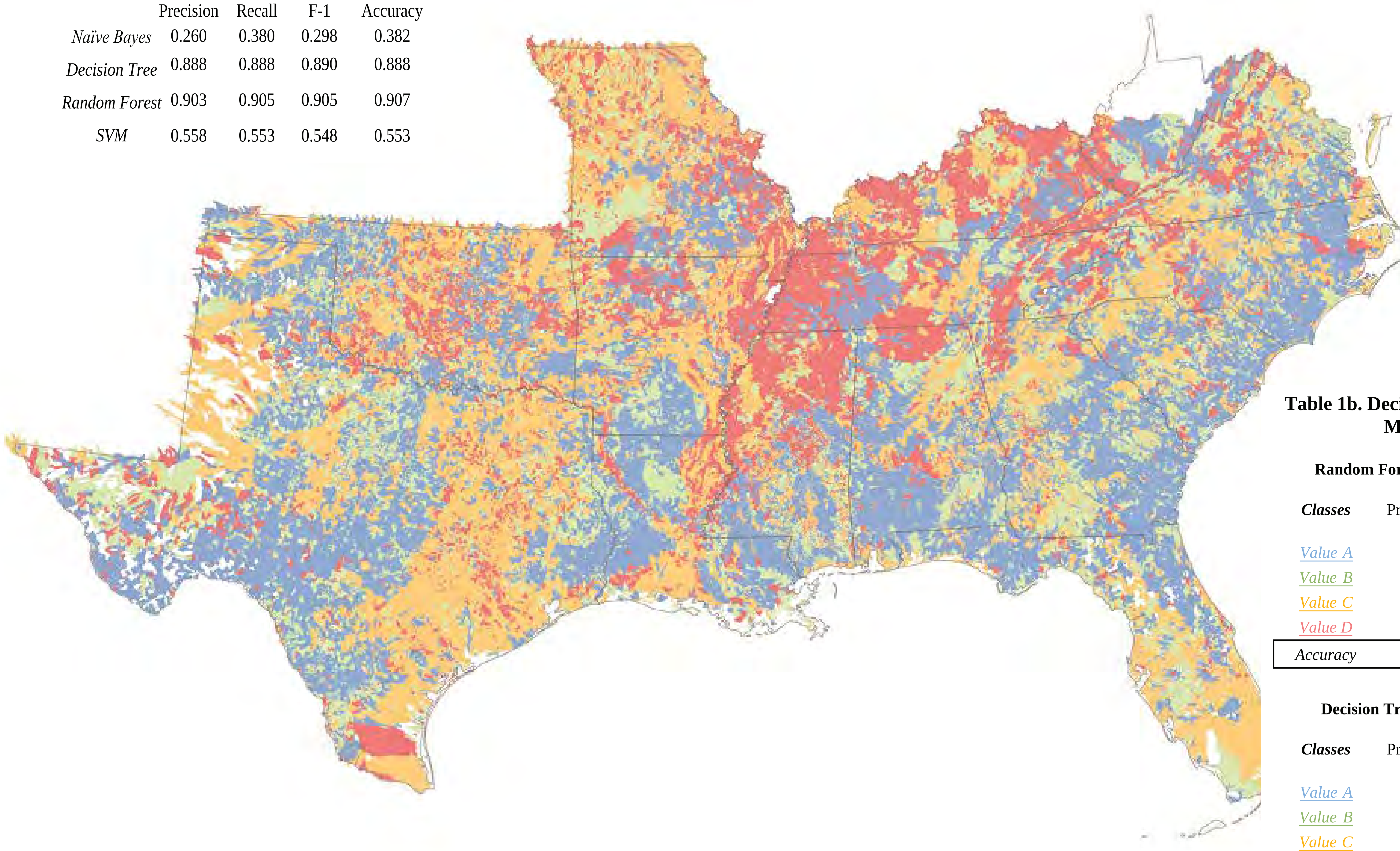

Table 1b. Decision Tree and Random Forest Model Comparisons

Random Forest Classifier Accuracy Report

Classes	Precision	Recall	F-1	Support
Value A	0.90	0.91	0.91	527
Value B	0.89	0.90	0.89	505
Value C	0.89	0.89	0.89	548
Value D	0.93	0.92	0.93	538
Accuracy				**0.907**

Decision Tree Classifier Accuracy Report

Classes	Precision	Recall	F-1	Support
Value A	0.90	0.88	0.89	527
Value B	0.86	0.89	0.88	505
Value C	0.88	0.87	0.88	548
Value D	0.91	0.91	0.91	538
Accuracy				**0.888**

REFERENCES

[1] "Achieving our conservation vision using strategic habitat conservation," Landscape Partnership. [Online]. Available: https://www.landscapepartnership.org/cooperative/our-guiding-principles/strategic-habitat-conservation-documents/achieving-our-conservation-vision-using-strategic-habitat-conservation.
[2] Gulf Coastal Plains and Ozarks LCC (GCPO LCC). [Online]. Available: Gulf Coastal Plains and Ozarks LCC CPA (databasin.org).
[3] D. J. Thornbrugh, S. G. Leibowitz, R. A. Hill, M. H. Weber, Z. C. Johnson, A. R. Olsen, J. E. Flotemersch, J. L. Stoddard, and D. V. Peck, "Mapping Watershed Integrity for the conterminous United States," Ecological Indicators, 09-Dec-2017.
[4] Pruitt, B.A., K.J. Killgore, and W.T. Slack. 2017. Multi-Scale Watershed Assessment User Guide, Engineer Research and Development Center. In support of Duck River Watershed Plan, Final Watershed Assessment, U.S. Army Corps of Engineers, Nashville District.
[5] Southeast Aquatic Resources Partnership (SARP). [Online]. Available: https://southeastaquatics.net/.
[6] "Estimated floodplain map for the Conterminous United States," EPA, 27-Aug-2018.
[7] "National Hydrography Dataset Plus Dataset," EPA. [Data set]. Available: https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data. [Accessed: 07-Mar-2022].
[8] Dewitz, J. (2021). National Land Cover Database (NLCD) 2019 Products [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54
[9] "feature_importances_", in Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
[10] Southeast Conservation Adaptation Strategy (SECAS). [Online]. Available: https:// secassoutheast.org.
[11] PangNing Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. "Classification: Alternative Techniques" in Introduction to Data Mining. Newyork, NY, USA.: Pearson, 2019, ch. 4, pp. 214-226.
[12] Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. (1984). Classification and regression trees. Pacific Grove: Wadsworth & Brooks.
[13] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
[14] Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (PDF). Machine Learning. 20 (3): 273–297.
[15] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
[16] "G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995."

ACKNOWLEDGEMENTS

This project was supported by the Southeast Aquatic Resources Partnership, the State of Mississippi Department of Wildlife, Fisheries, & Parks, Tennessee Tech University College of Engineering and School of Environmental Studies, as well as various members of the U.S. Fish & Wildlife Service Science Applications, Ecological Services, and Fisheries & Aquatics Conservation Programs. Without the contributions of all stakeholders, this project would not be possible.