

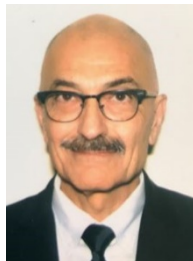
Financial Data Professional Institute

*The Global Designation for Finance Professionals in
a Data-Driven Industry*



Al Yazdani, PhD

*Chief Data Scientist & Founder, Calcolo Analytics LLC
Curriculum Committee Member, FDP Institute*



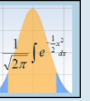
Moderator:

Mehrzad Mahdavi,
Executive Director, FDP Institute

www.fdpinstitute.org

July 8, 2020





Machine Learning Prediction of Recessions

An Imbalanced Classification Approach

Al Yazdani, PhD

Chief Data Scientist & Founder, Calcolo Analytics LLC

Curriculum Committee Member, FDP Institute



Background

- Predicting business cycles helps investors/economists/policy-makers seek alternative monetary and investment strategies.
- Research on forecasting US recessions goes back to early 1990s with emphasis on identifying financial drivers:
 - the slope of the yield curve (Estrella and Hardouvelis, 1991), stock market (Estrella and Mishkin, 1998), nonfarm payrolls and other employment and interest rate variables (Ng, 2014).
- The dominant modeling approach in the literature has been to estimate the probability of recessions using probit models with a binary target (e.g. recession = 1 and no recession = 0).

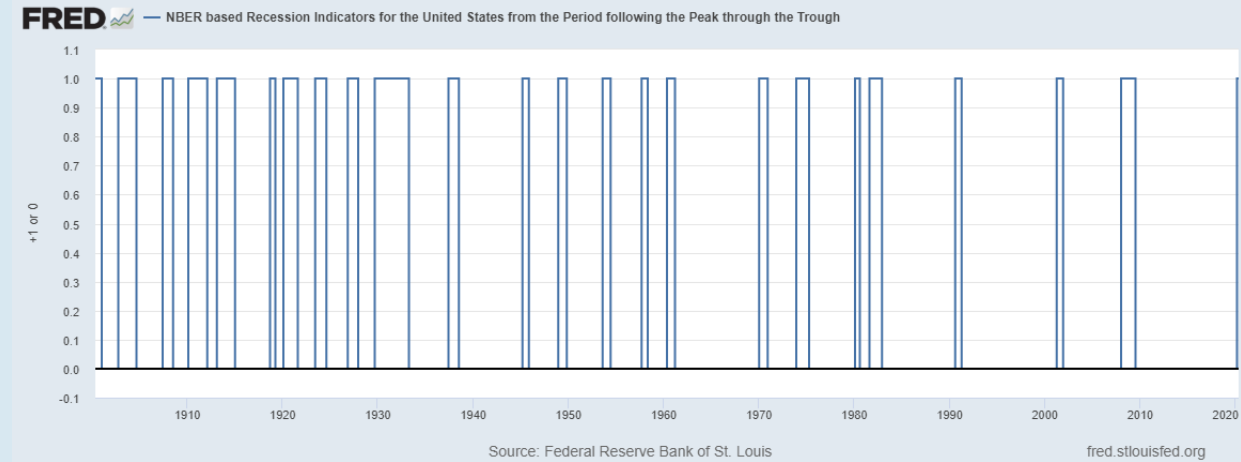


Can ML Help?

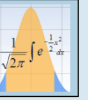
- Probit is a (generalized) linear model that use constant weight for predictors across all recessions.
- Machine Learning (ML) can help capturing nonlinear & interactive relationships among variables through dynamic data-driven mechanisms.
- Above problem can be treated using supervised ML (binary classification problem) to predict probability of recession in a given month given a number of financial variables.
- Due to the rare yet high-impact nature of the recessions, this problem is best addressed using an imbalanced-classification approach.

Data

- The US recession dates are obtained from the NBER database
- Predictors chosen from variables frequently reported in literature as informative about macro-economy and recessions
- Data spanning 01/1959-12/2019:
 - 732 months, 101 recessions (circa 14%)
 - 6:1 class ratio, moderate-severe imbalance

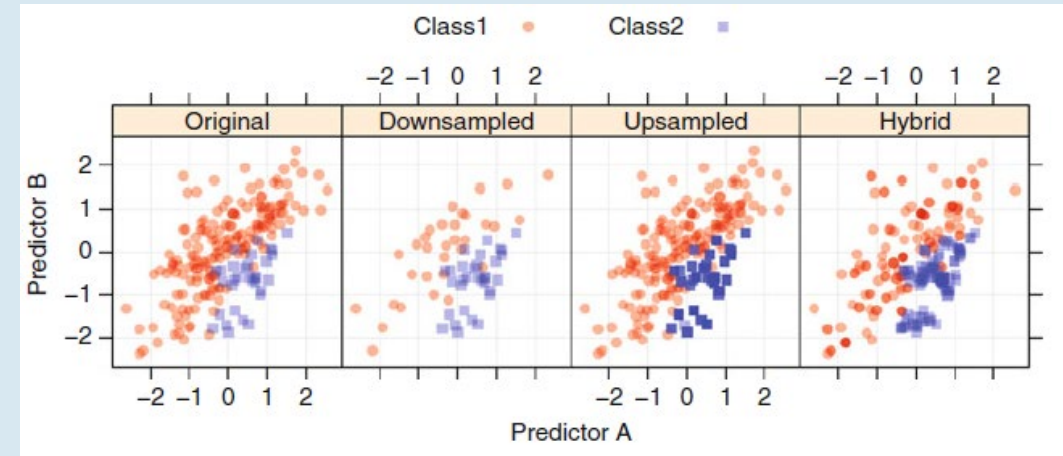


Series	Acronym	Transformation (lag in months)	Reported by
Federal Funds Rate	FEDFUNDS	DIFF (1M)	Sephton (2001), Ng (2014)
Industrial Production	INDPRO	DIFF_LOG (1M)	Sephton (2001), Chauvet and Piger (2008)
Total Nonfarm Payroll	PAYEMS	DIFF_LOG (1M)	Camacho et. al. (2012)
SP500 Index	SP500	DIFF_LOG (1M)	Estrella and Mishkin (1998), Qi (2001)
10 Year Treasury Bond	TY10	DIFF (1M)	Estrella and Mishkin (1998)
Unemployment Rate	UNEMPLOY	DIFF_LOG (1M)	Sephton (2001), Ng (2014)
Slope of the Yield Curve	YC	TY10 – FEDFUNDS (1M)	Estrella and Hardouvelis (1991), Estrella and Mishkin (1996), Wright (2006)



Pre-processing

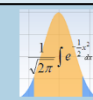
- To compensate for class-imbalance
 - Down-sampling the majority class (we employ this)
 - Up-sampling minority class (r. s. w. replacement)
 - Hybrid methods like “SMOTE”
- To split data for model training and evaluation (80/20 split)
 - Training sample, 01/1959-12/2006, 576 observations, ~14% flagged as recessions
 - Test sample 01/2007-12/2019, 156 observations, ~12% flagged as recessions
 - Train models on first part and evaluate on test data containing 2008-2009 crises
 - A walk-forward time-series cross-validation scheme employed to avoid look ahead bias





ML Models and Metrics

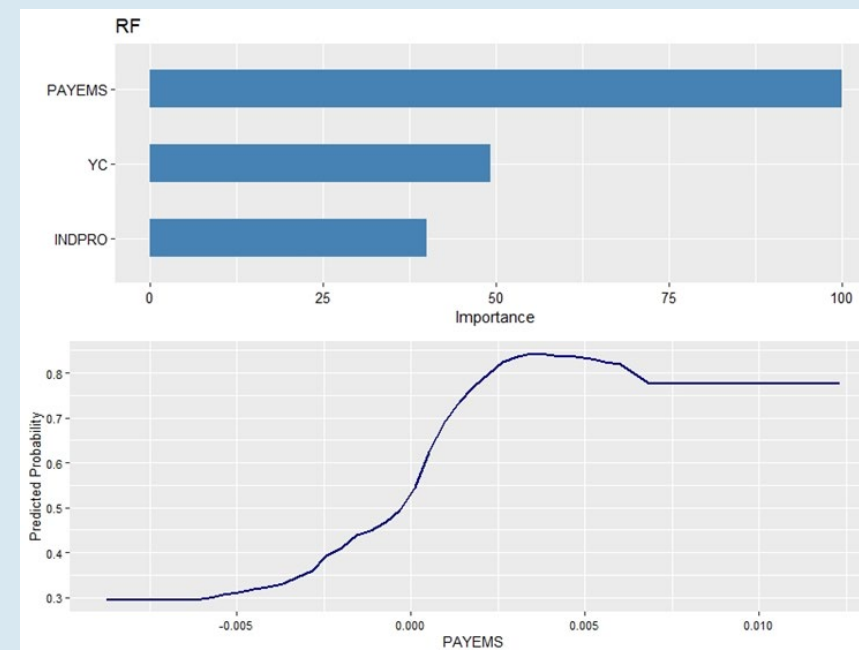
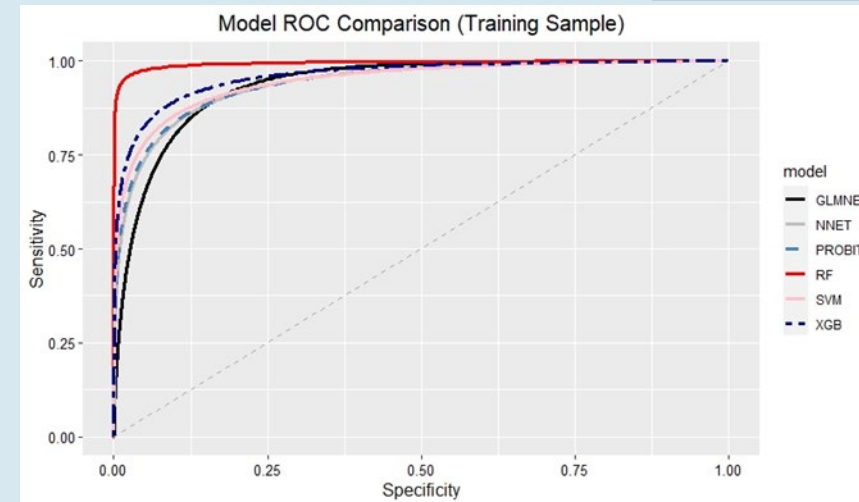
- The following Classification algorithms are employed
 - Probit benchmark model
 - GLMNET generalized linear model with regularization
 - SVM with Radial Basis Function
 - Random Forest and XGBoost ensemble with regularization
 - Single hidden layer feedforward Neural Network
- The following performance metrics are employed
 - Accuracy Rate, the proportion of all correctly predicted cases (positive or negative)
 - Precision, the ratio of true positives over sum of true positives and true negatives
 - Sensitivity or Recall, percentage of positive cases correctly predicted
 - Specificity, percentage of negative cases correctly predicted
 - Area Under ROC Curve (AUC), combining sensitivity and specificity, robust to class imbalances
 - F-Score, the harmonic mean of precision and recall
 - H-Measure, misclassification-cost weighted metric
 - Kolmogorov-Smirnov (KS), representing a model's ability to distinguish classes



Training Results

- The best performance overall belongs to a nonlinear “ensemble” RF model.

Model	Accuracy	AUC	F-Score	H-Measure	KS	Precision	Sensitivity	Specificity
PROBIT	87%	89%	53%	64%	79%	53%	93%	86%
GLMNET	88%	89%	54%	64%	79%	54%	91%	87%
SVM	90%	90%	59%	67%	80%	59%	90%	90%
RF	91%	95%	61%	78%	89%	61%	100%	89%
XGB	89%	90%	58%	66%	79%	58%	90%	89%
NNET	86%	89%	52%	62%	78%	51%	93%	85%

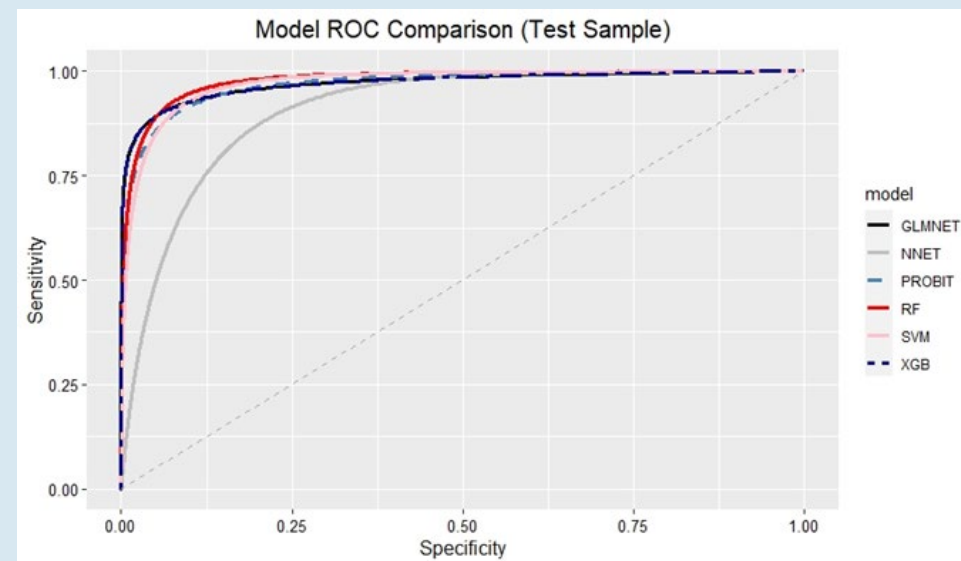




Test Results

- The best performance overall belongs to RF again.

Model	Accuracy	AUC	F-Score	H-Measure	KS	Precision	Sensitivity	Specificity
PROBIT	91%	90%	71%	70%	81%	59%	89%	91%
GLMNET	91%	95%	73%	79%	90%	58%	100%	90%
SVM	93%	91%	76%	74%	83%	65%	89%	93%
RF	94%	94%	80%	82%	89%	69%	95%	94%
XGB	94%	94%	78%	80%	88%	67%	95%	93%
NNET	92%	84%	68%	58%	68%	64%	74%	94%

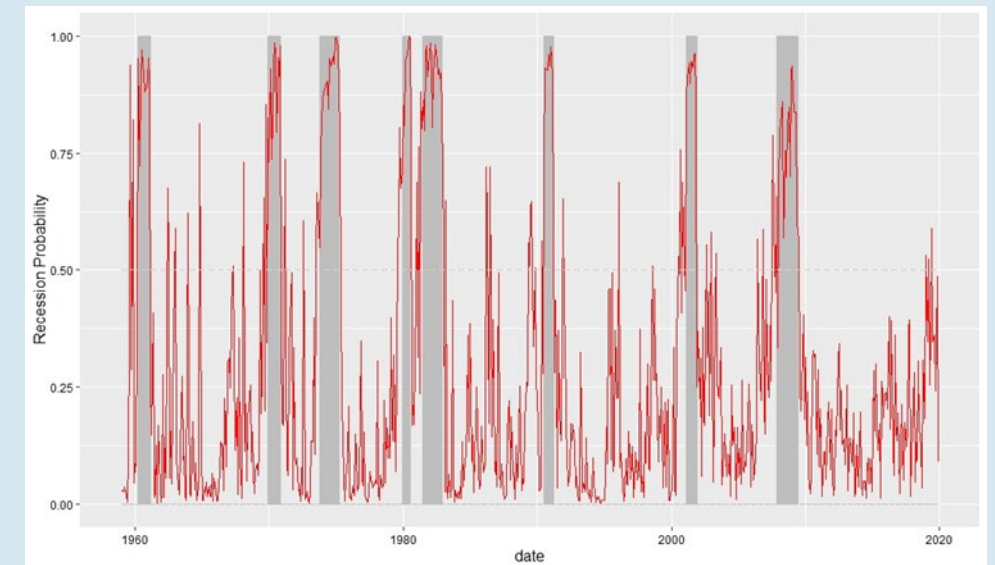


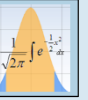


Full Sample Results

- RF has best overall performance
 - Near perfect sensitivity (correctly detecting all recessions)
 - Elevated recession probabilities late 2019-early 2020 (Trade-war, COVID-19)

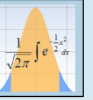
Model	Accuracy	AUC	F-Score	H-Measure	KS	Precision	Sensitivity	Specificity
PROBIT	88%	90%	68%	65%	79%	54%	92%	87%
GLMNET	88%	90%	69%	67%	81%	55%	93%	88%
SVM	90%	90%	72%	69%	81%	60%	90%	90%
RF	92%	95%	76%	79%	89%	62%	99%	90%
XGB	90%	91%	72%	69%	81%	59%	91%	90%
NNET	88%	88%	66%	62%	76%	53%	89%	87%





Summary and Conclusions

- Forecasting US business cycles & recessions is a major concern of investors and policy-makers alike
- Traditional models (e.g. linear probit) show limitations given the changing nature and drivers of recessions
- ML can outperform simpler recession prediction methods
 - through processing nonlinear & interactive relationships in data
 - low-frequency high-impact event predictions gain significantly through ML & sub-sampling mechanisms
- Data-driven predictive methods may better inform financial risk management strategies to prompt intervention policies



Useful Resources

- Estrella, A., Mishkin, F. S. (1998). "Predicting U.S. recessions: Financial variables as leading indicators," *The Review of Economics and Statistics*, 80(1), 45-61.
- Ng, S. (2014). "Viewpoint: Boosting recessions," *Canadian Journal of Economics*, 47(1), 1-34.
- Hastie, T., Tibshirani, R., Friedman, J., "The Elements of Statistical Learning, Data Mining, Inference, and Prediction", 2nd ed. Springer, 2008.
- Kuhn, M., and K. Johnson. *Applied Predictive Modeling*. New York: Springer, 2013.
- NBER Website: <https://www.nber.org/cycles/cyclesmain.html>
- FRED Website: <https://fred.stlouisfed.org/>
- The Journal of Financial Data Science: <https://jfds.pm-research.com/>
- FDP Curriculum: <https://fdpinstitute.org/>



Notes

- This work is to appear in *The Journal of Financial Data Science*, Sep. 2020.
- The material presented is for informational purposes only. The views and results presented are mine, are subject to change based on market and other conditions and factors, and may not necessarily reflect the views of my affiliations.

Q & A

Kind reminders of upcoming webinars as we go through the Q & A.
Add your questions in the chat room please.



WEBINAR SERIES
A Conversation With...



Dr. Gueorgui Konstantinov CAIA, FDP

Factor & Asset Allocation: The Role of Networks and LASSO

July 16, 2020 @ 11:00 AM EDT



WEBINAR SERIES
A Conversation With...



Michelle Seng Ah Lee
PhD Researcher, Compliant &
Accountable Systems Research Group
University of Cambridge

**Challenges of Algorithmic
Fairness in Financial Services**

July 29, 2020 @ 1:00 PM EDT



WEBINAR SERIES
A Conversation With...



Nigel Noyes
Principal Data Scientist
Quicken Loans

Data Science in Production

Aug 6, 2020 @ 1:00 PM EDT

Our upcoming webinars and webinar recordings can be found on

www.fdpinstitute.org/webinars

www.caia.org/caia-infoseries

Learn more about the FDP Institute at

www.fdpinstitute.org



In closing

Our upcoming webinars and webinar recordings can be found on

www.fdpinstitute.org/webinars

www.caia.org/caia-infoseries

Learn more about the FDP Institute at

www.fdpinstitute.org

