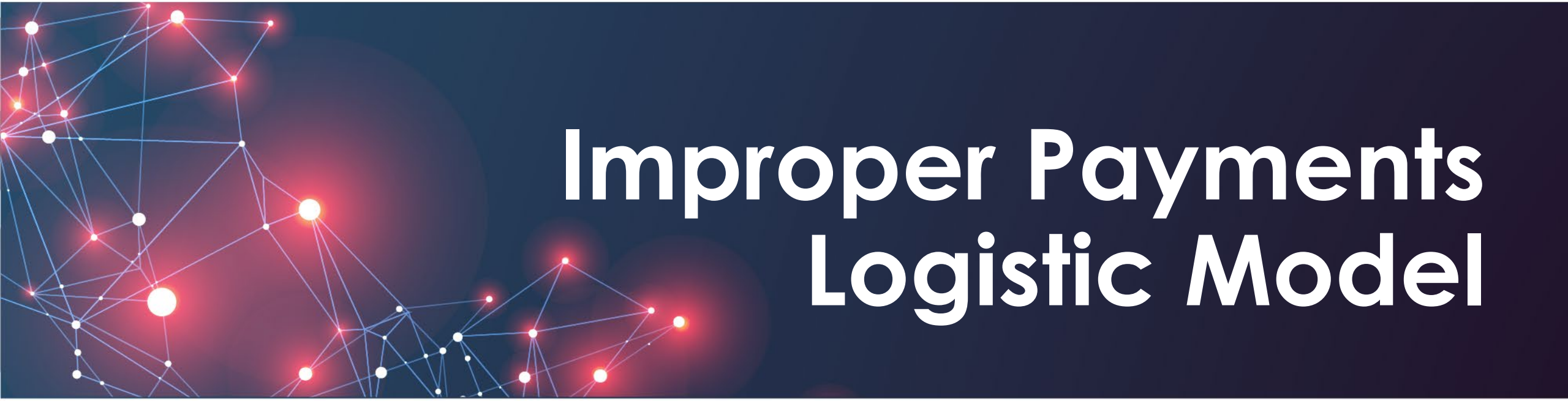
AGA®

TECHNOLOGY & TRANSFORMATION SUMMIT

November 18, 2020



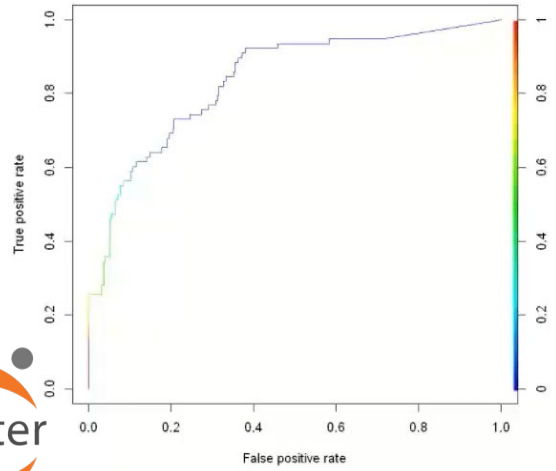
Improper Payments Logistic Model

National Science Foundation

AGA®

#TTS20

Improper Payments Logistic Model Why NSF?




```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5651  -0.3840  -0.3301  -0.2915   2.5149
```

AUDITYEAR	DBKEY	ELECAUDITSID	CFDA	FEDERALPROGRAMNAME	Findings	QCOSTS	TOTFEDEXPEND
<int>	<int>	<int>	<chr>	<chr>	<dbl>	<chr>	<dbl>
2013	67	15168345	47.074	NATIONAL SCIENCES FOUNDATION	0	N	6419632
2013	67	15168346	47.076	NATIONAL SCIENCES FOUNDATION	0	N	6419632
2013	83	14197645	47.049	MRI DEVELOPMENT OF ULTRACOLD POTASSIUM RYBERG SOURCE	0	N	11524319
2013	83	14197646	47.049	RUI RYBERG ATOMS & THEIR EFFECT ON ULTRA-COLD PLASMA DYNAMICS	0	N	11524319
2013	83	14197650	47.049	RUI SYNTHESES, STRUCTURES & REACTIONS OF STRAINED CYCLIC ALLENES	0	N	11524319
2013	83	14197652	47.049	CAREER: MOLECULAR DESIGN THROUGH OXACALIZARENES	0	N	11524319

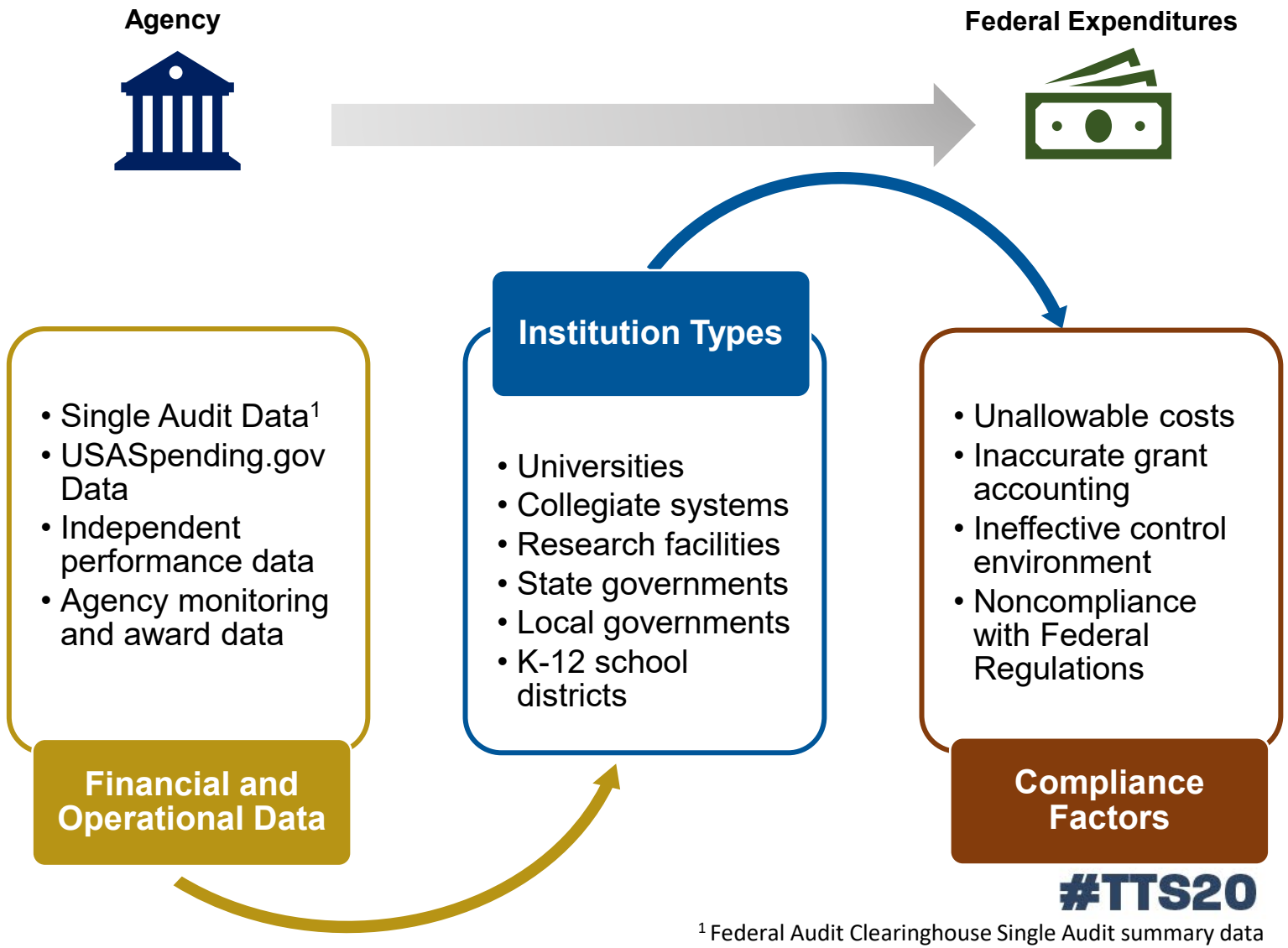
```
Coefficients:

(Intercept)
REPEAT_FINDINGS
program_Income
cooperative
AuditOriginal
Allowable_Costs
sqrt_TOTFEDEXPEND
Risk_CategoryB
Risk_CategoryC
Monitoring
```

- Introductory use case to begin implementing data science tools and techniques
- Open source – no licenses to procure, no hoops
- Turning data into information to drive decision making
- Professional development - upskilling for the workforce of the future

#TTS20

Improper Payments Logistic Model Challenges and Opportunities

- Management commitment to stewardship and oversight of grant funds
- Annual requirement for PIIA compliance and reporting
- Data sources were not fully integrated for monitoring activities
- Opportunity for continuous development for deploying data science tools



#TTS20

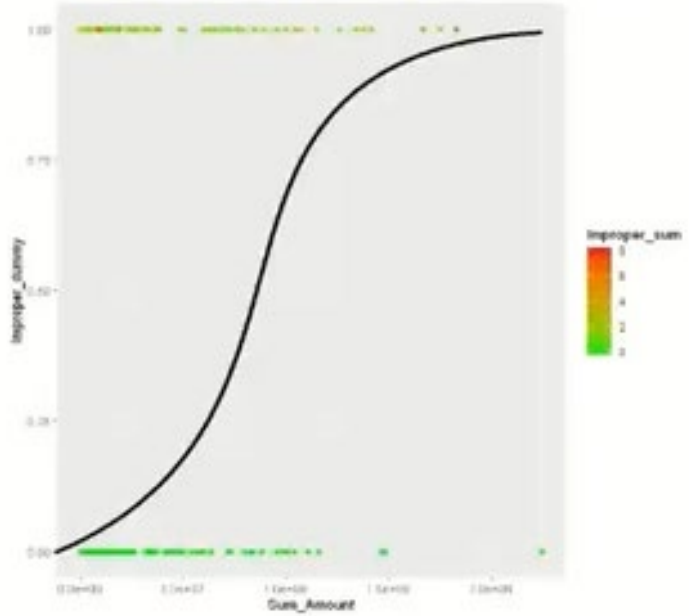
[1] Federal Audit Clearinghouse Single Audit summary data

Improper Payments Logistic Model Building the Solution

Model Development

- Created **inventory** of available data to help **understand risks** associated with institutions receiving federal funds
- Used **R** package dplyr to combine and manipulate data sets, design, and test the model
- Conducted tests on **model validity** including reviewing for correlation and multicollinearity, and bias/anomalies

```
In [11]: Logistic_demo <- readPNG("Logistic_demo.png")
         plot.new()
         rasterImage(Logistic_demo,0,0,1,1)
```



Equation for Probability of an Improper Payment Finding

$$log\frac{1}{1-p} = \beta X_1 + \beta X_2 + \cdots \beta X_9$$

Improper Payments Logistic Model Key Outcomes and Next Steps

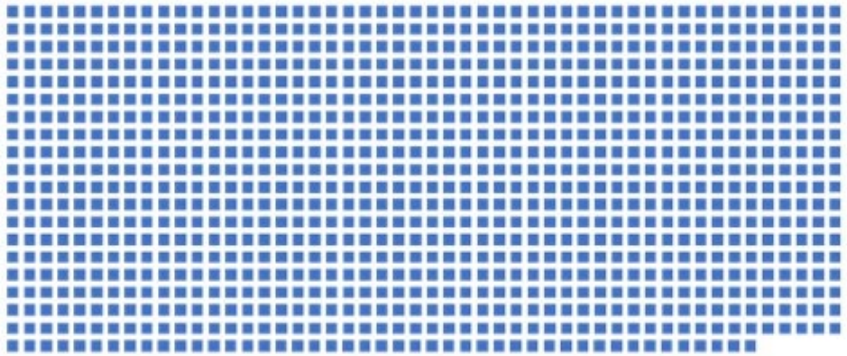
Model Results

Using the initial results as a basis for prediction, the model predicted with **87% accuracy** whether an institution would have an improper payment over a five-year period.

	Actual Positive	Actual Negative
Predicted Positive	**34 (8.9%)**	25 (5.7%)
Predicted Negative	39 (7.8%)	**339 (77.6%)**

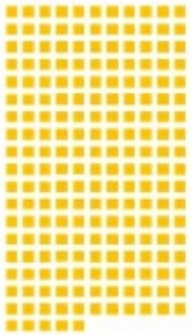
Compliance Summary

No Internal Control Findings



995

Material Weakness or Significant Deficiency



203

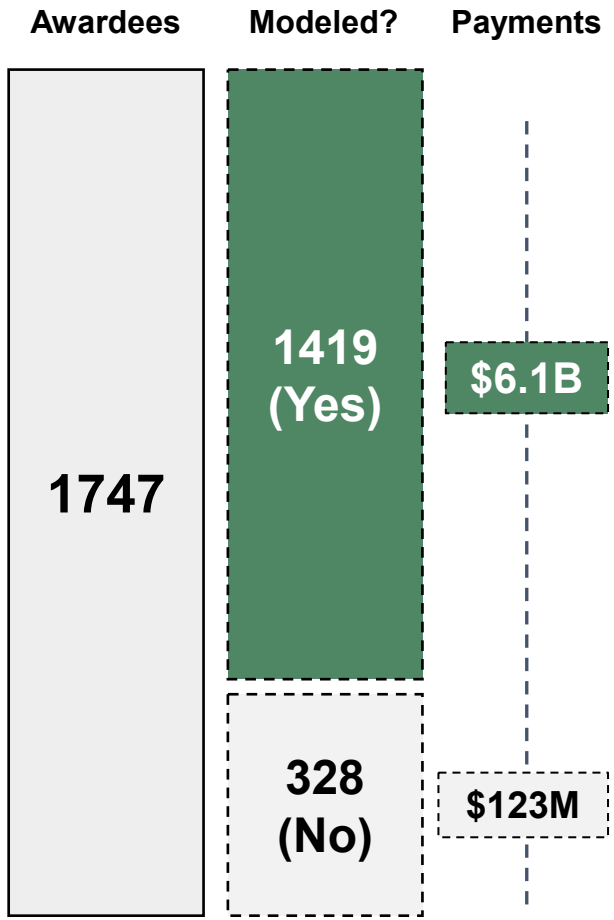
Improper Payments Logistic Model Considerations for the Community

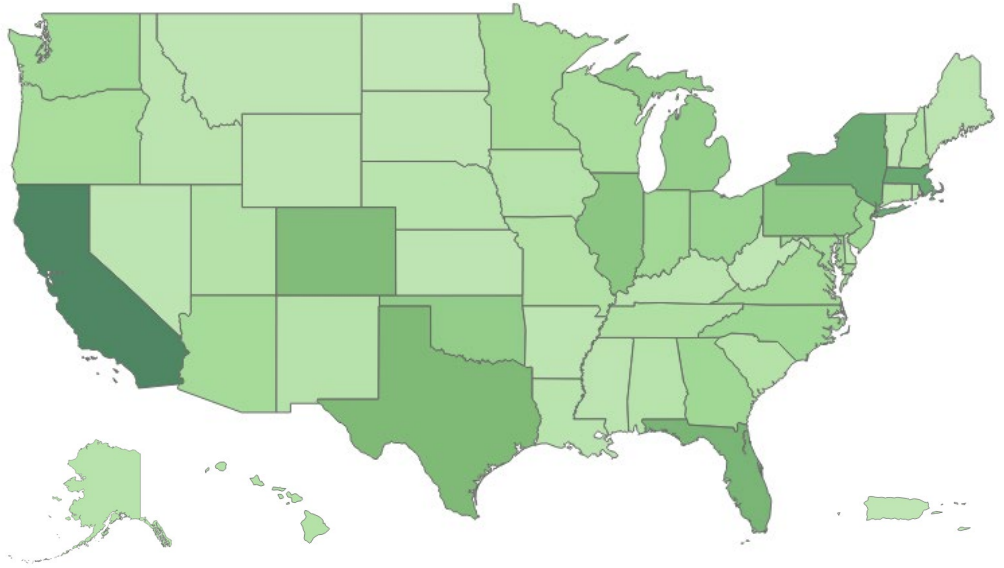
NSF Recipient Inclusion Criteria

- Payments in last 2 FY
- Single Audit in last 5 FY
- Included in annual risk assessment
- No missing variables

Transferability Considerations

- Average Size of Award
- Grantee Portfolio Size
- Recipient Types
- IP Root Causes
- Available Data Sources





Address gaps through other analyses

AGA®

NSF

#TTS20

Improper Payments Logistic Model
Crowdsourcing the Solution!

We have a **shared responsibility** to monitor the problem.

Why should we **stand alone** with the **solution**?

NEW!

No cost to implement!

Talk to your IT department about R as approved software.

Download today!

Immediate foundation to expand with your own data.

Review, collaborate, and share!

We all benefit when we can share our collective experiments.

https://github.com/NSF-DFM/AGA_TTS2020_DataSciencePilot

AGA

NSF

#TTS20