Predictive Insights for Secure Growth

By: Andrew Sims, Brett Olson, Dean Omirly, Isaac Wang, Marcus Steen

DEACON FINANCIAL SERVICES

Getting to know **Deacon Financial Services**

Who is DFS?

Getting to know **Deacon Financial Services**

Who is DFS?

- **Mid-Size** bank located in Charlotte, NC

Who is DFS?

- ❑ **Mid-Size** bank located in Charlotte, NC
- ❑ **~2.5 million** retail customers

Getting to know **Deacon Financial Services**

Who is DFS?

- ❑ **Mid-Size** bank located in Charlotte, NC
- ❑ **~2.5 million** retail customers
- ❑ Moving beyond **branch-only model**

Getting to know **Deacon Financial Services**

Customer Growth Crisis

Getting to know **Deacon Financial Services**

Customer Growth Crisis

- **Losing younger** customers to upcoming fintechs

Getting to know **Deacon Financial Services**

Customer Growth Crisis

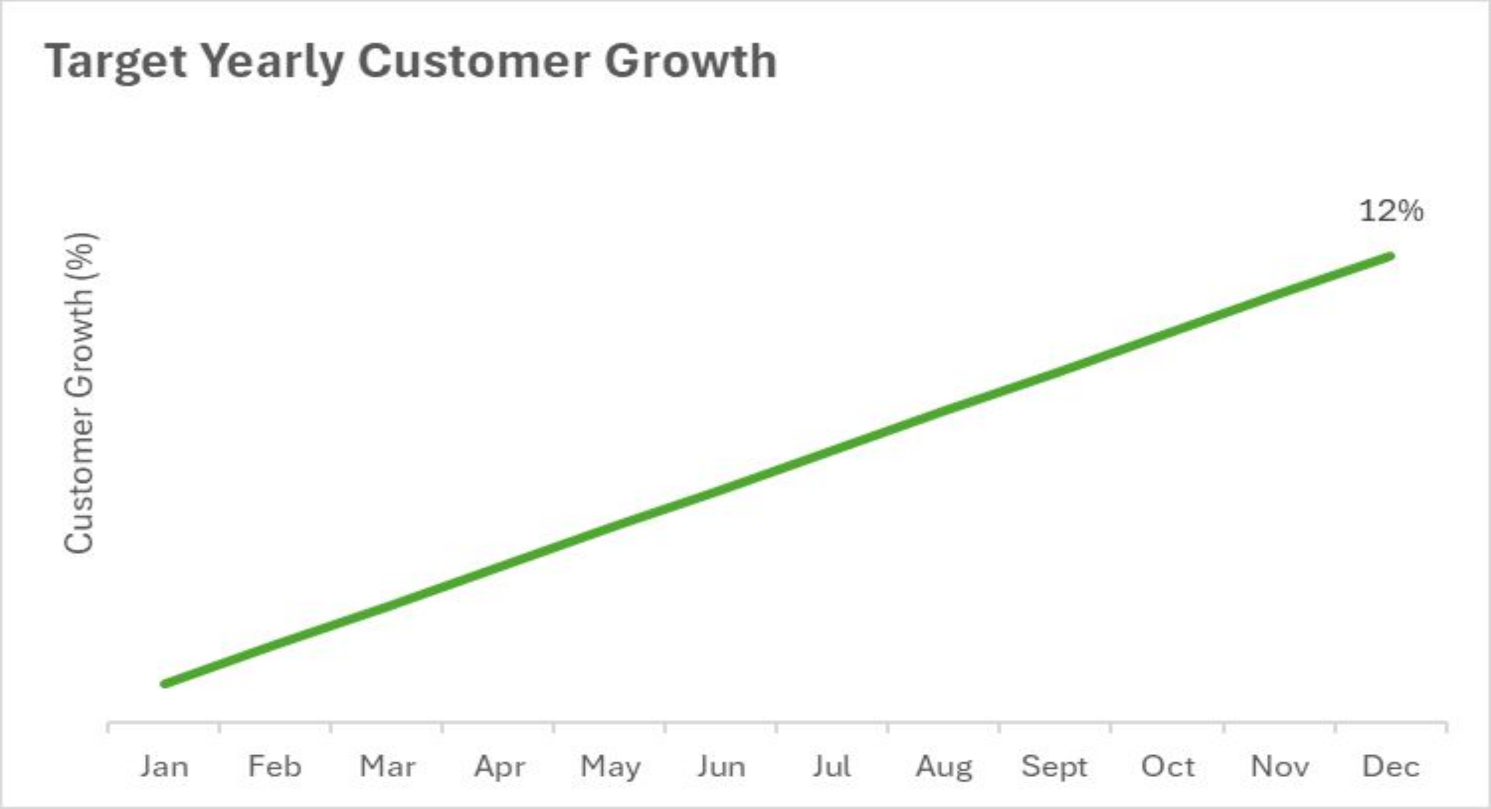
- ❑ **Losing younger** customers to upcoming fintechs
- ❑ Reduced **branch traffic**

Getting to know **Deacon Financial Services**

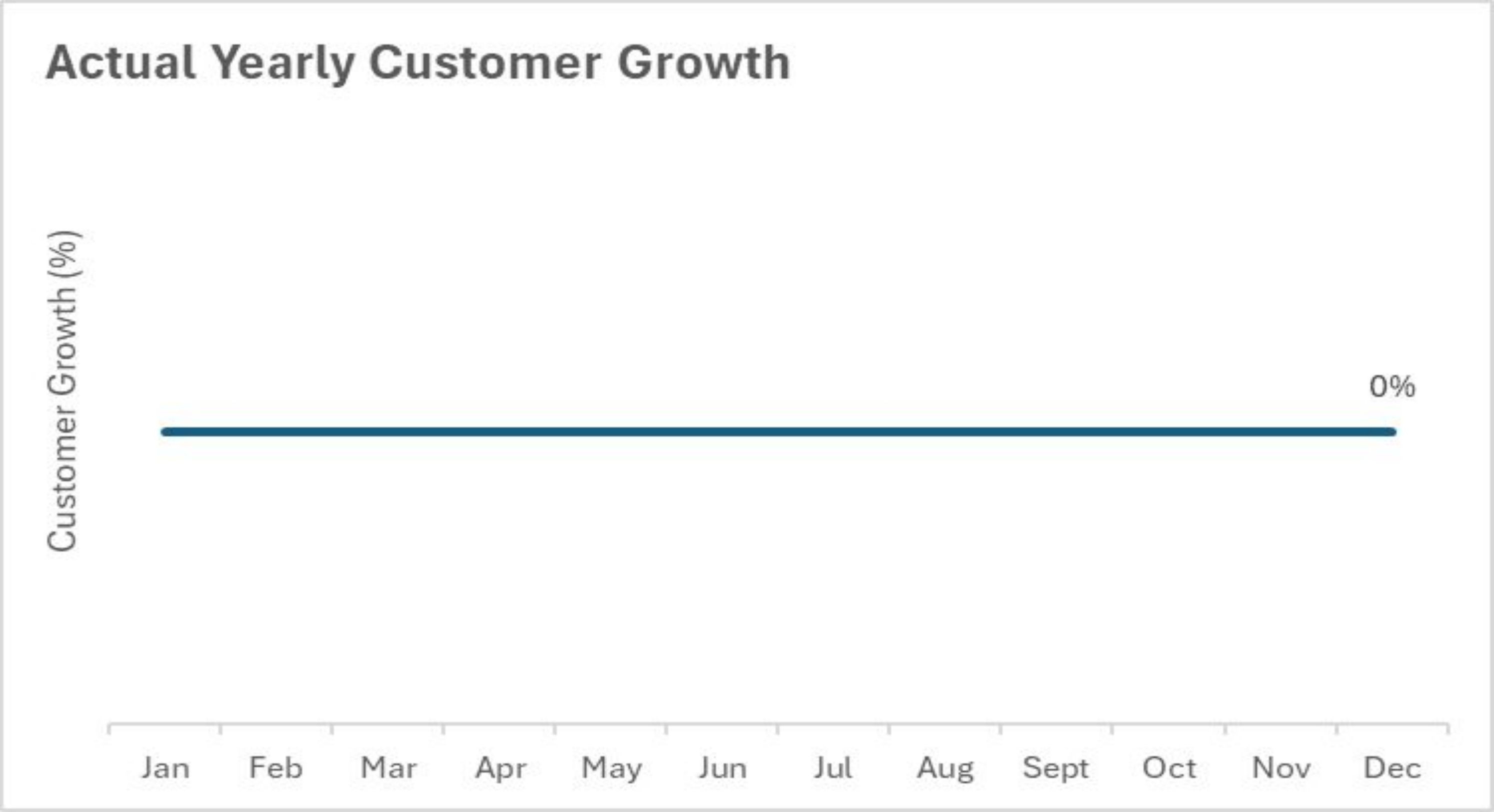
Customer Growth Crisis

- ❑ **Losing younger** customers to upcoming fintechs
- ❑ Reduced **branch traffic**
- ❑ Internal **friction**

Getting to know **Deacon Financial Services**



Getting to know **Deacon Financial Services**



Getting to know **Deacon Financial Services**

New Money Bonus

Getting to know **Deacon Financial Services**

New Money Bonus

- Offers **$200** to new customers

Getting to know **Deacon Financial Services**

New Money Bonus

- ❑ Offers **$200** to new customers
- ❑ **Fully digital**, low friction application

Getting to know **Deacon Financial Services**

New Money Bonus

- ❑ Offers **$200** to new customers
- ❑ **Fully digital**, low friction application
- ❑ Hyper-targeted **digital advertising**

The Core Problem

Campaign growth has led to **more fraud**

3-Month Results

- 8K→**30K** apps **(+375%)**

Campaign growth has led to **more fraud**

3-Month Results

- 8K→**30K** apps **(+375%)**
- Conversion 12% →**18%**
- CAC $165 (**40% better**)
- Regaining market share of **younger customers**

Campaign growth has led to **more fraud**

3-Month Results

- 8K→**30K** apps **(+375%)**
- Conversion 12% →**18%**
- CAC $165 (**40% better**)
- Regaining market share of **younger customers**

Hidden Crisis

- **Fraud rates** 2-3%→**12%**

Campaign growth has led to **more fraud**

3-Month Results

- 8K→**30K** apps **(+375%)**
- Conversion 12% →**18%**
- CAC $165 (**40% better**)
- Regaining market share of **younger customers**

Hidden Crisis

- **Fraud rates** 2-3%→**12%**
- False identities, account churning, strain on brand reputation
- **$1,200** in loss per fraud case

Campaign growth has led to **more fraud**

3-Month Results

- 8K→**30K** apps **(+375%)**
- Conversion 12% →**18%**
- CAC $165 (**40% better**)
- Regaining market share of **younger customers**

Hidden Crisis

- **Fraud rates** 2-3%→**12%**
- False identities, account churning, strain on brand reputation
- **$1,200** in loss per fraud case

Stakeholder Dilemma

- **Risk** wants to control fraud
- **Marketing** wants momentum
- **Operations** wants balance

DFS Leaders Emphasize **Different Priorities**

Marketing wants
Momentum

Jessica Medina
(Marketing)

- Sustain momentum
- Increase conversion
- CAC efficiency

DFS Leaders Emphasize **Different Priorities**

Marketing wants
Momentum

Jessica Medina
(Marketing)

- Sustain momentum
- Increase conversion
- CAC efficiency

Risk wants to
Control Fraud

Erin Woods
(Risk Management)

- Minimize fraud
- Regulatory compliance
- Loss Prevention

DFS Leaders Emphasize **Different Priorities**

Marketing wants *Momentum*

Jessica Medina
(Marketing)

- Sustain momentum
- Increase conversion
- CAC efficiency

Risk wants to *Control Fraud*

Erin Woods
(Risk Management)

- Minimize fraud
- Regulatory compliance
- Loss Prevention

Operations wants *Efficiency*

Carlos Jiménez
(Operations)

- Reduce call volume
- Clear process
- Sustainable growth

Project **Goal**

Build a predictive **fraud detection** model that **balances** customer growth while **reducing 12%** fraud rate

Preparing Data for Modeling

Initial dataset required **extensive** preprocessing

1. Initial Dataset

- ~1 million customer records | Approximately 35 customer variables | Application details, device signals, account behavior

Initial dataset required **extensive** preprocessing

1. Initial Dataset

- ~1 million customer records | Approximately 35 customer variables | Application details, device signals, account behavior

2. Data Cleaning

- Removed and imputed missing values | Created missing flag values | Fixed inconsistent formatting

Initial dataset required **extensive** preprocessing

1. *Initial Dataset*

- ~1 million customer records | Approximately 35 customer variables | Application details, device signals, account behavior

2. *Data Cleaning*

- Removed and imputed missing values | Created missing flag values | Fixed inconsistent formatting

3. *Feature Engineering*

- Created activity-ratio features | Built domain-driven fraud-risk flags | Added recency/frequency metrics

Feature engineering reveals **17 key** predictors

20 Features Created → 17 Significant for Fraud Prediction

Feature engineering reveals **17 key** predictors

20 Features Created → 17 Significant for Fraud Prediction

Multiple Emails Per Device (4-8 weeks)

- Fraud rate of **4.12%** vs 1.02% baseline

Feature engineering reveals **17 key** predictors

20 Features Created → 17 Significant for Fraud Prediction

Multiple Emails Per Device (4-8 weeks)

- Fraud rate of **4.12%** vs 1.02% baseline

Income/Salary Low Flags

- Financial profile **mismatches**

Feature engineering reveals **17 key** predictors

20 Features Created → 17 Significant for Fraud Prediction

Multiple Emails Per Device (4-8 weeks)

- Fraud rate of **4.12%** vs 1.02% baseline

Income/Salary Low Flags

- Financial profile **mismatches**

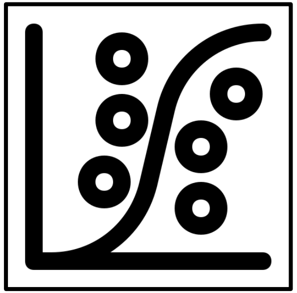
Zip Code Velocity

- **Rapid relocation** fraud signals

Exploring Model Development

Baseline models provide interpretable insights

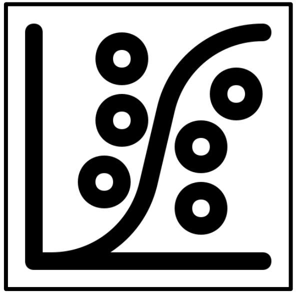
Logistic Regression



- **High** interpretability
- Odds Ratio
- Stable predictions
- Stakeholders **understand**

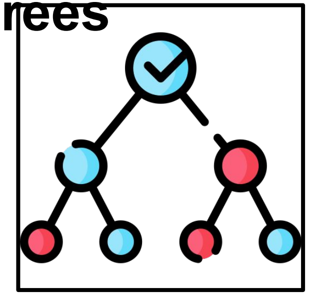
Baseline models provide interpretable insights

Logistic Regression



- **High** interpretability
- Odds Ratio
- Stable predictions
- Stakeholders **understand**

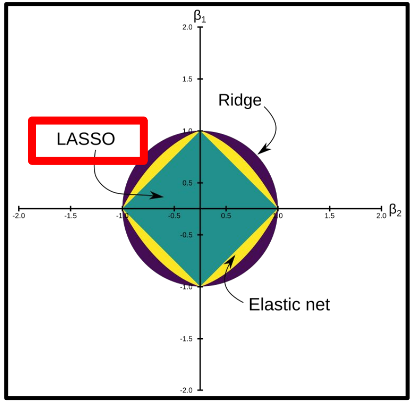
Decision Trees



- Hierarchical rules
- Easy to explain
- **Transparent** importance
- Stakeholders **understand**

Advanced models improve predictive accuracy

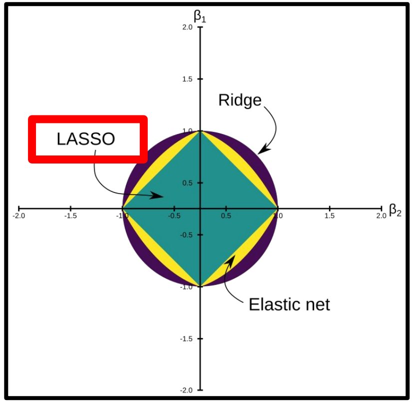
LASSO



- **Aggressive** feature selection
- Eliminates **weak** predictors
- Sparse and interpretable
- Prevents overfitting

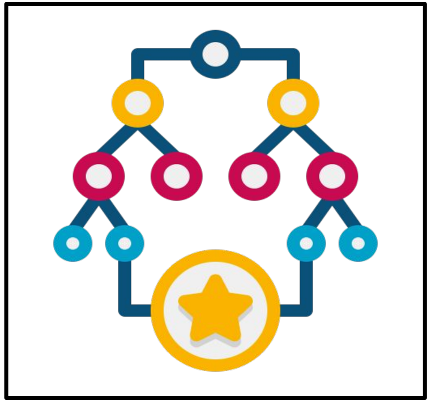
Advanced models improve predictive accuracy

LASSO



- **Aggressive** feature selection
- Eliminates **weak** predictors
- Sparse and interpretable
- Prevents overfitting

Random Forest



- **Multiple** decision trees
- Captures non-linear patterns
- **Votes** on predictions
- Handles complex interactions

Comparing Model Performance

LASSO leads in all advance metrics

Model	ROC-AUC	Youden	F1

LASSO leads in all advance metrics

Model	ROC-AUC	Youden	F1
Logistic Regression	.8462	.5378	.1625

LASSO leads in all advance metrics

Model	ROC-AUC	Youden	F1
Logistic Regression	.8462	.5378	.1625
Decision Tree	.7979	.4729	.1259

LASSO leads in all advance metrics

Model	ROC-AUC	Youden	F1
Logistic Regression	.8462	.5378	.1625
Decision Tree	.7979	.4729	.1259
Random Forest	.8765	.5964	.2252

LASSO leads in all advance metrics

Model	ROC-AUC	Youden	F1
Logistic Regression	.8462	.5378	.1625
Decision Tree	.7979	.4729	.1259
Random Forest	.8765	.5964	.2252
LASSO	.8835	.6045	.2307

Random forest ranks **best at predicting fraud**

Prediction Matrix

	Predicted Non-Fraud	*Predicted Fraud*
Actual Non-Fraud		
Actual Fraud		

Random forest ranks **best at predicting fraud**

Prediction Matrix

	Predicted Non-Fraud	*Predicted Fraud*
Actual Non-Fraud		
Actual Fraud	18.34%	**81.66%**

Random forest ranks **best at predicting fraud**

Prediction Matrix

	Predicted Non-Fraud	*Predicted Fraud*
Actual Non-Fraud	**77.94%**	22.06%
Actual Fraud	18.34%	**81.66%**

LASSO preserves **legitimate growth**

Prediction Matrix

	Predicted Non-Fraud	*Predicted Fraud*
Actual Non-Fraud		
Actual Fraud		

LASSO preserves **legitimate growth**

Prediction Matrix

	Predicted Non-Fraud	*Predicted Fraud*
Actual Non-Fraud	**81.65%**	18.35%
Actual Fraud		

LASSO preserves **legitimate growth**

Prediction Matrix

	Predicted Non-Fraud	Predicted Fraud
Actual Non-Fraud	**81.65%**	18.35%
Actual Fraud	21.24%	**78.76%**

Final Model Selection

Final Model: **LASSO**

Final Model: **LASSO**

Growth Preservation

Final Model: **LASSO**

Growth Preservation

Reduced Risk 12% to 2.4%

Final Model: **LASSO**

Growth Preservation

Reduced Risk

12% to 2.4%

Interpretability

Final Model: **LASSO**

Growth Preservation

Reduced Risk
12% to 2.4%

Interpretability

Stakeholder Alignment

Final Model: **LASSO**

Growth Preservation

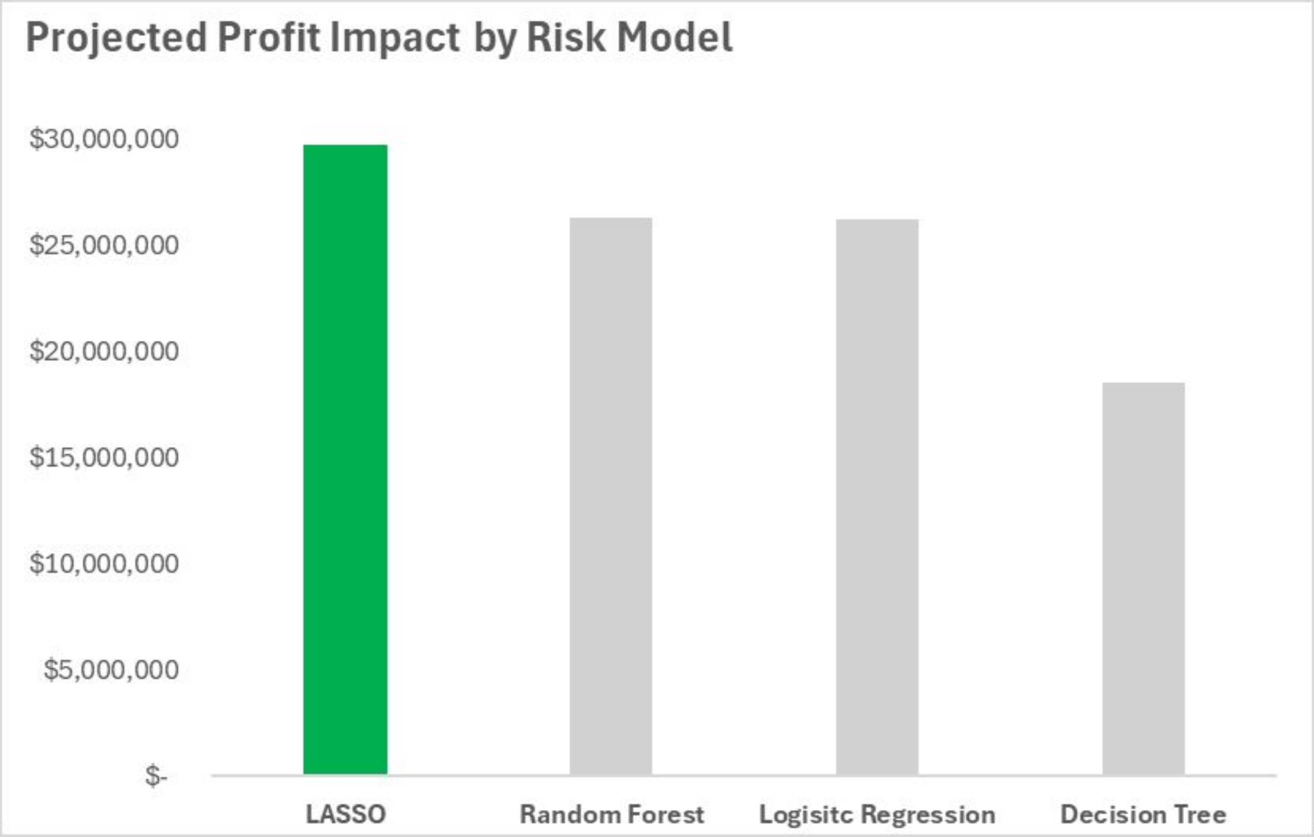
Reduced Risk 12% to 2.4%

Maximize Profit

Interpretability

Stakeholder Alignment

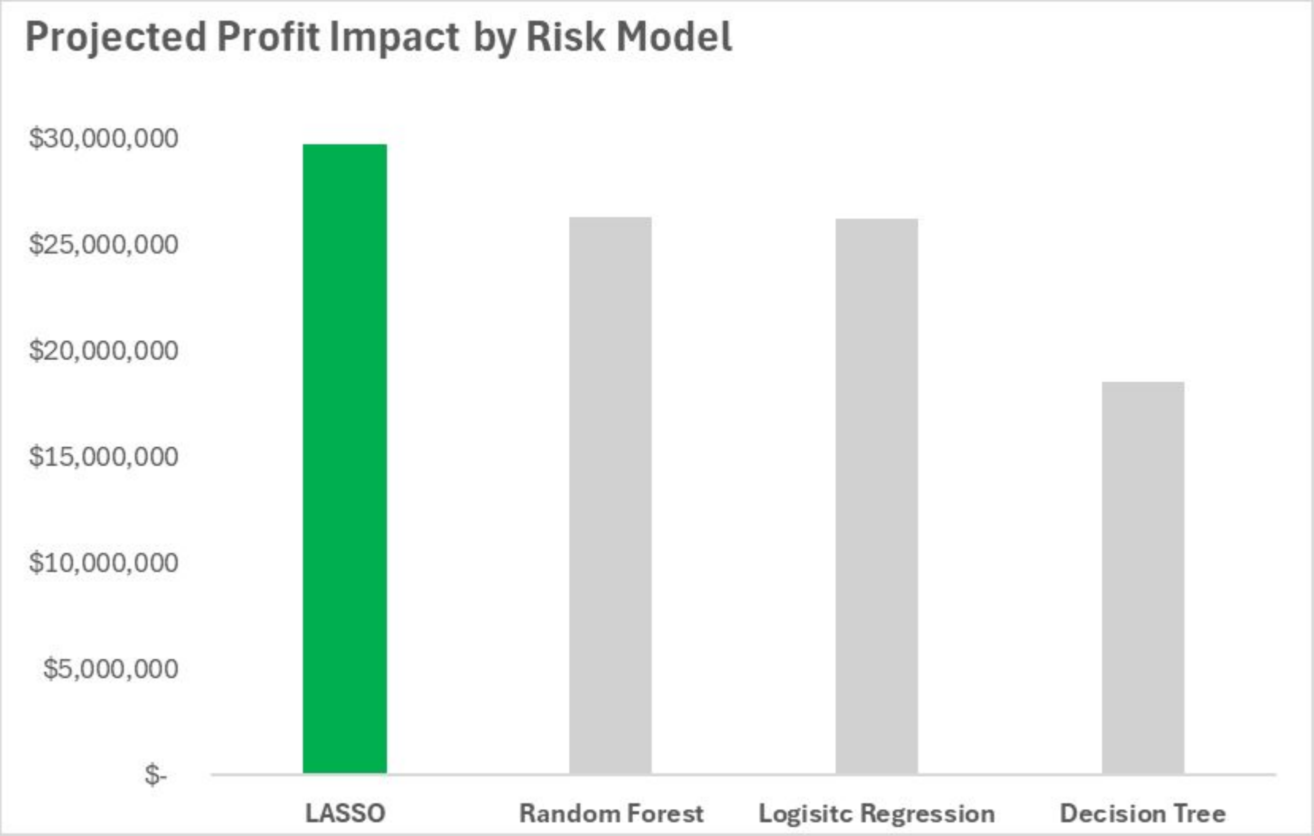
Final Model: **LASSO**



Final Model: **LASSO**

LASSO Potential Profit:

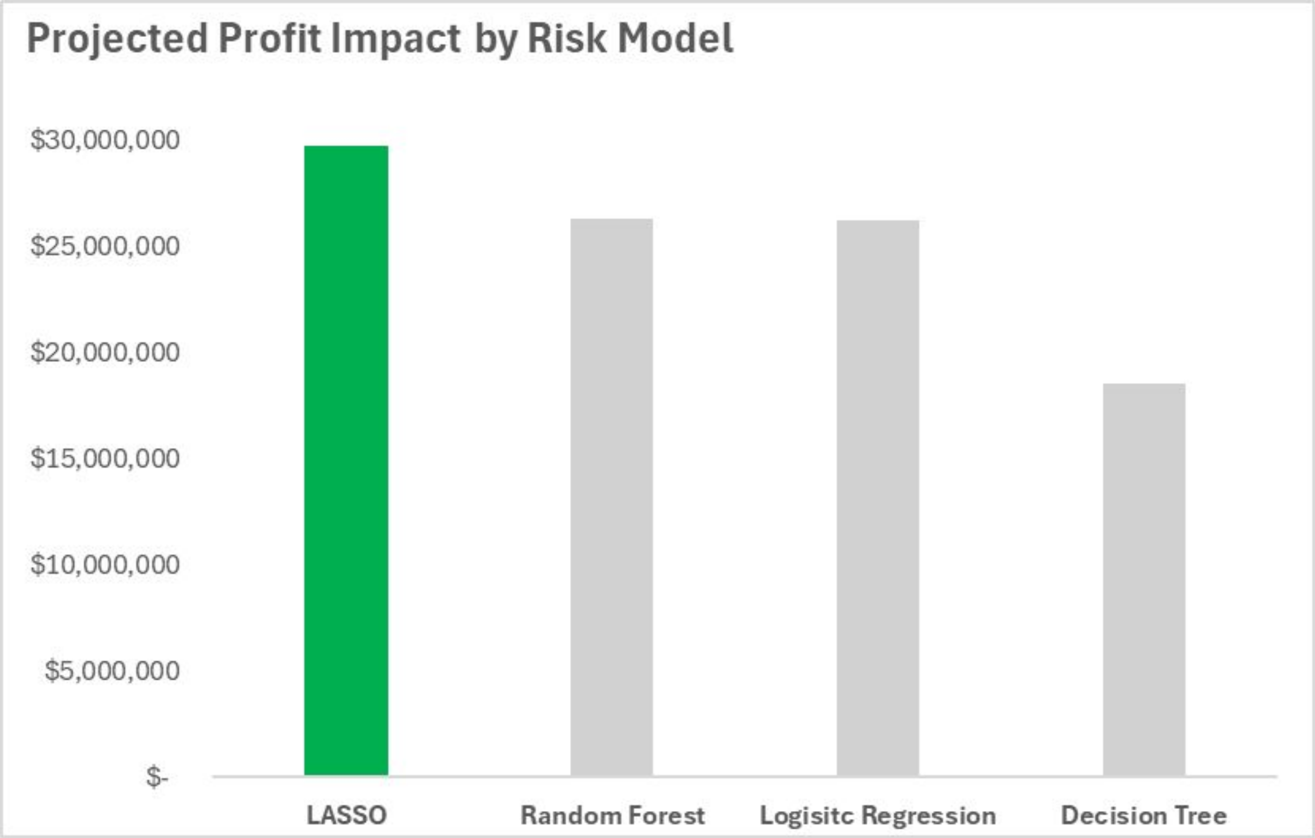
$29,752,400



Final Model: **LASSO**

Profit Advantage:

$3,417,600



Implementations & Next Steps

Staff Training to identify key fraud predictors

Low Risk

Prediction: **<0.005**

AUTO-APPROVE

Proceed to account setup

Minimal fraud indicators
Legitimate profile

Staff Training to identify key fraud predictors

Low Risk

Prediction: **<0.005**

AUTO-APPROVE

Proceed to account setup

Minimal fraud indicators
Legitimate profile

Medium Risk

Prediction: **0.005 - .011**

ACCOUNT VERIFICATION

In-Person screening

Some risk signals
Requires validation

Staff Training to identify key fraud predictors

Low Risk

Prediction: **<0.005**

AUTO-APPROVE

Proceed to account setup

Minimal fraud indicators
Legitimate profile

Medium Risk

Prediction: **0.005 - .011**

ACCOUNT VERIFICATION

In-Person screening

Some risk signals
Requires validation

High Risk

Prediction: **> .011**

AUTO-REJECT

Deny account setup

Multiple fraud indicators
Fraudulent Profile

Monthly reports to drive **threshold tuning**

KPI's of Fraud Prediction Model

Staff Training: Identify Key Fraud Predictors & Make Better Decisions

TOTAL FRAUD PREVENTION

200

Frauds prevented per 1,000 applications

CUSTOMER APPROVAL RATE

85%

Legitimate customers auto-approved or verified

LOW-RISK PROCESSING

850

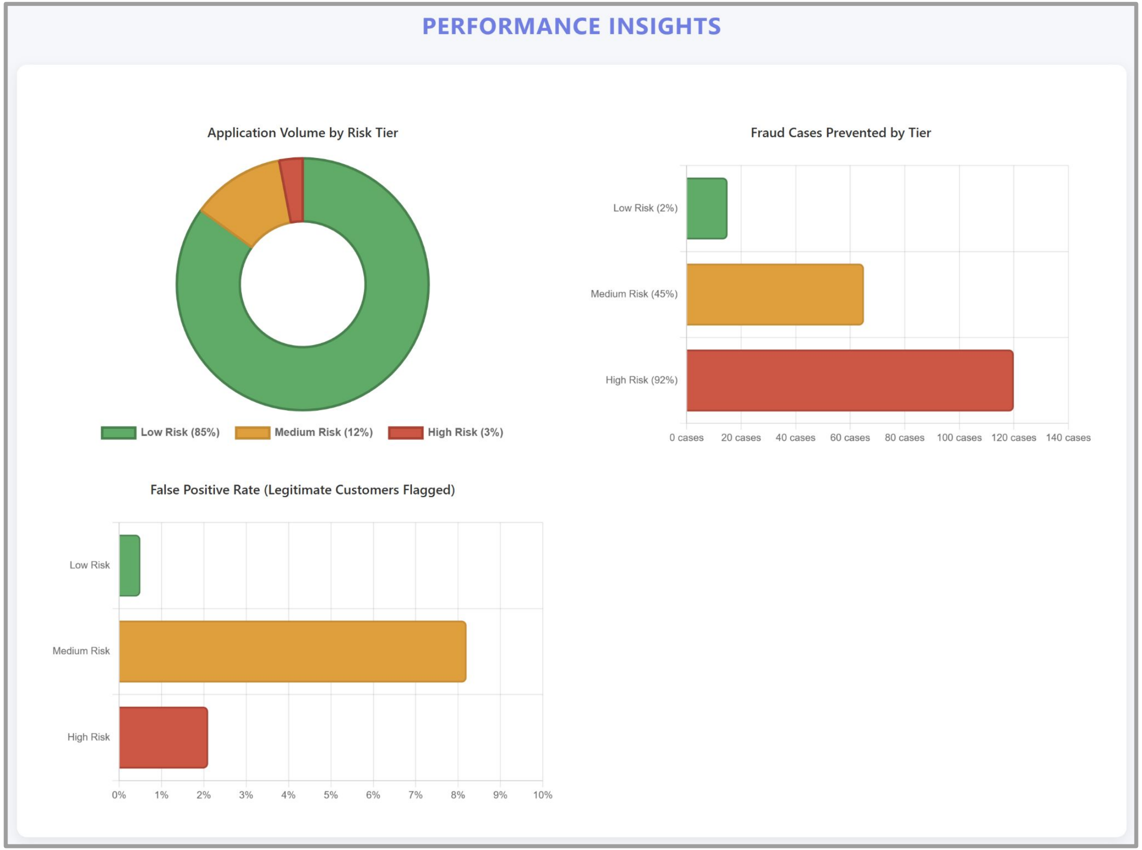
Applications auto-approved instantly (85%)

STAFF DECISION ZONE

120

Applications requiring human verification (12%)

Monthly reports to drive **threshold tuning**



Monthly reports to drive **threshold tuning**

CRITICAL INSIGHTS & FINDINGS

✓ Scale & Speed

85% of applications are low-risk and approved instantly. This means most legitimate customers get fast access to accounts without friction.

⚠️ The Critical Decision Zone

12% of applications (120 per 1,000) fall into the medium-risk "gray zone" where your staff's judgment determines outcomes. This is where training matters most.

✓ Fraud Prevention Power

200 frauds prevented per 1,000 applications through this three-tier system. The high-risk tier alone catches 92% of fraudsters with minimal false positives (2.1%).

⚠️ The Medium Tier Challenge

8.2% false positive rate in the medium tier means ~10 legitimate customers per 1,000 need extra verification. Balance is key—be thorough but fair.

✓ Staff Impact Metrics

- ✓ 45% fraud catch rate in medium tier through verification
- ✓ 90%+ staff accuracy achievable with proper training
- ✓ 2-minute reviews possible for low-risk applications

📍 Top Fraud Signals to Watch

- ✓ Housing Status BA
- ✓ Multiple emails per device
- ✓ Extreme credit scores
- ✓ No other credit cards
- ✓ Geographic inconsistencies

Expanding data to **boost predictive power**

Current Variables

Demographics

- Customer Age, Income Level, Salary

Activity

- Transaction Velocity, Banking History Months, Session Length

Credit Profile

- Risk Score, Credit Limit, Other cards

Device,Verification, Application

- Device OS, Emails/Device, Email Domain, Phone Validity, Payment Type, Status

Expanding data to **boost predictive power**

Current Variables

Demographics
- Customer Age, Income Level, Salary

Activity
- Transaction Velocity, Banking History Months, Session Length

Credit Profile
- Risk Score, Credit Limit, Other cards

Device,Verification, Application
- Device OS, Emails/Device, Email Domain, Phone Validity, Payment Type, Status

Recommended Additions

Social Network
- Social Connections, Multi-Account Rate, Device Migration

Geolocation
- IP Matching, Location Velocity, VPN Detection

Behavioral
- Login Patterns, Navigational Flows

Transaction Timing
- Spending Bursts, Sudden Increases in transfer amounts/frequency

Questions?

DEACON
FINANCIAL SERVICES

Appendix:

Model Paramters:		
Number of Customers within the last 12 months:		120,000
Number of Legitimate Customers in the last 12 months:		105,600
Number of Fraudulent Customers in the last 12 months:		14,400
Average yearly Customer value add:	$	500
Average amount of loss per fraud case:	$	1,200

LASSO:		
Number of Legitimate Customers predicted correctly:		86,222
Number of Legitimate Customers predicted incorrectly:		19,378
Number of Fraudulent Customers predicted incorrectly:		3,058
Money Made:	$	43,111,000
Money Lost:	$	13,358,600
Total Yearly Profit:	$	29,752,400

Logistic Regression:		
Number of Legitimate Customers predicted correctly:		83,360
Number of Legitimate Customers predicted incorrectly:		22,240
Number of Fraudulent Customers predicted incorrectly:		3,629
Money Made:	$	41,680,320
Money Lost:	$	15,474,800
Total Yearly Profit:	$	26,205,520

Random Forest:		
Number of Legitimate Customers predicted correctly:		82,304
Number of Legitimate Customers predicted incorrectly:		23,296
Number of Fraudulent Customers predicted incorrectly:		2,641
Money Made:	$	41,152,000
Money Lost:	$	14,817,200
Total Yearly Profit:	$	26,334,800

Decision Tree:		
Number of Legitimate Customers predicted correctly:		75,578
Number of Legitimate Customers predicted incorrectly:		30,022
Number of Fraudulent Customers predicted incorrectly:		3,502
Money Made:	$	37,788,960
Money Lost:	$	19,213,536
Total Yearly Profit:	$	18,575,424

Appendix:

Top 5 features with highest absolute Penalty Term for LASSO Regression

Rank	**Variable**	**Variable Definition**	**Absolute Penalty Term**
1	*prev_address_provided*	Whether the applicant provided their previous address	0.549254
2	*high_risk_housing*	Where applicant's current residential status is either 'BA' or 'BC'	0.424640
3	*has_other_cards_yes*	Whether the applicant has other cards from the same banking company	0.325271
4	*high_risk_device*	Operating system of the device used for the request is either Windows or Macintosh	0.299116
5	*multiple_emails_per_device*	Where device_distinct_emails_8w is greater than 2	0.290050