# APPLYING PREDICTIVE MODELING TECHNIQUES IN GOVERNMENT

§sas
THE POWER TO KNOW.

PA TechCon

# PREDICTIVE MODELS DEFINED

**A predictive model is a form of data mining that is concerned with the prediction of future probabilities and trends**



**NO MATCH** (68.3%)

**MATCH** (94.7%)

**NO MATCH** (44.0%)

**NO MATCH** (73.9%)

§sas | THE POWER TO KNOW.

**CONDITIONS UNDER WHICH IT IS BEST USED**



- ➢ Historical data
- ➢ Results data
- ➢ Past behavior is expected to continue in the future
- ➢ Timeliness matters

# KEEP CHILDREN SAFE

§sas
**THE POWER TO KNOW**

| **In the US, four children die every day as a result of child abuse**



- 3 million reports of child abuse and neglect, involving 6 million children filed annually
- Child abuse occurs at every socioeconomic level, across ethnic and cultural lines, within all religions and at all levels of education
- Annual cost of child abuse is $124B

**THE CHALLENGE**: *Can I identify which kids are at imminent risk of physical harm before they are injured or killed?*

**THE DATA** This government had over five million cases on reports of neglect and abuse, spanning a 6 year period

| Child Welfare Case Files | Child Death Review Database | Public Assistance Eligibility File |
|---|---|---|
| Source of all case information on referrals | 6 years of in-depth data on child fatalities | 7 million records |
| Excellent source of contacts and actions taken | Cause and manner of death | Good history of demographic data |
| 140+ tables from IT system | | Source of social network data |

§.sas | THE POWER TO KNOW.

**THE RESULTS** This government alerts social workers when a child faces imminent risk

## Social workers can act quickly to intervene on behalf of children



- Analyze entire caseload every day
- Alert social workers when at-risk children are identified
- Decision on how to intervene remains in hands of social worker
- Provides a non-intrusive method for evaluating the well-being of a large population of children

§sas | THE POWER TO KNOW.

*PREDICTIVE MODELING CAN…*

# FIGHT TAX FRAUD

**Fraud in the earned income tax credit is rampant, costing hundreds of millions of dollars**



- EITC has 3$^{rd}$ highest error rate among all federal government programs
- $15.6bn lost annually
- Fraud is so pervasive that it could threaten existence of the program

**THE CHALLENGE**: *Can I stop fraudulent tax returns before the refund is paid, without slowing down compliant returns?*

**THE DATA** This government had millions of tax returns submitted over 5 years, with thousands of confirmed fraud cases

| State income tax returns | Audit Selection Results | Federal income tax returns |
|---|---|---|
| 5 years of return data | Selected for audit | 3 years of historical return data |
| Average of 300 data elements captured for each tax return | Audited AND confirmed as non-compliant (including $'s assessed) | Confirmed fraud cases (including $'s assessed) |
| | Audited AND confirmed as compliant | |
| | Not audited | |

**THE RESULTS** | **This government stops tax refunds for suspicious taxpayers before they are paid**

- $50 million saved annually
- False positive rate dropped from 80% to 15%
- Identified 2 data elements that, in combination, are highly predictive of fraud
  - *ZIP code*
  - *Preparer indicator*

§sas | THE POWER TO KNOW.

**This government reversed the tremendous increase in EITC fraud, while getting money quickly to the neediest citizens**



- Analyze tax returns on a nightly basis
- Returns with high risk are automatically denied (no human intervention)
- Returns with medium risk are reviewed by an auditor
- No delay in paying refunds for compliant taxpayers

§sas | THE POWER TO KNOW.

*PREDICTIVE MODELING CAN...*

# STOP OPIOID ADDICTION

§sas
THE POWER TO KNOW®

## THE CHALLENGE

**Addiction to prescription opioids (pain killers) has tripled over a 15 year period**

- Millions of prescriptions filled each year for opioid pain killers
- 4,500+ deaths from overdose in 2014
- Rising costs from prescription drug addiction
  - *Addiction treatment*
  - *Law enforcement*
  - *Courts*
  - *Lost productivity*

**THE CHALLENGE**: *Can I identify patients who misuse prescription opioids before they slip into addiction?*

§sas | THE POWER TO KNOW.

**THE DATA** This government had a wealth of historical data on prescriptions filled and on death records

| Prescriptions Filled | Death Records | Addiction Claims |
|---|---|---|
| 10 years of prescription records | Cause of death | Claims for addiction-related services |
| In-state ONLY | Known reporting challenges | Medicaid claims ONLY |
| No unique identifier for patients | | |

# THE RESULTS

**This government alerts providers when a patient is deemed at high risk of addiction**

# HOW GOVERNMENT WORKS BETTER

**This government gives doctors the power to identify abusive behavior before a patient slides into addiction**

- Analyze new prescriptions every day
- Alert doctors when "risky" patients are identified
- Decision on prescribing remains in the hands of medical professionals
- Encourage patients to seek help <u>before</u> they become addicted

§sas | THE POWER TO KNOW.

# MANAGE RISK FROM LEGAL CLAIMS

§sas.
**THE POWER TO KNOW.**

**THE CHALLENGE**

**A government faces growing legal claims but has (very) limited resources to manage those claims effectively**

- 30,000 legal claims filed annually
- Payouts of over $500 million
- Few in-house attorneys
- Limited budget to hire outside counsel
- 50% of claims are never pursued by the claimant

**THE CHALLENGE**: *At the time a legal claim is filed, should I assign an attorney to handle the claim?*

**The government wanted a model that predicts the final disposition of a claim, achieving 90% or greater accuracy**

## 3 claim dispositions

- Frivolous

- Pre-litigation

- Post-litigation

## Claims data set

Received 23,129 claims

- 9,685 frivolous

- 4,656 pre-litigation

- 8,788 post-litigation

## THE RESULTS

**The predictive model achieved 91% accuracy for post-lit claims, but only 74% accuracy for Friv and Pre-Lit claims**

| TARGET | TRAINING DATA | VALIDATION DATA |
|---|---|---|
| FRIVOLOUS CLAIMS | 76% | 75% |
| POST-LITIGATION CLAIMS | 91% | 91% |
| PRE-LITIGATION CLAIMS | 76% | 74% |

§sas | THE POWER TO KNOW.

# THE RESULTS

**The model identified the five (5) data elements that are most significant in predicting disposition**



SAS Enterprise Miner Report
Selected Variable Importance
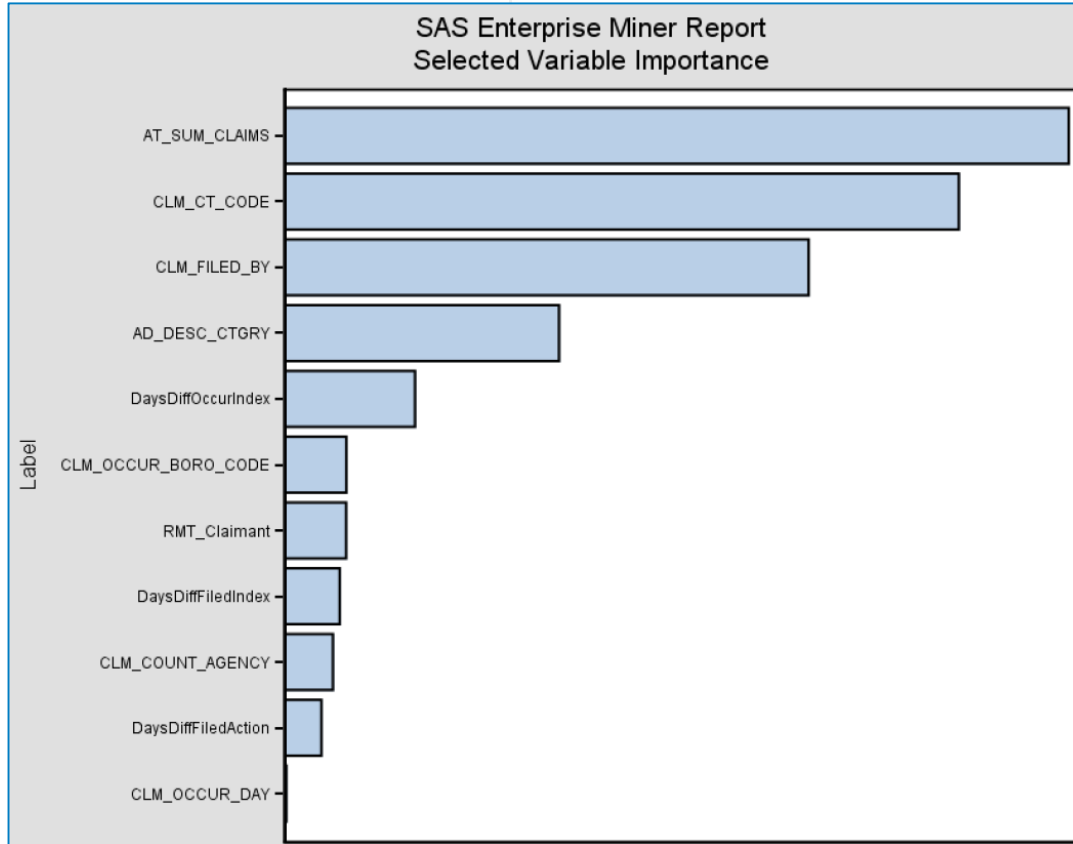
1. Total claims filed by the attorney (AT_SUM_CLAIMS)
2. Claim type code (CLM_CT_CODE)
3. Whether the claim is filed on behalf of the claimant by another party (CLM_FILED_BY)
4. Type of Gov't Agency (AD_DESC_CTGRY)
5. Number of days elapsed between claim occurrence and claim indexing (DaysDiffOccurIndex)

# We produced a decision tree to help understand the business logic identified by the model



| Node Id: | 1 | |
|---|---|---|
| Statistic | Train | Validation |
| FRIVOLOUS CLAIM: | 33.34% | 33.34% |
| POST-LIT CLAIM: | 33.33% | 33.33% |
| PRE-LIT CLAIM: | 33.33% | 33.33% |
| Count: | 13876.0000 | 9253.0000 |

Transformed DaysDiffFiledIndex

01:LOW-7.5 Or Missing

| Node Id: | 2 | |
|---|---|---|
| Statistic | Train | Validation |
| FRIVOLOUS CLAIM: | 39.88% | 39.73% |
| POST-LIT CLAIM: | 20.69% | 20.69% |
| PRE-LIT CLAIM: | 39.43% | 39.58% |
| Count: | 10827.0392 | 7226.4119 |

Transformed CLM_FILED_BY

SELF

| Node Id: | 4 | |
|---|---|---|
| Statistic | Train | Validation |
| FRIVOLOUS CLAIM: | 87.30% | 85.80% |

ATTORNEY, PARENT, _OTHER_ Or Missing

| Node Id: | 5 | |
|---|---|---|
| Statistic | Train | Validation |
| FRIVOLOUS CLAIM: | 30.18% | 30.59% |
| POST-LIT CLAIM: | 24.55% | 24.50% |
| PRE-LIT CLAIM: | 45.27% | 44.91% |
| Count: | 8987.2964 | 6030.3728 |

Transformed AT_SUM_CLAIMS

| Node Id: | 8 | |
|---|---|---|
| Statistic | Train | Validation |
| FRIVOLOUS CLAIM: | 91.71% | 89.52% |
| POST-LIT CLAIM: | 1.94% | 1.60% |
| PRE-LIT CLAIM: | 6.34% | 8.89% |
| Count: | 1670.4286 | 1099.2165 |

03:45-135.5, 04:135.5-266

| Node Id: | 11 | |
|---|---|---|
| Statistic | Train | Va |
| FRIVOLOUS CLAIM: | 16.24% | |
| POST-LIT CLAIM: | 20.43% | |
| PRE-LIT CLAIM: | 63.33% | |
| Count: | 3176.9116 | 2 |

# HOW GOVERNMENT WORKS BETTER

**This government quickly assigns legal claims to mitigate payout amounts**
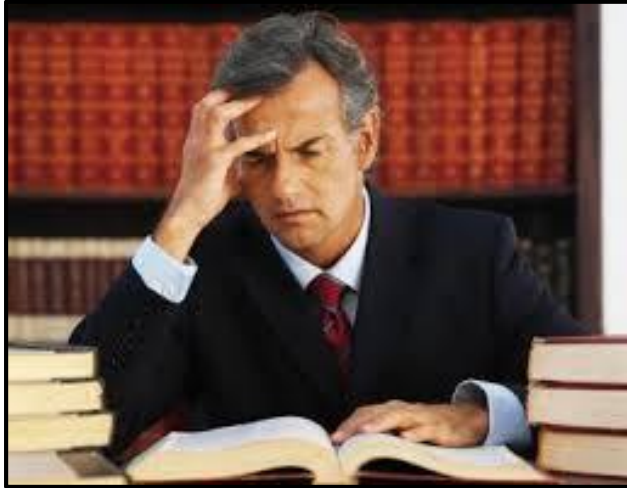


- Each claim is analyzed within minutes
- Claims are routed immediately for assignment
- Government saves money by knowing which claims to settle before trial

§sas | THE POWER TO KNOW.

*PREDICTIVE MODELING CAN…*

# TRANSFORM PENNSYLVANIA

Government leaders in Pennsylvania can use predictive modeling to transform how services are delivered

**PA**

**pennsylvania**

- **TRANSPORTATION**: Infrastructure maintenance and replacement
  - *Roads, bridges, sidewalks, water, sewer*
- **EDUCATION**: High school dropout risk
- **CORRECTIONS**: Recidivism risk
- **LICENSING/REGULATION**:  Inspections
  - *Restaurants, nursing homes, building sites*

# WHAT ARE YOUR IDEAS?

§sas | THE POWER TO KNOW.

**Please reach out if you would like to brainstorm on additional ways to use predictive modeling!**

**Shaun Barry**

Shaun.Barry@sas.com

(919) 531-9594

**Ed Haines**

Ed.Haines@sas.com

(717) 307-8186