

SAS User Group

What is normal behaviour?

July 2007

Introduction

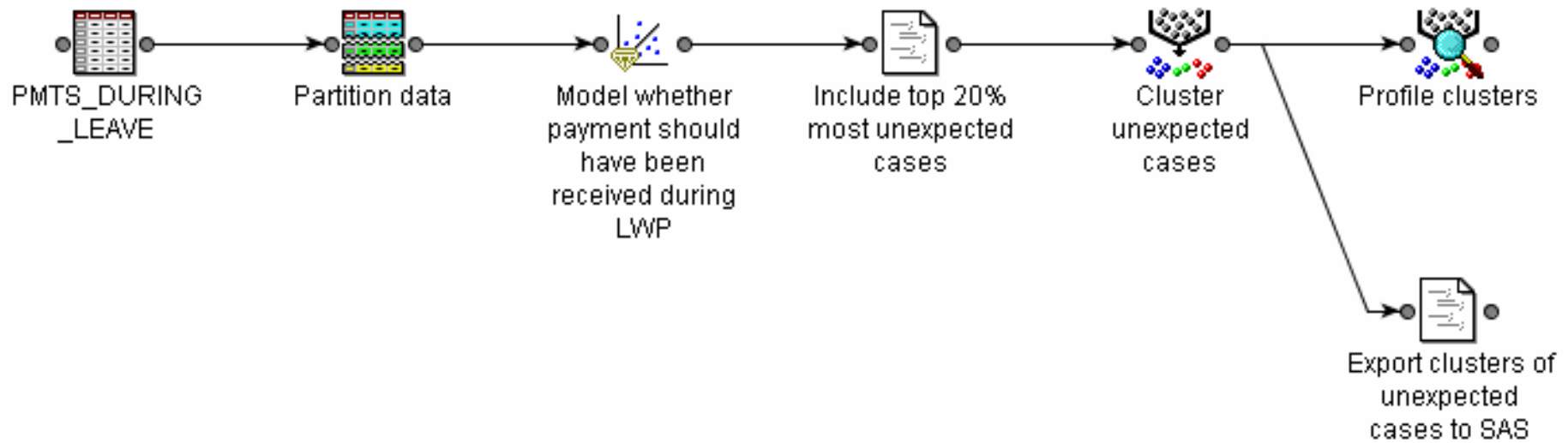
- A leading Australian financial institution wanted to understand how data mining techniques and SAS Enterprise Miner could strengthen their existing audit function
- In conjunction with the institution, we used data mining techniques to investigate several areas of the payroll function identified as potentially risky
- We use a “model of normality” approach that works without pre-conceived notions of suspicious behaviour or detailed knowledge of business rules
- Some of the processes we examined:
 - Payments made to staff while on unpaid leave
 - Employee exit payment calculations

Data

We used the following data for the analysis, on an anonymous basis:

- All extended leave taken for the past 3 years
- All salary payments for the past 3 years
- All employee exits for the past 3 years
- Exit payment breakdowns for a small sample of employees
- Leave balances on exit for a small sample of employees
- Audit log of record updates made by Payroll staff for the past 3 years
- Employment histories of all employees

Payments on unpaid leave: Approach



Payments on unpaid leave: Results

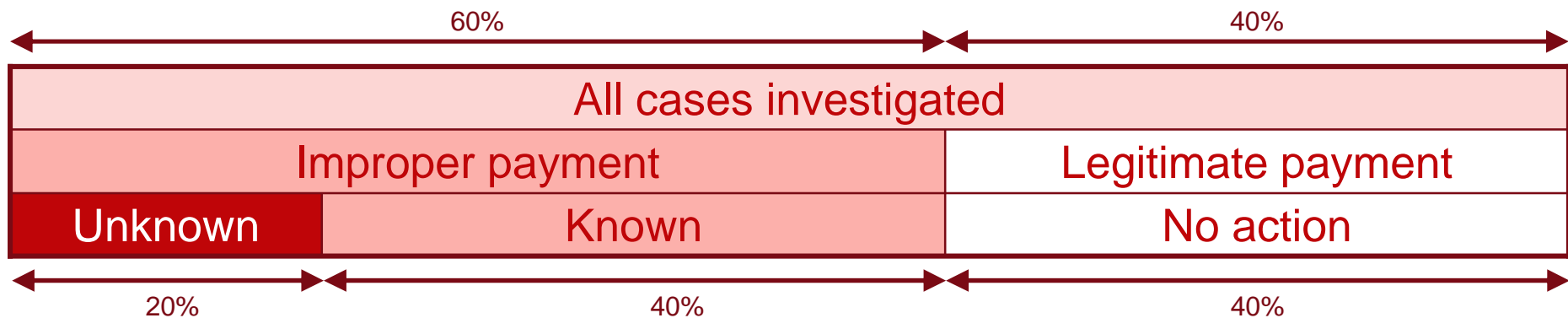
- The risk model can predict whether a given leave period should include a payment or not with 80% accuracy
- Clustering of the riskiest cases identifies four groups:

Segment 1	Highest-value cases, mostly categorised by repeated payments of similar amounts.
Segment 2	Some high-value cases. The largest payout is always in a certain month – probably the annual bonus payout – but 30% of cases have a second payment of a lesser sum.
Segment 3	Low to medium value cases, consisting of 1 to 3 payments. The biggest payment is in September, October, or November
Segment 4	Low to medium value cases, consisting of a few payments (mostly 1-3, up to 7) sometimes over a period of up to six months.

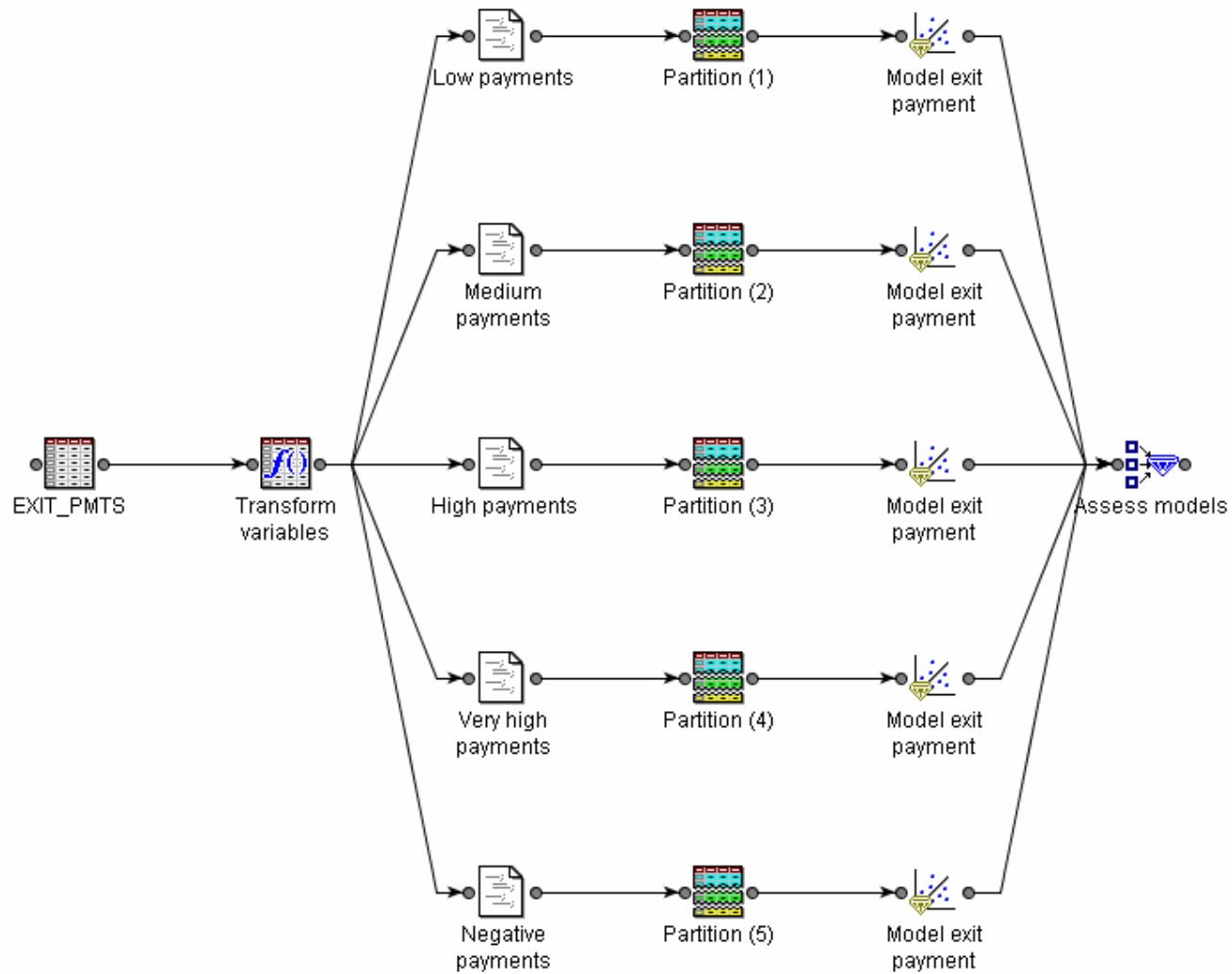
Payments on unpaid leave: Results (cont)

We took some high-risk cases to Payroll for a pilot investigation

- 60% of cases investigated were found to involve a payment in error
- Of those, in 65% of cases the overpayment was known to Payroll, and records showed that the money was usually recovered
- ... but in 35% of cases, the overpayment was unknown to Payroll



Exit payments: Approach



Exit payments: Results

The predictive accuracy of the models is sufficient for case selection, being generally around 50% of the group midpoint.

Group	Range	Midpoint	Average error in forecast
1	Low	\$1,000	\$300
2	Medium	\$2,000	\$800
3	High	\$20,000	\$10,000
4	Very high	\$750,000	\$70,000
Negative		\$-3,000	\$1,400

The next step is to select cases where the forecast and actual payments differ significantly and pass them to Payroll for investigation. This allows formation of a business case, as for payments while on unpaid leave.

Conclusions

The results of the project demonstrate the benefits of data mining in SAS Enterprise Miner:

- Data mining techniques identify improper activities with much higher precision than a traditional approach
- Data mining techniques also uncover patterns and trends in behaviour, that can be used to distinguish normal and acceptable activity from abnormal and improper activity, without prior knowledge of underlying business rules
- The capabilities of data mining can be validated externally, allowing estimation of the monetary benefits and development of a business case

Conclusions

“PwC's application of these techniques showed real value in both identifying issues we needed to address, and validating our existing controls. We are now looking to incorporate these tools and techniques into our audit and review functions.”

-- Manager, major Australian financial institution