

Effective deployment and use of predictive models

John McConnell – Professional Services

Rachel Clinton – Business Development

www.sv-europe.com

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email a PDF copy to you after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat facility if we run out of time we will follow up with you.





Predictive Analytics for Smarter Business



- Premium, accredited partner to IBM specialising in the SPSS Advanced Analytics suite.
- SAS Business Partner
- Expertise in open source analytical technologies & big data
- Team each has 15 to 20 years of experience working in the predictive analytic space



Today's Objectives

- Share some of our experiences and understanding about model deployment
- Discuss the broad options that exist



A proliferation of Data Science* tools

Figure 1. Magic Quadrant for Data Science Platforms



Some of these vendors offer open source/paid e.g. **RapidMiner**

Some are commercial vendors building on open source e.g. **Revolution** (now Microsoft) with **R**.

In addition we have the true "roll-your-own" open source: Chiefly **R** and **Python** but increasingly **SparkML**, **MS AzureML**, **Amazon ML**, etc.

*was "Advanced Analytics" pre 2017



The CRISP-DM process



Europe

When selecting among the tools we all tend to think about:

- Usability (mapped to available skills)
- Productivity
- Data Access and Preparation
- Analytical Range
- Price

The CRISP-DM process



6.Deployment (1)

- Could be as simple as a list of names and predictions/scores
 - E.g. a mailing list
- Could be as complex as a model encapsulated as a computer program and embedded in an operational system to predict in real time and automate decisions
 - E.g. a model embedded in a system which sends alerts and triggers
- Could be embedded in a **What-if?** simulator
- Important to distinguish between a model in the modelling and deployment phases
- Typically...
 - In the modelling phase many different models and modelling options are built and evaluated
 - In the deployment phase the winning model(s) are fixed
 - E.g. we deploy a decision tree with a fixed shape



An example modelling data window

SM

Eur

o p

a) Use Data we have on the customer to the <u>time before</u> the last period (e.g. month)



An example scoring (deployment) data window

SM

E **U** . г. a) Use Data up to the current point in time



6.Deployment (2) (Monitoring)

- If we did our job properly then the deployed model should correspond to what we saw in evaluation
 - Other factors may intervene
- Ongoing evaluation ("monitoring") still needs to happen if models are to be used over time
 - Some models have a longer shelf life than others
- More recently there has been some development of models which adapt/correct themselves to changing circumstances
 - Some level of re-modelling to improve accuracy
 - "Self adapting"
 - More commonly this is achieved through the concept of champion/challenger modelling or model refresh approaches



IBM's new ASUM-DM methodology extends CRISP-DM



http://datascience.ibm.com/blog/an-ode-to-the-analytics-grease-monkeys-analytics-deployment-roi/



Option 1 – Recode the model into something else



Simple models can be coded into other software

Call: lm(formula = CONSUME ~ PRICE + INC + TEMP + PRICEINCi, data = datavar) Residuals: Min Median 10 30 Max -0.0575279 -0.0163589 -0.0008483 0.0168662 0.0718922 Coefficients: Estimate Std. Error t value Pr(>ItI) (Intercept) 0.1570203 0.2324673 0.675 0.5058 PRICE -0.1636906 0.7438870 -0.220 0.8277 0.0012301 0.0012133 1.014 INC 0.3208 0.0028231 0.0004171 6.769 5.31e-07 *** TEMP -0.2786003 0.1344397 -2.072 PRICEINCi 0.0491 * Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03094 on 24 degrees of freedom Multiple R-squared: 0.7411, Adjusted R-squared: 0.698 F-statistic: 17.18 on 4 and 24 DF, p-value: 8.968e-07

- Linear Regression are a good example
- They fit easily into Excel formulae

Consume = 0.1570203 -0.1636906*PRICE + 0.0012301*INC + 0.0028231*TEMP -0.278600*PRICEINCi



Simple Decision Trees/Rule induction models

If HANDSET = ASAD170 or WC95 CHURNPROPENSITY=4.6%



IF HANDSET = BS210 and TARIFF = CAT100 or CAT50 CHURNPROPENSITY=93.8%



Recoding

- Introduces a time delay
- Can potentially introduce errors
- Becomes less appealing as models become more complex...



Option 2 – Export the model into a code format that can be plugged into other software



Exporting SQL is one option (sometimes)



<u>S</u> ave Node	
😻 Store Node	
Header and Footer	
Page Setup	
Print Previe <u>w</u>	•
<u>P</u> rint	•
Export Text	•
Export HTML	•
Export PMML	
Export <u>S</u> QL	
Publish for Server Scoring Adapter	



- Some tools support this; Python and R don't (to our current knowledge ... new packages are added all the time!)
- Caution around the differences in SQL between databases e.g. Oracle and IBM/DB2
- Some tools provide target specific SQL which can sometimes be published into the database

PMML is the standard export format



www.dmg.org

Predictive Model Markup Language

PMML 4.2 - General Structure

PMML uses XML to represent mining models. The structure of the models is described by an XML Schema. One or more mining models can be contained in a PMML document. A PMML document is an XML document with a root element of type PMML. The general structure of a PMML

document is:



The good news about PMML



http://zementis.com/resources/zementis-pmml-tools/

Most analytical tools – open or commercial – either export by default or 3rd party tools - in this case Zementis Py2PMML – plug the gap

R has its own PMML package https://cran.r-project.org/web/packages/pmml/index.html



The less good

Company / Project	Software	PMML Producer	PMML Consumer	Supported Model Type
ALPINE DATA	Alpine	PMML 4.1		Regression Models (Linear and Logistic) Naïve Bayes K-Means Alpine Forest SVM
Angoss	KnowledgeSTUDIO	PMML 3.2		Decision Trees Regression Models (Linear and Logistic) Neural Networks Clustering Models Rule Set Models (Scorecards)
	KnowledgeSEEKER	PMML 3.2		Decision Trees
	StrategyBUILDER	PMML 3.2		Decision Trees (Strategy Trees)
bigm	BigML Public API	PMML 4.1		Decision Trees (classification and regression)
	Strategy Tree Optimization	PMML 3.0, 3.1	PMML 3.0, 3.1	Decision Tree
	PowerCurve™ Strategy Management	PMML 3.0, 3.1, 4.0, 4.1	PMML 3.0, 3.1, 4.0, 4.1	Decision Tree
			PMML 3.0, 3.1, 4.0, 4.1	Regression Model

continues...

Coverage is not comprehensive ... particularly on the **Consumer** (deployment) side e.g. not all databases support PMML and/or the latest versions

Full table at: <u>http://dmg.org/pmml/products.html</u>



Zementis* has the most PMML support



Zementis ADAPA runs models in the cloud (with an Excel Add-in option)

Amazon Cloud, Microsoft Cloud, etc.

Or on-premise

*As of December 2016 Zementis is part of Software AG (<u>http://bit.ly/2yFfn7v</u>)



C	2)6	a 17 · (* ·	9.	Dat	laSets_Excel200	7.xisx [Read-	Only] - M	icrosoft Excel		
	2	Home In	sert Page Lay	out For	nulas Dat	a Review	vier Vier	w Predict	with AD	APA
C0	Setup nnecti icore T	Apply on Model our Data	Support About Blog Zementis Information							
_		J1	• (*	fx Adjust	ted					
	A	В	C	D	E	F	G	H	1	J
1	Age	Employmen	t Education	Marital	Occupation	Income	Sex	Deductions	Hours	Adjusted
2	38	Private	College	Unmarried	Service	81838	Female	0	72	0
3	35	Private	Associate	Absent	Transport	72099	Male	0	30	0
4	32	Private	HSgrad	Divorced	Clerical	154676.7	Male	0	40	0
5	45	Private	Bachelor	Married	Repair	27743.82	Male	0	55	1
6	60	Private	College	Married	Executive	7568.23	Male	0	40	1
7	74	Private	HSgrad	Married	Service	33144.4	Male	0	30	0
8	43	Private	Bachelor	Married	Executive	43391.17	Male	0	50	1
9	35	Private	Yr12	Married	Machinist	59906.65	Male	0	40	0
10	25	Private	Associate	Divorced	Clerical	126888.9	Female	0	40	0
11	22	Private	HSgrad	Absent	Sales	52466.49	Female	0	37	0
12	48	Private	College	Divorced	Service	291416.1	Female	0	35	0
13	60	Private	Vocational	Widowed	Clerical	24155.31	Male	0	40	0
14	21	Private	College	Absent	Service	143254.9	Female	0	35	0
15	21	Private	College	Absent	Machinist	120554.8	Male	0	40	0
16	50	Private	Master	Married	Executive	34919.16	Male	0	40	1
17	37	Private	HSgrad	Divorced	Executive	67176.79	Male	0	35	0
18	30	Consultant	HSgrad	Divorced	Repair	9608.48	Male	0	40	0
19	32	Private	HSgrad	Married	Machinist	12475.84	Male	0	40	0
20	65	SelfEmp	College	Married	Sales	32963.39	Male	0	40	0
21	28	Private	College	Married	Executive	31534 97	Male	0	55	0

So what are we missing?

- Until now we've only looked at how we can deploy the model itself
- Often modelling jobs have **data access/preparation/blending steps**
- We can recode these separately into a database using SQL
- But what if we want to automate this too?
- Some tools e.g. IBM/SPSS Solutions Publisher, allow us to export a complete job as an executable or as a high level programming language source like C or Java
- But more typically we need to consider the following options...



Option 3 – Using the same engine to deploy



The basic idea

- A. Start with the modelling "job"
 - You may need to adapt it to create a deployment/scoring version
 - Usually a simplified version
- B. Install the software on a deployment machine; R, Python, SAS, SPSS, WPS (SAS emulator), etc.
 - Or use the same machine
- C. Periodically run new data tł operational target





Which means we can usually schedule execution

		Create Basic Task Wizar	d		×
Summary					
Create a Basic Task					
Trigger	Name:	Test 88			
One Time	Description				
Action	Campuon				
Start a Program					
inish					
	Trigger:	One time; At 08:46 on 14/10/	2015		
	Trigger: Action:	One time; At 08:46 on 14/10/ Start a program; C:\Train\CF	2015 UK_Batch\BB.bat		
	Trigger: Action:	One time; At 08:46 on 14/10/ Start a program; C:\Train\CF	2015 UK_Batch\BB.bat		
	Trigger: Action: Open the	One time; At 08:46 on 14/10, Start a program; C:\Train\CF Properties dialog for this task	2015 UK_Batch\BB.bat when I click Finish	145	
	Trigger: Action: ☐ Open the When you cli	One time; At 08:46 on 14/10 Start a program; C:\Train\CF Properties dialog for this task ick Finish, the new task will be	2015 UK_Batch\BB.bat when I click Finish created and added to	o your Window:	s schedule.

As long as the analytical tool/engine allows **batch execution** (usually via the command line) we can use standard scheduling tools e.g. Windows Task Scheduler to run at a given time/periodically



Option 4 – Use a deployment platform



Platforms typically include

Ease of deployment	 Allow analysts/data scientists to easily deploy models and full jobs Portals installed by IT that give analyst access to IT ops
Support for multiple engines	•R, Python, SPSS, SAS, etc.
Monitor/report	 Continually model monitor performance in the field Link to BI/Visualization tools for MI reporting, Predictive KPIs, Forecasts, etc.
Alert	Notify when jobs fail or succeed Notify when model accuracy falls below accepted thresholds
Automated model refresh	•Champion/challenger •Adaptive models
Support real time	•As well as triggered – event-based scoring – e.g. trigger from call centre
Interface to operational systems	•e.g. ERP, eCommerce, CRM, Call Centre, HR, etc.
IT lifecycle/workflow support	•Development -> UAT -> Production
Support Decision Management	



SAS has Model Manager ... IBM/SPSS has C&DS*





And on into ...

Decision Management/Simulation/Optimisation

- Where business users can blend models and business rules to automate decisions
- A separate topic all of it's own
- James Taylor's site is peerless on this topic: <u>http://jtonedm.com/</u>

		e				Version: 3
ocabularies <	e	Rule Name	Operator	Term	Value	Rule Details
Enter search text	۳	Initialize				and an an
利用市のの			IF	Drag vocabular		Rule name:
			THEN	Action	"No Action"	Initialize
Telco	v	Unknown Phone Expired Contract				Record rule-fired data
Expired_Contr			IF	DEVICE_MANU	'Unknown'	Name:
Recent_Overa			AND	Contract_Tenure	<0	Detractor_offer_rules
Recent_Supp			THEN	Action	*Discount Phone	Rule folder name:
II Unknown_Dev ► ☆ Customer ► ☆ DecisionBuilderC	v	Overages	ELSE			Telco
			IF	ttl_overage_ch	>50	Rule set identifier:
	: •		THEN	Action	Retroactive New	10001
		Soon To Expire Contract	ELSE			Rule order: 1
			IF	Contract_Tenure	<5 and > 0	Description:
			THEN	Action	"New Contract Pr	
						Created by:
						SAS Demo User
						Created on:
						him 20, 2015 09:20 AM
						0011 00, 2010 00.00 Mil

SAS Decision Manager



Summary

- When selecting modelling tools it is important to think about the full analytical lifecycle
 - And how will you deploy?
- The **platforms** offer the most productivity and flexibility
 - But at a cost
 - We often find that the other options, e.g. batch mode, suffice as you scale up





Working with Smart Vision Europe Ltd.

- We can assist with every aspect of almost any analytical project
 - Selection, implementation & support of analytical tools
 - Guided consulting & training to develop in house skills
 - Identification & recruitment of analytical skills into your organisation
- As experts in Advanced Analytics / Predictive Analytics we will
 - deliver classroom training courses
 - offer side by side training support
 - offer "skills transfer" consulting
 - run booster and refresher sessions to get more from your existing technology
 - offer 'no strings attached' technical and business advice relating to analytical activities





Contact us:

+44 (0)207 786 3568 info@sv-europe.com Twitter: @sveurope Follow us on Linked In Sign up for our Newsletter

Thank you

www.sv-europe.com