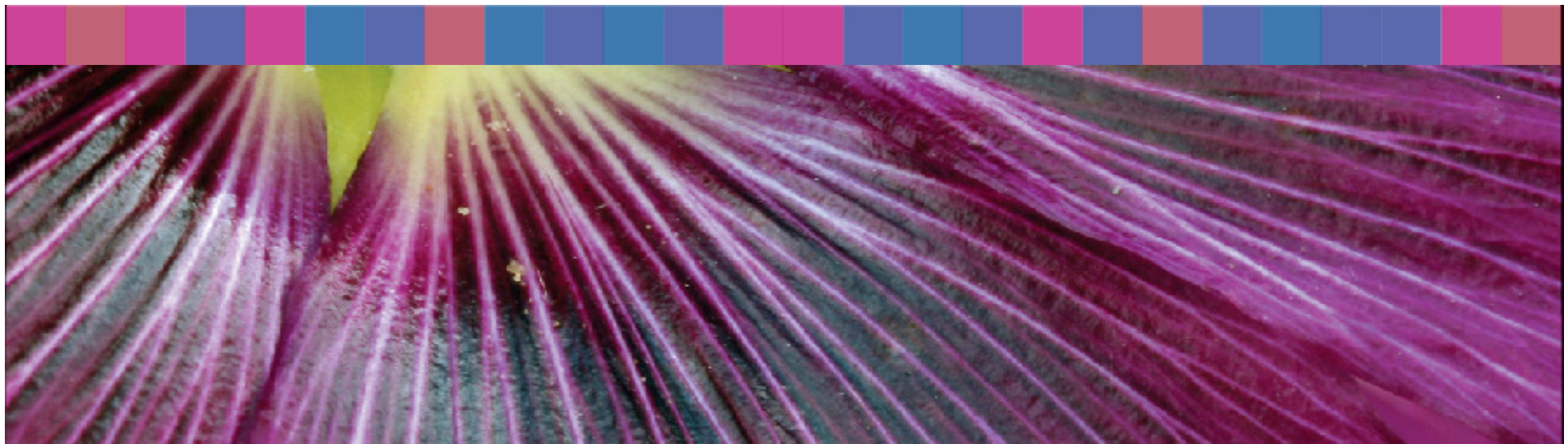




Sussex Health Informatics Service

Supporting healthcare with technology



Creating a risk prediction model

Colin Styles, Information Architect

Alison Smyth, Analytics Manager

How do data mining tools work?

Data mining tools work by looking at what has happened in the past, and identifying the important factors, or 'variables'

Once these factors are identified the model can be used to predict future events

- for example:

Woolly hat sales

Date	Forecast	Temp	Hat Sales
Monday	Sunny	16	12
Tuesday	Dull	6	30
Wednesday	Rain	2	15
Thursday	Sunny	12	25
Friday	Rain	16	5
Saturday	Rain	8	?????



Example - observations

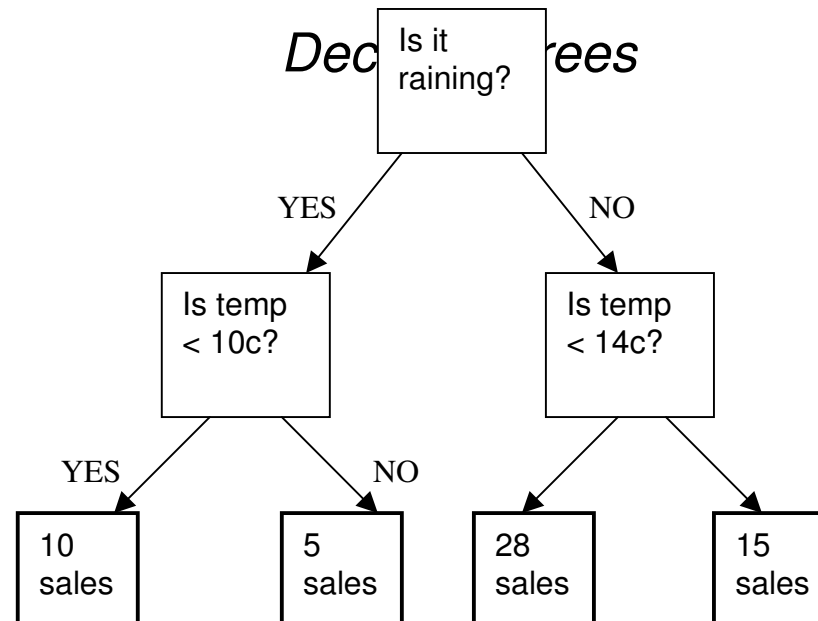
Woolly hats sell more on colder days

Whether it is sunny or dull makes no difference to sales

However, sales are much lower on rainy days

Bearing in mind these factors, we might predict to sell 10 hats on a cold, rainy day

Some data mining models



Simple to create and to understand, but not the most accurate

Our worked example is based on a decision tree model

Some data mining models

Nearest neighbour

Identifies which category a result is most likely to fall into

– eg ‘high sales’

Useful for grouping data, but does not give a specific answer as to how many sales

Some data mining models

Regression

Represents the prediction as a mathematical function of the input variable(s)

eg $\hat{\text{sales}} = 20 - \text{temp}/2$

Good for 'continuous' variables – eg temperature

Can be used alongside decision trees to give a more continuous spread of results

Some data mining models

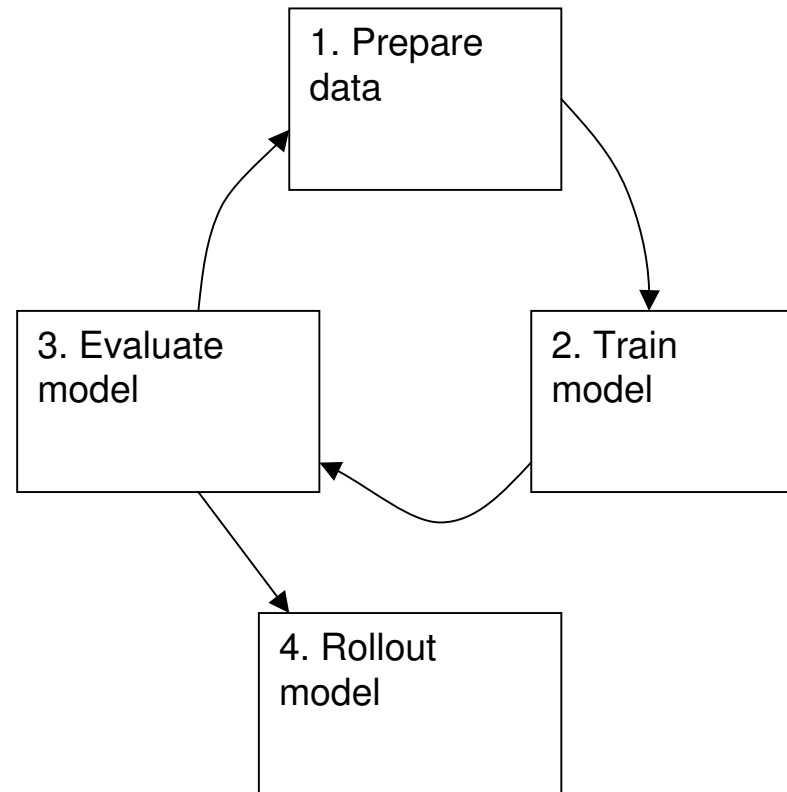
Neural network

Neural network mining models work by weighting the input variables to derive the best predictive result

They often give the best results, but it is harder to understand how the result was derived

We suggest building with decision trees initially, and then trying neural networks to see if the result improves

Four steps to creating your data mining model



Preparing data

Prior to data mining, you will need to get you data into a simple format that the data mining model can understand

This will typically involve:

- **Linking patient event records via a common patient identifier**
- **Aggregating patient events – eg count of outpatient attendances**
- **Also some data cleansing may be required (eg a future date of birth could produce some interesting predictions!)**

Training your data mining model

Demonstration using
SQL Server Analysis Services

Testing your results

There are two ways to test your results –

- Intuitive
- Statistical

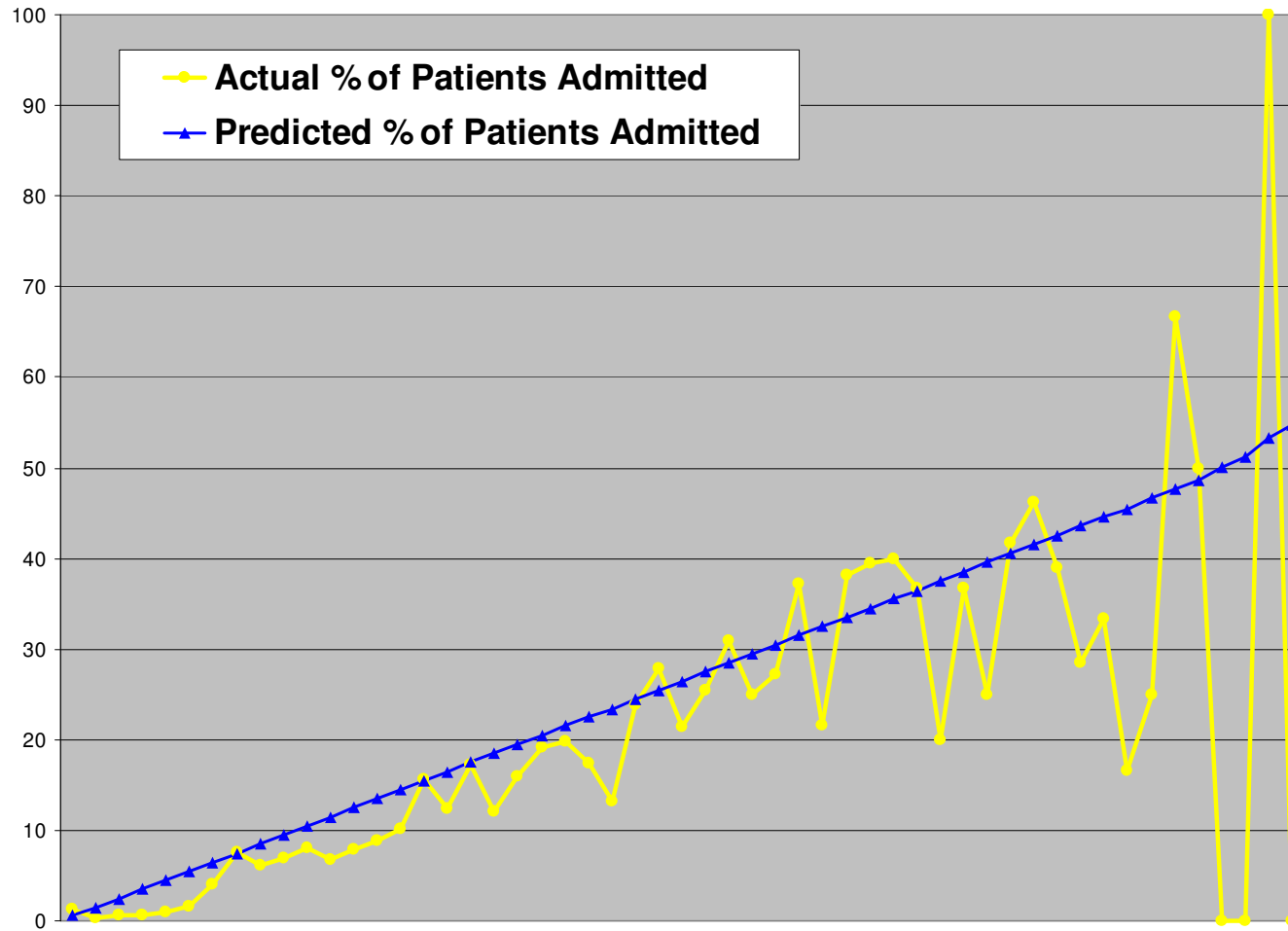
Intuitive is about looking at the results to see whether they feel right...

Do predicted results match your expectations?

Are the variables in the decision tree sensible?

Statistical evaluation is about using formal mathematical techniques eg..

Correlation



T-test

		Predicted		
		High risk	Not high risk	Total
Actual	Admitted	100	250	350
	Not Admitted	50	19600	19650
	Total	150	19850	20000

Positive Predictive Value = $100/150 = 67\%$

Iterative development

It is unlikely that your first attempt will give you the optimum model

You can dramatically change the model by changing the input variables and the type of data mining model used

Additional data sets are also likely to improve the model if available

Roll-out of the model

Once you are happy with your model it can be used to predict future events in exactly the same way as we predicted for the testing data

You need to decide how to provide the results to users
eg

- Excel spreadsheet
- Within existing system (eg PAS)
- Via bespoke tool eg web application

The End

Thank you for your attention