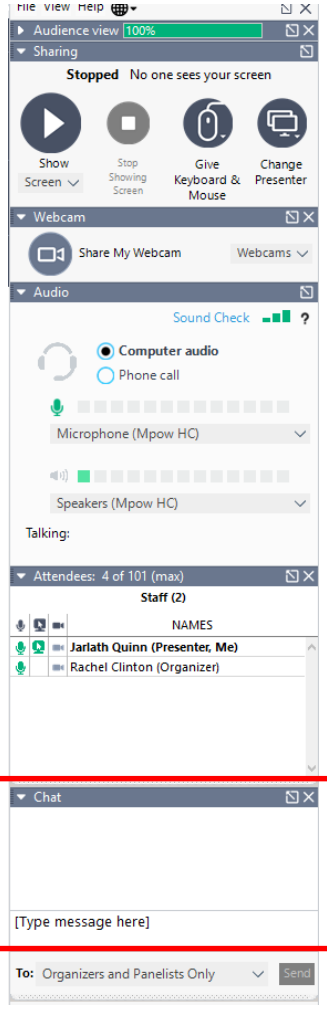
SMARTVISION Europe

Improving Predictive Models with IBM SPSS Modeler

Jarlath Quinn

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.



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

- Premier accredited partner to IBM and Predictive Solutions specialising in advanced analytics & big data technologies
- Work with open source technologies (R, Python, Spark etc.)
- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
  - Retail
  - Gaming
  - Utilities
  - Insurance
  - Telecommunications
  - Media
  - FMCG

Agenda

- Bootstrap aggregation
- Boosted models
- Feature engineering
- Ensemble models
- Meta-modelling
- Split-method models

Method 1:

Bootstrap Aggregation (Bagging)

Bootstrap Aggregation or 'Bagging'

- If I build a single predictive model on a given sample of data, how likely is it that I would get the same results using a slightly different sample?
- In fact, what if I took a random sub-sample comprised of 95% of the data and built a predictive model and then compared that model to another one built using a different random sub-sample based also based on 95% of the data?
- The fact is that models vary from one sample to another. So which one is best?
- Answer: All of them

Bootstrap Aggregation or 'Bagging'

- Bootstrap aggregation, also called bagging, is a random ensemble method designed to increase the stability and accuracy of models.
- It involves creating a series of models from the same training data set by randomly sampling with replacement the data. Sampling with replacement means that a specific row of data may appear more than once in the subsequent random sample.
- This means that each resultant model is trained against a slightly different sample of data. The resultant predictions from the multiple models are then all combined to create a single score

Method 2: Boosting

Boosting

- Why does a model accurately predict outcomes with some records in the dataset but not others?
- Is it simply random? Or are certain sub-sets of data harder to predict with a generic ‘one-size-fits-all’ model?
- What if we could build a predictive model that paid more attention to the parts of the dataset where it is least accurate?

Boosting

- Boosting is another ensemble model-building method that was designed to help develop strong classification models from weak classifiers
- Boosting methods focus on *error* (or misclassifications) that occur in prediction. After an initial model is built, the Boosting algorithm applies a series of weights to the data so that cases that were inaccurately predicted are given larger values and those that were accurately predicted smaller values.
- The classification algorithm is then re-applied to the data, but this time greater emphasis is given to correctly predicting the previously misclassified cases (i.e. those with the larger weights).
- The idea is that by repeatedly applying this approach, the algorithm attempts to hunt down the harder to classify cases.

Method 3: Feature Engineering

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Feature Engineering

- Rather than trying to find the best technique or the optimal parameters for a predictive model, perhaps the more sensible approach would be to create new structures or ‘features’ in the data to help the technique accurately predict the outcome in question.

Feature Engineering

- Re-scaling predictor fields
- Replacing missing values
- Excluding outliers and extreme values
- Creating new fields based on the ratio of one variable to another
- Using Factor Analysis/PCA to create new linear combinations of existing correlated variables
- Using Cluster Analysis to create groups in the data based on the similarity of cases

Method 4: Ensemble Models

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Ensemble Models

- Trying to find the ultimate modelling technique can be frustrating.
- You might find that no single method performs well across *all* the subgroups of the data.
- How about combining the predictions of different methods?
- You could predict outcomes based on the model with the highest confidence score, or just using the average probability from different models or perhaps calculate a weighted score.

Method 5: Meta-Models

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Meta Models

- What if you used the predictions from one model as an input variable for another predictive model?
- By adding the predictions generated by an initial modelling technique to an existing pool of predictor field, a second technique can then exploit these predictions to build a final, hopefully more accurate model.

Method 6: Split Models

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Split Models

- Split models or split population modelling is another technique that allows the user to build multiple models which can then be combined to create a single prediction.
- The idea with split modelling is that if the data represent different populations or contain separate groups that behave in very different ways, assuming that a single model can explain all the inherent variability across these distinct populations might be unreasonable.
- In which case, why not build separate local models for these key segments in the data and aggregate the resultant scores with the aim of increasing overall accuracy.

Smart Vision's Online Training Resources

To access many of these courses for free or at a greatly reduced price, use offer code: webinar100

Introduction to IBM SPSS Statistics course

£124.00

BUY THIS

Understanding and applying logistic regression techniques in SPSS Statistics

£75.00

BUY THIS

Understanding and Applying Linear Regression Techniques in SPSS Statistics

£75.00

BUY THIS

Building predictive models in SPSS Modeler

£75.00

BUY THIS

Statistical and significance testing in SPSS Statistics

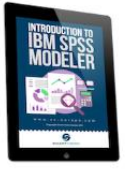
£75.00

BUY THIS

Working with decision trees in SPSS Statistics

£75.00

BUY THIS



Introduction to SPSS Modeler course

£124.00

BUY THIS

https://www.sv-europe.com/smart-vision-spss-courses/introduction-ibm-spss-statistics-course/

https://www.sv-europe.com/smart-vision-spss-courses/statistical-significance-testing-spss-statistics/

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Working with Smart Vision Europe Ltd.

- **Sourcing Software**
  - You can buy your analytical software from us often with discounts
  - Assist with selection, pilot, implementation & support of analytical tools
  - http://www.sv-europe.com/buy-spss-online/
- **Training and Consulting Services**
  - Guided consulting & training to develop in house skills
  - Delivery of classroom training courses / side by side training support
  - Identification & recruitment of analytical skills into your organisation
- **Advice and Support**
  - offer 'no strings attached' technical and business advice relating to analytical activities
  - Technical support services around SPSS

SMARTVISION Europe

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