

## DIRECTED RISK RESEARCH PROBLEM STATEMENT

<b>Risk Theme</b>	Credit Risk	<b>Problem Nr.</b>	PS15011		
<b>Client Name</b>	Wickes Robbertse	<b>Client Org.</b>	BAG		
<b>Designation</b>	Senior Risk Analyst   Model Validation				
<b>E-mail</b>	<a href="mailto:Wickes.Robbertse@absa.co.za">Wickes.Robbertse@absa.co.za</a>	<b>Tel (w)</b>	011-350-6044	<b>Mobile</b>	082-357-9712

**PROJECT TITLE:** Semi-supervised segmentation within a predictive modelling context in retail credit

### PROJECT GOAL:

The goal of the project is to propose a semi-supervised segmentation method and to do empirical studies on the performance of this method.

### HIGH LEVEL DESCRIPTION OF PROBLEM

Segmentation of predictive models has been well-established in the industry as part of the model development process. Segmentation is done with the aim of improving overall predictive performance, and is required when the characteristics available to model the outcome differ in some way between segments. Current segmentation techniques can be split into two main streams: The first stream is based on maximising target separation or impurity between segments through supervised classification, most commonly decision trees. The second stream defines segments by maximising the dissimilarity of the character distribution based on a distance function (for example Euclidean distance in cluster analysis).

Both streams have been successful in improving model accuracy, and depending on the application have outperformed each other in different environments; however they consider only one of two aspects, the dependent (target) variable, or the distribution of the independent variables to be used for modelling.

We require research into methodologies which balances both the use of the target variable as well as the distribution of the independent variables during segmentation. Such techniques are referred to as semi-supervised classification techniques.

### PROJECT OBJECTIVES

We require research on different segmentation techniques. Current techniques need to be compared. We also required research into the possibility of new techniques. The first objective is therefore to develop a new semi-supervised segmentation technique. The second objective will be to conduct an empirical study to compare the performance of current techniques as well as the proposed new technique. A third objective is to mathematically formulate this problem (semi-supervised segmentation within an predictive model context) as an explicit optimization problem and the fourth objective will be to empirically test how the optimization type of problem perform with other techniques.

### OUTPUTS REQUIRED

The outputs will be:

- Literature study of current semi-supervised classification techniques
- Develop a new semi-supervised classification technique
- Empirical studies to test the performance of this new technique.

### **STRATEGIC VALUE TO DIRECTED RISK RESEARCH**

This research will increase insight into segmentation techniques as well as a general improvement in retail credit risk modelling methodology.