DIRECTED RISK RESEARCH PROBLEM STATEMENT

Risk Theme	Credit Risk			Problem Nr.	PS15010

Client Name	Dries de Wet			Client Org.	XDS
Designation	Head: Analytics				
E-mail	ddewet@xds.co.za	Tel (w)	011 645 9100	Mobile	083 412 6015

PROJECT TITLE: Research in predictive modelling: Binning, Variable selection, Income modelling

PROJECT GOAL:

The goal of the project is to propose new methodologies (and compare existing methodologies) in the predictive modelling context, with a specific focus on:

- **Binning:**
  Binning (or discretization) concerns the process of transferring continuous variables into discrete counterparts.
- **Variable selection:**
  The problem of selecting subsets of variables, in regression or multivariate statistics, that contain most of the relevant information in the full data set.
- **Income modelling:**
  Income modelling is the process of building a predictive model to estimate the gross monthly income of individuals.

HIGH LEVEL DESCRIPTION OF PROBLEM

Predictive modeling is a process used in predictive analytics to create a statistical model of future behavior. Predictive models are widely used as analytical tools in retail credit.

Discretization/binning of variables have been well-established in the industry as part of the predictive modelling process. Some advantages of binning include: the scorecard format use binned variables; variables is more easily interpreted; and non-linear dependencies can be modeled using a linear relationship. Unfortunately binning is a time-consuming step and more effective and more automated methods for binning are required.

Another important phase in the predictive modelling process is the selection the subsets of variables that contain most of the relevant information in the full data set. We require that existing variable selection methodologies are compared and that new variable selection methodologies are developed.

In the field of retail credit, most predictive models involve extending credit to consumers. The National Credit Act requires that credit providers must properly test affordability before extending credit. An income model can be used to accurately predict the gross monthly income of individuals in order to support these affordability requirements. Various methods to model income need to be compared.

PROJECT OBJECTIVES

The following are important aspects in predictive modelling that requires further research:

- **Binning:** More effective and more automated ways to bin variables need to be researched.
- **Variable selection:** Several variable selection techniques need to be compared including stepwise regression, factor analysis, variable clustering, partial least squares and other state-of-the-art techniques.
- **Income modelling:** Investigate various research methodologies that can effectively be used in income modelling.

OUTPUTS REQUIRED

The output required from this research study is a formal report which discusses different Binning, Variable Selection, and Income Modelling techniques applied the retail credit environment. The report should highlight:

- current techniques available in Binning, Variable Selection, and Income Modelling (with application in retail credit), with their limitations and strengths
- compare different techniques in the retail credit environment, with recommendations as to best-of-breed from a business perspective
- Identify promising areas for further/future research

STRATEGIC VALUE TO DIRECTED RISK RESEARCH

This research will increase insight into Binning, Variable Selection, and Income Modelling techniques as well as a general improvement in retail credit risk modelling methodology.