IBM

Risk Scoring for a Loan Application on IBM System z

Running IBM SPSS Real-Time Analytics

- Using IBM SPSS Modeler for Analytics modeling
- Configuring risk assessment in SPSS Decision Management
- Real-time scoring using a System z host

Mike Ebbers
Keith Doan
Andrew Flatt

Redbooks

ibm.com/redbooks

International Technical Support Organization

Risk Scoring for a Loan Application on IBM System z: Running IBM SPSS Real-Time Analytics

October 2013

SG24-8153-00

Note: Before using this information and the product it supports, read the information in “Notices” on page v.

First Edition (October 2013)

This edition applies to Version 10 of IBM DB2 for z/OS, Version 4.1 of IBM CICS, and Version 15 of SPSS.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

CICS®
Cognos®
DB2®
IBM®
InfoSphere®
Redbooks®
Redbooks (logo) ®
SPSS®
System z®
WebSphere®
z/OS®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Preface

When ricocheting a solution that involves analytics, the mainframe might not be the first platform that comes to mind. However, the IBM® System z® group has developed some innovative solutions that include the well-respected mainframe benefits. This book describes a workshop that demonstrates the use of real-time advanced analytics for enhancing core banking decisions using a loan origination example. The workshop is a live hands-on experience of the entire process from analytics modeling to deployment of real-time scoring services for use on IBM z/OS®.

In this IBM Redbooks® publication, we include a facilitator guide chapter as well as a participant guide chapter. The facilitator guide includes information about the preparation, such as the needed material, resources, and steps to set up and run this workshop. The participant guide shows step-by-step the tasks for a successful learning experience. The goal of the first hands-on exercise is to learn how to use IBM SPSS® Modeler for Analytics modeling. This provides the basis for the next exercise “Configuring risk assessment in SPSS Decision Management”. In the third exercise, the participant experiences how real-time scoring can be implemented on a System z.

This publication is written for consultants, IT architects, and IT administrators who want to become familiar with SPSS and analytics solutions on the System z.

Authors

This book was produced by a team of specialists from around the world working at the IBM International Technical Support Organization, Poughkeepsie Center.

Mike Ebbers is an ITSO project leader in Poughkeepsie, NY.

Keith Doan is an IBM employee in Australia.

Andrew Flatt was an IBM employee in the United Kingdom when this book was written.

Thanks to the following people for their contributions to this project:

Richard Conway
IBM International Technical Support Organization, Poughkeepsie Center

Now you can become a published author, too!

Here’s an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Proof of Technology overview

The Proof of Technology (PoT) overview described in this IBM Redbooks publication demonstrates the use of real-time advanced analytics for enhancing core banking decisions, in this case loan origination. The PoT is a live hands-on experience of the entire process from analytics modeling to deployment of real-time scoring services for use on IBM z/OS.

Each participant will use a Windows 7 client for all development and deployment tasks.

1.1 Planning the Proof of Technology

Note: Read the information in this document carefully *before* planning a session with your prospect.

You need the following to conduct this PoT successfully:

- A facility with a classroom setup:
  - Overhead projector
  - Seating for a maximum of 20 participants (teams of two participants each)
  - A number of laptops or workstations with the appropriate client software
    - Either stand-alone or connected to an ESX server installed locally
    - Ethernet or wireless network connectivity to the Internet
    - At least 8 GB RAM, and ideally more
    - Free hard disk space of at least 60 GB
  - Internet connectivity
- Access to a System z environment with this host PoT software installed[1]
  - You need to reserve time and space on a System z host
  - You also need to request certificates and user IDs to be able to access the System z environment over the Internet from each participant workstation
- Staff to conduct the PoT

  This PoT is not like most other PoTs. It might require multiple subject matter experts (SMEs) to run. For the following areas, you need to arrange for an SME to be available to present and to assist in the hands-on sessions:
  - Module 1: Analytics modeling with IBM SPSS Modeler
    - IBM SPSS Modeler
  - Module 2: Configure the Risk Assessment in SPSS Decision Management
    - IBM SPSS Decision Management
  - Module 3: Configuration of the Risk Assessment for Real-Time Scoring
    - IBM SPSS Collaboration and Deployment Services
    - IBM CICS® TS on z/OS
    - IBM DB2® on z/OS
- Besides the above list of products, the facilitator team must be well versed in:
  - Service-oriented architecture (SOA), business process management (BPM), and Web Services concepts and technology
  - Data mining concepts
  - Using System z, z/OS, and Linux on System z, primarily to be able to start, stop, and restart servers, and do some troubleshooting

[1] Contact the Poughkeepsie ITSO using redbooks@us.ibm.com for suggestions.

1.2 Tasks to plan and execute the PoT

This PoT requires a number of tasks to complete before delivery can take place.

1.2.1 Several weeks before the PoT delivery

1. Read this facilitator guide thoroughly.
2. Check with the client on possible delivery dates.
3. Discuss with your client the agenda of the PoT and discuss which modules that the client wants to perform. Modules have to be executed in a sequential order so it would require some work to run a subsection of them.
4. Reserve classroom equipment at the local TEC center, or arrange for classroom equipment at the client site.

   Ideally, you run the PoT with teams of two persons each. Plan for one or two spare machines.
5. Reserve the PoT System z environment on your chosen host for the dates requested.
6. Request certificates or user IDs. You need a certificate or user ID for each classroom machine and a few spare ones. Therefore, if you plan to use 10 classroom machines, you should request at least 12 certificates or user IDs.
7. You need to install the PoT client software on each team's workstation.
8. Plan the required IBM staffing.

1.2.2 About one week before the PoT delivery

1. Obtain the following assets:
   a. Lab scripts
   b. Presentations
   c. Facilitator guide
   d. Workstation client software
2. Send the lab scripts and presentation slides to the reproduction department for duplication. Each participant requires a complete set of hardcopy material.

1.2.3 As soon as you have access to the System z system

STOP! Before proceeding, you need to ensure that the host names used in the labs point to the correct IP addresses.

Perform these steps as soon as possible, but no later than a few days before PoT delivery:

1. Ensure that the host names point to the correct IP addresses in the `etc/hosts` file in the client. Comment and uncomment entries as required and save the changes.
2. Test the connectivity with the System z environment:

   Open a Telnet, browser, or Personal Communications session and use the IP addresses as explained in 2.1, "Preparing the System z environment" on page 10. Ensure that you check connectivity with the z/OS logical partition (LPAR), as well as *both* z/Linux systems.
3. Test availability of all required software on System z, using the information in 2.1, "Preparing the System z environment" on page 10.

4. Eventually, clean the environment, again using the information in 2.1, “Preparing the System z environment” on page 10.
5. Run through the modules to ensure that you are familiar with the content and that the systems are working correctly.
6. Copy the client workstation software to the other machines.

1.2.4 Just before PoT delivery

Before delivery of PoT, proceed with the following steps:

1. Verify the license for Windows 7.
2. Verify the license for the SPSS software.

1.3 Flow

The PoT basically consists of presentations and practical modules. The time that is required for each presentation and module is variable. The length of the PoT is 4 - 6 hours, depending on how much you present and how smoothly the attendees go through the modules:

- Presentation 1: General SPSS Solution Overview
- Presentation 2: Business Scenario
- Presentation 3: Module Outline
- Module 1: Analytics modeling with IBM SPSS Modeler
- Module 2: Configure the Risk Assessment in SPSS Decision Management
- Presentation 4: Putting it all together
- Module 3: Configuration of the Risk Assessment for Real-Time Scoring

Figure 1-1 on page 5 shows a logical representation of the artifacts that are required for scoring data.

1.3.1 Design

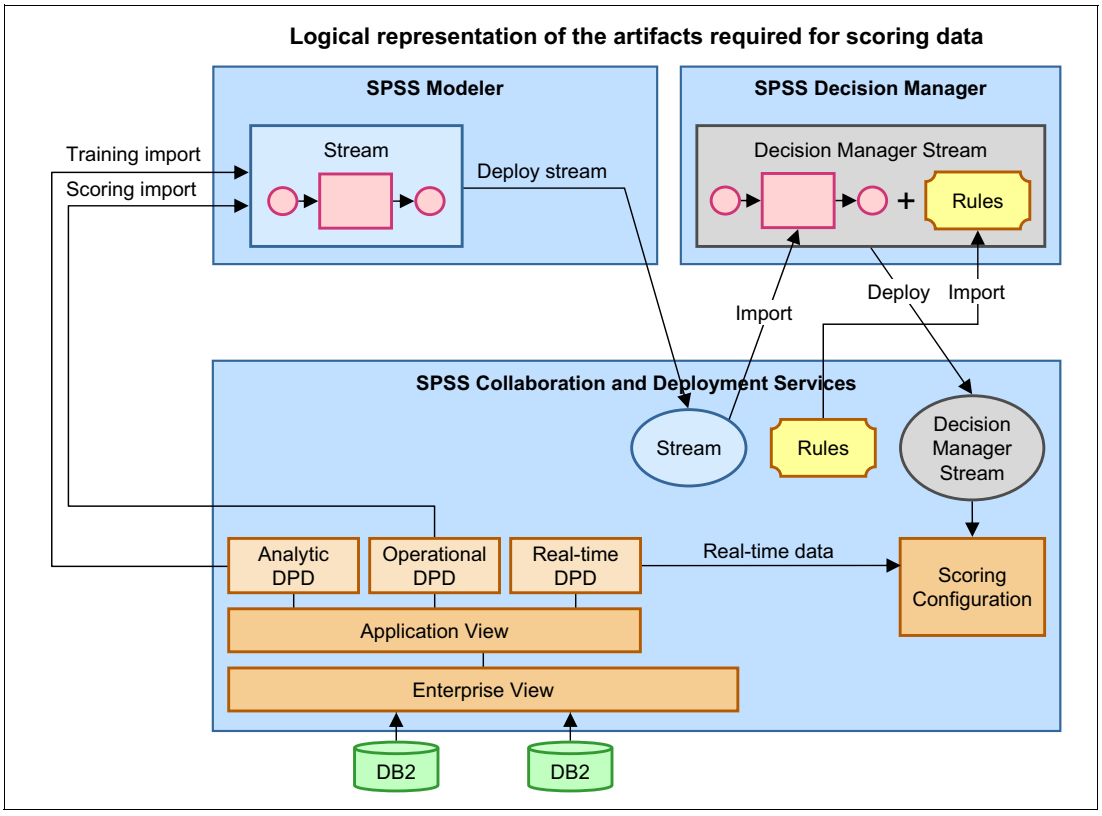


Figure 1-1 Logical representation of the artifacts required for scoring data

Figure 1-2 on page 6 shows a logical representation of the actual artifacts that are required for scoring data in these modules.

1.3.2 Actual implementation

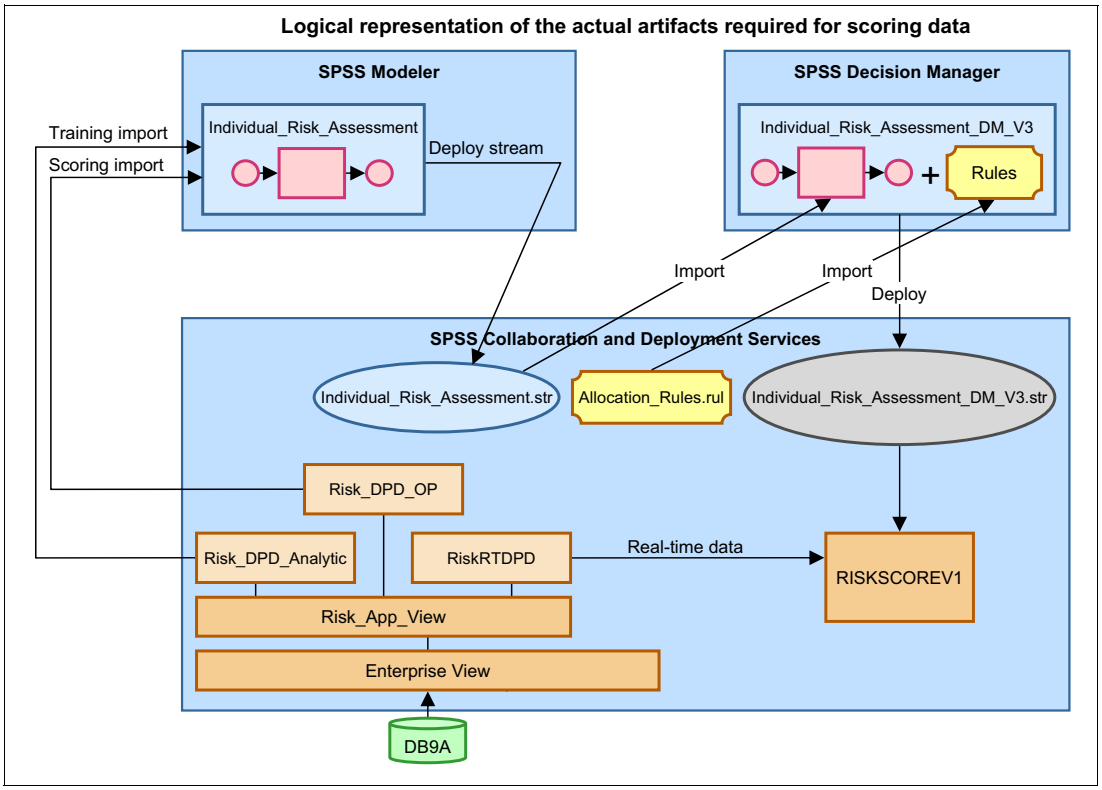


Figure 1-2 Logical representation of the actual artifacts required for scoring data in these modules

Figure 1-3 on page 7 shows a logical representation of the architecture that is required for real-time scoring from CICS.

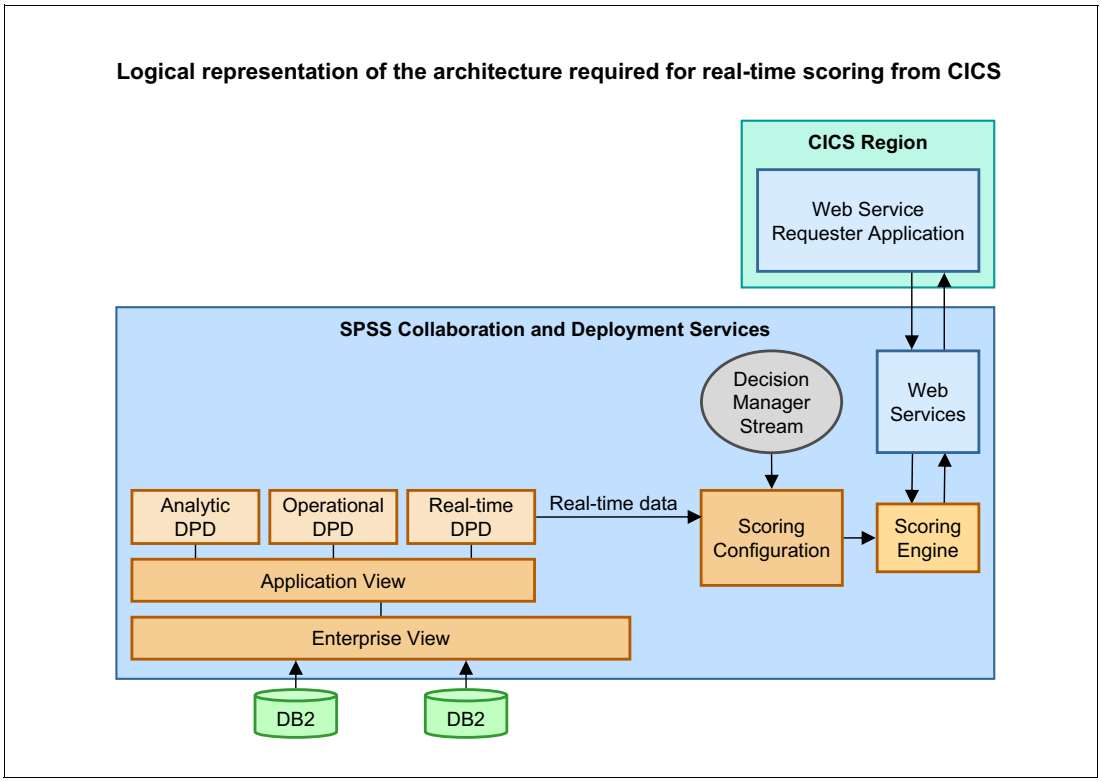


Figure 1-3 Logical representation of the architecture required for real-time scoring from CICS

1.4 Modules

The PoT is set up in a modular way. Each module has an “initial” state, which is the situation that the participant starts with, and a “solution” state, which is the situation that the participant should end with. Each module requires the initial state to be the solution state of the previous module.

The following modules are available.

Table 1-1 PoT modules

Module	Activities/topics
Module 1: Analytics modeling with IBM SPSS Modeler	► Import data into SPSS Modeler ► Create an SPSS Modeler analytics model ► Test an SPSS Modeler analytics model ► Deploy the SPSS Modeler analytics model
Module 2: Configure the Risk Assessment in SPSS Decision Management	► Configure a Risk Assessment application ► Connect Data ► Define Outcomes ► Configure Operational Decisions with Rules and Models ► Combine Rules and Analytical models to optimize decision outcomes ► Deploy a Risk Assessment application
Module 3: Configuration of the Risk Assessment for Real-Time Scoring	► Configure an SPSS Decision Management stream for real-time scoring ► Test the real-time scoring configuration ► Invoke the real-time scoring from a CICS TS Transaction

1.5 Handouts

Handouts need to be reproduced in hardcopy.

Note: Each participant needs to receive a hardcopy of all lab scripts scheduled. In most cases, it is acceptable to provide the presentations in softcopy format.

1.5.1 Presentations

Presentations are provided in PowerPoint format. Therefore, you can do your own customizing for your session. The presentation materials should be regarded as a “superset” so that you can reduce the content to your technical comfort level.

You can also replace the provided slide decks with your own slide decks, if they fit well.

1.6 Products used

Figure 1-4 illustrates which products are used and how they integrate with each other.

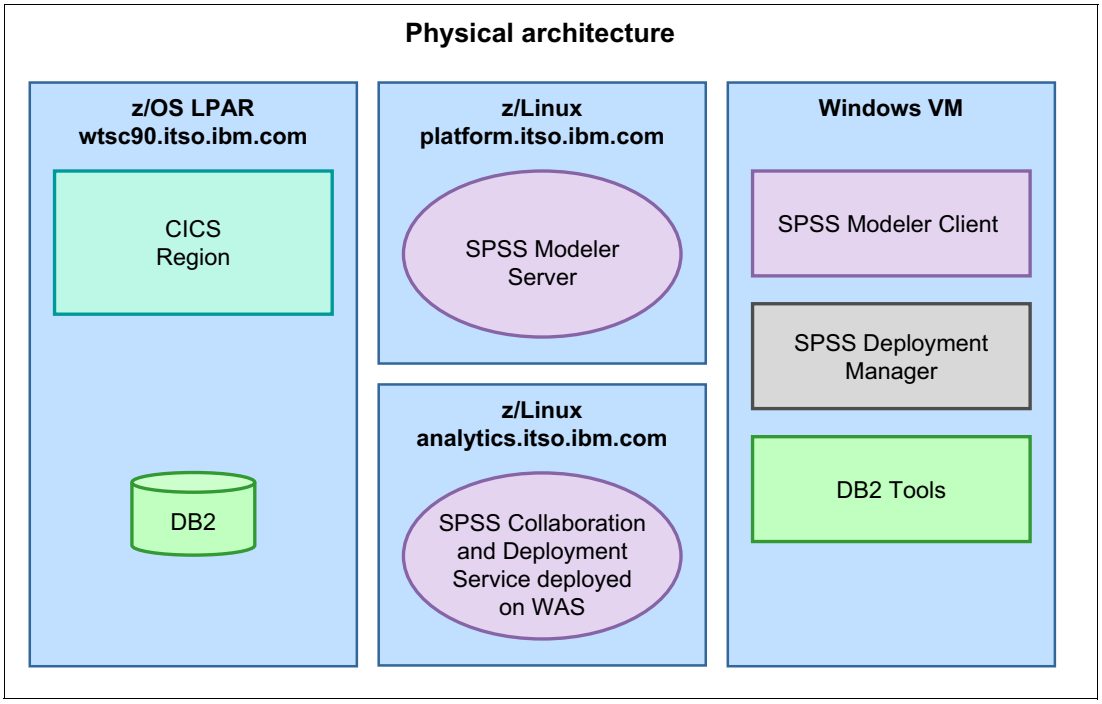


Figure 1-4 PoT hardware and software environment

2

PoT technical overview

The System z environment for this Proof of Technology (PoT) can be hosted on a standard System z platform that is running the required software. See 1.2, “Tasks to plan and execute the PoT” on page 3.

2.1 Preparing the System z environment

The System z environment consists of a z/OS logical partition (LPAR) and two Linux systems. The systems listed below are examples only, for your use in setting up your host system.

2.1.1 z/OS LPAR

Note: You need to map the host name used to the IP address that you are going to use in the `etc/hosts` file in the clients. If you are providing your own host, you need to add a line in `etc/hosts` to make the mapping. You need to comment/uncomment the entries depending on where you run the PoT.

The following software products are used on z/OS:

- CICS Transaction Server V4.1
- DB2 for z/OS V9.1

Note: The above software might or might not be running and you need to verify for each one whether it is running or not. Instructions follow below to start or restart parts of the environment.

STOP: If you have no experience with z/OS, contact somebody you know with z/OS experience.

Sample connection details

Host name ITSO	wtsc90.itso.ibm.com
IP address POK Yellow	10.52.52.121
User ID	ITSO99
Password	ITSO99

CICS Transaction Server

Each team uses the same CICS region, named *CITSO99*. Below are the characteristics. Substitute "*XX*" with the team number.

CICS host name	wtsc90.itso.ibm.com
CICS port	3099
CICS started task	SYS1.PROCLIB(CITSO99)
CICS start command	S CITSO99
CICS stop command	C CITSO99
CICS terminal	Point your Personal Communications (PCOMM) at wtsc90.itso.ibm.com, port 23, and enter CITSO99
CICS user ID	ITSOXX
CICS Password	ITSOXX (never expires)
CICS IBLA datasets	ITSO99.IBLA.*
CICS libs	CICSTS41

IBLA application

The ITSO Bank Loan Application (IBLA) is a custom-designed and simplified back-end application that provides certain functions in initiating and maintaining a loan.

Initialization of the CICS environment

Note: Prior to running the PoT, it is strongly recommended to reset the CICS environment. Follow the below procedure to do this.

- In data set ITSO99.IBLA.JCL, run jobs ITSO01L through ITSO12L.
- After starting/restarting each CICS region, you have to install the groups *IBLA* and *SPSS*:

```
CEDA I GRP(IBLA)
CEDA I GRP(SPSS)
```

Some troubleshooting techniques:

- SDSF LOG
- SDSF sysout of CICS region
- In CICS, CEMT I PROG(<*progname*>), to check transaction count

DB2

The operational data used by the IBLA application is stored in DB2 tables. All participants use the same set of tables. Below is the technical relevant information for the DB2 environment:

DB2 host name	wtsc90.itso.ibm.com
DB2 TCP port	38350
DB2 RES port	38351
DB2 location name	DB9A
DB2 SSN	DB9A
DB2 JDBC HFS	/usr/lpp/db2/db9a/db2910_jdbc
DB9A setup JCL	DB9AU.SDSNSAMP
DB9A ISPF panels	Enter 9A from main ISPF panel
DB9A volumes	STORCLAS=DB9A

Check if data exists in DB2 tables

Create the analytics table that is to be used for SPSS modeling and scoring by following these steps:

- Open SPSS Modeler and retrieve the Modeler stream from the Repository at, \Individual Risk Assessment\Data\CheckData.str.
- Within `CheckData.str`, run the database source nodes to check if data exists. If not, run the flow to populate required data.

2.1.2 z/Linux system: platform

Important: Use Putty to access the z/Linux systems (SSH on Port 22).

Note: You need to map the host name used to the IP address that you are going to use in the `etc/hosts` file in the clients. If you are providing your own host, you need to add a line in `etc/hosts` to make the mapping. You need to comment/uncomment the entries depending on where you run the PoT.

This system hosts SPSS Collaboration and Deployment Services running on IBM WebSphere® Application Server (WAS). The artifacts for this Linux system are:

Host name ITSO platform.itso.ibm.com

IP address POK Yellow 10.52.78.13

User ID root

Password rootpw

DB2

A DB2 database is also used to support SPSS Collaboration and Deployment Services. The artifacts are:

- Starting the DB2 server:
  - su db2inst1 (password `db2inst1`, if prompted)
  - db2stop
  - db2start
  - Exit to change back to root
- Product path:

  /code/IBM/db2/V9.5
- Instance path:

  /code/IBM/db2inst1

SPSS Collaboration and Deployment Services running on WebSphere Application Server

The artifacts for WAS are:

WAS admin console https://<IP address>:9043/ibm/console/logon.jsp

Path to profile /code/IBM/WebSphere/AppServer/profiles/AppSrv01

CnDS Username admin

CnDS Password passw0rd

Restarting the WAS server

A restart of SPSS Collaboration and Deployment Services might be necessary, which is achieved by restarting the WAS server hosting SPSS Collaboration and Deployment Services:

- cd /code/IBM/WebSphere/AppServer/profiles/AppSrv01/bin
- ./stopServer.sh server1
- ./startServer.sh server1

Initialization of the SPSS Collaboration and Deployment Services repository
After each PoT, the SPSS Collaboration and Deployment Services repository may need to be cleaned. This can be done manually very easily, as follows:

- Use SPSS Collaboration and Deployment Services Deployment Manager.
- Log on to SPSS Collaboration and Deployment Services using the CnDS username and password.
- Expand the content repository.
- Delete all the files under each ITSOXX folder, as shown in Figure 2-1.

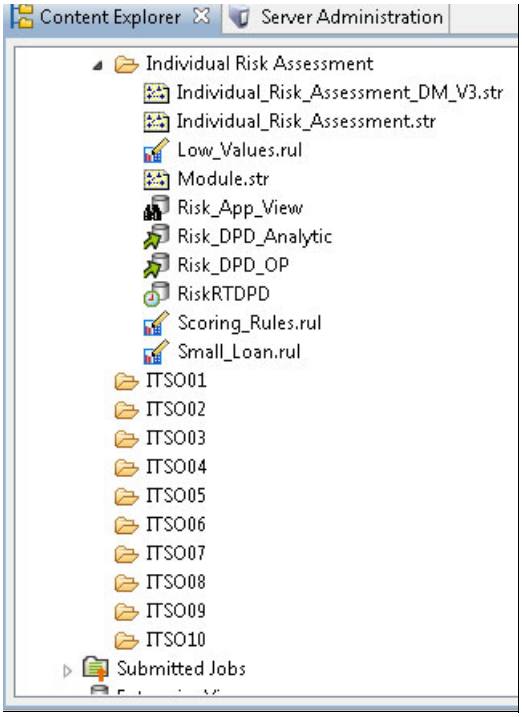


Figure 2-1 Content Repositories starting point

- Go to **View** → **Show View** → **Scoring**.
- Select any RISKSCORING**XX** configurations and right-click them.
- Click **Delete Configuration(s)**.

2.1.3 z/Linux system: analytics

Important: Use Putty to access the Linux systems (SSH on Port 22).

This system hosts SPSS Modeler Server. The artifacts for this Linux system are:

Host name ITSO	analytics.itso.ibm.com
IP address POK Yellow	10.52.78.12
User ID	root
Password	rootpw

SPSS Modeler Server

Perform the following actions to stop, start, or check if the SPSS Modeler Server is started.

Stopping the SPSS Modeler Server

- To stop the SPSS Modeler server run:

  /code/IBM/SPSS/ModelerServer/14.2/ # ./modelersrv.sh stop

Starting the SPSS Modeler Server:

- To start the SPSS Modeler server run:

  /code/IBM/SPSS/ModelerServer/14.2/ # ./modelersrv.sh start

Checking if the SPSS Modeler Server is started

- To check if the SPSS Modeler server has started to run:

  /code/IBM/SPSS/ModelerServer/14.2/ # ./modelersrv.sh list

2.2 Preparing the clients

This PoT uses software on a Windows workstation for all the client applications required.

2.2.1 Client software

- SPSS Modeler Client
- SPSS Collaboration and Deployment Services Deployment Manager
- DB2 Tools

2.3 Final checks

Here are some tasks for checking your setup.

2.3.1 Run the scoring transaction to test setup

To test that the scoring transaction can run, the first thing that you need to do is open a 3270 terminal to the mainframe. You do this by logging on to the CICS region so that you can invoke a transaction to call the scoring service. Use IBM Personal Communications.

__ 1. Double-click the wtsc90 session.

__ 2. Set up a connection to the z/OS machine.

__ 3. You should see a logon panel similar to Figure 2-2 on page 15.

__ 4. To log in to the CICS region, type “CITS099”.

__ 5. Press the right control button (Ctrl) to enter the command.

__ 6. Enter the user ID “ITSO99” and the password “ITSO99”.

__ 7. Press the right control button (Ctrl) to enter the logon credentials.

__ 8. You should see a message like **`DFHCE3549 Sign-on is complete`**, which informs you that your logon was successful.

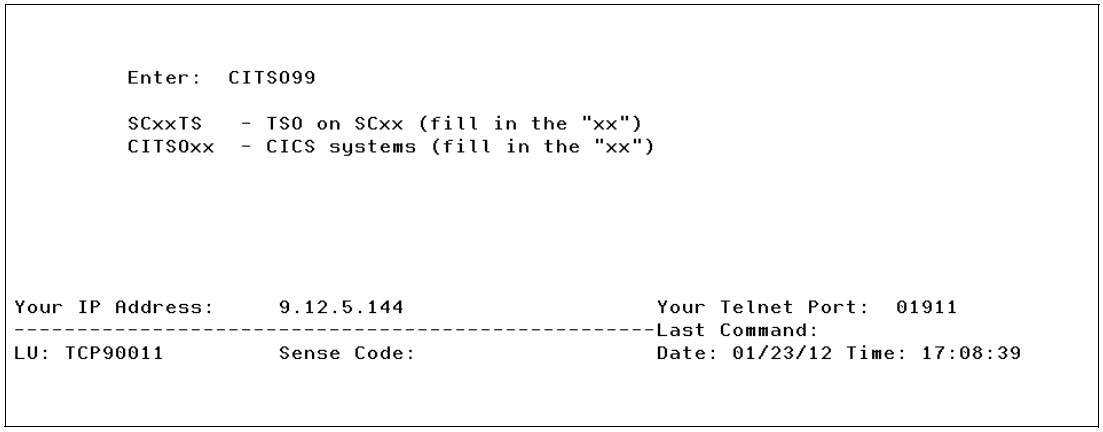


Figure 2-2 ITSO logon panel

Figure 2-3 shows the panel to sign on to CICS.

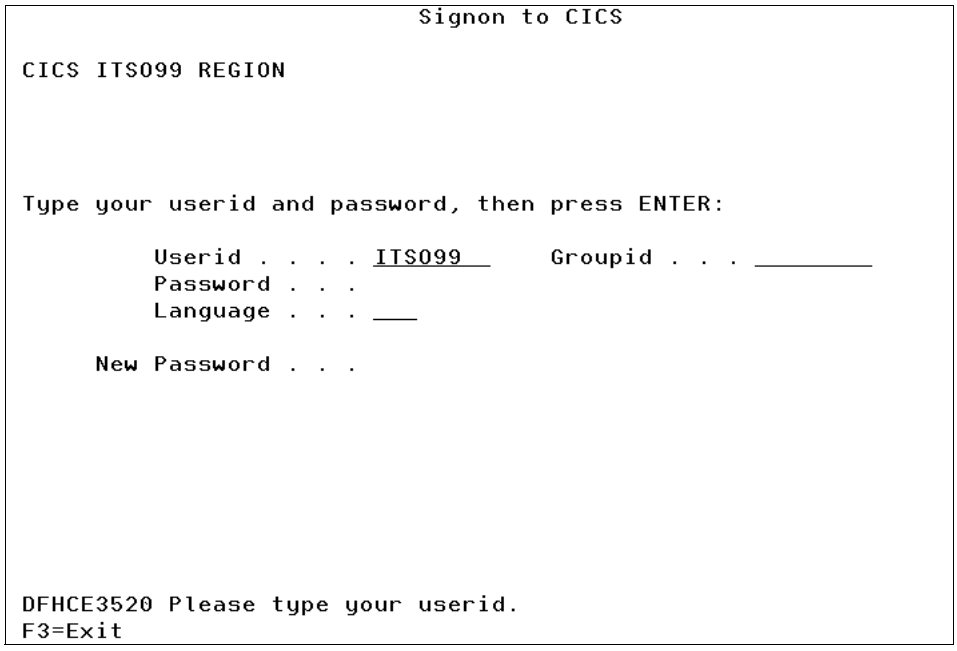


Figure 2-3 Sign on to CICS panel

__ 9. Press the pause button to clear the display.

__ 10. In the top right corner, type the transaction name “SCR1”, as shown in Figure 2-4 on page 16.

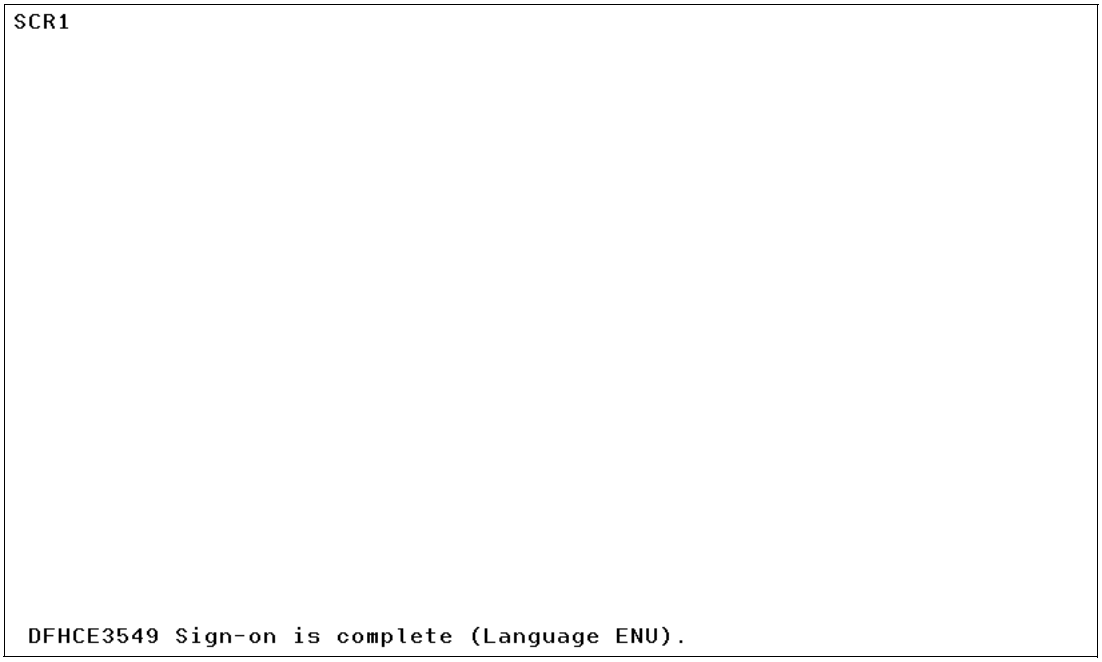


Figure 2-4 Entering the scoring transaction

__ 11. Press the right control button (Ctrl) to submit the transaction to CICS.

__ 12. You should now see the scoring application CICS user interface (UI). This will look similar to Figure 2-5.

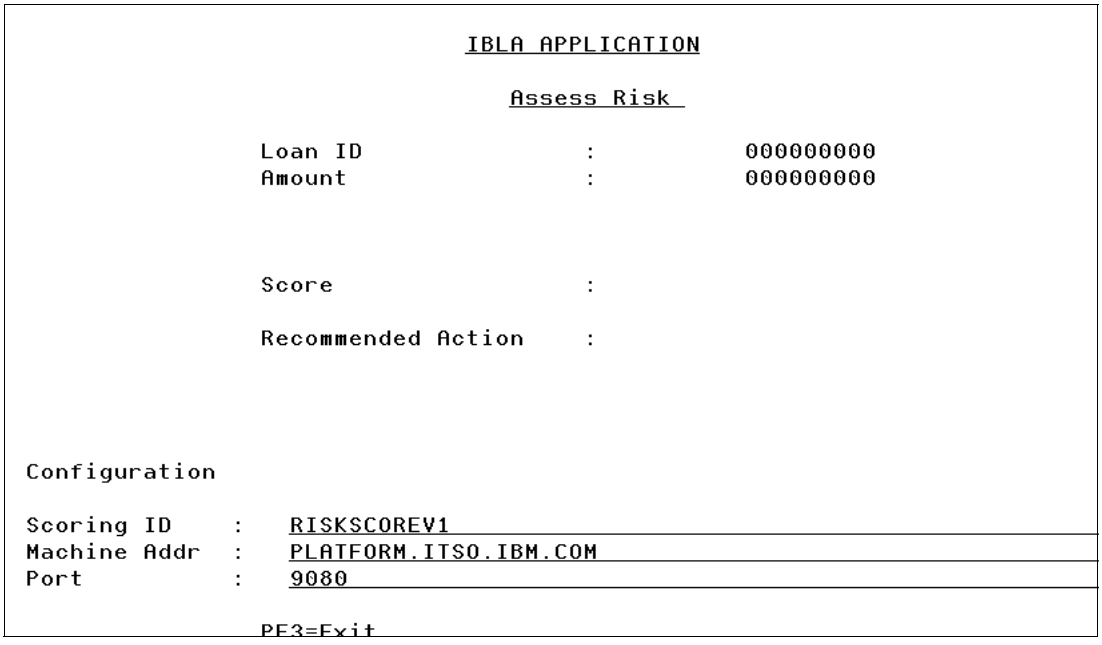


Figure 2-5 CICS Scoring 3270 user interface

__ 13. Enter a **Loan ID**. This can be a number 1 - 9.

__ 14. Enter an **Amount**.

__ 15. Press the right control button (Ctrl) to submit the inputs for scoring.

__ 16. You now can see that a scoring algorithm ran and returned a risk score and a recommend action. See Figure 2-6 for a successful scoring execution.

Note: The formatting causes the inputs to have leading zeros, that is, 000000003.

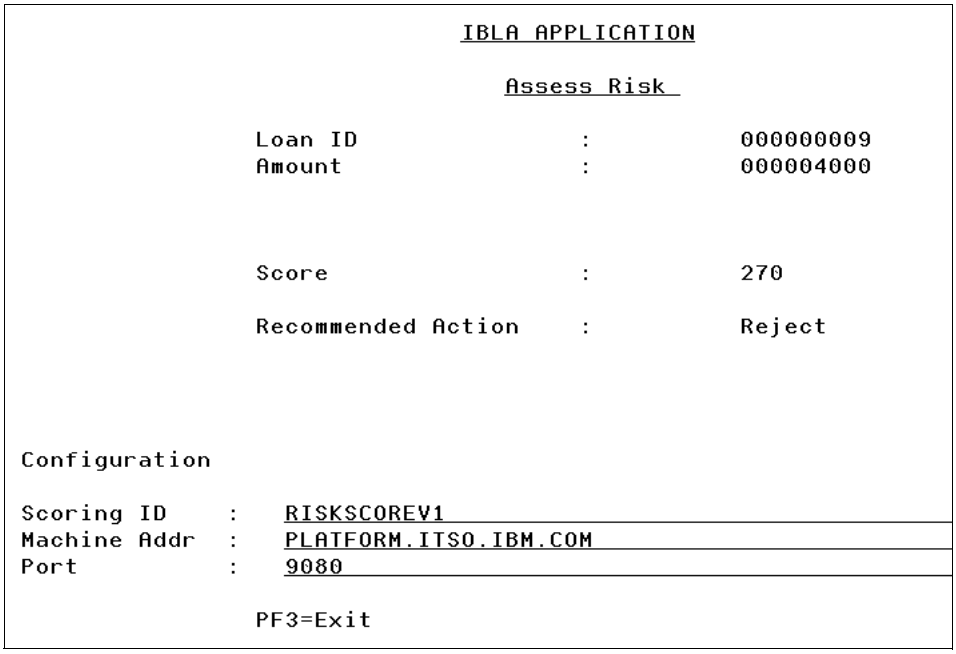


Figure 2-6 Successful scoring execution

Note: If you see an error message displayed in the application, check that the scoring configuration still exists and has access to the host files.

3

Preparations before you start

This chapter gives you information about how to use this guide and provides you with the information that you need to perform the tasks in the worksheet.

3.1 Important reading guidelines

It is important to read this section before you start any lab activities.

Attention: Do not start with any lab before you have read the instructions below.

- Review and complete the worksheet where necessary. Detach the sheet from the lab exercise book and place it on your table. You need to refer to the worksheet continuously.
- User IDs, host names, and other location-dependent variables are referred to as variable names in the lab scripts using the following font: *variable name*.

  Refer to your worksheet for the value of the variable to be used in your lab.
- When you see a check box symbol "__1", it means that you have to "do" something on your computer—not merely read the document.
- If you see user IDs, passwords, file names, and so on, with an "xx" or "XX" in it, it usually means that you need to replace those characters with the number of the user ID that has been assigned to you. If in doubt, ask the facilitator.
- If you see a discrepancy between the screen capture and the text explaining the step, use the information in the text. Some screen captures do not exactly reflect the names and values that are used in the labs.
- Most errors occur because of typing errors. Ensure that you type exactly the values as printed in the document. Most of the values are *case-sensitive*.

3.2 Worksheet

In the labs many variables are used, such as user IDs, passwords, and port numbers.

Instructions:

- Review and complete this worksheet where necessary *before* starting with the labs.
- Copy or detach this worksheet from your book and place it on your table for reference during the labs. You will need to refer to this worksheet frequently.

Your team number: _____

Use this team number as a substitute for '**XX**' when indicated.

3.3 General system information

Variable	Your value
Modeler Stream Name	Individual_Risk_Assessment.str
Decision Management URL	http://platform.itso.ibm.com:9080/DM/
DM User	admin
DM Password	passw0rd
CnDS User	admin
CnDS Password	passw0rd
Platform Hostname	platform.itso.ibm.com
Platform Port	9080
Group Folder	ITSO**nn**, where '**nn**' is your team number
Decision Management Project Name	Individual_Risk_Assessment_DM.str
Scoring ID	RISKSCORING**nn**, where '**nn**' is your team number (MUST BE ENTERED IN CAPITALS)
z/OS Hostname	wtsc90.itso.ibm.com
z/OS User ID	ITSO**nn**, where '**nn**' is your team number
z/OS Password	ITSO**nn**, where '**nn**' is your team number
CICS Region Name	CITSO99
Screen Size	32x80
Scoring Transaction	SCR1
Scoring Address	PLATFORM.ITSO.IBM.COM
Scoring Port	9080
z/Linux Username	root
z/Linux Password	rootpw

4

Worksheet

In the labs many variables are used, such as user IDs, passwords, and port numbers.

Instructions:

- Review and complete this worksheet where necessary *before* starting with the labs.
- Detach this worksheet from your book and place it on your table for reference during the labs. You will need to refer to this worksheet frequently.

Your team number: _____

Use this team number as a substitute for **'XX'** when indicated.

4.1 General system information

Variable	Your value
Modeler Stream Name	Individual_Risk_Assessment.str
Decision Management URL	http://platform.itso.ibm.com:9080/DM/
DM User	admin
DM Password	passw0rd
CnDS User	admin
CnDS Password	passw0rd
Platform Hostname	platform.itso.ibm.com
Platform Port	9080
Group Folder	ITSO**nn**, where '**nn**' is your team number
Decision Management Project Name	Individual_Risk_Assessment_DM.str
Scoring ID	RISKSCORING**nn**, where '**nn**' is your team number (MUST BE ENTERED IN CAPITALS)
z/OS Hostname	wtsc90.itso.ibm.com
z/OS User ID	ITSO**nn**, where '**nn**' is your team number
z/OS Password	ITSO**nn**, where '**nn**' is your team number
CICS Region Name	CITSO99
Screen Size	32x80
Scoring Transaction	SCR1
Scoring Address	PLATFORM.ITSO.IBM.COM
Scoring Port	9080
z/Linux Username	root
z/Linux Password	rootpw

5

Analytics modeling with IBM SPSS Modeler

This lab takes you through the steps of creating and deploying predictive models for analytical insight in our chosen loan risk assessment scenario using the IBM SPSS Modeler.

5.1 Background reading: Modeling and IBM SPSS Modeler

Data mining is a general term that refers to a variety of modeling techniques that identify nuggets of information in (large) bodies of data, without necessarily having preconceived notions about what will be discovered. Data mining extracts information in such a way that it can be used in areas such as decision support, prediction, forecasts, and estimation. Data is often voluminous but of low value and with little direct usefulness in its raw form. It is the hidden information in the data that has value.

Data mining is an interactive and iterative process. Success comes from combining your (or your expert's) knowledge of the data with advanced, active analysis techniques in which the computer identifies the underlying relationships and features in the data. The process of data mining generates models from historical data that are later used for predictions, pattern detection, and more. The technique for building these models is called *machine learning*, or *modeling*.

In this lab, we have a sample of historical loan data with known outcomes to create a machine learning model to predict how likely a new individual loan is approved based on its data-driven risk profile. To evaluate the model performance, we partition the existing data for “training” a model and then “testing” it against the performance metrics such as overall prediction accuracy.

5.1.1 Key questions for a modeling project

These are key questions before starting a modeling project:

1. Is data available?

   Data needs to be stored in an easily accessible format. Often the data is stored in different locations or formats that need to be pulled together before analysis. There are also potential limitations, such as legal or political reasons, why data cannot be accessed.

2. Does the data cover relevant factors?

   It is important that the data contains all the relevant factors and variables. Often, an objective of data mining is to help identify the relevant factors in the data. If thought is given to this question, you achieve a greater prediction accuracy.

3. Is the data erroneous?

   The more erroneous or missing data, the more difficult it will be to make accurate predictions. IBM SPSS Modeler capabilities have been shown to successfully handle data consisting of data made up by 50% errors.

4. Is there enough data?

   The answer depends on the individual problem. It is not often the amount of data that causes difficulties in modeling. It is the attempt to represent the target population and cover all possible outcomes.

5. Is the expertise on the data available?

   Successful modeling projects require domain expertise that is practical and relevant. We require knowledge about how the data was generated, the data characteristics, how the data is used, and what the intended use is of a modeling project. The domain expert guides the project in identifying relevant factors, helps interpreting the results, and sorts out the truly useful pieces of information from a business perspective.

Because the objective of this module is to illustrate the proof of technology, we have simplified the data for facilitating modeling and deployment activities for expediency in a training course.

5.1.2 Introduction to IBM SPSS Modeler

IBM SPSS Modeler is a data mining workbench that supports all the steps in the data mining process. IBM SPSS Modeler can run in local mode or distributed (client-server) mode. In this lab, you run IBM SPSS Modeler in distributed mode, as shown in Figure 5-1.

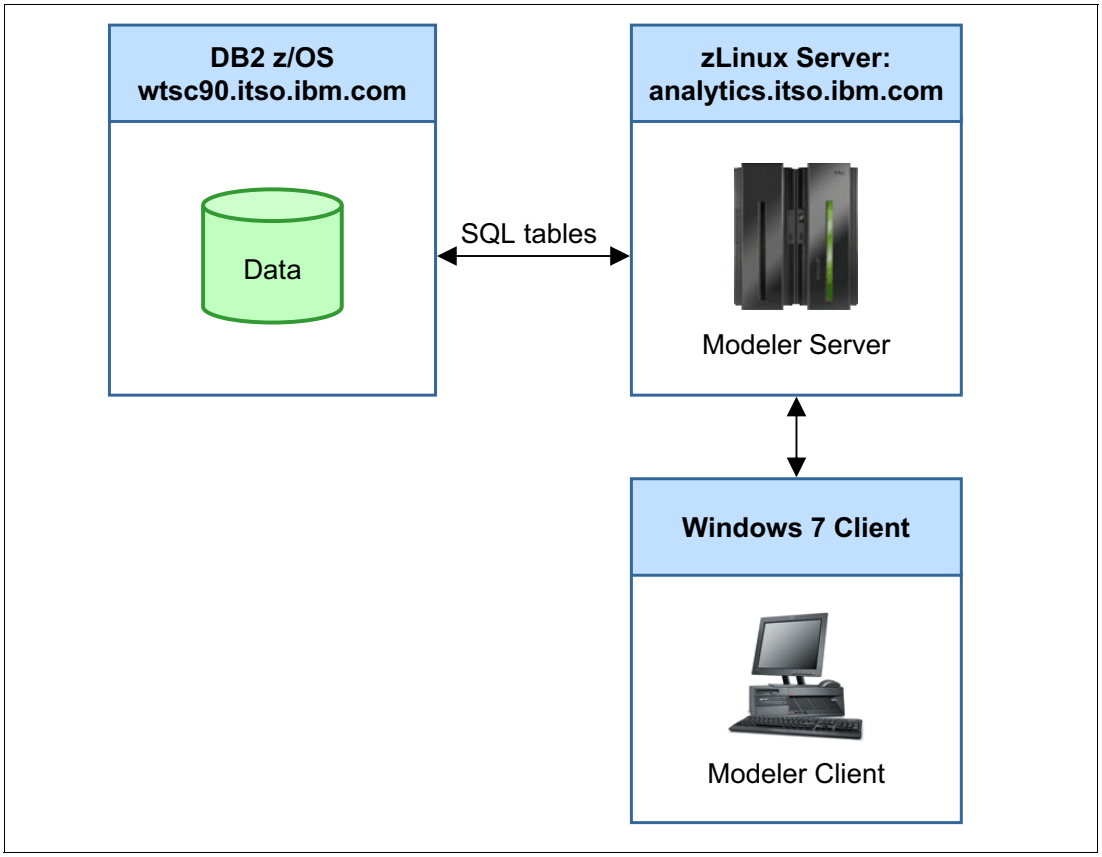


Figure 5-1 Modeler server topology

5.1.3 IBM SPSS Modeler benefits

IBM SPSS Modeler is easy to learn because it has an intuitive visual interface and requires no programming. The SPSS Modeler workbench offers a comprehensive range of data mining functions with powerful automation including automated data preparation and multi-model creation and evaluation.

The SPSS Modeler is based on an open and scalable architecture that allows for:

- SQL pushback support
- Maximized use of infrastructure with multithreading, clustering, and use of embedded algorithms (in-database mining)
- Integration with IBM technologies such as IBM Cognos® and IBM InfoSphere® Warehouse

SPSS Enterprise View

Enterprise View (*EV*) is a lightweight virtual data layer that can bring together multiple data sources in one virtual layer. This streamlines data access for the business and can be integrated with existing data management technologies of the enterprise—including enterprise data warehouses, entity management, and master data management.

EV consists of tables that list all data elements that can be used throughout the organization.

EV provides the necessary separation of concerns for analytics users. They do not have to consider physical data sources at modeling time. It decouples them from the impact of physical data sources changing.

When building individual analytical applications, only a subset of all the elements are required. These required elements are defined in the **Application View** by mapping to selected EV tables.

Adding a new data source such as a loan payment history, involves mapping the new source by defining a *Data Provider Definition* (*DPD*). When that has been defined, the new attributes can be leveraged within the analytics and operational environment for rules, arbitration inputs, or model inputs.

The DPD is the only component that actually knows where to query and collect the required data to pass to the Application View.

In this Proof of Technology (PoT), DPDs differ per environment: Training for analytics, scoring for operational, and real-time scoring where data is collected from a combination of inside the request message (which is defined as *context data*) and the database.

Functionally, *Training* represents the historical data with known outcomes. *Scoring* represents the loan transactions that require prediction through scoring, hence, without the predicted field. *Real-time scoring* represents the real-time data that requires the real-time input data by users for individual transaction scoring.

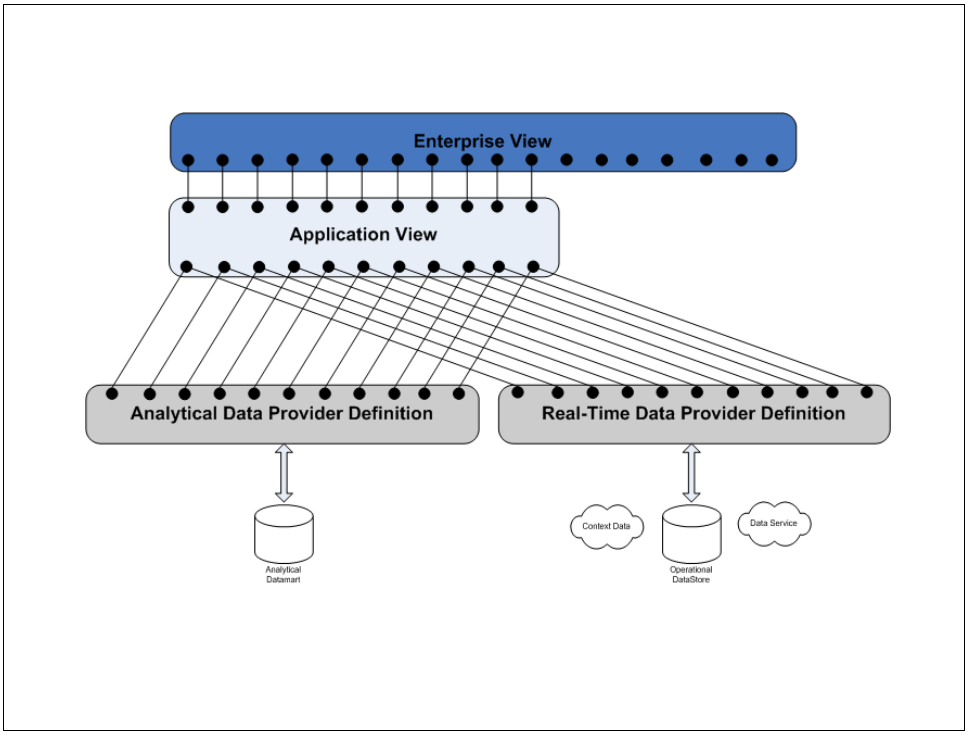


Figure 5-2 SPSS Enterprise View concepts

5.2 What this lab is about

In this lab, you retrieve loan data, prepare the data, build models, evaluate them, and then deploy them into the operational environment.

Important: Each time that you see this check box "__ 1", it means that you have to do something on your workstation in addition to reading the document.

5.2.1 What will you learn?

At the end of this lab exercise, you should have learned:

- How to create predictive models in IBM SPSS Modeler
- How to connect and retrieve enterprise data for the loan risk assessment scenario
- How to prepare data for modeling
- How to create models using the automated modeling feature
- How to evaluate models that meet your modeling objective
- How to deploy your chosen model to your operational environment for business application

5.2.2 Prerequisites

- You should have access to the client workstation where you will be performing the lab exercise. Check with your lab facilitator to get access.
- You should have received your team number from the lab facilitator. You will be using this number during lab and it will determine variables in the worksheet.

5.3 Building a predictive model

You open the SPSS Modeler workbench to create a workflow for building a model. You connect data to the training data view, a specific enterprise view for building model. You then partition the input data and set each data field to a role. When the data is ready, you create a model for the sample data using automated modeling. The resulted model will be evaluated to ensure that the most appropriate model is chosen for deployment.

Important: All activities are to be performed on the client.

5.3.1 Starting SPSS Modeler Workspace

__ 1. Start IBM SPSS Modeler: **Start** → **All Programs** → **IBM SPSS Modeler 14.2** → **IBM SPSS Modeler 14.2**.

__ 2. You are presented with the Server Login dialog. See Figure 5-3 on page 30.

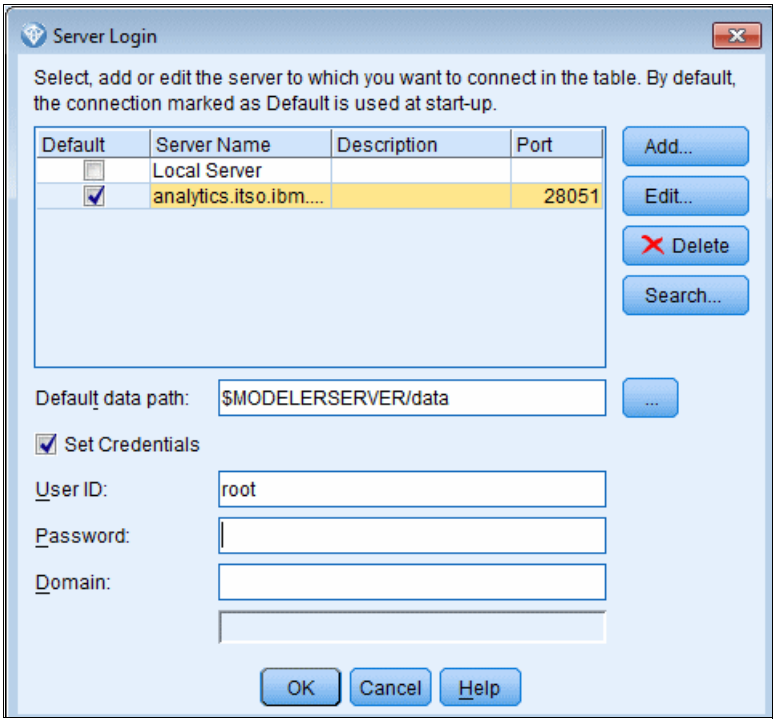


Figure 5-3 IBM SPSS Modeler server login

__ 3. Set the **User ID** to the *z/Linux Username* from the worksheet

__ 4. Set the **Password** to the *z/Linux Password* from the worksheet

__ 5. Click **OK**.

__ 6. You are then presented with the Modeler Workbench. See Figure 5-4 on page 31.

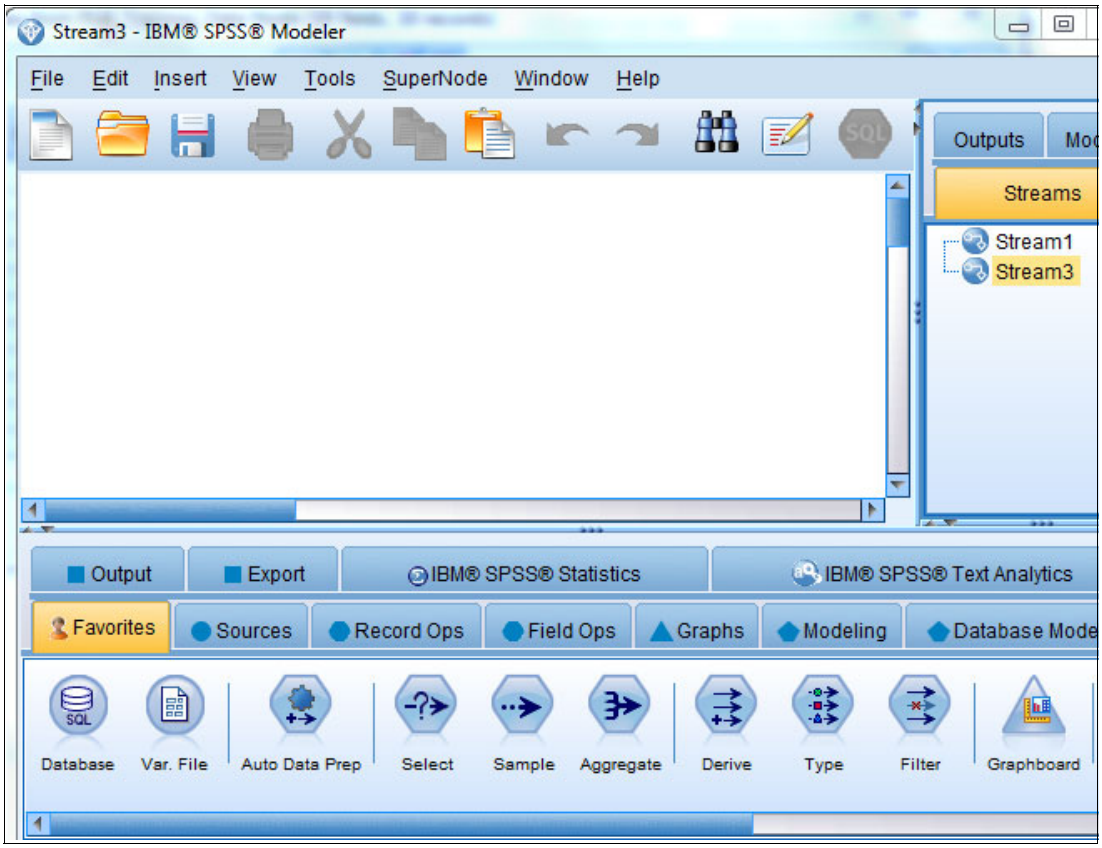


Figure 5-4 BM SPSS Modeler Workbench

5.3.2 Connecting to data

In this next step, you set up the connection and partition the data. This data includes history on payments, income, assets, age, occupation, and other demographic data to enable the bank to assess credit worthiness of the loan applicant. In the scope of this lab, the use of the EV data capability means that users do not have to be concerned with physical data sources (see “SPSS Enterprise View” on page 27 for more explanation).

You want to partition the data into two sets: Training and Testing. You want to only use 50% of the data for training so that you can see how the resulting models perform against the test data. If you were to train and test against the same data, you may end up with a model that is too specific to the training data and not generic enough to be applied in the application. Then, you provide types and roles to the data so that the model can use the inputs effectively to derive a target value.

Follow these steps:

__ 7. Double-click the **Enterprise View** node (under the **Sources** tab).

__ 8. Double-click the **Partition** node and **Type** node (under the **Field Ops** tab).

__ 9. Validate that you have a workflow that connects the **Enterprise View** node to the **Partition** node, then from the **Partition** node to the **Type** node. This configuration is shown in Figure 5-5 on page 32.

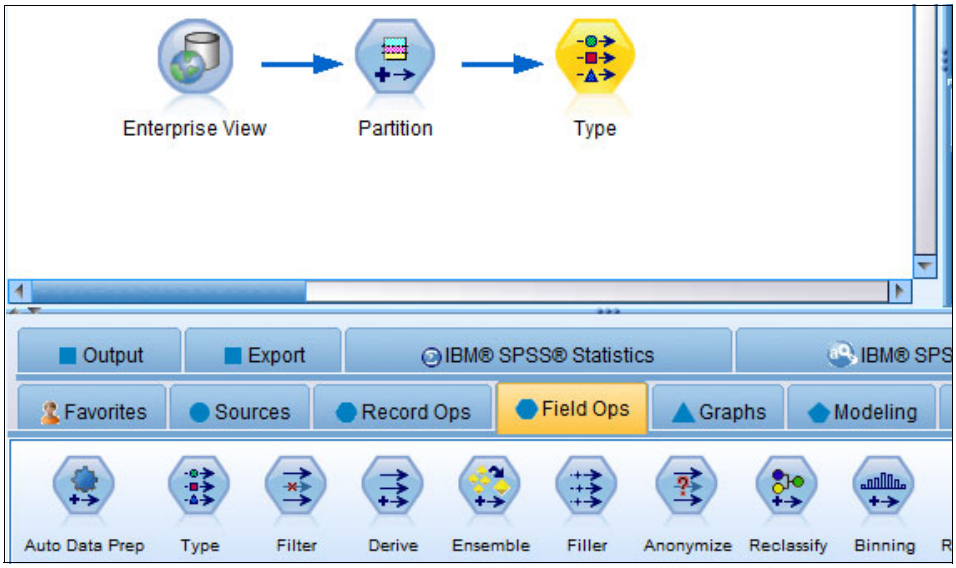


Figure 5-5 SPSS Modeler Workflow

Steps 10 - 18 on page 33 are to establish the connection to EV data that is managed by the repository:

__ 10. Double-click the **Enterprise View** to show the dialog to specify the application view and table from which to read data.

__ 11. Click the Connection drop-down menu.

__ 12. Select **Add/Edit a connection**, as shown in Figure 5-6.

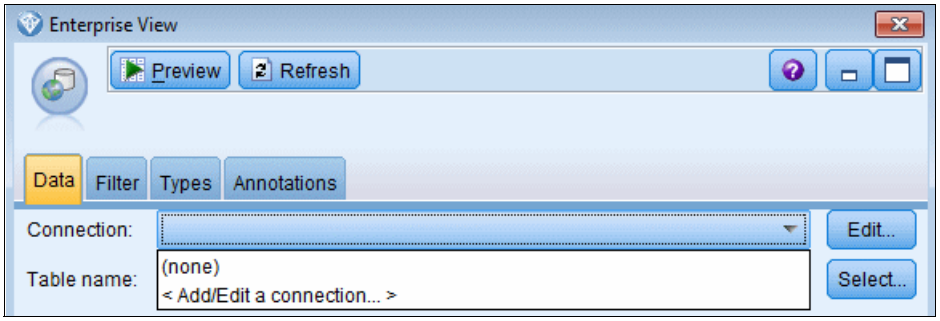


Figure 5-6 Enterprise View dialog

A Repository Server connection dialog opens, as shown in Figure 5-7.

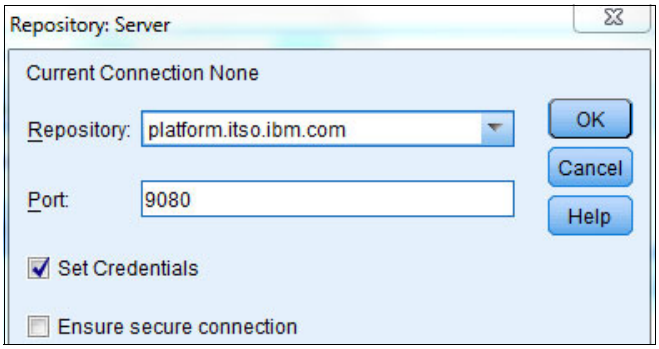


Figure 5-7 Repository Server connection dialog

__ 13. Enter the *Platform Hostname* from your worksheet in to the **Repository** input.

__ 14. Enter the *Platform Port* from your worksheet into the **Port** input.

__ 15. Click **OK**.

A Repository Credentials dialog opens, as shown in Figure 5-8.

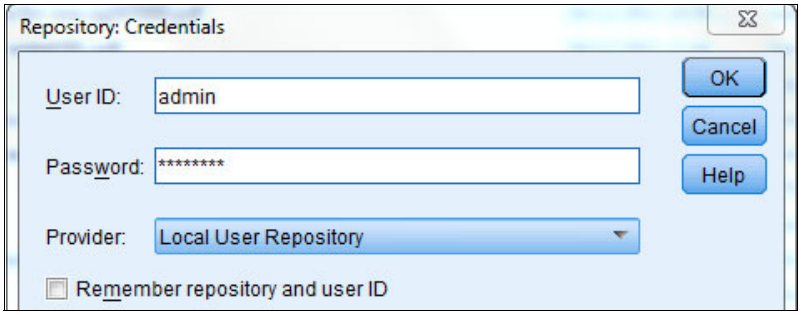


Figure 5-8 Repository Credentials dialog

__ 16. Enter the *CDS User* from your worksheet into the **User ID** input.

__ 17. Enter the *CDS Password* from your worksheet into the **Password** input.

__ 18. Click **OK**.

When the repository connection is established, an Enterprise View Connections dialog opens, as shown in Figure 5-9.

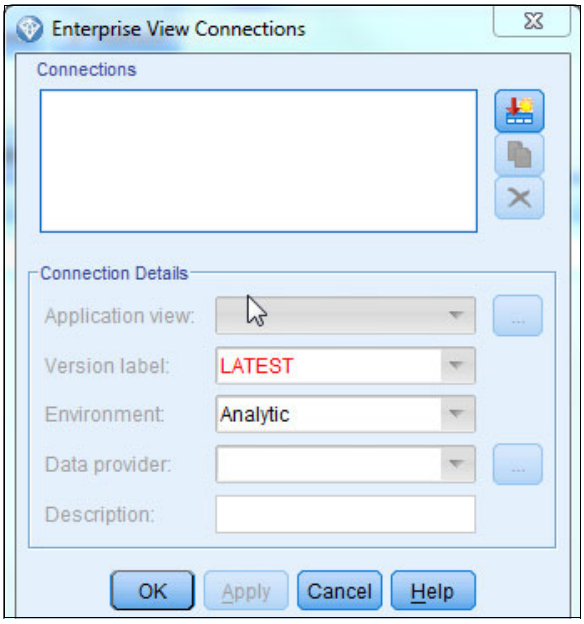


Figure 5-9 Enterprise View Connections dialog

Steps 19 - 28 on page 34 are to select the required training data for creating predictive models. The training data is represented by the EV: Risk_App_View\Risk_DPD_Analytics\Risk_Training_Data.

__ 19. Click the top-right icon (it has a red arrow in it).

__ 20. Open the **Individual Risk Assessment** folder.

__ 21. Click **Risk_App_View** to select it.

__ 22. Click **OK**.

__ 23. Set the data provider to **Risk_DPD_Analytic**.

__ 24. Check the configuration, as shown in Figure 5-10.

__ 25. Click **OK**.

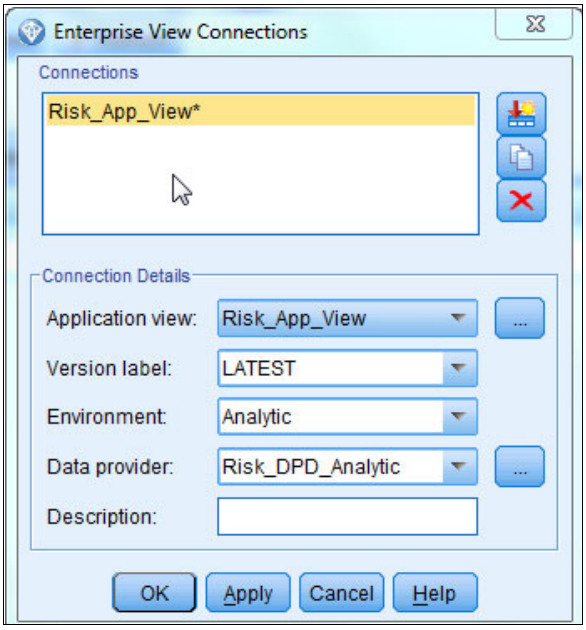


Figure 5-10 Enterprise View Connections dialog

__ 26. Press **Select** next to the Tables input.

__ 27. Click **Risk_Training_Data**, as shown in Figure 5-11 on page 35.

__ 28. Click **OK**.

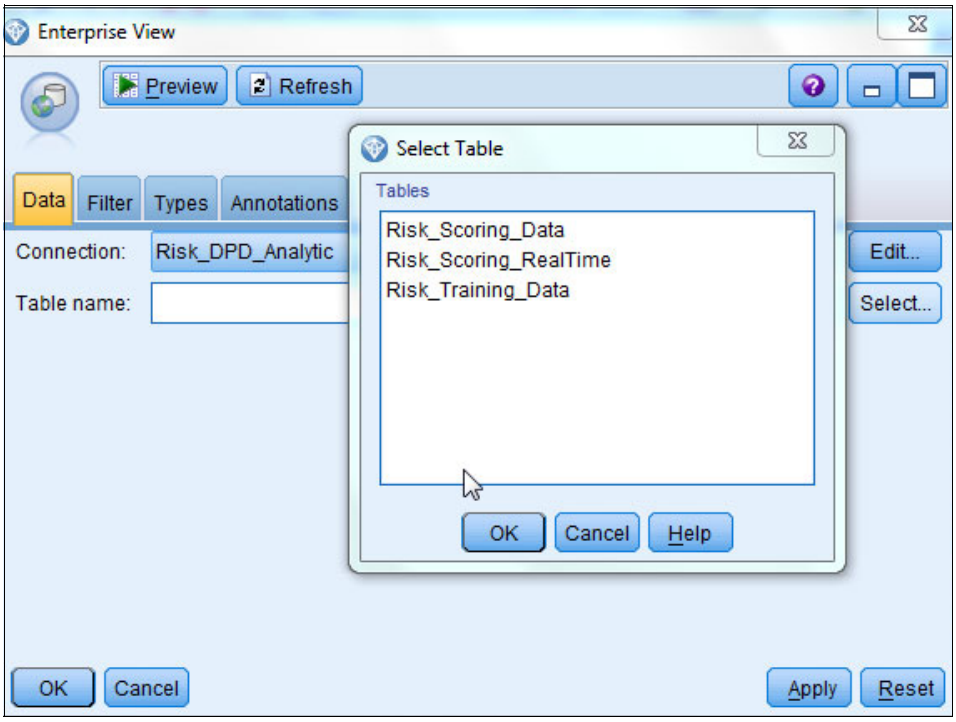


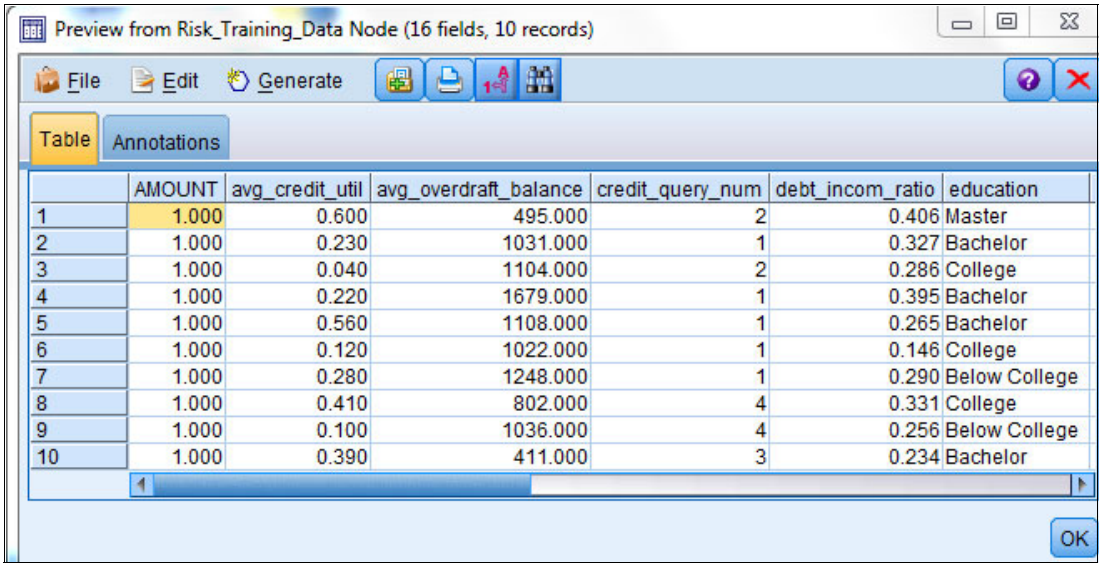
Figure 5-11 Enterprise View table dialog

Important: If you encounter a problem while connecting to the repository server, ask your lab facilitator to verify your firewall authentication.

__ 29. Now the **Connection** and **Table** are selected. Click **Preview** to view the first rows of input data.

__ 30. After examining, click **OK**. See Figure 5-12.

__ 31. Click **OK** in the Enterprise View dialog.



	AMOUNT	avg_credit_util	avg_overdraft_balance	credit_query_num	debt_incom_ratio	education
1	1.000	0.600	495.000	2	0.406	Master
2	1.000	0.230	1031.000	1	0.327	Bachelor
3	1.000	0.040	1104.000	2	0.286	College
4	1.000	0.220	1679.000	1	0.395	Bachelor
5	1.000	0.560	1108.000	1	0.265	Bachelor
6	1.000	0.120	1022.000	1	0.146	College
7	1.000	0.280	1248.000	1	0.290	Below College
8	1.000	0.410	802.000	4	0.331	College
9	1.000	0.100	1036.000	4	0.256	Below College
10	1.000	0.390	411.000	3	0.234	Bachelor

Figure 5-12 Preview the first rows of input data

__ 32. Double-click the **Partition** node to open the dialog for setting the data into two partitions: Training and Testing, as shown in Figure 5-13.

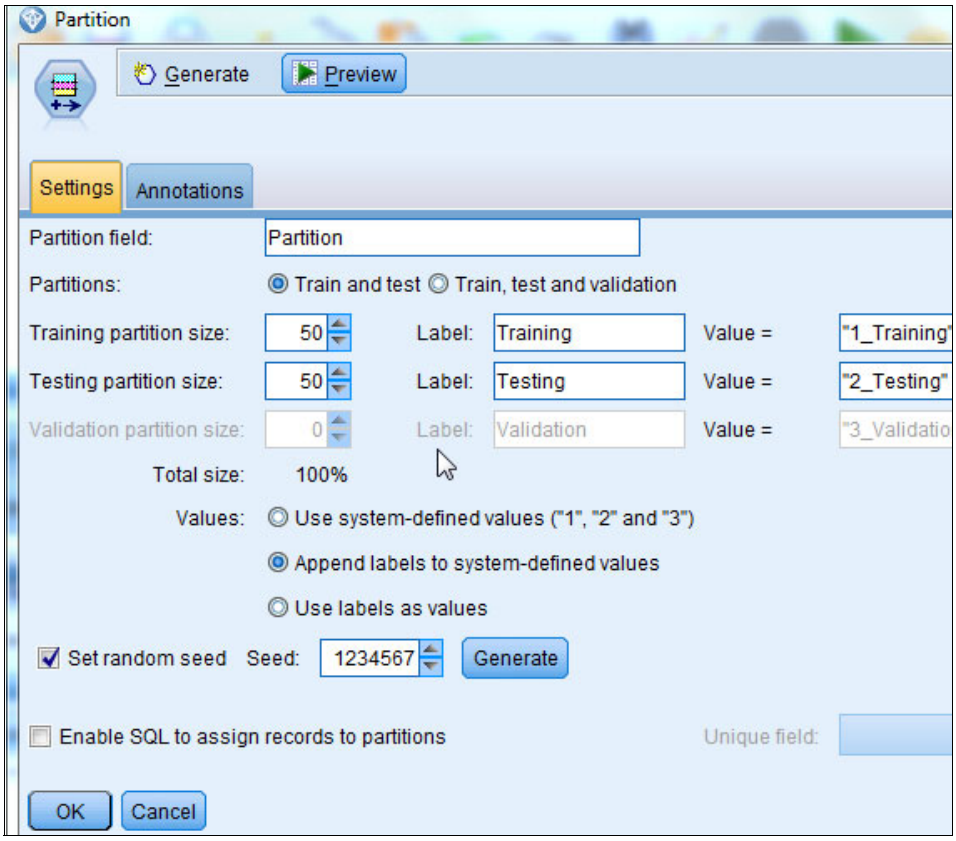


Figure 5-13 Partition node dialog

__ 33. Click **OK** to accept the default values.

Now you have the input data to be partitioned into 50% training and 50% testing. Steps 34 - 37 on page 37 enable the analyst to determine how each input data field will be used for creating predictive models.

__ 34. Double-click the **Type** node to open the dialog, as shown in Figure 5-14 on page 37.

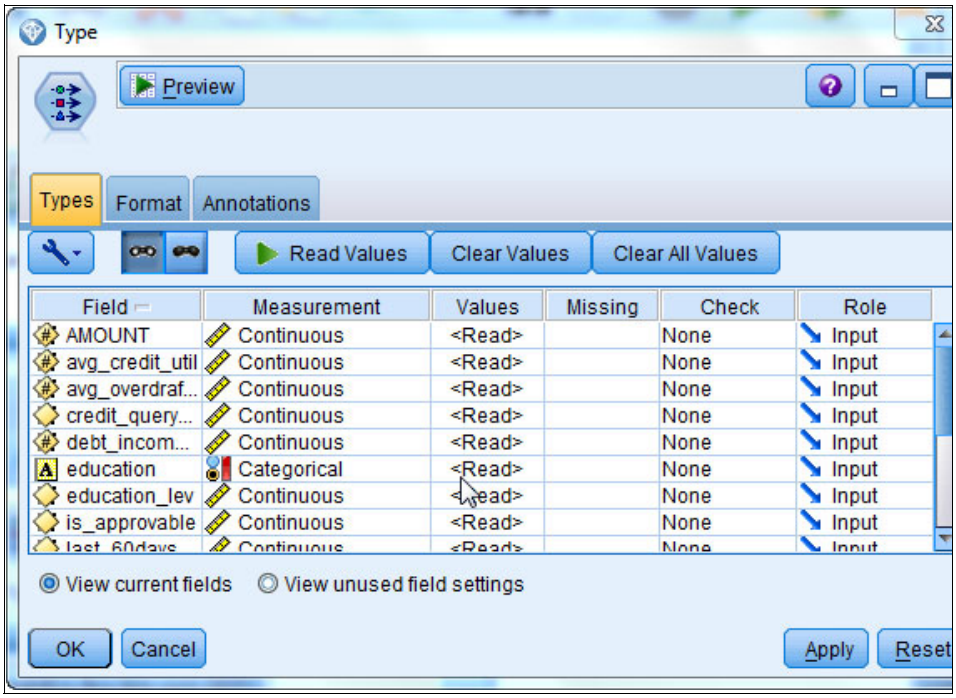


Figure 5-14 Type node dialog

__ 35. Locate the following data fields and set appropriate *Measurement* values from the available list, as shown in Figure 5-15 on page 38:

- **education** as **Typeless**
- **LOANID** as **Typeless**
- **residential** as **Typeless**
- **STATUS** as **Typeless**
- **is_approvable** as **Flag**

__ 36. For **is_approvable**, set the *Role* as **Target**.

__ 37. Click **OK**.

Note: is_approvable is the flag data field that we want to predict with the value `1` or `0` (`True` or `False`). Because other fields in the preceding list do not add value to the modeling, we ignore them by setting their roles as **Typeless**, as shown in Figure 5-15 on page 38.

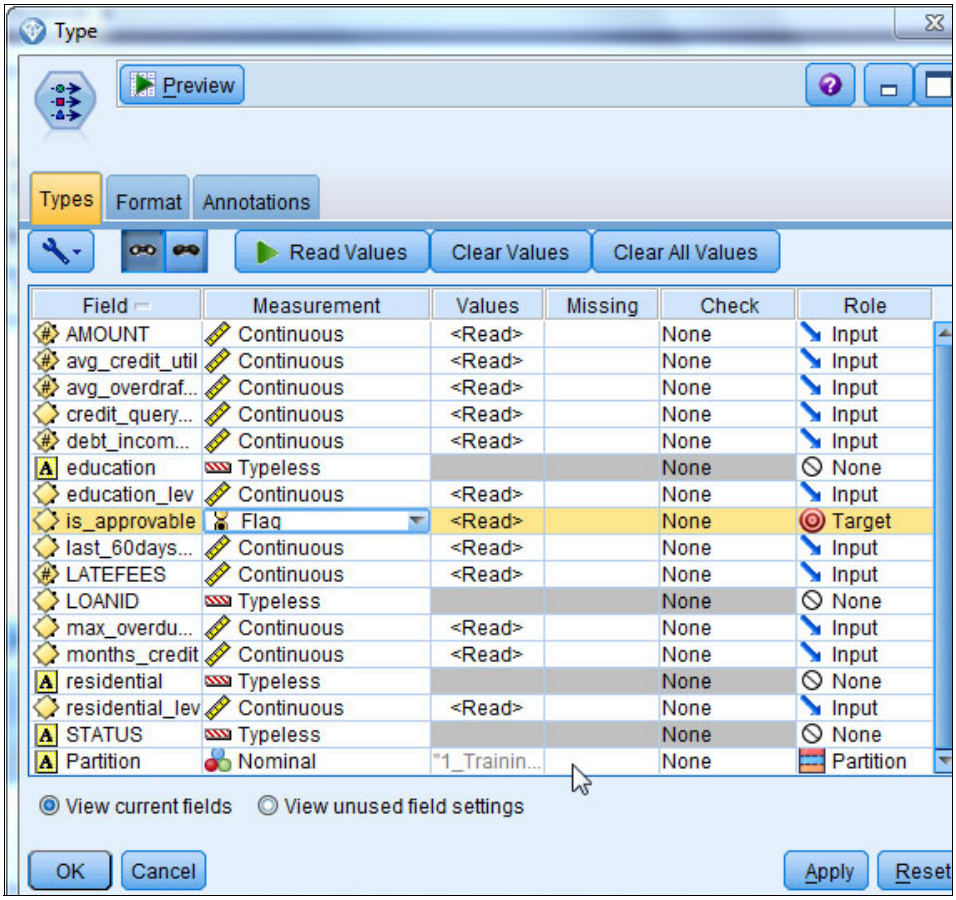


Figure 5-15 Enter Type node dialog

What you have done so far:

- In this section, you connected to an Enterprise View data source to retrieve an appropriate data view for modeling.
- Partitioned the input data for training and testing.
- Set the modeling role to each input data field.

5.4 Building a model

You should now be in a position to create a predictive model from the input data. In this section, we use the SPSS automated modeling capability so that you can use the modeler to find the best algorithm for the data. Follow these steps:

__ 1. Select the **Auto Classifier** node (under **Modeling** tab) to connect from the **Type** node, as shown in Figure 5-16 on page 39.

The *Auto Classifier* is an automated modeling node that generates prediction by combining the best ranked algorithms over the set of available algorithms.

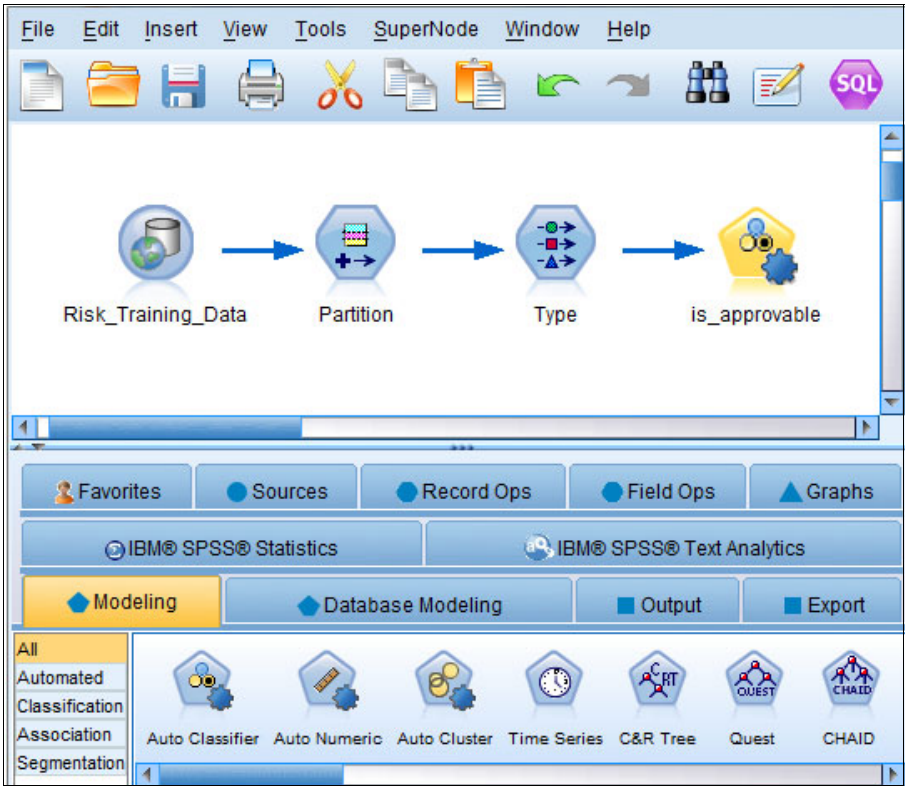


Figure 5-16 Modeler Workflow with Auto Classifier Modeling node

Note: The name of the Auto Classifier node becomes the name of the target field, in this case, `is_approvable`.

__ 2. Double-click the **is_approvable** node to open the dialog. See Figure 5-17 on page 40.

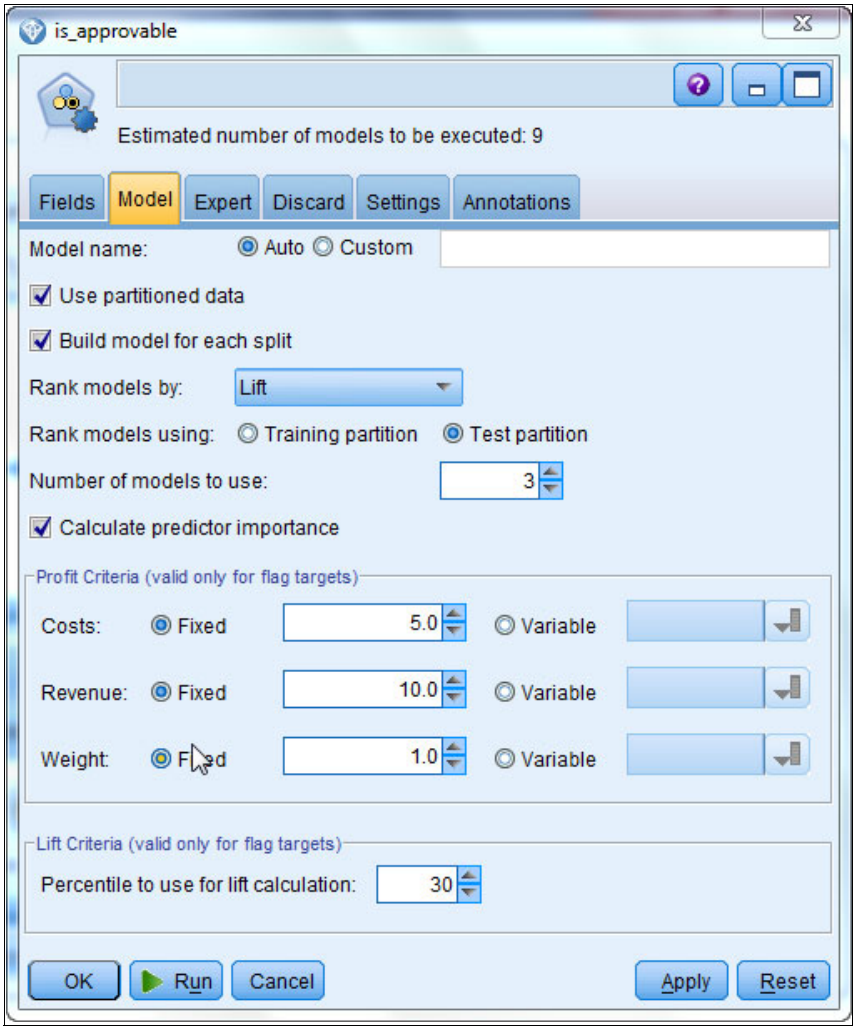


Figure 5-17 Model tab dialog of Auto Classifier

In the dialog above, the automated modeling node executes nine models. The criteria used to compare and rank models include overall accuracy, area under the receiver operating characteristic (ROC) curve, profit, lift, and number of fields. The three best ranked models are combined for the final model.

Note: In the Model tab that is shown in Figure 5-17, the Auto Classifier node uses partitioned data and retains the three best ranked models.

__ 3. Select the **Expert** tab. You see there are nine available algorithms to select from; hence, the Auto Classifier retains the best three out of nine models. See Figure 5-18 on page 41.

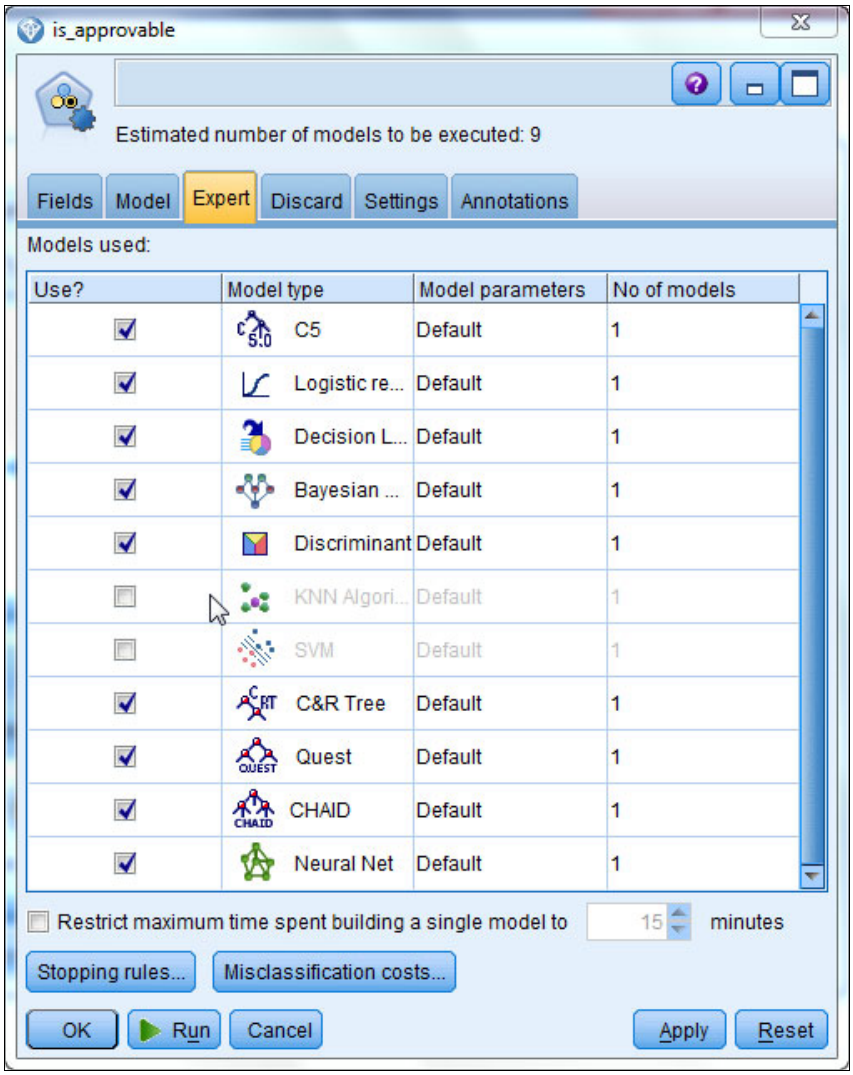


Figure 5-18 Expert tab dialog of Auto Classifier node

__ 4. Select **Run** to execute the node.

When the execution completes, the generated model appears as shown in Figure 5-19.

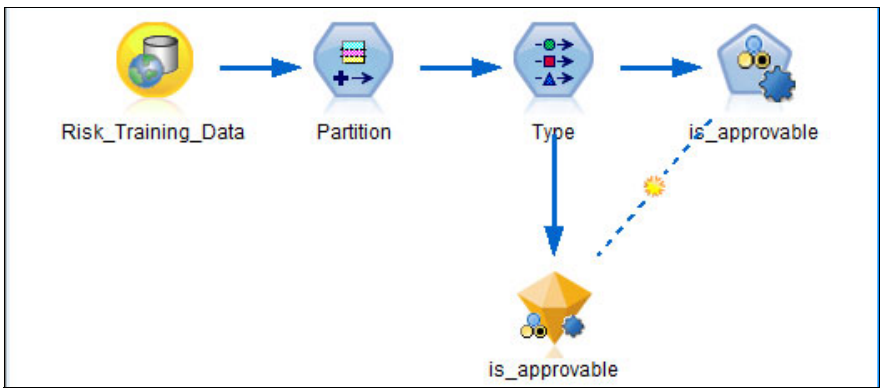


Figure 5-19 Generated Auto Classifier Model

__ 5. Double-click the generated **is_approvable** nugget node to view the modeling results. See Figure 5-20.

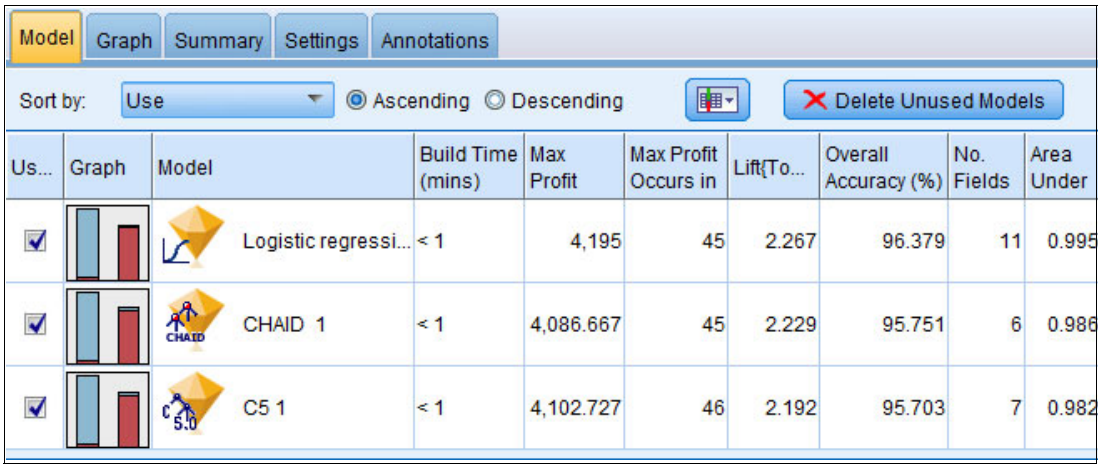


Figure 5-20 Auto Classifier modeling results

You see that the overall accuracy of the logistic regression is higher than the other two. However, it uses more data fields than the others: 11, versus 6 and 7.

Note: It is preferable for models to use fewer fields because it provides less demand for data, more efficient execution, and a more generic model. In short, you can achieve very similar results and require fewer fields.

Steps 6 - 8 are to decide how to combine the best three ranked algorithms for the final model:

__ 6. Click the **Settings** tab.

__ 7. Set the ensemble method to **Average raw propensity**.

__ 8. Click **OK**, as shown in Figure 5-21.

This method selection enables the approvable propensity score to be generated for later use in a loan risk assessment application.

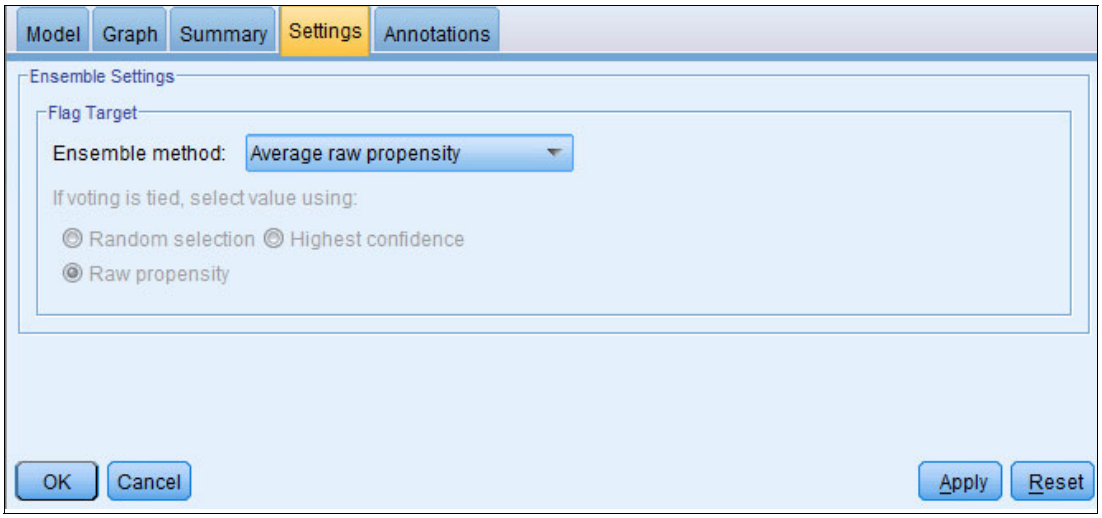


Figure 5-21 Selecting Ensemble method for automated modeling

Note: The resulted model is generated by the ensemble capability that combines the three best ranked models to obtain more accurate predictions than can be gained from any of the individual models. By combining predictions from multiple models, limitations in individual models may be avoided, resulting in a higher overall accuracy. Models that are combined in this manner typically perform at least as well as the best of the individual models, and often better.

__ 9. Click the **Output** tab.

__ 10. Double-click the **Analysis** node. This is used to evaluate the resulting model performance. See Figure 5-22.

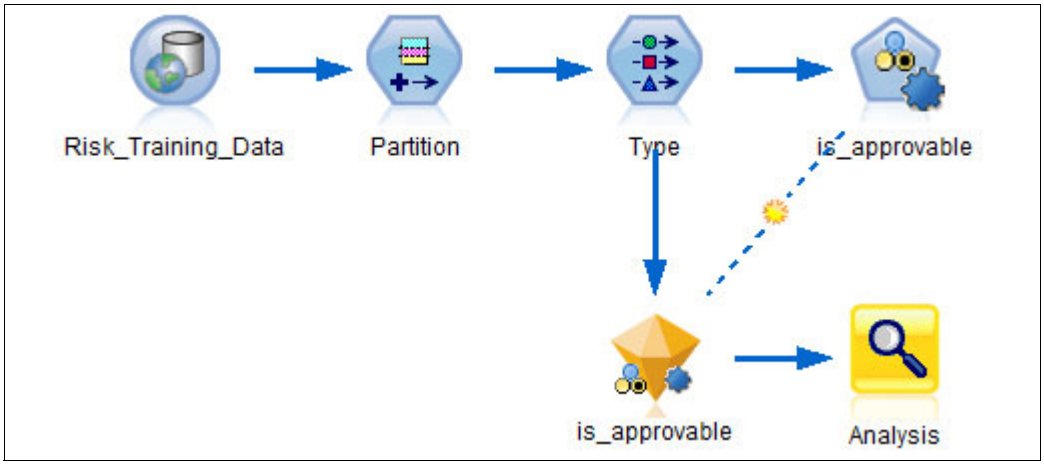


Figure 5-22 Evaluate resulting model by the Analysis node

__ 11. Double-click the **Analysis** node to open the dialog, as shown in Figure 5-23 on page 44.

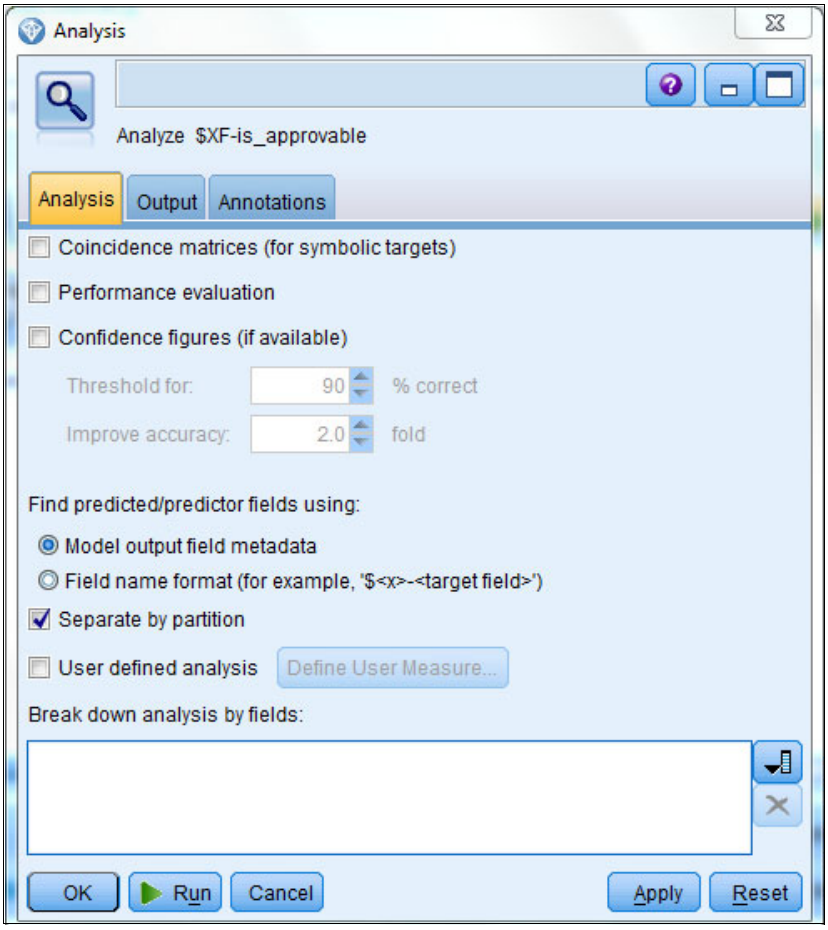


Figure 5-23 Analysis dialog

__ 12. Click **Run** to execute the node with the default values and the result, as shown in Figure 5-24.

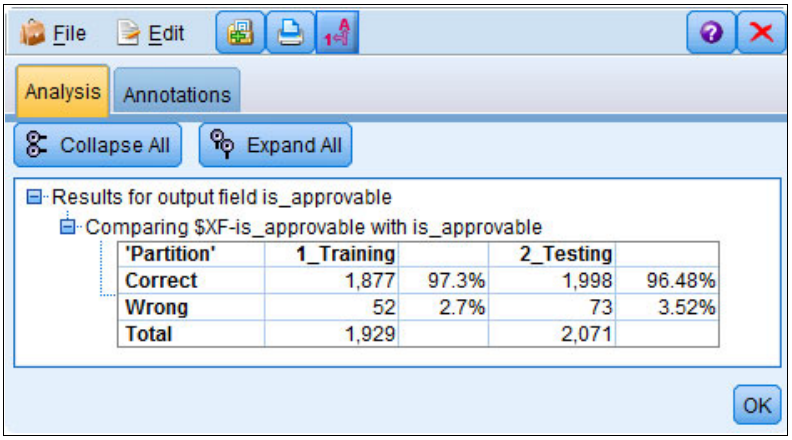


Figure 5-24 Model Analysis output

The Analysis output suggests that only 52 out of 1929 records, or 2.7% of the training data, are wrongly predicted by the model. Similarly, only 73 out of 2071, or 3.52% of the testing data, are wrongly predicted by the model.

With such low percentages of the model errors for both training and testing data, the resulted model should meet the performance requirement for the scenario.

__ 13. Click **OK**.

What you did in this section:

- In this section, you successfully created a predictive model for the sample data using the automated modeling capability via Auto Classifier node.
- You evaluated the performance of the model against the training and testing data using the Analysis node.

5.5 Deploying a model

In this section, we create a Scoring workflow that deploys the resulted model to your operational analytics environment for business applications.

5.5.1 Creating a scoring workflow

You will connect to the Scoring data view, which has all the data fields as the Training data view except the target field (`is_approvable`). The previous generated model from the Training workflow will then be used to process the input scoring data and then output to a Table node.

Steps 1 - 13 on page 33 are performed to select the required scoring data for executing resulted predictive models. The scoring data is represented by the EV: Risk_App_View\Risk_DPD_OP\Risk_Scoring_Data.

__ 1. Click the **Sources** tab.

__ 2. Double-click the **Enterprise View** data source to create a new one. This puts the Enterprise View somewhere on the canvas. If it is in the middle of the flow, you can click and drag it to somewhere convenient.

__ 3. Double-click it to open the dialog.

__ 4. Select to add a new Enterprise View connection by clicking the **Connection** drop-down box and selecting **Add/Edit a Connection**.

Figure 5-25 on page 46 shows how to add a new connection.

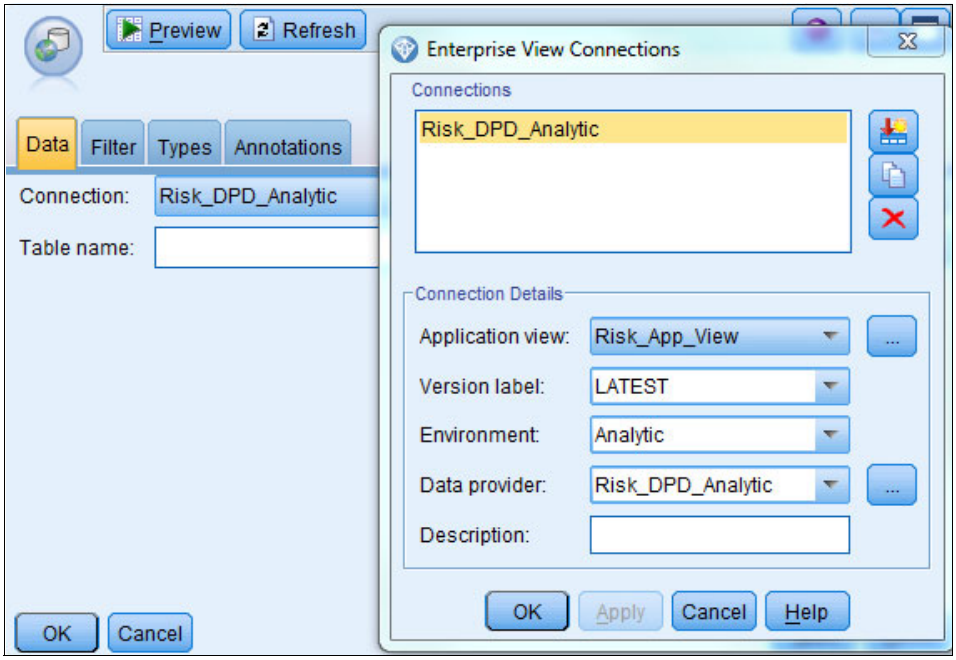


Figure 5-25 Add a new connection

__ 5. Click **Risk_DPD_Analytic** to select it.

__ 6. Set the **Application view** to **Risk_App_View**.

__ 7. Set the **Environment** to **Operational**.

__ 8. Set the **Data Provider** to **Risk_DPD_OP**, as shown in Figure 5-26.

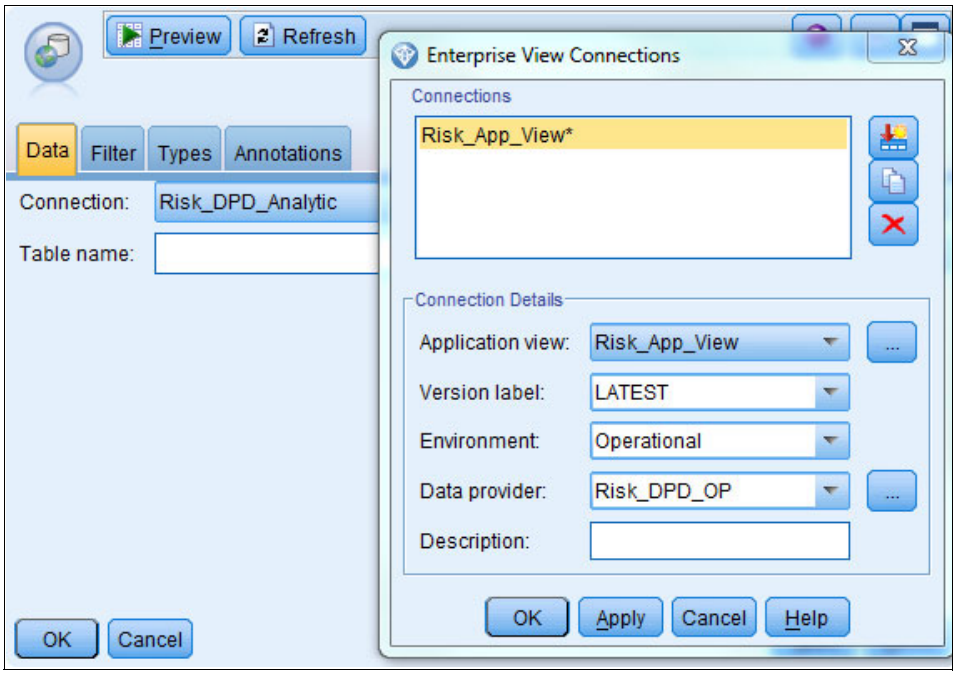


Figure 5-26 Add an Operational data view

__ 9. Click **OK**.

__ 10. Click the **Select** button next to the Tables input.

__ 11. Click **Risk_Scoring_Data** to select it.

__ 12. Click **OK** in the Tables view. See Figure 5-27.

__ 13. Click **OK**.

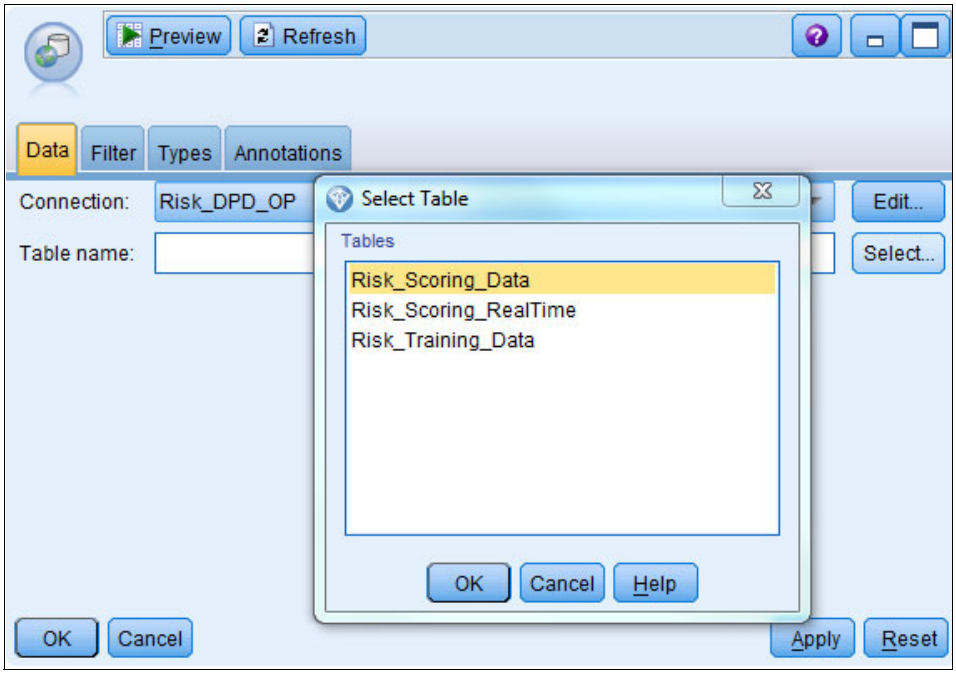


Figure 5-27 Select Table name

In steps 14 - 21 on page 48, you copy the resulted predictive model to include in the stream for scoring:

__ 14. Right-click the **is_approvable** nugget node.

__ 15. Click **Copy Node.**

__ 16. Right-click the *Canvas.*

__ 17. Click **Paste** to paste the nugget into the canvas.

__ 18. Drag the pasted node next to the Enterprise node that you just created.

__ 19. Right-click the Enterprise node that you just created.

__ 20. Click **Connect**.

__ 21. Now click the copied **is_approvable** node to connect, as shown in Figure 5-28.

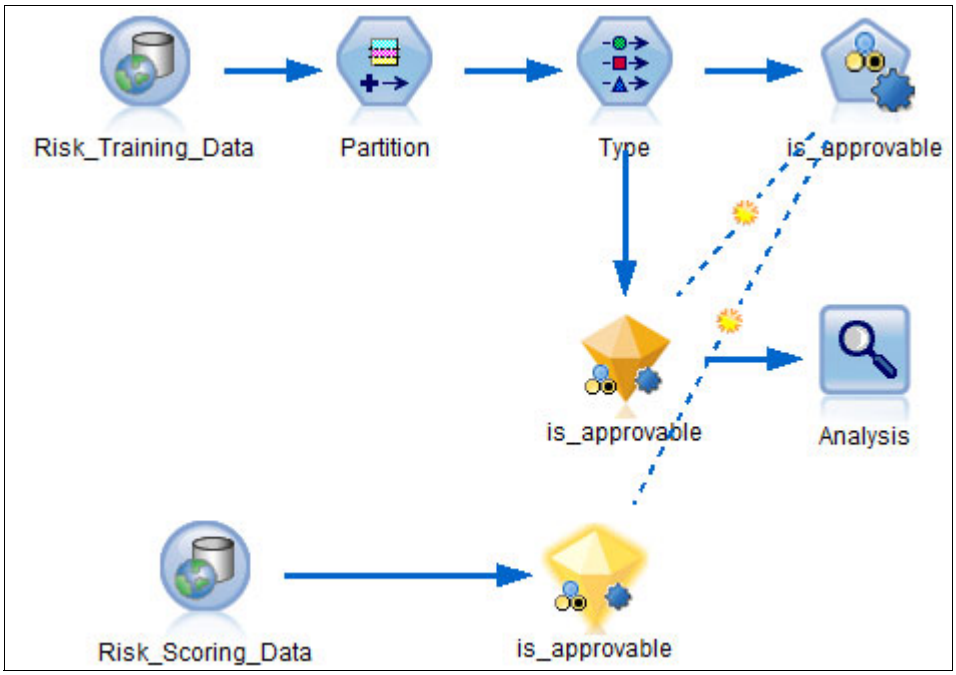


Figure 5-28 Create scoring workflow

__ 22. Click the **Output** tab.

__ 23. Click the copied **is_approvable** node to select it.

__ 24. Double-click the Table node. This connects to the **is_approvable** node, as shown in Figure 5-29.

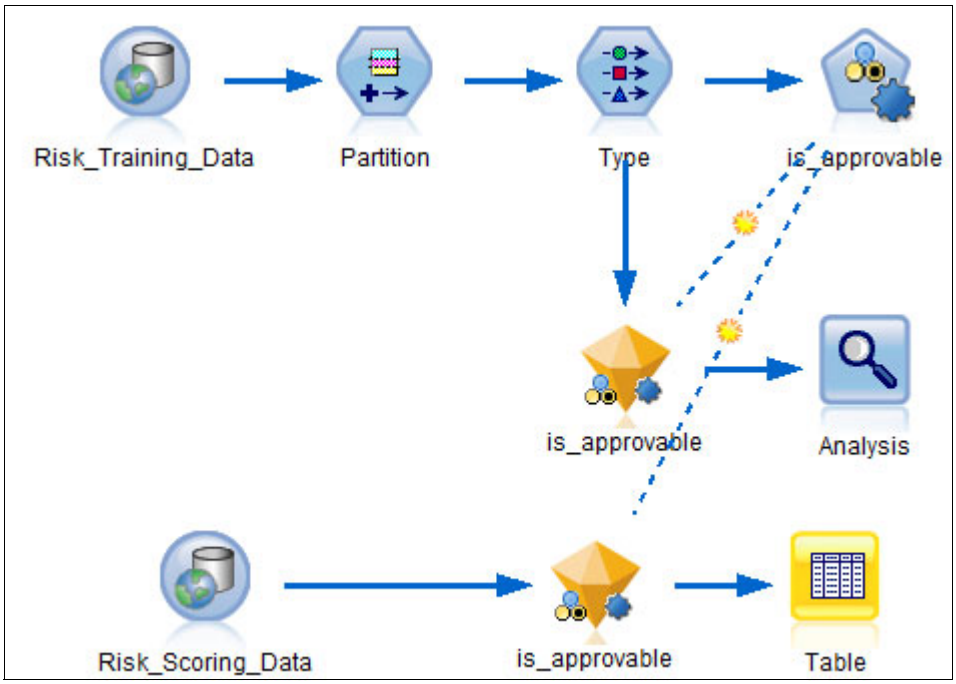


Figure 5-29 Add Table node to the scoring workflow

In steps 25 - 28, you execute the data to produce the predictive results that are captured in the data field $**XFRP-is_approvable** with values being close to “one”. This value indicates that it is more likely to be approved, and being close to “zero”, indicates that it is more likely to be rejected.

__ 25. Right-click the **Table** node to select **Run**.

__ 26. As the Scoring Table results open, scroll to the right side of the table to see the results, as shown in Figure 5-30.

__ 27. Click **OK** after you examine the table results.

LOANID	max_overdue...	months_credit	residential	residential_lev	...	$XF-is_approvable	$XFRP-is_approvable
1	6	60	Rental	2	...	0	0.112
4	2	63	Own	3	...	1	0.708
4	0	64	Other Ca...	1	...	1	0.985
1	0	68	Own	3	...	1	0.824
4	1	84	Own	3	...	1	0.985
4	0	44	Own	3	...	1	0.987
1	0	49	Own	3	...	1	0.987
4	0	78	Rental	2	...	1	0.877
1	2	16	Rental	2	...	1	0.652
1	0	75	Other Ca...	1	...	1	0.986
1	0	72	Own	3	...	1	0.987
1	0	42	Other Ca...	1	...	1	0.870
4	0	48	Own	3	...	1	0.951
3	0	43	Own	3	...	1	0.986
1	0	110	Rental	2	...	1	0.988
1	0	48	Own	3	...	1	0.958
1	1	22	Own	3	...	1	0.925
3	1	24	Own	3	...	1	0.797
4	0	30	Rental	2	...	1	0.961
4	0	35	Other Ca...	1	...	1	0.531

Figure 5-30 Scoring output

$XF-is_approvable value is the prediction of whether a loan is approvable.
$XFRP-is_approvable value is the calculated propensity for approving loans, which will be used for assessing an individual loan in the business application.

__ 28. Click **OK**.

5.5.2 Deploying scoring workflow

In this section, you deploy the Scoring workflow to your analytical platform. Follow these steps:

__ 29. Click **File**.

__ 30. Click **Store**.

__ 31. Click **Deploy**, as shown in Figure 5-31 on page 50.

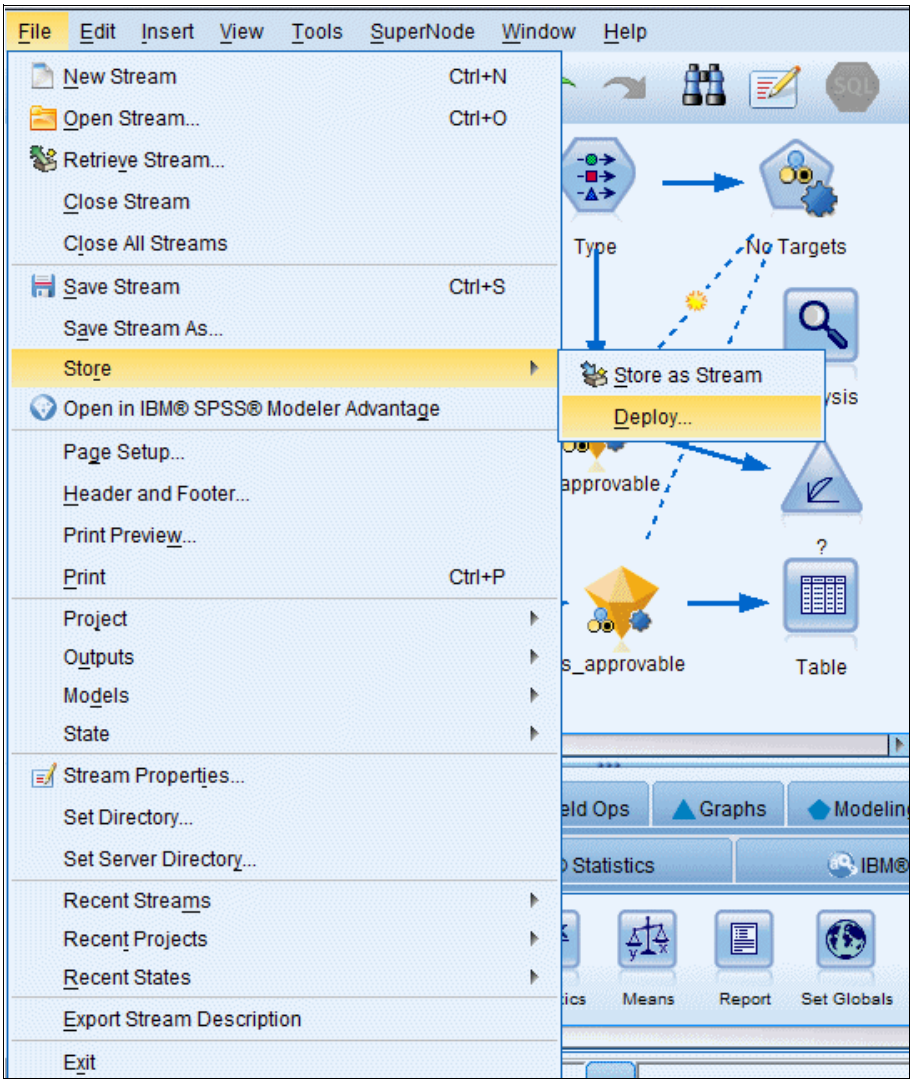


Figure 5-31 Deploy modeler stream to repository

In steps 32 - 39 on page 52, you deploy the scoring stream into the repository for operational application in the enterprise environment. When it is done, the predictive model can be used in a batch operation for scoring batch input transactions or in a real-time application for responding to an individual scoring request on demand.

__ 32. Select **Deployment** type as **Scoring Only**.

__ 33. Set the Scoring node as **Table**.

__ 34. Select **Deploy as stream**, as shown in Figure 5-32 on page 51.

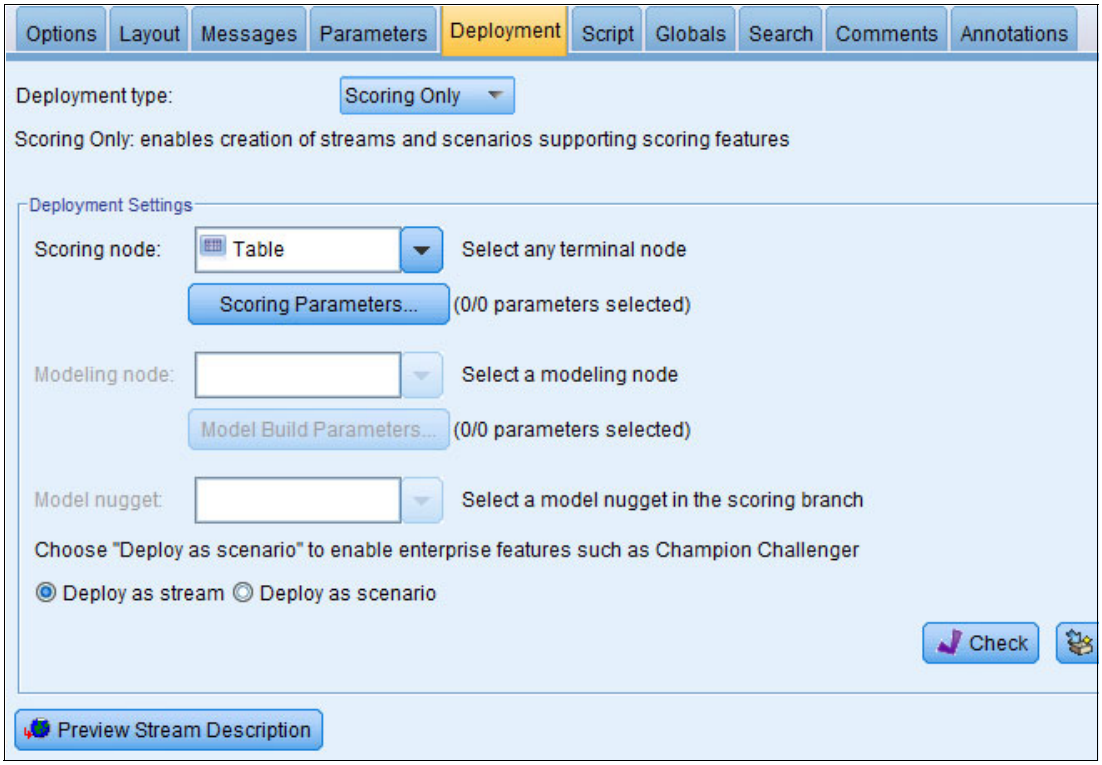


Figure 5-32 Deployment dialog

__ 35. Click **Store**.

__ 36. Go to your *Group Folder*, which is defined in your worksheet. See Figure 5-33 on page 52.

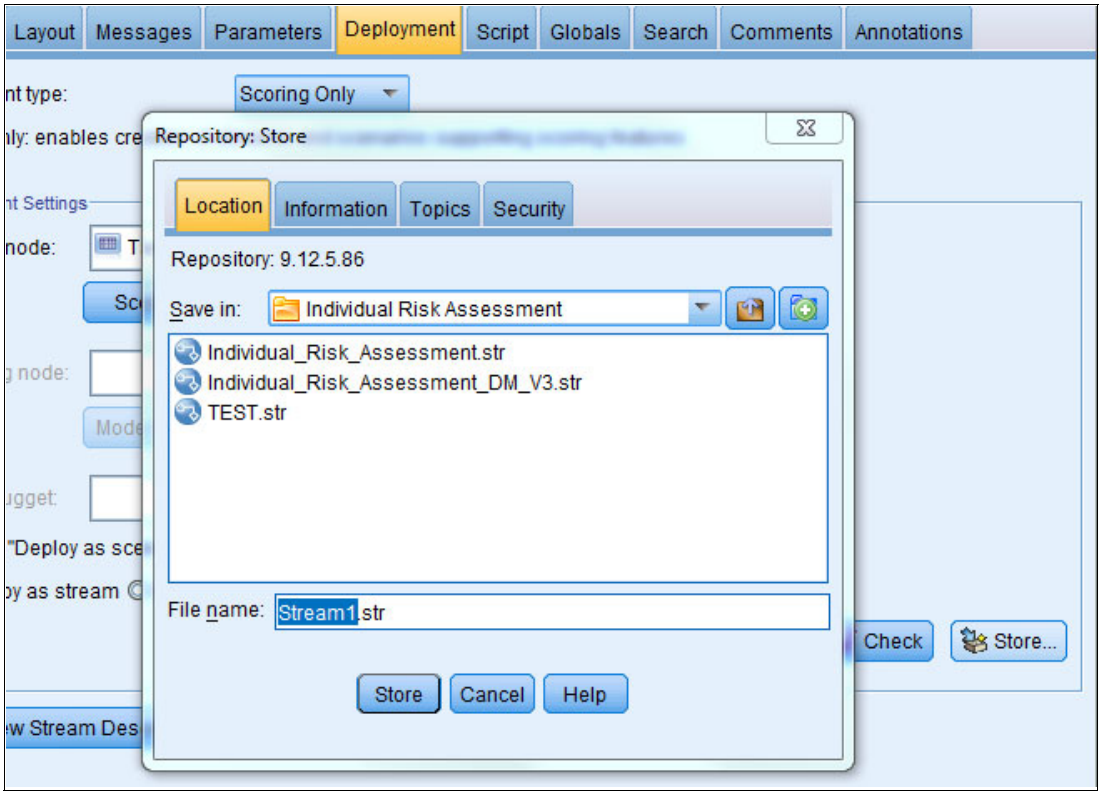


Figure 5-33 Store model stream into Repository

__ 37. Name the stream the *Modeler Stream Name* from your worksheet.

__ 38. Click **Store**.

__ 39. Click **OK** to complete the model deployment.

Congratulations, you have completed the lab.

__ 40. Select **File**.

What you did in this section:

- In this section, you successfully created a scoring workflow that uses your newly created model.
- You deployed the scoring workflow to your analytics platform to be ready for operational application.
- This model will be used in the next lab.

41. Select **Exit**.

42. Click **Exit**. You do not need to save the model because you deployed it to the repository.

5.6 Summary

You have seen how you can easily access data and build a good model, and then deploy it to be ready for use in operational application in the enterprise.

6

Configure the risk assessment in SPSS Decision Management

The lab guides you through a structured decision making with the seven steps of SPSS Decision Management (DM) to combine decision with advanced analytics.

6.1 Introduction to SPSS Decision Management

The end goal of advanced analytics is to be able to deploy and apply analytics insights at the very point of impact. In our lab scenario of assessing online loan applications, it means that the retail bank is able to gather all available intelligence to make decisions in real time.

There can be potentially thousands of incoming online applications across different types of loans and loan amounts. The bank needs to make a decision on whether to approve or reject an individual loan application in a short time. These decisions are operational and repeatable in nature and may require rules allowing rapid reaction to any market situation. The data for assessing a loan comes from a variety of sources, which may be used to augment the existing risk metrics that the bank uses. An example is FICO scores, with their own data-driven analytics and rules.

To handle such challenges, IBM SPSS Decision Management (SPSS DM) enables the automation of such high volume, repeatable decision-making by combining business analytics, business rules, optimization, and data management. SPSS DM drives the power of analytics into optimized decisions in the hand of business by combining analytics results with business knowledge. By providing completely configurable templates such as risk assessment, SPSS DM provides a business tool in a language that the business can understand.

6.1.1 Prerequisites

- You should have access to the client workstation where you will be performing the lab exercise. Check with your lab facilitator to get access.
- You should have received your team number from the lab facilitator. You will be using this number during the lab to determine which user ID and which CICS region to use, among other things.
- You should have already completed the modeling lab.

6.1.2 Create your SPSS Decision Management project

This section is to configure a Decision Management specification that combines business rules and predictive models that are created in the previous lab to arrive at an optimal recommendation for incoming loan applications.

Log on to SPSS Decision Management

The first thing to do is to log on to SPSS Decision Management. From Decision Management, you then perform the steps required to take an SPSS Modeler Model and business rules to create a new stream:

__ 1. Go to the SPSS Decision Management by entering the *Decision Management URL* from your worksheet.

> **Note:** The URL is case-sensitive.

__ 2. You should now see a login page very similar to Figure 6-1 on page 55.

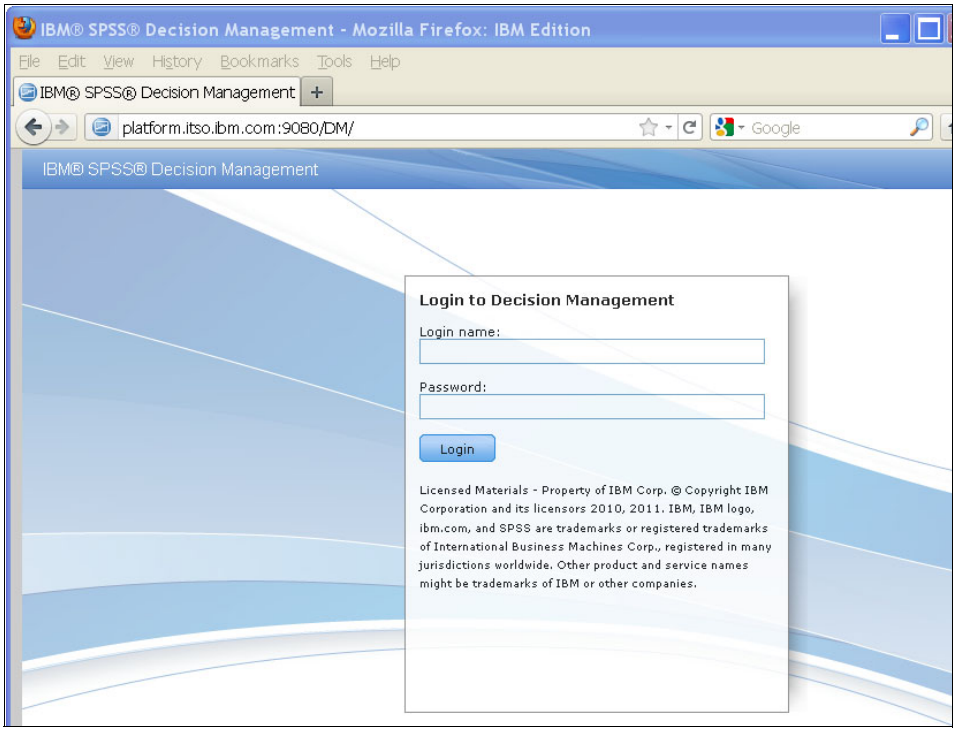


Figure 6-1 SPSS Decision Management login page

__ 3. Enter the *DM User* as mentioned in your worksheet.

__ 4. Enter the *DM Password* as mentioned in your worksheet.

The launch page appears to display configurable solution templates, which are shown in Figure 6-2. SPSS provides a number of vertical configurable solution templates. In this lab, we use the template for assessing credit risk and recommending the best action for incoming loan applications.

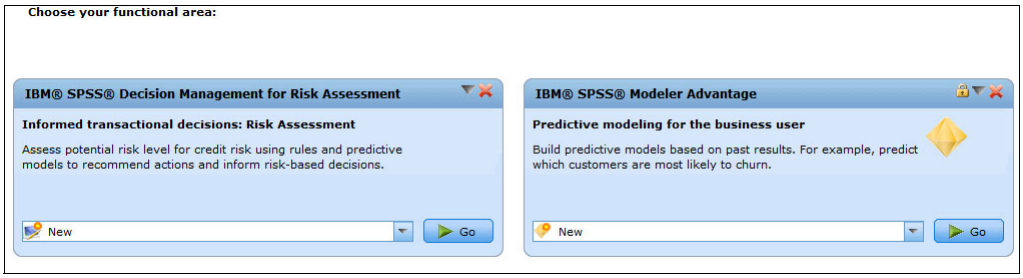


Figure 6-2 SPSS Decision Management launch page

On the **IBM SPSS Decision Management for Risk Assessment** template (shown in Figure 6-3 on page 56) perform the following steps:

__ 5. On the drop-down box, select **New**.

__ 6. Click **Go**.

Figure 6-3 shows the SPSS Decision Management for Risk Assessment portal.

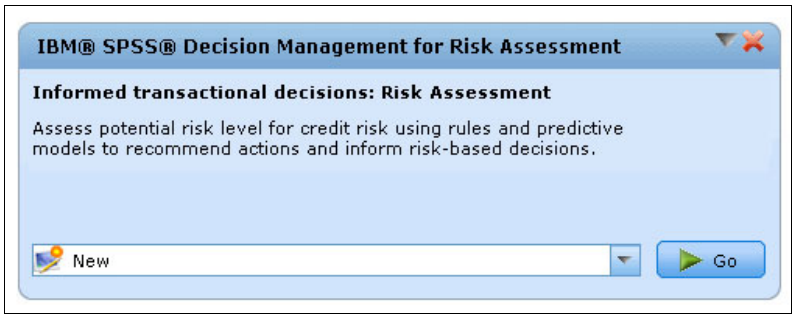


Figure 6-3 SPSS Decision Management for Risk Assessment portal

A flowchart appears as shown in Figure 6-4. The application template provides a step-by-step workflow, as represented by the icons on the home page. Click any icon to jump to that step.

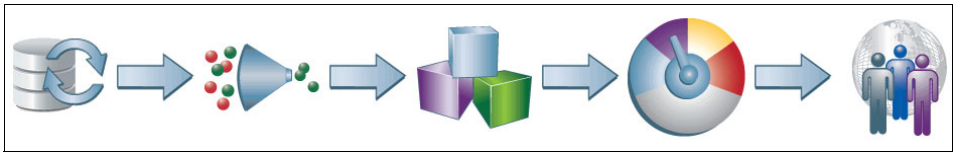


Figure 6-4 Steps in Risk Assessment Configuration

Figure 6-5 provides a summary of steps in a structured decision making for configuring risk assessment application.

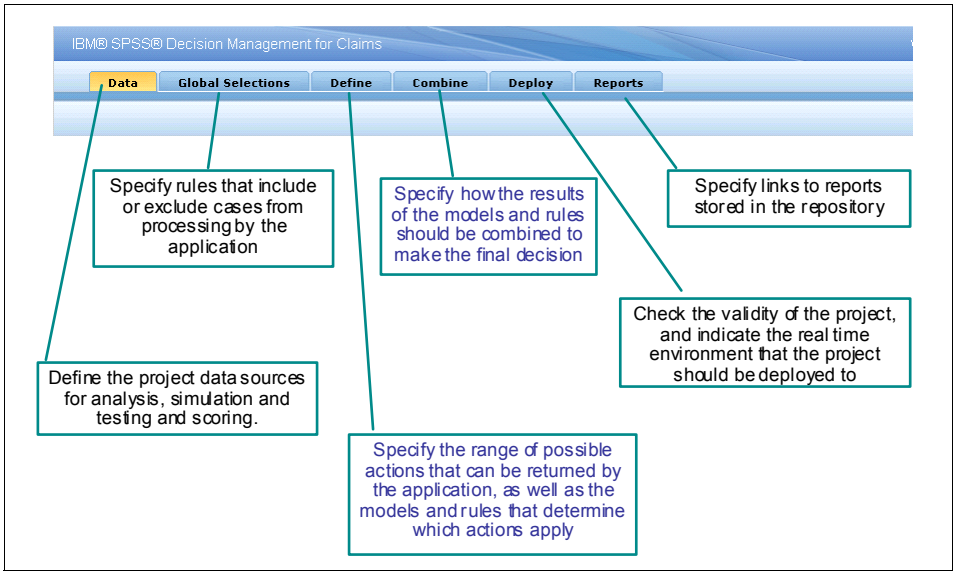


Figure 6-5 Structured decision making steps in the SPSS DM configurable solution template

Step 1: Connect to data

Take the following steps to connect to the data.

__ 1. Click the Data icon to connect to data source. See Figure 6-6 on page 57.

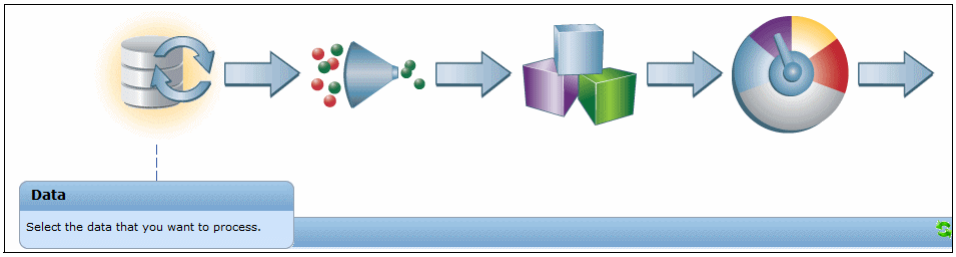


Figure 6-6 Connect to data

The Data section opens for entering details, as shown in Figure 6-7.

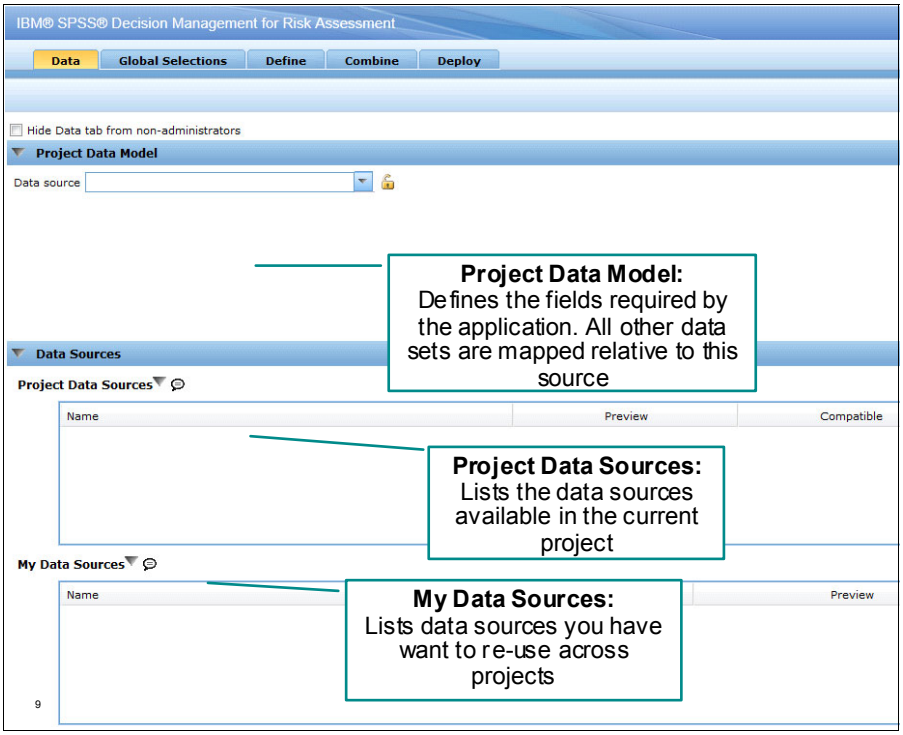


Figure 6-7 Data tab

You use the scoring data view that includes all data elements that are specifically required for operational risk assessment scoring. Follow these steps:

__ 2. Under **Project Data Model**, use the drop-down menu to select **Add a data source...**.

__ 3. Type the **DataSource name** to **RiskDataSource**.

__ 4. Select the radio box **Enterprise View** for the data source type.

__ 5. Next to Application View, click the **Browse** button.

__ 6. Navigate into the Individual_Risk_Assessment folder by clicking it.

__ 7. Select **Risk_App_View**.

Risk_App_View is the application view that includes all data elements that are required for Risk Assessment application. We have already defined this data view for expediency.

__ 8. Click **Open**.

__ 9. Select the *Table* **Risk_Scoring_Data**.

Risk_Scoring_Data is the Enterprise View table that specifically includes all data elements that are required for the operational scoring in this scenario.

__ 10. Select the *Data Provider* **Risk_DPD_OP**.

Risk_DPD_OP is the data provider that populates all required data elements for the operational scoring in this scenario.

__ 11. Check that the Environment field is set to *Operational*.

__ 12. Expand the **Specify Input Fields** toggle.

__ 13. Check that all the fields are selected.

__ 14. The configuration should look similar to Figure 6-8.

__ 15. Click **Save**.

__ 16. Click the **Save** icon**,** in the toolbar at the top**.** If you have not done this already, save the project in your *Group Folder*, which can be found in your worksheet. Call the project the *Decision Management Project Name* that is specified in the worksheet.

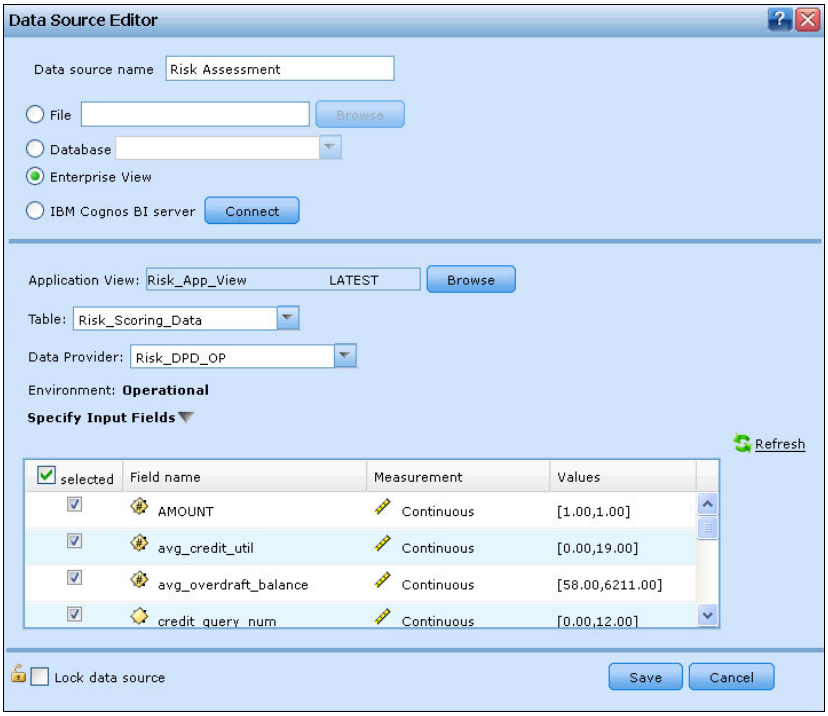


Figure 6-8 Data source configuration

__ 17. Check **Mark as Done** and click the *blue arrow* to go to the next step.

Step 2: Global Selections

The Global Selections dialog opens, as shown in Figure 6-9 on page 59.

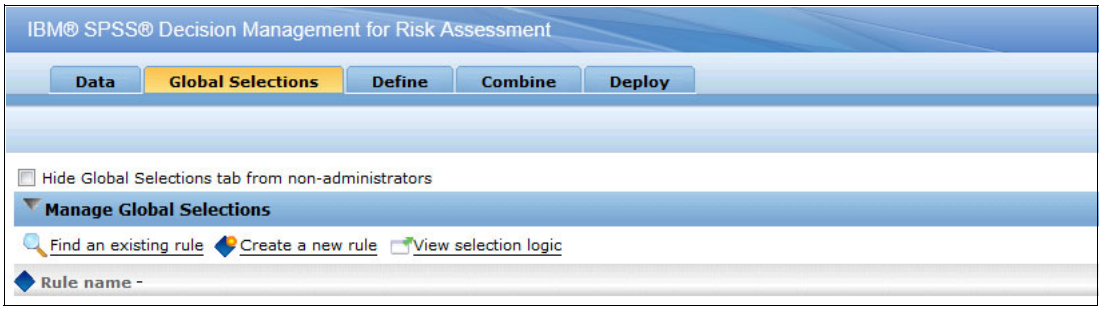


Figure 6-9 Global Selections dialog

This section allows business users to define the decision scope using *Global Selections*. The user sets the ground rules for the type of decision that will be rendered. At this point, exclusion or inclusion rules can be defined. For example, the user can include only those loan applications that are less than $500 K US. In this lab, we do not define such rules at this step, for simplicity.

__ 1. Check **Mark as Done** and click the *blue forward arrow* to go to the next step.

Step 3: Define outcomes

In this step, you define three possible actions for incoming transactions: approve, reject, or investigate.

The Define dialog opens as shown in Figure 6-10.

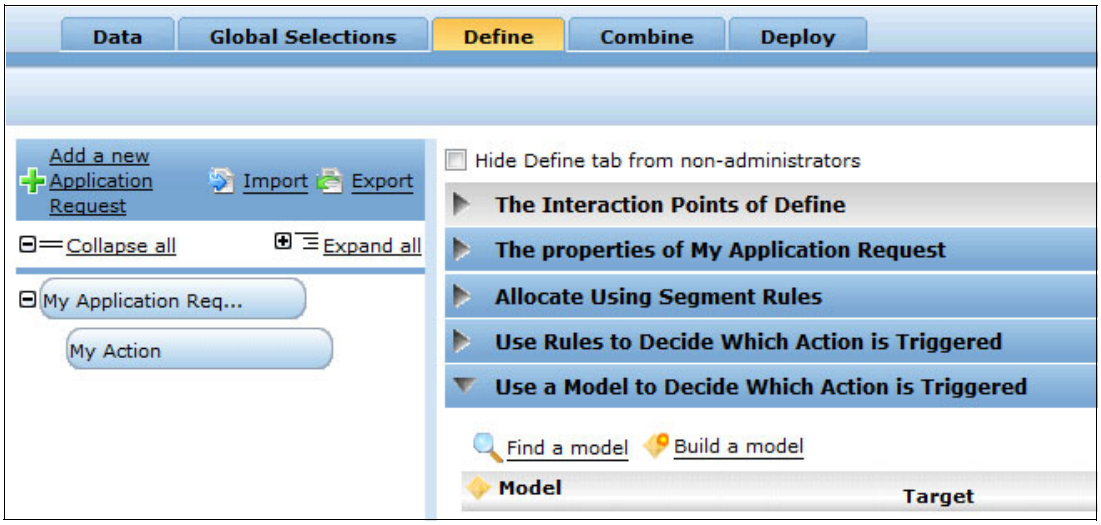


Figure 6-10 Define dialog

This section allows business users to define desired outcomes. The result of every decision is an outcome. The business defines what outcomes are expected in given conditions. In this lab, you define three operational decision outcomes for the application: approve, reject, or investigate.

__ 1. Look in the left panel. You see a panel that is used for defining decision outcomes, as shown in Figure 6-11 on page 60.

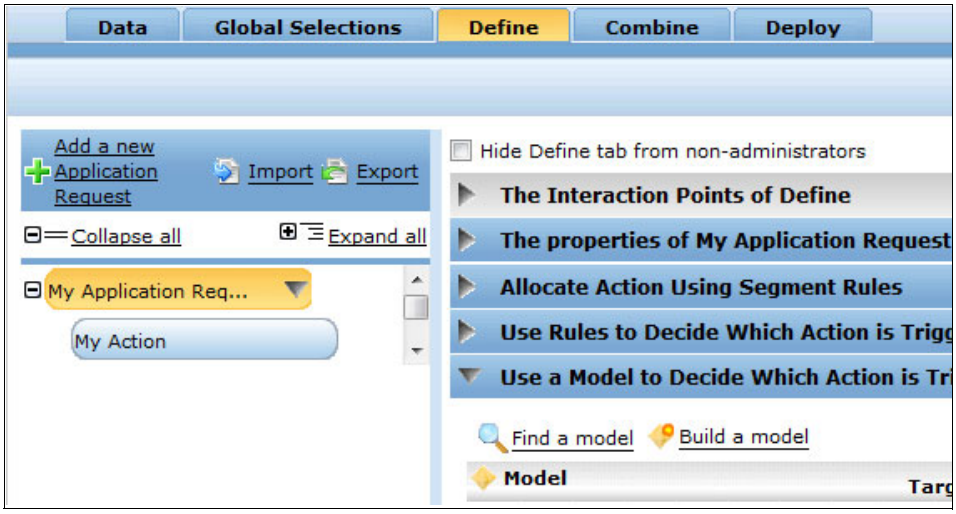


Figure 6-11 Define Decision Outcomes

__ 2. Click My Action.

__ 3. Click the triangle button.

__ 4. Select **Duplicate**, as shown in Figure 6-12**.**

__ 5. Click the triangle button again.

__ 6. Select **Duplicate** again, as shown in Figure 6-12.

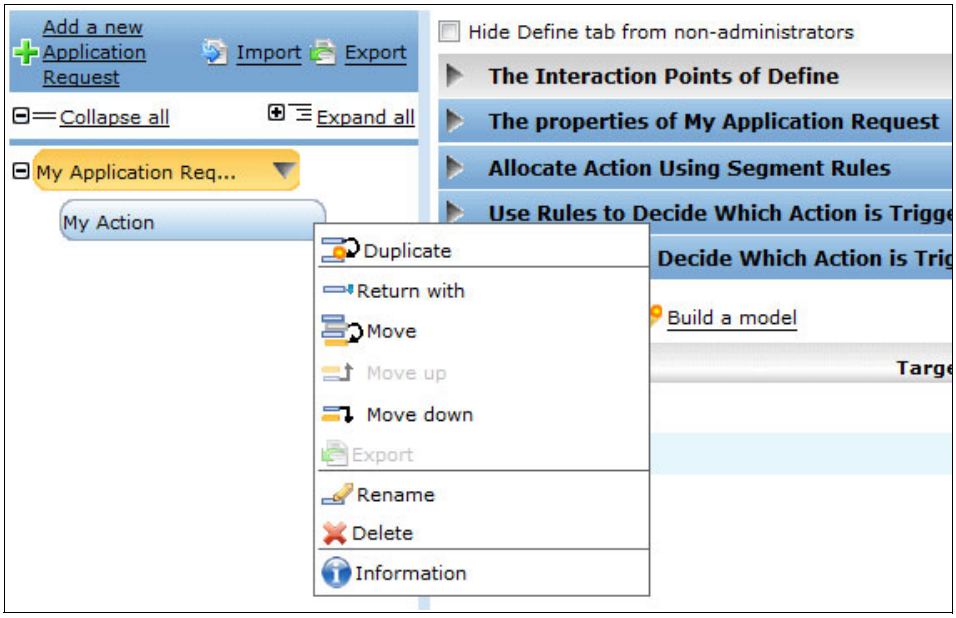


Figure 6-12 Duplicate actions

__ 7. You now have three actions, but you have to rename them. To rename each action:

- Click **My Action** to select it.
- Click the triangle button.
- Select **Rename**.

Figure 6-13 shows all three decision outcomes.

.

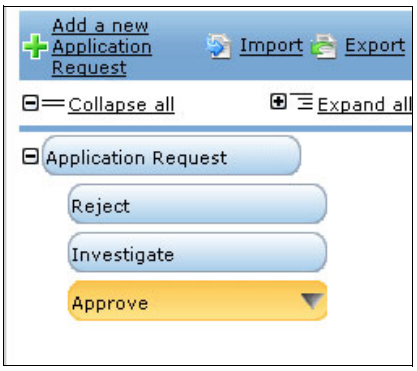


Figure 6-13 All three outcomes are now created

__ 8. Click the **save** icon in the top right. If you have not done this already, save the project in your *Group Folder*, which can be found in your worksheet. Call the project the *Decision Management Project Name* that is specified in the worksheet.

Step 4: Define operational decisions with rules and models

In this step, you allocate actions with rules and predictive models. The risk analyst configures the decision space by indicating what actions each loan application will be routed to.

__ 1. Click to select **My Application Request**, as shown in Figure 6-14.

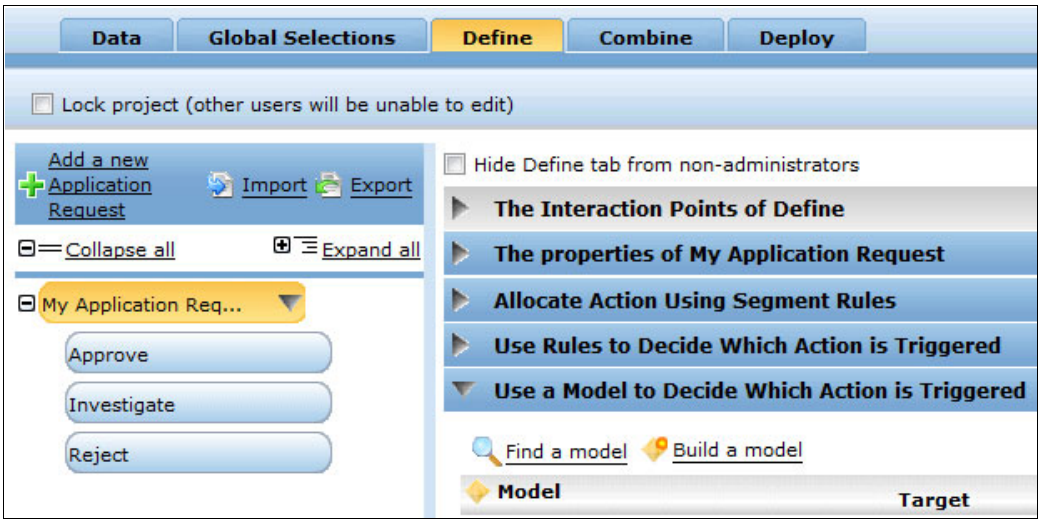


Figure 6-14 Define operational decisions with rules and models

In this lab, you will use different techniques:

- Allocate Action Using Segment Rules
- Use Rules to Decide Which Action is Triggered
- Use a Model to Decide Which Action is Triggered

In steps 2 on page 62 through 11 on page 63, you allocate actions by *Segment Rules*, which are high-value rules that will allocate an applicant to a particular action immediately and bypass any other allocation.

__ 2. Expand **Allocate Action Using Segment Rules** to open the details section, as shown in Figure 6-15**.**

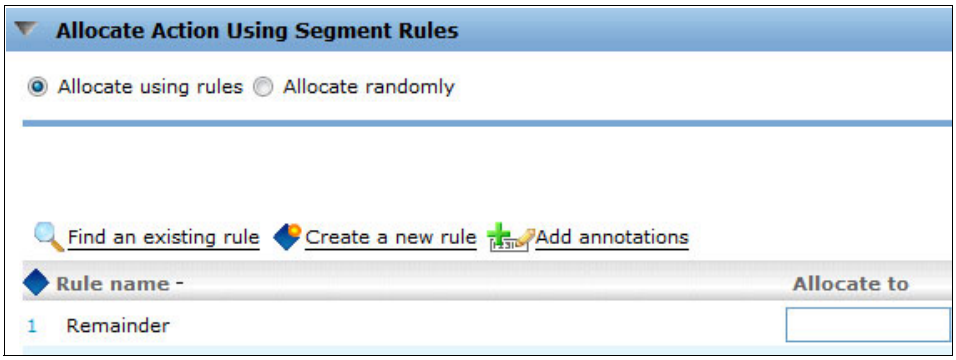


Figure 6-15 Allocate Action Using Segment Rules section

Instead of defining rules from scratch, we use the predefined rules for expediency.

In steps 3 through 8 on page 63, you define the Low_Values rule that seemed to be highly indicative of risk. If the applicant fits that rule, the loan will be rejected. This allocation provides the risk analyst with total control over his own risk-based processes.

__ 3. Click **Find an existing rule**.

__ 4. Select **Low_Values.rul** from the repository, as shown in Figure 6-16.

__ 5. Click **Open**.

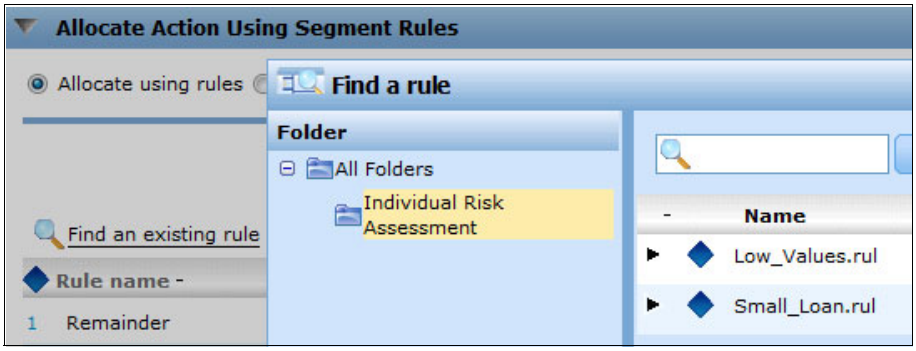


Figure 6-16 Retrieve existing rules from the repository

__ 6. Select the **Allocate To** drop-down list of Low_Values to select the action, **Reject**.

This rule means that if the applicant has a low education level (<1) and has a maximum overdue number greater than 3, the assessment system automatically rejects the loan.

Note: The rules that are used in the lab are made up for illustrating the technology with no reflection on actual bank rules. You can change data attributes and values to create your own custom rule.

__ 7. Click **Low_Values** to open the rule preview. This allows you to see the rule in more detail. See Figure 6-17.

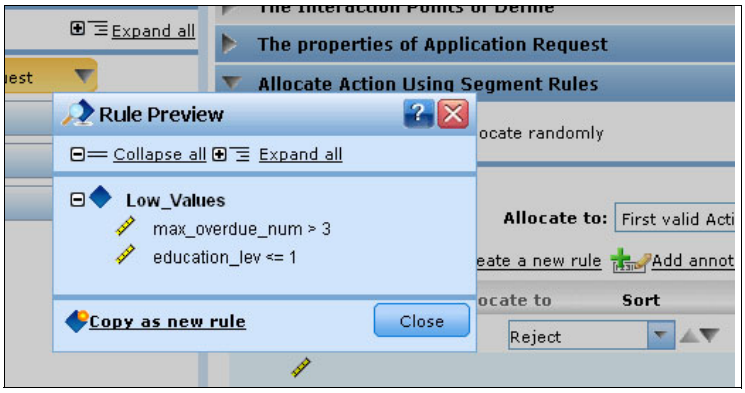


Figure 6-17 Low_Values rule preview

__ 8. Click **Close**.

In steps 9 through 11, you define the Small_Loan rule that automatically approves the loan application due to the small amount of the loan.

__ 9. Similarly, select **Small_Loan.rul** from the repository by using the previous steps.

The Small_Loan.rul means that if the applicant has a loan request of less than 500, automatically approve the loan.

__ 10. This time, set the **Allocate to** drop-down list of Small_Loan to **Approve**.

You should now have two allocation rules, as shown in Figure 6-18. If you are running ahead, spend some time trying to define your own allocation rules.

Figure 6-18 Allocation rules

__ 11. Close the **Allocate Action Using Segment Rules** section.

In steps 12 through 22 on page 66, you allocate actions by aggregating scores. A score is assigned to each business rule. The sum of these scores will be compared against the pre-determined thresholds for action allocation.

__ 12. Expand **Use Rules to Decide Which Action is Triggered** to open the details section, as shown in Figure 6-19 on page 64**.**

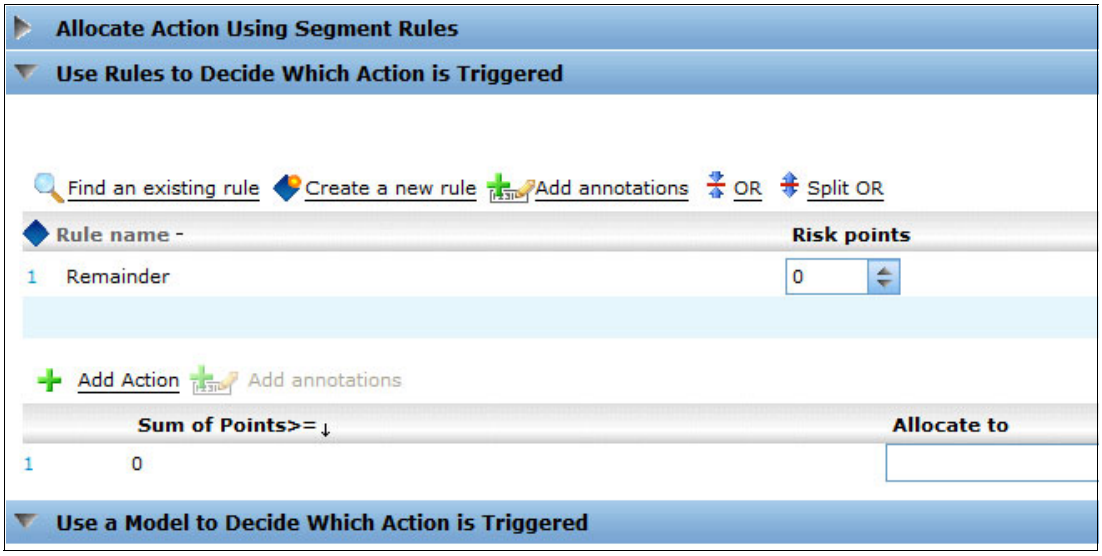


Figure 6-19 Use Rules to Decide Which Action is Triggered section

Again, instead of defining rules from scratch, we use the predefined rules for expediency:

__ 13. Click **Find an existing rule**.

__ 14. Select to retrieve the **Scoring_Rules.rul** from the repository.

__ 15. Click **Open**.

__ 16. When the dialog asks if the **Scoring_Rules.rul** is to be a **Convert** or **Replace,** select **Convert**.

The section now displays the imported rule set that will be used to add additional points to the application's risk score. You now configure the risk points for each rule.

Figure 6-20 on page 65 is a suggestion of how the risk points could be associated with the rules. In this step, you can experiment with assigning the risk points with your own scores for the rules. The allocation rules are based on common risk metrics such as credit history, debt-income ratio, and credit utilization. Thresholds have been established for these common metrics at the corporate level and for each key threshold, the risk analyst assigns points. In this example, the higher the point value, the better the risk.

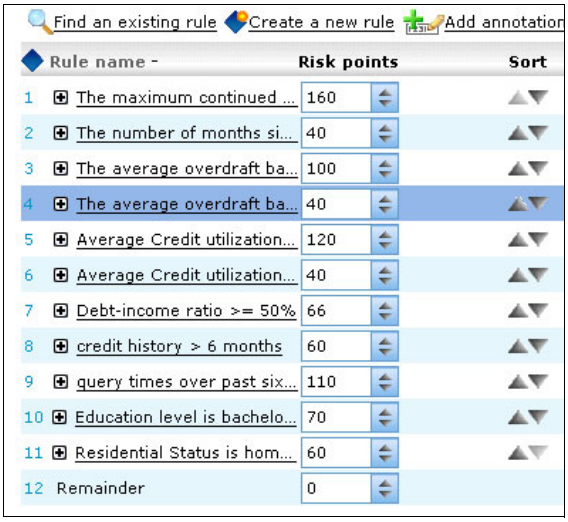


Figure 6-20 Example of Risk point allocation

You now have a score that is generated from the rules. Next, you need to associate an action with that score.

Steps 17 through 21 are to compare the sum of scores against the pre-determined thresholds. The higher that the sum of scores is, the higher credit-worthy the applicant is and hence, the more likely the application is to be approved.

__ 17. Under the rules, click the **Add Action** button twice.

__ 18. Set line 1 so that the *sum of the points* >= **400**; the action is to **Approve.**

__ 19. Set line 2 so that *the sum of points* >= **200**; the action is **Investigate.**

__ 20. For the remaining (line 3), the action is **Reject**.

__ 21. Click the **Save** icon (*the floppy disk)* in the top right to save the project.

The section should now look like Figure 6-21 on page 66.

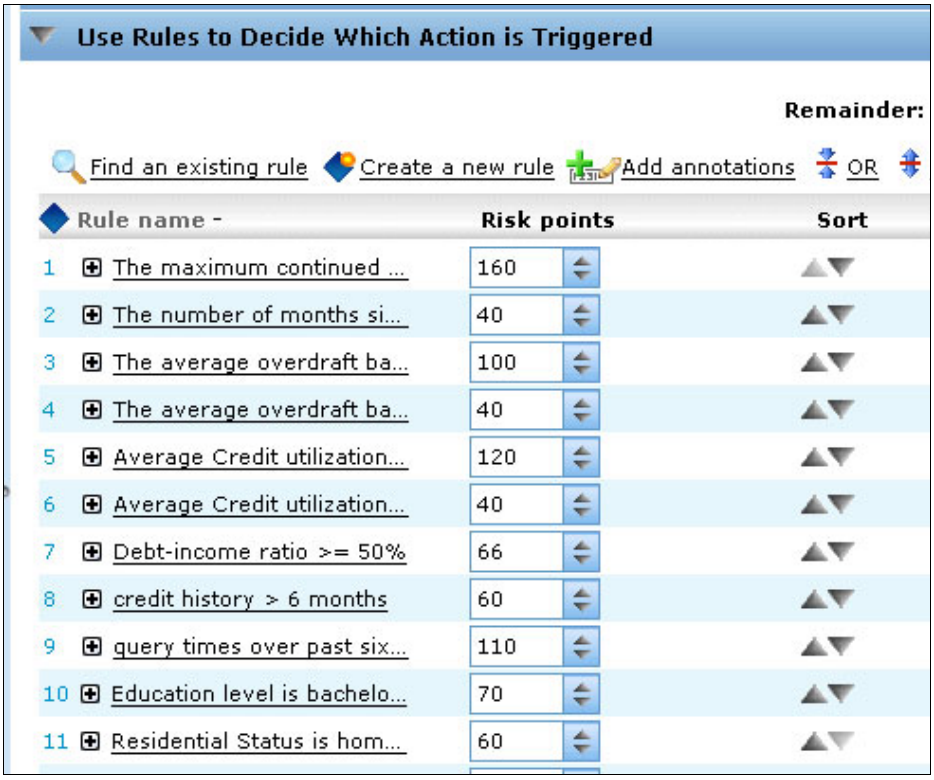


Figure 6-21 Use rules to decide which action is triggered

__ 22. Close the **Use Rules to Decide Which Action is Triggered** section.

You use the model that you created in the modeler module.

In steps 1 through 11, you allocate actions by comparing the model propensity with the pre-determined threshold. The higher the model propensity that is close to one is, the more likely the transaction is to be approved:

__ 1. Expand the **Use a Model to Decide Which Action is Triggered** section.

__ 2. Click **Find a model**.

__ 3. Browse into your group folder (specified in the worksheet) and select the model stream that you created earlier, called **Individual_Risk_Assessment.str**.

__ 4. Click **Open**.

__ 5. Click the drop-down menu under *Measure*.

__ 6. Select **Propensity**.

__ 7. Underneath the model, click **Add Action** two times.

__ 8. Set line 3 (the remainder) to **0** and the action to **Reject**.

__ 9. Set line 2 to **0.2** and the action to **Investigate**.

__ 10. Set line 1 to **0.8** and the action to **Approve**.

__ 11. *Save* the project.

The section should now look like Figure 6-22 on page 67. While rules capture organizational history and are based on things that the risk analyst has measured and observed, a model has been configured that predicts, in real time, whether a given applicant is approvable and calculates their approvable *propensity*. The *propensity* is used to allocate an action. Of those

applicants, the model indicates a high probability to be approvable and receive a more favorable outcome than those that have a lower probability to be approved.

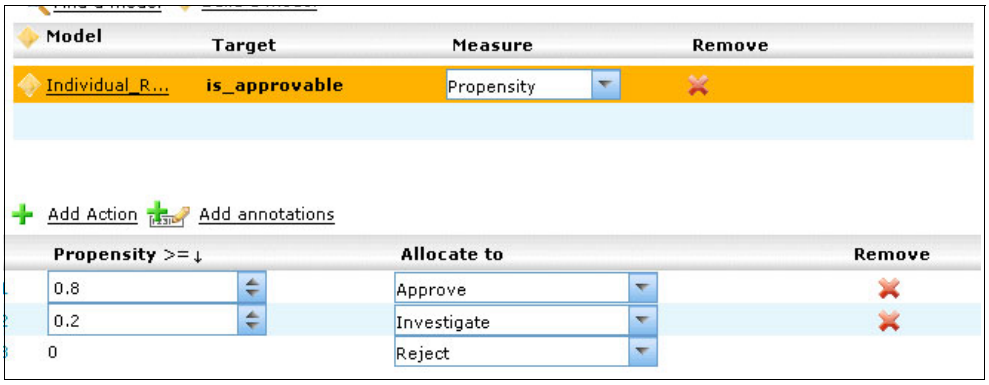


Figure 6-22 Use a model to decide which action is triggered

After you have defined the operational decisions with rules and models, you can simulate on your new configuration of Allocation, Model, and Rule for impact analysis. Business users effectively use this facility to conduct “What-If” simulation until the rule and model configuration is finalized. Follow these steps iteratively until the user is satisfied with the results:

__ 12. Click **Simulate** and then click **Run**. See the Simulation dialog, as shown in Figure 6-23.

Simulation within the define step shows how the allocations, rules, and models trigger across the population.

At the Application Request tab, you see the total number of cases used in this simulation.

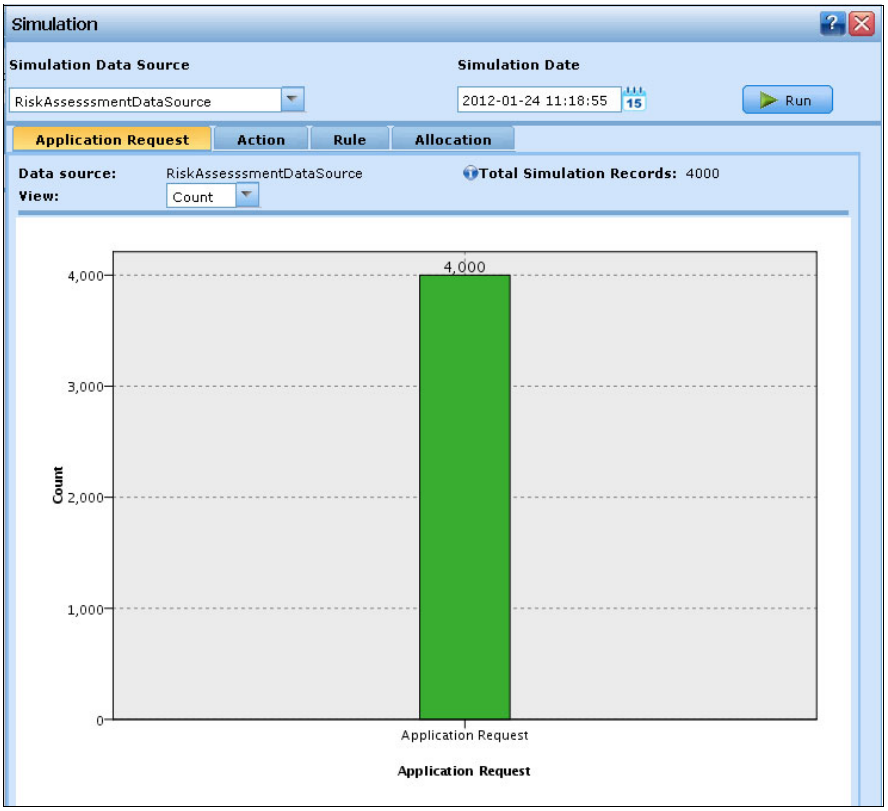


Figure 6-23 Total cases for simulation

Step 13 allows the risk analyst to compare the allocation results side by side.

__ 13. Select the **Action** tab to see the number of loan applications that fall into each of the outcomes that are based on **Allocation**, **Model**, and **Rule**, as shown in Figure 6-24.

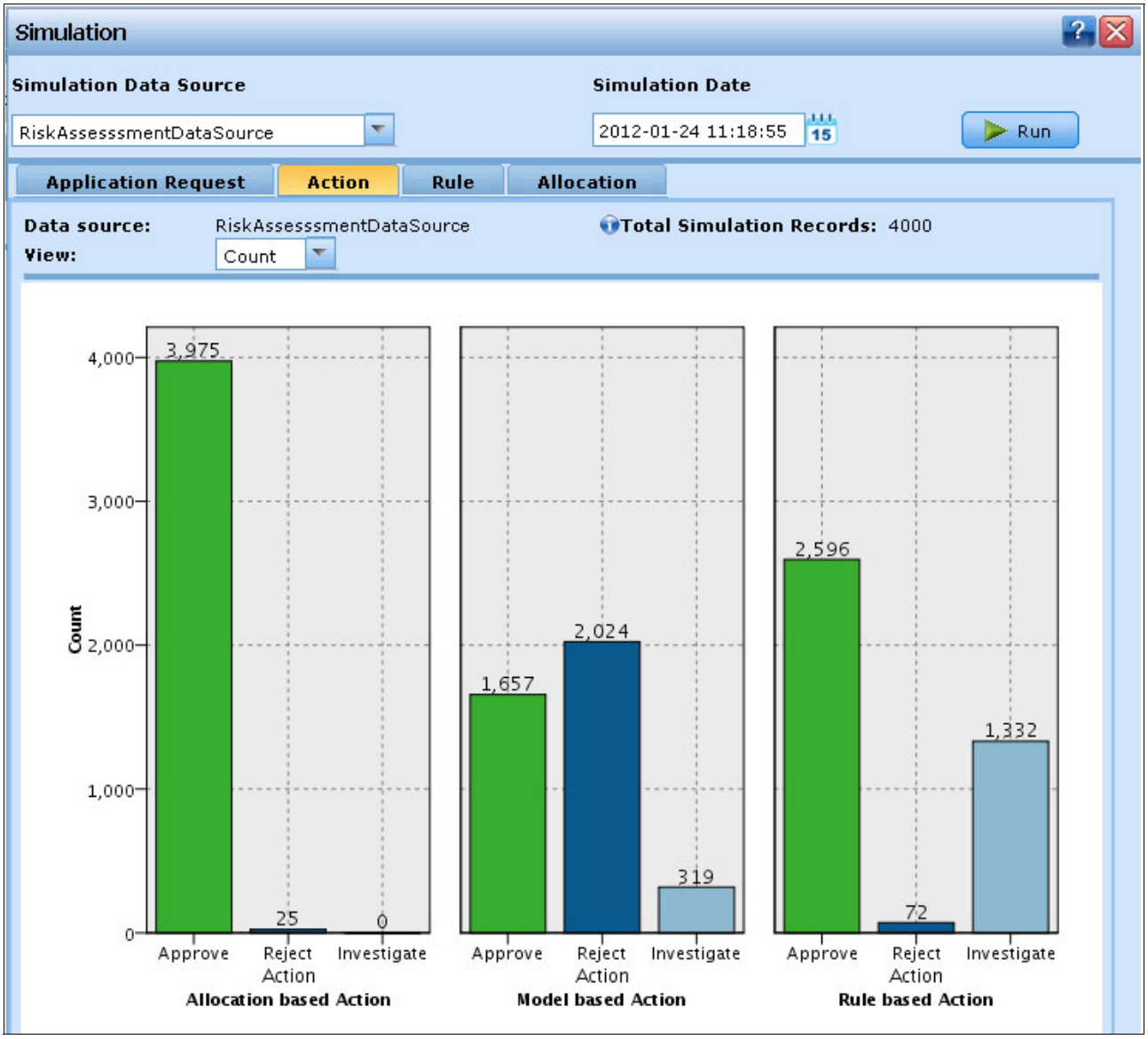


Figure 6-24 Action results

Steps 14 and 15 on page 70, allow the analyst to drill into the rules and the allocations by segment rules:

__ 14. Select the **Rule** tab to see the number of loan applications that fall into each Rule, as shown in Figure 6-25.

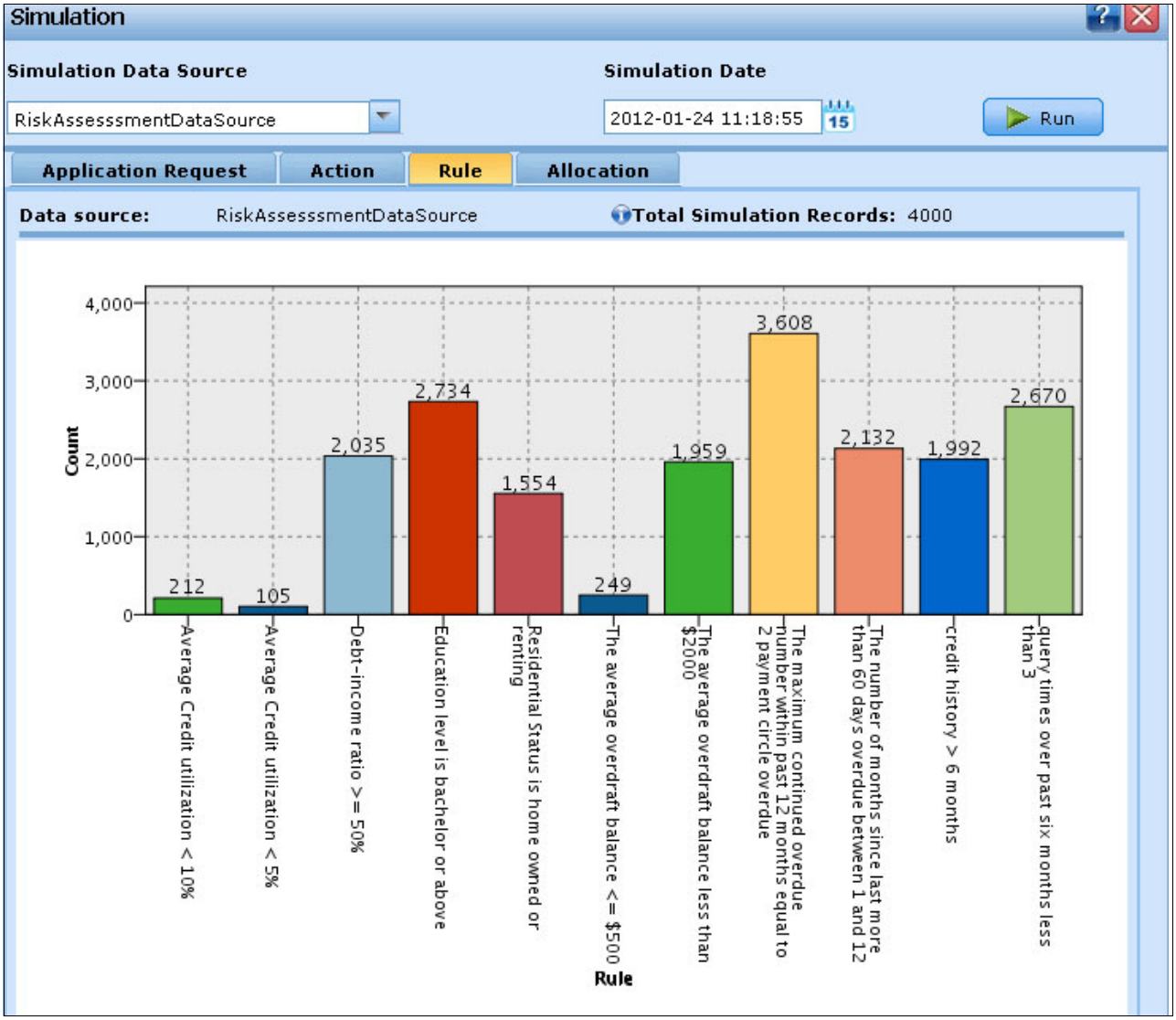


Figure 6-25 Rule tab

__ 15. Select the **Allocation** tab to see the number of loan applications that fall into each Allocation Rule, as shown in Figure 6-26.

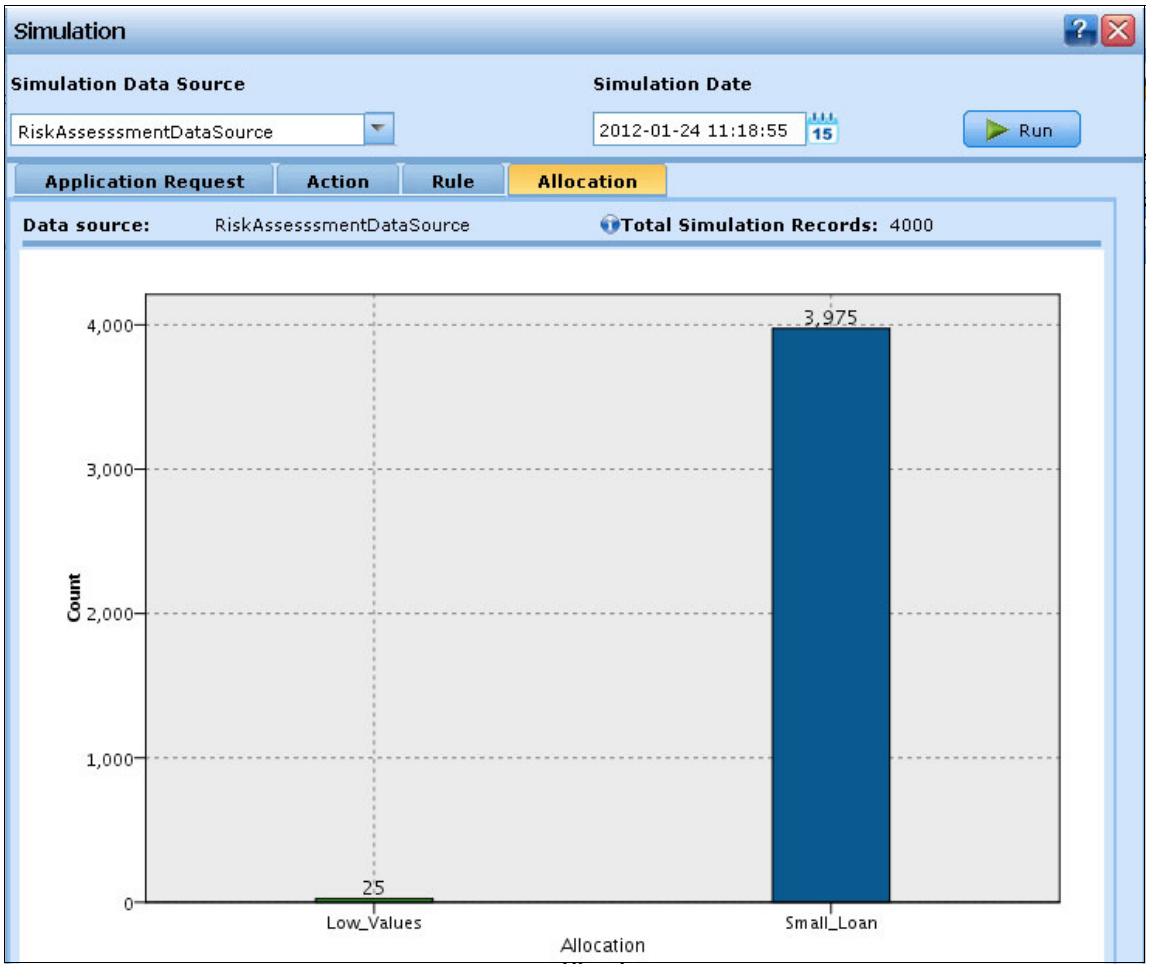


Figure 6-26 Allocation tab

When you have finished with simulation, close the simulation window and move on to the next section.

6.1.3 Combine

At the combine stage, the risk analyst indicates how the model outcome and the rule-based outcome should adjudicate. Simulation allows them to visualize the results and understand how many applicants, given their project data, would fit each of their configured actions.

The next steps are to specify how rules and models are combined together to determine the recommended action for each application. You can then use this to simulate the end result:

__ 1. Click the **Combine** tab.

__ 2. The first thing that we do is to assign colors to the values. Click the *white box* that is next to **Reject**.

__ 3. A color palette should appear, as in Figure 6-27 on page 71. Click *red*.

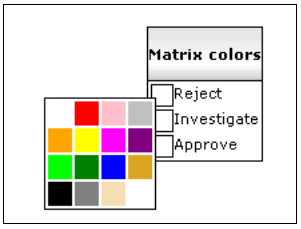


Figure 6-27 Action color palette

__ 4. Repeat for *yellow* for **Investigate**.

__ 5. Repeat for *green* for **Approve**.

The table should now look like Figure 6-28. The actions are color coded, which is a visual display of how your rules and models are combined together.

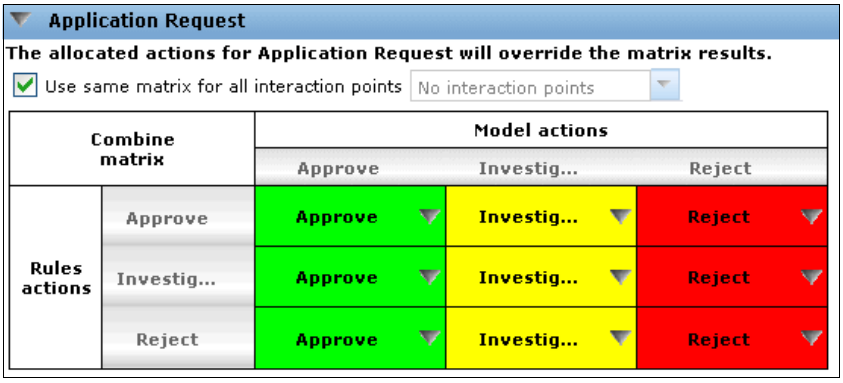


Figure 6-28 Application Request matrix

Step 6 allows the risk analyst to adjust the combination, or go back and adjust the rules and models, until the results are consistent with the organization's risk-tolerance and resource realities.

__ 6. At this step, click "**What-if**", then click **Run**. This runs a simulation on the combinations of rules and models. Changing the decision allocations in this table and using the simulation will help you arrive at the optimal configuration. We suggest the settings that are shown in Figure 6-29.

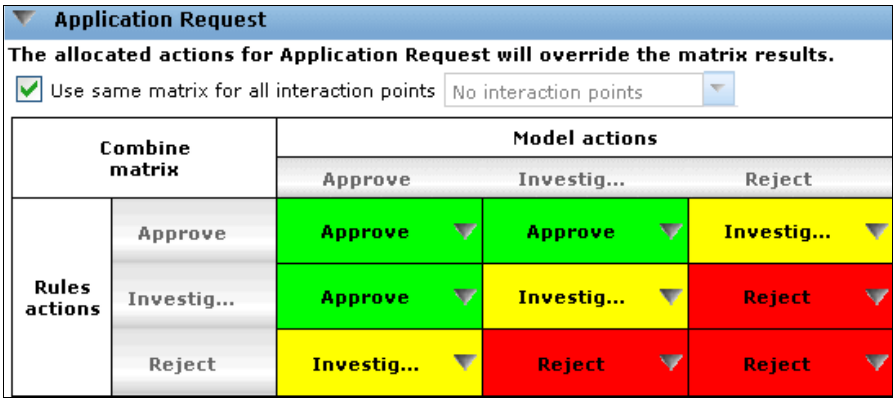


Figure 6-29 Customized Application Request matrix

__ 7. To change the decision allocations, click the triangle on the decision and select which action that you want to be performed. See Figure 6-30.

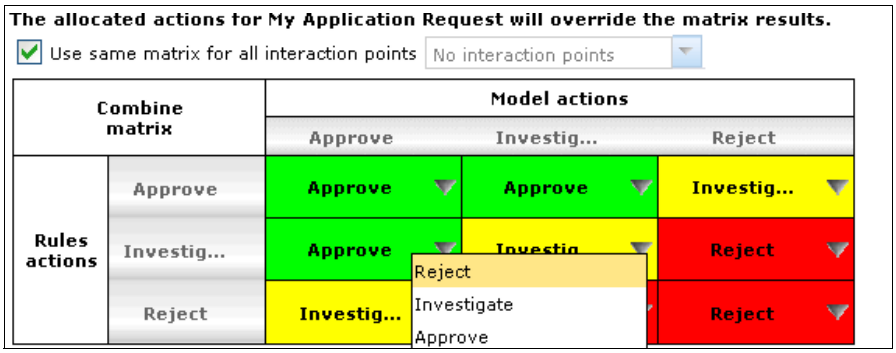


Figure 6-30 Selecting the decision combination

When you have finished experimenting, close the simulation box.

6.1.4 Deploy

After actions, rules and models are defined and the analyst has determined how rules and models interact. The project can be deployed. Steps 1 through 10 are to deploy the configuration to the repository.

Finally, you need to validate your current application processing project configuration and mark it as *ready to be deployed*:

__ 1. Click **Save**.

__ 2. Click the **Deploy** tab.

__ 3. Set **Deploy as** to **Deploy**.

__ 4. Click **Validate**.

__ 5. Check that the validation message reads: *The application workspace is valid for deployment*.

__ 6. Click **OK**.

__ 7. Click **Deploy**.

__ 8. If asked, click **Move**.

__ 9. Check that the deployment message reads: *The application workspace has been successfully deployed*.

__ 10. Click **OK**.

Congratulations. You are now finished with SPSS Decision Management; you can close that window.

6.1.5 Summary

In this module, you took the analytics model that you created in the first lab and enriched the outcome with business rules to arrive at an optimized configuration for the risk assessment through simulation.

After you configured the data, rules, and combination, you then deployed the SPSS DM stream to the repository to be used by the operational processes. In the next lab, you use this DM stream to create a scoring service, which can be called in real time.

7

Configuration of the risk assessment application for real-time scoring

In this part of the lab, you create a real-time scoring engine that is based on the configured SPSS Decision Management. Then, a CICS application can call a web service to request scoring in operational systems, informing decisions in either real time or batch. A risk analyst can change the parameters or configuration at any time and see results immediately. This is particularly useful if a new risk pattern has emerged and the risk analyst needs to quickly adjust to the new risk reality.

7.1 Introduction

The lab guides you through setting up, testing, and invoking the scoring service using an SPSS Decision Management stream and CICS Transaction Server.

At the end of this lab, you will know how to perform the following actions:

1. Configure SPSS Collaboration and Deployment Services for scoring
2. Use a CICS transaction to call the scoring service

7.1.1 Prerequisites

- You should have access to the client where you will be performing the lab exercise. Check with your lab facilitator for access.
- You should receive your team number from the lab facilitator. You will be using this number during lab to determine some of the worksheet parameters.
- You should have already completed the analytics modeling with IBM SPSS Modeler and configured the risk assessment in SPSS Decision Management modules.

7.1.2 Create a scoring configuration for the Decision Manager stream

Before a stream can be used for real-time scoring, you must define some supplemental information. The scoring configuration allows you to define which parameters, outputs, identification, real-time data provider, logging, and cache that you want the scoring to use. This allows for a single model to be used in a variety of scoring situations.

In the next steps, you create a scoring configuration for the Decision Manager stream that you created in the last module. The scoring configuration allows you to call the SPSS Collaboration and Deployment Services scoring service with the **LOANID** and **AMOUNT** parameters. The scoring configuration then uses the provided Real-Time Data Provider Definition to populate the rest of the scoring parameters with data from the database. We then specify the scoring to return the advised action and the risk score:

__ 1. Open **IBM SPSS Collaboration and Deployment Services** → **Deployment Manager 4.2.1** from the Windows Start menu.

__ 2. Right-click **CnDS**.

__ 3. Click **Log on as...**.

__ 4. Use the *CnDS User* and *CnDS Password* in your worksheet to authenticate.

__ 5. Expand **CnDS**.

__ 6. Expand **Content Repository**.

__ 7. Expand your *Group Folder* specified to you in your worksheet. You should be able to see the *SPSS Decision Management stream* that you just deployed, with the name *Individual_Risk_Assessment_DM.str*.

__ 8. Right-click the name and select **Configure Scoring**, as shown in Figure 7-1 on page 75.

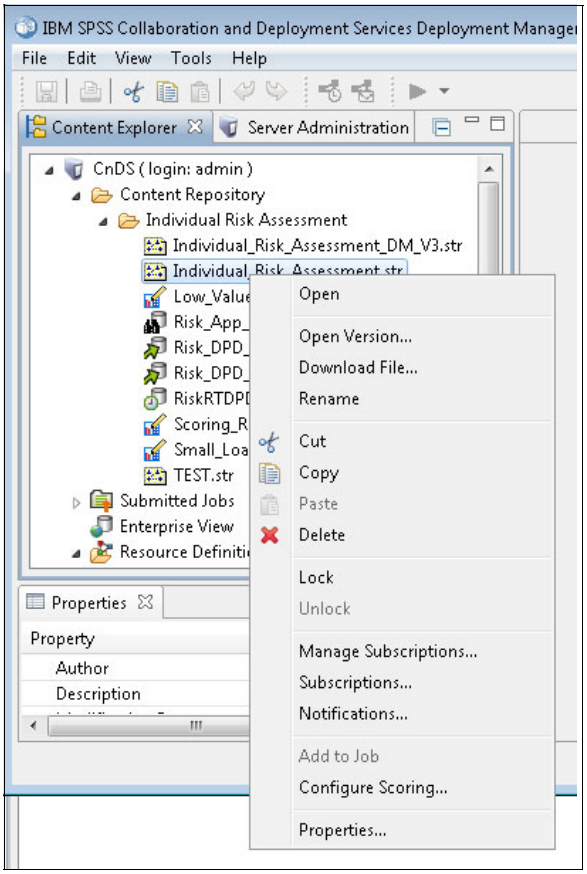


Figure 7-1 Configure Scoring option

__ 9. Enter the **Name** that is specified as the *Scoring ID* in the worksheet.

__ 10. Leave the **Label** as **LATEST**.

__ 11. Click **Next.**

__ 12. Leave the **Enable Interactive Score** check box *unchecked.*

__ 13. Click **Next**.

__ 14. In the Data Provider Settings dialog:

- Check **Use Data Provider**
- Select the **Data Provider**: **RiskRTDPD**
- Set the **Label** to **LATEST**
- Select the **Table**: **Risk_Scoring_Data**
- Select the **Key**: **Scoring_Key**

__ 15. The configuration should look similar to Figure 7-2 on page 76.

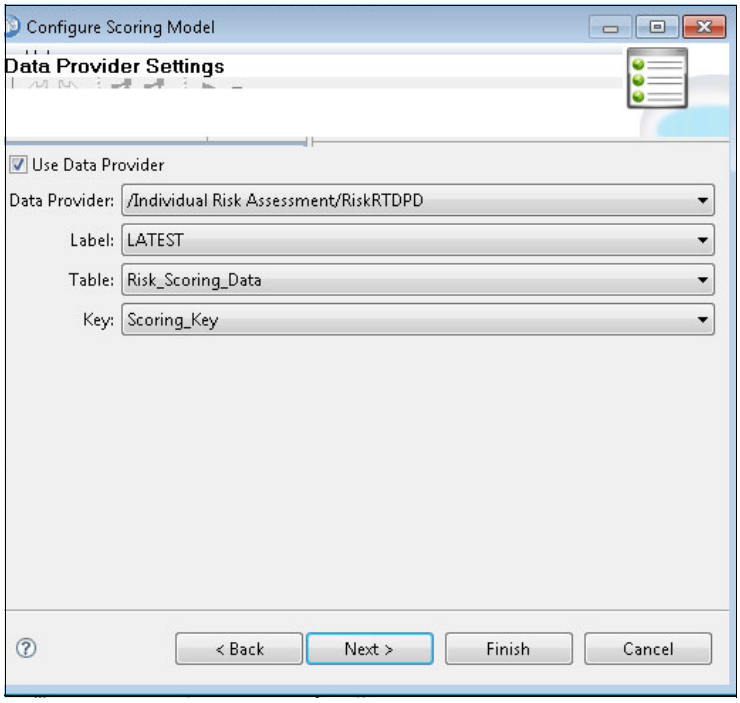


Figure 7-2 Data Provider settings

__ 16. Click **Next**.

__ 17. In the Input Data Order dialog:

- Move **LOANID** to the top of the list of inputs by selecting it and clicking **Up**.
- Move **AMOUNT** under **LOANID**, if it is not already there.

These are the most important inputs, as well as the only inputs that we include as real-time data. The data provider uses the **LOANID** to get the data to fill the rest of the columns.

See Figure 7-3 on page 77 for the configure scoring model input data order.

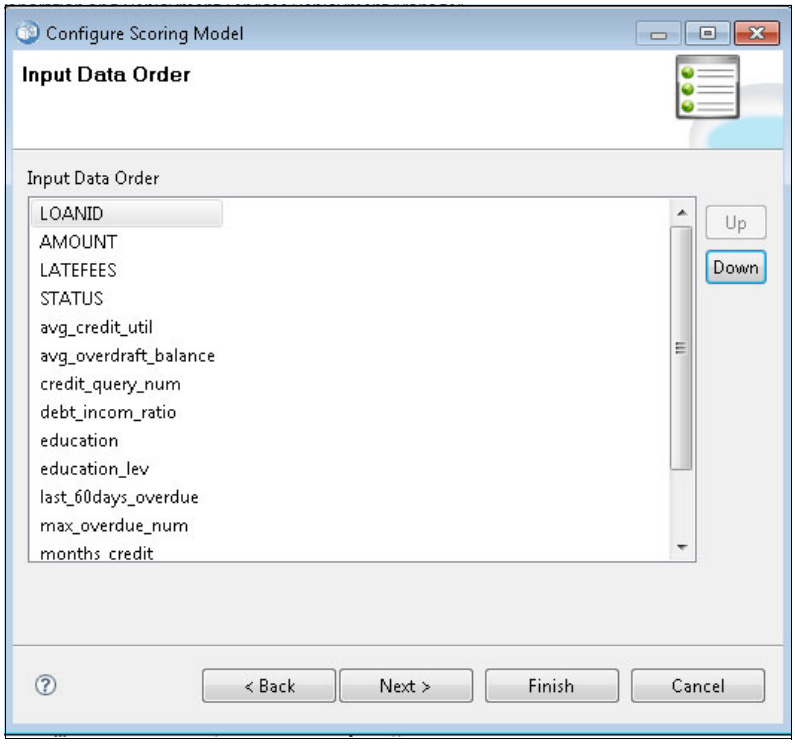


Figure 7-3 Configure scoring model input data order

__ 18. Click **Next**.

__ 19. In the Input Data Returned Settings dialog, leave the **Return model inputs in response** check box *unchecked.*

__ 20. Click **Next**.

__ 21. The only two fields that we want to return in this Proof of Technology (PoT) are **Aggregate-Value** and **Action**. Therefore, in the Output Data Returned Settings dialog, check these two fields, as shown in Figure 7-4 on page 78.

> **Shortcut:** Unchecking **Model Outputs** automatically unchecks all of the fields.

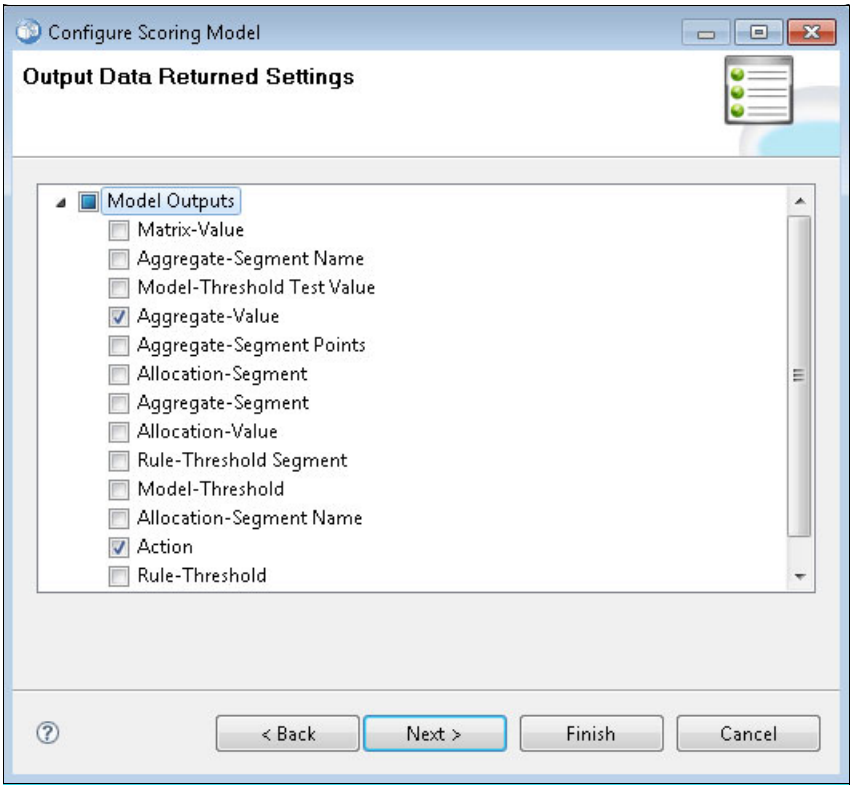


Figure 7-4 Configure scoring model output data

__ 22. Click **Finish**.

7.1.3 Use SPSS Collaboration and Deployment Services portal to test the scoring configuration

Now that you have defined the scoring configuration, you can test that you can use the model for real-time scoring using the deployment portal. The SPSS Collaboration and Deployment Services portal interface allows for all users with sufficient permissions to access the artifacts, logs, and results stored in the repository, as well as run jobs to generate scores, all through a thin client interface. In the next steps, we use the thin client to verify that we can use the scoring configuration to get scores from the Decision Manager stream.

__ 1. In your VM, open the deployment portal at http://<*Scoring Address*>:<*Scoring Port*>/peb. Where *Scoring Address* and *Scoring Port* are specified in your worksheet.

__ 2. Log in to the Collaboration and Deployment Services Deployment Portal with the *CnDS User* and *CnDS Password* in your worksheet.

__ 3. Click the **Content Repository** tab.

__ 4. Click your *Group Folder*, which is specified in your worksheet.

__ 5. Click your *Decision Management stream*, which is specified as the *Decision Management Project Name* in your worksheet.

__ 6. If you are prompted for your scoring configuration, select the one that you just created. You used the name specified by the *Scoring ID* variable in your worksheet.

__ 7. You are now able to see the inputs that are required for the scoring.

__ 8. Enter in a **LOANID** 1 - 9.

__ 9. Check the box in front of **AMOUNT**. By doing this, you are informing the scoring service that you want to use this value over the value in the database.

__ 10. Enter an amount.

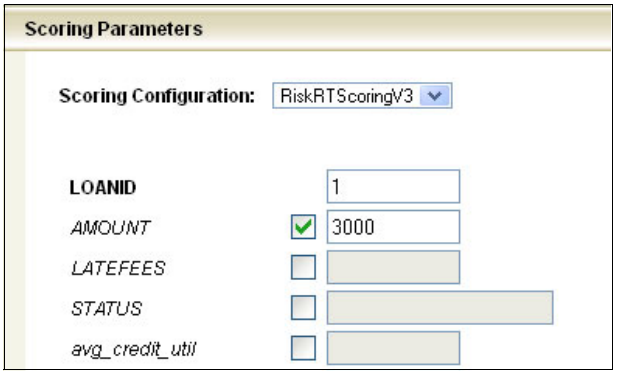


Figure 7-5 Entering scoring parameters

__ 11. Click **Score**. This button can be found at the bottom of all the inputs.

__ 12. You should see a result at the very bottom of the page, as shown in Figure 7-6.

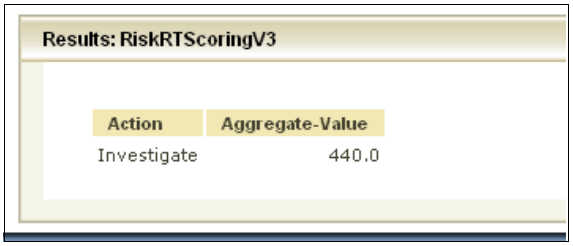


Figure 7-6 Successful scoring result

7.1.4 Call the scoring service from a CICS transaction

Introduction

Now that you have a scoring service that is configured and tested, the next step is to call the service in real time from a CICS transaction. There are many ways to create a CICS Web service requestor program. For this lab, the requestor application has already been written using the CICS Web services assistant tool.

The creation of the CICS Web services requestor involved the following steps (illustrated in Figure 7-7 on page 80):

1. Used SPSS Modeler to create a modeler stream and generate a predictive model from some business data that is stored in a DB2 database. Defined a scoring branch on the modeler stream.
2. Deployed the modeler stream to SPSS Collaboration and Deployment Services specifying the scoring node and parameters. We then enhanced the stream by using SPSS Decision Management.
3. Defined a scoring configuration in SPSS Collaboration and Deployment Services for the deployed stream. This allows the predictive model to be used for scoring via the SPSS Collaboration and Deployment Services scoring service.

4. Obtained the Web Services Description Language (WSDL) and schema documents, which define the generic SPSS Collaboration and Deployment Services scoring service. Used the CICS Web services assistant tool, DFHWS2LS, to generate a high-level language data structure and a web service binding file from a web service description.
5. Developed a CICS COBOL service requester application using the code that is generated by the DFHWS2LS tool to call the getScore operation on the scoring service.

Note: Only steps 4 and 5 would be required. Steps 1 - 3 have already been described in the previous models that were covered in the prerequisites section.

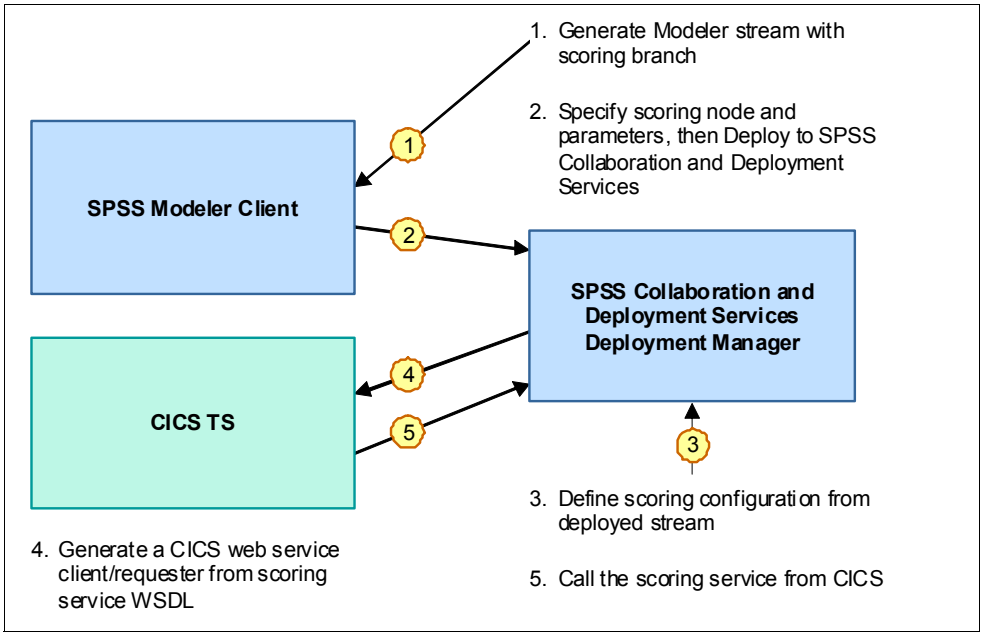


Figure 7-7 Creating the CICS scoring service requestor application

This application has already been deployed for you, so in the last part of this lab you will perform the following steps:

1. Log on to the CICS region on the mainframe using a 3270 terminal
2. Invoke the scoring application
3. Configure the application to point at your scoring service
4. Execute the scoring in real time from the CICS transaction

Log on to the CICS region on the mainframe using a 3270 terminal

The first thing that you need to do is open a 3270 terminal session to the mainframe. You will use this to log onto the CICS region so that you can invoke a transaction to call the scoring service. Your client workstation needs IBM Personal Communications for this:

__ 1. Open *IBM Personal Communications* and double-click **wtsc90**. See Figure 7-8 on page 81.

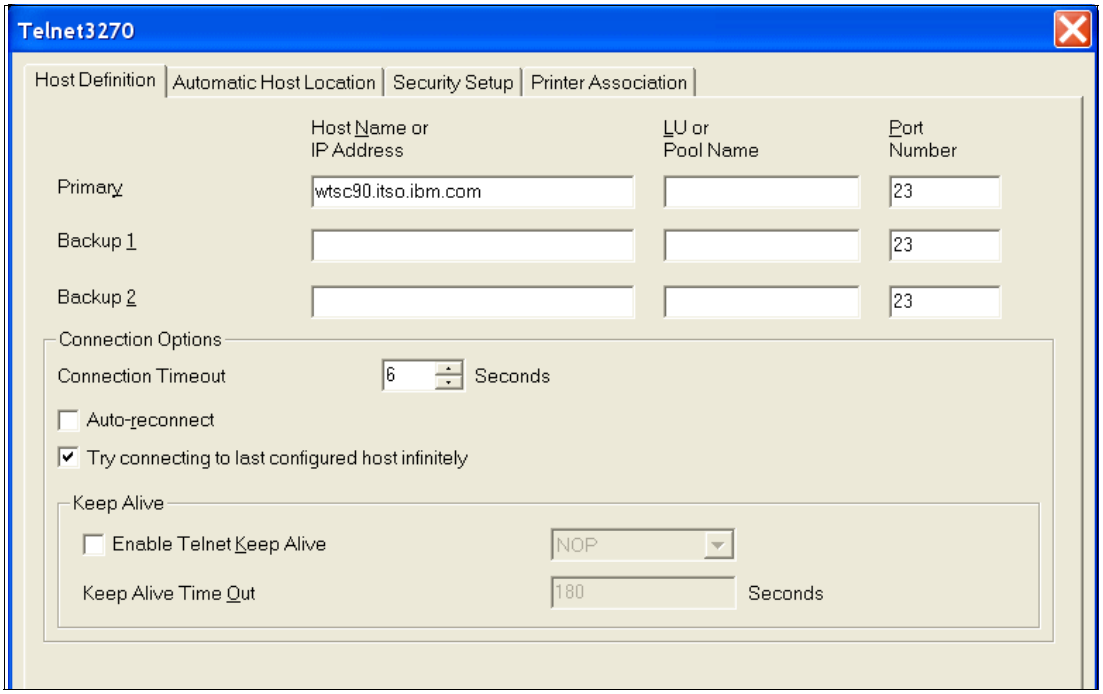


Figure 7-8 Link parameter setup panel

__ 2. Click **OK**.

You should see a logon window that is similar to Figure 7-9.

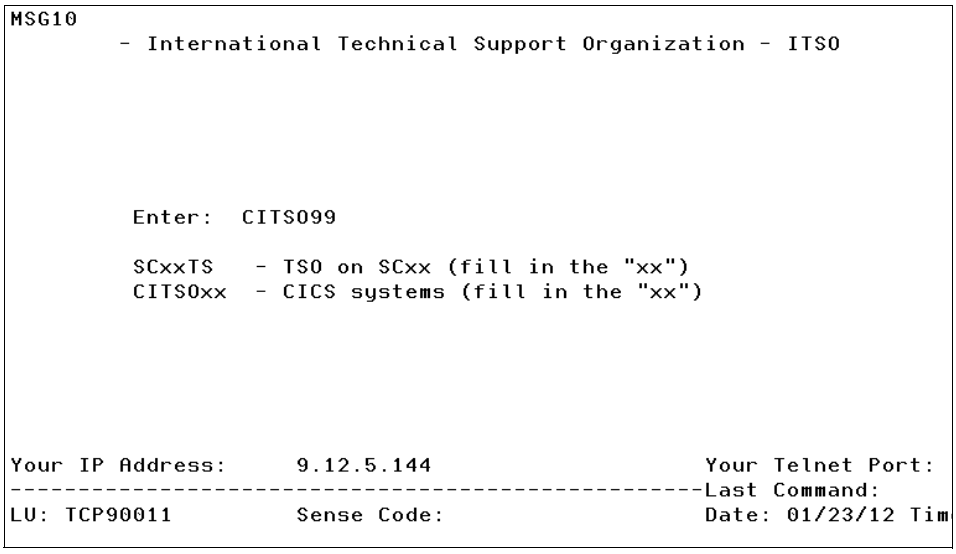


Figure 7-9 ITSO logon window

__ 3. To log in to the CICS region, type in the *CICS Region Name* from your worksheet.

__ 4. Press the *right control button (Ctrl)* to enter the command.

__ 5. Enter the *z/OS user ID* and *z/OS Password* from your worksheet, as shown in Figure 7-10 on page 82.

__ 6. Press the *right control button (Ctrl)* to enter the logon credentials.

__ 7. You should see a message such as *DFHCE3549 Sign-on is complete*, which informs you that your logon was successful.

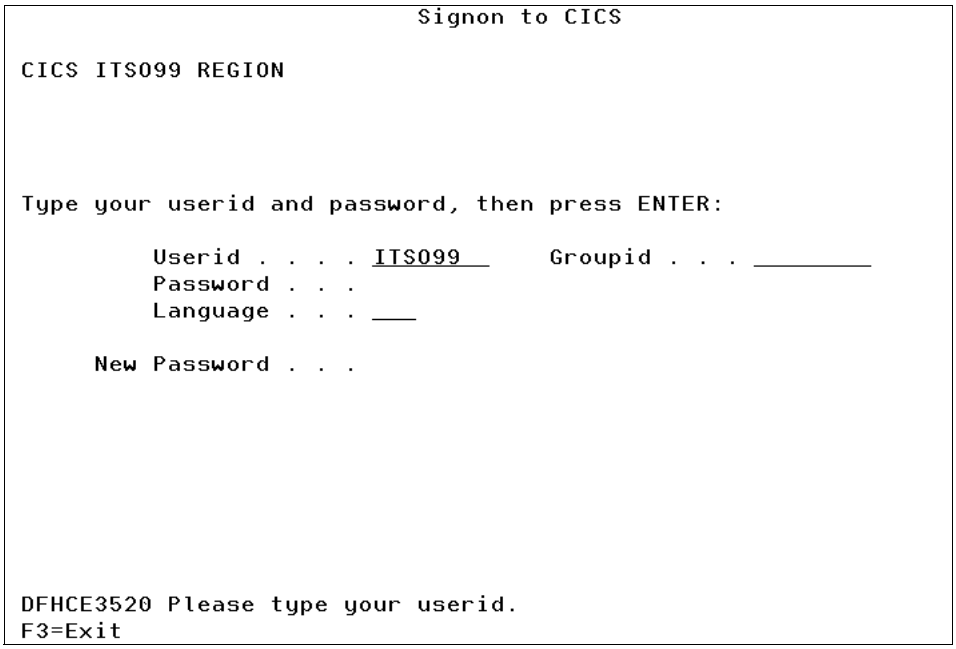


Figure 7-10 Signon to CICS window

Invoke the scoring transaction

The scoring transaction provides a CICS 3270 terminal user interface to collect real-time input for the scoring. This is one example of how to consummate a CICS transaction using real-time scoring.

To start the scoring application, take the following steps:

__ 1. Press the *pause button* to clear the display.

__ 2. In the top left corner, type the transaction name. This is the *Scoring Transaction* variable in your worksheet. See Figure 7-11 on page 83 for an example.

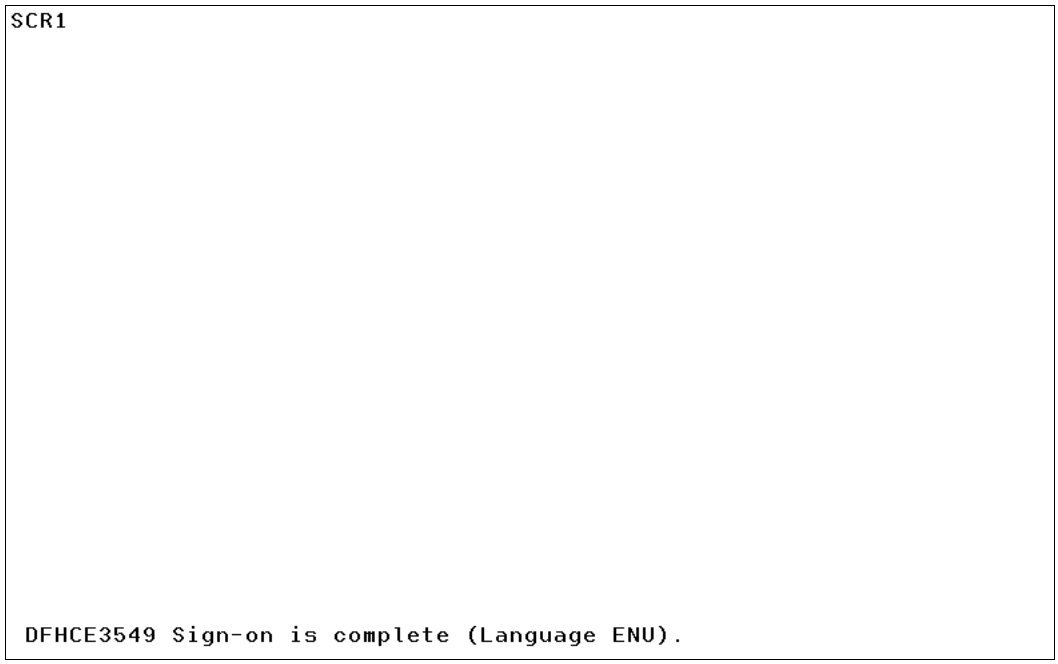


Figure 7-11 Entering the scoring transaction

__ 3. Press the *right control button (Ctrl)* to submit the transaction to CICS.

__ 4. You should now see the scoring application CICS user interface (UI). This will look similar to Figure 7-12.

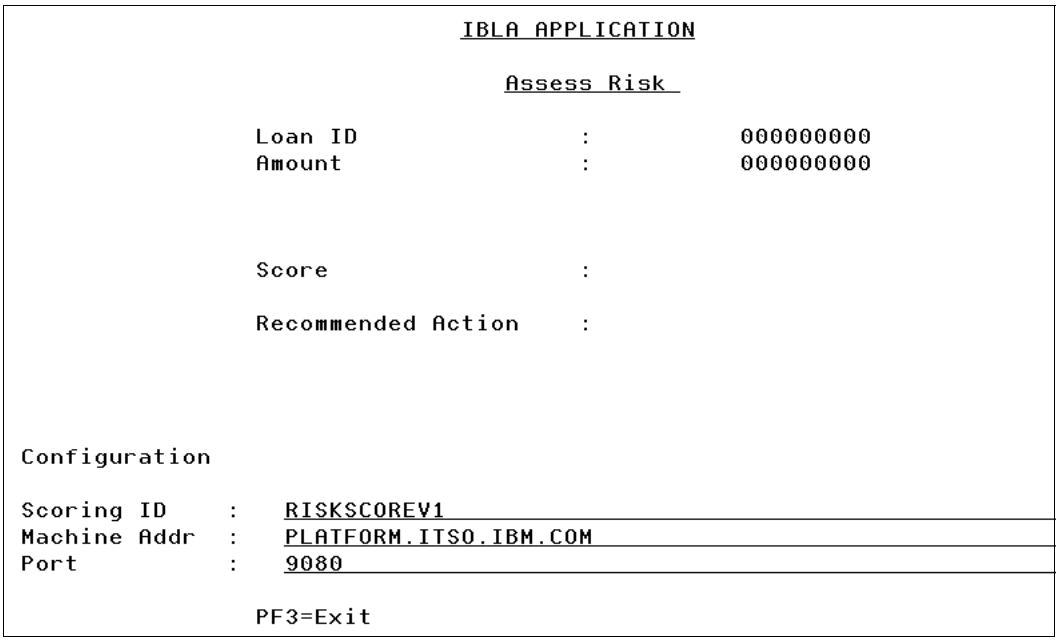


Figure 7-12 CICS scoring 3270 user interface

Configure the application to point at your scoring service

There is a configuration part at the bottom of the application window. This allows for you to point the CICS transaction at different SPSS Collaboration and Deployment Service servers and use the scoring configuration that you just made.

__ 1. Set the **Scoring ID** to the ID of the scoring configuration that you made. This is the same as the *Scoring ID* variable in your worksheet.

__ 2. Set the **Machine Addr** to the *Scoring Address* variable in your worksheet.

__ 3. Set the **Port** to the *Scoring Port* variable in your worksheet.

Execute the scoring in real time from the CICS transaction

Now that the scoring transaction has the correct scoring parameters, you can call your scoring configuration in real time from the CICS application. This is done by following these steps:

__ 1. Enter a **Loan ID**. This can be a number *between 1 and 9*.

__ 2. Enter an **Amount**.

__ 3. Press the *right control button (Ctrl)* to submit the inputs for scoring.

__ 4. You now can see that a scoring algorithm has run and returned a risk score and a recommend action.

Note: The input fields add leading zeros to the data values, such as 000000003.

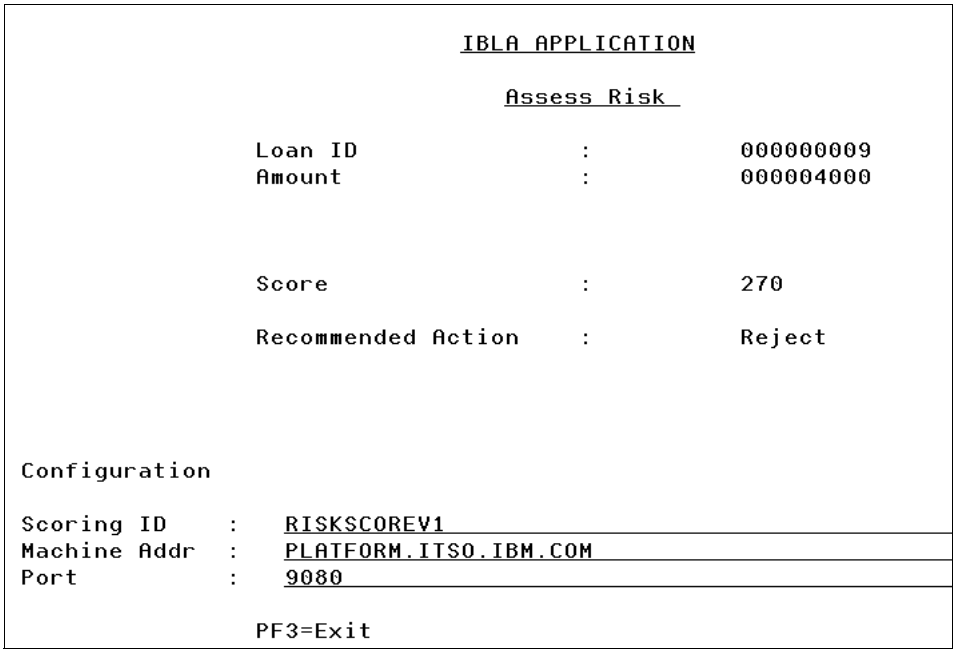


Figure 7-13 Successful scoring execution

Note: If you see an error message pop-up window in the application, check your configuration, your inputs, and that your scoring ID name is in capital letters. For example, **NON-ZERO RC FROM INVOKE SERVICE.**

7.1.5 Summary

In this module, you created a scoring configuration that allows you to use the Decision Management stream that you created to be called in real time. In this example, the call from real time is coming from a CICS transaction with real-time data provided on a CICS 3270 terminal user interface.

A

Additional material

This book refers to additional material that can be downloaded from the Internet as described in the following sections.

Locating the Web material

The Web material associated with this book is available in softcopy on the Internet from the IBM Redbooks Web server. Point your Web browser at:

ftp://www.redbooks.ibm.com/redbooks/SG248153

Alternatively, you can go to the IBM Redbooks website at:

ibm.com/redbooks

Select the **Additional materials** and open the directory that corresponds with the IBM Redbooks form number, SG248153.

Using the Web material

The additional Web material that accompanies this book includes the following files:

File name	*Description*
8153 POT.zip	Zipped Presentations and PDFs

System requirements for downloading the Web material

The Web material requires the following system configuration:

Hard disk space: 7 MB minimum
Operating System: Windows

Downloading and extracting the Web material

Create a subdirectory (folder) on your workstation, and extract the contents of the Web material.zip file into this folder.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topic in this document. Some publications referenced in this list might be available in softcopy only.

- *DB2 10 for z/OS Technical Overview*, SG24-7892
- *Using zEnterprise for Smart Analytics: Volume 2 Implementation*, SG24-8008

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft, and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- *DB2 10 for z/OS Installation and Migration Guide*, GC19-2974
- *DB2 for z/OS Application Programming Topics*, SG24-6300
- *IBM SPSS Modeler Server Scoring Adapter for DB2 on z/OS License Information*, GC19-3721
- *IBM SPSS Modeler Server Scoring Adapter for DB2 on z/OS Program Directory,* GI10-8919

Online resources

These websites are also relevant as further information sources:

- IBM Business Analytics on System z

  http://www.ibm.com/software/os/systemz/badw
- IBM DB2 Accessories Suite for z/OS - Software

  http://www.ibm.com/software/data/db2imstools/db2tools/accessories-suite/
- IBM SPSS software

  http://www.ibm.com/software/analytics/spss

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

(0.2”spine)
0.17”<->0.473”
90<->249 pages

Risk Scoring for a Loan Application on IBM System z

Running IBM SPSS Real-Time Analytics

Using IBM SPSS Modeler for Analytics modeling

Configuring risk assessment in SPSS Decision Management

Real-time scoring using a System z host

When ricocheting a solution that involves analytics, the mainframe might not be the first platform that comes to mind. However, the IBM System z group has developed some innovative solutions that include the well-respected mainframe benefits. This book describes a workshop that demonstrates the use of real-time advanced analytics for enhancing core banking decisions using a loan origination example. The workshop is a live hands-on experience of the entire process from analytics modeling to deployment of real-time scoring services for use on IBM z/OS.

In this IBM Redbooks publication, we include a facilitator guide chapter as well as a participant guide chapter. The facilitator guide includes information about the preparation, such as the needed material, resources, and steps to set up and run this workshop. The participant guide shows step-by-step the tasks for a successful learning experience. The goal of the first hands-on exercise is to learn how to use IBM SPSS Modeler for Analytics modeling. This provides the basis for the next exercise “Configuring risk assessment in SPSS Decision Management”. In the third exercise, the participant experiences how real-time scoring can be implemented on a System z.

This publication is written for consultants, IT architects, and IT administrators who want to become familiar with SPSS and analytics solutions on the System z.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8153-00 ISBN 0738438847