



Execution of R Built Predictive Solutions

Alex Guazzelli, PhD
VP, Analytics - Zementis, Inc.

useR! 2010

Exporting Models from R

Memory

Why?

Speed

Freedom

Transparency

Interoperability

Accessibility

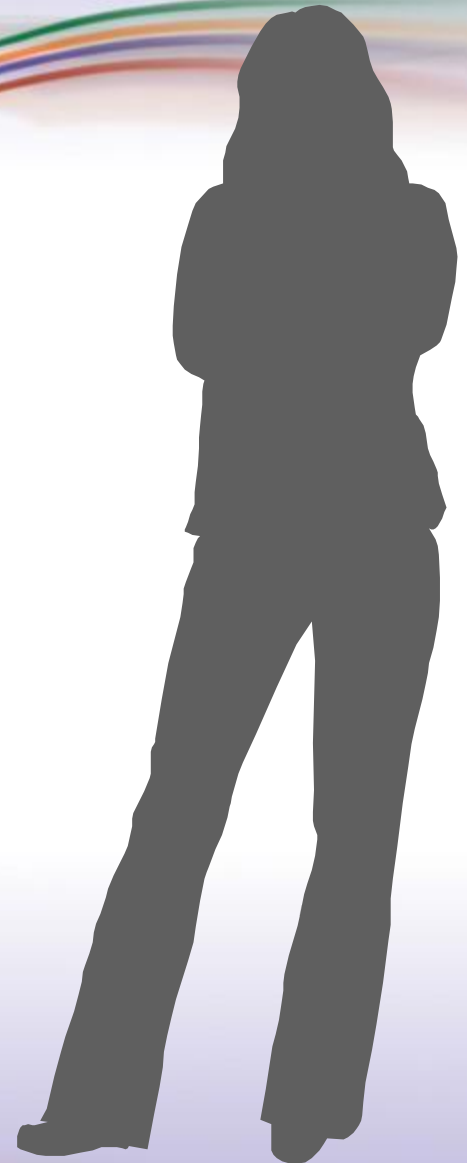
Because you can!



Exporting Models from R

How?

PMML

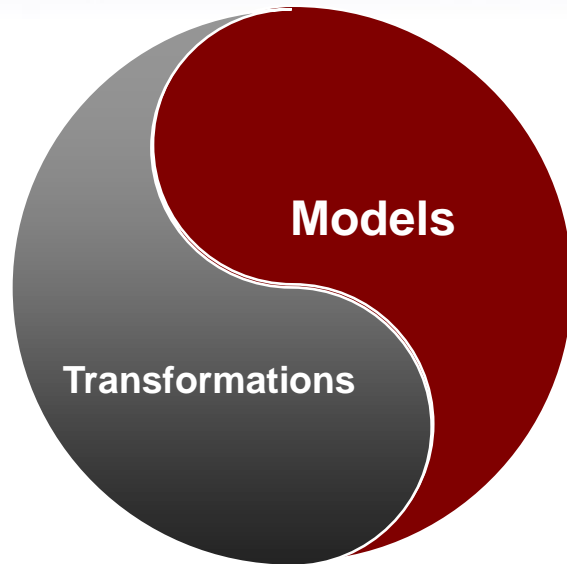




Predictive Model Markup Language (PMML)

- PMML is an XML-based language to
 - Define data mining models
 - Share models between compliant applications
- Standard for exchange of models to
 - Avoid proprietary issues and incompatibilities
 - Easily put models to work
- Clear separation of tasks
 - Model development vs. model execution
 - Scientists focus on building the best model
 - Eliminates need for custom model *deployment*

PMML Structure



PMML defines a standard not only to represent data-mining models, but also **data handling** and **data transformations** (pre- and post-processing)

- A **Data Dictionary** defines all the raw data fields (including missing value strategy and outlier treatment).
- Several **Data Transformations** strategies allow for intelligent extraction of feature detectors from raw data (“data massaging”).
- A comprehensive list of **Data-Mining Models** offers power and flexibility.
- Post-processing of results allow for tailored decisions.
- Model Explanation allows for performance evaluation.

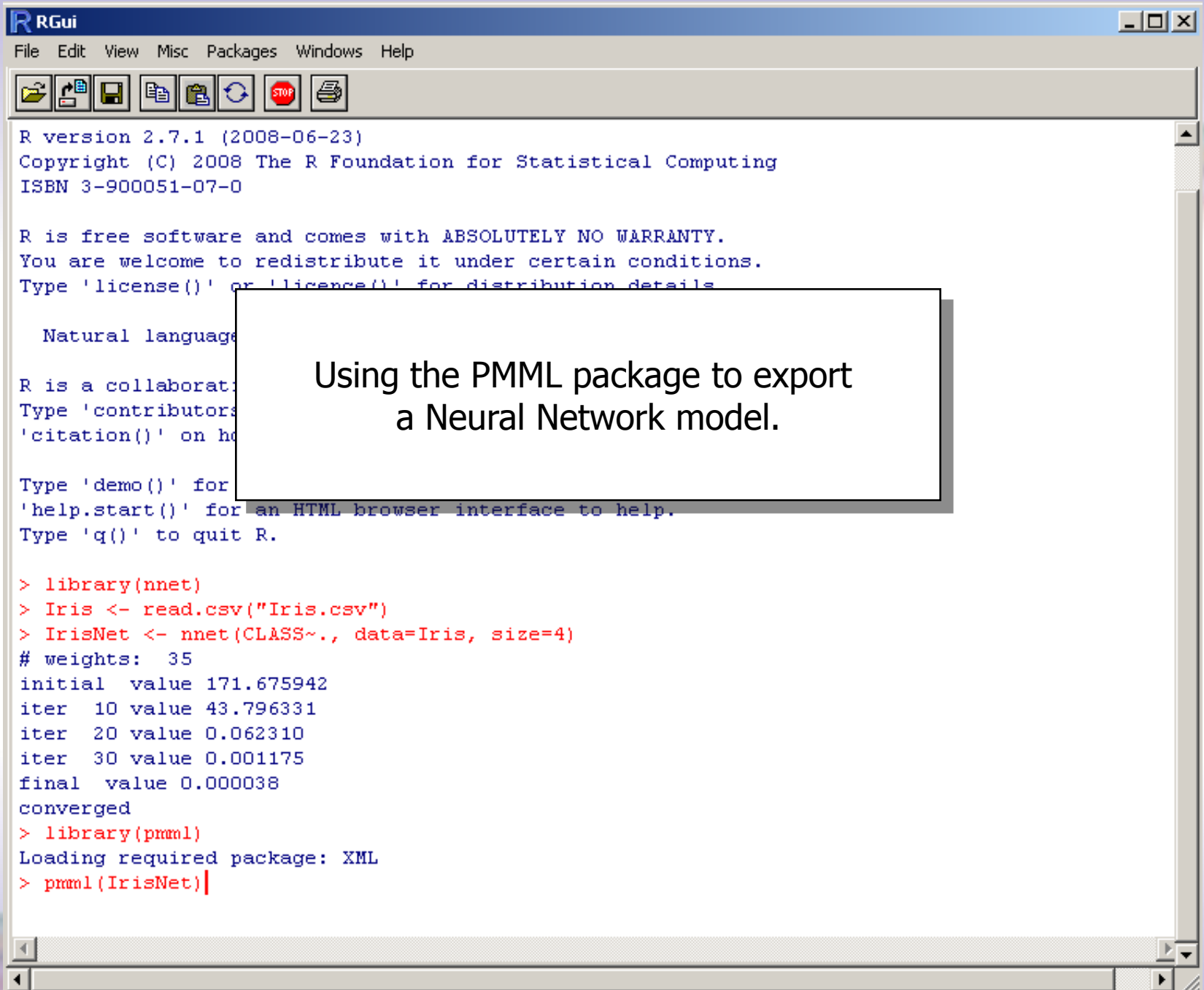
PMML

Industry Support



Matured and Supported by Industry

- Data Mining Group <http://www.dmg.org>
- Vendor independent consortium
- Mature standard
 - Current version 4.0
 - Active group and constant enhancements
- Industry supporters
 - Major Players: IBM/SPSS, Oracle, SAP, Microsoft
 - Analytics: KXEN, SAS, Salford, Togaware, Zementis
 - BI: Microstrategy, Teradata, Tibco, Pentaho
 - Open Source: R, KNIME, Rapid-I
 - Others: Equifax, FICO, Open Data Group, Visa, Pervasive, NASA



The image shows a screenshot of the RGui window. The window title is "RGui" and it has a menu bar with "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations. The main area is a text editor displaying the R startup message and a series of R commands and their output. A white text box with a black border is overlaid on the right side of the console, containing the text "Using the PMML package to export a Neural Network model." The R console output shows the version (2.7.1), copyright information, and a list of help topics. The user has entered several commands: loading the 'nnet' package, reading an 'Iris.csv' file, creating a neural network model 'IrisNet', and finally loading the 'pmml' package and calling 'pmml(IrisNet)'.

```
R version 2.7.1 (2008-06-23)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language processing help topics:
  help(package="nltools")

R is a collaborative project.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(nnet)
> Iris <- read.csv("Iris.csv")
> IrisNet <- nnet(CLASS~., data=Iris, size=4)
# weights: 35
initial value 171.675942
iter 10 value 43.796331
iter 20 value 0.062310
iter 30 value 0.001175
final value 0.000038
converged
> library(pmml)
Loading required package: XML
> pmml(IrisNet)|
```

Model is readily exported in PMML and ready to be used.

```
> pmml(IrisNet)
<PMML version="3.2" xmlns="http://www.dmg.org/PMML-3_2" xmlns:xsi="http://www.w3.org/2001/XMLSchema"
  <Header copyright="Copyright (c) 2008 Alex Guazzelli" description="Neural Network PMML Model"
    <Extension name="timestamp" value="2008-10-23 17:45:45" extender="Rattle"/>
    <Extension name="description" value="Alex Guazzelli" extender="Rattle"/>
    <Application name="Rattle/PMML" version="1.1.9"/>
  </Header>
  <DataDictionary numberOfFields="5">
    <DataField name="CLASS" optype="categorical" dataType="string">
      <Value value="Iris-setosa"/>
      <Value value="Iris-versic"/>
      <Value value="Iris-virgin"/>
    </DataField>
    <DataField name="SEPAL_LE" optype="continuous" dataType="double"/>
    <DataField name="SEPAL_WI" optype="continuous" dataType="double"/>
    <DataField name="PETAL_LE" optype="continuous" dataType="double"/>
    <DataField name="PETAL_WI" optype="continuous" dataType="double"/>
  </DataDictionary>
  <NeuralNetwork modelName="NeuralNet_model" functionName="classification" numberOfLayers="2">
    <MiningSchema>
      <MiningField name="CLASS" usageType="predicted"/>
      <MiningField name="SEPAL_LE" usageType="active"/>
      <MiningField name="SEPAL_WI" usageType="active"/>
      <MiningField name="PETAL_LE" usageType="active"/>
      <MiningField name="PETAL_WI" usageType="active"/>
    </MiningSchema>
    <NeuralInputs numberOfInputs="4">
      <NeuralInput id="1">
        <DerivedField name="derivedNI_SEPAL_LE" optype="continuous" dataType="double">
          <FieldRef field="SEPAL_LE"/>
        </DerivedField>
      </NeuralInput>
```


From R to PMML

Supported Packages/Objects



nnet: Neural Networks

hclust: Clustering

rpart: Decision Trees

arules: Association Rules

lm/glm: Regression

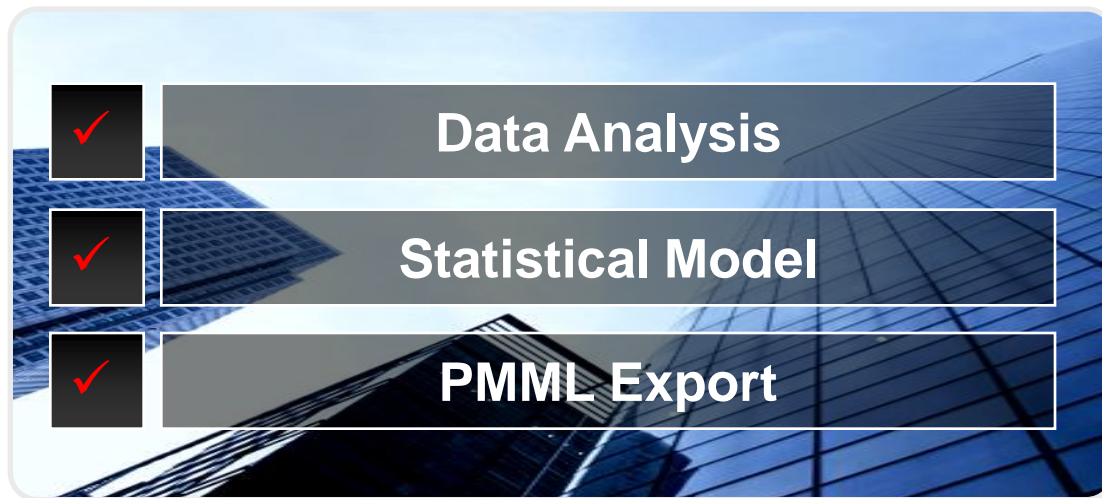
kmeans: Clustering

ksvm: SVMs

randomForest



Got Models...



What Now?



ADAPA

❖ ADAPA by Zementis

- ❖ Predictive Decisioning Platform
- ❖ PMML-based
- ❖ Drools to integrate business logic
- ❖ Scalable execution platform
- ❖ Real-time integration into business processes
- ❖ Accessible from anywhere
- ❖ Not a model development environment

❖ ADAPA on Amazon Elastic Compute Cloud

- ❖ Software as a service
- ❖ Up/Down scaling as needed
- ❖ Pay-as-you-go
 - ❖ Amazon Payments (\$.99 per hour)
- ❖ Amazon experience & reliability



From Model Building to Model Deployment

Model Building

PERVASIVE

IBM

KNIME

R

SALFORD SYSTEMS

sas

TIBCO
The Power of Now

RAPID MINER

SPSS

Microsoft
SQL Server

K
kxen
KNOWLEDGE
EXTRACTION
ENGINES

MicroStrategy

STATISTICA

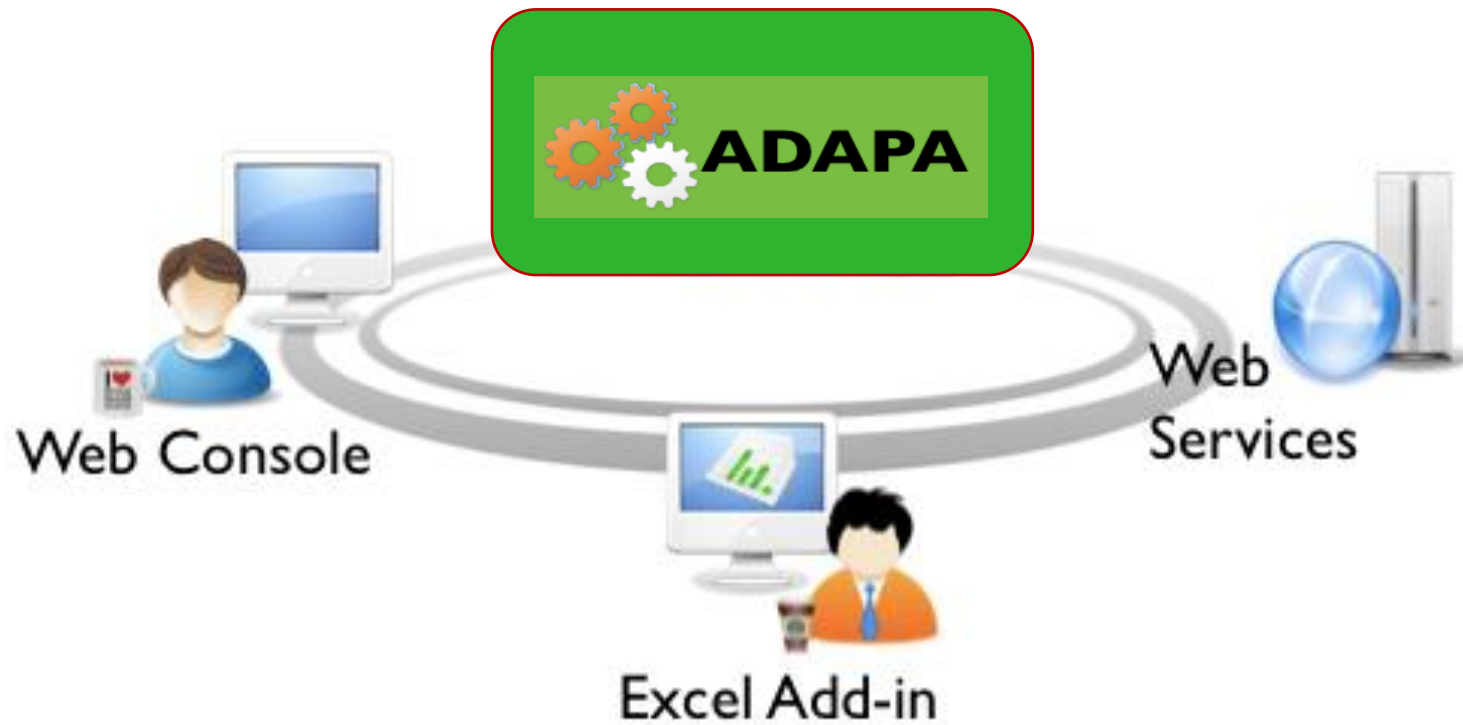
PMML
Predictive Model
Markup Language

Model Deployment

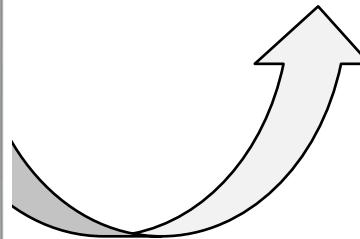
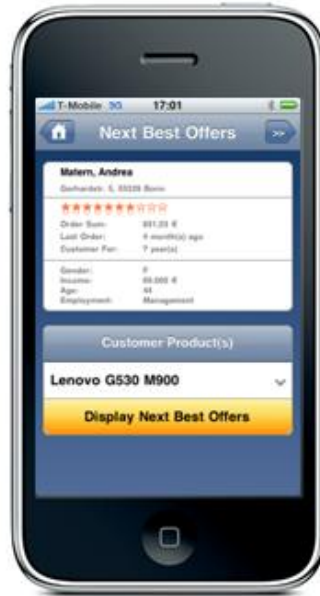
ADAPA

powered by
amazon
web services™

Model Execution



Model Execution via iPhone



Zementis Contributions

- **ADAPA:** A decision engine that deploys models expressed in PMML and executes them in real-time. Available for on-site and cloud deployments.
- **Excel Add-in:** Allows for scoring in ADAPA directly from within Excel.
- **PMML Converter:** Validates, converts, and corrects old and new PMML code. Available at the DMG website and at <http://www.zementis.com/pmml.htm>.
- **Contributing Member of the DMG:** Submitted several proposals for PMML 4.0 and already working with other members on PMML 4.1.
- Code contributor for the **R PMML package** (available on CRAN).
- **PMML Articles:** R Journal and SIGKDD Explorations Newsletter. Available for downloading at <http://www.zementis.com/manual.htm>
- **PMML Book:** Available on Amazon.com.
- **PMML Blogs:** Several blogs on PMML topics (<http://adapasupport.zementis.com> and <http://www.predictive-analytics.info>).



Thank You!

E-mail: *info@zementis.com*

U.S.A Headquarters

6125 Cornerstone Court East
Suite 250
San Diego, CA, 92121

Tel: +1 619 330-0780
Fax: +1 858 535-0227

Asia Office

19/F., Unit A
Ho Lee Commercial Building
38-44 D'Aguiar Street
Central, Hong Kong (S.A.R.)

Tel: +852 2868-0878
Fax: +852 2845-6027