**PREDICTIVE MODELING FUNDAMENTALS**

### About this course

**Learning Objectives**

**In this course you will learn about:**

* Introduction to Data Mining
* CRISP-DM Methodology
* Introduction to IBM SPSS Modeler - predictive data mining workbench
* SPSS Modeler interface

Ce cours vous permettra de vous familiariser avec :

* Introduction à l'extraction de données
* Méthodologie du CRISP-DM
* Introduction à IBM SPSS Modeler - atelier d'exploration prédictive de données
* Interface du modélisateur SPSS

# About the Software

IBM SPSS Modeler is a comprehensive predictive analytics platform designed to bring predictive intelligence to decisions made by individuals, by groups, by systems – by your enterprise as a whole.  The following video will provide an overview of the product.

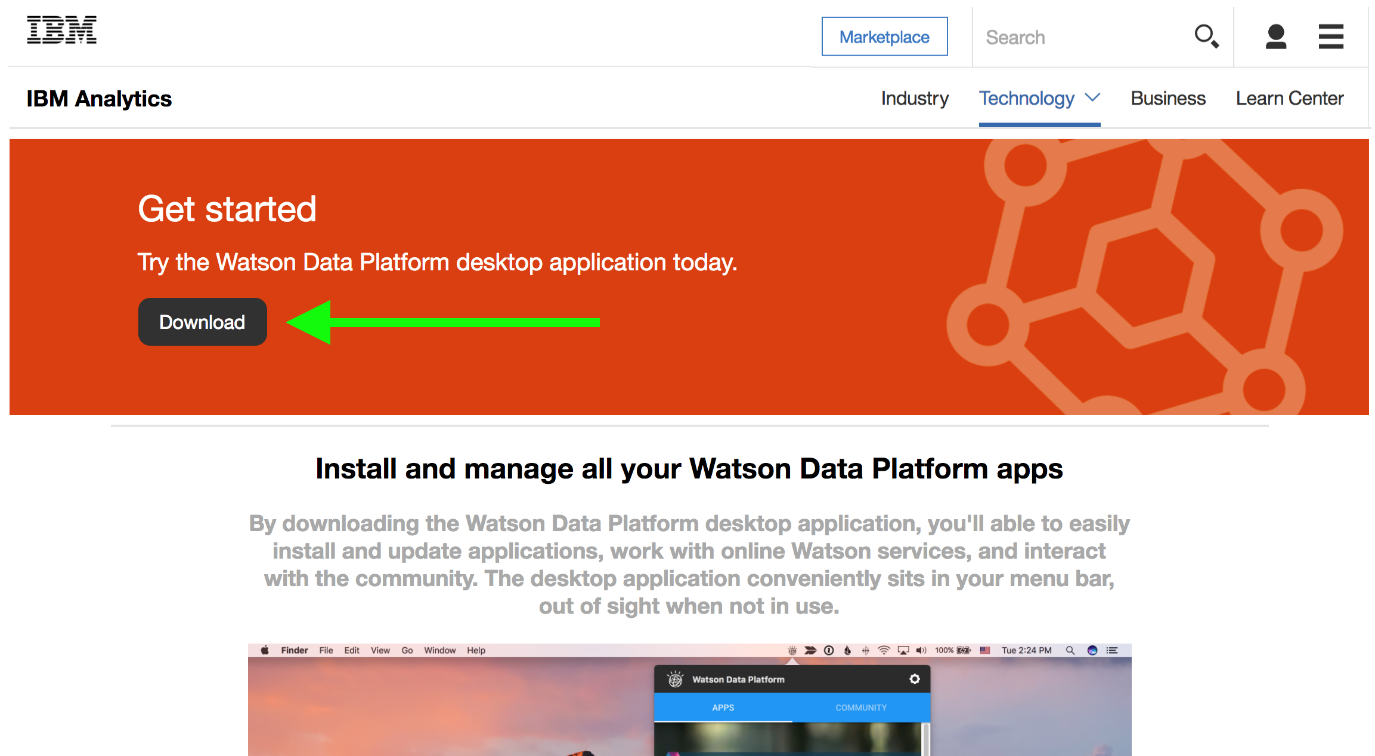
confidence it's what you want when you're making critical business decisions it can often be difficult to predict the best direction to go if your data is unclear or is growing too fast to manage with ease IBM SPSS modeler gives you the ability to successfully manage and harness your data this innovative analytics platform is designed to help you make confident business decisions with solid predictive intelligence to back you up modeler Minds the data and provide you with the capabilities to quickly build accurate predictive models using an intuitive code optional program users at different skill levels can easily gather data with the intuitive work palette by dragging and dropping fields along a customizable analytical path you can also simplify the modeling process by integrating open source languages such as our Python and SPARC use modeler to connect with distributed file systems such as Hadoop to leverage data on a massive scale now it's easy to organize key concepts utilizing both structured and unstructured data and use the data to make more informed decisions modeler tilizes common skills experimental algorithms and analytical techniques while providing strong governance and deployment stability and scalability suitable for any enterprise environment using entity analytics advanced linguistic technologies and natural language processing it automatically extracts and organizes key concepts to create meaningful category restructuring this helps remove duplicates and find any hidden relationships within the data identity resolution is vital to customer relationship management fraud detection anti money laundering and security modeling these techniques help you create comprehensive accurate profiles that become the basis for powerful predictive models with this knowledge you can improve business processes make more informed decisions and take advantage of profitable opportunities with modelur is available in three editions IBM SPSS modeler professional is the base Edition and includes advanced algorithms data manipulation and automated modeling and easy to use data preparation techniques IBM SPSS modeler premium has all the features of the professional editions plus text and entity analytics as the top-of-the-line Edition IBM SPSS modeler gold extends the premium version with analytical decision management capabilities advanced model management and more to provide you everything you need to successfully recommend actions in real time these powerful tools help your organization make the right decisions every time all additions a modeler can be fully integrated with other IBM products and solutions and are easily deployed on-site using cloud or hybriddeployment methods so what's the right solution for you desktop cloud or hybrid are you looking to integrate your predictive intelligence throughout your entire organization learn more about how IBM spss modeler gives you the tools to dig deep into your data to proactively and consistently reduce costs and increase profitability

### IBM SPSS Modeler : The power of predictive intelligence

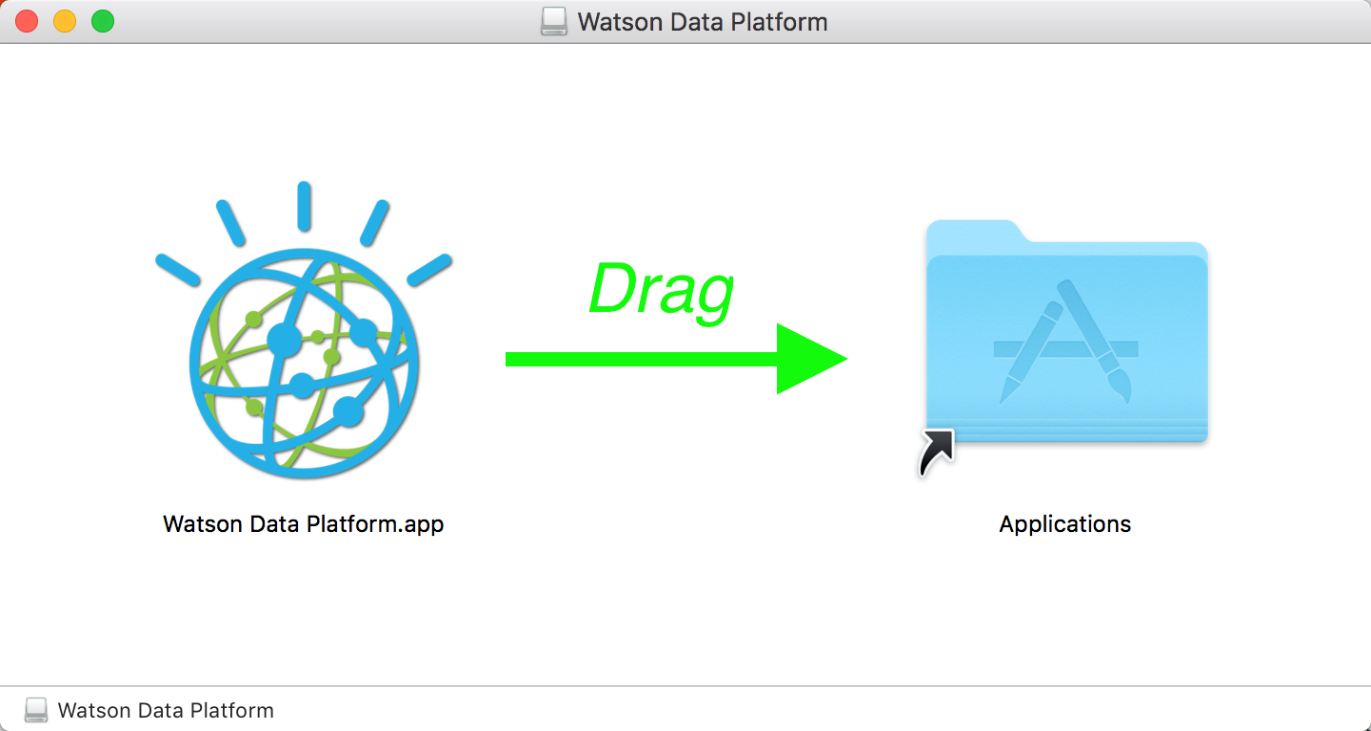
### Trial software

#### **For Mac OSX users (3 steps):**

**Step 1:** Go to: [SPSS Modeler](http://bit.ly/2p0Pj5v) and click "**Download**" to download the **Watson Data Platform app.**

[](http://watson-data-platform.online/)

**Step 2:** Open the file, and install the **Watson Data Platform** app by dragging and drop it into your Applications folder.



**Step 3:** Open the **Watson Data Platform app.** To install SPSS Modeler, simply click on the **"Try"** button and proceed through the installation instructions.

#### For Windows users:

Download the free trial version of IBM SPSS Modeler here : [IBM SPSS Modeler software](https://ibm.co/2pZLyxk)

### Community Support

Visit the [Predictive Analytics community](https://developer.ibm.com/predictiveanalytics/" \t "_blank) for up-to-date information about Predictive Analytics such as blogs, discussions, and more!

### Module 1: Introduction to Data Mining

### Learning objectives

In this lesson you will learn about:

* Introduction to Data Mining
* CRISP-DM Methodology
* Introduction to IBM SPSS Modeler - predictive data mining workbench
* SPSS Modeler Interface

Dans cette leçon, vous en apprendrez plus sur :

* Introduction à l'extraction de données
* Méthodologie du CRISP-DM
* Introduction à IBM SPSS Modeler - atelier d'exploration prédictive de données
* Interface du modeleur SPSS

### Introduction to Data Mining

and welcome to the first less and predictive modeling fundamentals one

class this is a lesson 1 and the introduction to our courts so my name is

Michaela carriage product marketing manager with a game predictive analytics

been enjoyed by Armando Ruiz who is the product manager for IBM predictive

analytics so in today's class we will set the stage for what we're going to be

talking about for the remainder of the course will do an introduction to data

mining and how it applies to business problems and will discuss the software

tools SPSS modeler a tool that where we'll be using throughout the course of

the session you'll be able to understand better understanding of what they did

when he is and what it isn't

understand the Christie a methodology which we will get into where and then

you will start accessing exploring around the SPSS modeler work so again

the agenda for today we just covered of course review

we're going to dive into the MIT followed up by Prince p.m. there will do

an introduction to SPSS modeler as a test of that next step will be a

first-lap which is its rich defense can still smile or trial and you know taking

the first steps within this will be an opportunity for you to

review laptop so just to put things in perspective to understand why we are

doing this now why we're doing this course why there are so many courses

predictive analytics and data mining popping up all over is the fact that

data is being generated an astounding rate and it's becoming cheaper and

cheaper to generated and collected once you have upstate at the question becomes

how do you expect consents from a tiny body to protect did in 2000 and

analytics they already are driving decision-making for many organizations

and business leaders say their insights are integral to their decision-making

and was time the investment and the focus analytics and extent extended used

to have a latex is only going to grow so data mining so what is the meaning you

know where the information age and a lot of the days of arable so the question

becomes how do we consume this data how do we translate to dating to information

usable hard-driving sent from it as well as data mining is the process of

discovering insights hidden parents friends and relationships from a deal

and we can accept various types of knowledge from the data so we can do

descriptive analytics reaches you know what has happened and why did it happen

this is basically business intelligence and it is predictive analytics

with the tailor what is likely to happen next in this is what's going to be the

focus off power plus just so what are some of the things are converted to

common types of algorithms to predict analytics and it to separate into

soldiers association rules which are rules that indicate relationships for

example people who buy diapers also baby people who do acts also do I things

happen in conjunction with him see how for example different product placements

can work on target on sale can influence of products products go to candidates to

this classification which is finding a model to describe the data and classify

them into service category so for example you have a data and drivers and

will provide demographic information with their driving habits and with past

expiration of whether they know you're likely to drink and drive and do not so

with this information we can identify what would properties make a person more

likely to be contrived and if we were gonna a new person we haven't seen

before

business person to interview likely to drink and it is documentation which is

being people by similarly we don't know what these people have done in the past

so we don't have a particular available and which was segmented them up but we

are broken breaking them up into crab clusters based on their similarity

amongst each other you know for example you have children single college

students of an empty nesters absolutely why they have mining important again

well it is a function agnostic and industry agnostic to the process

it can help organizations drive value regardless of whether focuses and

regardless what the use of data mining you within the organization they can

focus on working in attracting grown retainer customers they can optimize

operations taking signed out

instances of fraud can help manage risk etc so did in my news is as prevalent to

the entire business cycle and it can make a positive impact on the

organization regardless what industry it's and and what function it's

implemented so what what is the remaining it's not due to keeping about

the mining is that it allows us to uncover information and insisted he

resembles dividing up customers of a company according to the general that's

not data mining because we have this information sort of US sales of a

company that's instigation again no new information is created if we sort of

student debt ap stats not data mining nothing new happens pretty can be

predicting the outcome of tossing a fair pair of died while it may seem like its

data mining because it has predicting it but to the latter is shared by every

outcome has the same capabilities really known information get some tough but

looking at predicting future stock price of a company's historical records with

time-series data mining group customers buy this similarity that's also didn't

practice so what we're looking at predictive models and data mining

algorithms data mining algorithms the help us create predictive models by

analyzing the data to time attackers within a to help us extract insights and

then

d'amato's that we derive the algorithm statement contains the parameters of

we've covered that we've identified and then we use them to make predictions

about future cases so some cases its customers are likely to turn which

customers a more hands now he was going to be the credit risk or is likely to be

malignant or benign so do the questions directly gallery in an algorithm runs on

a set of historic game and tomorrow's general parameters and then that apply

to new instance that we haven't seen before and make a prediction about you

through the data mining process that uses and that is you know he widely used

in industry the Chris p.m. process to process industry practices for data

mining interest taxes and continues through burns and improve still

commemoration of insider to keep our citizens just to this just go wanting to

shoot pool and it goes in a circle and continuously updates and as you update

from 18 you may have to go back to previous steps and it's comprised of the

following steps as the first one is business understanding the question is

what should be accomplished from a business perspective that's what it is

did understanding acquiring the data needed to accomplish the objective

that's what is data preparation once we understand the data collecting including

the team that's the process then we may transform an aggregated for analysis

from the next one is the modeling stage will recite the technique they've been

trying to model asset accuracy then devaluation stay in the evaluation stage

we see whether the model we've built our business objectives and finally the

deployment stage is the strategy crowd actually deploying the model and get to

that will be using drugs of course his IBM

SPSS modeler so what you're looking at right now the screenshot thats the

modeler mentioned it's a comprehensive predictive analytics platform and it's

designed to bring predictive intelligence 22 decisions made by

individuals or groups and systems the Enterprise IT marketing about of scales

from desktop deployments to integration without operations system provides a

range of advanced analytics including text analytics Identity analytics social

network analysis decision as optimization and it's built with the

business user name so marketers and analysts you can use it and they don't

need to run complex algorithms with scripts they don't need to know how to

do that you'll have to be a programmer to specify smarter for those who want to

take advantage of the more complex tools or customization options that's

obviously terrible and integrates well with open source software and result of

strong and powerful data manipulation and programs available will give you

more than later to give an example there's the trip we can do turn around

us as we continue this support vector machines there's various kinds of a

cessation outcomes so it's really really powerful and comprehensive what

in the next day the next step

this lesson will be going out to actually doing a lab pans and a-level

install my workbench and see how it actually works so thank you and we'll

see next

### IBM SPSS Modeler Interface

welcome back to the predictive modeling fundamentals one source of this is less

than one the continuation in order now going to explore the SPSS modeler

workbench so the agenda for this course is we're gonna look over the model it

and its case understand what streams are due for building the predictive models

in the poll it's called I don't worry about it what that means we'll discuss

it in the condition while and then we're going to do a lab services again this is

tomorrow where Tennessee's this is what you will see when you open up my father

and as you can see it has a nice visual and a freeze in the white states have

called the screen campus this is where actually you're going to be doing your

predictive models at the bottom you see pledged wears a coat with tabs across

town and with different notes in each node's symbolizes a particular step in

the process

pulling in data to transforming Dana emerging data mining models doing grass

etc these are cold and so that's your gonna drop them into this canvas and

connected together to build a model and the upper right corner of this is where

you can a half you want streams outputs and so the great thing about building

predictive models with my daughter is that its visual and it's based on these

items were no she picked up to those from Apple n drop them on the workbench

can edit their attributes and customize them and Internet tend to each other to

actually build predictive modeling stream that has a very true flow from

the collection of data to the processing of data to the modeling of data to then

output ten you can see at the bottom of the screen shot sample stream where you

were taken some data

you're a driving issue of setting the tax return and connecting 42 Morrow

decision tree to it so the streams that evil this once you take these notes in a

canvas you can come together and then this is again this is a slow day through

these operations are notes to about the destination which can be attempts to

chart tomorrow

exports another format to can be put back into database into darkness behind

for visualization you can push out to do whatever it so there's several rules for

connecting various the source nodes that they are those for actually taking in

the data sources and bring them the starting point of a streams and other

notes can be connected after the source notes but I think for them then we have

an immediate knows which are returning future operations for the air

conditioned transformation processing P processing those can be connected before

and after them in the middle of the new terminal nodes amount of modeling note

grasp no doubt the notes etc so other notes can be connected before these

notes but not after terminals music about to enter into it if left button

new snow traction placement positioning rate no buttons were invoking context

menus to edit and in middle button to connect two notes various are really

great help menu available at times with your two trials not alert you will have

all the information around Marler around here to properties you will also include

a sample dataset and tutorials to walk through and actually do simple things to

practice on your own so over the place you know they have to stop source notes

record operations for instance graphs will go over there

would be using all of them and discourse so and if you work tomorrow taps the

continuous supply routes because there's so many different algorithms available

and if you could enter do feel that automated subtab we have automated miles

diesel models wear but it helps you ask aspiring helps you pick the best model

for UT 84404 what you're trying to do ya wanna classifier note on American Idol

costs and tension makes it will go through all of them and now that we're

gonna start with the first lab installation of IBM SPSS modeler than

getting familiar with the work

## Lab 1: Installation of SPSS Modeler

https://www.ibm.com/tryspss

### SPSS Installation Video

hi welcome to predictive modeling fundamentals one I will be covering 25

labs that are included with this training series this is lap one where we

will cover the installation of SPSS modeler and the first steps with getting

started so to get SPSS modeler you wanna go to IBM dot com slash try SPSS and it

will take you to a page that has a lot of information about not only IBM SPSS

modeler but also IBM SPSS Statistics and so you can learn more about the two

different products here and have an easy way to get to the links to get a trial

so once you navigate to modeler section just click on try and you'll be taken to

a forum where few fields in your information and once you complete that

then you move on to a page where you can choose whether you want to install the

32 bit or 64 bit version of the software and once you have that installed will

just go over a couple of things for getting familiar with the software ok so

once you've completed your insulation and you and I am SPSS modeler you should

see something like this basically you have an empty blank canvas where you can

add notes which can be found at the bottom portion of the screen so the the

tab that you see the bottom of the screen are called pallets so throughout

this series will refer to things such as their record ops palette the field on

style in modeling talent and so on so these are split up into one of these you

can easily find the note that you're looking forward to perform the task to

do so

projects going to start with bringing a datasource in two modeler so just to

show you that you can easily add a no to your canvas by double clicking and you

once you have the money came as you can move them around once you've selected to

lead to get rid of it or you can right-click and you have a lot of

options for what you want to do with it just to show you really quickly we're

going to be building in this series this this is called a stream so basically

what this does is it brings a dataset into modeler and I will be going over

building all of this but you bring in your dataset we do some processing steps

and clean up a little bit so it's ready to be put to a model and we're gonna use

a couple different model building techniques in this tutorial to show you

what options you have available so first we me but I'll just stick regression

model and then we're gonna be doing a automatic bottle building process where

it builds many different models using different algorithms and we can see

which perform the best regards to accuracy and tourism analysis on that

and we'll do that on our training data first and then come back and

and complete the same processor user model that we've built with our testing

data so this is what is going back at the end and along the way

go through each step

### Module 2: The Data Mining Process

### Learning objectives

In this lesson you will learn about:

* The Data mining process :
  + Business Understanding
  + Data Understanding
  + Data Preparation

Dans cette leçon, vous en apprendrez plus sur :

* Le processus d'exploration des données :
  + La compréhension des affaires
  + Compréhension des données
  + Préparation des données

### The Data Mining Process

hello and welcome back to the predictive modeling fundamentals 14 this is a

lesson to taking today are you can electoral-rich product marketing manager

for IBM predictive analytics and Arman worries product manager for IBM

predictive analytics so in this lesson we will pick up where we left off we

will discuss the first steps of the first-team methodology we discuss data

preparation and data preprocessing and then we will perform data preparation

was SPSS model so our agenda for today's we're gonna start with the first half of

Chris p.m. to business understanding and we will discuss the use case that we

will be using throughout the rest of this course we will go over there and

standing and go to tools for data extortion available in my earlier and

then we'll go on to test preparation and processing and we'll review the tools

that are smaller force to be able to do that and then we're going to a lab band

so the first half of christiana get it is business understanding without

knowing what exactly directives and requirements are for the project

chances for success

attending my task reviews tell you have to be able to know what you're going to

do and you have to know what you gonna look for and you have to do what data

you're gonna need and why you doing its owner keys were looking at the sinking

sinking of the Titanic use cases these shoes case Titanic ship collided with a

sturdy and most people thrive because of the lack of rest because of a lack of

light jacket so most likely to survive more like they just have a few others

where women children and the upper class and our challenge project will be to

analyze a date and price per passenger as whether they're likely to survive but

now we're going to be doing it with a damn I saw this dataset 1010 dictator

comes from capitol and can go a couple proves to you know get startled to mine

project place for the only competition for a lot of data sets a publicly

available there and it's a great way to learn and explore they get there will

you doing so for our cases to studio saturated source the training testing

dataset so we don't have two separate or attached to these two sectors

subsections so some of the tools available energy business model for

exploring the available in the underside free agency can get the data you have a

grasp or that allows you to pick our various brands based on what they do you

have two different price and distributions histograms and web brass

and rallies across that when your value so this is something we'll be using

throughout the course so when it comes to data preparation this is where we

create an initial rocky and often times that provision tasks take a while to 32

responsible steps can be repeated in four times because oftentimes the data

and real four of his dirty it's incomplete and borrow actually values

are lacking for example you cannot measure you can have a chamber of

nothing which is obviously not correct or does not have found you too can be

noisy which means there is a result they can be consistent when it comes to many

meanings 04 discrepancies between different table so we have to account

for that time we have to take care of before we can run a model that actually

makes sense

so the key task and data processing data cleansing so filling in missing values

did that's no easy identifying and moving out fliers resolving

inconsistencies 10 this data integration was just an ablation of data from

different data bases science intensive have to transform the team by

normalizing marketing sometimes if you have a deal for example you can have a

rating date of the hearing today because it's gonna be so large

thousand dollars in damage so we normalize the static transform its 021

attributes 10 2010 we have to perform data reduction which is done by

Principal Component Analysis factor analysis basically you are kidding

reduced representation positive you're trying to capture as much capability in

this much information you didn't participate in a smaller size which is

necessary to pass trusting stuff especially when you're working with

extremely low to handle noisy team and one of them is bidding you can sort the

data and the partition it into men's deaths

number of customers with C span you can also call austerity to identify where

you know where the deputy to how they live together as distinct can help us to

trust each other so they can also help us identify outlines and then obviously

expert August crucial because of my room can detect potential outliers person

that has 2008 price subject matter expert actually knows that kato knows

the stakes so how fires did objects are very different from you know the general

representation of the tapes of this stand out and sometimes fires are needed

and we have a has to be corrected but sometimes hours actually bring

invaluable information and they have to be included in the model and this is

where subject matter expertise especially really comes into play when

it comes to did after formation this couple hours of suits moves noisy data

moving all the noise we can sometimes aggregated data I can we can normalize

it and in the end max ko you know everything from 0 to for example how

does he score normalization allows us to do that we can create new features are

attributes with her principal components analysis so some other tools that are

available to us for dinner provisional IRA certificate to type this allows us

to find what type of each attribute

continues data set income page is a categorical string values the flag yes

and no word no

particular order was a tapper the type node structure is you have you can say

you did to be an input which is used as an input or predictors of the moderate

acne to target which is the outcome that we're trying to cross the picket

sometimes Bebo so used and specifically association rules from outside counsel

not to sometimes it can take food from marlin or petition that means we're

using this variable to separate it into something for training testing for

volatility and other important issue to deal with when preparing it is handling

missing values and it's important to handle the table before him to

understand how to handle this tentative quality details you know obviously

profiteering escape the garbage in garbage out a result of modern art of

you can do to help business decisions and files and fields of missing data can

produce results if they're not identified the analysis so there's

different types missing you can have an older you or no value from the military

you can have a blank space for a string variable presentation of users specified

missing counties are going to be so for example for children that they can be

covered in 1999 wishes shows you that there is a really need to be out of 10

per episode and this noted by itself

hopes you handle the data preparation tasks helps you to analyze data identify

missing

values normalized fear today if you want to do he hopes you take care of some of

the spacing work which can be if you attempt to me

number of record operations options exist tomorrow as we continue to extract

transform and load ETL process who is he can you know sample taken sordid balance

it out you have a certain class that's under represented you can worst arafat

aggregated data driving new variables Ferguson variables out reclassified from

vehicles being partitioned into training and testing the worst critic

attributable to a series of dummy variables so a lot of things that this

is a lot of things available

powerful attempts of these operations with you didn't know either creating

with Myler is claiming which the expression building would you can't beat

you can be used for example the revenue actually it is really do love you can

you know you can see you can't do everything with point-and-click you can

check it to view is an integer variable you can convert available Tuesday 23

some different variables and there's a lot of tools available here for you with

this expression

so glad we're very loud too we're gonna load attending to explore it and where

appropriate for modeling and we will see you in the course

### Lab 2 - Data Exploration and Preparation

predictive modeling fundamentals on in this tutorial what we're gonna be doing

is getting our data for a project we're going to put it in that space is modular

and start doing some preparation so the first step of this is getting our

dataset and there's a link in the document a tutorial that goes along with

the series that directs you to find the training and testing data sets but I

have that open here so it's just casual dot com slash see isn't cat's eyes

Titanic / data and so you should see something like I have on the screen here

what this is is a dataset of the passengers on the Titanic was some of

the information about them such as what they paid for a ticket for sex their age

where they were staying at on the Titanic and things like that and our

class variable we're trying to predict is whether they survived or died when

the Titanic sank so you just come out here it's to CSV files that you to

download so you wanna get the train CSV and the trustee as well as download

those files to where you work out of

and then we'll be ready to move forward I'm so once you have those downloaded

you're gonna want to go back to SPSS modeler and you can start a new stream

and what we're going to do is under the sources Palin we have a final note and

this is what you can use for importing site files into modeler so the dialogue

here it's very easy I was his final just to the right of that there's a button to

open a browser to select the father to load so just navigate to where you have

your training trial first that's we want to get in

so find this year and really quickly a nice feature we have as we can preview

the data so what we do this we can see the attributes that we have for the

dataset and their shoes in a show you the top 10 rows in the CSV file so we

have no idea where they survived the class which you can see websites in the

network right because we have a peek last name but looks like the clash as

part of the name and so that happens sometimes when you're a CSV file Wenders

quotes and the reader doesn't know exactly how to handle it so that's

alright so what we want to do here is the double quotes at the bottom

we're just gonna switch that from discarded paper discard so that's

changed

get a preview again you can see with the preview that price which in double

quotes two-parent discard 36 their data so now everything will be expected so

passenger deal experts are very busy 01 we have a class for one of the

passengers at the name the sex age and a few other details here

ok so we can click OK and now we see that this is pretty nice the nose been

remade the name of a file so fast now we know the desire training data and so we

can do now if we go to that output power had a table node and I just did that by

double-clicking with training notes elected so different way that you could

do that is to drag this on the road onto the campus and you can right click your

input node or sellers and click Connect and click on where you tryna connected

to another way that you can connect nose and you can get rid of connections by

right clicking into leading so another way that is a little quicker a little

bit short cut if you're using a mouse and you have school in the center if you

click down on that while

connect you can see that draws a line and by releasing it next to the note I

want to go out to its gonna make that connection there so it's a nice shortcut

should make easy connections between notes ok now put it will just show which

is a visual way basically to have a great of the data within modeler so I

added that made the connection right click on it and click to run and this

shows us instead of just the top 10 rows we have all of our data and a nice table

so this is good just to get a sense of what we're gonna do it did you working

with but it's not really good for doing any kind of adjustments and you can see

here that we have some no values so this is kind of a first step and we're going

to continue on with doing some more analysis for their data so we can add we

can do now is exploring since we have a data model exploring so if you go to the

output Talat you can add a data audit node and so once again we can connect

our input data to the data audit noted and you can see right away

detective there's 12 fields and I just double click that and if I click Run

this information about the data that we have so there's a nice way to do some

initial exploratory analysis and data we see the graph we have some histograms

here today

age a tribute is normally distributed are categorical variables and the psych

sex is pretty well balanced and you can see that the details you have with the

descriptive statistics you know I was gonna be there for your continuous

variables but any categorical that's going to be we do have here for sex we

have no tells you the name of categories for cabin we have a hundred and forty

eight so that's all

may not be a good predictor for us but we do see that the age that the mean or

average ages twenty 9.67 we have quality which will ensure before we have 75%

complete fields 20% are complete records this is definitely something you want to

do every time you start a project we're doing check to make sure that here

data's high-quality ok so so that's just an exploratory analysis

the next step we want to do is do some or do some preparation of our data so if

you go to the field ops palette for 11 ill C type node so what this does

and dispensing double-clicking I pretty much the campus and what this does is

this gives us our field governor tribute or variable you want to say it

our measurement type so if it's a continuous categorical or flag also a

nominal and the role that it has so some things that we can do at this point

there was some work as we can just say that the survived variable is our target

so that means that as we build models it's going to automatically know that we

want to be predicting the survived that we want to predict that one is here

whether the passengers survived your not so this is just steps are gonna go

through this kind of thing so at this point we're gonna do is actually remove

some of these pay given the role of known so you can see this is probably

intuitive but things like me that's not going to be a good predictor for us

ticket we can switch that to known that will be a good predictor for us and

they're also gonna more cabinet

in part is being done so we would go to model

gonna be taking our imports and I'm trying to predict our target and then

click Apply here and ok

so the next thing we want to do is do some more data preparation but this time

we're gonna be using a Autodata prep note that can be found at field ops so

I'm just gonna make this up on the canvas and now if I connect these two

and go in and I have options here and just want to point out that you can see

that there's a red triangle here and you'll see that once we complete this

process that's going to change so in the toilet as guides steps for how we would

have all the setup we want to keep it set as balance for speed and accuracy

when we go to settings were gonna go to prepare inputs and targets and here is

some checkboxes these all should be pressure for what we're doing here I'm

going to uncheck to reorder nominal fields because we want to keep them in

the same order to as to make it easier we can leave the other box is checked

the other thing we want to do on this screen is unchecked the transform

continuous field so if we leave this checked it's what this note will do is

normalize our continuous variables to put them on a disk or transformations on

a scale based on our standard deviation and mean that we have calculated here so

just to keep everything we don't do any transformations British go unchecked

this then click Apply here and let's click on analyzed data to run this note

and we have some new transformed classes and we also have aged transformed so

this this did work for us and so we can click OK and that completed our data

preparation steps for for this lab

### Module 3: Modeling Techniques

### Learning objectives

In this lesson you will learn about:

* Introduction to Common Modeling Techniques
* Cluster Analysis (Unsupervised Learning)
* Classification & Prediction (Supervised Learning)
* Classification - Training & Testing
* Sampling Data in Classification
* Predictive Modeling Algorithms in SPSS Modeler
* Automated Selection of Algorithms

Dans cette leçon, vous en apprendrez plus sur :

* Introduction aux techniques de modélisation communes
* Analyse des grappes (apprentissage non supervisé)
* Classification & Prédiction (Apprentissage supervisé)
* Classification - Formation et tests
* Données d'échantillonnage dans la classification
* Algorithmes de modélisation prédictive dans SPSS Modeler
* Sélection automatisée des algorithmes

### Modeling Techniques

both come to the predictive modeling Fundamentals class myself mechanic image

can and joined by a man whose product manager for IBM predictive analytics so

in this session we are going to introduce some common modeling

techniques were appointed to discuss the difference between supervised and

unsupervised worried and we will understand the algorithms that are

viewable IBM Business Partner so the agenda for today again we're going to

introduce two techniques were Canada Trust provides learning which is

questioned analysis discuss supervised learning a classic Eastern prediction

will go further into PATH station the training and testing down south

something we're seeing the previous like chair around at 8 a.m. accused case that

it would gain from the Titanic use case will talk about sampling data predictive

modeling algorithms and SPSS modeler and I don't get too excited about the times

so some of the common modeling techniques available to us if we break

down into pickpockets we have supervised learning which describes into English

classes for future predictions based training to become america's decision

trees regression analysis can use neighbors neural networks so you have a

tariff fit you have enough practice trip to China

so now you're going to build a model based on that particular train them on

that day in fact and then you can applies to a new it's a new customer we

don't know whether that customer will turn out so we're going to make a

prediction for that thirteen example supervised learning that there's a

supervised learning we're going to analyze data where these labels we r

comes up unknown to creep groups and classes for objects turn to each other

with a group of dissimilar to other groups have questioned out to some some

of the common message that we have our team is questioning hierarchical

questioning to stop fussing and other associations were analyzed therefore

events in instances that occur together for example she died percent beer

company purchased together probably toothpaste and toothbrushes also

purchased together so we look for these instances as some of the comments field

below are a priori ok so unsupervised learning to your customers whose

behavior collection of data and within to these questions the data points are

similar to one another within the same question but at the same time being are

dissimilar to appoint any other costs and cost analysis allows us to group a

dataset into these objects into these clusters and the classes are not

provided we don't know if what Chrysler Group outcome certainly the deal was

blocked so we lift tomorrow

group based on their difference dissimilarity to travel rather than

burning from Justin Bannan the flying this morning to new teen when it comes

to supervised learning crustacean prediction we have to do two things of

classification were ridiculously Journal turn fraud fraud purchase yes or no it

doesn't have to be by any

it could be multiple outcomes you can buy we can predict where customers can

buy three items that customers can buy more than three items our customers can

buy more than five but less than 10 and it took root structure across station

model based on the training set and use it to question you did this prediction

for Milan continuous fearful some predicting an owner missing out so

classification is for predicting vehicles prediction is for prosecution

prediction supervised learning in general is training and testing is very

important so we want to split heart into training second attempt except for a

plan tomorrow we are proud we want to put the 256 I seventy-five percent for

training to 24 percent 25% for testing so first we train the model on the

training dataset with existing crisis and then we tested the remainder of the

data sector was not included in the training is detected and this allows us

to evaluate them I'll compare the accuracy compared to model how it

performed in the training and testing data set up to a procedure which is the

percentage of families currently classified by tomorrow and you know we

want to see high accuracy for not just testing get a proper training and

sometimes if we have problems here we do really well with training for 10

tomorrow does not do with testing data that it's an offer

overfishing problem the model doesn't generalizable to new data that means

it's probably too complex and it starts to caption worries in the dance of the

teachers so at this point we have to go back and revisit tomorrow and may be

removed from the site which is transmitted to the future unknown

objects so we are too simple to track for for it and

dressed gatien why because an interesting reaction because we want to

deal with a smaller subset of a really really big deal except as representative

of the population so there's different approaches to patent just take a simple

template 30% of original sample which sometimes me that be appropriate for

their ballast for example we're talking to predict where they're actually going

to be benign or malignant and we're gonna have mostly but I cases but

somewhat and it's going to be reading for us to accurately predict both cases

end up with three percent of the cases were gonna do really well with

predicting bank is prepared to be able to pacify the case is cracked and so we

have a couple examples of past two temples which uses simple groups or

clusters or we can stratified sample sori select samples independently you

know not overlap acceptance straps for example men and women to sample in April

proportions or you know certain regions socioeconomic group so that our sample

proportions of these groups are appropriately represented we're

maintaining the original proportions of those variables and her has a lot a lot

a lot of different groups available for all needs of grannies from of

classification and prediction algorithms different decision tree algorithms

regression analysis neural networks generalized linear models country cuts

russians for survival data support vector machines vision to retain your

neighbor so there's a lot of actors material for clustering also we have

k-means Kohonen to step down early detection algorithm that helps to

identify potential for ice and associations algorithms farmer field so

what happens if you don't know what's coming out going to help you carry candy

store which which algorithm should you pick which have room is right for you

12 miler there are automated algorithms available whereas this is part of a

select the best algorithms for your project given you know you're trying to

protect so we have had a classifier algorithm for America for predicting

continued skis or cluster and penchant for forecasting pictures so with that

let go ahead with a third lap and here we're going to build the logistic

regression model for a tenant Dana and then we're also gonna use are mining

feature of

## Lab 3: Building Models

welcome back what we're going to do is continue our tutorial series where we

are using SPSS modeler to build models that will predict whether passengers on

the Titanic survivor died so we're going to follow along if you haven't seen the

previous tutorials you want to go back because we're just going to kind of jump

in the middle ear and we're doing in this tutorial is going over building our

first model so what we're going to do is use the logistic regression note to

create a model are going to validate the model that we created and then we're

going to use the auto modeling tools that are available in its business model

to calculate or create many different models and then develop evaluate that

downward from that to select the best option for us ok so let's get started so

you want to start with the stream that we've been working on the previous

tutorials the next thing that will want to add to this is going to be in field

ops were you want to add a filter know once you open that up you can see a list

of variables that we have in our dataset so what we want to do is read name age

which was the one transformed variable was gonna rename it to be aged sort of

age transform we're going to once again go through the variables that we don't

want to include in our model and we're gonna just click on the center column to

not include them so we're going to filter them out

last name to get involved are all the features that we want to exclude ok so

quick reply here ok now now that is ready to go to be put into a model so

under the modeling palette here I'm going to select the logistic note here

trying to the canvas and now I can connect the filter no to the logistic

model okay and you can see that this was renamed as a survived because that's

what that's the target feature that were trying to predict so this there's a lot

of settings that you can adjust and this node but we're going to keep it simple

here and switch to binomial the procedure he says we're just predicting

a binary outcome so we have a winner 0 is not multiple classes that were trying

to predict so that selected him click right here and you'll see that the

screen the stream that we've created everything that impacts that model

endowed chair in Korean and so we can see what's being affected so just to

explain that those initial exploratory analysis notes that we added to dataset

when we first started those are still great because they're not being executed

only 13 executed as our dead up pre processing steps and then a model so

what has been created here is this golden nugget which is our model

so you can double-click downing and see some more information so just looking at

the Summary tab this is information about what's been used to create a model

for Advanced tab is going to have more information for those who want to get

kind of importance to see the details were exactly how model performed just

going down we want to look at here it's okay for classification table so what

this is if you're not familiar so this is your kitchen table or miss

classification matrix is gonna give us very easily to see is the overall

percentage that accuracy of our model we see her there were 79.1 percent accurate

in predicting weather and passenger survived the Titanic sinking or they

died so you can see here what was predicted here the column shows that

zero or one was predicted and will actually was observed this is actual

values so the trace of this matrix for the diagonal here this shows us all of

our correct values so we correctly predicted for injured and 75 passengers

to be zeros and 232 be ones and then missing misclassified the the 112 and

the seventy-four year so this tells us that you know you can see her little

better predicting are survived

zeros and another one so so you can use this to do a lot of valuation to see how

your models performing looking at different metrics ok so so on the output

palette what we want to do is add analysis note just double checking and

that is going to be connected to her or model so once opined that i'm gonna run

this and it's gonna give us a little bit which is another way to look at our data

limit easier to see that we have a total corrective 705 with 17.1 percent

accuracy and 20.8% it is incorrect

this is a different way just so you know we have different options within modeler

to look at our results ok so the next thing we can do for after we have our

model is you can add a table here and I'm going to connect this to a model and

just run this table and see what we get here ok so you can see that once again

we have all of our the day they were used that went into the model and then

we have some additional columns here so this what this is showing us is the

dollar sign elder survived that shows the predicted outcome and the dogs NOP

shows the predicted probability for that outcome looking at that you can see that

we have some more data for the results actually legit logistic regression

ok so now we have that created let's let's build another model for modeling

palette and just going to school to the far left here and we're gonna have to

auto classifier Turkey and this is so we're doing as a classification task by

trying to take our data and see if we can predict the class outcome I was your

0414 the passenger survival rate so this auto classification is going to use a

lot of different techniques that are used for classification and rather than

us doing wanna tell him it's going to just do all of them and tell us which

performs the best she can see here at the bottom here just to give you some

idea that all these different note at the bottom of the screen they cannot be

used to classification they have their strengths and weaknesses and a lot of it

depends on what data you're working with the best technique but if you're not an

expert and you're not sure about which one is the best to use the auto

classification will take some of this work out of it to tell you you know

based on these different models this will perform the best ok so we have this

on-screen we can feel this is all good just want to show you that expert tab

will show you whether we want to use the different file types so you can see

these are all the different model tests we do it we are including logistic

regression we're doing decision trees reviewing Beijing so we're using some

problem facing network discriminant analysis and neural net I'm so there's a

lot of different things you can do with one shot with 1 note

rather than going in adding all these different things hiroshima testing in

trouble by one so let's get this going warning quick run here and you can see

our stream has been highlighted and so there's they're building its building on

nine different models for us so that's pretty quick and now we have our new

model notice that's that's as the output of them in the sea so we're just going

here by double-clicking this is gonna show you a graph for the top models and

you can see here the accuracy so we're going to use that as our performance

metric for these models we see that our C five decision tree performed the best

it's giving us eighty-six percent accurate so that's a really good score

and we can see here that that came in second and logistic regression in the

third and then once again you can see that

79.2 services aggression 10 matches up with what we did the first time but this

just ran it among every other model so

ok they're just going back and let's take a look more like this

see five so this if I decision tree that was created we have the full model here

so you can see by expanding this on the model screen model temps we have the

rules where where the decision trees making it split I'm so this is really

helpful to try to get some intuition about trying to understand your data

more to see why why certain things are being grouped together where the split

is taking place

viewer this is a really great way to visualize the decision tree and so crazy

about a little bit here we can see how the splits are occurring

and how many terminal least we have least nodes and see how this all looks

so that should do the same thing now with our output model so regarded output

palette and once again I'm very enjoyable analysis and node and

bookstore survived the model that we created and I'll just run this now we

see that we have a three percent correct and 16.9%

and so you know one way to evaluation models to compare the classification

matrix but another we can do that is also using the area under the ROC curve

and that can be done with just a few clicks in SPSS modeler so do that now so

their crafts we have evaluation and I managed a job that on the campus and now

what we want to do is connect our our logistic regression model nugget to the

auto classified nugget and I'll just show you how to do that so this is our

logistic regression and I'm just gonna mouse scroll wheel and drop it on the

survived that was done by the auto classification and I'm going to replace

CCA now we have connection here and now they could connections to the evaluation

so connecting two models and connecting that to our area under the curve so just

double clicking this what we want to do now is the chart type and said it gains

we're gonna switch this to receiver operating characteristic and now the

beds we don't need to make any other changes and I'll quickly run

and so this output if you're unfamiliar with ROC curve basically what we want is

our online to stay as close as possible to the far left and upper sides or edges

of of this craft so you can see here that the excess survived which is our

auto classification model output you can see that that is closer to where we want

which is that far left and up

upper side of this graph and our dollar sign lol which is our logistic model

that's a little bit closer or farther away from what we wanna see you there is

another way to visualize the performance of our model so that does it for this

utero in this week we created ten different models so first we could one

note and modeler and then we created nine or nine more with the click of a

button using auto classification note and we did evaluation looking at the

classification major to see how accurate the model and also use graphical

representation to analyze the area to compare two to four models

### Module 4: Model Evaluation

### Learning objectives

In this lesson you will learn about:

* Metrics for Performance Evaluation
* Accuracy as Performance Evaluation tool
* Overcoming Limitations of Accuracy Measure
* ROC Curves

Dans cette leçon, vous en apprendrez plus sur :

* Les mesures pour l'évaluation des performances
* La précision comme outil d'évaluation des performances
* Surmonter les limites de la mesure de précision
* Courbes ROC

### Model Evaluation

and welcome to predictive modeling fundamentals one question for you today

I myself we can encourage product marketing manager for IBM predictive

analytics and their monies product manager for IBM products hours in this

session we're going to understand time to come it took me two metrics for

classification model evaluation gonna pick up where we left off in the

previous class we're gonna talk about a plan to take tomorrow off at NTNU Tina

and then we're going to use this disparate ourselves part of performance

and accuracy

jennifer is begin by review some of the concepts from my speed record data

mining tools training and testing data sampling data federal government to

metrics for performance evaluation discussed actors as performance

evaluation and other measures for overcoming ruled petitions of accuracy

we're going to talk about our secrets for measuring how well it performs and

we're gonna go into a lab previous reports have money to squeak and broken

up into three main categories supervised learning which is classification of

production where we look at historical data an attempt to predict outcomes or

crafts items into groups based on historical day so we're going to

describe the distinguished classes for future prediction classification

classification section of suppressing deals with both categorical data

production deals with continuous efforts we employ a decision trees kenyans

neighbors at work then we use regression for prediction degrees of separation

working where we could do

we cannot know belongs to you to create groups and classes for objects with

creating these customers there have been a similarity but also have high interest

in their class objective view similar with each other for a few different to

have to check out of cost accounting methods we use our team's quest 30

question to step up and then there's association rules were reality for event

occur to campus for a couple with two things commonly by introduction and

common method we use acronyms prosecution prediction we evaluate final

performance we put fear into training to approximately $0.60 to $0.70 going into

training and 24 percent went at some point so we train the model on

retrieving section of historical deal with existing classes at school

supervisor and then we can smile did not include testing and attract attention

and this variation allows us to compare how will tomorrow

performed at the training set and the technical you want to make sure to see

if that is generalizable to new data so we don't have the overfitting problem

where they're really well in the training to the people and the test to

see high accuracy for training and testing and then we used a smile for

questions thanks Peter known so sometimes it's important for us to

sample did it because we're doing such a matter of fact

we need to some Italian we need to deal with smaller subset of it and if such a

way that the subsidies representative large population of Argentina side so we

can sometimes take a sample taken the time to stop appropriate for the

balanced where we have a large study determined from one class only here

until this coming from another crisis wears too much more rare of classes

going important for us to accurately predict in this case will resolve this

with complex samples were stratified temples where we maintain the original

portions of prices can key so for example if we have percent positive and

negative cases in the regional powers for a couple also have 99% positive

cases and 10% mixed case once we have built a model it was looking at products

for customers and prospects would like to return and we want to see how well it

does so what's the one way to do that that's true actress generally do that by

looking at the confusion matrix matrix that was classified our forces how r

here so we have two positives and negatives to accurately classified as

either yes or no and then lost that case and false positives and it doesn't have

to be a binary outcome can be a multi-dimensional data we can be

multiple classes but worth looking to compare their versatile actress is a

performance evaluation is basically a measure that would look how many total

cases a clear pic less having a degree actually classified US troops and maybe

between accurately classify that's true positives best price for that is divided

by the total number of times has its limitations even though it's very

popular tool for example you have no negative cases technical cases in the

two important for us to accurately predict everything negative

a model is going to predict everything to be positive would have really high

accuracy and working with model did really well but in reality if it failed

cases they did not field especially for predicting

tumors and is extremely important for us to identify their career case because

it's easy to overlook that there's some other measures to deal with it for

example this procession which is true positives divided by the sum of

positives and false positives of the how many people can we prove we put it to be

true then restive city which is true positives divided by 2 percent plus

false negatives how many of the outcomes actually positive to predict to be

positive and negative divided by two neighbors 130 miles we can compare we

can compare them from the performance for example we ran one final question

and other one that decision for NYC how to perform in comparison to jump suit in

using ROC curve receiver operating characteristic curve to present a

performance for binary classification model and its ability to distinguish

false positives and true so there is a draft of the event his relation to the

graph that allows us to build this metric and in the sample she we have

streamlined the replay and we have two miles of build a blue-eyed represent a

decision tree and a green line 32 percent in which is the creation of

green line representing a decision tree here has done better

logistic regression because it's closer to the upper left corner that's what we

want to smile to be want to subscribe to this party and close to the upper left

corner as possible so that let's move on to the next level we're gonna perform

and we'll see in the next class

### Modeling for Predicitions

welcome back to predictive modeling fundamental is one and this video we're

going over for scoring test data so in this tutorial what we're going to do is

kind of just keep building off stream that we've been working on the past few

labs so the last laugh we finally built some models and so we did some

evaluation so now what we really want to do is actually use our model that we

created and we want to see and limit our predictions with only those that were at

least 80% confident confident about and kind of go to next freshly I'm using our

model in a web application so they'll be in the next tutorial so this is your

work is going after testing data modeler and limit or disciplinary actions ok so

you should have this stream something looks like this stream on your in your

campus if you are just jumping in

want to go back to the first tutorials the first lives and either watch the

videos or go through the steps on your own I'm seeing get to this point we're

just going to continue moving forward with this so just to get started here

what you want to do next

we want to have our testing CSV file that you downloaded in the second lab

testing dataset for the Titanic data and we're gonna add that to modeler so

really the same as we did the first time you want to go to the sources palette at

the bottom of your screen just click and drag the wire file onto the canvas and

I'm going to double click that once again the same process as before I'm

gonna selected testing CSV file click Preview and we see once again this post

incorrectly so that's alright we know how to handle this in the double quotes

at the bottom of this dialog box we're gonna switched from discs are two-parent

discard and now just for that one change you can see now named as being read

incorrectly entered it looks pretty good

get something quick reply

and click OK and now you can see basically we're set up for now is we can

do something similar to what we did with our training data just with our testing

data although at this point we already built our model and so we've already

given the stream we've told it was important to look for so long as they're

testing data has in place it's going to be easy for us just to apply our model

so what we can do now just something I've been pointed out so far in the

environment of SPSS modeler that you see on the far right side at the top we have

a different types here and so these are pretty easy come in handy so if you're

working on multiple projects you can have multiple streams open and you can

quickly jump between them by clicking here it's gonna give you all the tables

of the graphs of the previews achieve created so that's a quick really good

thing to jump active you just what you've been working on and then models

is the third time here and this is what we're going to use at this point so

since our other classification model perform better let's go for it that way

and so this you can think of this just like something from the palace at the

bottom and check and drop it here on campus now we have this connected I'm

just gonna make a connection here directly between the testing data and

that model and to get an output just to kind of visualize something here I'm

gonna had a table to this also just click Iran and see if this works ok so

that was really quick

just to show you so if you are just joining are you did you see the last

video when you're first created the model especially the opacification and

it's running through you know a big data set and many different many different

models are trying to test out it can take some time so now the advantage of

what what we have here is rather than creating a model again we can use it

during our testing dataset and that was almost instantaneous to get those

results and you can see here we have are predicted class and our confidence so if

you recall the original the testing dataset it didn't have an actual class I

survived or died so 10 so all we had was passenger I D through and barked and

just by with this data monitor model now we have predicted 10 and the probability

for that something that you might want to do if you're doing this kind of work

is that ok that's good but I don't only one apply this model early wanna make a

prediction over a certain percentage confident and my results so that's what

we're gonna do now to do that kind of Jack so we can do is under our record

hops we have a no thats select and this is really powerful because this is where

you can select only certain roads are instances that meet your criteria so

what we can do and reconnect the filing no to the select and going in here this

dialogue is but here it's kind of a calculator lets you make your finals so

there's different ways you can use you can type it in manually what we can do

here though is the XFC survived that was the probabilities number that I showed

you from the table and so that's our variable and so what we want to do is

make sure that we only keep those that are greater 8.8

so we want to make sure that we're at least 80% confident in our results and

we can click check at the bottom of the just gonna do a throwing error let us

know via something totally wrong but that looks good so we can look ok here

and now it's at another table here so if I could help lead and tables going to

make that connection ok let's just run this right now you can see if you recall

from our previous table we have all the all the outputs and its 418 on this

table that we just selected where we live at it we only 273 and looking at

the far right column I just doing a spot check here you can see that everything

is higher than 20 which is what we want because the predictions that we feel

confident that was it for this tutorial on the next tutorial gonna be doing is

setting up a woman's account and we're going to use our model to put it onto a

web application or we can use it

dying dynamically on our website so we'll see you then

### Module 5: Deployment on IBM Bluemix

### Learning Objectives

In this lesson you will learn about:

* Scoring new data
* Deployment of the Predictive Model
* What is IBM Bluemix?
* Predictive Modeling service: Deployment in the Cloud
* SPSS Collaboration and Deployment Services

Dans cette leçon, vous en apprendrez plus sur :

* Evaluation des nouvelles données
* Déploiement du modèle prédictif
* Qu'est-ce que le Bluemix d'IBM ?
* Un service de modélisation prédictive : Déploiement dans le Cloud
* Services de collaboration et de déploiement SPSS

### Deployment on IBM Bluemix

additional employment services so into this agenda will cover my point is

whether the second will deploy of a mile and Wyoming terry's how to implement the

crowd and finally we will be able to talk about collaboration and Deployment

Services so when we're working in a data mining project when we use only the way

we split our data between training and we leaving

Train Your Mother and usually when you do is just be the data in a way that you

use 80% for example to train them and putting both 10 2010 what they used to

train them all the more important for you to know more accurate gonna be your

you're breaking while so wine is made them all those using the training and

the second is going on at this thing said

and compared them with the observed values so they can talk on the door

dataset is not big enough

you may have wrong room balance oh and in case we have her on the phone records

so we're splitting a d20 so we have a hundred records for the training and to

any

and 200 records for the testing going back to the crease methodology we saw

before that we have sixteen deaths West business understanding trying to figure

out what we're trying to do in this project

second they don't understanding what is made available and how can I get into my

environment the first episode of preparation which is basically getting

that they are ready for modeling then we come to modern fire and so actually

creating the model is generally not the end of the project because of the way

and what we try to do here is to deploy these smart model into a reason to my

man so we can get some benefit from that so for the moment they are different

solutions provided by APM they work well with SPSS let me point out here when he

says basis

solution for we share Lumix service available in the cloud and the thirties

I V YM

blaming terrorists I'm going to focus now on the cloud offering and I do it I

V emblem thanks for those who doesn't know what I V emblems of blood from a

self-service application hosting environment so they view here is that

you will be able to deploy applications and don't have to spend weeks or so the

application developers will focus only on their business logic on on the Nov 10

you don't have to be so the application developer for example what happened to

worry about how to install or manage the runtimes framework and libraries and

we're opening a bunch of different viruses there that are easy to manage

and you can tell that killed them as you know so I is based on Monday which is

open source and it is a very strong and growing community so you can access it

on the net and in there you will find many different and that when we are we

are interested who do they look pretty another service offered to do the

developers and at the time the weather to integrate capabilities into

applications so basically went to the end and that's what we will be going

into the blood stream to develop smaller and you uploaded into the area and its

gonna provide you the great point APA's so you will be able to call the into

your applications deployed so it's very simple to create the ferry you upload

the file and the service provider directly the point finally we have

deployments have a larger solution of the possible

and frame and we are in this training but this is so that I know you love

number five

according to the dentist uses my mother into the IBM Cloud and you will use this

application thank you very much

## Lab 5: Deployment of model on IBM Bluemix

predictive modeling fundamentals one this is and in the slurry going to go

over the deployment of your model on IBM so this is the fifth part of the serious

in the first four parts we went over the installation of SPSS modeler getting

started getting familiar with using the software and then we built our first

model where we used simple data that was from goal.com on the regarding the

Titanic and we built in model and now we're ready to use our model and web

application so if you are just joining it is your first video word I'm not

gonna be going over where how we got to this point so you want to go back and

read the first tutorials or watch the previous videos to kind of get to this

point so what I want to cover here is what we need to do with modelers so that

we can use what we just created in a blue makes application then I'm gonna

show you IBM blue makes briefly and just kind of do a walk here for what needs to

be done to use our the model we created with the application so first we need to

do is get our stream ready and we want to do is have a simplified stream which

is our testing stream without the data preparation and we're going to do now is

make this our scoring branch so if you want to use a stream that you created on

web service a good mix of scoring branch so we we do this is all you have to do

so that the table that you want its output from the embargo and we want to

keep our select in here because we only one that we want to make predictions on

those that were at least 80% confident about if you recall from the previous

video and so by right clicking the table we just clicked you Scott uses corn

bridge and then uses this is highlighted as are so this time while we've been

working I haven't saved my stream it so let's just do that so we can use these

extreme

web application to put this on my desktop

so what's going to stop and say this is the Titanic and we're going to say this

ok so what we need to do as good as marlin done there we can get on Luminex

so just go up to make stand and login if you haven't used it before you can

create an account for free it's really have an account it's free to do a lot of

different things especially if you're just testing out the service and trying

to see how everything works it's pretty cool for all the tools that are

available on it in this tutorial when I can focus on everything you can do which

is one small aspect of how you can incorporate your models on Blu mix

login here something that you click on dashboard of the top 10 logged in and

this is what it looks like one year you have an account on the site and I did

this previously but let's just go from scratch that create an app here so

undercounter have just click Create and is going to be for a website so well and

we can click liberty for a job for getting started just click Continue here

and then we can give it a name so i'm gona just called this Titanic

but what we all we really want to do is add our ability to do a model on our

weather so we can do that by going to catalog at the top of the screen and

this catalog

this is all the services that are come built in with you make sure there's a

ton of different things you can do with a click of a button but let's just go

here on the left side we can filter some things out so it's only look at the data

and analytics and what we want to do is take this predictive analytics click on

that ok so now what I want to do is touch by tectonic demo for the app here

and click Create

get a question to restaged applications so just quickly stage that's just going

to get their application rate ago probably going to do that because you

said the service so what's that stage the wall or go forward

and so under services you can just click on predictive analytics and that is

going to take you to the predictive analytics service that we had it so they

do is our new model stream you can see we can just drag our stream to this

location and it'll uploaded or I can do

/ file here I just say this one of my desktop every working on so I tended to

give us an option we want to put in a context I T and you want to keep track

of this because we use it and the upcoming steps so I'm cynical this

titanic

ok and quick deploy

the context I D and we have our our stream file here and this is our models

that we have active use for this for the next appeared this is a website that was

built to do a demonstration for this so I'm gonna go to SPSS training and

there's a website that's built where we can put in our context and access key if

we get from our index dashboard and we can get our results for our predictions

here so just packing up his stuff wanna go back to dashboard application over

you want to show credentials for predictive analytics here and here we

did is our access key interviewer that we need to make a copy of it was the URL

so she remember our context idea was taped an egg and access key is given

here

so we put in this application is using the Stream there we built to make a

prediction for the probability that the passenger would have survived or died so

and what we have here are the different features are the predictors and how that

affects our prediction okay so it's just to see if we have a 10 year olds ok so

you will get some error messages out of bounds said something that could be

improved in a future iterations for something like this you see we have over

20 percent in 20 year old 87.9 percent chance that they would have died from it

and see if we can get someone is a bit older worst higher percentage email

maybe ok that's out of its shell

for females 43 with no siblings on board

83.7 percent chance of surviving higher fare so the higher fare brought of the

probability of survival little bit ok and $52 fair even higher probability see

we switched to an email that shows us some challenges with some of the stuff

but the general idea is pretty cool that we're using a model that we've built on

our web application ok so that completes the stories for the predictive

fundamentals one so basically in this series we are familiar we installed a

special electro we use it to build multiple models were used as models for

deploying a real weather station so what will your phone work more comfortable

using this software and how quickly it is to get a lot accomplished so long

with this training and I wish you look as you continue on with SPSS modeler R