

PHI-Integration.com

Simple Credit Approval Model using Decision Tree in Azure Machine Learning

Machine Learning Introduction Series

Contents

Pendahuluan	2
Azure Machine Learning	3
Case: Credit Approval.....	4
Upload Dataset	5
Membuat Experiment dan Memasukkan Dataset	7
Melihat Isi Data	9
Menambahkan Module Split Data	11
Melihat Hasil pada "Split Data"	13
Membuat Model Decision Tree	13
Melakukan Prediksi pada Testing Set.....	16
Melakukan Evaluasi Model	17
Membuat Web Service	18
Deploy Web Service	18
Menggunakan Web Services pada Excel	20
Kesimpulan	21

Pendahuluan

Sistem IT di organisasi dan perusahaan di dunia saat ini termasuk di Indonesia menghasilkan data yang sangat banyak tapi minim informasi. Istilahnya "data rich but information poor". Sebagai contoh, data penjualan terkumpul sangat banyak tapi kita tidak tahu mana sebenarnya produk yang jika dipaketkan akan menghasilkan penjualan lebih besar.

Seperti kata peribahasa, "Knowledge is Power" dimana orang yang memiliki pengetahuan akan memiliki kekuasaan atau kontrol yang besar. Namun tanpa informasi yang bermanfaat tentu tidak ada *knowledge* atau wisdom baru yang didapatkan.

Untungnya, saat ini telah ada ilmu Data Science dan teknologi Big Data yang mampu menghasilkan informasi tersebut dengan mengolah "lautan" data tersebut secara otomatis.

Namun kadangkala ilmu Data Science ini menjadi momok karena harus memiliki campuran pengetahuan akan statistik, bisnis dan *computer science*. Lalu kemudian banyak yang bingung, bagaimana memulainya?

Tapi jangan kuatir, saat banyak teknologi yang makin meringankan beban untuk tetap mampu produktif menggali informasi jauh lebih baik dengan belajar secara bertahap. Salah satu teknologi yang mampu memudahkan tersebut adalah Azure Machine Learning.

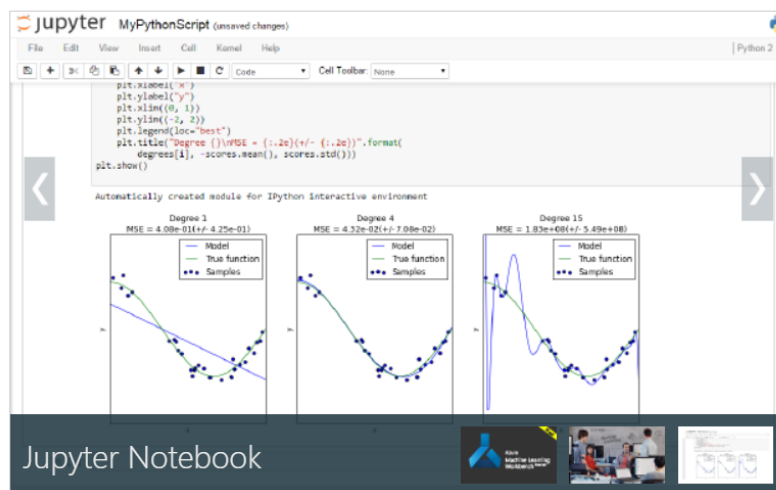
Azure Machine Learning

Azure Machine Learning adalah solusi machine learning berbasis cloud yang disediakan oleh Microsoft.

Untuk merancang dan menguji flow machine learning ini Anda dapat menggunakan apa yang dinamakan Azure Machine Learning Studio yang terdapat pada url berikut.

<https://studio.azureml.net/>

Anda perlu registrasi dan sign in terlebih dahulu sebelum bisa menggunakan.



Welcome to Azure Machine Learning

Try it for free

No [Azure subscription](#)? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

[Sign In](#)

Not an Azure ML user?

[Sign up here](#)

[Pricing & FAQ](#)

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

Cakupan dari tutorial adalah praktek langsung sehingga Anda mendapatkan gambaran mengenai penggunaan Azure Machine Learning. Untuk konsep dan lainnya mengenai machine learning tidak termasuk dalam pembahasan.

Mari kita mulai praktekkan *step by step* dengan menggunakan case dari platform DQLab.id.

Case: Credit Approval

Untuk memulai kita akan, kita akan menggunakan contoh kasus persetujuan pengajuan kredit (credit approval) dari industri keuangan.

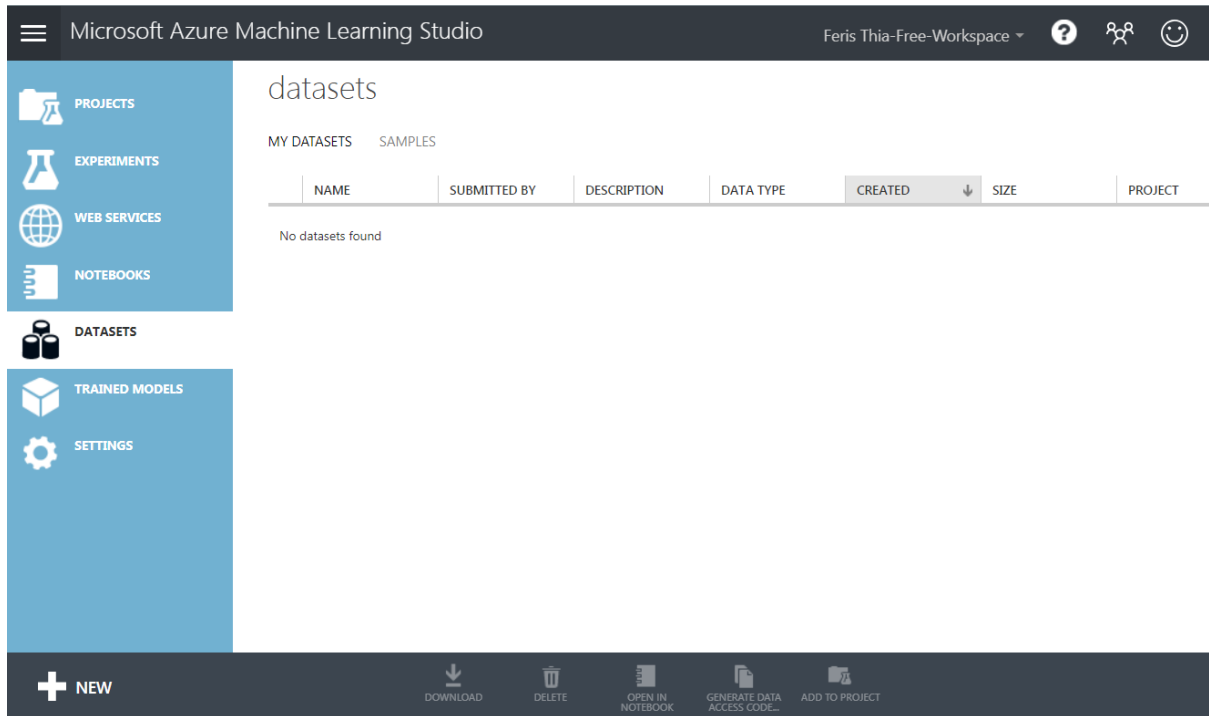
Tahapan-tahapan pembuatan solusi machine learning adalah sebagai berikut:

- Upload Dataset ke Azure Machine Learning
- Membuat Experiment
- Membaca Dataset di dalam Experiment
- Membagi Dataset sebagai Training dan Testing Set
- Menjalankan dan Mereview Hasil
- Setup dan deploy web service "Credit Approval"
- Menggunakan web service

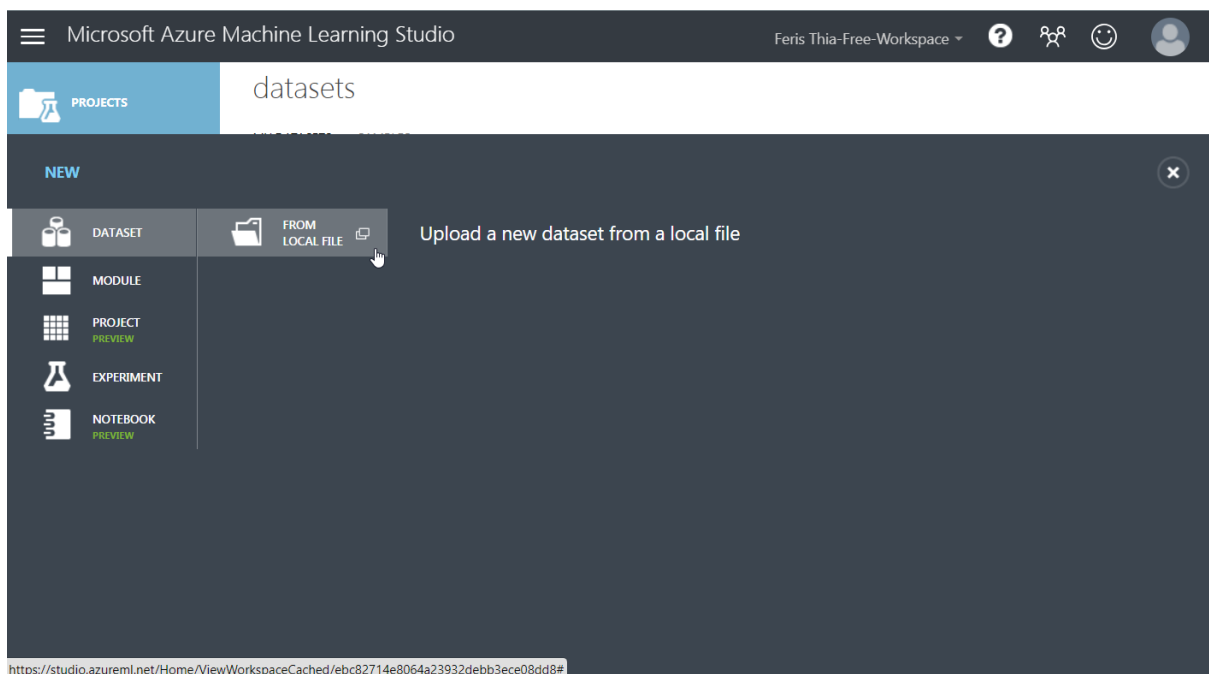
Upload Dataset

Langkah pertama yang perlu Anda lakukan adalah mengupload dataset yang dapat Anda download di <https://www.phi-integration.com/dataset/bank-data.csv>. File ini merupakan modifikasi sedikit dari file asli yang diambil dari url <http://facweb.cs.depaul.edu>.

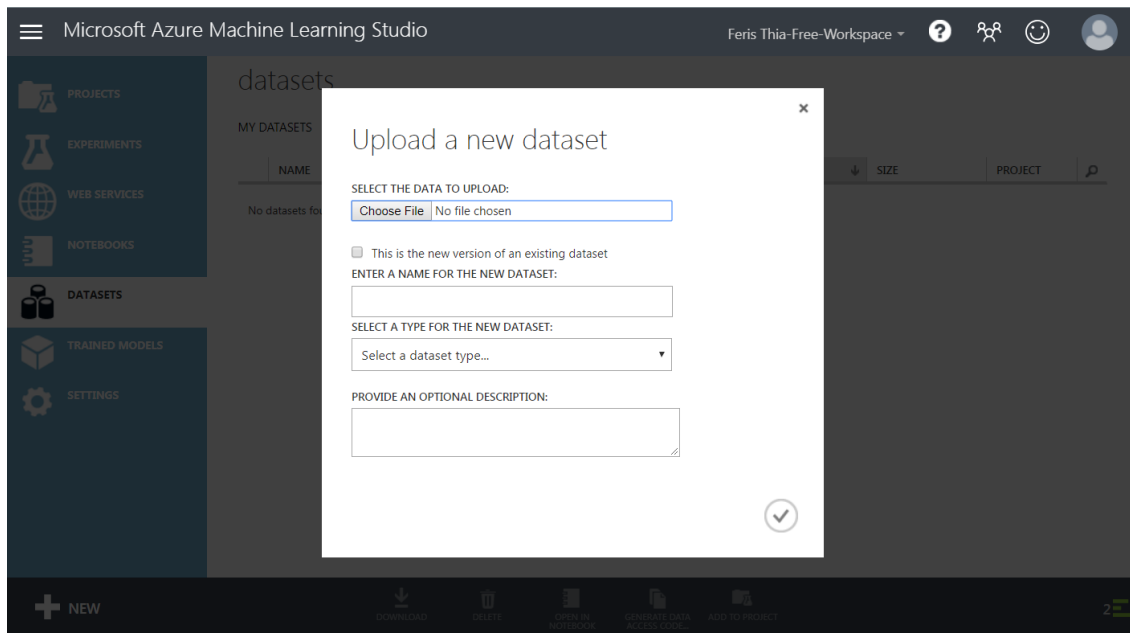
Pada Azure Studio, klik menu Datasets sehingga terlihat pada gambar berikut.



Klik tombol New pada bagian bawah kiri menu.



Muncul dialog sebagai berikut, upload file yang telah Anda download sebelumnya.



Akan muncul isian sebagai berikut, klik tombol centang.

Upload a new dataset

SELECT THE DATA TO UPLOAD:
Choose File bank-data.csv

☐ This is the new version of an existing dataset

ENTER A NAME FOR THE NEW DATASET:
bank-data.csv

SELECT A TYPE FOR THE NEW DATASET:
Generic CSV File with a header (.csv)

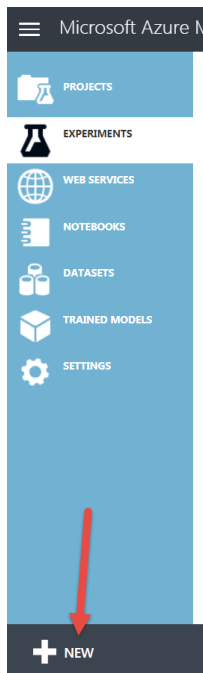
PROVIDE AN OPTIONAL DESCRIPTION:
Dataset untuk membuat model Credit Approval

Jika berhasil, maka akan muncul daftar dataset baru dengan nama bank-data.csv.

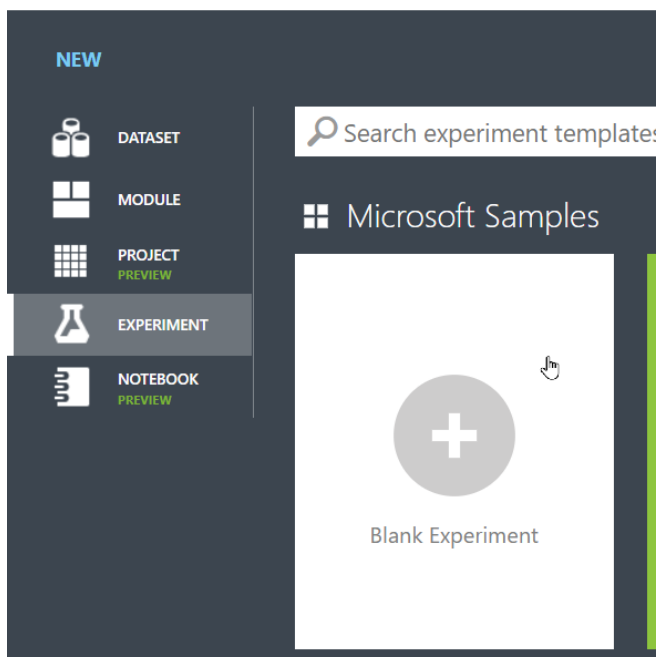
MY DATASETS		SAMPLES		
	NAME	SUBMITTED BY	DESCRIPTION	DATA TYPE
<input type="checkbox"/>	bank-data.csv	feris	Dataset untuk membuat model Credit Approval	GenericCSV

Membuat Experiment dan Memasukkan Dataset

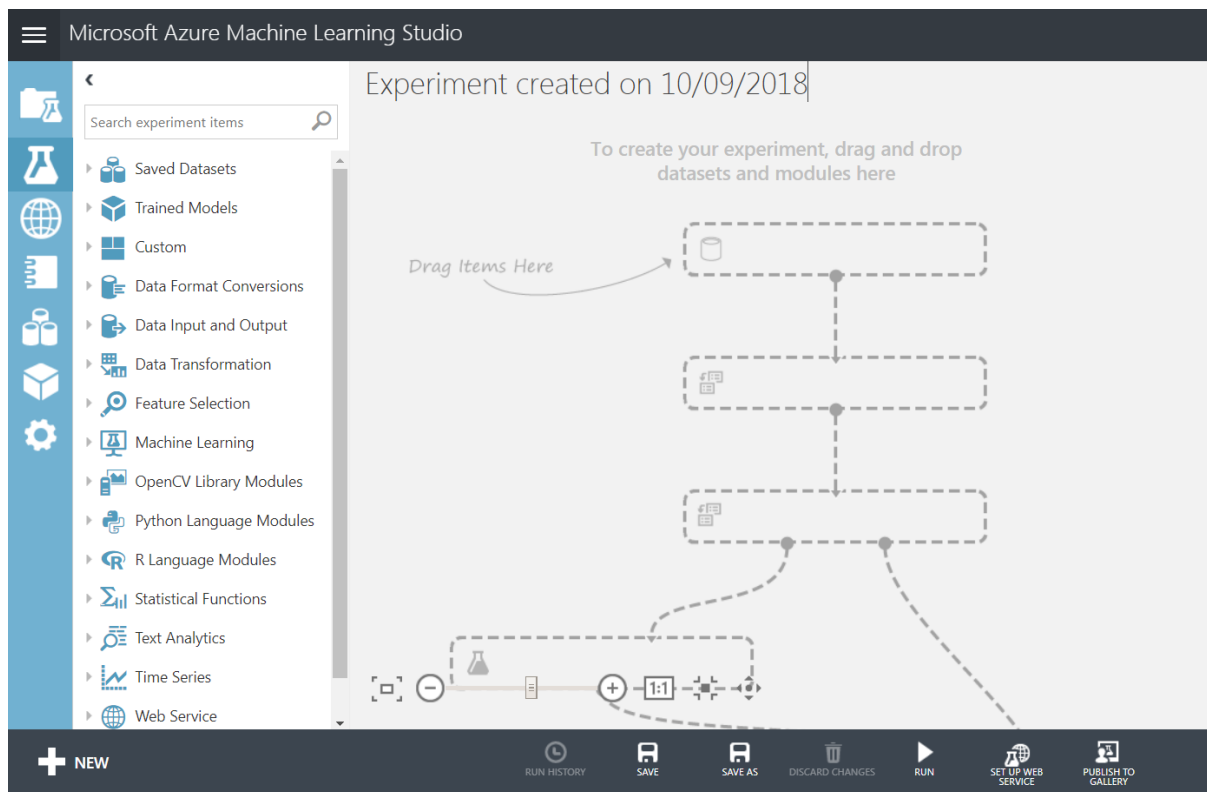
Tahap berikutnya, kita akan membuat experiment. Klik menu Experiments dan klik tombol New seperti pada gambar berikut.



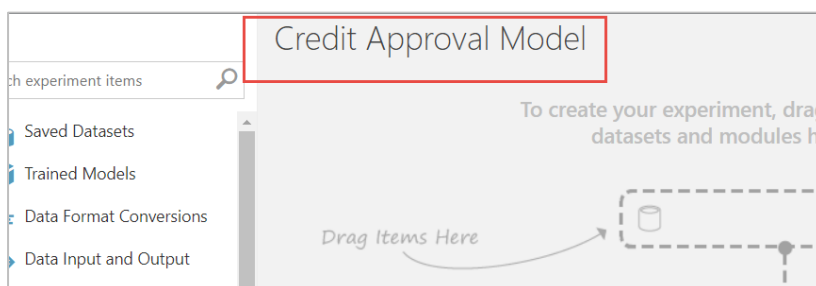
Dan pilih template Blank Experiment.



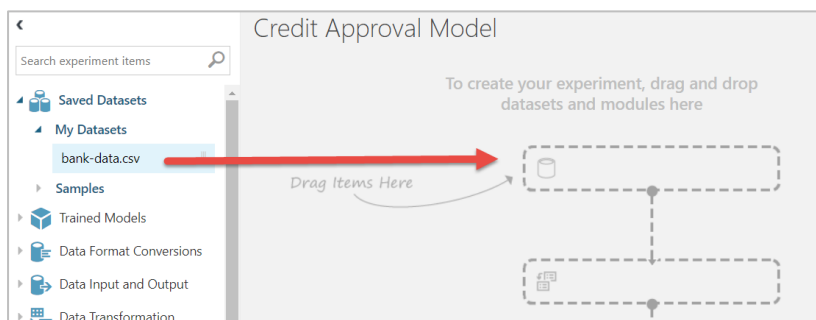
Dalam beberapa saat akan muncul workspace experiment yang terlihat seperti pada gambar berikut. Pada bagian kiri panel terdapat banyak icon-icon yang bisa di-drag ke dalam workspace, icon ini disebut sebagai module.



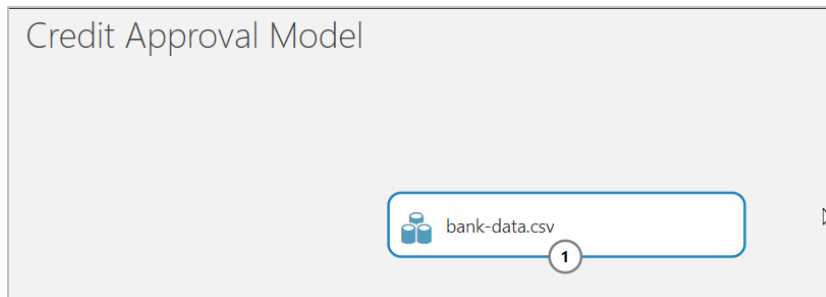
Rubah judul dengan klik pada bagian tanggal dari workspace, dan berikan judul baru dengan nama "Credit Approval Model".



Cari dan drag module pad menu Saved Datasets → My Datasets → "bank-data.csv" ke dalam workspace.

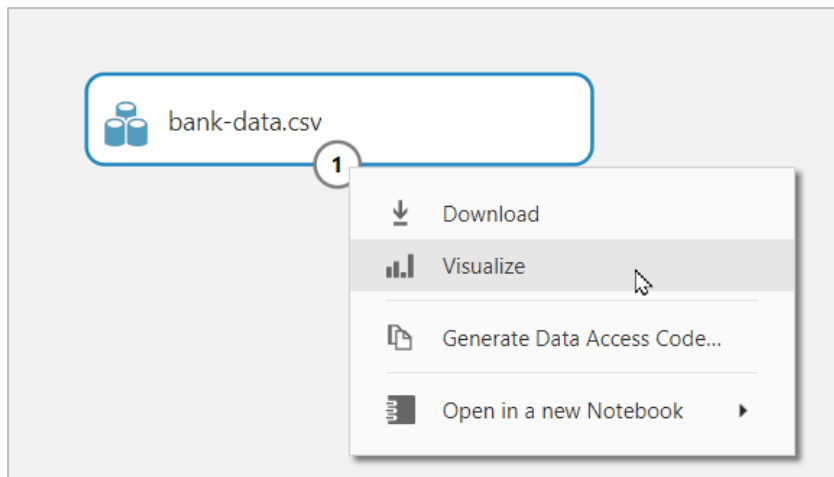


Jika berhasil akan muncul tampilan di workspace sebagai berikut.



Melihat Isi Data

Untuk melihat preview data dari dataset kita, klik pada bulatan kecil – yang mewakili port data - bernomor 1 pada kotak "bank-data.csv", pilih Visualize.





Akan muncul beberapa preview data seperti terlihat pada gambar berikut, yaitu:

- Jumlah baris data terbaca sebanyak 600 (bagian **rows**).
- Jumlah kolom sebanyak 12 (bagian **columns**).
- Nama-nama kolomnya adalah sebagai berikut:
 - id: nomor unik untuk identifikasi aplikasi pengajuan.
 - age: umur dari pengaju aplikasi kredit (dalam angka tahun).
 - sex: jenis kelamin dari pengaju aplikasi kredit (MALE / FEMALE).
 - region: daerah dimana pengaju aplikasi tinggal (inner_city/rural/suburban/town).
 - Income: jumlah pendapatan dari pengaju aplikasi (dalam angka).
 - married : apakah pengaju aplikasi telah menikah (YES/NO).
 - children: jumlah anak yang dimiliki (dalam angka).
 - car: apakah pengaju aplikasi memiliki mobil (YES/NO).
 - save_acct: apakah pengaju aplikasi memiliki rekening tabungan account (YES/NO).
 - current_acct: apakah pengaju aplikasi memiliki rekening di bank yang sama saat ini (YES/NO).
 - mortgage: apakah pengaju aplikasi memiliki kredit kepemilikan rumah (YES/NO).
 - loan_approved: apakah pengaju aplikasi disetujui aplikasinya (YES/NO).
- Distribusi sepuluh nilai terbesar yang terlihat sebagai histogram kecil pada kolom.
- Tampilan seratus baris data pertama, gunakan scrollbar di samping kanan untuk melihat seratus data tersebut secara detail.

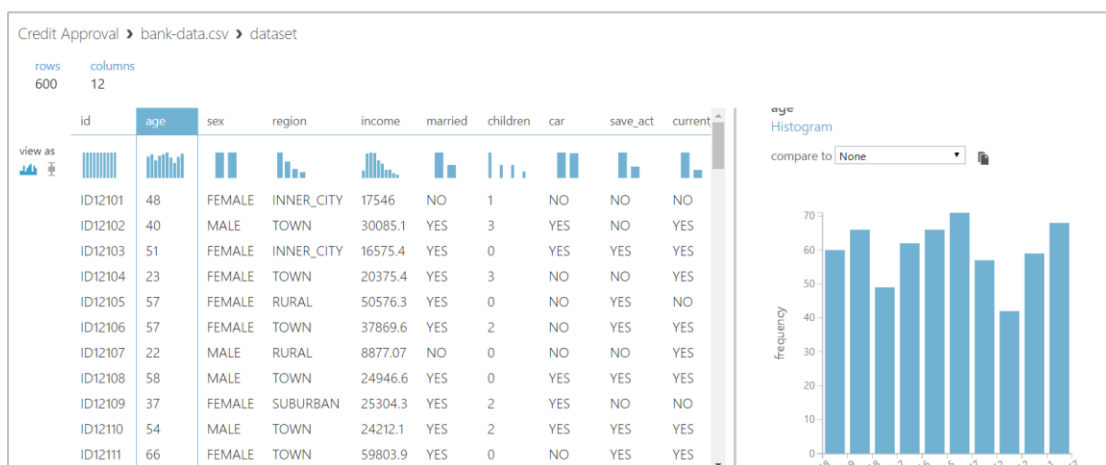
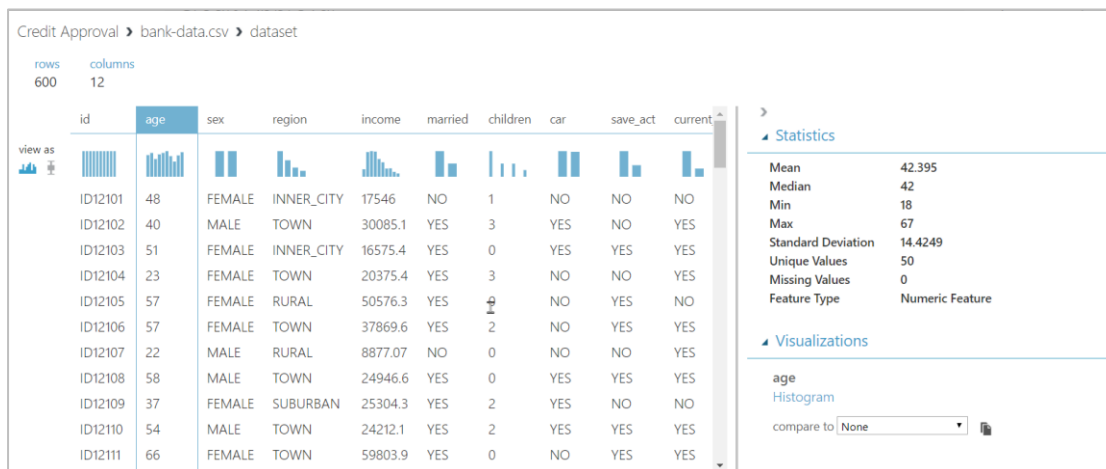
Credit Approval Model > bank-data.csv > dataset

rows: 600, columns: 12

view as:  

id	age	sex	region	income	married	children	car	save_act	curr
ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	NO
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO	YES
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES
ID12105	57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO
ID12106	57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES
ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO	YES

Sekarang klik header kolom "Age", pada bagian panel kanan akan muncul informasi distribusi nilai terkait kolom tersebut untuk dapat kita analisa.



Ada dua bagian dari informasi tersebut yaitu:

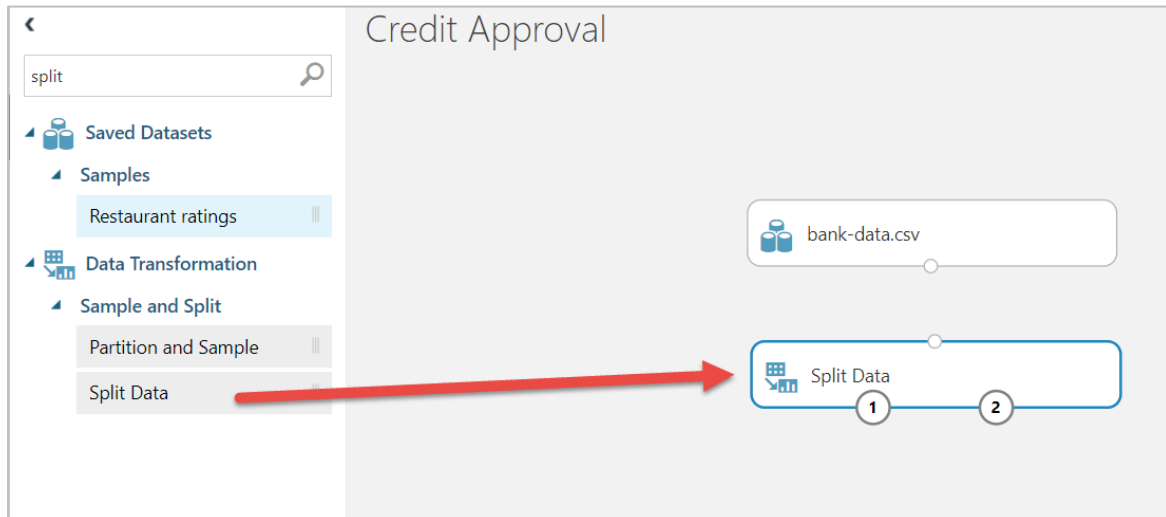
- Statistik, berisi informasi distribusi angka. Sebagian informasi yang terlihat pada gambar ada 50 nilai unique (Unique Values), tidak ada missing value, dan tipe datanya adalah angka (Numeric feature).
- Visualisasi, yang merupakan chart yang menunjukkan karakteristik kolom tersebut. Pada gambar terlihat histogram dari sepuluh rentang umur dengan frekuensi data.

Dengan cara yang sama cobalah analisa kolom lainnya dari preview dataset tersebut.

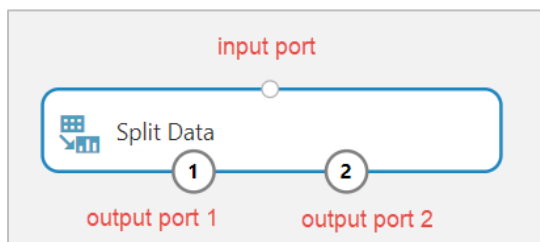
Menambahkan Module Split Data

Kita akan melakukan pembagian data menjadi training set dan testing set untuk algoritma decision tree. Training set adalah data yang akan digunakan untuk melatih algoritma membentuk model, sedangkan testing set adalah data uji untuk melihat sejauh mana akurasi didapatkan.

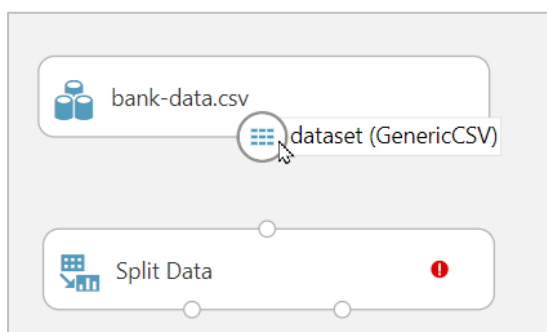
Pada workspace kita sekarang tambahkan module "Split Data" dari kategori "Data Transformation" → "Sample and Split" pada workspace kita.

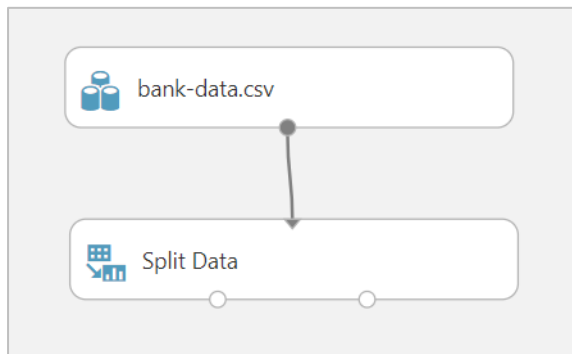
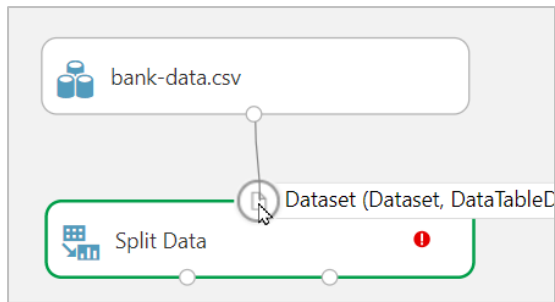


Terlihat ada satu port input dan dua port output dari module "Split Data".



Hubungkan kedua module dengan cara dengan cara mengklik bagian output port dari dataset "bank-data.csv", drag ke input port dari module "Split Data" dan lepaskan seperti terlihat pada beberapa langkah pada gambar berikut.

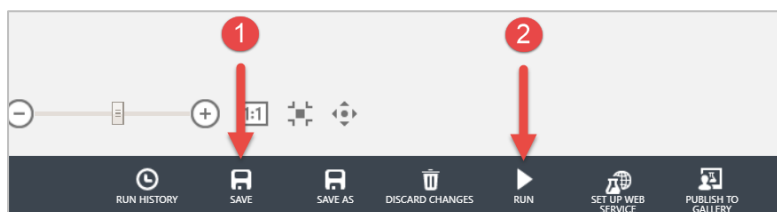




Klik module "Split Data", dan pada panel kanan masukkan konfigurasi berikut.

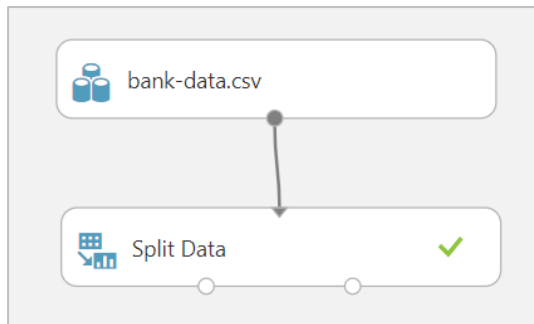
- Fraction of rows in the first output dataset: 0.8. Ini artinya 80% data akan diberikan output port 1, sisa 20% ke output port 2.
- Randomized split: false (tidak dicentang). Ini artinya data dibagi secara teratur, jadi tidak akan acak (random).

Simpan dan jalankan experiment kita melalui tombol "Save" dan "Run" yang terdapat di panel bawah workspace kita.



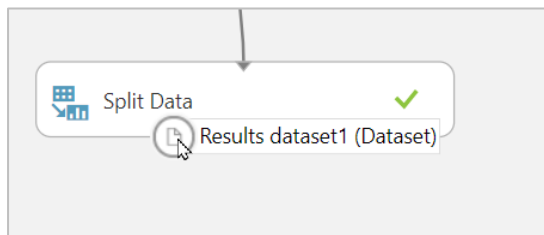
Azure Machine Learning Studio akan melakukan proses pembacaan data dan eksekusi modul "Split Data".

Tunggu sampai module "Split Data" memiliki tanda centang hijau, yang artinya eksekusi telah selesai.



Melihat Hasil pada "Split Data"

Sekarang kita akan melihat hasil eksekusi sejauh ini untuk module "Split Data". Klik pada port output 1 dan pilih Visualize.



Akan terlihat hasil sebagai berikut, dimana baris data adalah 480 atau 80% dari jumlah data keseluruhan sebanyak 600. Porsi data ini yang akan jadi training set.

Credit Approval > Split Data > Results dataset1

rows	columns								
480	12								
		id	age	sex	region	income	married	children	car
									save_a
view as									
		ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO
		ID12102	40	MALE	TOWN	30085.1	YES	3	YES

Sekarang klik pada port output 2 dan pilih Visualize. Terlihat jumlah data yang divisualisasikan adalah 120 atau 20% dari 600.

Credit Approval > Split Data > Results dataset2

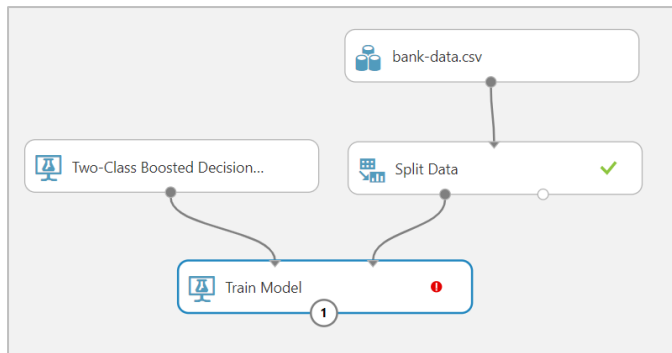
rows	columns								
120	12								
		id	age	sex	region	income	married	children	car
									save_a
view as									
		ID12581	30	MALE	RURAL	11536.2	YES	2	NO
		ID12582	47	MALE	INNER_CITY	44658.6	NO	2	YES

Terlihat ada dua rule yang ditampilkan dimana left hand side dari rule tersebut mengandung item "Dompert Flip Cover". Terlihat juga metrik Support, Confidence dan Lift terkait rule yang ditampilkan.

Membuat Model Decision Tree

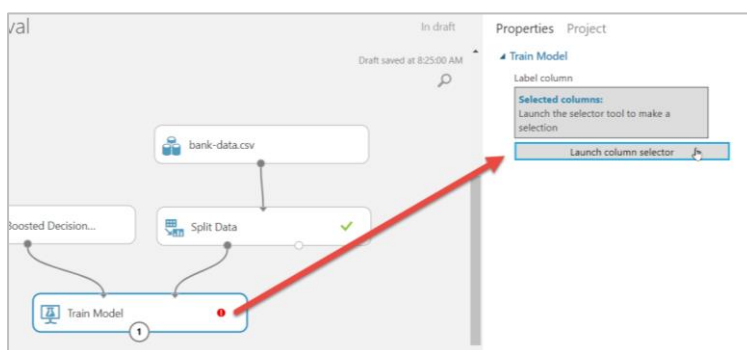
Untuk membuat model decision tree, maka diperlukan dua module dari kategori "Initialize Model" untuk memilih algoritma dan **"Train Model"** untuk membentuk model. Dan untuk decision tree sendiri, masih banyak algoritma yang bisa digunakan. Disini kita akan menggunakan module **"Two-Class Boosted Decision Tree"**.

Masukkan module "Two-Class Boosted Decision Tree" dan "Train Model" ke dalam workspace. Hubungkan semua module sehingga terlihat seperti pada gambar berikut.

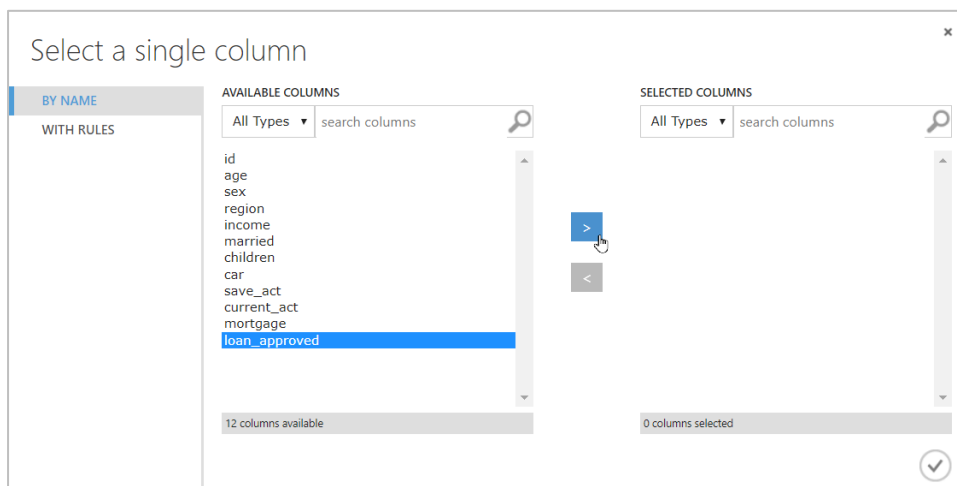


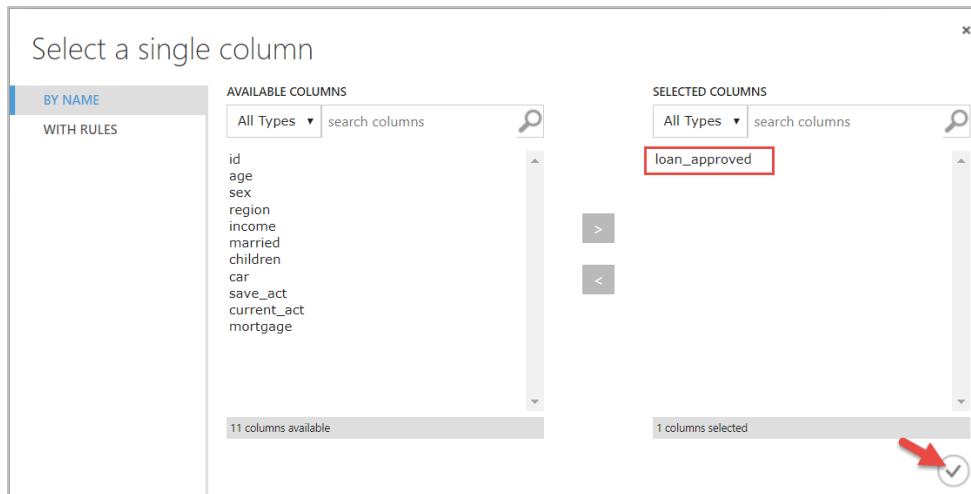
Perhatikan ada tanda merah di module "Train Model", ini karena kita belum memasukkan kolom yang akan diprediksi (Label Column).

Klik pada module tersebut, dan pada panel kanan klik "Launch column selector".

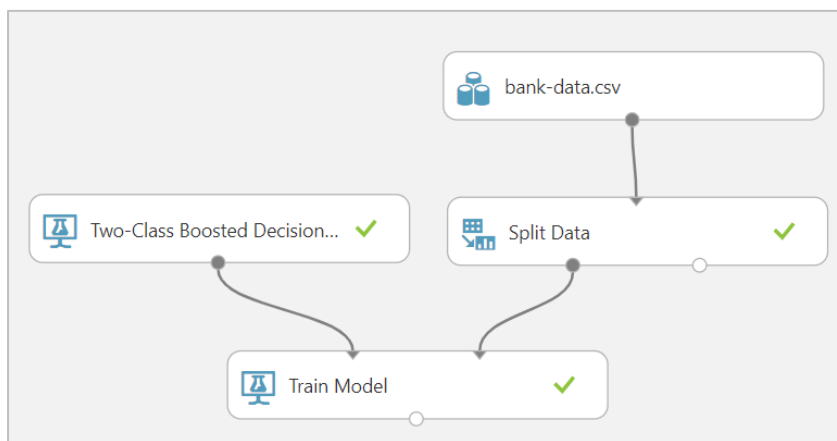


Pada saat muncul dialog "Select a single column", pilih kolom "loan_approved" dari bagian "Available Columns" untuk dimasukkan ke bagian "Selected Columns".

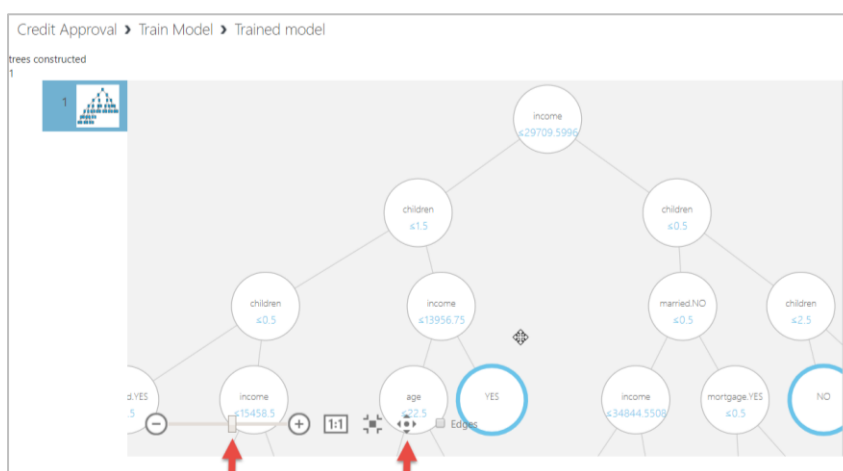




Klik pada module "Two-Class Boosted Decision Tree" dan rubah konfigurasi "Number of trees constructed" ke nilai 1. Simpan dan jalankan experiment kita melalui tombol "Save" dan "Run" yang terdapat di panel bawah workspace kita.

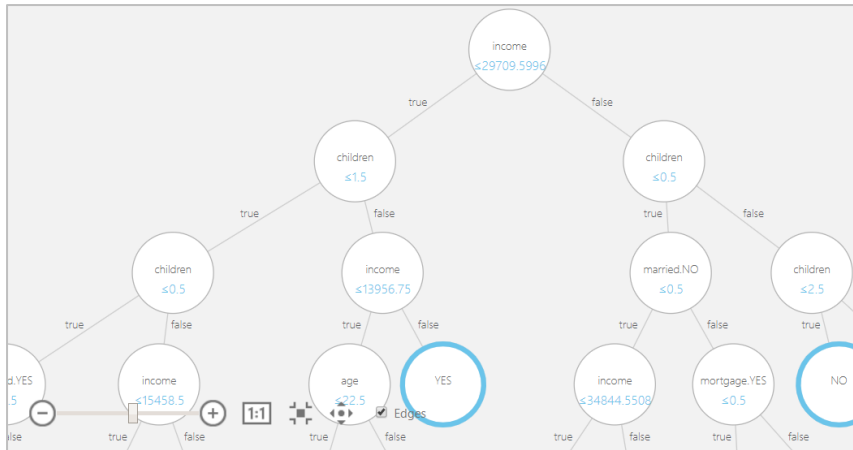


Klik output port dari "Train Model", dan pilih Visualize. Akan muncul satu decision tree dengan node berisi informasi variable dan nilai, serta cabang keputusannya.



Atur tampilan dengan tombol slider dan pan yang ditunjukkan dengan tanda panah.

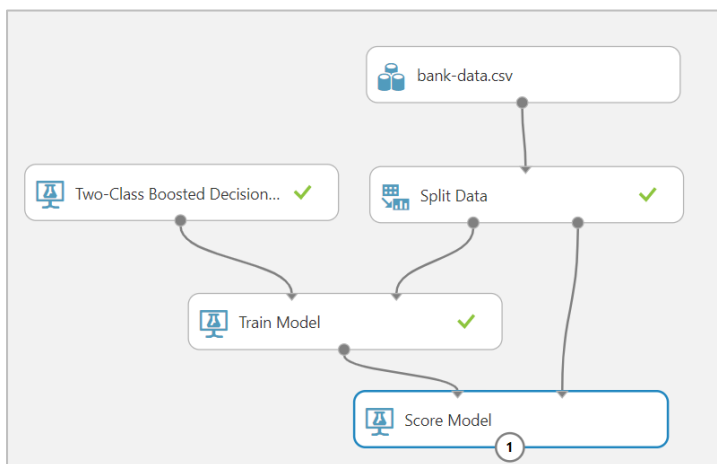
Klik opsi "Edges", maka akan muncul kondisi True dan False seperti berikut.



Tutup dialog hasil output ini.

Melakukan Prediksi pada Testing Set

Masukkan module "Score Model" dan susun workspace sehingga seperti berikut ini.



Tidak ada yang perlu dikonfigurasi lebih lanjut, simpan dan jalankan experiment kita ini. Setelah selesai, klik output port dari module "Score Model", dan pilih Visualize.

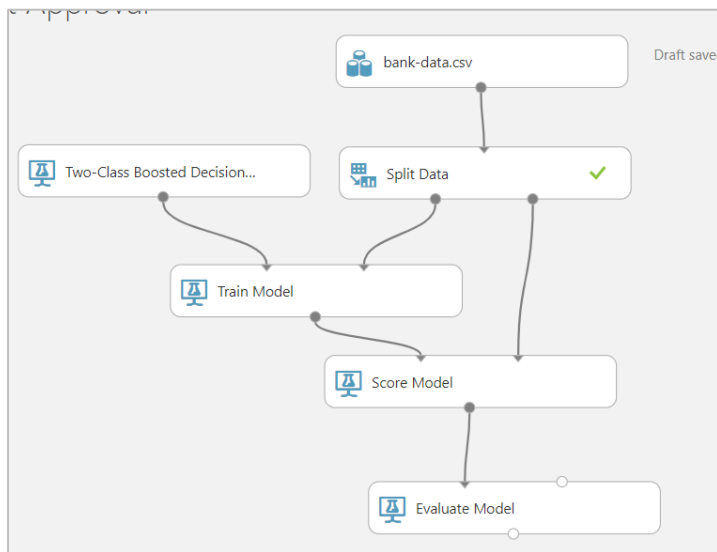
Credit Approval > Score Model > Scored dataset

income	married	children	car	save_act	current_act	mortgage	loan_approved	Scored Labels	Scored Probabilities
11536.2	YES	2	NO	NO	YES	YES	NO	NO	0.284724
44658.6	NO	2	YES	YES	YES	NO	YES	YES	0.94643
32762.5	NO	0	NO	YES	YES	YES	NO	NO	0.149262
16403.8	YES	0	YES	YES	NO	NO	NO	NO	0.157662
21184.7	YES	1	YES	YES	YES	NO	YES	YES	0.940262
49917.3	YES	0	YES	YES	YES	YES	NO	NO	0.056915
21623.8	YES	0	NO	YES	NO	YES	NO	NO	0.057399
16625.9	YES	1	NO	NO	YES	NO	YES	YES	0.940262
14014.5	YES	3	NO	YES	NO	NO	NO	NO	0.0616
20409.3	YES	2	NO	YES	NO	NO	YES	NO	0.0616

Terlihat ada kolom baru "Scored Labels" yang merupakan kolom prediksi dari model untuk kolom "Loan Approved". Pada gambar telah dibuat kotak perbandingan terhadap isi kedua kolom tersebut, terlihat pada gambar ada satu baris data dimana nilai kolom prediksi "Scored labels" dan aslinya, kolom "Loan Approved" berbeda nilai. Tutup Dialog tersebut.

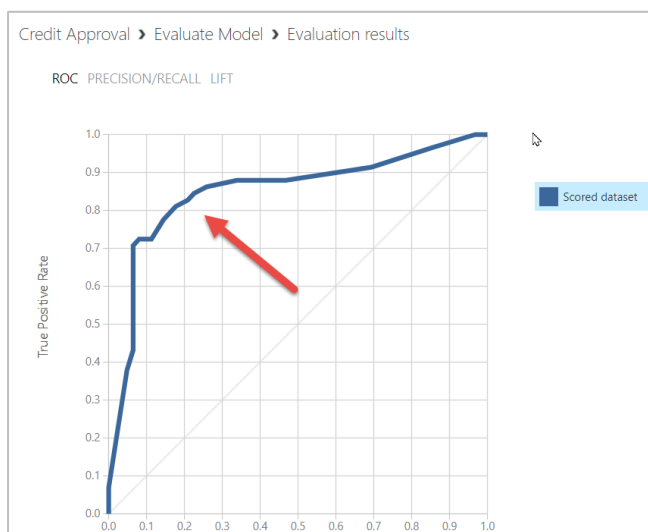
Melakukan Evaluasi Model

Sekarang kita perlu Masukkan module "Score Model" dan susun workspace sehingga seperti berikut ini.



Tidak ada yang perlu dikonfigurasi lebih lanjut, simpan dan jalankan experiment kita ini. Setelah selesai, klik output port dari module "Score Model", dan pilih Visualize.

Akan muncul ROC Curve dan Confusion Matrix seperti gambar berikut. Karena di atas kiri diagonal, maka model masih bisa dianggap bagus.



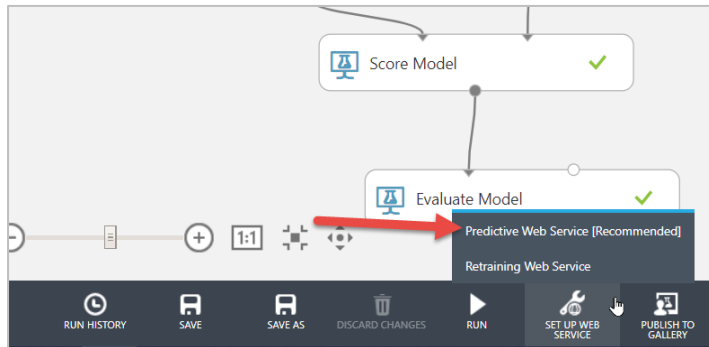
False Positive Rate			
True Positive	False Negative	Accuracy	Precision
42	16	0.825	0.894
False Positive	True Negative	Recall	F1 Score
5	57	0.724	0.800
Positive Label	Negative Label		
YES	NO		

Tutup dialog tersebut.

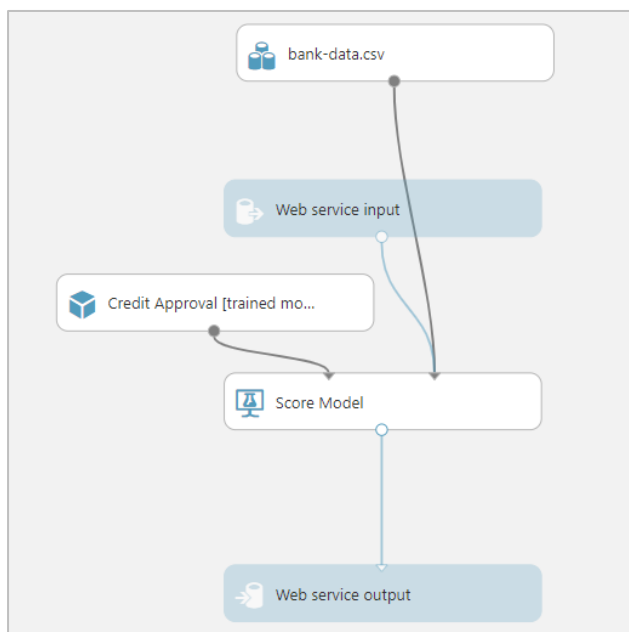
Membuat Web Service

Agar model tersebut dapat digunakan, harus ada API (Application Programming Interface) dari model tersebut ke aplikasi. Dan web service adalah mekanisme yang paling populer.

Azure Machine Learning menyediakan fitur tersebut yang dapat digunakan dengan mudah. Caranya adalah klik "Set Up Web Services".



Dalam beberapa saat akan muncul tab baru bernama "Predictive Experiment" dengan tampilan susunan modules yang agak berbeda, dengan tambahan module "Web Services Input" – yang akan menggantikan "bank-data.csv" – dan "Web Services Output".



Deploy Web Service

Sekarang kita sudah siap untuk deploy Web Services tersebut untuk dapat digunakan lebih lanjut, jalankan predictive experiment ini dengan Run sekali. Setelah selesai, klik tombol "deploy web service".

Dalam beberapa saat akan muncul halaman web services dengan detail API Key dan link untuk menggunakan jenis servicenya, apakah satu persatu (Request/Response) atau dalam kelompok besar data (Batch). Namun untuk sesi ini, kita gunakan form test yang langsung disediakan oleh Azure. Klik link Test (bukan tombol biru Test) yang ada di Request / Response.

credit approval [predictive exp.]

DASHBOARD CONFIGURATION

General **New Web Services Experience** *preview*

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

Sfpd+P/GNx/Er7qjibYEYyo/4kG3loEY7zZL9ODW+j6ZWc0o0fGWUbPFnDjqy7t/CmgX+QPbzNziulFAQxB5sQ==

Default Endpoint

API HELP PAGE	TEST	APPS
REQUEST/RESPONSE	Test Test <i>preview</i>	Excel 2013 or later Excel 2010 or earlier workbook
BATCH EXECUTION	Test <i>preview</i>	Excel 2013 or later workbook

Pada saat form sudah muncul masukkan data berikut.

Quickstart	Dashboard	Batch Request Log
id	<input type="text" value="ID001"/>	
age	<input type="text" value="33"/>	
sex	<input type="text" value="MALE"/>	
region	<input type="text" value="INNER_CITY"/>	
income	<input type="text" value="100000"/>	
married	<input type="text" value="NO"/>	
children	<input type="text" value="3"/>	
car	<input type="text" value="NO"/>	
save_act	<input type="text" value="NO"/>	
current_act	<input type="text" value="NO"/>	
mortgage	<input type="text" value="NO"/>	
an_approved	<input type="text"/>	

Kemudian klik tombol "Test Request-Response". Dalam waktu singkat akan muncul hasil di panel kanan.

married	NO
children	3
car	NO
save_act	NO
current_act	NO
mortgage	NO
loan_approved	
Scored Labels	YES
Scored Probabilities	0.87503719329834

Terlihat hasilnya adalah "YES", yang artinya dengan data aplikasi yang kita masukkan maka kreditnya langsung disetujui.

Cobalah test dengan berbagai data yang lain.

Menggunakan Web Services pada Excel

Untuk user yang terbiasa dengan Excel, maka bisa menggunakan service ini dengan mudah juga. User tinggal masukan data2nya yang mungkin sudah ada sebelumnya untuk dianalisa segera secara otomatis.

Kembali ke halaman Web Services, download file Excel untuk tipe API Request/Response.

credit approval [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) preview

Published experiment
[View snapshot](#) [View latest](#)

Description
 No description provided for this web service.

API key
 Sfpd+P/GNx/Er7qjibYEYyo/4kG3loEY7zZL9ODW+j6ZWco0fGWUbpFnDjqy7t/CmgX+QPbzNziulFAQxB5sQ==

Default Endpoint


API HELP PAGE	TEST	APPS
REQUEST/RESPONSE	Test Test <small>preview</small>	Excel 2013 or later Excel 2010 or earlier workbook
BATCH EXECUTION	Test <small>preview</small>	Excel 2013 or later workbook

Buka file tersebut dengan aplikasi Excel Anda. Anda harus mengaktifkan makro agar file ini bisa digunakan. Masukkan data berikut pada bagian header berwarna biru.

PARAMETERS											
id	age	sex	region	income	married	children	car	save_act	current_act	mortgage	loan_approved
ID0001	33	MALE	INNER_CITY	100000	NO	3	NO	NO	NO	NO	0

Hasil prediksi akan didapatkan sebagai berikut dimana dimunculkan nilai YES.

PREDICTED VALUES													
loan_id	age	sex	region	income	married	children	car	save_act	current_act	mortgage	loan_approved	ScoredLabels	ScoredProbabilities
0 ID	33	MALE	INNER_CITY	100000	NO	3	NO	NO	NO	NO	0	YES	0.875037193



Kesimpulan

Azure Machine Learning Studio adalah tool yang sangat lengkap dan memudahkan kita untuk mengembangkan model machine learning dan mengujinya.

Dengan waktu singkat kita dapat menarik data, melakukan transformasi data seperti pembagian data, membuat model dan mengevaluasinya.

Tutorial singkat ini menunjukkan semua hal tersebut dengan case persetujuan kredit. Pengujian bahkan dilakukan dengan tool yang sangat kita kenal, yaitu Excel.